Towards a New Ontology of Polling Inaccuracy: The Benefits of Conceiving of Elections as Heterogenous Phenomena for the Study of Pre-election Polling Error

Jack Tudor

Submitted to Swansea University in fulfilment of the requirements for the degree of

Doctor of Philosophy

Swansea University

2023

Abstract

A puzzle exists at the heart of pre-election polling. Despite continual methodological improvement and repeated attempts to identify and correct issues laid bare by misprediction, average polling accuracy has not notably improved since the conclusion of the Second World War. In this thesis, I contend that this is the result of a poll-level focus within the study of polling error that is both incommensurate with its evolution over time and the nature of the elections that polls seek to predict. I hold that differences between elections stand as a plausible source of polling error and situate them within a novel four-level model of sources of polling error. By establishing the heterogenous nature of elections as phenomena and its expected impact on polling error, I propose a new election-level ontology through which the inaccuracy of polls can be understood. I test the empirical validity of this new ontology by using a novel multi-level model to analyse error across the most expansive polling dataset assembled to date, encompassing 11,832 in-campaign polls conducted in 497 elections across 83 countries, finding that membership within different elections meaningfully impacts polling error variation. With the empirical validity of my proposed ontology established, I engage in an exploratory analysis of its benefits, finding electoral characteristics to be useful in the prediction of polling error. Ultimately, I conclude that the adoption of a new, multi-level ontology of polling error centred on the importance of electoral heterogeneity not only offers a more comprehensive theoretical account of its sources than current understandings, but is also more specifically tailored to the reality of pre-election polling than existing alternatives. I also contend that it offers pronounced practical benefits, illuminating those circumstances in which polling error is likely to vary.

Declarations and Statements

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed: Jack Tudor (candidate)

Date: 30/08/2023

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in footnotes. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed: Jack Tudor (candidate)

Date: 30/08/2023

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed: Jack Tudor (candidate)

Date: 30/08/2023

Table of Contents

Abstract ... i

Declarations and Statements ... ii

Dedication ... iii

Acknowledgements ... iv

List of Tables ... v

List of Figures ... viii

List of Acronyms ... x

Chapter 1 – The More Things Change, The More They Stay the Same: The Puzzle at the Heart of Pre-election Polling ... 1

Chapter 2 – A Rising Challenge to the Dominant Paradigm: Reviewing the Presence of Election-level Explanations in Assessments of Polling Error ... 18

Chapter 3 – Beyond a Poll-level Understanding of Error: Conceiving of Elections as Sources of Polling Inaccuracy ... 67

3.1: Identifying and Bounding the Predictive Utility of Pre-election Polls ... 70

3.2: Marginality or Missing the Mark? Conceptualising Polling Inaccuracy ... 84

3.3: The Dominant Poll-level Understanding of Polling Inaccuracy ... 99

3.4: Compositional Heterogeneity: Elections as Sources of Polling Inaccuracy ... 123

Chapter 4 – Assessing the Importance of Election-level Differences for Polling Error: Decomposing Error Variance Using a Novel Multi-level Model ... 158

4.1: Data Composition, Hierarchically Nested Error, and Multi-level Modelling ... 160

4.2: Approaches to Decomposing Multi-level Variance Components ... 189

4.3: Operationalising Polling Error ... 203

4.4: Descriptive Analysis: Establishing the Basis for Election-level Investigation ... 220

4.5: Measuring the Effect of Election-level Differences on Polling Error Variance ... 233

Chapter 5 – Which Differences Make a Difference? Identifying Plausible Election-level Predictors of Polling Error Variance 256

5.1: Identifying Likely Election-level Predictors of Polling Error Variance 257

5.2: The Expected Impact of Election-level Differences in Interaction 305

5.3: Controlling for Predictor Variables from Additional Grouping Levels 318

Chapter 6 – From Predictors to Predictions: Assessing the Ability of Election-level Differences to Predict Variation in Polling Error 327

6.1: Modelling Procedures, Test Statistics, and Dealing with Missing Data 330

6.2: Prediction Model Outputs 351

6.3: Exploring Different Modelling Approaches and Emergent Themes 422

Chapter 7 – A Foundation to Build Upon: Towards Ontological and Epistemological Re-orientation in the Study of Pre-election Polling Error 443

Appendix A – Additional Calculative Techniques and Robustness Checks to Support the Inferential Analysis in Chapter 4 457

A1: Alternative Approaches to Partitioning Group-level Variance 457

A2: Additional ICC Estimates for Two-level Models 459

A3: Additional ICC Estimates for Three-level Models 464

A4: Additional ICC Estimates for Four-level Models 470

A5: The Variation of Election-level ICC Values Across Countries 478

Appendix B – Additional Figures and Predictive Models to Support the Inferential Analysis in Chapter 5 479

B1: Additional Additive Prediction Models for Measures of Distributive Polling Error 479

B2: Additional Additive Prediction Models for Measures of Bounded Polling Error 485

B3: Additional Additive Prediction Models for Measures of Polling Bias 495

B4: Additional Additive Prediction Models for Measures of Substantive Polling Error 510

B5: Additional Interactive Prediction Models for All Error Measures 515

B6: The Robustness of Additional Interactive Prediction Models to Controls 536

Bibliography 564

Dedication

This thesis is dedicated to my late grandfather, William Conrad Mills, who would have been proud to see it completed.

Acknowledgements

This thesis would never have existed were it not for the support of Dr. Matthew Wall who has tirelessly championed it since its inception. Equally, it would never have been finished without the endless generosity of Dr. Kevin Fahey. I owe them both a profound debt that I can never truly repay. Instead, I can only promise to pay their kindness forward. I would also like to thank Prof. Robert Ford for his thorough and extremely useful feedback on this thesis. His insights have served to strengthen it considerably and I am extraordinarily grateful for the care and attention he paid to its development. Similarly, I am grateful to Dr. Bryn Willcock for his thoughtful and constructive feedback on a past iteration of this work which has served to improve its rigour and manner of presentation.

List of Tables

Table 1 - Pearson's correlations for continuous measures of polling error 218

Table 2 - Fligner-Killeen tests of variance homogeneity 222

Table 3 – Kruskal-Wallis test 224

Table 4 – Chi-squared tests of independence 225

Table 5 – Two-level ICCs for continuous error measures using MLE and RMLE 234

Table 6 – Two-level ICCS for continuous error measures using half student-t priors 236

Table 7 – Two-level ICCs for binary error measures using Laplace approximation 238

Table 8 – Two-level ICCs for binary error measures using half student-t priors 240

Table 9 – Three-level ICCs for continuous error measures using MLE and RMLE 242

Table 10 – Three-level ICCs for continuous error measures using half student-t priors 243

Table 11 – Three-level ICCs for binary error measures using Laplace approximation 245

Table 12 – Three-level ICCs for binary error measures using half student-t priors 246

Table 13 – Four-level ICCs for continuous error measures using MLE and RMLE 247

Table 14 – Four-level ICCs for continuous error measures using half student-t priors 249

Table 15 – Four-level ICCs for binary error measures using Laplace approximation 251

Table 16 – Four-level ICCs for binary error measures using half student-t priors 252

Table 17 – The number of first and second round elections within my dataset 270

Table 18 – The number of legislative and presidential elections within my dataset 272

Table 19 – The number of scheduled and snap elections within my dataset 293

Table 20 – Summary statistics for control variables 320

Table 21 – Comparison of partisanship and turnout across countries 347

Table 22 – Comparison of ENEP and late decision-making across countries 348

Table 23 – Comparison of partisanship and late decision-making 349

Table 24 – Additive prediction model outputs for MAE across whole dataset 353

Table 25 – Additive prediction model outputs for MAE across partisanship subset 355

Table 26 – Additive prediction model outputs for MAE across late decision-making subset 357

Table 27 – Additive prediction model outputs for MAE across ideological positioning subset 359

Table 28 – Additive prediction model outputs for MAE across reduced subset of data....... 361

Table 29 – Average per-variable RMSE reductions across additive, continuous prediction models.. 365

Table 30 – Average per-variable change in correct classification across additive, binary prediction models.. 372

Table 31 – Average prediction improvements across all additive models........................... 374

Table 32 – Interactive prediction model outputs for MAE across whole dataset 377

Table 33 – Interactive prediction model outputs for MAE across partisanship subset......... 378

Table 34 – Interactive prediction model outputs for MAE across late decision-making subset 380

Table 35 – Average per-variable RMSE reductions across interactive, continuous prediction models relative to additive models 381

Table 36 – Average per-variable change in correct classification across interactive, binary prediction models relative to additive models 387

Table 37 – Average prediction improvements across all interactive models........................ 389

Table 38 – Prediction model outputs for MAE in the presence of controls across whole dataset 397

Table 39 – Prediction model outputs for MAE in the presence of controls across partisanship subset 400

Table 40 – Prediction model outputs for MAE in the presence of controls across late decision-making subset 402

Table 41 – Unpacking the effect of overfitting in models tasked with predicting MAE across the subset of data containing late decision-making and controls 403

Table 42 – Average RMSE reduction associated with election-level variables in models containing controls across all continuous measures of polling error 405

Table 43 – Average classification improvements associated with election-level variables in models containing controls across all binary measures of polling error 407

Table 44 – The predictive benefit of election-level variables across all measures of polling error in the presence of controls 408

Table 45 – The raw reduction in RMSE associated with election-level variables across alternate modelling specifications 425

Table 46 – The average percentage point improvements in RMSE afforded by alternative models over linear specifications 427

Table 47 – The percentage point improvement to classification accuracy associated with election-level variables in the presence of controls across alternative model specifications 430

Table 48 – The average percentage point improvements in classification accuracy afforded by alternative models over linear specifications 431

Table 49 – Variation in the utility of election-level variables as predictors of continuous measures of polling error across country-level groupings 436

Table 50 – Variation in the importance of election-level variables as predictors of variance in binary measures of polling error across country-level groupings 439

List of Figures

Figure 1 – Chicago Tribune 'Dewey Defeats Truman' headline 25

Figure 2 – Newsweek Magazine 'Madam President' special issue 53

Figure 3 – The total survey error framework 101

Figure 4 – A visual representation of random error 103

Figure 5 – A visual representation of systematic error 107

Figure 6 – The average MAE exhibited by polls over time 116

Figure 7 – Four-level model of sources of polling error 174

Figure 8 – Two-level model of polling error 178

Figure 9 – Tudor and Wall's three-level model of polling error 181

Figure 10 – FiveThirytEight's three-level model of national polling error 184

Figure 11 – Distributions of continuous measures of polling error 221

Figure 12 – Average distributive polling error over time 226

Figure 13 – Average bounded polling error over time 227

Figure 14 – Average polling bias over time 228

Figure 15 – The proportion of positive SBP and LVRC values across all elections 229

Figure 16 – World map visualising average MAE by country 231

Figure 17 – Average MAE and APB across polling organisations 232

Figure 18 – The margin of victory across elections 261

Figure 19 – ENEP values across elections 264

Figure 20 – ENEP change between subsequent elections 267

Figure 21 – Voter turnout across elections 275

Figure 22 – Changes in voter turnout between subsequent elections 279

Figure 23 – The strength of partisan loyalty across elections...281

Figure 24 – The extent of late decision-making across elections 285

Figure 25 – The standard deviation of the ideological position of parties and candidate between elections...298

Figure 26 – The number of registered voters across elections..300

Figure 27 – Pearson's correlations between election-level variables....................................303

Figure 28 – Visualisation of the four-level model of sources of polling error 319

Figure 29 – Illustration of repeated 10-fold cross validation..334

Figure 30 – The four-level structure of sources of polling error emphasising the nested relationship between elections and countries..434

List of Acronyms

ABI 1	average bounded inaccuracy (first operationalisation)
ABI 2	average bounded inaccuracy (second operationalisation)
AGHQ	adaptive Gauss-Hermite quadrature
ANOVA	analysis of variance
APB	average party bias
CCA	complete case analysis
CSES	Comparative Study of Electoral Systems
DIM	difference in margin
ENEP	effective number of elective parties
GDP	gross domestic product
GINI	Gini coefficient
GLM	generalised linear model
ICC	intra-class correlation coefficient
LASSO	least absolute shrinkage and selection operator
LPB	leading party bias
LVRC	largest vote share recipient correct
MAE	mean absolute error
MAR	missing at random
MCAR	missing completely at random
MCMC	Markov chain Monte Carlo
MLE	maximum likelihood estimation
MNAR	missing not at random
MI	multiple imputation
OLS	ordinary least squares
RMLE	restricted maximum likelihood estimation
RMSE	root mean square error
SBP	significantly biased poll

Chapter 1 – The More Things Change, The More They Stay the Same: The Puzzle at the Heart of Pre-election Polling

> *"Sometimes a normal problem, one that ought [ostensibly] to be solvable by known rules and procedures, resists the reiterated onslaught of the ablest members of the group within whose competence it falls".*[1]
>
> \- Thomas Kuhn (1962)

Predicting the future is widely recognised to be a difficult task. Indeed, variants of the phrase 'it is difficult to make predictions, especially about the future' have appeared in works across a wide range of academic disciplines.[2] Though the difficulty of prediction is a multi-disciplinary concern, the social sciences have a particularly troubled relationship with the future. From difficulties in conflict and intelligence prediction,[3] to infamous failures of economic prediction,[4] the future has proven difficult to predict across a range of social scientific fields. While predictive accuracy is improving in certain areas,[5] consistently accurate predictions remain elusive to this day.[6]

Like other social scientific disciplines, political science is no stranger to the difficulties of prediction. Nowhere is this made more apparent than by the history of pre-election polling. Though the purpose of pre-election polls and their manner of analysis varies depending on the point in time at which they are conducted, when they are conducted in reasonable proximity to

---

[1] Thomas Kuhn, *The Structure of Scientific Revolutions*', (Chicago: The University of Chicago Press, 1962), p. 5.
[2] H. H. Wandall, 'Medical Education in Denmark', *Academic Medicine*, 36.9 (1961), 1059 – 1078 (p. 1069); Mark Kac, 'Some Reflections of a Mathematician on the Nature and Role of Statistics', *Advances in Applied Probability*, 7 (1975), 5 – 11 (p. 5); George J. Annas, 'Precatory Prediction and Mindless Mimicry: The Case of Mary O'Connor', *The Hastings Center Report*, 18.6 (1988), 31 – 33 (p. 33).
[3] Robert Jervis, *Why Intelligence Fails: Lessons from the Iranian Revolution and the Iraq War*, (New York: Cornell University Press, 2011), pp. 1 – 5; Kai Jager, 'Not a New Gold Standard: Even Big Data Cannot Predict the Future', *Critical Review*, 28.3 (2016), 335 – 355 (p. 335).
[4] David Colander and others, 'The Financial Crisis and the Systemic Failure of the Economics Profession', *Critical Review*, 21.2 (2009), 249 – 267 (p. 249).
[5] Thomas Chadefaux, 'Conflict Forecasting and its Limits', *Data Science*, 1.1 (2017), 7 – 17 (p. 7).
[6] Spyros Makridakis, Rob J. Hyndman, and Fotios Petropoulos, 'Forecasting in Social Settings: The State of the Art', *International Journal of Forecasting,* 36 (2020), 15 – 28 (p. 26).

an election, they can be understood to provide predictions of the likely distribution of vote shares on election day based on the voting intention of a sample of surveyed individuals. These predicted vote share distributions have not always aligned with reality on election day, leading to misprediction. Indeed, in recent years, pre-election polls have failed to accurately predict the hung parliament in the 2017 UK general election,[7] the success of Donald Trump in the 2016 US presidential election,[8] and the substantive outcome of the 2015 UK general election.[9] Even in cases where they correctly identify the substantive outcome of elections, polls have been prone to high levels of error. This was apparent in both the 2020 US presidential election,[10] as well as the second round of the 2017 French presidential contest.[11]

Over the course of its history, the study of pre-election polling has principally been driven by instances of misprediction.[12] These instances of misprediction have spurred investigations into sources of polling error that have tended to identify poll-level issues as the causes of predictive failure.[13] That is, they identify deficiencies in the methods undergirding pre-election polls as the drivers of their erroneous predictions. On the basis of these perceived deficiencies, assessments of polling failures largely recommend methodological revisions to improve predictive accuracy moving forward.[14]

[7] Bertram Düring and Oliver Wright, ‘On a Kinetic Opinion Formation Model for Pre-election Polling’, *Philosophical Transactions of the Royal Society A*, 380.2224 (2022), 1 – 20 (p. 1).
[8] Ramin Skibba, ‘Why the Polls Missed Trump’, *Nature*, 539.17 (2016), 339 (p. 339).
[9] Jonathan Mellon and Christopher Prosser, ‘Missing Nonvoters and Misweighted Samples: Explaining the 2015 Great British Polling Miss’, *Public Opinion Quarterly*, 81.3 (2017), 661 – 687 (p. 661).
[10] Costas Panagopoulos, ‘Polls and Elections: Accuracy and Bias in the 2020 U.S. General Election Polls’, *Presidential Studies Quarterly*, 51.1 (2021), 214 – 227 (p. 214).
[11] Jack Tudor, *Are Certain Elections More Predictable than Others? A Series of Bivariate Analyses of the Impact of Electoral Characteristics Upon the Predictability of British National Elections* (2017), <https://ssrn.com/abstract=3601958> [accessed 24/08/2023].
[12] Ibid.
[13] Jack Tudor and Matthew Wall, ‘A Moving Target? An Analysis of Electoral Context and Variation in Polling Accuracy Across Post-war British General Elections’, *Journal of Elections, Public Opinion and Parties*, (2021), 1 – 24 (p. 3).
[14] Ibid.

This poll-level approach to understanding and resolving polling error finds its roots in the 1936 US presidential election with the infamous predictive failure of *The Literary Digest*.[15] While *The Literary Digest* failed to predict the election using polls predicated on large-scale convenience sampling, competing 'scientific' polls conducted by Gallup, Crossley, and Roper successfully predicted the outcome using polls based on representative quota sampling.[16] Successful electoral prediction, it seemed, was attainable through adjustments to polling methodology.

Since the success of Gallup and company in 1936, pre-election polls have correctly predicted the outcome of a range of elections.[17] These successes have often been facilitated by methodological adjustments or innovations at the poll level,[18] ostensibly confirming earlier successes and solidifying the attendant poll-level approach to understanding polling error. As such, poll-level alterations have proven capable of improving the accuracy of pre-election polls. However, under the poll-level approach to understanding polling inaccuracy, the average predictive accuracy of pre-election polls has not meaningfully improved since 1945.[19] This largely static level of accuracy has persisted despite an iterative process of methodological revision aimed at reducing polling error driven by responses to a number of notable

[15] The Literary Digest, 'Landon, 1,293,669; Roosevelt, 972,897: Final Returns in the Digest's Poll of Ten Million Voters', *The Literary Digest*, 31 October 1936, Topics of the Day, pp. 5-6

[16] Daniel Katz and Hadley Cantril, 'Public Opinion Polls', *Sociometry*, 1 (1937), 155-179 (p. 158).

[17] Ivor Crewe, 'The Opinion Polls: The Election They Got (Almost) Right', *Parliamentary Affairs*, 58.4 (2005), 684 – 698 (p. 684); Ivor Crewe, 'The Opinion Polls: Confidence Restored?', *Parliamentary Affairs*, 50.4 (1997), 569 – 586 (p. 569); Leo Bogart, 'Politics, Polls, and Poltergeists', *Society*, 35.4 (1998), 8 – 16 (p. 8); Robert Worcester, 'Political Polling: 95% Expertise and 5% Luck', *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 159.1 (1996), 5 – 20 (p. 5); Joe Twyman, 'Getting it Right: YouGov and Online Survey Research in Britain', *Journal of Elections, Public Opinion and Parties*, 18.4 (2008). 343 – 354 (p. 343); Robert Northcott, 'Opinion Polling and Election Predictions', *Philosophy of Science*, 82.5 (2015), 1260 – 1271 (p. 1261); Humphrey Taylor and others, 'The Record of Internet-based Opinion Polls in Predicting the Results of 72 Races in the November 2000 US Elections', *International Journal of Market Research*, 43.2 (2001), 127 – 135 (p. 128).

[18] George Gallup, 'The Gallup Poll and the 1950 Election', *The Public Opinion Quarterly*, 15.1 (1951), 16 – 22 (p. 17); Twyman, p. 343; Northcott, p. 1261; Taylor and others, p. 133.

[19] Jennings and Wlezien, 'Election Polling Errors Across Time and Space', p. 280.

mispredictions.[20] This exposes the puzzle at the heart of pre-election polling: despite continual innovation and change at the poll level in response to identified sources of inaccuracy, average polling error remains largely unchanged.

The lack of meaningful improvement in polling accuracy over time under the poll-level approach to understanding polling error speaks to its insufficiency in isolation, but does not necessarily indicate that it has proven ineffective. Indeed, the poll-level improvements implemented since 1945 have not occurred in a vacuum but, rather, have occurred across a period of time in which a range of other forces have conspired to make accurately predicting elections through the use of polling more difficult, especially when accuracy is considered as an average across cases. Issues in established democracies such as declining response rates,[21] reductions in the strength of partisan loyalty amongst voters,[22] and increases in electoral volatility have all conspired to make polling more difficult over time.[23] In addition to this, polling has expanded to encompass emerging democracies over time. Not only does polling in these new countries present a diverse range of challenges that may serve to drive prediction

[20] Frederick Mosteller and others, *The Pre-election Polls of 1948: Report to the Committee on Analysis of Pre-election Polls and Forecasts*, (New York: Social Science Research Council, 1949), pp. 174 – 289; Patrick Sturgis and others, 'An Assessment of the Causes of the Errors in the 2015 UK General Election Opinion Polls', *Journal of the Royal Statistical Society: Statistics in Society Series A*, 181.3 (2018), 757 – 781, (pp. 757 – 781); Roger Jowell and others, 'The 1992 British Election: The Failure of the Polls', *Public Opinion Quarterly*, 57.2 (1993), 238 – 263 (pp. 238 – 262); Rami Zeedan, 'The 2016 US Presidential Elections: What Went Wrong in Pre-election Polls? Demographics Help to Explain', *J – Multidisciplinary Scientific Journal*, 2.1 (2019), 84 – 101 (pp. 84 – 101); Mellon and Prosser, pp. 661 – 687; Mark Pickup and others, 'Why Did the Polls Overestimate Liberal Democrat Support? Sources of Polling Error in the 2010 British General Election', *Journal of Elections, Public Opinion and Parties*, 21.2 (2011), 179 – 209 (pp. 179 – 209); Claire Durand and André Blais, 'Quebec 2018: A Failure of the Polls?', Canadian Journal of Political Science, 53.1 (2020), 133 – 150 (p. 133); Claire Durand, 'The Polls of the 2007 French Presidential Campaign: Were Lessons Learned from the 2002 Catastrophe?', International Journal of Public Opinion Research, 20.3 (2008), 275 – 298 (p. 296); Claire Durand and André Blais, 'Why Did the Polls Go Wrong in the 1998 Quebec Election? The Answer from Post-election Polls', Bulletin de Méthodologie Sociologique, 62 (1999), 43 – 47 (pp. 43 – 47); Claire Durand and André Blais, 'Why Did the Polls Go Wrong in the 1998 Quebec Election? The Answer from Post-election Polls', Bulletin de Méthodologie Sociologique, 62 (1999), 43 – 47 (pp. 43 – 47); Colin Rallings and Michael Thrasher, 'Opinion Polling and the Aftermath of the 1992 General Election', Contemporary British History, 7.1 (1993), 187 – 197 (p. 190).
[21] Michael W. Traugott, 'Can We Trust the Polls? It All Depends', *The Brookings Review*, 21.3 (2003), 8-11 (p. 9).
[22] Stephen C. Craig, 'The Decline of Partisanship in the United States: A Re-examination of the Neutrality Hypothesis', *Political Behavior*, 7 (1985), 57-78 (p. 57).
[23] Mogens N. Pederson, 'The Dynamics of European Party Systems: Changing Patterns of Electoral Volatility', *European Journal of Political Research*, 7 (1979), 1-26 (p. 1).

error, including issues of local cooperation with polling organisations,[24] developmental and technological barriers to contacting representative samples of individuals,[25] and difficulties in eliciting voting intentions in societies in which the expression of political opinions was previously the focus of repression,[26] but their incorporation alongside established democracies over time stands to increase average polling error.

When taken in tandem, the issues arising over time in both established and emerging democracies lend themselves to the expectation of an increase in polling error over my studied timeframe in line with the emergence of difficulties and the increasing heterogeneity of the cases in which pre-election polling is undertaken. That this has not occurred and, instead, polling error has remained largely static speaks to a Red Queen problem.[27] Specifically, that it has been necessary for the polling industry to implement a near-continual set of poll-level revisions and improvements over time simply to keep the level of polling accuracy stable. To this end, the poll-level approach to identifying and remedying issues of polling accuracy can be said to have been successful. However, that the process of engaging in poll-level improvements is only allowing the industry to maintain the status quo when it comes to polling accuracy suggests that sources of error exist beyond the confines of polls that affect their ability to render accurate predictions. This suggestion animates the research contained within this thesis.

[24] Mitchell A. Seligson, 'Improving the Quality of Survey Research in Democratizing Countries', *PS: Political Science*, 38.1 (2005), 51-56 (p. 51).
[25] Robert Mattes, 'Public Opinion Research in Emerging Democracies' in *The SAGE Handbook of Public Opinion Research*, ed. by Michael W. Traugott and Wolfgang Donsbach, (London: SAGE Publications, 2007), 113-122 (pp. 116-117).
[26] Kwasi Ansu-Kyeremeh, 'The Challenges of Surveying Public Opinion in an Emerging Democracy', *International Journal of Public Opinion Research*, 11.1 (1999), 59-74 (p. 61).
[27] Alex Coram, 'The Red Queen and the Dynamics of Resource Spending in Party Competition', *British Journal of Political Science*, 40.2 (2010), 469-475 (p. 470).

In this thesis, I contend that differences between elections exist as a plausible source of polling error that rest within a broader four-level structure of sources of error. Specifically, I argue that differences in the characteristics possessed by elections not only lead to environments that are variously conducive to polling error, but also have the potential to bear upon the projection mechanisms used by pre-election polls to render predictions and, through doing so, possess the ability to make polling error more or less likely. As they possess the ability to create environments that are more conducive to polling error, I also contend that electoral characteristics ought to be predictive of the degree to which polling error varies. Through the use of a novel, four-level multi-level model alongside an array of (non-)parametric prediction models, I demonstrate that these contentions are borne out empirically and are robust to a range of model specifications. On the basis of this, I call for ontological and epistemological re-orientation in the study of pre-election polling error towards a recognition of the importance of electoral heterogeneity as a driver of error variance.

Thesis Structure and Research Questions

To unpack my contentions and substantiate them empirically, this thesis progresses through five substantive chapters and a conclusion that centre on three research questions. In chapter two, directly following this introduction, I provide the foundation for my election-level investigation of polling error by addressing the academic literature surrounding pre-election polling error from the birth of 'scientific' pre-election polling in 1936, through to works produced in 2022. In addressing the literature, I identify that while the poll-level approach to understanding polling error has been dominant throughout its history, and remains preponderant to this day, an alternative, election-level understanding of polling error has existed within scholarship since the 1930s. Throughout the chapter, I chart the rise of this election-level approach to polling error through five distinct waves of literature, each characterised by a greater focus on election-level sources of error than the last.

While I note its rising status, I also recognise that the election-level approach to understanding polling error remains underdeveloped, with two significant gaps existing at its core. The first of these gaps is theoretical in nature. Though it has existed since the 1930s, the election-level approach to understanding polling error presently lacks a comprehensive framework outlining both the expectation of electoral heterogeneity and its likely impact on polling error. Current understandings of electoral heterogeneity rest on assertions of compositional variance that remain opaque and underexplored,[28] with explorations of the importance of election-level differences universally failing to establish why these differences occur, or even why their presence can be expected.[29] To fill this gap in the literature, I provide the first comprehensive approach to conceiving of elections as heterogenous phenomena that can be expected to bear on the error exhibited by polls.

The second gap in the literature is empirical in nature. To date, no studies have empirically assessed the benefit of adopting an election-level approach to understanding polling error. While a limited number of prospective benefits have been proposed,[30] and calls for the adoption of an election-level understanding of polling error have been made,[31] no work has been conducted to identify the tangible benefits of adopting such an approach. To remedy this, within this thesis, I demonstrate that adopting an election-level understanding of polling error is beneficial for identifying those circumstances in which polling error is likely to vary. Due to this, I contend that it not only allows for a deeper understanding of why polling error varied

[28] Tudor and Wall, p. 5.
[29] Sohlberg and Branham, p. 11
[30] Ibid.
[31] Tudor and Wall, pp. 18 – 19.

between past elections but can, when built upon by future work, lead to election-level differences being used to inform and improve poll-based predictions.

In chapter three, I fill the theoretical gap in the literature by establishing the rationale for conceiving of polling error as a function of differences between elections. To achieve this, I address my first research question (RQ_1):

> $\mathbf{RQ_1}$**:** To what degree can variance in polling error be expected to be a function of differences between elections?

To rationalise the discussion of polling error, I begin by establishing the conditions under which pre-election polls can be understood to be predictive undertakings. I define pre-election polls as voting intention polls conducted shortly prior to election day, as distinct from post-election polls and more general public opinion surveys. I contend that only polls that adequately capture the constellation of factors at play on election day can reasonably be treated as predictive of electoral outcomes. To this end, for the purposes of understanding the prediction error they present, I confine my focus to in-campaign polls to ensure that they satisfy this criterion and capture meaningfully focused and activated voter sentiment. Ultimately, I situate those polls that can be considered predictive of elections at the intersection of three prominent approaches to understanding future outcomes: forecasts, predictions, and projections. From this, I contend that pre-election polls are best understood as future-orientated predictions that rest on a series of projections.

With the predictive nature of pre-election polls established, I move to conceptualise polling inaccuracy. I identify that, in attempting to predict electoral outcomes, pre-election polls provide three pieces of information: estimated vote share distributions, implied electoral outcomes, and estimates of the uncertainty surrounding their estimates. On the basis of these pieces of information, I derive three conceptualisations of polling error. The first of these is distributive error and centres on the difference between the estimated vote share distributions provided by polls and the distribution of votes on election day. The second is bounded error and concerns the degree to which election results fall outside of the margin of error surrounding polling estimates. The third focuses on the substantive accuracy of pre-election polling estimates. While the substantive implications of vote share distributions for government formation vary between electoral systems, rendering the creation of a universally tractable measure of substantive polling accuracy difficult,[32] I conceive of substantive error in terms of whether a poll correctly identifies the party or candidate in receipt of the largest share of the vote.

After establishing my conceptualisations of polling error, I explore the dominant poll-level understanding of how this error comes about. I identify that polling error is principally understood to be a function of random and systematic errors at the poll level arising from mechanisms and processes represented by the total survey error framework. Following this, I demonstrate that conceiving of polling error solely as a function of random and systematic errors at the poll level is insufficient in isolation by examining the largely static nature of polling error over time. Through exploring a series of possible explanations for the disconnect between continual methodological improvement at the poll level and static polling error over time, I conclude that conceiving of error in solely poll-level terms presents an incomplete picture of its determinants. I contend that, for the disconnect at the heart of polling error to be rationalised, sources of error beyond those identified at the poll level must be bearing upon their accuracy. Ultimately, I hold that the characteristics of the elections that polls strive to predict stand as plausible drivers of the prediction error they present.

[32] Tudor and Wall, p. 9.

In the final section of chapter three, I directly address my first research question and establish the degree to which polling error can be expected to be a function of electoral characteristics. I identify that the study of pre-election polling error is principally based on an epistemology that holds that knowledge of polling error is principally gained through unpacking the methods employed by polls in past instances of misprediction. Further to this, I identify that, ontologically, polling is often predicated on the implicit understanding of elections as homogenous phenomena. I hold that this ontology is incommensurate with the reality of elections, as they exist as heterogenous phenomena which differ compositionally between cases. To illustrate the heterogeneity of elections, I provide a series of examples outlining that, while the core nature of their characteristics remains consistent, the prominence and magnitude of these characteristics varies between contests. Importantly, I identify that the replication of the values taken by electoral characteristics across cases is unlikely, establishing the expectation of heterogeneity between cases.

With the compositional heterogeneity of elections established, I identify the mechanisms through which it can be expected to affect polling error and provide substantive plausibility for these expectations by demonstrating that past polling failures have regularly occurred alongside pronounced instances of electoral heterogeneity. On the basis of the expectations, I form my first hypothesis: membership within different elections will affect the degree to which polls exhibit error. I proceed to test this hypothesis in the following chapter.

In chapter four, I address the degree to which theoretical expectations concerning the impact of electoral heterogeneity on polling error variance can be empirically validated. To achieve this, I answer my second research questions (RQ_2):

> $\mathbf{RQ_2}$**:** To what extent can the expectation that variance in polling error exists as a function of differences between electoral characteristics be validated empirically?

To begin answering this question, I set out the nature and scope of the polling dataset against which my theoretical expectations will be tested. Ultimately, I identify that my dataset solely comprises in-campaign polls – that is, polls conducted within the official campaign period of my studied elections – as the voting intentions captured by these polls are elicited from suitably primed and electorally aware respondents such that they can reasonably be thought of as predictions of future voting behaviour. I also establish that the dataset collated for use in this thesis stands as the most geographically expansive set of polling data gathered in political science to date. I then move to describe the fundamentally multi-level nature of polling error, identifying that sources of error exist within four distinct and interconnected grouping levels. I unpack the advantages of adopting a four-level approach to understanding pre-election polling and its attendant sources of error in comparison to existing multi-level approaches that employ a diminished number of levels.

Given the multi-level nature of polling error, I contend that a multi-level approach to its analysis is necessary. To this end, I outline a multi-level modelling approach to assessing the effect of the four grouping levels within my data on the variance exhibited by polling error. I identify that in order to assess the aggregate effect of different grouping levels, the parameters of interest within multi-level models are the between- and within-group variance terms. While I recognise that a range of approaches exist to extracting and decomposing these variance terms, I establish the intra-class correlation coefficient (ICC) as the principal tool of multi-level variance decomposition used within this thesis. I note that approaches to estimating the ICC and the model parameters on which its estimation rests vary between model concerned with continuous and binary outcome variables. In unpacking the differences between these approaches, I outline the manner in which the ICC is calculated for both my continuous and binary multi-level models and identify maximum likelihood estimation (MLE) and Laplace approximation as the principal approaches to the estimation of the parameters on which these calculations rest.

With the approaches to ICC calculation and parameter estimation established, I outline the manner in which I operationalise polling error for use in models. Across my distributive, bounded, and substantive conceptualisations of polling error, I provide eight approaches to measurement. The five approaches taken to measuring distributive polling inaccuracy are taken from the literature, while the three remaining approaches to measuring bounded and substantive inaccuracy exist as novel measurement strategies designed for use within this thesis.

After outlining my approaches to measuring polling error, I provide descriptive analysis of my dataset to establish the statistical basis for the investigation of election-level polling error. I demonstrate that differences between election-level grouping exist as statistically significant drivers of polling error variance. I further illustrate the likely importance of election-level differences for polling accuracy by providing descriptive plots of the error associated with different elections over time, noting pronounced variability between cases. In addition to illuminating the importance of the election-level, I also recognise the need to control for pollster- and country-level differences by visualising the variation in polling error across cases.

Once my descriptive analysis is completed, I move to establish the importance of election-level differences for polling error and, in so doing, answer my second research question. By presenting the results from a series of two-, three-, and four-level multi-level models, I demonstrate that election-level differences consistently account for a significant proportion of polling error variation, even in the presence of controls. I find that, in isolation, election-level differences account for between 32% and 76% of polling error variance across all measures in two-level models and account for between 28% and 60% in the presence of pollster- and country-level control groups. From this, I not only accept my first hypothesis that membership within different elections affects the degree to which polls exhibits error, but also answer my second research question and provide the expectation that polling error exists as a function of electoral characteristics with empirical validation.

In chapters five and six, I build upon the findings that result from multi-level variance decomposition and isolate those election-level differences that can be expected to bear most closely on polling error variance. I hold that as differences between elections affect the degree to which polling error varies, they ought logically to be predictive of its occurrence. To assess this contention, I answer my third research question (RQ_3):

> $\mathbf{RQ_3}$**:** To what degree can differences between electoral characteristics aid in the prediction of polling error variance?

To answer this question, in chapter five I identify a series of individual differences between elections that can be expected to bear upon polling error variance and unpack the mechanisms through which they can be expected to do so across my differing conceptualisations of error. I recognise that differences between elections can also be expected to affect the degree to which polling error varies when considered in interaction with one another. From this, I isolate a series of two- and three-way interactions between election-level differences that can be expected to impact the degree to which polling error varies. On the basis of the expected impact of election-level differences on polling error variance, I establish my second hypothesis: that election-level variables will aid models in predicting polling error variance both additively and interactively, proving most useful in the case of substantive error and least useful in the case of bias.

Given the four-level nature of sources of polling error, I recognise the importance of controlling for the impact of difference housed within alternate grouping level to meaningfully test this hypothesis. To this end, I isolate a series of predictor variables housed within the poll-, pollster-

, and country-level groupings of my four-level model that can be expected to bear upon the degree to which polling error varies. These variables are included in later analysis as controls.

In chapter six, I provide a series of prediction models to empirically test my second hypothesis. Through the use of additive and interactive prediction models, I demonstrate that election-level variables are useful predictors of the variance exhibited by each of my measures of polling error. I establish that these findings are not only robust to the presence of controls, but also to alternative modelling specifications. From this, I find strong substantive support for my second hypothesis and answer my third research question.

Finally, in chapter seven, I conclude the thesis by bringing together the findings presented in the preceding chapters. I summarise these findings and relate them back to the research questions that rest at the core of this thesis. I follow this by outlining the theoretical and practical importance of the work presented in this thesis, as well as identifying avenues for future research that come about because of it.

The Contributions Made by This Thesis

In total, I provide six contributions of varying size to the study of pre-election polling error through the production of this thesis. These contributions are theoretical, methodological, and empirical in nature. My first theoretical contribution concerns providing the first comprehensive framework for understanding elections as heterogenous phenomena that can be expected to affect polling error. Though alternative accounts of anticipated electoral heterogeneity and its likely impact on polling error have been provided in past research,[33] their elaboration is limited. Present understandings of electoral heterogeneity rest on assertions of compositional variance that remain unelaborated and underexplored,[34] with calls for the

[33] Tudor and Wall, p. 5.
[34] Ibid.

investigation of the impact of election-level differences on polling error universally failing to establish why these differences occur, or even why they can be expected.[35] Given these limitations, my theoretical contribution frames the importance of adopting a theory of polling error that embraces electoral heterogeneity and provides the first comprehensive, phenomenon-level account of why this heterogeneity is likely.

My second theoretical contribution centres on providing a novel four-level approach to understanding sources of polling error. Though multi-level approaches to the analysis of polling error have been used in previous academic research,[36] as well as in analysis conducted by pre-election polling aggregators,[37] these approaches employ no more than three levels. My approach to analysis is the first to contain discrete poll-, pollster-, election-, and country-level grouping factors. I contend that this is the most appropriate method for the multi-level modelling of polling error, as it better mirrors the reality of it sources than previous approaches.

The first methodological contribution made by this thesis centres on developing three novel approaches to measuring polling error. I develop two new limit-based measures of the extent to which pre-election polls exhibit error beyond the boundaries set by their margins of error and formalise a novel binary approach to measuring whether polls correctly predict the recipient of the largest share of the vote. While existing approaches to understanding polling error beyond the stated margin of error principally centre on decomposing bias and variance components,[38] my newly developed measurement strategies provide direct measures of the

[35] Sohlberg and Branham, p. 11
[36] Tudor and Wall, p. 13.
[37] FiveThirtyEight, *How FiveThirtyEight's House, Senate and Governor Models Work* (2020), <https://fivethirtyeight.com/methodology/how-fivethirtyeights-house-and-senate-models-work/> [accessed 8 February 2022].
[38] Peter Selb and others, *Bias and Variance in Multiparty Election Polls*, <https://www.polver.uni-konstanz.de/typo3temp/secure_downloads/105568/0/f8a48094815e56ff0112346ba44c2136cc6e487a/Disentangling.pdf> [accessed 12/12/2021]; Houshmad Shirani-Mehr and others, 'Disentangling Bias and Variance in Election Polls', *Journal of the American Statistical Association 113.552* (2018), 607 – 614 (p. 607).

extent to which the error exhibited by polls exceeds their associated margins of error. By contrast, my novel binary measure of polling error presents an approach to measuring the substantive inaccuracy of pre-election polls that, at a high level of abstraction, overcomes several of the issues of cross-case tractability associated with such measures, which often require case-specific solutions.[39]

Beyond my theoretical and methodological contributions, I also provide three empirical contributions to the study of pre-election polling error through the production of this thesis. The first of these contributions relates to empirical data. Through the production of this thesis, I provide and draw upon what is, to the best of knowledge, the most geographically expansive polling dataset assembled to data in political science. It comprises 11,832 in-campaign polls conducted in 497 elections across 83 countries and stands as the only existing dataset to contain polls spanning all six populated continents of the world. While other large-scale, international polling datasets exist,[40] the geographical scope of their polling data is much reduced in comparison to the novel dataset collated for use within this thesis.[41]

The second and third empirical contributions of this thesis relate to its findings. Through my analysis, I produce the first findings that indicate that the effect of election-level differences on polling error variance is robust to the presence of both country- and pollster-level controls. I also produce the first findings that demonstrate the practical utility of adopting an election-level understanding of polling error. I show that electoral characteristics are useful predictors

[39] Tudor and Wall, pp. 9 – 10.
[40] Jennings, Will and Christopher Wlezien, *Replication Data for: Election Polling Errors Across Time and Space* (2018), <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/8421DX> [accessed 31 January, 2022]; Ryan Kennedy, Stefan Wojcik, and David Lazer, 'Improving Election Prediction Internationally', *Science*, 355.6324 (2017), 515 – 520 (p. 515).
[41] Ibid.

of the polling error variance, both individually and in interaction with one another, and are therefore capable of illuminating those circumstances in which error variance is more likely.

To provide the foundation on which these contributions are based, the following chapter reviews the academic literature concerning pre-election polling error written to date. In so doing, it provides the context in which the work contained within this thesis is situated and demonstrates the gaps in scholarship that it intends to fill.

Chapter 2 – A Rising Challenge to the Dominant Paradigm: Reviewing the Presence of Election-level Explanations in Assessments of Polling Error

"Developing an accurate [poll-based] prediction model, one that will work in a particular country in the particular context of a particular election, is always going to be difficult".[42]

- Simon Atkinson (2017)

Past assessments of polling mispredictions have largely adopted a poll-level approach to understanding error. Under this approach, the sources of error that cause misprediction are believed to stem from issues within polls themselves, most notably the methods that underpin them. While recent scholarship has identified that polling accuracy has not become worse under this approach,[43] it has equally failed to bring about meaningful improvement.[44] The lack of improvement under this poll-level approach begs the question of whether its use in isolation is sufficient to fully understand misprediction, or whether its combination with an alternative approach would provide greater insight.

Within this review, I identify that an alternative, election-level approach to understanding polling error has been growing in the literature since the 1930s, with the speed of its growth increasing in recent years. Under this approach, polling error is believed to stem from the characteristics of the elections in which polls are conducted, such as the number of parties contesting them or differences in turnout between them. Despite its presence and growing prominence, this election-level approach is still in its nascent stages and remains largely marginalised in favour of its poll-level counterpart. As a consequence of this, it remains

[42] Simon Atkinson, 'The Opinion Polls: In Praise of Measurement', *International Journal of Market Research*, 59.4 (2017), 405 – 407 (p. 406).

[43] Will Jennings and Christopher Wlezien, 'Election Polling Errors Across Time and Space', *Nature Human Behaviour*, 2.4 (2018), 276 – 283 (p. 280); Christopher Prosser and Jonathan Mellon, 'The Twilight of the Polls? A Review of Trends in Polling Accuracy and the Causes of Polling Misses', *Government and Opposition*, 53.4 (2018), 757 – 790 (p. 757).

[44] Jennings and Wlezien, 'Election Polling Errors Across Time and Space', p. 280.

underexplored. As the election-level approach is gaining momentum within the literature, I hold that the discipline is now primed for a large-scale analysis of the effect of election-level factors on polling accuracy and contend that substantive plausibility for such an approach can be found in the historical polling misses that form the core of this review.

By way of structure, I identify five waves of post-election analyses encompassing 26 polling failures in national legislative and presidential elections across 12 countries. These waves exist as collections of temporally proximal polling failures, with the post-mortem analyses in each wave possessing greater reference to election-level factors than the one preceding it. I identify that whilst the earlier waves of post-mortem analyses were Americentric in focus, due to the earlier emergence and maturation of pre-election polling in the US, subsequent waves became more international in focus as polling became a more widespread and internationally normalised practice.

I show that assessments of polling misses have predominantly adopted poll-level approaches to understanding error, with election-level considerations remaining significantly underexplored by comparison. Nevertheless, I demonstrate that the potential for election-level characteristics to affect polling error has been recognised within the literature since the 1930s and has been slowly growing over time. I conclude the first section by contending that election-level assessments of polling error are beginning to enter the academic mainstream, priming the polling literature for an in-depth analysis of their nature and importance.

Before beginning the review, it is important to unpack some of the terminology on which the discussion at its core rests. The review focuses on two broad conceptions of polling error: quantitative inaccuracy and qualitative inaccuracy. Quantitative inaccuracy concerns percentage point differences between the estimated vote share distributions provided by polls and the actual vote share distributions that emerge on election day. By contrast, qualitative inaccuracy concerns whether the leading party or candidate identified in a poll goes on to win the election. This qualitative form of polling inaccuracy is often referred to as substantive inaccuracy, as it concerns the ability of polls to correctly predict the substantive outcome of the elections to which they pertain. While the methodological revisions undertaken by pollsters that serve as one of the foci of the review often focus on improving quantitative inaccuracy, instances of qualitative inaccuracy have served more widely as drivers of the post-election assessments of polling error on which it centres, especially those conducted by the media.[45]

In addition to different conceptions of error, the discussion in the subsequent literature review applies a series of labels to the sources of polling inaccuracy identified by post-election assessments. It chiefly focuses on two labels: poll-level and election-level. Poll-level sources of error are those which relate to the methods inherent within, and nature of, polls themselves. Examples of poll-level sources of error include sample size and the number of days prior to an election that fieldwork is conducted. By contrast, election-level sources of error concern characteristics possessed by elections that serve to drive polling inaccuracy. Examples of such characteristics include pronounced shifts in turnout between elections, or marked differences in the extent of late decision-making amongst the electorate between contests. Beyond these labels, the literature review occasionally makes reference to pollster- and country-level sources of error. Pollster-levels sources of error concern drivers of inaccuracy inherent within decisions made, or actions undertaken, by polling organisations. These include the potential for herding – the phenomenon in which pollsters monitor and are often inclined to match the predictions rendered by other organisations[46] – and differences in the specific models employed by pollsters. Similarly, country-level sources of error concern characteristics possessed by

[45] Tudor, p. 37.
[46] Paul Whiteley, 'Why Did the Polls Get It Wrong in the 2015 General Election? Evaluating the Inquiry into Pre-election Polls', *The Political Quarterly*, 87.3 (2016), 437-442 (p. 438).

countries that can be expected to bear on the accuracy of polls. These include the imposition of polling moratoriums by certain countries, representing periods of time prior to an election during which pre-election polling is forbidden,[47] and the different electoral systems employed by states. Each of these sources of polling error is further unpacked later in the thesis in the discussion of the hierarchical nature of polling inaccuracy. With the terminology on which the literature review is based laid out, I move to address the first wave of post-elections assessments of polling error within the literature.

The First Wave of Post-election Assessments of Polling Inaccuracy

While the roots of pre-election polling can be traced back to the 1824 U.S. presidential election,[48] the first wave of substantive post-election assessments of polling inaccuracy began in 1936 with the predictive failure of the *Literary Digest*. After successfully predicting the five preceding US presidential elections,[49] the *Digest* confidently predicted that Alf Landon (R)

[47] Tim Bale, 'Restricting the Broadcast and Publication of Pre-election and Exit Polls: Some Selected Examples', *Representation*, 39.1 (2002), 15 – 22 (p. 15).
[48] Tom W. Smith, 'The First Straw? A Study of the Origins of Election Polls', *Public Opinion Quarterly,* 54.1 (1990), 21-36 (p. 28); James W. Tankard Jr., 'Public Opinion Polling by Newspapers in the Presidential Election Campaign of 1824', *Journalism Quarterly*, 49 (1972), 361-365 (p. 361); George Gallup and Saul Rae, *The Pulse of Democracy: The Public-Opinion Poll and How It Works*, (New York: Simon and Schuster, 1940), p.3; George Gallup, *The Sophisticated Poll Watcher's Guide*, (Princeton: Princeton Opinion Press, 1972), p. 240; William Lydgate, *What America Thinks*, (New York: Thomas Y. Crowell, 1944), pp. 1-3; Mildred Parten, *Surveys, Polls, and Samples: Practical Procedures*, (New York: Harper and Brothers, 1950), pp. 1-2; Susan Herbst, *Numbered Voices: How Opinion Polling Has Shaped American Politics*, (Chicago: University of Chicago Press, 1995), p.79; John Gray Geer, *Public Opinion and Polling Around the World: A Historical Encyclopaedia*, (Santa Barbara: ABC-CLIO, 2004), p. 39; Graham R. Walden, *Public Opinion Polls and Survey Research: A Selective Annotated Bibliography of U.S. Guides and Studies from the 1980s*, (New York: Routledge, 2014), p. xiii; Bill Jones and Dennis Kavanagh, *British Politics Today*, (Manchester: Manchester University Press, 2003), p. 75; Nick Moon, *Opinion Polls: History, Theory and Practice*, (Manchester: Manchester University Press, 1999), p. 6.
[49] The Literary Digest, 'Political Reports from 3,000 Communities', *The Literary Digest*, 28 October 1916, Topics of the Day, p. 1087; The Literary Digest, 'How the Straws Say the Election Will Go', *The Literary Digest*, 23 October 1920, Topics of the Day, p. 14; The Literary Digest, '2,386,052 Straws Forecast Tuesday's Tempest', *The Literary Digest*, 1 November 1924, Topics of the Day, pp. 5-8; The Literary Digest, 'Final Returns in the Digest's Presidential Poll', *The Literary Digest*, 3 November 1928, Topics of the Day, pp. 5-7; The Literary Digest, 'Roosevelt Bags 41 States Out of 48', *The Literary Digest*, 5 November 1932, Topics of the Day, pp. 8-9.
The Literary Digest, 'Landon, 1,293,669; Roosevelt, 972,897: Final Returns in the Digest's Poll of Ten Million Voters', *The Literary Digest*, 31 October 1936, Topics of the Day, pp. 5-6.

would defeat Franklin D. Roosevelt (D) in the 1936 presidential election.[50] On election day, Roosevelt won in a landslide, carrying 46 states on his way to the White House.[51] The failure of the *Digest* was made all the more stark by the success of competing polls conducted by Gallup, Crossley, and Roper.[52] Owing to its past success, the failure of the *Digest* came as a shock and triggered a series of analyses into the causes of its misprediction. As the *Digest* relied on large-scale convenience sampling whilst Gallup, Crossley, and Roper used quota sampling,[53] poll-level issues surrounding sampling methods became the primary focus of these analyses. The *Digest*'s own enquiry into the misprediction concluded that unrepresentative sampling due to issues of survey response was to blame for the failure.[54] Later academic analyses corroborated this finding, identifying the over-representation of higher income voters in the *Digest*'s sample as the cause of the predictive failure.[55]

Though the adoption of a poll-level approach by those assessing the misprediction of 1936 is understandable, as the success of Gallup, Crossley, and Roper's quota sampling made the case for the misprediction existing as a methods-based problem seem self-evident, the performance of these competing polls was not as convincing as it might have been. Although they correctly called Roosevelt as the winner of the election, they overestimated his vote share by five percentage points.[56] For this reason, while differing sampling methods clearly affected the

[50] The Literary Digest, 'Landon, 1,293,669; Roosevelt, 972,897: Final Returns in the Digest's Poll of Ten Million Voters', pp. 5-6.
[51] David A. Hopkins, *Red Fighting Blue: How Geography and Electoral Rules Polarize American Politics*, (Cambridge: Cambridge University Press, 2017), pp. 64 – 98.
[52] Sarah E. Igo, '"A Gold Mine and a Tool for Democracy": George Gallup, Elmo Roper, and the Business of Scientific Polling, 1935 – 1955', *Journal of the History of the Behavioral Sciences*, 42.2 (2006), 109 – 134 (p. 109); D. W. Moore, *The Superpollsters: How they Measure and Manipulate Public Opinion in America*, (New York: Four Walls Eight Windows, 1992), pp. 31 – 55; M. Wheeler, *Lies, Damn Lies, and Statistics: The Manipulation of Public Opinion Polls in America*, (New York: Liveright, 1976), pp. 67 – 70.
[53] Archibald M. Crossley, 'Straw Polls in 1936', *Public Opinion Quarterly*, 1.1 (1937), 24 – 35 (p. 24); Katz and Cantril, p. 158.
[54] The Literary Digest, 'What Went Wrong with the Polls?', *The Literary Digest*, 14 November 1936, pp. 7-8.
[55] Harold F. Gosnell and Sebastian de Grazia, 'A Critique of Polling Methods', *The Public Opinion Quarterly*, 6.3 (1942), 378-390 (p. 378); Katz and Cantril, p. 167; Crossley, p. 29.
[56] Peverill Squire, 'Why the Literary Digest Failed', *Public Opinion Quarterly*, 52.1 (1988), 125 – 133 (p. 129).

ability of polls to correctly call the substantive outcome of the election, they did not provide a complete understanding of the causes of misprediction. This brings into question whether an alternative approach could have illuminated additional sources of predictive error to provide a more complete understanding of the causes of misprediction.

An alternative election-level understanding of poll-based misprediction was beginning to emerge within the literature prior to the *Digest*'s failure. In 1932, Robinson acknowledged the importance of election-level factors for polling accuracy. Amongst a broader review of poll-level sources of error, including issues of sampling and survey mode, he emphasised the importance of the closeness of the race between candidates within an election. In addition to arguing that pollsters ought to use the closeness of an election to inform the uncertainty surrounding their predictions, he noted that slimmer margins between leading candidates increased the likelihood of polling error.[57] Through this, he produced the first indication that the characteristics of elections mattered for the predictive accuracy of polls.

Despite the existence of this alternative election-level approach, it was not adopted by any assessments of the misprediction of 1936. It must be noted, though, that the election was not conducive to its application. Not only did the unanimous and seemingly self-evident conclusions surrounding unrepresentative sampling not provide much scope for election-level enquiry, but the success of quota sampling rendered the failure of the *Digest* an outlying case. As Gallup, Crossley, and Roper had succeeded, it was more reasonable to assert that the source of error lay within the *Digest* itself, rather than within the characteristics possessed by the election. In later analysis, it was found that if all those who had received the *Digest*'s poll had responded, it would have correctly predicted the election,[58] all but confirming that its failure

[57] Claude E. Robinson, *Straw Votes: A Study of Political Prediction*, (New York: Columbia University Press, 1932), p. 129.
[58] Squire, p. 125.

was a poll-level phenomenon. Even if Robinson's focus on the importance of the closeness of the race had been used in an attempt to understand the overestimation of those polls using quota sampling, it is doubtful whether it would have had a significant impact, as Roosevelt won the election by a significant margin, garnering a share of the vote 24% larger than that of Landon.[59]

The predictive failure of the *Literary Digest* in 1936 therefore stands as an important moment in pre-election polling's long history of misprediction. The success of the quota sampling methods used by Gallup, Roper, and Crossley and the seemingly unambiguous conclusion that convenience sampling caused the *Digest*'s failure came to establish the dominance of the poll-level approach to understanding polling misses moving forward.

While polls correctly predicted the results of the 1940 and 1944 US presidential elections,[60] the literature following these elections still sought to identify sources of predictive error. Within this literature, the dominant poll-level focus remained evident, with papers addressing problems in the techniques associated with projecting Electoral College votes from raw vote shares,[61] as well as issues concerning the use of weightings and post-survey adjustments.[62] Others drew attention to the wording of questions used within pre-election polls,[63] as well as the potential presence of partisan house effects.[64] However, despite the predominant poll-level focus, certain literature produced in the wake of the 1944 election focused on the impact of election-level factors on polling accuracy. Indeed, the American Institute of Public Opinion

[59] Laura Crowell, 'Franklin D. Roosevelt's Audience Persuasion in the 1936 Campaign', *Speech Monographs*, 17.1 (1950), 48 – 64 (p. 48).
[60] Daniel Katz, 'The Public Opinion Polls and the 1940 Election', *Public Opinion Quarterly*, 5.1 (1941), 52 – 78 (p. 52); Katz, 'The Polls and the 1944 Election', p. 468.
[61] Henry B. Mann, 'On a Problem of Estimation Occurring in Public Opinion Polls', *The Annals of Mathematical Statistics*, 16.1 (1945), 85 – 90 (p. 85).
[62] Daniel Katz, 'The Polls and the 1944 Election', *Public Opinion Quarterly*, 8.4 (1944), 468 – 482 (p. 482).
[63] Gordon. M. Connelly, 'The Questions the Polls Ask: Now Let's Look at the Real Problem: Validity', *Public Opinion Quarterly*, 9.1 (1945), 51 – 60 (p. 51); Hadley Cantril, 'Do Different Polls Get the Same Results?', *Public Opinion Quarterly*, 9.1 (145), 61 – 69 (p. 61); Henry C. Freiberg, 'The Problem of Validity vs. Reliability in Public Opinion Polls', *Public Opinion Quarterly*, 6.1 (1942), 87 – 98 (p. 87).
[64] Eric F. Goldman, 'Poll on the Polls', *The Public Opinion Quarterly*, 8.4 (1945), 461 – 467 (p. 461).

identified the difficulty of predicting low turnout elections.[65] However, election-level analyses were still comparatively scarce in the face of poll-focused questioning.

The focus on predictive error intensified when polls failed to predict Harry S. Truman as the winner of the 1948 US presidential election. Instead, they predicted that Thomas E. Dewey would be victorious, a prediction that was emblazoned boldly on the front page of the *Chicago Daily Tribune* (see: Figure 1).[66]

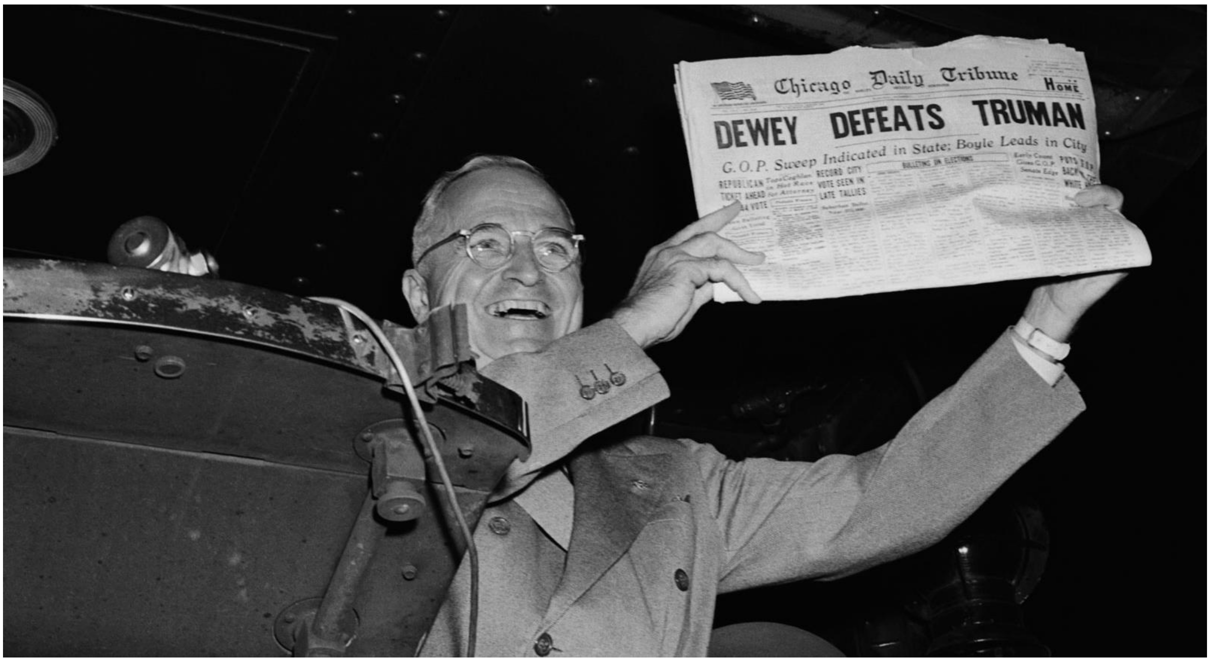

Figure 1: Famously wrong – following the U.S. presidential election of 1948, Harry S. Truman holds aloft the Chicago Daily Tribune headline predicting his defeat on the basis of unfavourable polls. Truman won the election by 114 Electoral College votes.[67]

This errant prediction brought about a series of analyses into its causes. Chief amongst the assessments of the 1948 misprediction was the enquiry launched by the Social Science Research Council.[68] Although it adopted a solely poll-level approach to the failure, it expanded

[65] Edward G. Benson, Cyrus C. Young, Clyde A. Syze, 'Polling Lessons from the 1944 Election', *Public Opinion Quarterly*, 9.4 (1945), 467 – 484 (p. 484).
[66] Chicago Daily Tribune, 'Dewey Defeats Truman', *Chicago Daily Tribune*, 3 November 1948, p. 1.
[67] Nicholas Lai, *Clinton Defeats Trump: Polling Failure or Media Failure?* (2018), <https://bpr.berkeley.edu/2018/05/22/clinton-defeats-trump-polling-failure-or-media-failure/> [accessed 25/07/2020].
[68] Mosteller and others, pp. 174 – 289.

the scope of enquiry. In characterising sources of error, it identified that polls possessed an intricate series of steps, any one of which could give rise to error and misprediction.[69] Of these error-prone steps, the enquiry focused its criticism on the inadequate allocation of undecided voters,[70] failure to remove non-voters from samples,[71] unrepresentative sampling,[72] and failure to capture late swings in voting intention.[73] It concluded that sampling and interviewer error, along with inadequate allocation of undecided voters and a failure to foresee a late swing in voting intention, were the principal causes of the polling miss.[74]

The poll-level focus of the enquiry by the Social Science Research Council was mirrored by other assessments of the 1948 predictive failure. In much the same way as 1936, several works pointed to sampling issues, including unrepresentative samples and problematic sampling methods, as the cause of misprediction.[75] Others expanded poll-level enquiry further and blamed complex question wording and biases within the interview process for the failure.[76]

Though the assessments of the 1948 misprediction placed blame at the feet of a variety of sources of error, these sources were invariably poll-level in nature. As such, they dismissed the alternative election-level approach to misprediction that had been growing in depth during the years since 1936. Through his work on the 1942 US mid-term elections, Robinson built upon earlier work by the American Institute of Public Opinion, recognising the importance of an additional election-level factor for misprediction: turnout.[77] He noted that fluctuations in

---

[69] Ibid., p. 6.
[70] Ibid., pp. 190 – 210.
[71] Ibid., p. 300
[72] Ibid.
[73] Ibid., p. 251.
[74] Ibid., p. 290.
[75] Leo P. Crespi, 'The Interview Effect in Polling', *Public Opinion Quarterly*, 12.1 (1948), 99-111 (p. 110).
[76] Fay Terris, 'Are Poll Questions Too Difficult?', *Public Opinion Quarterly*, 13.2 (1949), 314-319 (p. 314); Lindsay Rogers, *The Pollsters: Public Opinion, Politics, and Democratic Leadership*, (New York: Alfred A. Knopf, 1949), p. 124.
[77] Claude E. Robinson, 'Pre-election Polls in the 1942 Elections', *Public Opinion Quarterly*, 7.1 (1943), 139 – 144 (p. 139).

turnout levels profoundly affected the vote shares received by differing parties.[78] As they affect vote share distributions, these fluctuations necessarily affect electoral outcomes and attempts to accurately predict them. Due to this, he identified that error occurs if polls fail to foresee shifts in turnout.[79] While attempts to foresee such shifts were being undertaken by both polling organisations and universities at the time, they were in their nascent stages.[80]

Following the predictive failure of 1948, the 1950s saw the production of a series of works affirming the poll-level approach to assessing instances of misprediction. Not only did this period see a triumphant Gallup state that the issues presented by late swings, undecided voters, and turnout projection had all been resolved through method-based revisions,[81] but authors variously attributed polling inaccuracy to poll-level issues of sampling,[82] in-house adjustment methods,[83] interview biases,[84] issues of survey response,[85] and problematic question wording.[86] Select works even served to summarise the range of poll-level issues deemed responsible for misprediction.[87] Despite this, select analyses recognised the impact of differences between elections as drivers of polling error. Specifically, works identified the ability for shifts in turnout between elections to confound the turnout projection mechanisms on which polls result and undermine their accuracy.[88]

[78] Ibid., p. 140 – 141.
[79] Ibid., p. 141.
[80] Ibid., p. 143.
[81] Gallup, 'The Gallup Poll and the 1950 Election', p. 17.
[82] Charles F. Haner and Norman C. Meir, 'The Adaptability of Area-Probability Sampling to Public Opinion Measurement', *Public Opinion Quarterly,* 15.2 (1951), 335 – 352 (pp. 335-352).
[83] H. Field, 'The Accuracy of Pre-election Polls', *Bulletin of the Oxford University Institute of Economics and Statistics*, 13.7 (1951), 201 – 222 (p. 221).
[84] Donald T. Campbell, 'Systematic Error on the Part of Human Links in Communication Systems', *Information and Control*, 1 (1958), 334 – 369 (p. 351).
[85] Hugh J. Parry and Helen M. Crossley, 'Validity of Responses to Survey Question', *Public Opinion Quarterly*, 14 (1950), 61 – 80 (p. 80).
[86] W. Robson, 'The Survey on Nationalisation', *Political Quarterly*, 30.2 (1959), 111 (p. 111).
[87] Louis Harris, 'Election Polling and Research', *Public Opinion Quarterly*, 21.1 (1957), 108 – 116 (p. 115).
[88] Mungo Miller, 'The Waukegan Study of Voter Turnout Prediction', *Public Opinion Quarterly*, 16.3 (1952), 381 – 398 (p. 398).

The 1960s saw a continuation of the focus on poll-level explanations of misprediction within academic literature. Works focused on issues of survey response,[89] the impact of biases present in the interview process,[90] problematic sampling procedures,[91] along with the effect of question wording on response validity.[92] However, despite the predominantly poll-level nature of surrounding literature, in their assessment of polling conducted for the 1960 US presidential election, Hennessy and Hennessy built upon the election-level work of Robinson by identifying that pre-election polls do not predict close elections well.[93] They contended that in elections in which the margin between the candidates was smaller than the margin of error inherent within polls, their success was largely based on luck, with the predictions they render being no better than hunches.[94]

Though additional election-level works were slowly beginning to appear in the literature, the predominantly poll-level focus of the first wave of post-election assessments of polling inaccuracy continued into the 1970s. This focus was particularly pronounced when pre-election polls errantly predicted that the Labour Party would win the 1970 British general election.[95] Assessments of the predictive failure predominantly adopted a poll-level approach. The post-election Nuffield study investigated several sources of polling error, concluding that while the presence of undecided voters was problematic, the inability of polls to detect a late swing in voting intention was the primary cause of the misprediction.[96] Their conclusion was later

89 Paul Perry, 'Election Survey Procedures of Gallup Poll', *Public Opinion Quarterly*, 24 (1960), 531 – 542 (p. 534).
90 Archibald M. Crossley and Helen M. Crossley, 'Polling in 1968', *The Public Opinion Quarterly*, 33.1 (1969), 1-16 (p. 1).
91 D. E. G. Plowman, 'Public Opinion and the Polls', *The British Journal of Sociology*. 13.4 (1962), 331 – 349 (p. 331); Leslie Kish, *Survey Sampling*, (New York: John Wiley and Sons, 1965), pp. 510 – 532.
92 Elisabeth Noelle-Neumann, 'About Methodological Progress in Survey Research', *Dialectica*, 16.4 (1962), 307 – 328 (pp. 307 – 328).
93 Bernard C. Hennessy and Erna R. Hennessy, 'The Prediction of Close Elections: Comments on Some 1960 Polls', *The Public Opinion Quarterly*, 25.3 (1961), 405-411 (p. 409).
94 Ibid., p. 411.
95 Tudor and Wall, p. 1.
96 David Butler and M. Pinto-Duschinsky, *The British General Election of 1970*, (London: Macmillan, 1971), pp. 180 – 185.

corroborated by the Market Research Society,[97] while alternative accounts accorded blame to issues of survey response, question wording, and the presence of a bandwagon effect in which individuals are more likely to support candidates or parties that are ahead in the polls.[98]

Although poll-level factors were deemed responsible for the predictive failure by most assessments, a number of works drew attention to the importance of the role of low turnout.[99] Despite this, its impact was questioned by Abrams and subjected to post-hoc analysis which discounted it as a significant contributory factor.[100] Instead, he drew further attention to the issues presented by undecided and capricious voters.[101]

The marginalisation of election-level explanations was also evident in assessments of the failure of polls to correctly predict the outcome of the subsequent British general election, held in February 1974.[102] In the years since the failure of 1970, the polling literature continued to focus on poll-level explanations of misprediction, citing the importance of sampling procedures,[103] the inability of polls to capture late swings in voter sentiment,[104] and the interview process itself as significant determinants of error.[105]

97 Market Research Society, *Public Opinion Polling on the 1970 Election*, (London: MRS, 1972), pp. 1 – 5.
98 Stephen P. Koff, 'Public Opinion Polls and the 1970 British Election', *Il Politico*, 37.3 (1972), 466 – 482 (pp. 478 – 479).
99 Richard Hodder-Williams, *Public Opinion Polls and British Politics*, (London: Routledge, 1970), p. 92; Richard Rose, *The Polls and the 1970 General Election*, Occasional Paper No. 7, (Strathclyde: University of Strathclyde Survey Research Centre, 1970) p. 1; F. Teer and J. Spence, *Political Opinion Polls*, (London: Hutchinson, 1973), pp. 1 – 240.
100 Mark Abrams, 'The Opinion Polls and the 1970 British General Election', *The Public Opinion Quarterly*, 34.3 (1970), 317-324 (p. 321).
101 Ibid, pp. 323 – 324.
102 Tudor and Wall, p. 1.
103 Paul Perry, 'A Comparison of the Voting Preferences of Likely Voters and Likely Non-voters', *Public Opinion Quarterly*, 37.1 (1973), 99 – 109 (p. 105).
104 Stephen E. Fienberg, 'The Sun-Times Straw Poll, 1968 and 1970: A Statistical Appraisal', *Journal of the American Statistical Association*, 67.338 (1972), 292 – 297 (p. 296).
105 Robert E. Kraut and John B. McConahay, 'How Being Interviewed Affects Voting: An Experiment', *Public Opinion Quarterly*, 37.3 (1973), 398 – 406 (p. 396); Robert M. Worcester, 'Winning in the Rain', *New Society*, 31.643 (1975), 394 (p. 394).

In keeping with the literature of the time, assessments of the failure of polls to correctly predict the British general election of February 1974 adopted a poll-level approach. Analyses variously identified problems with votes-to-seats transformation methods and in-house adjustment mechanisms as the root causes of the polling miss.[106] Though much was made of the rise of the Liberal party in wider literature,[107] their improved performance was not directly applied to understandings of the misprediction.

While the cases addressed to this point could lead to the suspicion that the identified themes merely exist as idiosyncrasies of post-election analyses of polling performance in anglophone elections, the misprediction of the French general election of 1978 serves to dispel such concerns. In the time since the British polling miss four years earlier, the literature had further entrenched along poll-level lines, focusing on the importance of the time between the completion of a poll and election day,[108] the revision of polling methods,[109] increasing refusal rates,[110] and the difficulty of question comprehension for polling accuracy.[111] Pursuant with its dominance in the literature, assessments of the failure of polls to correctly predict the French general election of 1978 unanimously adopted a poll-level approach, identifying problematic sampling methods, the presence of undecided voters, and flawed turnout projection

[106] Richard Rose, 'The Polls and Election Forecasting in February 1974' in H. Penniman (ed.), *Britain at the Polls* (Washington DC: AEI, 1974), pp. 109 – 130.
[107] Peter H. Lemiuex, 'Political Issues and Liberal Support in the February 1974, British General Election', *Political Studies*, 25.3 (1977), 323 – 342 (pp. 323 – 342); Ivor Crewe, Bo Särlvik, and James Alt, 'Partisan Dealignment in Britain 1964 – 1974, *British Journal of Political Science*, 7.2 (1977), 129 – 190 (p. 131).
[108] Marcus Felson and Seymour Sudman, 'The Accuracy of Presidential Preference Primary Polls', *Public Opinion Quarterly*, 39.2 (1975), 232 – 236 (p. 232).
[109] George Gallup, 'Pollsters, Not Prophets', *Society* (13 (1976), 19 – 23 (p. 22).
[110] Seymour Martin Lipset, 'The Wavering Polls', *The Public Interest*, 43 (1976), 70 – 89 (p. 70); Seymour Martin Lipset, 'Interpreting the Polls', in *Advances in Consumer Research*, ed. by Beverlee B. Anderson, (Cincinnati: Association for Consumer Research, 1976), pp. 17 – 23.
[111] Jean M. Converse, 'Predicting No Opinion in Polls', *Public Opinion Quarterly*, 40 (1976), 515 – 530 (p. 515).

mechanisms as prominent sources of polling error,[112] along with the failure of polls to detect a late swing in voting intention.[113]

Analyses conducted in the wake of the 1978 French general election marked the end of the first wave of post-election assessments of polling inaccuracy. The first wave of literature was marked by its near universal focus on poll-level explanations for predictive failures. Though several election-level explanations for polling misses were proposed during its 42-year span, such explanations were comparatively rare and only existed in their nascent stages. By contrast, the second wave of post-election analyses which began in the wake of the Japanese general election of 1979 was characterised by a growing focus on election-level explanations for polling failures, with a recognition of the importance of electoral characteristics starting to take root more widely within the literature.

The Second Wave of Post-election Assessments of Polling Inaccuracy

Despite the continued dominance of poll-level understandings of predictive failure, assessments of the failure of polls to correctly predict the Japanese House of Representatives election in 1979 represented a step towards the normalisation of election-level alternatives. Indeed, these assessments served as an inflection point within the polling literature, shifting its focus, and leading to a second wave of post-election assessments of polling inaccuracy.

The polls conducted for the 1979 Japanese election vastly overestimated the performance of the Liberal Democratic Party.[114] Several hypotheses were posited for the misprediction, some of which centred on country-level drivers of error. Specifically, it was recognised that the vote

[112] Paul Perry, 'Certain Problems in Election Survey Methodology', *Public Opinion Quarterly*, 43.3 (1979), 312 – 325 (p. 312).
[113] Vincent Wright, 'The French General Election of March 1978: La Divine Surprise', *West European Politics*, 1.3 (1978), 24 – 52 (p. 41).
[114] Hans. H. Baerwald, 'Japan's 35th House of Representatives Election: The LDP Toys with a Return to 1954', *Asian Survey*, 20.3 (1980), 257 – 268 (p. 260).

share predictions rendered by polls were undermined by poorly targeted endorsements of candidates by the Liberal Democratic party.[115] The largely multi-member district system employed in Japanese House of Representatives elections is such that if a party endorses too many candidates in a given district, it risks diffusing its overall vote share to such a degree that none of its selected candidates are elected (or each receives a diminished share of the vote).[116] In the case of the 1979 election, the Liberal Democratic party was found to have endorsed too many candidates, distributing its overall vote share too widely, resulting in its chosen candidates failing to secure election in at least 17 districts.[117] This resulted in the Liberal Democratic party losing seats in the legislature that polls had predicted they would win, resulting in substantive error at the district-level.

Country-level factors were not the sole focus of polling post-mortems in the wake of the 1979 Japanese House of Representatives election. Indeed, election-level factors were also identified as drivers of polling error. Specifically, analyses focused on the unexpectedly low turnout in the elections, considered to be an artefact of the air of inevitability surrounding the victory of the Liberal Democratic Party, along with poor weather.[118] The impact of unexpected turnout levels on polling accuracy speaks to the reliance of poll-based predictions on anticipated behaviours being borne out on election day, as divergence alters the composition and intention of the voting population, negatively effecting the likelihood of predictive accuracy.

While the identification of electoral determinants of error in the Japanese case was encouraging for the progression of election-level understandings of misprediction, its dismissal in lieu of poll-level alternatives in most prior cases speaks to its outlying position. The dominance of the

[115] Ibid., pp. 260 – 261.
[116] Ibid.
[117] Ibid.
[118] Ibid.

poll-level approach to analyses of misprediction continued in the following year with the failure of polls to correctly predict the outcome of the 1980 US presidential election.

Such was the failure of pre-election polls in the 1980 US presidential election that reporting told of an atmosphere of 'backbiting, mudslinging and mutual criticism' amongst pollsters.[119] Despite the purported accuracy of private polls,[120] publicly available polls unanimously failed to correctly predict the magnitude of Ronald Reagan's victory.[121] Assessments published in popular publications in the immediate aftermath of the miss were quick to adopt a poll-level approach to identifying likely sources of error. These ranged from the failure of polls to account for a bandwagon effect in which voters are more inclined to support candidates who are ahead in the polls and a late swing in voting intention towards the end of the campaign, to problems of question order and untruthful responses.[122] Subsequent academic analyses followed suit, offering poll-level explanations surrounding inaccurate turnout projections,[123] the high number of undecided voters,[124] failure to detect voters who were unwilling to disclose their true voting intention when polled,[125] issues of sampling,[126] the impact of a bandwagon effect in which

[119] Lydia Saad, *Late Upsets Are Rare, but Have Happened* (2008), <https://news.gallup.com/poll/111451/late-upsets-rare-happened.aspx> [accessed 21/05/2020].

[120] Warren J. Mitofsky, 'The 1980 Pre-election Polls: A Review of Disparate Methods and Results', *The Proceedings of the Survey Research Methods Section of the American Statistical Association*, 4 (1981), 47 – 52 (p. 50).

[121] Gallup, *Gallup Presidential Election Trial-Heat Trends, 1936 – 2008* (2008), <https://news.gallup.com/poll/110548/gallup-presidential-election-trial-heat-trends.aspx> [accessed 21/05/2020]; Louis Harris, *Reagan Now Far Behind Carter as the President's Lead Continues to Grow*, *The Chicago Tribune*, 3 January 1980, ABC News-Harris Survey, pp. 1 – 3; John F. Stacks, 'Nation: Where the Polls Went Wrong', *Time Magazine*, 1 December 1980, pp. 21 – 24; Everett Carll Ladd, 'The Brittle Mandate: Electoral Dealignment and the 1980 Presidential Election', *Political Science Quarterly*, 96.1 (1981), 1 – 25 (p. 11).

[122] Stacks, p. 21.

[123] Andrew Kohut, 'A Review of the Gallup Pre-election Methodology in 1980', *The Proceedings of the Survey Research Methods Section of the American Statistical Association*, 4 (1981), 41 – 46 (p. 41); Mitofsky, 'The 1980 Pre-election Polls', p. 47.

[124] Ian Fenwick and others, 'Classifying Undecided Voters in Pre-election Polls', *Public Opinion Quarterly*, 46 (1982), 383 – 391 (p. 383); Kohut, p. 41.

[125] Kohut, p. 41; Mitofsky, 'The 1980 Pre-election Polls', p. 47.

[126] Mitofsky, 'The 1980 Pre-election Polls', p. 47.

voters are more likely to support candidates ahead in the polls,[127] and undetected late swing.[128] Not only did poll-level explanations dominate analyses of the 1980 misprediction, but the importance of the most prominent election-level factor under consideration, divergent turnout, was refuted and disproven,[129] though this finding was in keeping with the negligible decrease in turnout from the 1976 contest.[130] Despite this, undetected late swing born of late decision-making amongst the electorate was identified as an important driver of polling error.[131] As the extent of late decision-making has the potential to vary between elections, it stands as an election-level driver. When this is considered alongside the focus on the impact of turnout levels, the potential for election-level factors to bear on polling error was clearly recognised in assessments of polling accuracy following the 1980 US presidential election.

Two years later, the 1982 US mid-term elections witnessed another instance of poll-based misprediction. Inaccuracy was rife amongst the polls conducted for the mid-terms,[132] leading to a series of post-election assessments of their failure. These assessments largely adopted a poll-level approach to error, asserting that polls failed due to inaccurately predicting divergent voting behaviour between urban and rural areas[133] – an issue which suggests the presence of

[127] Andrew Skalaban, 'Do Polls Affect Elections? Some 1980 Evidence', *Political Behaviour*, 10.2 (1988), 136 – 150 (p. 136).
[128] Richard Brody and Lawrence Rothenberg, 'The Instability of Partisanship: An Analysis of the 1980 Presidential Election', *British Journal of Political Science*, 18 (1988), 445 – 465 (pp. 445 – 465); Gregory Markus, 'Political Attitudes During an Election Year: A Report on the 1980 NES Panel Study', *American Political Science Review*, 76 (1982), 538 – 560 (pp. 538 – 560).
[129] Kohut, p. 44.
[130] The American Presidency Project, *Voter Turnout in Presidential Elections* (2020), <https://www.presidency.ucsb.edu/statistics/data/voter-turnout-in-presidential-elections> [accessed 04/07/2020].
[131] Richard Brody and Lawrence Rothenberg, 'The Instability of Partisanship: An Analysis of the 1980 Presidential Election', *British Journal of Political Science*, 18 (1988), 445 – 465 (pp. 445 – 465); Gregory Markus, 'Political Attitudes During an Election Year: A Report on the 1980 NES Panel Study', *American Political Science Review*, 76 (1982), 538 – 560 (pp. 538 – 560).
[132] Burns W. Roper, 'Election Poll Errors: A Problem for Us All', *AAPOR News*, 10.2 (1983), 1 (p. 1).
[133] Andrew Kohut, 'Illinois Politics Confounds the Polls', *Public Opinion*, 5 (1983), 42 – 43 (pp. 42 – 43).

either coverage or non-response error – as well as failing to detect a late swing in voting intention and employing survey questions that were problematic in both wording and order.[134]

While the poll-level understandings of misprediction again dominated post-election assessments of polling error, limited election-level enquiries were undertaken. Not only did certain analyses focus on the impact of abnormally large turnout, but also the impact of strong partisanship in the form of dedicated straight-ticket voting.[135] The assessment of the impact of abnormally large turnout was a logical election-level enquiry, as past elections of the same kind are used as markers for expected turnout, and turnout levels in US mid-term elections are consistently low.[136] On this basis, the turnout projections employed by polls would be predicated on the expectation of low turnout and would therefore be susceptible to pronounced inaccuracy if this expectation was subverted by high turnout on election day.

Despite the heightened likelihood of the relevance of divergent turnout to mid-term elections, they were found to be of little consequence in the case of 1982.[137] However, this is not to say that election-level explanations were dismissed in their entirety. It was concluded that the strength of partisan loyalty exhibited in the election was the principal determinant of the misprediction as it brought about an upsurge in straight-ticket voting in direct contradiction to polling responses.[138]

[134] Richard Day and Kurt M. Becker, 'Pre-election Polling in the 1982 Illinois Gubernatorial Contest', *Public Opinion Quarterly*, 48.3 (1984), 606 – 614 (p. 608); Irving Crespi and Dwight Morris, 'Question Order and the Measurement of Candidate Preference in the 1982 Connecticut Elections', *Public Opinion Quarterly*, 48.3 (1984), 578-591 (pp. 578-591).
[135] Day and Becker, p. 608.
[136] Angus Campbell, 'Surge and Decline: A Study of Electoral Change', in *Elections and Political Order*, ed. by Angus Campbell and others, (New York: Wiley, 1966), p. 51; Robert A. Jackson, 'Differential Influences on Participation in Midterm Versus Presidential Elections', *The Social Science Journal*, 37.3 (2000), 385 – 402 (p. 385).
[137] Day and Becker, p. 610.
[138] Ibid., pp. 612 – 613.

Wider literature suggests that the role of partisanship as a determinant of polling error is particularly germane to mid-term elections, as they present lower information environments than presidential contests.[139] Not only are the responses gathered in such an environment largely the product of guesswork,[140] but low information electoral environments often result in the use of partisanship as a heuristic to aid voters in their decision-making at the ballot box.[141] As such, not only were the responses to polls less reliable due to the low-information environment, increasing the likelihood of misprediction, but they presented the ideal scenario for a reversion to partisanship to confound polling predictions.

The recognition of stronger-than-anticipated partisanship as the principal determinant of polls' failure to predict the 1982 US mid-term elections represented a clear turning point within the literature. While the second wave of post-election assessments of polling inaccuracy began with the recognition of the importance of electoral context for polling error in Japan, no assessments had yet conclusively identified election-level explanations for misprediction within Anglophone elections. Indeed, election-level explanations of polling misses in Western democracies more broadly had, until 1982, existed as secondary considerations or assumed the role of addenda to otherwise poll-focused literature. For assessments to settle on an election-level explanation for polling failure was a significant step and represented the growth in the scope of enquiry evident throughout the second wave of literature.

Though the second wave of post-election assessments of polling inaccuracy was bookended by election-level explanations of misprediction, demonstrating a greater prominence of election-level enquiries than the first wave, growth in the popularity of this approach remained slow. Nevertheless, the third wave of assessments in the literature would begin to solidify the

[139] Jackson, p. 385.
[140] Brian F. Schaffner and Matthew J. Streb, 'The Partisan Heuristic in Low-information Elections', *Public Opinion Quarterly*, 66 (2002), 559 – 581 (p. 561).
[141] Ibid., p. 578.

election-level approach championed in the case of the 1982 mid-term elections and begin to establish the basis on which future enquires would build.

The Third Wave of Post-election Assessments of Polling Inaccuracy

In the years following the misprediction of the 1982 US mid-terms, the academic literature again began to coalesce around poll-level explanations of error. Analyses of polling inaccuracy focused on the point in a campaign at which a poll was conducted,[142] issues born of question wording,[143] errors in turnout projection and likely vote estimation,[144] along with failures of sampling procedure.[145]

Despite the ostensible return to a poll-level understanding of misprediction, in 1986, Buchanan noted that reductively assessing polling inaccuracy in terms of sampling techniques was not sufficient to capture the wide range of methods employed by pollsters and, therefore, the full gamut of possible sources of error.[146] While advocating for wider enquiry into the determinants of polling error, Buchanan stopped short of calling for electoral characteristics to be included in analyses of misprediction. Nevertheless, his work betrays the beginning of the recognition of the need to move beyond narrow, procedurally orientated poll-level enquiry. This desire to broaden the scope of assessments into misprediction would come to characterise the third wave of literature.

[142] Kurt Lang and Gladys Engel Lang, 'The Impact of Polls on Public Opinion', *The Annals of the American Academy of Political and Social Science*, 472.1 (1984), 129-142 (p. 135); Stephen Borrelli, Brad Lockerbie, and Richard G. Niemi, 'Why the Democrat-Republican Partisanship Gap Varies from Poll to Poll', *The Public Opinion Quarterly*, 51.1 (1987), 115 – 119 (p. 119).
[143] Burns W. Roper, 'Are Polls Accurate?', *The Annals of the American Academy of Political and Social Science*, 472.1 (1984), 24 – 34 (p. 24); Alexander Rosenberg, *Philosophy of Social Science*, (Boulder: Westview Press, 1988), pp. 102 – 103; Borrelli, Lockerbie, and Niemi, pp. 115 – 119.
[144] Michael W. Traugott and Clyde Tucker, 'Strategies for Predicting Whether a Citizen Will Vote and Estimation of Electoral Outcomes', *Public Opinion Quarterly*, 48.1 (1984), 330 – 343 (pp. 330 – 343); Irving Crespi, *Pre-Election Polling: Sources of Accuracy and Error*, (New York: The Russell Sage Foundation, 1988), p. 170.
[145] Burns W. Roper, 'Are Polls Accurate?', pp. 24 - 34; Borrelli, Lockerbie, and Niemi, pp. 115 – 119.
[146] William Buchanan, 'Electoral Predictions: An Empirical Assessment', *The Public Opinion Quarterly*, 50.2 (1986), 222 -227 (p. 222).

Two years later, in his work on pre-election polling, Crespi acknowledged the complimentary nature of election- and poll-level sources of error. Specifically, he noted the effect of the visibility of an election – how aware the voting population is of its central issues and candidates – on the prevalence of late decision-making. He contended that the lower the visibility of an election, and therefore the less aware a voting population is of candidates and their positions, the later in the election cycle they actively begin the decision-making process.[147] This late decision-making increases the likelihood of a late swing in voting intention, contingent on polls being conducted earlier in the campaign. As such, Crespi illustrated that an idiosyncrasy of low visibility elections increases the prevalence of a poll-level source of error.

While the previous two waves of literature were relatively narrow in their geographic scope, the third wave of post-election assessments of polling inaccuracy began in earnest with an assessment of the 1990 Nicaraguan general election in which polls failed to predict the victory of Violeta Chamorro over incumbent Daniel Ortega.[148] The misprediction gave rise to a series of analyses that largely espoused poll-level understandings of the failure. The inability of polls to capture a late swing in voting intention,[149] the misallocation of non-respondents,[150] the inapplicability of North American-style polling procedures to a fundamentally different society, unrepresentative sampling, inadequate screening of likely voters, issues of question wording,[151] and issues of interview bias were all variously blamed for the failure.[152]

[147] Crespi, ‘Pre-election Polling’, p. 127.
[148] Katherine Bischoping and Howard Schuman, ‘Pens and Polls in Nicaragua: An Analysis of the 1990 Pre-election Surveys’, *American Journal of Political Science*, 36.2 (1992), 331 – 350 (p. 331).
[149] Simon Dyson, ‘Polls Apart? The 1990 Nicaraguan and 1992 British General Elections’, *The Political Quarterly*, 65.4 (1994), 425 – 431 (p. 431).
[150] Leslie Anderson, ‘Surprises and Secrets: Lessons from the 1990 Nicaraguan Election’, *Studies in Comparative International Development*, 27.3 (1992), 93 – 119 (p. 106).
[151] Peter V. Miller, ‘Which Side Are You On? The 1990 Nicaraguan Poll Debacle’, *Public Opinion Quarterly*, 55 (1991), 281 – 302 (pp. 288 – 291).
[152] Bischoping and Schuman, pp. 331 – 350; Leslie Anderson, ‘Neutrality and Bias in the 1990 Nicaraguan Pre-election Polls: A Comment on Bischoping and Schuman’, *American Journal of Political Science*, 38.2 (1994), 486 – 494 (pp. 486 – 493).

Despite the prominence of poll-level conclusions, election-level considerations played a complimentary role in assessments of inaccuracy. Due to the authoritarian nature of the incumbent, Daniel Ortega, levels of reported partisan loyalty amongst the electorate were found to have confounded polling estimates. Pre-election polls were perceived by voters to be partisan exercises conducted on behalf of the government – far from the dispassionate social scientific ideal – resulting in significant pressure to register support for the incumbent through fear of reprisal.[153] Consequently, the high level of false partisan loyalty recorded by polls, itself an artefact of the electoral context in which polls were conducted, was blamed in part for their failure.

1992 witnessed a significant polling failure in the United Kingdom, with the overwhelming majority of polls failing to correctly predict the Conservative victory in the general election of the same year. Indeed, the performance of the polls was worse than it had been for decades,[154] with the election being referred to as the Waterloo of public opinion polling (one of at least three polling failures to have been identified as Waterloos).[155] Such was the scope of the failure that it brough about large-scale post-election analyses which predominantly focused on poll-level sources of error. The inability to detect late swing in voting intention, poor handling of undecided voters, issues of non- and untruthful response, along with inadequate quota controls and bias within samples where all identified as significant determinants of error.[156]

[153] Miller, p. 350.
[154] Robert Worcester, 'The Polls Have a Lot to Answer For', *British Public Opinion*, 15 (1992), 1 – 3 (p. 1).
[155] David Butler and Dennis Kavanagh, *The British General Election of 1992*, (London: Macmillan, 1992), p. 135; Everett C. Ladd, 'The Election Polls: An American Waterloo', *Chronicle of Higher Education*, November 22, 1996, p. A52; Everett C. Ladd, 'The Pollsters' Waterloo', *Wall Street Journal*, November 19, 1996, p. A22; Owen Dudley Edwards, 'Remembering the Kennedys', *Journal of American Studies*, 18.3 (1984), 405 – 423 (p. 411).
[156] Market Research Society, *Report of the MRS Inquiry Into the 1992 General Election Opinion Polls*, (London: MRS, 1992), pp. 3 – 5; Ivor Crewe, 'A Nation of Liars? Opinion Polls and the 1992 Election', *International Journal of Market Research*, 35.4 (1993), 1 – 19 (pp. 1 – 19); Butler and Kavanagh, pp. 141 – 143; Rallings and Thrasher, p. 190; Jowell and others, pp. 238 – 262.

In spite of the dominance of poll-level conclusions, Crewe remarked that to solely focus on internal issues within polling while assessing the misprediction would be to disregard their success in previous elections. Instead, he asserted that it was incumbent upon analysts to adopt an election-level approach and assess the differences between the election of 1992 and previous contests,[157] a point agreed upon by Smith.[158] The presence of a greater number of shy voters – itself the result of social desirability bias surrounding the Conservative party – was identified as a key and impactful difference between the 1992 election and previous contests.[159]

Indeed, additional election-level explanations were considered in broader post-election assessments in the form of turnout. It was held that turnout of pro-Conservative voters was considerably higher than expected, contributing to the misprediction by confounding turnout projections.[160]

The widespread recognition of the importance of electoral characteristics determinants of polling error in the 1992 UK general election demonstrated that election-level assessments of polling inaccuracy were increasing in frequency and prominence. Moreover, the assertions of Crewe and Smith served to formalise the earlier suggestions of Buchanan, calling for a broadening of the scope of analyses into polling failures.

The increased election-level focus of the third wave of post-election assessments of polling error continued in 1997 with the failure of polls to correctly predict the French legislative election. In the years since the UK polling failure of 1992, literature renewed its general focus on poll-level sources of inaccuracy. Works identified issues of survey response,[161] question

---

[157] Crewe, p. 7.

[158] T. M. F. Smith, 'Public Opinion Polls: The UK General Election, 1992', *Journal of the Royal Statistical Society*, 159.3 (1996), 535 – 545 (p. 535).

[159] Crewe, p. 7; Smith, p. 540.

[160] Butler and Kavanagh, pp. 142 – 143; Market Research Society, *Report of the MRS Inquiry*, pp. 3 – 5.

[161] J. Zaller and S. Feldman, 'A Simple Theory of Survey Response: Answering Questions versus Revealing Preferences', *American Journal of Political Science*, 36.3 (1992), 579 – 616 (p. 579).

wording,[162] the duration and timing of field work,[163] along with the bias that resulted from the use of quota sampling as key determinants of prediction error.[164]

Despite the dominant focus on poll-level sources of inaccuracy, two pieces of prominent election-level literature were published during this time, the first of which was that of Gelman and King. In their analysis of the variability of American presidential polls, they identified a series of election-level determinants of polling error.[165] Most directly, they held that extremely close elections are more difficult to predict than less closely fought contests due to predictions being statistically indistinguishable from fifty-percent likelihoods.[166] Owing to their fast-paced nature, they also contended that primaries are more difficult to predict than presidential elections, as voting preferences shift substantially on the basis of singular events due to the issue proximity of candidates.[167] Finally, they noted the impact of low visibility elections and uneven campaigns on the accuracy of polls, corroborating the earlier work of Crespi and analyses of the 1980 US presidential election, respectively.[168]

The second election-level work was penned by Beltrán and Valdivia. Drawing on conclusions reached in the wake of the 1990 Nicaraguan election, they investigated the hypothesis that election-level factors were more significant determinants of error in countries that were not completely democratic.[169] They found no support for this hypothesis in the case of Mexico or

[162] Susan Herbst, 'Surveys in the Public Sphere: Applying Bourdieu's Critique of Opinion Polls', *International Journal of Public Opinion Research*, 4.3 (1992), 220 – 229 (pp. 220 – 229).
[163] Richard R. Lau, 'An Analysis of the Accuracy of Trial Heat Polls During the 1992 Presidential Election', *The Public Opinion Quarterly*, 58.1 (1994), 2 – 20 (p. 2).
[164] Peter Lynn and Roger Jowell, 'How Might Opinion Polls be Improved? The Case for Probability Sampling', *Journal of the Royal Statistical Society*, 159.1 (1996), 21 – 28 (pp. 21 – 28).
[165] Andrew Gelman and Gary King, 'Why Are American Presidential Election Campaign Polls So Variable When Votes are so Predictable?', *British Journal of Political Science*, 23.4 (1993), 409 – 551 (p. 419).
[166] Ibid.
[167] Ibid., pp. 419 – 420.
[168] Ibid., p. 420.
[169] Ulises Beltrán and Marcos Valdivia, 'Accuracy and Error in Electoral Forecasts: The Case of Mexico', *International Journal of Public Opinion Research*, 11.2 (1999), 115 – 134 (p. 115).

El Salvador,[170] though held that authoritarianism undoubtedly affected Nicaraguan polling accuracy.[171] Instead, they noted that a combination of poll- and election-level factors were responsible for polling error in the Mexican elections of 1994 and 1997. Specifically, they contended that problematic sampling, issues of non-response, and the allocation of undecided voters combined with the low information nature of Mexican elections and their inherently uneven campaigns to bring about polling inaccuracy.[172] In so doing, they not only demonstrated the differing effect of election-level sources of error between countries, but also their importance in combination with poll-level factors.

Although the prominence of election-level enquiry was rising in the wider polling literature, assessments of the failure of polls to predict the 1997 French legislative election unanimously adopted poll-level understandings of the misprediction. Prior to election day, polling had indicated rising support for the ruling right-wing majority amongst the French electorate,[173] with *Le Monde* reporting that there was no doubt that they would form the new government.[174] This confidence was ultimately misplaced, as the Socialist Party emerged from the election as the largest party in direct contradiction to predictions.[175] Analyses into the predictive failure identified issues of sampling, an undetected late swing in voting intention, and widespread item non-response on questions of voting intention as substantial determinants of polling error.[176]

The following year saw polls fail to predict another Francophone contest: the Quebec general election in Canada.[177] During the last week of the campaign, polls unanimously and confidently

[170] Ibid., p. 132.
[171] Ibid., p. 120.
[172] Ibid., p. 132.
[173] Paul Hainsworth, 'The Return of the Left: the 1997 French Parliamentary Election', *Parliamentary Affairs*, 51.1 (1998), 71 – 83 (p. 72).
[174] Bruno Jérôme, Véronique Jérôme, and Michael S. Lewis-Beck, 'Polls Fail in France: Forecasts of the 1997 Legislative Election', *International Journal of Forecasting*, 15 (1999), 163 – 174 (p. 163).
[175] Hainsworth, p. 81.
[176] Jérôme, Jérôme, and Lewis-Beck, p. 173.
[177] Durand and Blais, pp. 43 – 47.

predicted that Parti Québécois would be victorious with a margin of at least five points, only to find the Liberal Party with a plurality of votes come election day.[178] Assessments of the misprediction once again predominantly adopted a poll-level approach to understanding the miss. Unforeseen late swing in voting intention, the overconfidence of polling projections resulting in the widespread abstention of Parti Québécois supporters,[179] problematic sampling frames, issues of non-response,[180] and an inherent polling bias against the Liberal Party were all identified as likely causes of the predictive failure.[181]

While the dominance of the poll-level approach was plain to see, one election-level factor was addressed in post-election assessments of the 1998 Quebec general election: turnout. Differential turnout was believed that have occurred as an artefact of the overconfidence of the polls. The overconfidence of their predictions regarding the Parti Québécois was argued to have made its supporters less likely to turnout than their Liberal Party counterparts. Due to its ability to alter the voting population on election day significantly from polling projections, this differential turnout was proposed as a key source of polling error.[182] Although it was a plausible determinant of error, the impact of differential turnout on the election result was questioned in subsequent literature.[183]

[178] Ibid., p. 43.
[179] Ibid.
[180] Claire Durand, André Blais, and Sebastien Vachon, 'Accounting for Biases in Election Surveys: The Case of the 1998 Quebec Election', *Journal of Official Statistics*, 18.1 (2002), 25 – 44 (p. 28).
[181] Claire Durand, 'Are Polls Biased Against Quebec Liberals?', *Policy Options Montreal*, 23.3 (2002), 51 – 56 (p. 51).
[182] Durand, Blais, and Vachon, p. 28.
[183] James P. Allan, Marc J. O'Reilly, and Richard Vengroff, 'The Election Everybody Won? The Impact of Party System Change, Voter Turnout, and Strategic Voting in the 1998 Quebec Election', *American Review of Canadian Studies*, 30.4 (2000), 497 – 519 (p. 497).

The third wave post-election assessments of polling inaccuracy continued in 2002 in the wake of a trio of mispredictions in Ireland,[184] France,[185] and Hungary.[186] Since the polling failure in Quebec four years previously, the literature had once again predominantly focused on poll-level sources of error. Design effects between polling organisations,[187] over-inflated turnout projections,[188] differential refusal,[189] the distortion of polling mechanisms due to declining response rates,[190] and undecided voter allocation methods were variously identified as crucial determinants of polling inaccuracy.[191]

Despite the dominance of poll-level assessments of inaccuracy, two pieces of literature focusing on election-level sources of error were published during this time. The first was produced by Endersby, Galatas, and Rackaway.[192] Therein they tested the hypothesis that slimmer election margins should motivate turnout and therefore affect electoral outcomes, noting the lack of literature on the relationship. Through an analysis of the 1993 and 1997 Canadian federal elections, they found that the closeness of the elections did affect levels of voter participation, concluding that it was of importance in the determination of election

[184] Gail McElroy and Michael Marsh, 'Why Opinion Polls Got It Wrong in 2002', in *How Ireland Voted 2002*, ed. by Michael Gallagher, Paul Mitchell, and Michael Marsh, (Basingstoke: Palgrave Macmillan, 2003), p. 159.
[185] Claire Durand, André Blais, and Mylène Larochelle, 'The Polls in the 2002 French Presidential Election: An Autopsy', *Public Opinion Quarterly*, 68.4 (2004), 602 – 622 (p. 602).
[186] Tamás Bodor, 'Hungary's "Black Sunday" of Public Opinion Research: The Anatomy of a Failed Election Forecast', *International Journal of Public Opinion Research*, 24.4 (2012), 450 – 471 (pp. 450 – 471).
[187] Erikson and Wlezien, 'Presidential Polls as a Time Series', p. 163.
[188] R. F. Belli and others, 'Reducing Over-reporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring', *Public Opinion Quarterly*, 63.1 (1999), 90-108 (pp. 90-108).
[189] P. J. Brown, D. Firth, and C. D. Payne, 'Forecasting on British Election Night 1997', *Journal of the Royal Statistical Society*, 162 (1999), 211 – 226 (p. 211).
[190] Barry Burden, 'Voter Turnout and the National Election Studies', *Political Analysis*, 8.4 (2000), 389 – 398 (pp. 389 – 398).
[191] Penny S. Visser and others, 'Improving Election Forecasting: Allocation of Undecided Respondents, Identification of Likely Voters, and Response Order Effects', in *Election Polls, the News Media, and Democracy*, ed. by P. Lavrakas and M. Traugott, (New York: Chatham House, 2000), p. 256; Traugott, 'Assessing Poll Performance in the 2000 Campaign', p. 394.
[192] J. W. Endersby, S. E. Galatas, and C. B. Rackaway, 'Closeness Counts in Canada: Voter Participation in the 1993 and 1997 Federal Elections', *The Journal of Politics*, 64.2 (2002), 610 – 631 (pp. 610 – 631).

outcomes.[193] As the outcomes of elections represent the results that polls attempt to predict and the figures against which they are judged, by the reckoning of Endersby and company, the degree of marginality should therefore affect polling accuracy.

The second election-level work was that of Schaffner and Streb concerning the impact of partisanship.[194] Through an analysis of low-information elections within the United States, they concluded that polling predictions rendered in elections characterised by stronger partisan loyalties are more accurate than those conducted in elections characterised by weaker partisanship.[195]

Despite the growth of election-level analyses prior to their occurrence, assessments of the trio of mispredictions in 2002 predominantly adopted poll-level approaches to understanding the misses. While assessments of the overestimation of the Fianna Fáil seat share in the Irish general election dismissed the impact of a late swing in voting intention, errors in turnout prediction, and non-response bias, they concluded that the failure to draw representative samples,[196] along with constituency-level sampling error accounted for the miss.[197]

Assessments of the underestimation of the National Front by polls in the French presidential election also adopted a poll-level approach, citing issues of question wording, the use of quota sampling, an over-reliance on landline-based interviews, the statistical adjustments made by pollsters, the high number of non-disclosures, and the impact of social desirability bias on response validity as the likely determinants of error.[198]

---

[193] Ibid.
[194] Schaffner and Streb, p. 562
[195] Ibid., p. 578.
[196] McElroy and Marsh, p. 164 – 166.
[197] Pat Lyons, 'Public Opinion in the Republic of Ireland – 2002', *Irish Political Studies*, 18 (2003), 6 – 23 (p. 10).
[198] Durand, Blais, and Larochelle, pp. 609 – 617.

The poll-level theme continued in assessments of the Hungarian 'Black Friday' polling miss. Following twelve years of polling accuracy,[199] the victory of the Socialist Party over Fidesz in 2002 came as a universal shock.[200] Post-election polling assessments identified poor-quality samples, inadequate likely voter models, differential non-response, and a failure to recognise a late swing in voting intention as the primary sources of error.[201]

So, while the third wave of post-election assessments of polling inaccuracy was generally characterised by a greater focus on the importance of election-level factors than previous waves, widening the scope of analyses into polling error and bringing thoughts of broader causes of error into the mainstream, it ended with an apparent reversion to the mean. In spite of the work of Buchanan, Crewe, and Smith, as well as the conclusions drawn from elections in Quebec and Nicaragua, three prominent post-election assessments of polling error reverted back to well-established poll-level understandings of misprediction. Despite this, the foundation for election-level enquiry established by the third wave of literature would quickly be built on by scholars, allowing in-depth election-level assessments of polling inaccuracy to emerge within a fourth wave of literature.

The Fourth Wave of Post-election Assessments of Polling Inaccuracy

In addressing polling accuracy across Portuguese general elections, Magalhães identified the importance of a range election-level factors.[202] In so doing, he catalysed the fourth wave of post-election analyses within the literature which was characterised by a deeper focus on the importance of electoral context for polling error, leading to the slow normalisation of the

[199] Csaba Nikolenyi, 'Strategic Co-ordination in the 2002 Hungarian Election', *Europe-Asia Studies*, 567 (2004), 1041 – 1058 (p. 1041).
[200] Kenneth Benoit, 'Like Déjà Vu All Over Again: The Hungarian Parliamentary Elections of 2002', *Journal of Communist Studies and Transition Politics*, 18.4 (2002), 119 – 133 (p. 119).
[201] Bodor, pp. 450 – 452.
[202] Pedro C. Magalhães, 'Pre-election Polls in Portugal: Accuracy, Bias, and Sources of Error', *International Journal of Public Opinion Research*, 17.4 (2005), 399 – 421 (p. 399).

approach. He held that ‘errors seem to be caused by specific features of elections themselves’ and proceeded to identify a series of election-level factors that profoundly affected the accuracy of Portuguese polls.[203] Chief amongst these were volatile turnout levels and the closeness of the electoral contests which were found to account for polling error, even when controlling for poll-level considerations, such as sample size, sampling design, or survey mode.[204] While previous studies has identified the importance of election-level factors, or stated that election-level factors matter *in tandem* with poll-level influences, Magalhães was amongst the first to demonstrate empirically that election-level factors existed as significant determinants of polling error, even when controlling for more conventionally recognised, poll-level causes.

In 2006, polls severely overestimated the margin of victory of the left-wing coalition in the Italian general election.[205] In the four years since the trio of polling misses in 2002, the wider literature continued to focus on poll-level determinants of predictive error. Particular attention was paid to issues of sampling, likely voter screening,[206] house effects,[207] and the allocation of undecided voters as determinants of inaccuracy.[208]

Despite the election-level findings of Magalhães, assessments of the 2006 Italian polling failure adopted a largely poll-level approach. Analyses identified the importance of coverage error, sample size, issues of non-response, and the failure to screen likely voters as the predominant

[203] Ibid., p. 418.
[204] Ibid.
[205] James L. Newell, ‘The Italian Election of May 2006: Myths and Realities’, *West European Politics*, 29.4 (2006), 802 – 813 (p. 802); Geoff Andrews, ‘The Italian General Election of 2006’, *Representation*, 42.3 (2006), 253 – 260 (p. 257).
[206] Jay DeSart and Thomas Holbrook, ‘Campaigns, Polls, and the States: Assessing the Accuracy of State-wide Presidential Trial-heat Polls’, *Political Research Quarterly*, 56.4 (2003), 431 – 439 (pp. 431 – 439).
[207] Michael W. Traugott, ‘The Accuracy of the National Pre-election Polls in the 2004 Presidential Election’, *Public Opinion Quarterly*, 69.5 (2005), 642 – 654 (p. 643).
[208] António José da Cruz Belo, ‘Accuracy of Polls in Portugal’, paper delivered at the WAPOR 58th Annual Conference (2005), pp. 13 – 14.

sources of polling error.[209] While their focus lay primarily on poll-level determinants of error, post-election assessments identified one significant country-level factor: the imposition of a moratorium on polling by the state.[210] Italy puts in place a particularly long two-week moratorium on the publication of polls prior to election day.[211] As the proximity of a poll to election day had long been understood to be a determinant of accuracy,[212] analyses held that the inability to conducted polls in the closing weeks of the election necessarily affected accuracy. Moreover, they contended that, as the moratorium prevented polls from being able to capture late decision-making amongst the electorate, it increased the likelihood of an undetected late swing in voting intention.[213] In so doing, not only did they recognise the importance of country-level factors as drivers of polling error, but they drew attention to the importance of late decision-making as an impactful source of error – a source of error that varies between elections, situating it at the election-level – but did not explicitly frame it in these terms.

Though the recognition of the impact of the polling moratorium was encouraging, analyses of the polling failure in the 2006 Italian general election still predominantly adopted poll-level approaches. In this sense, the beginning of the fourth wave of post-election assessments of polling inaccuracy closely resembled the third wave, with election-level enquiries existing as accepted approaches, but being wholly overshadowed by their poll-level counterparts. However, the fourth wave would see the dominance of the poll-level approach wane with the

[209] Mario Callegaro and Gaincarlo Gasperoni, 'Accuracy of Pre-election Polls for the 2006 Italian Parliamentary Election: Too Close to Call', *International Journal of Public Opinion Research*, 20.2 (2008), 148 – 170 (pp. 165 – 166); Laura Fumagalli and Emanuela Sala, 'The Total Survey Error Paradigm and Pre-election Polls: The Case of the 2006 Italian General Elections', *ISER Working Paper Series*, 29 (2011), 1 – 41 (p. 32).
[210] Callegaro and Gasperoni, pp. 162 – 164.
[211] Tom W. Smith, 'Freedom to Conduct Public Opinion Polls Around the World', *International Journal of Public Opinion Research*, 16 (2004), 215 – 233 (p. 216).
[212] Crespi, pp. 4 – 5; DeSart and Holbrook, p. 432; Lau, p. 2.
[213] Callegaro and Gasperoni, p. 163.

publication of an increased number of election-level works, several of which were prominently positioned.

The new decade ushered in the next wave of post-election analyses as polls overestimated the vote share received by the Liberal Democrats in the 2010 UK general election.[214] Though the fourth wave would eventually see the dominance of the poll-level approach challenged through the normalisation of election-level discussions, the literature penned since the Italian polling failure four years previously focused primarily on poll-level sources of error. The importance of the lead time between a poll and election day, the misallocation of undecided voters,[215] issues of sampling, problematic in-house adjustment mechanisms,[216] inadequate likely voter estimation, and differing survey modes were all variously addressed.[217] Despite the rise of election-level works in the mid-2000s, it is notable that no such works were produced between the mispredictions of 2006 and 2010.

The dominant focus on poll-level sources error was reinforced by the assessment of the failure of polls to accurately predict the 2010 UK general election by Pickup and others.[218] They dismissed the influence of unrecognised late swing, but could not wholeheartedly conclude that the polling miss was the result of methodological deficiencies within polling.[219] Despite this uncertainty, they ultimately urged future analysis to focus on the impact of house effect and the variation of methodologies between polling organisations.[220]

---

[214] Pickup and others, pp. 179 – 209.
[215] Chien-chou Su and Mandy Sha, 'Poll Accuracy Measures in a Quasi-two-way Election: An Empirical Assessment of Taiwan's 2000 Presidential Election', *NPUST Humanities and Social Sciences Jikan*, 19.4 (2007), 505 – 527 (pp. 524 – 525).
[216] Durand, 'The Polls of the 2007 French Presidential Campaign', p. 296.
[217] Costas Panagopoulos, 'Pre-election Poll Accuracy in the 2008 General Elections', *Presidential Studies Quarterly*, 39.4 (2009), 896 – 907 (p. 906).
[218] Pickup and others, p. 200.
[219] Ibid.
[220] Ibid., pp. 200 – 201.

Though analysis of the 2010 UK polling miss saw reversion to poll-level explanations of error, in the following year Cosciug produced a novel, election-level study of polling accuracy within Romania.[221] As Romanian parliamentary and presidential elections operate under different electoral systems – proportionally representative and majoritarian, respectively[222] – he tested their variable effect of these two systems on polling error. In his analysis, he found that polling error was substantially more pronounced under the majoritarian system than it was under proportional representation.[223]

While the work of Cosciug contributed to the establishment of election-level analyses within the polling literature, the 2013 analysis conducted by Vignati and Gasperoni further normalised the recognition of extra-methodological factors as determinants of polling error. Through an assessment the predictive performance of pre-election polls in relation to Italian regional and senate elections in 2010 and 2013, they concluded that the methodological differences between polling organisations was not a significant determinant of predictive efficacy.[224] Through doing so, they acknowledged the need to look beyond poll-level differences and broaden the scope of analyses of polling error.

The growing election-level focus of assessments of polling error was further evident in the following year with the work of Wright, Farrar, and Russell. Through an analysis of the accuracy of pre-election polling in New Zealand, they questioned whether non-sampling error varied between different voting populations and electorates, or was consistent throughout pre-election polling efforts, opening the door for the analysis of inter-election variation.[225]

---

[221] Anatolie Cosciug, 'The Influence of the Electoral System on the Quality of Electoral Predictions: Romanian Elections: 1992 – 2009', *Studia UBB Politica*, 1 (2011), 79 – 118 (pp. 79 – 118).
[222] Ibid., pp. 92 – 93.
[223] Ibid., pp. 113 – 114.
[224] Rinaldo Vignati and Giancarlo Gasperoni, 'The Predictive Ability of Pre-election Polls in Italy: A Regional Focus', *Statistica Applicata – Italian Journal of Applied Statistics*, 25.3 (2013), 287 – 302 (p. 301).
[225] Malcom J. Wright, David P. Farrar, and Deborah F. Russell, 'Polling Accuracy in a Multiparty Election', *International Journal of Public Opinion Research*, 26.1 (2014), 113 – 124 (p. 122).

In 2015, Coletto and Breguet assessed the accuracy of Canadian pre-election polls, addressing the nine provincial elections held between 2011 and 2013.[226] In their analysis, not only did they find that methodological aspects of polling were crucial determinants of accuracy, including sample size and survey mode, but also that differences between elections were strong predictors of polling error.[227] Specifically, they found that absolute change in voter turnout between elections and the change in the percentage distribution of vote share between elections were both significant determinants of polling error.[228]

In the following year, the British Polling Council published a report into the misprediction of the 2015 UK general election.[229] The report systematically addressed an exhaustive array of prospective causes for the polling miss.[230] Of these causes, it concluded that the unrepresentative sampling techniques used by polling organisations were responsible for the predictive failure.[231] While its analysis predominantly centred on poll-level causes of error,[232] the report entertained explanations that were election-level in nature, such as the effect of late swing (indicating the presence of last minute decision-making within the electorate) and marked differences in turnout between contests, though found little evidence of their impact.[233] That a significant post-mortem analysis into polling failure encapsulated election-level explanations represented significant progress from the normalisation of the approach within the literature.

[226] David Coletto and Bryan Breguet, 'The Accuracy of Public Polls in Provincial Elections', *Canadian Political Science Review*, 9.1 (2015), 41 – 54 (pp. 41 – 54).
[227] Ibid., p. 41.
[228] Ibid.
[229] Sturgis and others, pp. 757 – 781.
[230] Ibid., pp. 766 – 777.
[231] Ibid., p. 777.
[232] Ibid., pp. 757 – 781.
[233] Ibid.

Following the report by the British Polling Council, Mellon and Prosser conducted further analysis into the failure of the polls in 2015, emphasising the underestimation of the Conservative Party vote share and the overestimation of that of the Labour Party.[234] In focusing on unrepresentative sampling, unforeseen late swing, the elicitation of untruthful responses, and the disproportionate movement of undecided voters confounding the allocation strategies used by polls,[235] this post-election assessment largely conformed to the dominant poll-level focus. However, through addressing the possibility of levels of turnout in the election to undermine the turnout projection mechanisms on which polls rest, Mellon and Prosser recognised the potential for election-level characteristics to bear on polling error.

In assessing the prospective causes of error identified at the poll and election levels, Mellon and Prosser found little evidence that unrecognised late swing, unforeseen turnout anomalies, respondent untruthfulness, or uniform decision-making on the part of undecided voters were major determinants of the polling failure. Instead, they laid significant blame at the feet of problematic sampling and weighting procedures, holding that weighting to population targets without correcting for turnout was the primary cause of polling inaccuracy.[236] Therefore, while their analysis focused on a greater number of poll-level factors than it did election-level determinants of polling error, their final conclusion clearly pointed to the importance of election-level characteristics as drivers of polling error. Specifically, they explicitly recognised the potential for the levels of turnout exhibited in an election to undermine the weighting procedures on which polls rest, leading to prediction error.

[234] Mellon and Prosser, pp. 661 – 687.
[235] Ibid., pp. 662 – 664.
[236] Ibid., p. 683.

Figure 2: Echoes of 'Dewey defeats Truman' – on the eve of the 2016 U.S. presidential election, Hillary Clinton signs one of the 125,000 issues of Newsweek distributed during the closing stages of the campaign amid the air of inevitability afforded to her victory by pre-election polls. These copies were later recalled due to her shocking loss to Donald Trump.[237]

Pre-election polls unanimously agreed that Hillary Clinton was likely to win the 2016 US presidential election. Such was the confidence of polls, that many poll-based models placed Clinton's chances of winning at over 90%.[238] From this, the outcome of the election was widely treated as a foregone conclusion (see: Figure 2). Ultimately, the pre-election polls missed the mark and Donald Trump was elected the 45th president of the United States.

The failure of polls to predict the election of Trump, saw a prominent focus on methodological re-evaluation within the literature. In the immediate aftermath of the misprediction, post-

[237] Rob Crilly, *Newsweek Recalls 125,000 Copies of its Souvenir Madam President Issue* (2016), <https://www.telegraph.co.uk/news/2016/11/10/newsweek-recalls-125000-copies-of-its-souvenir-madam-president-i/> [accessed 02/08/2020].

[238] Sam Wang, *Looking Ahead* (2016), <https://election.princeton.edu/articles/looking-ahead/> [accessed 14/07/2022]; Maurice Tamman, *Clinton Has 90 Percent Chance of Winning: Reuters/Ipsos States of the Nation* (2016), <https://www.reuters.com/article/us-usa-election-poll-idUSKBN1322J1> [accessed 14/07/2022]; The New York Times, *Presidential Forecast: Hillary Clinton Has a 91% Chance to Win* (2016), <https://www.nytimes.com/newsgraphics/2016/10/18/presidential-forecast-updates/newsletter.html> [accessed 14/07/2022].

mortem analyses were conducted extensively online focusing on issues of non-response bias and undetected shy voters.[239] Later academic work also focused on methodological issues including the adjustment procedures underpinning herding,[240] issues of weighting, and an undetected late swing in voting intention.[241] In-line with this poll-level focus, the post-election enquiry into polling error commissioned by the American Association for Public Opinion Research concluded that unrepresentative sampling, specifically the over-representation of highly educated individuals, was largely to blame for the polling miss.[242]

In spite of the predominant poll-level focus, select works in the wake of the 2016 US presidential election furthered the slow normalisation of election-level analysis, positing that polling accuracy could fluctuate between elections as a function of late decision-making within the electorate or differing levels of turnout.[243] Whilst findings in relation to late decision-making were inconclusive, the turnout difference between 2012 and 2016 was deemed a significant determinant of polling error.[244] Moreover, Kennedy and others questioned whether the constellation of election-level factors that arose in the 2016 presidential election could be repeated in future elections, noting their significance in determining its outcome.[245]

With the election-level analyses conducted in the wake of the 2016 US presidential election, the fourth wave of post-election assessments of polling inaccuracy came to an end. Compared to preceding waves, the fourth wave was characterised by far greater engagement with election-

[239] Peter K. Enns and others, *Why the Polls Missed in 2016: Was it Shy Trump Supporters After All?* (2016), <https://www.washingtonpost.com/news/monkey-cage/wp/2016/12/13/why-the-polls-missed-in-2016-was-it-shy-trump-supporters-after-all/> [accessed 11/06/2020]; Harry Enten, *Shy Voters Probably Aren't Why the Polls Missed Trump* (2016), <https://fivethirtyeight.com/features/shy-voters-probably-arent-why-the-polls-missed-trump/> [accessed 11/06/2020].
[240] Costas Panagopoulos, Kyle Endres, and Aaron C. Weinschenk, 'Pre-election Poll Accuracy and Bias in the 2016 U.S. General Elections', *Journal of Elections, Public Opinion and Parties*, 28.2 (2018), 157 – 172 (p. 170).
[241] Courtney Kennedy and others, 'An Evaluation of the 2016 Election Polls in the United States', *Public Opinion Quarterly*, 82.1 (2018), 1 – 33 (p. 30).
[242] Kennedy and others, p. 2.
[243] Panagopoulos, Enders, and Weinschenk, p. 169.
[244] Kennedy and others, p. 30.
[245] Ibid.

level explanations of polling error, leading to the slow normalisation of the approach within the literature. This continuing normalisation would see the assessment of election-level characteristics as determinants of polling error enter the academic mainstream in the fifth (and current) wave of literature.

The Fifth Wave of Post-election Assessments of Polling Inaccuracy

Signs that election-level analyses would begin to enter the academic mainstream began to emerge following the 2016 US presidential election. Indeed, while a great number of post-mortem analyses were produced in the wake of the failure of polls to predict the outcome of the 2017 UK general election which focused on issues surrounding the approaches to data adjustment employed by polls,[246] thereby conforming to the well-established poll-level understanding of error, others focused on election-level factors. In their comprehensive review of the general election, Cowley and Kavanagh identified a single culprit for the misprediction: the turnout projection mechanisms used by polling organisations.[247] They argued that turnout adjustments amplified polling error and that without them, the pre-election polls would have largely resembled the election results.[248] As such, they viewed the problem of misprediction as a problem born of the (in)ability of polls to accurately capture election-level characteristics, specifically the turnout levels that they exhibit.

Later academic work on the 2017 UK general election addressed the predictive success of the multi-level regression and post-stratification (MRP) model employed by YouGov.[249] The MRP model succeeded where many alternative attempts at prediction failed due to its ability to

[246] Peter Kellner, *General Election Polls 2017: How the Pollsters Got it Wrong* (2017), <https://www.standard.co.uk/news/politics/general-election-polls-how-the-pollsters-got-it-wrong-a3560936.html> [accessed 11/06/2020].
[247] Philip Cowley and Dennis Kavanagh, *The British General Election of 2017*, (London: Palgrave Macmillan, 2018), p. 271.
[248] Ibid.
[249] Will Jennings, 'The Polls in 2017' in *Political Communication in Britain*, ed. by D. Wring, R. Mortimore, and S. Atkinson, (London: Palgrave MacMillan, 2019), p. 212.

capture granular constituency effects – effects made particularly impactful by the first-past-the post electoral system employed within the United Kingdom[250] -- thereby placing the focus of predictive success on the importance of capturing country-level sources of error. As the MRP model employed by YouGov differed from the approaches used by other pollsters, its success also underscored the importance of decisions at the pollster-level as drivers of polling (in)accuracy.

Despite the seeming retrenchment of poll-level explanations of misprediction, the 2017 UK general election saw Simon Atkinson, the chief knowledge officer at Ipsos Mori, explicitly concede the importance of differing electoral contexts, and variable election-level factors, for polling error.[251] This concession came alongside the publication of several substantial assessments of the impact of election-level factors on polling error. Tudor provided suggestive evidence for the impact of an array of election-level factors on polling error within the United Kingdom,[252] while Luengo and Peláez-Berbel analysed the importance of electoral volatility (defined as the changeable nature of vote distributions between elections) and party-system fragmentation for polling error in a comparative, international context.[253] Whilst Tudor found statistically significant relationships between polling error and both election type and the effective number of electoral parties,[254] Luengo and Peláez-Berbell found electoral volatility to be a significant determinant of error, akin to the earlier work of Coletto and Breguet.[255]

In 2018, the inclusion of election-level factors within post-election analyses continued to establish itself within the psephological mainstream. Durand and others conducted a post-

[250] Ibid.
[251] Atkinson, p. 406.
[252] Tudor, p. 41.
[253] Óscar G. Luengo and Jaime Peláez-Berbell, 'Exploring the Accuracy of Electoral Polls During Campaigns in 2016: Only Bad Press?', *Contemporary Social Science*, 14.1 (2017), 43 – 53 (pp. 43 – 53).
[254] Tudor, p. 73.
[255] Luengo and Peláez-Berbell, p. 50.

mortem analysis of the failure of the polls to correctly predict the Chilean presidential election of the preceding year. They identified that Chilean commentators focused on well-established poll-level explanations for the miss, ranging from problematic question ordering to unrepresentative sampling.[256] However, they cited the potential importance of election-level factors as determinants of polling error, drawing attention to compulsory voter registration and a 15-day moratorium on polling prior to election day.[257] Ultimately they concluded that changes in electoral law between elections compounded issues of likely voter estimation and resulted in the misprediction.[258]

In the same year, Castillo-Manzano, López-Valpuesta, and Pozo-Barajas drew novel conclusions in their analysis of polling accuracy in the 2016 Spanish general election. They noted that polling error was not so much influenced by the number of parties in an electoral system, but rather the number of *new* parties for which no relevant past performance data is available.[259] Interestingly, they identified that this problem could not be resolved using traditional, methodologically focused approaches to the reduction of polling inaccuracy.[260] While the earlier work of Durand and Blais posited that poll- and election-level determinants of polling error were complementary, the work of Castillo-Manzano, López-Valpuesta, and Pozo-Barajas indicated that, in certain instances, the two factor types can be mutually exclusive, in keeping with the earlier findings of Magalhaes.

[256] Claire Durand and others, 'Report of the WAPOR Committee Reviewing the Pre-election Polls in the 2017 Presidential Election in Chile', *WAPOR* (2018), 1 – 39 (pp. 9 – 10).
[257] Ibid., p. 10.
[258] Ibid., p. 33.
[259] José I. Castillo-Manzano, Lourdes López-Valpuesta, and Rafael Pozo-Barajas, 'At a Time of Insurgent Parties, Can Societies Believe in Election Polls? The Spanish Experience', *Revista de Economía Aplicada*, 78.26 (2018), 81 – 97 (p. 93).
[260] Ibid.

2018 also saw the analysis of historical polling error by Jennings and Wlezien.[261] Motivated by a series of contemporaneous polling failures, and the general perception that polling error was increasing over time, they analysed the average error exhibited by over 30,000 polls from 351 general elections in over 45 countries. Not only did they identify that polling accuracy has not notably improved in 80 years,[262] but also that both the election type and electoral system exist as prominent determinants of polling error. Specifically, they found that error is generally lower in both proportionally representative systems and presidential elections.[263] They also posited that volatility between elections, either in terms of voters' allegiances or differing levels of turnout, could affect polling error in line with the earlier work of both Coletto and Breguet and Luengo and Peláez-Berbell.[264] Ultimately, giving its wide-ranging scope and the prestige of the journal in which it was published (*Nature Human Behaviour*), the work of Jennings and Wlezien helped to draw attention to election-level assessments of polling error and establish their position in mainstream literature.

In the same year, the need for greater emphasis on the role of election-level differences in the determination of polling error was recognised by Shirani-Mehr and others in their work on sources of bias and variance in polling estimates.[265] They identified that differences in error between elections varied to a greater extent than would be expected from sampling error alone,[266] finding substantial election-level bias in the polls studied.[267] While acknowledging the continued importance of poll-level effects on error,[268] they presented marked and unexpected changes in turnout between elections as a potential source of error,[269] arguing for

[261] Jennings and Wlezien, 'Election Polling Errors Across Time and Space', pp. 276 – 283.
[262] Ibid., p. 280.
[263] Ibid., p. 282.
[264] Ibid., p. 283.
[265] Shirani-Mehr and others, p. 607.
[266] Ibid.
[267] Ibid., p. 613.
[268] Ibid., p. 614.
[269] Ibid.

the impact of such election-level differences to be factored into the margins of error accompanying polls.[270]

While 2018 was notable for the strides taken towards an election-level view of sources of polling error, elements of the literature remained largely poll-level in their focus. Prosser and Mellon conducted their own review of historical polling accuracy in the UK and US, along with the purported causes of misprediction.[271] In-line with the work of Jennings and Wlezien, they found that in spite of high profile polling misses in recent years, the error exhibited pre-election polling had remained stable over time (and therefore not improved).[272] However, they largely attributed polling error to methodological factors, systematically identifying their role in past instances of misprediction.[273] Further literature unpacked a range of international polling failures from preceding years, drawing attention to issues of sampling, non-response, untruthful responses, and flawed survey questions as important causes of misprediction.[274] However, they conceded that instances of late swing in support for parties and candidates resulting from last minute shifts in decision-making among the electorate were of consequence for polling error.[275]

2019 saw pre-election polls incorrectly predict that the Labor Party would win the Australian federal election.[276] Given the scale of the miss – no polls correctly predicted the winner of the election[277] – it spurred a series of retrospective analyses. These analyses focused on the impact

[270] Ibid.
[271] Prosser and Mellon, pp. 757 – 790.
[272] Ibid., p. 780.
[273] Ibid, pp. 780 – 781.
[274] Ron S. Kenett, Danny Pfeffermann, and David M. Steinberg, 'Election Polls – A Survey, A Critique, and Proposals', *Annual Review of Statistics and its Application*, 5.1 (2018), 1 – 24 (pp. 12 – 14).
[275] Ibid., p. 14.
[276] Luke Mansillo and Simon Jackman, 'National Polling and Other Disasters', in *Morrison's Miracle: The 2019 Australian Federal Election*, ed.by Anika Gauja, Marian Sawer, and Marian Simms (Canberra: Australian National University Press, 2020), p. 184.
[277] Murray Goot, 'How Good are the Polls? Australian Election Predictions, 1993 – 2019', *Australian Journal of Political Science*, 56.1 (2021), 35 – 55 (p. 35).

of differences in methods between polling organisations,[278] herding within the polling industry,[279] and issues of survey mode.[280] Further assessments focused on the processes of data collection and weighting, arguing that issues cannot be adequately identified and remedied given the lack of transparency surrounding them.[281] As such, the focus on the methods and practical decisions that underpin polls again found itself at the forefront of attempts to understand a polling miss. However, election-level factors were not altogether absent in post-election assessments. Indeed, analyses published in the media during the immediate aftermath of the polling miss addressed potential issues of late decision-making amongst the electorate,[282] little evidence was found to support their effect on polling error.

Literature produced in 2019 also saw a series retrospective analyses of the misprediction of the 2016 US presidential election.[283] These assessments centred on the impact of undecided voters, finding that larger numbers of undecided voters coupled with inadequate approaches to their allocation accounted for a significant proportion of error.[284] They also perpetuated calls for poll-level methodological re-evaluation to reduce polling error, citing demographically unrepresentative samples at the state level as a significant determinant of misprediction, especially in key Electoral College states within the US.[285]

While many poll-level understandings of prediction error were present in the literature, 2019 also saw further recognition of the importance of electoral characteristics. Giuliani argued that

[278] Mansillo and Jackman, p. 136.
[279] Ibid., p. 142.
[280] Ibid., p. 145; Goot, 'How Good are the Polls?', p. 53.
[281] John Stirton, 'Why the Polls Got it Wrong', in *From Turnbull to Morrison: The Trust Divide*, ed. by Mark Evans, Michelle Grattan, and Brendan McCaffrie, (Melbourne: Melbourne University Press, 2019), pp. 310 – 313.
[282] Murray Goot, *Did Late Deciders Confound the Polls?* (2019), <https://insidestory.org.au/did-late-deciders-confound-the-polls/> [accessed 20/08/2022].
[283] Joshua J. Bon, Timothy Ballard, and Bernard Baffour, 'Polling Bias and Undecided Voter Allocations: US Presidential Elections, 2004 – 2016, *Journal of the Royal Statistical Society*, 182.2 (2019), 467 – 493 (pp. 467 – 493); Zeedan, pp. 84 – 101.
[284] Bon, Ballard, and Baffour, p. 483.
[285] Zeedan, p. 63.

differential turnout between elections served a key driver of polling error.[286] In assessing past US presidential elections, Kenett and Redman contended that the likely outcome of individual contests is contingent on characteristics inherited from the nature and duration of the existing administration, as well as from the candidates contesting it, especially if these candidates represent the continuation of a perceived political dynasty, noting that voters in the US are keen to avoid the concentration of power in the hands of a given party for extended periods of time.[287] Indeed, they recognised that, in elections that occur after one party has held the presidency for two terms, voters are more likely to support the opposing party.[288] They ultimately argued that, had polling organisations been cognisant of the impact of these election-specific characteristics in 2016, they would have realised that Trump was the probable victor and would, therefore, have been less likely to render inaccurate predictions.

In 2020, Durand and Blais adopted an interesting take on the performance of pre-election polls during the 2018 Quebec general election in Canada. They drew a distinction between a polling *miss* and a polling *failure*.[289] They defined a polling failure as being methodological in nature, namely a misprediction caused by poll-level methodological deficiencies. However, they defined a polling miss to be a misprediction due to volatile voting behaviour and its changeable nature between elections. This distinction was quite profound in terms of understanding poll-based misprediction, as it posited that it was perfectly possible for polls to be inaccurate despite not possessing any notable methodological deficiencies. This opened the door to a conceptualisation of polling error entirely removed from methodology, building on earlier works, such as that of Castillo-Manzano, López-Valpuesta, and Pozo-Barajas.

[286] Marco Giuliani, 'Making Sense of Pollsters' Errors. An Analysis of the 2014 Second-order European Election Predictions', *Journal of Elections, Public Opinions and Parties* 29.2 (2019), 162 – 178 (p. 169).
[287] Ron S. Kenett and Thomas C. Redman, *The Real Work of Data Science: Turning Data into Information, Better Decisions, and Stronger Organisations*, (Chichester: Wiley, 2019), pp. 9 – 10.
[288] Ibid.
[289] Durand and Blais, p. 133.

2020 also saw Sohlberg and Branham build upon the earlier election-level work of Jennings and Wlezien.[290] In assessing the impact of a range of election-level variables on polling error, they concluded that large shifts in party support from one election to the next increase the likelihood of error.[291] They ultimately argued that knowledge of shifts in support between elections ought to be used to inform the uncertainty surrounding poll-based predictions, tempering the confidence with which they are treated.

In the following year, Tudor and Wall took a substantial step towards establishing the importance of the election-level for polling error. Through a multi-level assessment of the impact of the election in which a poll was conducted on the error it exhibited, they were able to demonstrate that differences in characteristics between elections accounted for a significant proportion of polling error variance both internationally and within the UK.[292] From this, they concluded that characteristic differences between elections ought to be included within future analyses if polling error.[293]

Additional election-level findings were put forward in 2021 by Lloyd and Turgeon who concluded that elections characterised by larger numbers of undecided voters and lower information campaign environments were more prone to pronounced polling errors.[294] Moreover, in an assessments of sources of polling error in the 2020 US presidential election, Costas Panagopoulos identified the presence of a high proportion of late decision-making amongst voters in an election as a potential source of inaccuracy.[295] Similarly, in his assessment of polling in both the 2016 and 2020 US presidential elections, Gelman recognised the

[290] Sohlberg and Branham, pp. 1 – 13.
[291] Ibid., p. 13.
[292] Tudor and Wall, p. 1.
[293] Ibid.
[294] Ryan Lloyd and Mathieu Turgeon, 'Polling in New Democracies and Electoral Malpractice: The Case of Brazil', *International Journal of Public Opinion Research*, 33.4 (2021), 1039 – 1049 (p. 1039).
[295] Panagopoulos, 'Accuracy and Bias in the 2020 U.S. General Election Polls, p. 225.

potential for differential turnout and late decision-making between elections to affect polling error, acknowledging their variable impact across contests.[296]

Despite the continuing rise of election-level explanations for polling error, not all findings published in 2021 were positive. While acknowledging the potential for turnout to affect levels of polling error, along with recognising the lack of previous literature systematically testing the issue, Daoust found no evidence that the quality of poll-based predictions existed as a function of turnout across a wide-ranging international analysis.[297]

While literature published in 2021 saw election-level assessments of polling error firmly entrench themselves within the academic mainstream, poll-level understandings of misprediction continued to emerge. Works continued to investigate the impact of differing survey modes on polling error,[298] while additional post-mortems into polling in the 2020 US presidential election addressed issues of non-response.[299]

Literature published in 2022 saw poll-level determinants of error remain a prominent feature of analysis. Further assessments of sources of error in the 2020 US presidential election drew attention to missing demographics within samples and issues with likely voter modelling,[300] as well as issues of non-response as determinants of polling error.[301] Analyses of polling error in German federal elections also adopted poll-level approaches, identifying issues of survey

[296] Andrew Gelman, 'Failure and Success in Political Polling and Election Forecasting', *Statistics and Public Policy*, 8.1 (2021), 67 – 72 (p. 70).
[297] Jean-Francois Daoust, 'Blame it on Turnout? Citizens' Participation and Polls' Accuracy', *The British Journal of Politics and International Relations*, 23.4 (2021), 736 – 747 (p. 736).
[298] Claire Durand and Timothy P. Johnson, 'Review: What About Modes? Differences between Modes in the 21st Century's Electoral Polls across Four Countries', *Public Opinion Quarterly*, 85.1 (2021), 183 – 222 (p. 183).
[299] Ole J. Forsberg, 'US Election Polls: A Quick Postmortem', *Significance*, 18.1 (2021), 4 – 5 (p. 4).
[300] Natalie Jackson and Michael S. Lewis-Beck, 'Causes of 2020 Polling Error', in *Polarization and Political Party Factions in the 2020 Election*, ed. by Jennifer C. Lucas, Tauna S, Sisco, and Christopher J. Galdieri London: Rowman and Littlefield, 2022), p. 140.
[301] Joshua D. Clinton, John S. Lapinski, and Marc J. Trussler, 'Reluctant Republicans, Eager Democrats? Partisan Nonresponse and the Accuracy of 2020 Presidential Pre-election Telephone Polls', *Public Opinion Quarterly*, 86.2 (2022), 247 – 269 (p. 247).

design, the handling of untruthful responses, and the assignation of undecided voters as crucial determinants of error.[302]

In an analysis of polling error across Italian general elections, De Stefano, Pauli and Torelli adopted a multi-level approach to decomposing the impact of house effects as a function of time.[303] Despite adopting a multi-level approach to polling error, akin to that used by Tudor and Wall in the previous year,[304] by focusing on house effects, the work of De Stefano and company principally decomposed error into poll- and pollster-level groupings. Due to this, it did not include an election-based grouping level. Nevertheless, the use of a multi-level strategy betrays a growing recognition of the need to account for sources of error beyond the methodological underpinnings of polls.

Recognition of the impact of election-level factors in the determination of polling error also remained present in literature published in 2022. While predominantly focusing on the impact of question wording of polling error, analyses of polling performance in elections in Sweden and the Netherlands recognised the impact of the varying number of political parties between elections on the ease of poll-based prediction.[305]

Given the prominent rise in election-level assessments of polling error during the recent fifth wave of literature and the nascent stage at which much of this work finds itself, I contend that the study of pre-election polling error is primed for the comprehensive, foundational study of the impact of electoral characteristics that I provide through this thesis. In building upon the research addressed within this review, I identify two prominent gaps within the existing

[302] Alexander Bauer and others, 'Mundus Vult Decipi, Ergo Decipiatur: Visual Communication of Uncertainty in Election Polls', *PS: Political Science & Politics*, 55.1 (2022), 102 – 108 (p. 103).
[303] Domenico De Stefano, Francesco Pauli, and Nicola Torelli, 'Pre-electoral Polls Variability: A Hierarchical Bayesian Model to Assess the Roles of House Effects with Application to Italian Elections', *The Annals of Applied Statistics*, 16.1 (2022), 460 – 476 (p. 460).
[304] Tudor and Wall, p. 1.
[305] Wandi Bruine de Bruin and others, 'Asking About Social Circles Improves Election Predictions Even with Many Political Parties', *International Journal of Public Opinion Research*, 34.1 (2022), 1 – 11 (p. 1).

literature that need to be filled to fully substantiate an election-level understanding of polling error and address them in the following subsection.

Theoretical and Empirical Gaps in the Literature

While the present state of polling literature is such that election-level explanations of error are more common and widely accepted than they have been at any point previously, with the fifth wave of post-election assessments of polling inaccuracy seeing them firmly embedded within the mainstream, it remains an emergent and underdeveloped area of study when compared to its poll-level counterpart. Due to this, two prominent gaps exist within the literature that I fill through the production of this thesis.

The first of these gaps is theoretical in nature. Though the recognition of impact of election-level factors on polling error has been growing for the past ninety years, it still lacks a comprehensively elaborated theoretical basis. Though expectations of electoral heterogeneity have been provided in past research,[306] they are done so with little in the way of substantiation. Present understandings of electoral heterogeneity rest on assertions of compositional variance that remain unelaborated and underexplored.[307] Similarly, calls for further research into the impact of election-level differences on polling error within the literature fail to establish why these differences occur, or even to suggest why they might be likely.[308] Given these issues, my theoretical contribution frames the importance of adopting a theory of polling error that embraces electoral heterogeneity and provides the first comprehensive, phenomenon-level account of why this heterogeneity is likely. In so doing, I fill theoretical lacuna that is present in the literature and provide future works with a foundation to build upon and a framework to work within.

[306] Tudor and Wall, p. 5.
[307] Ibid.
[308] Sohlberg and Branham, p. 11

The second gap in the literature is empirical in nature. Though the investigation of election-level sources of polling error exists as a rising research focus, to date no studies have empirically assessed the benefit of adopting an election-level approach to the understanding of polling error. While prospective benefits have been proposed,[309] and calls for the adoption of an election-level understanding of polling error have been made,[310] presently no work has been conducted to identify the tangible benefits of adopting such an approach. To provide the beginnings of an evidentiary basis in support of the adoption of an election-level approach, I take the first steps towards identifying its empirical benefits for the study of polling error. Specifically, I identify election-level variables as useful predictors of the extent of polling error, allowing for a better understanding of those circumstances in which polls are more likely to offer erroneous predictions and, in so doing, allowing predictive expectations to be tempered ahead of election day.

In the following chapter, I address the theoretical gap in the literature by developing a theoretical framework for conceiving of elections as sources of polling error. To demonstrate the need for a re-orientation of how scholars approach the study of polling error, I address the disconnect between the dominant poll-level focus within the literature and the progression of polling error over time. I contend that this disconnect suggests that factors beyond the poll level affect polling error and that, due to their nature as phenomena and the impact of their characteristics on the projection mechanisms underpinning polling predictions, elections stand as likely determinants of error.

[309] Sohlberg and Branham, p. 11.
[310] Tudor and Wall, p. 1.

Chapter 3 – Beyond a Poll-level Understanding of Error: Conceiving of Elections as Sources of Polling Inaccuracy

"The assumption of constancy may adequately describe data from physical, natural, and most engineering related applications, but it fails to capture the essence of [social scientific] data which changes continually and is inherently unstable".[311]

- Spyros Makridakis (1981)

While attempts to conceive of polling error as a function of election-level characteristics are beginning to enter the mainstream literature, the empirical relationship between these characteristics and polling inaccuracy has yet to be fully explored and lacks a sound theoretical foundation. In this chapter, I establish a theoretical basis on which to reasonably assert that elections and their attendant characteristics exist as meaningful drivers of pre-election polling inaccuracy that warrant specific attention. In doing so, I address my first research question.

I approach this chapter from first principles and split it into four sections. In the first, to rationalise the discussion of inaccuracy, I establish that pre-election polls conducted in reasonable proximity to an election exist as future-orientated undertakings. I develop this point by situating them at the intersection of three approaches to understanding future outcomes: forecasting, projection, and prediction. With the ontological position of thesis regarding the nature of polls established, I move in the second section to conceptualise polling inaccuracy. I identify that in attempting to predict future electoral outcomes, pre-election polls provide three pieces of information: estimated vote share distributions, implied electoral outcomes, and estimates of the uncertainty surrounding their measures. I demonstrate that these pieces of information can be used to conceptualise polling inaccuracy as either a distributive, bounded,

[311] Spyros Makridakis, 'Forecasting Accuracy and the Assumption of Constancy', *Omega*, 9.3 (1981), 307 – 311 (p. 307).

or substantive consideration and break down the benefits and shortcomings of each of these approaches.

In the third section I unpack the way in which the inaccuracy exhibited by polls is understood under the dominant paradigm. I identify that polling inaccuracy is presently conceived of as a function of random and systematic errors at the poll level and explore the characteristics of both types of error. I illustrate that both random and systematic polling errors exist as products of the processes of measurement and representation that underpin pre-election polling through use of the total survey error framework. Following this, I demonstrate that the understanding of polling inaccuracy solely as a function of random and systematic errors at the poll level is insufficient by illustrating its incongruence with the historical performance of polls. Through an assessment of a series of possible explanations for this disconnect, I conclude that conceiving of polling inaccuracy solely as a function of random and systematic errors at the poll level presents an incomplete picture of its determinants. I therefore contend that factors beyond the poll level contribute to polling inaccuracy and posit electoral characteristics as plausible examples of such factors.

In the final section, I outline the theoretical rationale for why the characteristics possessed by elections can be expected to exist as determinants of polling inaccuracy. I begin by addressing the manner in which the predictability of phenomena is directly affected by the variables that they comprise. With reference to the Popperian clock-to-cloud continuum,[312] I illustrate that certain phenomena can be more clock- or cloud-like, and therefore more or less conducive to accurate prediction, on the basis of the characteristics that they comprise. I also demonstrate

[312] Karl R. Popper, *Objective Knowledge: An Evolutionary Approach*, (Oxford: Clarendon Press, 1972), pp. 207 – 208.

that the nature of characteristics as either constants or variables necessary influences and bounds their predictability.

With a framework in which to assess the predictability of phenomena established, I move to address how the determinants of variable predictability apply to elections ontologically. In so doing, I identify elections as compositionally heterogenous phenomena that, while exhibiting fundamentally similar ontologies, possess varying levels of clock- and cloud-like characteristics. I establish the importance of these compositional differences for polling error by unpacking not only their effect on the predictability of elections as phenomena, but also their likely impact on the projection mechanisms undergirding poll-based prediction that results from the incommensurably of these mechanisms with a heterogenous ontology of elections.

To illustrate the substantive plausibility of the proposed impact of electoral characteristics on polling error, I demonstrate that a range of past polling failures have occurred in electoral environments comprising characteristics that would be expected to affect error. I counted that the occurrence of error in the presence of these characteristics, while far from definitive proof in and of itself, reifies their potential as determinants of inaccuracy and further warrants their investigation.

Given the changeable nature of elections as phenomena, their expected impact on polling error, and the substantive plausibility of this impact in light of past polling failures, I form my first hypothesis: membership within different elections will affect the degree to which polls exhibit error. I proceed to test this hypothesis in later analysis.

3.1: Identifying and Bounding the Predictive Utility of Pre-election Polls

Pre-election polls exist as representations of the voting intention of a sample of respondents who are ideally representative of the larger voting population. As chapter two made evident, a vast literature surrounds the predictive inaccuracy of pre-election polls. Necessarily, any assessment of the predictive inaccuracy of polls is based on one key assumption: that pre-election polls are predictive of future electoral outcomes. To unpack the validity of this assumption and to provide it with a degree of bounding, two questions must be answered. Firstly, can polls be understood to be predictive of future election outcomes? Secondly, if they can be understood in this way, which polls can reasonably be treated as predictions? In the following sub-section, I take both of these questions in turn. I begin by identifying that only polls that successfully capture a constellation of factors that is representative of the environment at play on election day can reasonably be argued to be predictive of its outcome. Following this, I recognise that only polls conducted within reasonable proximity of election day are capable of satisfying this criterion. I situate this conclusion within the wider literature with which it agrees, providing the temporal bounds in which pre-election polls are generally taken to offer information sufficient to be predictive of electoral outcomes.

Identifying the Characteristics of Predictively Useful Polls

Before unpacking the degree to which polls can be considered predictive of electoral outcomes, it is important to note that, when discussing pre-election polls in the context of this thesis, I am referring specifically to voting intention polls conducted shortly prior to election day, as distinct from polls conducted years prior to elections, voter expectation polls, and other election-orientated polling, such as exit polls. It is also important to identify voting intention polls as a specific form of broader public opinion surveys. Public opinion surveys seek to make inferences about a target population by drawing a sample from it and eliciting responses from those included through the use of questionnaires. Such surveys can be used to measure the attitudes, beliefs, and behaviour of respondents across a wide range of subject matters in relation to either the past, present, or future. Voting intention polls exist as a subset of public opinion surveys. While they are based on the same fundamental methodology, voting intention polls conducted shortly prior to an election distinguish themselves from broader public opinion surveys through their predominant focus on the intended voting behaviour of respondents in an upcoming election and by virtue of the central characteristic of their output: an estimated vote share distribution for that election.

Voting intention polls typically ask voters the question, 'if the election were held today, for whom would you vote?'.[313] This question is designed to elicit the current voting intention of individuals.[314] Given this, it is generally accepted that results presented by pre-election polls are snapshots of public opinion at singular moments in time.[315] The snapshots of public opinion provided by polls lend themselves directly to time series analysis of political trends,[316] horse race coverage of campaigns,[317] as well as assessments of the impact of campaign events on public opinion and are widely used for these purposes.[318]

[313] Perry, p. 312.

[314] S. G. Kou and Michael E. Sobel, 'Forecasting the Vote: A Theoretical Comparison of Election Markets and Public Opinion Polls', *Political Analysis*, 12 (2004), 277-295 (p. 277).

[315] Mark Blumenthal, 'Polls, Forecasts, and Aggregators', *Political Science and Politics*, 47.2 (2014), 297 – 300 (p. 298); Simon Jackson, 'Pooling the Polls Over an Election Campaign', *Australian Journal of Political* Science 40.4 (2005), 499 – 517 (p. 500); Thomas E. Patterson, 'Of Polls, Mountains: U.S. Journalists and Their Use of Election Surveys', *Public Opinion Quarterly*, 69.5 (2005), 716 – 724 (p. 720); David L. Paletz and Others, 'Polls in the Media: Content, Credibility, and Consequences', *Public Opinion Quarterly*, 44 (1980),495 – 513 (p. 496); June Woong Rhee, 'How Polls Drive Campaign Coverage: The Gallup/CNN/USA Today's Coverage of the 1992 Presidential Campaign', *Political Communication*, 13.2 (1996), 213 – 229 (p. 213)..

[316] Robert S. Erikson and Christopher Wlezien, 'Presidential Polls as a Time Series: The Case of 1996', *The Public Opinion Quarterly*, 63.2 (1999), 163-177 (p. 163).

[317] C. Anthony Broh, 'Horse-Race Journalism: Reporting the Polls in the 1976 Presidential Election', *Public Opinion Quarterly*, 44.4 (1980), 514 – 529 (p. 514).

[318] Christopher Wlezien and Robert S. Erikson, 'Campaign Effects in Theory and Practice', *American Politics Research*, 29.5 (2001), 419 – 436 (p. 419); Daron R. Shaw, 'A Study of Presidential Campaign Event Effects from 1952 to 1992', *The Journal of Politics*, 61.2 (1999), 387 – 422 (p. 387).

While polls are widely used to capture snapshots of political opinion, especially in periods during which elections are not imminent,[319] the information on voting intention provided by polls is also used to predict electoral outcomes.[320] This necessarily entails the application of the information provided by polls to future outcomes. The ability for polls to be used in this manner is contingent on the conditions that they capture and, by extension, the time at which they are conducted. For polls to be reasonably considered predictive of electoral outcomes, the sentiment that they capture must be able to be mapped on to future voting behaviour defensibly. It is unreasonable to expect polls conducted significantly in advance of election day to meet this criterion. The reasoning for this concerns the related issues of the crystallisation of voting intention, the solidification of electoral environments, and the time horizons over which voters can be said to be meaningfully aware of elections. I take each of these in turn.

It is widely acknowledged that voting intention does not crystallise into sentiment that meaningfully reflects future electoral behaviour until the later stages of election cycles.[321] Before voters have been exposed to candidates, informed as to their policy agendas, and had their latent partisanship activated by campaigning, the sentiment that they provide to polls offers little useful information on their likely voting behaviour.[322] In the absence of these stimuli, the intentions voiced by voters in response to polls will not be reflective of their true preferences, as these would yet to have been activated. Therefore, polls must capture sentiment

[319] Christina Holtz-Bacha, 'Polls, Media and the Political System', in *Opinion Polls and the Media: Reflecting and Shaping Public Opinion* ed. by Christina Holtz-Bacha and Jesper Stromback (Basingstoke: Palgrave Macmillan, 2012), 267-281 (pp. 267-281).

[320] George Gallup, 'Polls and the Political Process – Past, Present, and Future', *Public Opinion Quarterly*, 29.4 (1965), 544 – 549 (pp. 544 – 549); Nicolas Sauger, 'Assessing the Accuracy of Polls for the French Presidential Election: The 2007 Experience', *French Politics*, 6 (2008), 116 – 136 (pp. 116 – 136); Philip E. Converse and Michael W. Traugott, 'Assessing the Accuracy of Polls and Surveys', *Science* 234.4780 (1986), 1094 – 1098 (pp. 1094 – 1098).

[321] Gelman and King, p. 409; Robert S. Erikson, Costas Panagopoulos, and Christopher Wlezien, 'The Crystallization of Voter Preferences During the 2008 Presidential Campaign', *Presidential Studies Quarterly*, 40.3 (2010), 482-496 (pp. 482-483).

[322] Gelman and King, pp. 409 – 410.

from sufficiently informed and primed respondents in order to provide information that can reasonably be said to be predictive of electoral outcomes.

In addition to exposure to the above stimuli, respondents must possess knowledge of the parties and candidates contesting an election to provide meaningful sentiment regarding them.[323] If respondents lack this knowledge, polls will capture sentiment from individuals who are being asked to make decisions between unknown choices, rendering it unreliable and unreflective of their eventual informed or activated voting preferences. To this end, polls must question respondents in environments that are sufficiently reflective of the choice they will face on election day to elicit sentiment that can reasonably be taken to be predictive of their voting behaviour.

In a similar, though more general, vein, polls must engage with respondents who are meaningfully aware of an upcoming election in order to provide measures of sentiment that reasonably pertain to it. While this captures both of the preceding points, insofar as respondents must be aware of the parties and candidates contesting an election and the issue positions that they represent, it also encompasses something more rudimentary. Even in the presence of campaign forces that inform, crystallise, and activate electoral preferences, a significant portion of voters often fail to decide upon their voting intention until late in the election cycle.[324] Though some of this late decision-making may be accounted for by strategic voting,[325] along with differences in the amount of candidate- or party-centric information that voters deem

[323] Alan I. Abramowitz, 'Visibility, Electability, and Candidate Choice in a Presidential Primary Election: A Test of Competing Models', *The Journal of Politics*, 51.4 (1989), 977-992 (p. 979).
[324] Simon Willocq, 'Explaining Time of Vote Decision: The Socio-Structural, Attitudinal, and Contextual Determinants of Late Deciding', *Political Studies Review*, 17.1 (2019), 53-64 (p. 53).
[325] S. Kirkpatrick, 'Political Attitudes and Behaviour: Some Consequences of Attitudinal Ordering', in *Political Attitudes & Public Opinion* ed. by D. Nimmo and C. Bonjean, (New York: David McKay Company, 1972), 386-404 (p. 386); G. A. Irwin and J. J. van Holsteyn, 'What are they waiting for? Strategic Information for Late Deciding Voters', *International Journal of Public Opinion Research*, 20.4 (2008), 483-493 (p. 483).

necessary to arrive at a decision,[326] it nevertheless speaks to the difficulties in decision-making presented by voters during portions of election cycles in which factors actively conspire to elicit it. It is therefore unreasonable to expect voters to form meaningful opinions regarding elections that have not only yet to solidify in terms of specifics, but that do not actively factor into or bear with any urgency or immediacy on their lives. Given this, polls must interrogate voting intention at a point in time at which voters are suitably aware of a given election in order to capture information that can be considered predictive of its outcome. This contention is explored further later in the thesis when the sample of polls for analysis is specified.

Public opinion is also likely to change over long enough time horizons due to the emergence of new parties or candidates;[327] the occurrence of endogenous shocks, such as political scandals,[328] or exogenous shocks, such as wars and economic crises;[329] and even changes to electoral and political systems.[330] Polls conducted too far in advance of election day therefore run the risk of capturing sentiment that is vulnerable to substantial change over time, reducing the degree to which it can reasonably be applied to future electoral behaviour. Due to this, in order to be considered meaningfully predictive of electoral outcomes, polls must be conducted in sufficient proximity to an election such that the voting intentions they capture can reliably

[326] Brian Brox and Joseph Giammo, 'Late Deciders in U.S. Presidential Elections', *American Review of Politics*, 30 (2009), 333-355 (p. 337).
[327] P. Ignazi, 'The Crisis of Parties and the Rise of New Political Parties', *Party Politics*, 2.4 (1996), 549-566 (p. 549).
[328] Robert D. Highfill, 'Effects of News of Crime and Scandal upon Public Opinion', *Journal of Criminal Law and Criminology*, 17.1 (1926), 40-103 (pp. 41-42); Arthur H. Miller, 'Sex, Politics, and Public Opinion: What Political Scientists Really Learned from the Clinton-Lewinsky Scandal', *PS: Political Science & Politics*, 32.4 (1999), 721-729 (p. 721).
[329] Shanto Iyengar and Adam Simon, 'News Coverage of the Gulf Crisis and Public Opinion: A Study of Agenda-setting, Priming, and Framing', *Communication Research*, 20.3 (1993), 365-383 (p. 365); Timothy Hellwig and Eva Coffey, 'Public Opinion, Party Messages, and Responsibility for the Financial Crisis in Britain', *Electoral Studies*, 30.3 (2011), 417-426 (p. 417).
[330] Christopher Wlezien and Stuart N. Soroka, 'Electoral Systems and Opinion Representation', *Representation*, 51.3 (2015), 273-285 (p. 273).

be applied to its outcome without fear of substantial change throwing these sentiments out of alignment with reality.

Each of these factors speaks to the importance of the time at which polls are conducted for the extent to which they can be considered predictive of electoral outcomes. For polls to provide information that successfully captures meaningfully informed and activated voting intentions, an array of electoral characteristics that is reasonably reflective of the situation faced by voters on election day, and is provided by respondents who are reliably aware of an upcoming contest, it is necessary for them to be conducted in relatively close proximity to the elections to which they relate. Indeed, agreement exists in the wider literature that only polls conducted relatively close to election day can reasonably be thought of as predictive of elections,[331] with predictive accuracy increasing as the time to election day decreases.[332] In the following sub-section, I unpack the span of time prior to election day over which polls can reasonably be thought of as predictive of electoral outcomes.

Isolating the Timespan Over Which Polls Are Predictively Useful

Much work has been dedicated to the study of the time prior to election day over which polls are capable of providing reliable predictions of electoral outcomes.[333] Assessments of election forecasts have provided a range of results, with findings varying between cases. For example, in the case of elections in the United Kingdom, forecasts have been found to usefully relate to electoral outcomes as far as six to twelve months ahead of election day.[334] In other cases, this

[331] Will Jennings, Michael Lewis-Beck, and Christopher Wlezien, 'Election Forecasting: Too Far Out?', *International Journal of Forecasting*, 36.3 (2020), 949 – 962 (p. 960); Sturgis and others, p. 765; Sohlberg and Branham, p. 6; Shirani-Mehr and others, p. 5; Jennings and Wlezien, p. 3; Tudor and Wall, p. 1.
[332] Edward Kaplan and Arnold Barnett, 'A New Approach to Estimating the Probability of Winning the Presidency', Operations Research, 51.1 (2003), 32-40 (pp. 32-40); Tudor, p. 19.
[333] Jennings, Lewis-Beck, and Wlezien, p. 950.
[334] Michael S. Lewis-Beck, 'Election Forecasting: Principles and Practice', *British Journal of Politics and International Relations*, 7 (2005), 145-164 (p. 151).

span narrows to three to six months.[335] While a degree of this variation is the result of differences between countries,[336] some of it is driven by the differential impact of political systems on the crystallisation of voting intention, as the electoral preferences of voters come into focus earlier in parliamentary systems than their presidential alternatives.[337] To this end, polls are able to more accurately predict the results of parliamentary elections farther out from election day than they are presidential contests.[338]

Though election forecasts have been found to be effective twelve months out from election day, on average, the vote share estimates provided by polls only begin to meaningfully correlate with election results 200 days prior to election day.[339] In the case of presidential elections, polls conducted two to three months ahead of election day have been found to be predictive of electoral outcomes, offering performance that is (reasonably) comparable to those conducted closer to the end of the campaign.[340] In the case of legislative elections, polls conducted up to five months prior to election day have been found to be reasonably predictive of electoral outcomes.[341] The difference in the timespans over which polls prove predictively useful is again driven by the fact that voting preferences crystallise sooner in presidential systems than they do in legislative systems.[342]

It is clear, then, that while polls conducted closer to election day yield more accurate predictions of electoral outcomes that those conducted farther out, with accuracy decreasing at a rate of one percentage point per month,[343] it is nevertheless possible for polls conducted

[335] Jennings, Lewis-Beck, and Wlezien, p. 950.
[336] Philip E. Converse, 'Of Time and Partisan Stability', *Comparative Political Studies*, 2.2 (1969), 139–71 (p. 139).
[337] Will Jennings and Christopher Wlezien, 'The Timeline of Elections: A Comparative Perspective', *American Journal of Political Science*, 60 (2016), 219 – 233 (p. 231).
[338] Jennings, Lewis-Beck, and Wlezien, p. 950.
[339] Ibid., p. 952.
[340] Ibid., p. 956.
[341] Ibid.
[342] Jennings and Wlezien, p. 231.
[343] Ibid., p. 951.

earlier in the election cycle to prove predictively useful. Specifically, polls conducted within three months of presidential elections and five months of legislative elections capture voting intention that can be used to effectively predict behaviour on election day. Ultimately, this provides a degree of bounding to the timespan over which polls can be taken to be reasonably predictive of electoral results. I return to the question of appropriately predictive timespans later in the thesis when specifying the sample of polls used for analysis.

If polls conducted in reasonable proximity to election day can be considered to be predictive of electoral outcomes, how can the estimates of future voting behaviour they offer be understood? Estimations of future outcomes take three primary forms: forecasts, projections, and predictions.[344] In the following sub-section, I unpack these approaches and situate polls conducted shortly prior to elections within them.

The Predictive Intersectionality of Polls

Despite existing as distinct approaches, forecasts, projections, and predictions are often referred to in interchangeable or contradictory ways.[345] In spite of this, it is possible to discern the characteristics that render them distinct. Through the identification of these characteristics, it is clear that pre-election polls exist at the conceptual intersection of the three approaches.

The approach which flows most directly from the discussion of pre-election polls as future-orientated undertakings is forecasting. Forecasts exclusively seek to foresee events before they occur.[346] In so doing, they conceive the future to be absolute and unknown. That is, the future is defined strictly as a point in time which occurs after the forecast has been made. Resultantly,

[344] Barbara Adam and Chris Groves, *Future Matters: Action, Knowledge, Ethics*, (Leiden: Brill, 2007), p. 25; Thomas L. Saaty and Luis G. Vargas, *Prediction, Projection and Forecasting*, (Boston: Kluwer Academic Publishers, 1991), p. 1.
[345] William Ascher, 'Political Forecasting: The Missing Link', *Journal of Forecasting*, 1 (1982), 227 – 239 (pp. 237 – 230); Adam Tsakalidis and others, 'Predicting Elections for Multiple Countries Using Twitter and Polls', *IEEE Intelligent Systems*, 30.2 (2015), 10 – 17 (p. 10).
[346] Lewis-Beck, p. 145.

the forecaster has no verifiable knowledge of the future outcome of interest, as it has yet to occur. Applying this to polls, and therefore understanding them to be forecasts, is dependent on the point at time at which they are conducted. If polls are conducted and analysed as snapshots of public opinion at a given moment in time, they necessarily do not seek to foresee future outcomes and therefore cannot be understood to be forecasts, as the future is of little interest to them. However, polls conducted shortly before election day capture voting sentiment that can be, and often is, used to foresee future electoral behaviour. When polls are used in this manner, the future outcome to which they are being applied, that of an upcoming election, necessarily occurs after they are conducted. Owing to this, they conceive of the future in absolute terms and seek to foresee events before they occur, thereby lending themselves to being understood as forecasts.

Though forecasting speaks to the focus of pre-election polls conducted shortly before election day, in practical terms, they rest on a series of projections. Projections concern the estimation of future outcomes through the identification and application of prevailing trends to foresee future outcomes.[347] The act of projection does not simply constitute atheoretical extrapolation of the present on to the future, rather, it is informed by theories concerning the evolution of variables and phenomena over time,[348] and is often the result of multivariate statistical modelling which form the platform on which trends are understood and reconstructed.[349] As the trends used in projection are being mapped on to outcomes forward in time to estimate as yet unknown outcomes, projections conceive of the future in the same absolute manner as forecasts. At their core, the estimated vote share distributions produced by pre-election polls

[347] Steven Camarota and Karen Zeigler, *Projecting the 2012 Hispanic Vote Shares Nationally and in Battleground States* (2012), <https://cis.org/sites/cis.org/files/projecting-hispanic-vote-2012.pdf> [accessed 29/08/2020].
[348] Ibid.
[349] Sean Jeremy Westwood, Solomon Messing, and Yphtach Lelkes, 'Projecting Confidence: How the Probabilistic Horse Race Confuses and Demobilises the Public', *The Journal of Politics*, 82.4 (2020), 1530-1544 (p. 1530).

are projections of the responses gathered from a sample of individuals onto a population of interest. Inherent within the probability sampling ideal on which polls rest is the understanding that a randomly selected sample allows for population-level inferences to be made with a stipulated degree of uncertainty. This rests on the assumption that the trends present in the preferences of this sample are representative of the target population and can therefore be extrapolated onto it successfully. Despite this, polls often fail to conform to the ideal of random sampling and, instead, rely on theories of voter behaviour and complex modelling to map voting intention on to the future electoral outcomes.[350] As such, in practical terms, polls conform closely to the tenets of projection, insofar as the estimated vote share distributions they produce rely, in some cases implicitly and in others more explicitly, on the projection of trends identified by theory and statistical modelling.

As pre-election polls are conducted and published prior to elections, the population of interest onto which samples are projected – those individuals who turn out to vote on election day – always resides in the future. Consequently, through the estimation of vote share distributions, the responses gathered by polls during their fieldwork dates are projected forward in time to estimate future behaviour. As attempts to estimate future outcomes, the accuracy of the vote share distributions produced by polls is therefore reliant on the continuation of the trends on which they are based. The duration for which these trends must hold decreases as election campaigns progress, producing well-understood improvements to estimative accuracy.[351]

Given that the target population of pre-election polls solely comprises those individuals who will vote on election day, efforts are taken to ensure that estimated vote share distributions are representative of this. As not all respondents who state that they intend to vote go on to do

[350] Fisher and others, p. 250.
[351] Jennings, Lewis-Beck, and Wlezien, p. 960; Erikson and Wlezien, p. 90.

so,[352] likely voter models are employed to identify those respondents most likely to turn out and responses are weighted accordingly.[353] While respondents are often classified as likely or non-voters through the question-based processes of screening and scaling, with non-voters excluded from samples,[354] likely voters are also identified on the basis of past voting behaviour.[355] Though issues of false recall are problematic,[356] respondents who attest to having voted in past elections are generally considered more likely to do so moving forward. As such, the inclusion of respondents within samples, is, at least partially, based on the projection of their past behaviour onto the present. Moreover, undecided voters that arise in samples are also often allocated to parties on the basis of their past voting behaviour,[357] again projecting past behaviour on to future actions. As the estimated vote share distributions produced by polls are based on these samples, and therefore encompass the approaches to inclusion and attribution applied to the individuals within them, they can be understood as projections which are themselves partially based on earlier instances of projection.

While the focus and mechanisms inherent within pre-election polls speak to their relationship with both forecasting and projection, their association with prediction is more complex. This complexity stems from the fact that prediction is the subject of a dual conceptualisation, taking both pragmatic and scientific forms.[358]

[352] Perry, 'Election Survey Procedures', p. 534.
[353] D. Sunshine Hillygus, 'The Evolution of Election Polling in the United States', *Public Opinion Quarterly*, 75 (2011), 962 – 981 (pp. 962 – 981); Ron Kenett, Danny Pfeffermann, and David Steingberg, 'Election Polls – A Survey, A Critique, and Proposals', *Annual Review of Statistics and Its Application*, 5 (2018), 1 – 24 (pp. 1 – 24).
[354] Gary Langer and Daniel M. Merkle, 'Likely Voter', in *Encyclopedia of Survey Research Methods*, ed. by Paul J. Lavrakas, (Thousand Oaks, Sage Publications, 2008), pp. 425 – 426.
[355] John Curtice and Nick Sparrow, 'The Past Matters: Eliminating the Pro-Labour Bias in British Opinion Polls', *International Journal of Market Research*, 52.2 (2010), 169 – 189 (p. 174).
[356] Hilde T. Himmelweit and others, 'Memory of Past Vote: Implications of a Study of Bias in Recall', *British Journal of Political Science*, 8.3 (1978), 365 – 375 (p. 365).
[357] Janet A. Hoek and Philip J. Gendall, 'A New Method of Predicting Voting Behaviour', *International Journal of Market Research*, 35.4 (1993), 1 – 14 (p. 3).
[358] Keith Dowding and Charles Miller, 'On Prediction in Political Science', *European Journal of Political Research*, 58 (2019), 1001 – 1018 (p. 1001).

Pragmatic prediction concerns the probabilistic prediction of future outcomes ahead of time and does not principally concern itself with testing a given theoretical understanding of a phenomenon.[359] By contrast, scientific predictions principally focus on testing the implication of a theoretical model – that *x* should logically lead to *y* – with predictions existing as the expected outputs of these models.[360] While a notional distinction exists between these two forms of prediction, they differ more in focus and application than they do in kind. Though pragmatic predictions do not necessary concern themselves with testing theoretical expectations, in practice, the reasonable prediction of future outcomes often relies, implicitly or explicitly, on statistical models that themselves capture interactions and association between variables driven by theory.[361] Similarly, while scientific prediction chiefly concerns itself with assessing the predictive utility of theories, in practice, the use of theories to predict outcomes will necessarily be subject to empirical calibration to optimise performance.[362]

In terms of application, much in the same way as both forecasts and projections, pragmatic predictions conceive of the future in absolute terms, as it necessarily occurs after they are rendered and, therefore, sits outside of the parameters of the prediction model. This stands in contrast to predictions of known phenomena conducted after the fact used to test and calibrate the performance of models.[363] Ultimately, the accuracy of pragmatic predictions is gauged via the proximity of predicted outcomes to actual events.[364]

---

[359] Ibid, pp. 1001 – 1002.
[360] Ibid.
[361] Alan I. Abramowitz, 'An Improved Model for Predicting Presidential Election Outcomes', *PS: Political Science and Politics*, 21.4 (1988), 843-847 (p. 843).
[362] Claudia Werker and Thomas Brenner, 'Empirical Calibration of Simulation Models', *ECIS Working Paper Series*, 200413 (2004), 1 – 30 (pp. 19 – 20).
[363] Simon Jackman and Gary N. Marks, 'Forecasting Australian elections: 1993, and all that', *Australian Journal of Political Science*, 29.2 (1994), 277-291 (p. 277); Michael S. Lewis-Beck and Tom W. Rice, 'Forecasting presidential elections: a comparison of naïve models', *Political Behavior,* 6.1 (1984), 9-21 (p. 9); Michael S. Lewis-Beck and Tom W. Rice, 'Forecasting U.S. House Elections', *Legislative Studies Quarterly*, 9.3 (1984), 475-486 (p. 475).
[364] Dowding and Miller, pp. 1001 – 1002.

By contrast, scientific predictions largely concern the logical implication of a theoretical model – that *x* should logically lead to *y* – with predictions existing strictly as the expected outputs of these models.[365] As such, they are partially explanatory in nature, seeking to account for the manner in which outcomes are brought about.[366] Crucially, scientific predictions do not conceive of the future in absolute terms. Rather, they treat the future as a relative concept. Put simply, scientific predictions are not necessary concerned with the unknown future. This is because, when conceived of in relative terms, the future can be defined in relation to a specified past time point. From the vantage point of the present, any event that occurs after another in the past does so in the relative future of the first event. Therefore, multiple relative futures exist to be predicted, as all points in time possess different futures relative to one another. This allows scientific predictions to test the veracity of theoretical relationships in relation to the relative futures of past timepoints which, necessarily, have already occurred and are known. Consequently, the future in question does not necessarily occur after a scientific prediction is rendered, nor does it necessarily sit beyond the parameters of prediction models. Rather, scientific predictions can be made of known phenomena in the past to test theories or better understand the mechanisms that brought them about.

Of the two conceptualisations of prediction, pre-election polls lend themselves most readily to pragmatic prediction. Not only do they conceive of the future in absolute terms, but they also seek to foresee election results ahead of time and are judged by the degree to which their predicted vote share distributions resemble these results. Despite this, pre-election polls can also be understood as tests of theories. At a prosaic level, polls represent tests of sampling theory. That is, their predicted vote share distributions test the ability of samples to provide

[365] Ibid.
[366] Keith Dowding, *The Philosophy and Methods of Political Science*, (New York: Palgrave Macmillan, 2016), pp. 36 – 67.

valid inferences about a target population. As these distributions are predicated on the extrapolation of prevailing trends, pre-election polls can also be seen as tests of the validity of projection as a means of foreseeing future outcomes. In this sense, pre-election polls also lend themselves to being understood as scientific predictions.

Despite this, pre-election polls are not explicitly explanatory in nature. While vote share distributions rest on both sampling theory and projection – thereby tacitly holding that the predicted election outcome will come about due to the validity and stability of inferences made from the intentions of a representative sample of respondents – polls do not directly address *why* these predicted outcomes will occur. Instead, such explanation is reserved for post-mortem analyses, undermining the extent to which polls can be understood to be scientific predictions. Nevertheless, to differing degrees, pre-election polls encapsulate the traits of both pragmatic and scientific prediction and can therefore be understood to be predictions.

Owing to their ability to be understood as forecasts, projections, and predictions, I contend that pre-election polls sit at the intersection of these three approaches to the estimation of future outcomes, encompassing traits associated with each approach simultaneously. Due to their foci and mechanisms, I hold that pre-election polls are best understood as future-orientated predictions that rest on a series of projections and understand them to exist as such moving forward.

With the predictive nature of pre-election polls and their relationship to future outcomes established, I move to address the ways in which the success of their attempts to foresee these outcomes can be conceptualised. I contend that the information provided by pre-election polls lends itself to three conceptualisations of accuracy: distributional, bounded, and substantive. In the following section, I unpack the nature of these conceptualisations and their utility in assessing the predictive performance of polls.

3.2: Marginality or Missing the Mark? Conceptualising Polling Inaccuracy

In attempting to predict electoral outcomes, pre-election polls provide three pieces of information: estimated vote share distributions, implied electoral outcomes, and estimates of the uncertainty surrounding their estimates. On the basis of these factors, polling inaccuracy lends itself to three forms of conceptualisation: in distributive terms, as a function of the disparity between predicted and actual vote share distributions; as a bounded issue concerning margins of error; and substantively, in relation to the ability of polls to predict substantive electoral outcomes.

In this section, I unpack these three conceptualisations into their sub-forms and identify their relative advantages and disadvantages. I identify that while the distributive conceptualisation of polling inaccuracy is the only approach with broad analytical tractability, it presents profound issues surrounding its disregard of substantive electoral outcomes, intolerance of random error, and the extent of its interpretability in isolation. While I note that these shortcomings can be resolved by employing alternative bounded and substantive conceptualisations, I demonstrate that these approaches possess even deeper issues concerning a lack of tractability and representativeness even when they are considered in combination.

Distributive Inaccuracy

As pre-election polls provide estimates of the likely vote share distribution on election day, their inaccuracy can be conceptualised as a function of the disparity between these estimates and official election results. Under this conceptualisation, any disparity between predicted and actual vote share distributions, irrespective of direction, constitutes inaccuracy. The severity of this inaccuracy is a matter of degree, increasing in tandem with the size of the disparity between distributions.

Conceptualising the inaccuracy of pre-election polls as a function of the disparity between predicted and actual vote share distributions possesses several advantages. It is even-handed in its judgement of inaccuracy, as it accounts equally for both over- and under-estimation. Moreover, not only does this make it possible to know whether a poll is incorrect in an absolute sense, but also allows for the severity of inaccuracy to be gauged relative to other polls. Most importantly, as all polls provide predicted vote share distributions and all elections provide finalised distributions in the form of results, it possesses universal analytical tractability.

Despite its benefits, conceptualising polling inaccuracy as a function of the disparity between predicted and actual vote share distributions possesses shortcomings. The most prosaic of these is the stringency of its singular necessary and sufficient condition. As any disparity between predicted and actual vote shares is considered to represent inaccuracy, it presupposes that the foremost aim of pre-election polls is not to correctly predict substantive electoral outcomes, but rather to predict vote share distributions. While the two may seem synonymous, they are in fact discrete considerations and in need of distinction. It is possible for one poll to fail to accurately predict the results of an election in terms of its vote share distribution, yet correctly predict its substantive political outcome. Likewise, another may predict a vote share distribution that closely resembles the result of an election yet fail to predict its substantive outcome. This was evidenced by past elections in France and the US, respectively. In the 2017 French presidential election, polls unanimously correctly called Macron as the victor of the second-round run-off but underestimated his vote share by an average of 10 percentage points.[367] Conversely, in the 2004 US presidential election, several polls correctly predicted Kerry's vote share, but incorrectly predicted that he would win the election.[368]

In much the same way as it is important to note the distinction between vote share distributions and substantive electoral outcomes, it is necessary to distinguish inaccuracy from bias. Inaccuracy concerns the disparity between predicted and actual vote share distributions in the case of both individual polls and aggregations. In the aggregate, inaccuracy may comprise instances of both over- and under-estimation simultaneously, thereby rendering it inherently directionless. By contrast, though bias also pertains to the disparity between predicted and actual vote share distributions, it cannot be applied to individual polls. Bias concerns consistent over- or under-estimation on the part of collections of polls. Therefore, it relates to the presence of estimative trends and can only exist in the aggregate. Moreover, as it rests on the existence of these trends, bias is inherently directional, relating to instances in which either over- *or* under-estimation is preponderant.

In addition to being conceptually distinct, inaccuracy and bias are also practically discrete. More specifically, an aggregation of polls may be inaccurate without being biased and vice versa. The vote share estimates provided by a series of polls may each vary from the finalised results in differing directions to differing extents, resulting in no overall directional bias despite the estimates remaining inaccurate. Moreover, the vote share estimates provided by distinct clusters of polls, such as those conducted by specific polling organisations, may each possess their own directional biases as a result of house effects or other systematic errors, but these biases may cancel one another out in the round, leading to no overall directional bias and

[367] Harry Enten, *Macron Won, But the French Polls Were Way Off* (2017), <https://fivethirtyeight.com/features/macron-won-but-the-french-polls-were-way-off/> [accessed 17/07/2020].
[368] Real Clear Politics, General Election: Bush vs. Kerry (2004), <https://www.realclearpolitics.com/epolls/2004/president/us/general_election_bush_vs_kerry-939.html> [accessed 17/07/2020].

rendering their mean estimate accurate. It is also possible for pre-election polls to contain internal biases and still render accurate vote share estimations, provided that these biases are not correlated with the outcome of interest. Owing to their distinct nature, without a consideration of bias, distributive inaccuracy cannot be held to be a rounded measure for evaluating misestimation. As such, the presence of consistent directionality within the disparities between predicted and actual vote share distributions must be considered to evaluate the presence of bias.

Whether it is concerned with the identification of inaccuracy or bias, focusing solely on the disparity between predicted and actual vote shares fails to fully account for the traits possessed by vote share distributions which are also important in the determination of polling inaccuracy. In addition to providing percentage estimates of support, vote share distributions also identify likely winners in the form of the party or candidate with the largest share of the vote and the anticipated government formation through the ordering of competitors. As such, they provide three possible sources of inaccuracy when considered relative to election results. Consequently, to discount these additional sources of disparity and solely focus on percentage point deviation is not only to provide a partial account of vote share distributions but, by extension, to offer an incomplete assessment of polling inaccuracy.

By solely focusing on the disparity between predicted and actual vote share distributions, the distributive conceptualisation of inaccuracy is incongruent with the reality of pre-election polling. All polls possess a margin of error – bounds within which their predictions are expected to lie due to random sampling error – and therefore encompass a degree of expected inaccuracy. Though the continuous conceptualisation does not account for this, it is central to a tolerance-based conceptualisation of polling inaccuracy.

Bounded Inaccuracy

The estimated vote shares provided by polls are accompanied by margins of error. These represent the random error associated with vote share estimations on the basis of their sample size and indicate, to a given degree of confidence, the number of percentage points surrounding the election result within which polling estimates are expected to be. For example, the margin of error with 95% confidence for a randomly sampled poll of sample size 800 is ±3.5 percentage points.[369] Consequently, the point estimates provided by polls possess bounds within which they are likely to fluctuate.

The existence of this likely fluctuation lends itself to the delineation of polling inaccuracy into two variants: expected and unexpected. Expected inaccuracy represents the likely fluctuation of predicted point estimates within their margins of error, while unexpected inaccuracy represents their deviation beyond these tolerance parameters. A bounded conceptualisation of polling inaccuracy can be constructed around these expected and unexpected forms of inaccuracy. Most intuitively, this bounded conceptualisation of polling inaccuracy can be understood in binary terms. If the error presented by a poll falls outside of its stated margin of error, it can be understood to be inaccurate. However, if the error presented is within the margin of error, a poll cannot be understood to be inaccurate, insofar as it was no more inaccurate than its bounds made clear. Bounded inaccuracy can also be understood continuously as a function of the degree to which the finalised error presented by a poll falls outside of its margin of error. Under this understanding, the severity of polling inaccuracy would increase in tandem with the extent to which the error of poll lay outside of the margin of error.

Though conceptualising polling inaccuracy in bounded terms seems intuitive, as it is ostensibly more congruent with the practical reality of polling than its distributive counterpart, it possesses

[369] Shirani-Mehr and others, p. 607.

considerable shortcomings. The use of margins of error to caveat the understanding of polling inaccuracy presupposes that they are truly representative of the uncertainty surrounding the point estimates provided by polls. Unfortunately, they are not. Polls generate population-level vote estimations by taking a weighted average of responses from a random sample in order to correct for known differences between it and the target population. This generates two statistics: an election outcome point estimate and the estimated error surrounding this point estimate, inclusive of the effect of weighting.[370] The margin of error only represents the uncertainty introduced by random sampling error and excludes the known influence of design effects, such as weighting procedures, along with sources of non-sampling error which would increase its range.[371] As such, it not fully representative of the uncertainty surrounding vote share point estimates. Resultantly, conceptualising polling inaccuracy as a tolerance-based consideration affords too much prominence to an unrepresentative statistic and produces an unrealistic distinction between expected and unexpected inaccuracy.

In addition to the unrepresentative nature of its central focus, conceptualising polling inaccuracy as a tolerance-based issue shares many of the same issues as its continuous counterpart. Most straightforwardly, it is possible for a poll to exhibit inaccuracy within its stated margin of error and still fail to correctly predict the substantive political outcome of an election. This is most likely if the vote share point estimates are significantly erroneous in and of themselves, or when the percentage difference between the vote shares of the leading candidates is smaller than the margin of error. Moreover, while it accounts for a greater number of the characteristics possessed by vote share distributions, it still fails to account for the

[370] Ibid.

[371] Peter V. Miller, 'The Authority and Limitations of Polls', in *Navigating Public Opinion: Polls, Policy and the Future of American Democracy*, ed. by Jeff Manza, Fay Lomax Cook, and Benjamin J. Page, (New York: Oxford University Press, 2002), p. 225.

ordering of point estimates and its relevance for substantive political outcomes. As such, it too presupposes that the prediction of election results is of greater importance than their outcomes.

The issues surrounding the dismissal of outcome-orientated predictions can be resolved by conceptualising the inaccuracy of polls dichotomously as a function of their ability to predict substantive political outcomes. Though this does eliminate the restrictive focus on results, it presents its own shortcomings which undermine its ability to provide a rounded and tractable account of polling inaccuracy.

Substantive Inaccuracy

Implicit within the estimated vote share distributions provided by polls are predictions of the substantive political outcomes of elections. Most visibly, these take the form of the party or candidate predicted to receive the largest share of the vote and the ordering of the vote shares thereafter. Dependent on electoral system and margin of victory, these variously represent the anticipated winner of an election and likely government formation. As such, the inaccuracy of polls can be conceptualised as a function of their ability to correctly predict these substantive political outcomes. This conceptualisation lends itself to two understandings. The first and most intuitive is binary in nature. Pre-election polls are deemed accurate if they correctly predict the winner or finalised government formation of an election, and incorrect if they do not, irrespective of margin.[372] The second is continuous in nature and treats substantive inaccuracy as a matter of degree. As the substantive outcome of elections is determined by the number of parliamentary seats or electoral college votes accrued by parties or candidates, the

[372] Spencer Kimball, '2016 Presidential State-wide Polling – A Substandard Performance: A Proposal and Application for Evaluating Pre-election Poll Accuracy', *American Behavioral Scientist*, 63.7 (2019), 768 – 788 (p. 768); Rami Zeedan, 'The 2016 US Presidential Elections: What Went Wrong in Pre-election Polls? Demographics Help to Explain', *J – Multidisciplinary Scientific Journal*, 2.1 (2019), 84 – 101 (p. 92); Tudor and Wall, p. 1.

substantive inaccuracy of polls can also be understood as a function of the degree to which they correctly predict these totals.

While it can be argued that pre-election polls do not, in fact, render either seat or electoral college predictions, both are informed by and contingent upon the popular vote share distribution within an election. As raw vote shares are invariably subject to transformation in order to determine the substantive outcome of elections, in estimating the distribution of these vote shares, pre-election polls implicitly provide insight into the likely substantive outcomes of the elections for which they are conducted. Though this insight is clearer under differing electoral systems and in the case of elections characterised by wider vote share margins between leading parties and candidates, the distribution of either parliamentary seats or electoral college votes implicit within polling estimates can be, and routinely is, determined through the use of secondary transformations.[373] Resultantly, through both these secondary transformations and the connection between vote share distributions and substantive electoral outcomes, the degree to which the substantive outcomes projected by polls correspond to reality can be established and used to establish their (in)accuracy.

Conceptualising the inaccuracy of pre-election polls in substantive terms arguably provides the most salient measure of their success, as future-orientated predictions are widely considered to be aids to decision-making and the planning of future actions.[374] Though their number, process

---

[373] Stephen D. Fisher and others, 'From Polls to Votes to Seats: Forecasting the 2010 British Election', *Electoral Studies*, 30 (2011), 250 – 257 (p. 250).; Robert Ford and others, 'From Polls to Votes to Seats: Forecasting the 2015 British General Election', *Electoral Studies*, 41 (2016), 244 – 249 (p. 244); Souren Soumbatiants, Henry W. Chappel Jr., and Eric Johnson, 'Using State Polls to Forecast U.S. Presidential Election Outcomes', *Public Choice*, 127 (2006), 207 – 223 (p. 207); Paul F. Whiteley, 'Forecasting Seats from Votes in British General Elections', *British Journal of Politics and International Relations*, 7 (2005), 165 – 173 (p. 165); Edward R. Tufte, 'The Relationship Between Seats and Votes', *The American Political Science Review*, 67.2 (1973), 540 – 554 (pp. 540 – 541); Joseph Bafumi, Robert S. Erikson, and Christopher Wlezien, 'Forecasting House Sears from Generic Congressional Polls: The 2010 Midterm Election', *PS: Political Science and Politics*, 43.4 (2010), 633 – 636 (p. 633).

[374] Jennifer L. Castle, Michael P. Clements, and David F. Hendry, *Forecasting: An Essential Introduction*, (New Haven: Yale University Press, 2019), p. 64.

of determination, and manner of distribution differs between systems, elections fundamentally comprise winners and losers. Those who win go on to shape the legislative, economic, and diplomatic future of their states for years to come, while the losers do not. As candidates often stand on polar platforms, the effects of their respective victories differ greatly. In much the same way as individuals rely on accurate weather forecasts to determine whether they need to prepare for rain, pre-election polls provide glimpses into the likely political future and all that it entails, allowing both domestic and international actors to plan accordingly. Given its centrality to future-orientated decision-making and its far-reaching ramifications, the prediction of the substantive political outcome of an election arguably stands above all factors as the most practically important determinant of polling inaccuracy.

In addition to its importance, incorrectly predicting the substantive political outcome of an election is also the most influential form of polling inaccuracy. Not only is it the preponderant catalyst for large-scale enquiries into predictive failures, as the previous chapter made clear, but it also plays an outsized role in the determination of narratives concerning polling accuracy, irrespective of the degree to which the vote share distribution of an election was (in)correctly predicted. This was made evident by the coverage of polling accuracy during past elections in the US,[375] France,[376] and UK,[377] respectively.

Despite its importance and influential nature, conceptualising inaccuracy solely as the inability of polls to correctly call the substantive political outcome of an election presupposes that this is their only objective. While the correct prediction of electoral outcomes is undoubtedly a key determinant of their accuracy, it only constitutes one of the three pieces of information that they provide. If the identification of the victor or outcome truly were their sole objective, then

[375] Nicholas A. Valentino, John Leslie King, and Walter W. Hill, 'Polling and Prediction in the 2016 Presidential Election', *Computer*, 50.5 (2017), 110 – 115 (p. 112).
[376] Enten, *Macron Won*.
[377] Tudor and Wall, p. 1.

the provision of estimated vote shares and their uncertainty would be moot. As it fails to consider these additional facets of polling, the substantive conceptualisation only provides a partial representation of polling inaccuracy.

The issues inherent within the substantive conceptualisation of polling inaccuracy are not limited to its inability to fully reflect polling inaccuracy. Its most basic shortcoming is that, in the manner in which it is operationalised in this thesis, it precludes a nuanced understanding of inaccuracy. As inaccuracy is treated as a binary concept – polls either are or are not inaccurate – its severity comparative to other polls cannot be discerned beyond a simple black and white distinction. Accordingly, while it still permits aggregate assessment, it reduces the utility of comparison, as all instances of inaccuracy are deemed equally problematic. As conceded above, this is not a shortcoming of conceiving of polling error in substantive terms, but is rather an artefact of the manner in which I measure substantive polling error. Substantive error could be treated as a matter of degree. That is, the extent to which polls fail to predict the winner of an election could be taken into account. Under this approach, a poll that predicts 49% of the vote for a candidate that receives 51% would be treated more favourably than a poll that predicts 30% for a candidate who receives 70%. Future studies ought, therefore, attempt to broaden the manner in which substantive error is operationalised to further assess the degree to which election-level factors bear upon it.

More broadly, it also possesses issues which are contingent on how the prediction of substantive political outcomes is defined. If it is defined as the correct prediction of seats-based outcomes via votes-to-seats transformations, then it suffers difficulties of tractability. Though seat distributions are, to a greater or lesser degree, a function of national vote share distributions, their explicit provision within pre-election polls is not widespread (though it must be noted that polls conducted in certain countries, such as the Netherlands, Israel, and India do routinely provide seat share estimates as recognised later in this thesis). Rather, the provision of seat share estimates often requires a secondary level of data transformation which can be conducted in a variety of ways, each providing different outcomes.[378] As the procedures used to transform vote shares into seat distributions differ between electoral systems and therefore countries,[379] the relationship between the national vote share distributions provided by polls and substantive, seat-based outcomes varies internationally, undermining its cross case comparability and consequently its tractability.

If the prediction of substantive political outcomes is defined as the prediction of the party with the largest share of the vote, the issues of tractability remain, and it suffers from being overly permissive of overestimation. Although all elections contain a party or parties in possession of the largest share of the vote, ostensibly making it a tractable measure, it is not always directly linked to the determination of their substantive political outcomes. For example, the party or candidate with the largest share of the vote did not go on to win the US presidential elections of 2000 and 2016,[380] the Canadian federal election of 2019,[381] or the UK general election of February 1974.[382] Additionally, as it pays no mind to margin, severe over- and under-estimations are considered as accurate as predictions more reflective of reality, provided they successfully predict the largest party. In this sense, it represents the mirror image of its

[378] Liezl van Eck, Stephanus E. Visagie, and Hendrik C. de Kock, 'Fairness of Seat Allocation Methods in Proportional Representation', *ORiON*, 21.2 (2005), 93 – 110 (pp. 94 – 97).

[379] Kai-Friederike Oelbermann, Antonio Palomares, and Friedrich Pukelsheim, 'The 2009 European Parliament Elections: From Votes to Seats in 27 Ways', *European Electoral Studies*, 5.1 (2010), 148 – 182 (p. 150); Tufte, p. 541.

[380] Danny L. McDonald and others, *Federal Elections 2000: Elections Results for the U.S. President, the U.S. Senate, and the U.S. House of Representatives*, (Washington D.C., Federal Election Commission, 2001), p. 11; Steven T. Walther and others, *Federal Elections 2016: Election Results for the U.S. President, the U.S. Senate, and the U.S. House of Representatives*, (Washington D.C., Federal Election Commission, 2017), p. 5.

[381] Elections Canada, *October 21st 2019 Federal Election: Election Results*, <https://enr.elections.ca/National.aspx?lang=e> [accessed 09/08/2020].

[382] UK Political Info, *1974 February General Election Results Summary*, <http://www.ukpolitical.info/1974Feb.htm> [accessed 09/08/2020].

continuous counterpart, as it is solely concerned with prediction of outcomes at the expense of results.

Clearly then, each of the outlined conceptualisations possesses shortcomings and neither presents a complete account of the ways in which it is possible for polls to be inaccurate. As each of the individual conceptualisations of polling inaccuracy possesses issues which are ostensibly resolved by their counterparts, it stands to reason that their combination could produce a more defensible conceptualisation. Moreover, as polls present three indices of predictive inaccuracy, a conceptualisation incorporating more than one factor would necessarily offer a more complete conceptualisation than individually focused alternatives.

Could Conceptual Combination Resolve Individual Shortcomings?

While the combination of the outlined conceptualisations of polling error is possible and offers prospective benefits, it also presents significant limitations and impracticalities. For example, as their substantive outcomes are determined by the same mechanisms and vote share distributions are a universal feature of my studied pre-election polls, it would be possible to combine the substantive and distributive conceptualisations of inaccuracy for polls held under the same electoral system. This would allow for the conceptualisation of inaccuracy relative to both substantive electoral outcomes and estimated vote share distributions, partially overcoming the shortcomings presented by each conceptualisation in isolation. Despite this, their combination still fails to overcome the predominant shortcoming of the substantive conceptualisation: analytical tractability between cases. While they can be combined successfully in analyses incorporating one electoral system,[383] the nature of substantive political outcomes is such that their determination differs between systems. As such, the

[383] Tudor and Wall, p. 1.

benchmark against which their (in)accurate prediction is conceived varies, precluding tractability.

Additionally, the distributive and bounded conceptualisations of polling inaccuracy could be combined to combat the overly stringent approach to error taken by the former. However, in order for this to provide a true representation of the uncertainty surrounding point estimates and allow for a defensibly nuanced approach to polling inaccuracy, the information inherent within reported margins of error would need to radically change. While attempts are being made to fully identify the uncertainty associated with polling estimates,[384] they are in their nascent stages. Accordingly, such a hybridised approach is not practically feasible at present.

Normatively, a rounded conceptualisation of polling inaccuracy ought to comprise each of the distributive, bounded, and substantive approaches in order to account for the full gamut of ways in which pre-election polls can be inaccurate. However, as the preceding examples have shown, conceptual combination is difficult and far from a silver bullet. Any conceptual combination involving substantive inaccuracy will be faced with the insurmountable issue of analytical tractability across cases, whilst any attempt to include bounded inaccuracy will be met with the present limitations of reported margins of error. Consequently, combination cannot overcome the limitations presented by conceptualisations of polling inaccuracy in isolation.

Despite the intractability of their associated shortcomings, my distributive, bounded, and substantive conceptualisations of polling inaccuracy possess a degree of inter-connection. In the follow sub-section, I unpack the nature of this inter-connection and establish the degree to which my conceptualisations of polling inaccuracy inform one another.

[384] Shirani-Mehr and others, pp. 607 – 614.

The Inter-connection Between My Conceptualisations of Polling Inaccuracy

The nature of distributive, bounded, and substantive polling inaccuracy is such that they present varying degrees of inter-connection. Most directly, distributive and bounded polling inaccuracy are closely connected. Polls cannot present bounded inaccuracy in the absence of distributive inaccuracy. If the vote share distribution predicted by a poll does not deviate from the results seen on election day, it cannot breach the error bounds established by its margin of error, as it is not erroneous. As such, distributive inaccuracy is a necessary condition for bounded inaccuracy.

While bounded inaccuracy cannot exist without distributive inaccuracy, the presence of distributive inaccuracy is not necessarily sufficient to bring it about. That is, it is possible for a poll to present distributive inaccuracy yet be boundedly accurate. As bounded inaccuracy requires distributive inaccuracy to cross a threshold established by the margin of error surrounding a poll, exhibiting distributive inaccuracy below this threshold is insufficient to bring it about.

In light of the threshold associated with bounded inaccuracy, greater distributive inaccuracy increases the likelihood of its occurrence, as polls are more likely to present errors sufficient to exceed the bounds established by their margins of error when their vote share predictions deviate from reality to a greater degree. In a similar vein, widespread bounded inaccuracy speaks to large-scale distributive inaccuracy, with the vote share predictions provided by polls broadly exceeding the tolerances provided to them by their stated margins of error.

While the relationship between distributive and bounded polling inaccuracy is often reciprocal, the same cannot be said of the relationship between distributive and substantive inaccuracy. Distributive inaccuracy remains a necessary condition of substantive inaccuracy, as polls cannot present substantive mispredictions unless their vote share predictions are erroneous.

Equally, increased distributive error lends itself to an increased probability of polls presenting substantively incorrect predictions, as polls are more likely to present substantively incorrect predictions if their predicted vote share distributions miss the mark by a greater margin. However, substantive misprediction does not necessarily entail large-scale distributive error. One poll may correctly predict the winner of an election but overestimate their share of the vote by ten points – a phenomenon seen in the 2017 French presidential election[385] – while another may incorrectly predict the winner of an election, but only underestimate the share of the vote received by the eventual winner by two points. The first is substantively accurate but presents high distributive error, while the second is substantively inaccurate yet presents low distributive error.

The above example also speaks to the relationship between substantive and bounded polling inaccuracy. As bounded inaccuracy occurs after the distributive error exhibited by a poll crosses the margin of error determined by its sample size (±3% for a typical poll with a sample size of 1,000 respondents[386]), its presence lends itself to substantive misprediction in elections characterised by margins of victory smaller than the margins of error associated with polls. However, the presence of bounded inaccuracy does not, ipso facto, lead to the presence of substantive inaccuracy. Given that a poll may correctly call the substantive outcome of an election yet overestimate the winning vote share by 10 points – a margin far greater than the margins of error typically associated with polls – highlights this disconnect. The poll in question presents clear bounded inaccuracy, but remains substantively correct.

385 Harry Enten, *Macron Won, But the French Polls Were Way Off* (2017), <https://fivethirtyeight.com/features/macron-won-but-the-french-polls-were-way-off/> [accessed 17/07/2020].

386 Anthony Wells, *Understanding Margin of Error* (2011), <https://yougov.co.uk/topics/politics/articles-reports/2011/11/21/understanding-margin-error> [accessed 01/09/2022].

In addition to informing one another to varying degrees, each of my conceptualisations of polling inaccuracy possess common causes. That is, they are each informed by the random or systematic errors present in the pre-election polling process, either in isolation or combination. In the following section, I address the dominant poll-level understanding of polling error, inclusive of random and systematic errors born of polling practices, and its connection with my three conceptualisations of polling inaccuracy.

3.3: The Dominant Poll-level Understanding of Polling Inaccuracy

Under the present paradigm, polling inaccuracy is conceptualised as a function of two factors: random and systematic errors at the poll level.[387] In this section, I explore this conceptualisation by unpacking the forms of error at its core, addressing their causes, and identifying their limitations. I begin by identifying that both random and systematic polling error are conceived of as products of the processes underpinning pre-election polling, represented by the total survey error framework. Thereafter, I unpack the characteristics of random and systematic error in pre-election polls, attribute them to the earlier conceptualisations of polling inaccuracy, and address systematically the specific issues within the total survey error framework that give rise to them. Moreover, I address the ways in which they can be, and are, mitigated and corrected for. Through this, I show that the present conceptualisation of polling inaccuracy is exclusively poll-level and comprises fundamentally surmountable elements.

With the nature and sources random and systematic polling error established, I illustrate that the dominant poll-level conceptualisation of polling inaccuracy possesses a central shortcoming. This shortcoming concerns the fact that, under its use, the accuracy of polling has not noticeably improved. I identify that the inaccuracy of global pre-election polls has remained

[387] Gary King, Robert O. Keohane, Sidney Verba, *Designing Social Inquiry*, (Princeton: Princeton University Press, 1994), pp. 155 – 158.

largely unchanged over the past 75 years. Through doing so, I note that the effort to reduce polling inaccuracy through iterative methodological revision over time has proven ineffective and resulted in a lack of progress.

Once this lack of progress has been identified, not only do I contend that it is problematic in and of itself, but I hold that it is incongruent with the nature of the present conceptualisation of polling inaccuracy. This assertion rests on the contention that, as they can be either mitigated or corrected for after the fact, if polling inaccuracy truly were solely a function of random and systematic errors at the poll level, then the extensive past efforts to improve it through the iterative modification of methods would have yielded discernible improvement over time. As this has not occurred, I present a series of possible explanations for why this is the case. Through their assessment, I conclude that the current conceptualisation provides an incomplete account of the factors affecting polling inaccuracy, as it cannot reasonably be reconciled with the performance of polls.

Understanding Poll-level Sources of Error: The Total Survey Error Framework

Under the current paradigm, polling inaccuracy is largely conceived of as a function of random and systemic errors at the poll level. The sources of these random and systematic errors are identified and understood through the total survey error framework.[388] Within this framework, random and systematic errors are products of issues arising from the processes undergirding pre-election polling.[389] Figure 3 displays the total survey error framework and the sources of random and systematic error that it houses to allow them to be unpacked.

[388] Robert M. Groves and Lars Lyberg, 'Total Survey Error: Past, Present, and Future', *Public Opinion Quarterly*, 74.5 (2010), 849 – 879 (p. 849).

[389] James Raymer and Philip Howell Rees, 'Framework for Guiding the Development and Improvement of Population Statistics in the United Kingdom', *Journal of Official Statistics*, 31.4 (2015), 699 – 722 (p. 710).

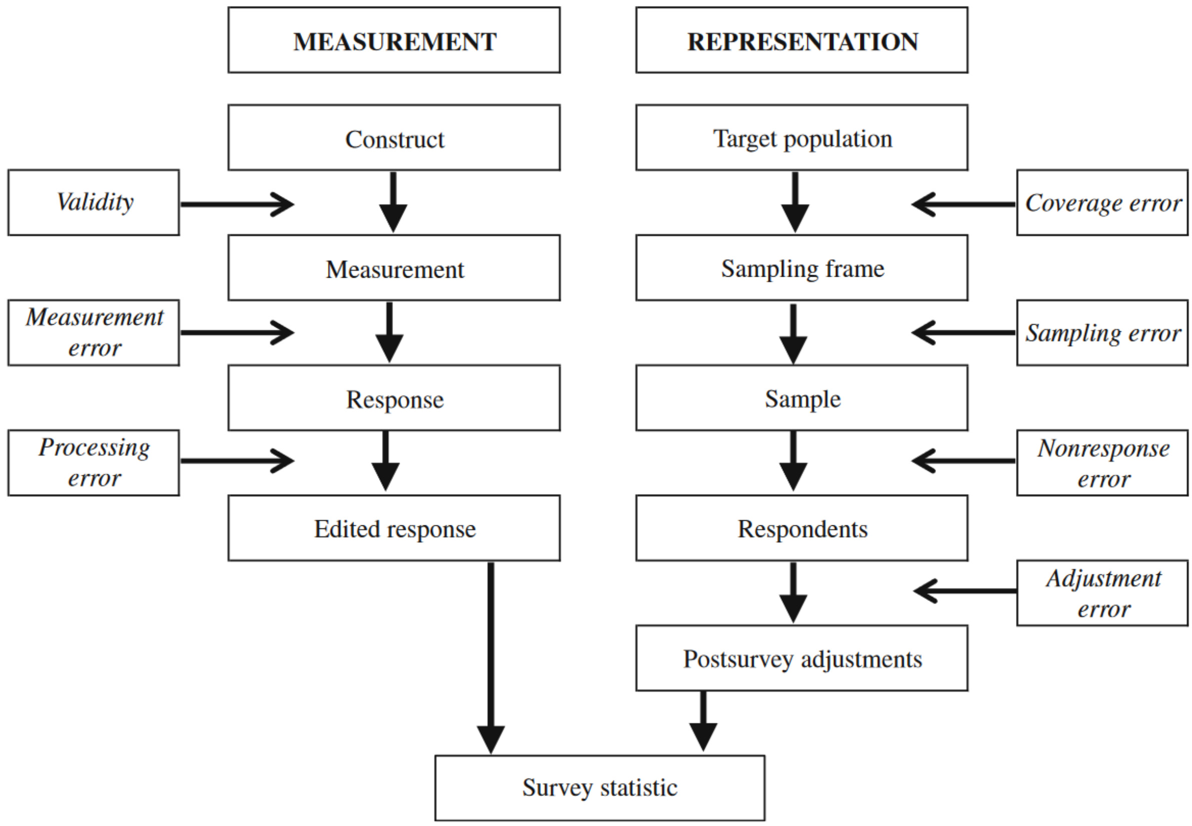

Figure 3: The total survey error framework, encompassing survey processes and associated sources of both random and systematic error. Adapted from Raymer and Rees.[390]

While the sources of error housed within the total survey error framework are often grouped into instances of sampling and non-sampling error,[391] it is more intuitive to unpack them further into issues of measurement and representation (as displayed in Figure 3). Issues of measurement surround the process of gathering, measuring and processing survey responses, while those of representation concern the degree to which the sample taken reflects the target population.[392] Relating errors directly to these processes removes sampling as the central grouping factor and allows for the exploration of seven discrete sources of error.

[390] Raymer and Rees, p. 710.
[391] Paul P. Biemer, 'Total Survey Error: Design, Implementation, and Evaluation', *Public Opinion Quarterly*, 74.5 (2010), 817 – 848 (p. 822).
[392] Marek Fuchs, 'Total Survey Error (TSE)', in *Encyclopedia of Survey Research Methods*, ed. by Paul J. Lavrakas, (Thousand Oaks, Sage Publications, 2008), p. 897.

The total survey framework encapsulates the processes of measurement and representation that undergird pre-election polling. Each process moves in tandem, eventually culminating in the production of a survey statistic in the form of the estimated vote share distribution provided by a poll. Each step of these parallel processes variously presents the possibility for random and systematic errors to occur at the poll level and, subsequently, for polling inaccuracy to arise. In order to understand how this process occurs, I move to define both random and systematic error, and link them both to the discrete steps of the total survey error framework and the previously addressed conceptualisations of polling inaccuracy.

Random Errors at the Poll Level

Random error refers to the imprecision inherent in repeated measurements of the same value or set of values, which is caused by the marginal variation, without pattern or direction, of each measurement from the one preceding it.[393] In survey research, random error arises as an artefact of the sampling process.[394] Within the total survey error framework, it is identified as a result of sampling variance.[395] As samples are inherently imperfect and incomplete representations of a target population, each sample drawn provides slightly different population estimates due to its differing composition.[396] Each poll therefore presents differing vote share estimates simply as a result of the random error inherent within samples, affecting the precision with which they can render predictions.

[393] Paul J. Lavrakas, 'Random Error', in *Encyclopedia of Survey Research Methods*, ed. by Paul J, Lavrakas, (Thousand Oaks: Sage Publications, 2008), pp. 678 – 679.
[394] Hillygus, p. 965.
[395] Donald P. Green, Alan S. Gerber, and Suzanna L. De Boef, 'Tracking Opinion Over Time: A Method for Reducing Sampling Error', *Public Opinion Quarterly*, 63.2 (1999), 178 – 192 (p. 178); Gerald C. Wright, 'Errors in Measuring Vote Choice in the National Election Studies', *American Journal of Political Science*, 37.1 (1993), 291 – 316 (p. 296); Karol Krotki, 'Sampling Error', in *Encyclopedia of Survey Research Methods*, ed. by Paul J. Lavrakas, (Thousand Oaks, Sage Publications, 2008), pp. 785 – 789.
[396] Tom Louwerse, 'Improving Opinion Poll Reporting: The Irish Polling Indicator', *Irish Political Studies*, 31.4 (2016), 541 – 566 (p. 542).

As random error born of the sampling process leads to random variations in the vote share estimates provided by polls, it necessary affects the degree to which these estimates accurately reflect finalised vote share distributions. In this way, it bears upon my distributive conceptualisation of polling error. The effect of random error on the vote share estimates provided by pre-election polls is represented visually in Figure 4:

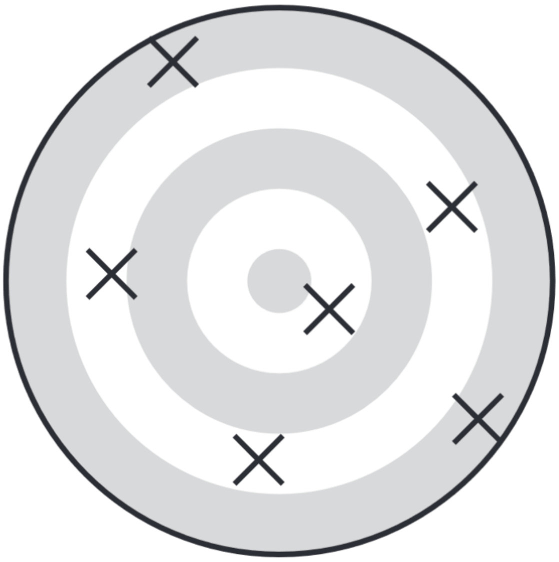

Figure 4: A visual representation of the effect of random error on the accuracy and precision of polling estimates, with estimates represented by crosses and the true population value by the bullseye. Adapted from the work of Lohr.[397]

The expected extent of random error born of the sampling process is represented by the margin of error associated with a poll.[398] The severity of this error is inversely dependent on sample size,[399] and can be mitigated using aggregation, along with the implementation of clustering and stratification procedures.[400] The sampling variance associated with polls can also be reduced through their aggregation. The law of large numbers holds that the average of a sufficiently large number of samples – represented in this case by individual polls – will

[397] Sharon L. Lohr, *Sampling: Design and Analysis*, (Boca Raton: CRC Press, 2019), p. 32.
[398] James W. Stoutenborough, 'Margin of Error (MOE)', in *Encyclopedia of Survey Research Methods*, ed. by Paul J. Lavrakas, (Thousand Oaks, Sage Publications. 2008), pp. 450 – 451.
[399] Shirani-Mer and others, p. 607.
[400] Krotki, p. 786.

converge on the population mean. Accordingly, the greater the number of polls, the closer the average of their estimated vote share distributions will be to the true population value, reducing random error. Additionally, the central limit theorem posits that the means of these sample values will tend towards a normal distribution about the population mean when $n \geq 30$. This allows for the simple calculation of the standard deviation which, due to convergence on the population mean, is an unbiased estimator of its population equivalent. In turn, this can be used to calculate the population-level uncertainty generated by random error, represented by the margin of error.

Given its representation within the margins of error surrounding polling estimates, random error bears closely upon my bounded conceptualisation of polling error. Its relationship with sample size is such that polls conducted using a smaller sample will present a larger margin of error but will be more susceptible to random error.[401] In this way, while it is necessary for them to present a greater amount of error to exceed their stated margins, they are more likely to present this error as a result of a diminished sample that is less likely to be representative of the voting population.

Though the random error associated with polls can be readily identified and mitigated, it can never be eliminated entirely. As they are not censuses comprising all individuals of interest, any poll, or series of polls, regardless of sample size or aggregation, still represents a sample of the target population. Therefore, they still possess random sampling error. Despite its constant presence, random error is theoretically unlikely to account for large-scale instances of polling inaccuracy in and of itself. By its nature, each instance of random error possesses the same probability of being either positive or negative and of taking on any real number as its value. If polls were only subject to random error and the accuracy of the predictions rendered

[401] Hillygus, p. 665.

by a sufficiently large number of them was measured as function of the percentage point disparity between predicted and actual vote share distributions, then the results would cancel out to perfect predictive accuracy. That is, the number of over-estimations would exactly equal the number of under-estimations, rendering the average prediction perfectly accurate.

In practice, it is neither economically nor temporally feasible to conduct the number of polls required for random error to cancel out exactly, as it would require a near-infinite number of values. However, it is possible to conduct a sufficiently large number of polls such that random error values prove insignificant through large-scale cancellation. This assumption governs the treatment of random error and is used widely in political science.[402]

While it is the most significant, the random sampling process is not the only source of random error in polling. It may also come about due to random response error,[403] variously born of issues surrounding the interpretation of survey questions on the part of respondents and random instances of refusal brought upon by personality and mood[404] or random interviewer error, which is brought about by mistakes in data entry or the idiosyncrasies of questioning.[405] Moreover, isolated instances of processing error, such as mistakes in transcription, instances of errant data entry, may also result in random error.[406] However, by virtue of being random,

[402] Richard R. Lau and David P. Redlawsk, 'Advantages and Disadvantages of Cognitive Heuristics in Political Decision Making', *American Journal of Political Science*, 45.4 (2001), 951 – 971 (p. 952); Shaun Ratcliff, 'Voter Behaviour', in *Australian Politics and Policy: Senior Edition*, ed. by Peter J. Chen and others, (Sydney: Sydney University Press, 2019), p. 475; John Bartle, Agusti Bosch and Lluís Orriols, 'The Policy Mood in Spain: The Thermostat in a Warm Climate, 1978 – 2017', *European Political Science Review*, 12.2 (2020), 133 – 153 (p. 137).
[403] Lloyd B. Brown and Henry W. Chappell Jr., 'Forecasting Presidential Elections Using History and Polls', *International Journal of* Forecasting, 15.2 (1999), 127 – 135 (p. 129).
[404] Harper W. Boyd Jr., and Ralph Westfall, 'Interviewers as a Source of Error in Surveys', *Journal of Marketing*, 19.4 (1955), 311 – 324 (p. 313).
[405] Woody Carter, 'Interviewer-related Error', in *Encyclopedia of Survey Research Methods*, ed. by Paul J. Lavrakas, (Thousand Oaks, Sage Publications, 2008), p. 378.
[406] Fuchs, p. 900.

these errors also cancel out to insignificance over a large enough number of instances and are often mitigated by a combination of training and weighting procedures.[407]

Accordingly, not only is random error identifiable in single polls and easily mitigated in the aggregate, but it should also be insignificant when considered in the aggregate. Given this information, random error is highly unlikely to be responsible for the large-scale, collective polling errors warranting post-election enquiries addressed in the previous chapter. The fact that such large, surprising errors occur speaks to the existence of additional forms of error which are both more significant and not the object of mitigation by mathematical laws.

Systematic Errors at the Poll Level

The additional source of polling inaccuracy under the current paradigm is systematic error at the poll level. Systematic error is the opposite of its random counterpart, as it concerns the consistent over- or under-estimation of population values.[408] As such, it is non-random and possesses directionality. Due to its non-random nature, it is compounded over the course of repeated measurement.[409] Moreover, its inherent directionality biases polling estimates in a certain direction, often causing them to cluster away from the true population value. Resultantly, while the estimates possess precision, they are universally and similarly inaccurate, as demonstrated below in Figure 5.

[407] Carter, p. 378.
[408] Karen Long Jusko, 'Systematic Error', in *Encyclopedia of Survey Research Methods*, ed. by Paul J. Lavrakas, (Thousand Oaks, Sage Publications, 2008), pp. 869 – 870.
[409] Louwerse, p. 543.

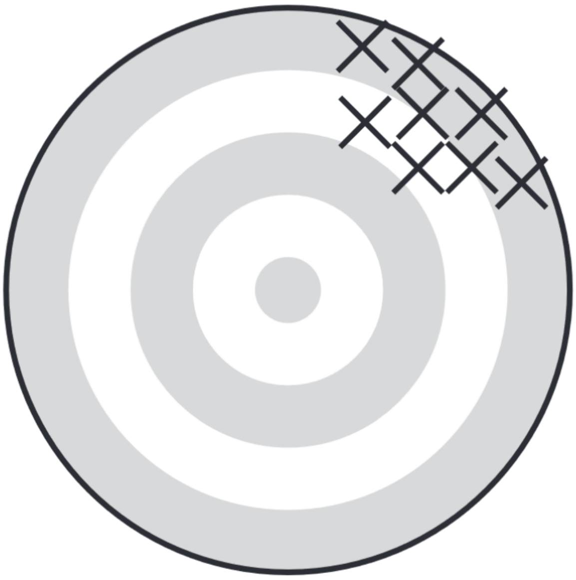

Figure 5: A visual representation of the effect of systematic error on the accuracy and precision of polls, with estimates represented by crosses and the true population value by the bullseye. Adapted from the work of Lohr.[410]

Under the present understanding of polling inaccuracy, it is systematic error at the poll level that brings about misprediction beyond the margin of error and, when replicated between polls, causes widespread inaccuracy. These systematic errors at the poll level can be understood through the lens of bias. Bias can be introduced to pre-election polling by the processes undertaken by polling organisations. Actions such as herding – the process of polling organisations altering their methodologies to produce results that better reflect prevailing wisdom[411] – can bias results, leading to a clustering of estimates that present low variance,[412] and result in a greater likelihood of systematic error. Similarly, the partisan alignment of polling organisations and the nature of the sponsors commissioning polls from these organisations can both systematically affect the nature of their predictions and increase the likelihood that they present systematic error.[413]

[410] Lohr, p. 32.
[411] Prosser and Mellon, p. 776.
[412] Sturgis et al., p. 5.
[413] Jacob Shamir, 'Pre-election Polls in Israel: Structural Constraints on Accuracy', *Public Opinion Quarterly*, 50.1 (1986), 62 – 75 (p. 62); Ivor Crewe, 'The Opinion Polls: Still Biased to Labour', *Parliamentary Affairs*, 54.4 (2001), 650 – 665 (p. 650).

Though it is strongly associated with bias due to its inherent directionality,[414] it is possible for systematically erroneous estimations to be equidistantly distributed around the true population value and to therefore be unbiased in the aggregate. Despite this, unbiased systematic errors are unlikely in polling due to the nature of herding, where pollsters cluster together by matching the prevailing wisdom of other firms, especially towards the end of electoral cycles.[415] Through this, systematic error is not only more likely to be biased, as estimates will be clustered around one another, but also has the potential to affect polls even in the absence of endogenous causes. It is also possible for biased systematic errors to come about in the absence of herding, as pollsters may use common methodologies which, when they contain an inherent directional bias that leads to problems in estimation, may result in bias that drives systematic error.

While random error cancels out in the aggregate, predictive inaccuracy caused by systemic error can only be resolved by identifying and correcting its cause. Moreover, while the presence of random error in polling is largely the result of a singular factor, to the extent that they possess the ability to consistently over- or under-estimate public opinion relative to its population level, all methods and processes undergirding polls have the potential to introduce systematic error.[416] Their likelihood of doing so, however, is a function of a wide range of factors. As such, while systematic error can be addressed in general terms, such as the broad identification of national organisational biases,[417] or the tendency of certain pollsters to over-estimate vote shares either due to partisan biases and house effects or over-weighting on the basis of past mistakes,[418] it is

[414] Simon Jackman, 'Pooling the Polls Over an Election Campaign', *Australian Journal of Political Science*, 40.4 (2005), 499 – 517 (p. 501); Jeffrey A. Stec, 'Bias', in *Encyclopedia of Survey Research Methods*, ed. by Paul J. Lavrakas, (Thousand Oaks, Sage Publications, 2008), pp. 56 – 60.
[415] Michael W. Traugott and Paul J. Lavrakas, *The Voter's Guide to Election Polls*, (Morrisville: Lulu Press, 2016), p. x; Whiteley, p. 439; Sturgis et al., p. 10; Panagopoulos, Enders, and Weinschenk, p. 170.
[416] Fisher and others, p. 250.
[417] Crewe, 'The Opinion Polls: Still Biased to Labour', p. 650; Lynn and Jowell, p. 23; Butler and Kavanagh, p. 143.
[418] Mark Pickup and Richard Johnston, 'Campaign Trial Heats as Election Forecasts: Measurement Error and Bias in 2004 Presidential Campaign Polls', *International Journal of Forecasting*, 24.2 (2008), 272 – 284 (pp. 272 – 284); Jakob Bergman and Björn Holmquist, 'Poll of Polls: A Compositional Loss Model', *Scandinavian Journal of Statistics*, 41.2 (2014), 301 – 310 (p. 307); Jackman, p. 500.

best understood through its direct attribution to the process underpinning polling outlined through the total survey error framework.

Each element of the measurement process underpinning pre-election polling possesses the potential to introduce bias and, therefore, systematic error. In the first instance, systematic errors may arise as a product of issues of validity, better understood as specification error. Specification error occurs in polling when what is being measured differs from what should have been measured.[419] For example, specification error may arise in polling when a respondent's understanding of a question differs from its intended purpose.[420] In this case, the response provided by the respondent is not capturing what the question is designed to elicit due to misinterpretation. If specification error such as this occurs, the estimates provided by a poll are based on incorrect parameters, undermining the validity of any inferences made after their collection. As it is often the result of incorrect questionnaire design,[421] specification error pervades the polls which it affects, leading to consistent and therefore systematic error in their estimates.

Systematic errors may also be born of instances of measurement error. Measurement error is the product of survey instruments, broadly conceived. These instruments include the interviewers who interact with and question respondents, the respondents themselves, the questionnaire, and the mode of data collection.[422] The interviewers who collect the raw data upon which polls are based can themselves inject bias into the process through the idiosyncrasies inherent within their approach questioning or the imposition of their own political leanings.[423] The wording of questions may prime individuals such that they tend

[419] Biemer, 'Total Survey Error', p. 821.
[420] Shirani-Mehr and others, p. 608.
[421] Ibid.
[422] Robert M. Groves, *Survey Errors and Survey Costs*, (Hoboken: John Wiley and Sons, 2004), p. vi.
[423] Carter, p. 378.

towards a certain response or may be worded in such a way as to confuse or suppress the true sentiments of the respondent. Respondents can and often will lie in response to questions, especially those which may elicit a response deemed undesirable in the presence of an atmosphere of social desirability bias.[424] In the case of election polling, individuals may be reluctant to admit their intention to vote for parties or candidates that are deemed socially undesirable, instead electing to provide untruthful – and, often, more socially desirable – responses when polled.[425] If these untruthful response go undetected and those providing them turn out in favour of their true, unrevealed preference on election day, then this phenomenon has the potential to lead to systematic polling error.

Likewise, differing modes of data collection media variously lend themselves to systematic error. Telephone polling often gathers consistently less accurate data than its face-to-face alternative,[426] opt-in online polls present more unrepresentative participant pools than alternative modes due to issues of self-selection bias in samples,[427] while mail-in polling suffers from extremely poor response rates,[428] reducing its representativeness. Such characteristics necessarily bias the estimates provided by pre-election polls, causing them to be systematically erroneous.

Instances of processing error may also bring about systematic error. Processing error has the potential to come about after raw polling data has been collected, but before finalised polling figures have been published.[429] Errors may be introduced when the answers provided by

[424] Noelle-Neumann, p. 307.
[425] Ibid.
[426] Paul Biemer, 'Measuring Data Quality', in *Telephone Survey Methodology*, ed. by Robert Groves and others, (New York: John Wiley and Sons, 1988), pp. 273 – 82.
[427] Eszter Hargittai and Gokce Karaoglu, 'Biases of Online Political Polls: Who Participates?', *Socius: Sociological Research for a Dynamic World*, 4 (2018), 1 – 7 (p. 5).
[428] Lara M. Greaves and others, 'How low can we go? Declining survey response rates to New Zealand electoral roll mail surveys over three decades', *Political Science*, 72.3 (2020), 228-244 (p. 228).
[429] Daniel Kasprzyk and Lee Giesbrecht, 'Reporting Sources of Error in U.S. Federal Government Surveys', *Journal of Official Statistics*, 19.4 (2003), 343 – 363 (p. 356).

respondents are compiled and coded within statistical software if mistakes are made by those inputting them. If these errors occur in a consistent direction, then they have the potential to bias results, bringing about systematic polling error.[430]

Systematic errors may also arise through issues inherent within the process of representation. The most prosaic of these is coverage error, also known as frame error.[431] As surveying every individual within the target population of a poll is neither economically nor temporally viable, a sample must be drawn to represent it. Within this process, coverage error stems from three issues: under-coverage, over-coverage, and duplication. Under-coverage concerns the failure to give certain individuals within a population of interest a chance of selection within a sample due to issues with the sampling frame.[432] For example, a sampling frame comprising all households with landline phones for random digit dialling will exclude those individuals who only own mobile phones. By contrast, over-coverage concerns the inclusion of individuals who do not belong in the target population,[433] such as individuals who are not registered to vote.[434] While duplication exists as a further issue concerning the repeated inclusion of individuals already captured within a sampling frame,[435] it is unlikely to serve as a meaningful driver of error in national polling. Ultimately, issues of under- and over-coverage have the potential to result in sampling frames that are unrepresentative of the target population and, when present to significant degrees, bring about consistently biased and therefore systematically erroneous estimates.

430 Ibid.
431 Biemer, p. 824.
432 Groves, p. 95.
433 Krotki, p. 788.
434 Lynn and Jowell, p. 23.
435 Samuel J. Best and Benjamin Radcliff, *Polling America: An Encyclopedia of Public Opinion*, (Westport: Greenwood Publishing Group, 2005), p. 134.

Once the sampling frame has been created, a sample is drawn from the target population. This gives rise to sampling error which may also bring about systematic error through sampling bias. Sampling bias concerns instances in which individuals are systematically more likely to be selected in a sample than others. Though it is linked to coverage issues and problems with the sampling frame,[436] the manner in which individuals are sampled may bias their chance of inclusion. To provide a real-world example, during the 1948 US presidential election, nationwide polling conducted via telephone implied that Dewey would defeat Truman.[437] By reaching out to respondents via telephone, these polls necessarily relied on samples of telephone owners. In 1948, household telephones were a relatively new and expensive technology, meaning that they were largely the preserve of the wealthy.[438] To this end, relying on a sample of telephone owners was such that wealthy Americans were systematically more likely to be included in the sample than less affluent individuals.

Instances of sampling bias ultimately result in an unrepresentative sample being drawn from the target population.[439] An unrepresentative sample is one that does not compositionally represent the target population. Accordingly, unrepresentativeness is the result of either the under- or over-representation of certain socio-demographic factors amongst the individuals comprising a sample. As it is no longer reflective of the target population, unrepresentative sampling introduces systematic error into a poll. In case of the 1948 US presidential election, wealthy Americans were more likely to support Dewey than less affluent individuals,[440] biasing

[436] Allan L. McCutcheon, 'Sampling Bias', in *Encyclopedia of Survey Research Methods*, ed. by Paul J. Lavrakas, (Thousand Oaks, Sage Publications, 2008), p. 784.
[437] Duminda N. Wijeysundera and Sindhu R. Johnson, 'Surveys, Samples, and Botched US Presidential Predictions', *European Journal of Anaesthesiology*, 29.10 (2012), 462-464 (p. 463).
[438] Ibid.
[439] McCutcheon, p. 240.
[440] Bernard Friedenson, 'Dewey Defeats Truman and Cancer Statistics', *Journal of the National Cancer Institute*, 101.16 (2009), 1157 – 1157 (p. 1157).

the findings of telephone polls in which they were over-represented, leading to systematic polling error.

Systematic error is also linked to issues of non-response which have the potential to bring about systematic errors. Non-response error exists in two forms: unit and item. Unit non-response is born of the failure to garner useful responses from all individuals within a sample, as some either do not respond or provide responses that are unusable.[441] This can be due to refusal on the part of the respondent, the second is non-contact in which polls do not manage to reach individuals, and the third concerns the physical inability of interviewers to communicate with respondents due to issues of language or comprehension.[442] The absence of these responses can lead to non-response bias if those who do not respond to polls hold voting intentions that differ systematically from those who do. The presence of this bias results in the misestimation of population characteristics on the basis of the sample of responses gathered, leading once again to sampling bias and systematically erroneous estimations. The ability for non-response to bias the findings of polls, leading to systematic error was made clear in the 2016 US presidential election. Non-university educated voters were less likely to respond to polls than those with university degrees, and were more likely to support Donald Trump.[443] As such, in cases where polls failed to recognise this and weight responses on the basis of education, non-college educated voters were underrepresented, leading to non-response bias, as the voting intention of those who did not respond to polls differed systematically from those who did. This non-response bias, compounded by issues of weighting, was identified as one of the principal drivers of polling error in the 2016 US presidential election.[444]

[441] Groves, p. 133.
[442] Daniel M. Merkle, 'Nonresponse Bias', in in *Encyclopedia of Survey Research Methods*, ed. by Paul J. Lavrakas, (Thousand Oaks, Sage Publications, 2008), p. 531.
[443] Robert Y. Shapiro, 'What We Relearned and Learned from the 2016 Elections: Comment on Gelman and Azari', *Statistics and Public Policy*, 4.1 (2017), 1-3 (p. 2).
[444] Ibid.

Non-response can also bring about random error,[445] often born of random instances of refusal brought upon by the personality or mood of respondents.[446] The clearest example of non-response bringing about random error in polling is when the missing responses that result from non-response are distributed randomly.[447] If systematic differences in voting intention are not present between those who respond to polls and those who do not, then the polling error that may result from non-response is more likely to be random in nature than systematic, even in cases of low response rates, rendering it less problematic.[448]

The final stage at which systematic error may arise is that of post-survey adjustment. While post-survey adjustments are intended to correct for biases introduced by previous forms of errors, such as non-response, they lend themselves to adjustment error. The weighting approaches that they employ have the potential to bias polling estimates if they are incorrect. Moreover, other forms of adjustment, such as the process of altering estimates to resemble rivals known as herding, also possess the potential to exacerbate systematic errors.

Systematic error bears upon each of my three conceptualisations of polling inaccuracy. In isolation, it bears most closely on substantive polling error. While it is possible for random error to result in substantive misprediction in the case of singular polls in particularly close elections, it is unlikely to bring about widespread substantive misprediction due to its cancellation in the aggregate. As systematic error does not cancel out in the aggregate and is, instead, compounded, it is more likely to drive substantive polling inaccuracy. Despite this, factors beyond systematic bias may bring about substantive polling error that are important to recognise. Significant late swings in voting intention or substantial amounts of late decision-

[445] Brown and Chappell Jr., p. 129.
[446] Boyd Jr. and Westfall, p. 313.
[447] René Bautista and others, ‘Studying Nonresponse in Mexican Exit Polls’, *International Journal of Public Opinion Research*, 19.4 (2007), 492-503 (p. 492).
[448] Ibid.

making amongst the electorate may occur on the eve of an election, beyond the ability of polls to capture, and result in widespread substantive error.

Both systematic and random error combine to bring about distributive inaccuracy. That is, as both systematic and random error increase, so too does the disparity between predicted and actual vote share distributions, leading to greater distributive inaccuracy. Systematic error also increases the likelihood of polls exhibiting bounded inaccuracy. As the margin of error surrounding a poll captures the expected effect of random sampling error, a poll is unlikely to present bounded inaccuracy as a result of random error. However, if systematic error is present, resulting in consistent over- or under-estimation, a poll is more likely to exceed its stated margin of error and present bounded inaccuracy.

While systematic error can be expected to bear upon each of my conceptualisations of polling error, its sources can be identified and corrected after the fact. Indeed, as unpacked in the literature review, a process of iterative methodological revision aimed at reducing systematic polling error has been ongoing in the wake of predictive failures for over eighty years. While this process has allowed polling accuracy to remain stable in the face of emerging issues, it has not proven sufficient to bring about a meaningful increase in polling accuracy over time. This problematises the current poll-level understanding of polling error and brings into question whether it presents a complete picture of the factors that bear upon it. In the following sub-section, I unpack the limitations of the current poll-level understanding of polling error, recognise the existence of alternative sources of error, and propose an alternative approach to understanding the determinants of polling inaccuracy.

An Incomplete Picture: Issues with the Present Understanding of Sources of Polling Error

From the total survey error framework, it is clear that polling error is principally understood as function of random and systematic errors born of issues inherent within the processes underpinning pre-election polls. While this conceptualisation is ostensibly supported by the routine identification of systematic errors as the determinants of large-scale instances of inaccuracy, as explored in the previous chapter, it suffers from one key shortcoming: while this approach has proven able to maintain stable levels of polling accuracy in the face of emerging difficulties over time, it has not proven sufficient to bring about a meaningful improvement to polling accuracy.

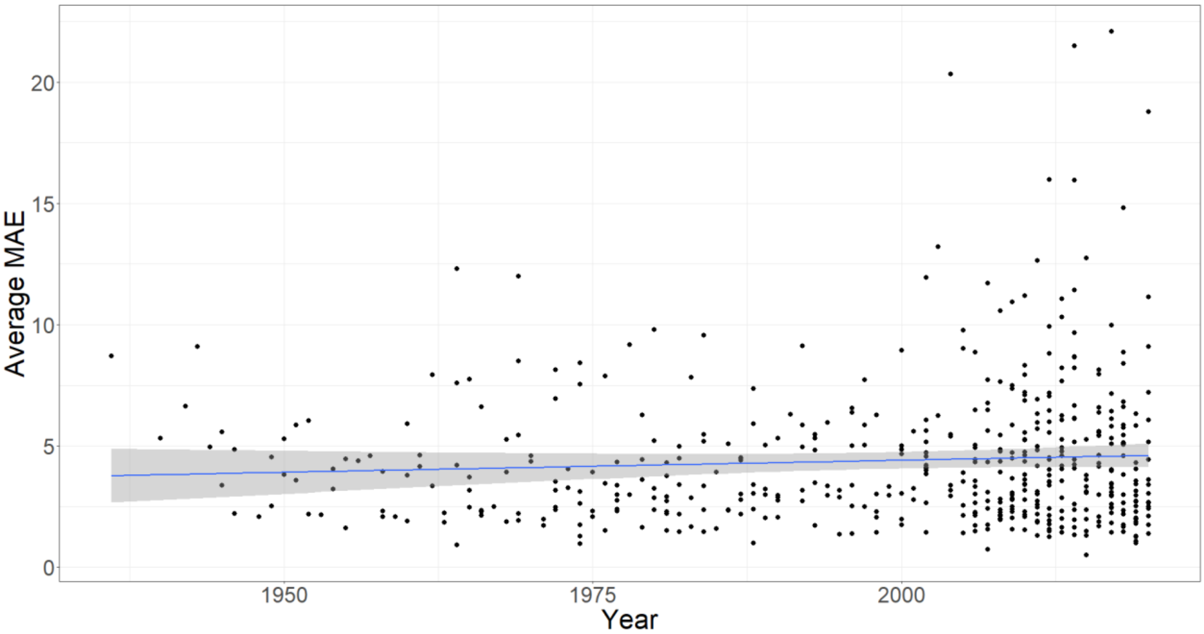

Figure 6: The average MAE exhibited by polls across my 497 elections from 1936 to 2020. The trend line represents the evolution of average MAE over time and is accompanied by a 99% confidence interval.

Figure 6 displays the average mean absolute error (MAE) exhibited by polls across my 497 studied elections from 1936 to 2020. From the trend line in the figure, is clear that the average error of polls has not meaningfully increased over this 84-year period, remaining largely stable. Though the confidence interval associated with the trend line narrows over time as more data points become available, its uppermost extent remains mostly static over time. This finding of broadly stable error over time is in keeping with previous research.[449]

[449] Jennings and Wlezien, 'Election Polling Errors Across Time and Space', p. 280.

The relative stability of average polling error over time also betrays the fact that, when considered globally, the performance of polls has not meaningfully improved over time. As made clear earlier in the thesis, the static nature of polling error over time does not necessarily indicate that the poll-level approach to improving polling accuracy has been ineffective. Though Figure 6 displays relatively static levels of polling error over time, it also charts the emergence of an increasingly challenging global polling environment. It is clear that the number of elections for which polls are conducted has risen over time. This is largely the result of the spread of pre-election polling to new countries over time which are often newly democratised.[450] This rise in cases is accompanied by an increase in the variance exhibited by polling error. Increased variance such as this suggests that the new cases to which polling organisations have turned their attention often present challenging and more error-prone polling environments.

The indication of the increased difficulty posed by the application of polling to new cases over time provided by Figure 6 is afforded credence by issues commonly faced by polling organisations when expending their operations to newly democratised states, including issues of local cooperation in survey processes and fieldwork,[451] barriers to contacting representative samples of individuals that arise through developmental and technological disparities,[452] and issues of response born of the difficulty in eliciting political preferences from individuals in societies where the expression of political opinions was previously the focus of state repression.[453] As such, it may be that polling error has actually decreased in certain countries or regions over time, but that this change is not visible in the aggregate due to the challenges

[450] Mattes, p. 114.
[451] Seligson, p. 51.
[452] Mattes, pp. 116-117.
[453] Ansu-Kyeremeh, p. 61.

posed by the application of polling to new cases. This contention is explored further later in the thesis.

The application of polling to an increasing number of difficult cases over time lends itself to understanding the largely static nature of global polling accuracy. That the accuracy of pre-election polls has remained stable over time despite the increasing range and difficulty of the cases to which they are applied suggests that the process of poll-level improvements identified earlier in this thesis has proven effective in mitigating these difficulties in the aggregate, allowing polling error to remain stable. When this is coupled with recognised issues that have risen over time within established democracies, such as the decline in response rates,[454] it is likely that with the rise of challenges over time, polling at large has faced a Red Queen problem, insofar as it is forced to work ever harder to maintain its performance. As such, it is impressive that the polling industry has managed to maintain a stable level of performance over time.

While the efforts of the polling industry in the face of a likely Red Queen problem are impressive, the largely static nature of polling accuracy over time despite concerted efforts to correct sources of error at the poll level nevertheless presents a puzzle. Ultimately, the largely static level of average polling accuracy over time lends itself to a range of variously likely explanations. I list these below and then assess each in turn.

1) Random and systematic errors at the poll level are the sole determinants of inaccuracy, but polling organisations have ignored the recommendations of post-mortems, leading to the sources of systematic error responsible for inaccuracy remaining unresolved.

[454] Zhenkun Zhou and others, ‘Why Polls Fail to Predict Elections’, *Journal of Big Data*, 8 (2021), 1-28 (p. 1).

2) Random and systematic errors at the poll level are the sole determinants of inaccuracy but, while changes have been made in light of failures, the polling for each election displays different forms of systematic error, leading to continued inaccuracy.

3) Random and systematic errors at the poll level are the sole determinants of inaccuracy, but sources of error change over time, limiting the extent to which efforts to correct them are able to bring about long-term improvements.

4) Random and systematic errors at the poll level are not the sole determinants of polling error, as the changing composition of the countries in which polling is undertaken over time introduces additional country-level sources of error, offsetting improvements to polling accuracy made by poll-level revisions.

5) Random and systematic errors at the poll level are not the sole determinants of polling error as, in addition to the emergence of new country-level drivers of error over time, differences between elections *within* countries serve to bring about polling error, offsetting improvements to polling accuracy that have been gained by methodological revision at the poll level.

The first explanation can be ruled out with ease. Not only do polling organisations strive to produce the most accurate predictions possible, facing ignominy and damage to their reputations when they fail,[455] but they have demonstrably acted on past failures, iteratively updating their methods in response to identified systematic errors. Accordingly, not only is it in their benefit to want to improve, but actions have been taken to resolve systematic errors. Moreover, even if no attempts to resolve systematic errors were made, the considerable increase in the number of polls conducted over time would reduce random error in the aggregate, as per the law of large numbers. Consequently, if random and systematic errors at

[455] John Curtice, 'So How Well Did They Do? The Polls in the 1997 Election', *Journal of the Market Research Society*, 39.3 (1997), 449-461 (p. 449).

the poll level truly were the sole determinants of polling inaccuracy, then at the very least a gradual reduction in inaccuracy would be expected over time due to diminishing random error, even if systematic errors remained unaddressed.

The actions taken by polling organisations to remedy past failings leads into the second explanation. Although changes have undoubtedly been made to the methods and processes underpinning polling in light of past failures, post-mortems invariably identify different combinations of systematic errors as being responsible for instances of misprediction, as made evident in the previous chapter. Accordingly, it could be argued that corrections made on the basis of past errors may not necessarily lend themselves to the improvement of predictions in the present, thereby undermining the extent to which they are able to bring about improvements in accuracy.

The third explanation makes clear the potential for particularly stubborn issues at the poll level to offset improvements that result from methodological revision at the poll level. The identification of issues at the poll level does not necessarily furnish the polling industry with the ability to resolve them. Some issues, while recognised, may simply be insoluble. Others, while theoretically soluble, may require infeasible levels of resources or time to resolve, rendering their solution impractical. Relatedly, the identification and ostensible correction of poll-level issues does not necessarily eliminate these issues permanently, as their nature or extent may change over time or differ between cases, undermining efforts to resolve them or rendering past solutions unsuitable.

While these insoluble and evolving issues certainly have the potential to offset improvements in polling accuracy that arise from methodological revision at the poll level, that they serve to bring about static error over time in isolation is not a compelling argument when applied to a heterogenous dataset which varies notably in the countries it comprises over time. In such a dataset, it is fundamentally unreasonable to assert that differences in cases over time do not bear on polling error, especially when the variance associated with this error can be seen to rise in tandem with the application of pre-election polling to new cases. As such, while the third explanation makes clear the importance of insoluble and evolving issues at the poll level for the nature of polling error over time, the premise on which it rests – namely, that random and systematic errors at the poll level stand as the sole determinants of polling error – does not reasonably account for the observed static trend of polling error over time.

As the contention that inaccuracy is solely a function of random and systematic errors at the poll level cannot be used to reasonably explain static error over time, the fourth explanation contends that, while the mitigation and correction of random and systematic errors at the poll level are undoubtedly important, it would be remiss to fail to account for the impact of country-level effects on polling error, especially those that arise from the iterative expansion of polling to newly democratised states over time. Indeed, the difficulties of expanding polling to such states is well-recognised, with issues concerning levels of local cooperation with polling organisations,[456] developmental and technological barriers to contacting representative samples of individuals,[457] and difficulties in eliciting political preferences in societies in which the expression of political opinions was previously the focus of repression each being recognised as key contributors to survey error.[458] These issues are such that the incorporation of newly democratised states alongside established democracies over time stands to increase average polling error. In so doing, they stand to offset improvements to polling accuracy that may have been observed in other states as a result of methodological innovation at the poll level, thereby contributing to the largely static nature of average polling error over time.

[456] Seligson, p. 51.
[457] Mattes, pp. 116-117.
[458] Ansu-Kyeremeh, p. 61.

The fifth explanation builds on its predecessor and acknowledges that while sources of polling error at both the poll and country levels each stand as important contributors to polling inaccuracy over time, another possibility exists. It may be that differences between elections *within* the countries to which polling is applied over time stand as a source of polling error. As illustrated in the literature review, while this approach to understanding polling error has existed since the 1930s, it has only recently come into the political scientific mainstream. Despite the nascent nature of election-level analysis, empirical evidence suggests that differences between election meaningfully impact the variance exhibited by polling error.[459] As election-level factors have been largely unincorporated in past assessments of polling error and, by extension, were not widely identified as sources of error requiring attention moving forward, it may be that the error brought about by differences between elections has served to offset improvements to polling accuracy made through poll-level revisions, contributing to the largely static trend of average polling error over time.

By unpacking these explanations, it is clear that it would be unreasonable to assert that the largely static nature of polling accuracy over time is a function of any individual source of error occluding improvement born of methodological revision at the poll level. Rather, it is evident that polling error is plausibly driven by a range of poll-, country-, and election-level factors. While I take care to recognise the importance of drivers of polling error that rest at the poll and country levels, along with those housed at the pollster level, incorporating each of these into later analysis, I principally focus on the impact of differences housed at the election level as drivers of polling error within this thesis. This focus is motivated by the fact that the understanding of election-level factors as drivers of polling is both theoretically and

[459] Tudor and Wall, p. 1.

empirically underdeveloped when compared to its counterparts, as it presently exists at a nascent stage as made evident by the literature review.

With the broad focus of this thesis established, I further unpack the contention that the characteristics possessed by elections represent plausible drivers of variability in polling error. In the following section, I explore the ontological origins of this contention and the challenge it poses to the assumptions that have underpinned previous attempts to understand polling inaccuracy.

3.4: Compositional Heterogeneity: Elections as Sources of Polling Inaccuracy

The contention that differences between elections stand as plausible drivers of polling error is motivated by ontological problems in the theoretical foundations of pre-election polling. That the prevailing approach to understanding and correcting polling inaccuracy has not yielded meaningful reductions in error over time speaks to a deficiency in its theoretical underpinnings. Specifically, it speaks to issues in the assumptions undergirding the approach that has, implicitly or otherwise, served as the basis for analyses of polling inaccuracy to date.

In the following section, I identify that, epistemologically, the study of pre-election polling error is principally based on the understanding that knowledge of polling error and approaches to its avoidance can be gained through unpacking the methods employed by polls in past instances of misprediction. Ontologically, I recognise that polling is often predicated on the implicit understanding of elections as homogenous phenomena. Though, I take care to recognise that this stance is not universal, with emergent studies recognising that adopting a heterogenous understanding of elections as heterogenous phenomena better reflects their nature and the manner they bear on polling error. Given the benefits of adopting a heterogenous view of elections, I call for ontological re-orientation.

To illustrate the need for this re-orientation, I identify the mechanisms through which the heterogeneity of elections can be expected to affect polling inaccuracy and demonstrate that past polling failures have regularly occurred alongside pronounced instances of electoral heterogeneity. On the basis of the likely and substantively plausible impact of the heterogenous nature of elections on polling error, I contend that the study of polling ought to engage in both epistemological and ontological re-orientation to better represent the determinants of error.

The Epistemological and Ontological Foundations of the Study of Polling Error

The modern practice of pre-election polling has employed a consistent approach over its 86-year history. The scientific basis of modern pre-election polling was established by the success of the randomly sampled polls employed by Gallup, Crossley, and Roper in 1936 at the expense of *The Literary Digest*.[460] The puzzle of predicting voting intention had, ostensibly, been found to be soluble on the basis of statistical methods. Four years later, Gallup professed to know what voters were thinking, ‘on the basis not of guesswork, but of facts’.[461] In basing the prediction of public opinion on empirical fact and statistical analysis, an inherently social phenomenon was being interrogated using approaches typically reserved for the natural sciences.

The scientific approach championed by Gallup and company has been apparent throughout pre-election polling and assessments of its performance ever since. Beyond the use of scientific methods to interrogate public opinion, two core assumptions have undergirded the practice of pre-election polling: cross-case generalisability and the continuity of knowledge between phenomena. Implicit within these assumptions is the homogenisation of phenomena. For approaches and understandings to be generalisable and for past knowledge to apply to future

[460] Igo, p. 109.
[461] George Gallup and Saul Forbes Rae, *The Pulse of Democracy: The Public Opinion Poll and How it Works*, (New York: Simon and Schuster, 1940), p. v.

cases, each instance to which they apply must be akin to all others. Equally, for knowledge to be iteratively and continually relevant, the phenomena from which it is derived and to which it is later applied must resemble one another.

As evidenced by the literature review, the development of scholarship on pre-election polling has adhered to these assumptions. In the wake of predictive failures, post-mortems are launched, problems are identified, and methods are updated to avoid these problems moving forward. The iterative revision of methods on the basis of past mistakes is undertaken in search of ever-improving predictive accuracy. This process speaks not only to the epistemological foundations of the study of polling, but also the ontological assumptions underpinning it. That an understanding of polling failures has routinely and predominantly been sought by unpacking polling mechanisms denotes an epistemological understanding that knowledge concerning polling error is gained from polls themselves.

The iterative approach to understanding and resolving polling error by focusing on issues at the poll-level also suggests an ontological foundation centred on the implicit understanding of elections as homogenous phenomena. That is, by ignoring election-level factors in post-election assessments of polling error, analyses implicitly hold that elections bear on polling error to the same degree, as different contests are not afforded different treatment, homogenising them as phenomena. It is important to underscore that this process of homogenisation is implicit and is not an active process employed in assessments of polling error, nor is it universal. Indeed, as shown in the literature review, emergent studies have begun to recognise elections as heterogenous phenomena, moving away from the implicit homogenisation of elections that underpins much of past study.

As identified in the previous section, poll-based approaches to mitigating polling error have failed to yield a meaningful increase in predictive accuracy over the past 77 years, undermining the plausibility of the assumption of continuity between elections on which they rest. I contend that the assumption of continuity between elections is incommensurate with their heterogenous nature as phenomena. This is not a particularly controversial contention, as the progression of social phenomena over time has long been understood to be characterised by a lack of constancy.[462] I hold that this heterogeneity is likely to bear upon polling error and therefore argue that ontological re-orientation away from a focus on continuity between elections towards a recognition of heterogeneity between cases more accurately represents the nature of polling error and allows for a better recognition of those factors that bear upon the accuracy of polls and hamper their performance.

In the following sub-section, I unpack the rationale behind the need for ontological re-orientation in the understanding of polling error by breaking down the heterogenous nature of elections as phenomena. I demonstrate that, compositionally, no two elections are likely to be identical, introducing differences between cases, and undermining the validity of assumptions of continuity. From this, I hold that adopting an ontological view of elections as heterogenous phenomena is more commensurate with their nature and can be used to better understand those circumstances in which polling error is likely to vary.

Heterogeneity Between Cases: The Need for Epistemological and Ontological Re-orientation

The heterogeneous nature of elections is not the result of differences in the core characteristics that they comprise, but from the manner in which these characteristics are arranged between cases. In this sub-section, I identify that while each election contains the same core characteristics, no two are exactly alike in composition. I contend that these differences in

[462] Makridakis, p. 307.

composition lead to subtle, though impactful differences between elections that bear upon polling error.

All democratic elections share a core of generalisable characteristics, as contained within minimalist, electorally focused conceptions of democracy.[463] These characteristics principally concern actors and processes. Meaningful competition between multiple parties or candidates,[464] the equal participation of a widely enfranchised electorate,[465] and the freedom to support and vote for any given candidate or party (and to have that vote counted) all rest at the core of democratic elections as phenomena.[466]

While these characteristics are always present, their prominence varies between cases. The number of parties or candidates contesting an election will often vary, so too will the degree to which the electorate turns out to vote. The loyalties present within the electorate, and therefore the likely target of voters' support, will also ebb and flow between contests. This changeability in the prominence of the core characteristics gives rise to heterogeneity between cases. While one election may comprise a high degree of partisanship, a large number of competing parties, and significant voter turnout, another may exhibit diminished partisanship, fewer parties, and lower turnout. Though both elections possess the same core variables, they do so to differing degrees. In this way, elections can be characteristically similar, but compositionally distinct.

The magnitude of compositional differences will necessarily vary between cases. Successive elections within the same country are more likely to be compositionally similar due to sharing similar electorates and system-level constraints. For example, the number of parties or

[463] Andrea Kendall-Taylor, Natasha Lindstaedt, and Erica Frantz, *Democracies and Authoritarian Regimes* (Oxford: Oxford University Press, 2019), p. 18.
[464] Michael Coppedge, 'Democracy and Dimensions: Comments on Munck and Verkuilen', *Comparative Political Studies*, 35.1 (2002), 35 – 39 (p. 36).
[465] Marc Buhlmann and others, 'The Democracy Barometer: A New Instrument to Measure the Quality of Democracy and its Potential for Comparative Research', *European Political Science*, 11 (2012), 519 – 536 (p. 526).
[466] Coppedge and others, p. 255.

candidates contesting elections is a partial function of the electoral system in which they operate,[467] and changes to electoral systems are rare.[468] Despite this, party systems are often subject to change between elections, even as the overarching electoral system remains stable.[469] Indeed, election systems often serve to bound the extent to which party systems are able to change. Systems such as proportional representation with low thresholds for representation in the legislature are more conducive to the emergence of new parties,[470] while more restrictive systems such as first past the post reduce the likelihood of party emergence.[471]

Given the potential for the emergence of new parties between elections, while successive elections in the same country are more likely to be compositionally similar than temporally successive elections in different countries, they are nevertheless often subject to change between contests. Indeed, in presidential systems with term limits, though successive elections may be contested by the same candidates, once an incumbent president has served the maximum number of terms, a new candidate for a given party must contest the next election, bringing about compositional change between contests.

On the other hand, elections held in different countries, or at greater temporal distances from one another, are more likely to present striking compositional differences. As elections in different countries often operate according to differing electoral systems, the number of parties or candidates contesting them will vary according to the system in place. New political parties

[467] Rein Taagepera and Bernard Grofman, 'Rethinking Duverger's Law: Predicting the Effective Number of Parties in Plurality and PR Systems – Parties Minus Issues Equals One', *European Journal of Political Research*, 13 (1985), 341 – 352 (p. 341).
[468] Pippa Norris, 'Choosing Electoral Systems: Proportional, Majoritarian, and Mixed Systems', *International Political Science Review*, 18.3 (1997), 297 – 312 (p. 297).
[469] Michael Laver and Kenneth Benoit, 'The Evolution of Party Systems between Elections', *American Journal of Political Science*, 47.2 (2003), 215-233 (p. 215).
[470] Alex B. Rivard, 'It is not me, it is you: The emergence of secessionist parties in Western democracies', *Nations and Nationalism*, (2023), 1-20 (p. 6).
[471] Ethan Scheiner, 'The electoral system and Japan's partial transformation: party system consolidation without policy realignment', *Journal of East Asian Studies*, 12.3 (2012), 351-380 (p. 356).

emerge over time,[472] leading to differences in the number of parties contesting elections. The emergence of new parties stands to re-align partisan loyalties amongst electorates, altering their nature. The strength of the partisan loyalties of electorates also varies between countries, as well as waxing and waning over time. So too does the motivation and composition of electorates, leading to differences in turnout levels both between countries and over time.

Despite the variable likelihood of similarity between cases, the probability of two elections possessing identical compositions is vanishingly slim. To illustrate how unlikely exact compositional duplication between cases is, I present a toy model capturing the characteristics possessed by two discrete electoral contests. Reductively, each of these contests will contain variable degrees of partisan sentiment within the electorate, varying proportions of late decision-making, and differing levels of turnout on election day. If these characteristics are measured as a proportion of the electorate, then each is a continuous variable hypothetically able assume any value from 0 to 100. While this holds in theory, it is implausible in reality, as certain values for partisanship, late decision-making, and turnout are considerably more likely than others, bounding the degree to which they vary between cases. For example, a certain proportion of individuals will always turn out for general elections, precluding a value of zero. Similarly, not all voters turn out on election day, even in countries in which compulsory voting is mandated by law,[473] precluding a value of one hundred. Turnout values typically over more restricted ranges, as illustrated by the ranges seen in US presidential elections (between 40%

[472] Simon Hug, *Altering Party Systems: Strategic Behaviour and the Emergence of New Political Parties in Western Democracies* (Ann Arbor: The University of Michigan Press, 2001), pp. 11 – 12; Heloise Nez, 'Podemos: The Emergence of a New Political Party in Spain', in *Contemporary Voting in Europe: Patterns and Trends*, ed. By Alexis Chommeloux and Elizabeth Gibson-Morgan (London: Springer Nature, 2017), p. 113.
[473] Jonathon Louth and Lisa Hill, 'Compulsory Voting in Australia: Turnout with and Without it', *Australian Review of Public Affairs*, 6.1 (2005), 25 – 37 (pp. 26 – 27).

and 80% in the period 1828-1992) and parliamentary elections across Europe (~30% to ~92% in the period 1979-2004).[474]

To capture values that are reasonably representative of reality, the values across which turnout, late decision-making, and partisanship are allowed to vary within my toy model are bounded by the values present in my dataset. Specifically, turnout is allowed to vary between 33% and 96%, late decision-making is allowed to vary between 2% and 50%, and levels of strong partisanship is allowed to vary between 3% and 77%. For ease of illustration, each variable is measured on an integer scale. Therefore, the probability of any given turnout value is $\frac{1}{63}$, the probability of any late decision-making value is $\frac{1}{48}$, and the probability of any given partisanship value is $\frac{1}{74}$. The probability of any individual combination of the three given variables occurring is calculated as shown in equation 1 where $\mathrm{P}(V_{\mathrm{n}})$ is the probability of an individual variable value occurring.

$$\mathrm{P}(V_1 \text{ and } V_2 \text{ and } V_3) = \mathrm{P}(V_1) * \mathrm{P}(V_2) * \mathrm{P}(V_3) \tag{1}$$

Multiplying the individual value probabilities together yields a cumulative probability of $\frac{1}{223{,}776}$ that any given combination of values occurs. So, even in this reductive example, any specific combination of these election-level variable values occurs in isolation only once in over two-hundred and twenty thousand permutations. As such, elections are unlikely to represent the same constellation of characteristics.

While my toy model suggests that exact replication of sets of characteristics between elections is unlikely, it possesses two key shortcomings: the granularity of its measure of change and its

[474] Peter F. Nardulli, Jon K. Dalager, and Donald E. Greco, 'Voter Turnout in U.S. Presidential Elections: An Historical View and Some Speculation', *PS: Political Science and Politics*, 29.3 (1996), 480 – 490 (p. 481); Richard S. Flickinger and Donley T. Studlar, 'One Europe, Many Electorates? Models of Turnout in European Parliament Elections After 2004', *Comparative Political Studies*, 40.4 (2007), 383-404 (p. 385).

implicit assumption of independence between cases. I take each of these in turn. The vanishingly slim probability of the exact replication of characteristics between elections displayed by my toy model is predicated on the existence of *any* difference in the characteristics possessed by elections, irrespective of magnitude. Although this captures the presence of differences, it does not account for whether these differences are meaningful. It may be that while two elections present different sets of characteristics, the differences present are so small as to be of little substantive consequence. For example, a one-point shift in partisanship between elections does not represent a substantial change to the electoral environment and is unlikely to bear on the error exhibited by polls to a meaningful degree. However, a thirty-point shift in partisanship between contests represents a considerable change to the electoral landscape and is more likely to affect the accuracy of polls.

Precedent for large-scale changes in characteristics between elections can be found in a range of real-world cases. For example, levels of partisanship declined sharply between the 1964 and 1966 UK general elections.[475] Similarly, the extent of late decision-making in the electorate declined sharply between the 1992 and 1996 US presidential elections and was followed by a sharp rise between the contests of 1996 and 2000.[476] Levels of turnout also declined sharply between the 1975 and 1978 New Zealand general elections, only to rise sharply between the 1978 and 1981 contests.[477] In each case, the longer-term trends in which these notable changes rest speak to the small-scale, though near-constant, change in electoral characteristics between contests that the earlier toy model captures. While these small-scale changes may be unlikely to bring about polling error, the occurrence of large-scale shifts in characteristics between

[475] Paul R. Abramson, 'Generational Replacement and Partisan Dealignment in Britain and the United States', *British Journal of Political Science*, 8.4 (1978), 505-509 (p. 508).
[476] Brian Box and Joseph Giammo, 'Late Deciders in U.S. Presidential Elections', *The American Review of Politics*, 30 (2009), 333-355 (p. 353).
[477] Jack Nagel, 'Voter Turnout in New Zealand General Elections, 1928-1988', *Political Science*, 40.2 (1988), 16-38 (p. 20).

elections makes real the potential for differences between contests to be sufficient to stand as drivers of polling error.

Though my toy model demonstrates that the probability of any set of election-level characteristics occurring in isolation is slim, it is unrealistic to hold that the characteristics that define elections come about in a vacuum. Indeed, a degree of path dependency can be expected between contests that affects the likelihood of certain election-level characteristics occurring. This path dependency may be representative of established behaviours in given countries or may reflect the continuation of electorally salient factors across contests. By way of an example, consider turnout across sequential elections in the same country. High turnout amongst the electorate in one election speaks to the likelihood of high turnout in the next. The high levels of turnout may speak to a politically engaged and motivated populace that is likely to turn out to a similar degree in subsequent elections. Equally, low turnout in one election may speak to an unmotivated populace that is unlikely to turn out to a significant extent in subsequent elections. Indeed, the tendency for voters to turn out in similar numbers between elections is evident in the progression of turnout levels in parliamentary elections across Europe.[478]

While broad expectations of path dependency can be held, it is important to recognise that its presence across elections may be conditional. In the above example, high turnout in a given election may be the result of a particularly compelling candidate, such as the sharp rise in African American turnout in the 2008 US presidential election in support of Barack Obama.[479] Equally, it may be the result of elections fought on issues that are considered particularly salient

[478] Nicola Maggini, 'The evolution of turnout in European elections from 1979 to 2009', in *The European Parliament Elections of 2014*, ed. by Lorenzo De Sio, Vincenzo Emanuele, and Nicola Maggini, (Rome: CISE, 2014), 31 – 36 (p. 34).

[479] Tasha S. Philpot, Daron R. Shaw, and Ernest B. McGowen, 'Winning the Race: Black Voter Turnout in the 2008 Presidential Election', *Public Opinion Quarterly*, 73.5 (2009). 995-1022 (p. 995).

by voters.[480] As such, the continuation of high turnout between elections may be dependent on subsequent contests being fought by the same candidates or on issues deemed similarly important by the voting population. If these conditions are not met and the electorate is not sufficiently enthused, turnout may drop considerably between contests,[481] leading them to encompass notably different characteristics. Indeed, significant shifts in turnout between contests are not unprecedented, as made evident by the sharp decline in turnout between the 1975 and 1978 general elections in New Zealand.[482] Considerable changes in turnout between elections have the potential to affect the accuracy of the vote share estimates provided by polls by confounding the turnout models on which they rest, as these models often project past turnout levels on to future elections due to the general tendency of path dependency to constrain change between cases.[483]

The conditional nature of path dependency across election characteristics is also evident in the evolution of partisanship between contests. While a degree partisanship will be baked in, with members of the voting population remaining unconditionally loyal to political parties over significant spans of time,[484] some will be more conditional in nature, hinging on specific candidates or the policy platforms adopted by parties.[485] This conditional partisanship is only likely to hold across elections that encompass the candidates or issues that drive it. Instances of change in party leadership or the ideological re-orientation of parties, which itself may be

---

[480] Daniela Braun and others, 'Issues that mobilize Europe. The role of key policy issues for voter turnout in the 2019 European Parliament election', *European Union Politics*, 23.1 (2022), 120-140 (p. 120).
[481] Thomas E. Cavanagh, 'Changes in American Voter Turnout, 1964-1976', *Political Science Quarterly*, 96.1 (1981), 53-65 (p. 56).
[482] Nagel, p. 20.
[483] Anthony Rentsch, Brian F. Schaffner, and Justin H. Gross, 'The Elusive Likely Voter: Improving Electoral Predictions with More Informed Vote-propensity Models', *Public Opinion Quarterly*, 83.4 (2019), 782 – 804 (p. 786).
[484] Alberto Ardevol-Abreu and Homero Gil de Zuniga, 'Obstinate partisanship: political discussion attributes effects on the development of unconditional party loyalty', *International Journal of Communication*, 14 (2020), 324-345 (p.324).
[485] Jonathan Mummolo, Erik Peterson, and Sean Westwood, 'The Limits of Partisan Loyalty', *Political Behaviour*, 43 (2021), 949-972 (p. 949).

an artefact of leadership change,[486] may cause this partisanship to wane, resulting in a notable change in its extent in the following election. The potential for significant shifts in levels of partisanship between sequential elections can be seen in the sharp drop in strong partisan sentiment that occurred between the 1964 and 1966 UK general elections.[487] As considerable shifts in partisanship between elections will meaningfully alter the proportion of individuals likely to vote, as partisans are more likely to vote than unaligned individuals,[488] they represent significant changes to the electoral environment that may impact the accuracy of the vote share estimates provided by polls by confounding the turnout models on which they rest. I further unpack the impact of changes in partisanship on the accuracy of polls in the following sub-section.

The presence of path dependency necessarily makes significant shifts in electoral characteristics between contests less likely, as past electoral environments often inform present circumstances and constrain the degree of change observed between cases. This is illustrated by the relative stability of electoral characteristics between contests in given countries, made evident by the stability of turnout levels in countries across Europe over time.[489] However, the examples provided make real the potential for substantial shifts in electoral characteristics to occur between elections and bear upon the accuracy of the vote share estimates provided by polls.

Given that two elections are unlikely to possess exactly the same constellation of characteristics, they can be expected to vary compositionally on a case-by-case basis. While some of the differences between cases may be so small as to be unlikely to affect the likelihood

[486] Jim Buller and Toby S. James, 'Statecraft and the assessment of national political leaders: The case of New Labour and Tony Blair', *The British Journal of Politics and International Relations*, 14.4 (2012), 534-555 (p. 550).
[487] Abramson, p. 508.
[488] Eli G. Rau, 'Partisanship as Cause, Not Consequence, of Participation', *Comparative Political Studies*, 55.6 (2022), 1021 – 1058 (p. 1021).
[489] Maggini, p. 34.

of polling error, or may be constrained by the path dependency the often governs electoral characteristics, others will be of a sufficient magnitude to plausibly bear upon polls to a degree sufficient to impact their accuracy. A process of investigation that homogenises elections is incapable of capturing these differences and therefore risks missing a potential source of polling error. As such, heterogeneity between cases should be embraced as it not only better reflects the nature of the phenomena of interest, but also opens the door for investigating the extent to which differences between elections bear on polling error. In the following section, I illustrate the manner in which the heterogeneity between elections can be expected to affect polling error and demonstrate the practical need for theoretical realignment.

Why is the Varying Composition of Elections Likely to Affect Polling Accuracy?

There are two principal mechanisms through which the differing composition of elections can be expected to affect the accuracy of polls. The first concerns the values assumed by variables in a given composition. The values taken by characteristics bound elections as phenomena and determine the ease by which they can be predicted. This intuition is based on the recognition of the predictive importance of the characteristics possessed by phenomena within wider disciplines. This understanding finds its roots in metaphysics with the work of Karl Popper who held that phenomena exist on a continuum.[490] At one extreme of this continuum, phenomena resemble clocks and possess regular, orderly characteristics, lending themselves to prediction. On the other, they resemble clouds and comprise disorderly, irregular characteristics, rendering their prediction difficult.[491] A phenomenon comprising a larger number of clock-like characteristics will necessarily lend itself more readily to accurate prediction than one that possesses a greater number of cloud-like characteristics.

[490] Popper, pp. 207 – 208.
[491] Ibid.

The position of phenomena along the continuum is based on the number and severity of the clock- and cloud-like variables that they possess. It is possible for phenomena of the same type to differ in the degree to which they resemble clocks or clouds if the nature of the variables they comprise changes between instances. This change may be the result of the varying presence or absence of variables in their entirety, or differences in the magnitude of the same constellation of variables between cases. Popper framed the variability of erstwhile similar phenomena using cars. Though their mechanical underpinnings do not differ fundamentally, the performance of one brand of car may be further towards the clock-like end of the continuum in terms of its predictability than a less reliably constructed vehicle due to differences in the quality of their internal components.[492] I posit that elections can be understood through a similar lens. While each contest is ostensibly similar, possessing the same set of core characteristics, the values assumed by these characteristics differ from case to case, rendering certain elections more conducive to accurate prediction than others.

The variable prominence of characteristics between elections is of significance for their ability to be accurately predicted when considered in terms of the clock-to-cloud continuum. In the same vein as the variability between cars expounded by Popper, the prominence or relative absence of certain electoral characteristics alters the degree to which elections are clock- or cloud-like and, therefore, the extent to which their future behaviour and outcomes can be accurately predicted. For example, the number and size of the parties contesting an election affect the degree to which it can be consider clock- or cloud-like. An election contested by two large parties and ten smaller parties presents an environment in which fewer parties are vying for meaningful shares of the vote than an election contested by five medium-sized parties. In this example, despite containing a greater number of parties, the first election presents a more

[492] Ibid.

clock-like environment, lending itself to greater polling accuracy, as the accuracy of the vote share estimates that they provide is mostly contingent on capturing the performance of the two large parties and is only minimally affected by the remaining smaller parties. By contrast, in the second election, the accuracy of the vote share estimates provided by polls is contingent on successfully capturing the substantial shares of the vote obtained by five similarly sized parties, each of which could vary considerably. This presents a more cloud-like environment, increasing the likelihood that the vote share estimates provided by polls miss the mark.

Similarly, differences in the extent of partisanship across elections is likely to bear polling accuracy. A larger degree of strong partisan sentiment amongst the electorate is likely to increase the ability of pre-election polls to accurately predict the vote share distribution of an election, as the future decision-making of voters will largely break down along established party lines, rendering it more clock-like. By contrast, an election characterised by minimal or diminished partisanship

The influence exerted by the prominence or relative absence of characteristics upon the accuracy with which elections can be predicted is more easily understood when they are considered as systems. As elections comprise a large number of interconnected characteristics and evolve over a given period of time, they fundamentally exist as dynamic systems.[493] If the outcome of an election is, at least partially, a function of the characteristics it comprises, the nature of these components can be expected to affect its evolution. Electoral characteristics take two forms: constants and variables. The value of constants necessarily remains unchanged over the course of an election, while the value of variables changes. Given their static nature, constants place overarching constraints on the evolution of elections over time. For example, a given election will be contested by a set number of parties (barring last-minute withdrawals or

[493] Bennett I. Bertenthal, 'Dynamical systems: It's about time', in *Data Analytic Techniques for Dynamical Systems*, ed. by S. M. Boker and M. J. Wenger, (Hillside, NJ: Erlbaum, 2007), pp. 1 – 24.

mergers). This number places constraints on the degree to which the vote share in an election can be fragmented, with a larger number of parties permitting a greater amount of fragmentation than a smaller number. The accurate prediction of the finalised vote share distribution can be expected to be consistently more difficult in elections contested by a larger number of parties, given its heightened fragmentation and the need to render a greater number of predictions. This difficulty lends itself to increased polling error.

While constants bound the overall evolution of elections, variables determine their fluctuation over time. The values taken by variables render them variously clock- or cloud-like, with their changeable nature between time points serving to exacerbate or diminish these characteristics, directly affecting the predictability of the election to which they belong. For example, the proportion of undecided voters encompassed by an election variously lends itself to more clock- and cloud-like prediction environments. The number of undecided voters typically declines over the course of an election campaign as voters' partisan preferences are primed and their decision-making begins to crystallise.[494] An election in which a greater proportion of voters remain undecided after this process will present a more changeable and cloud-like environment than an election in which a smaller proportion of voters remain undecided, lending itself to greater polling error.

The compositional differences between elections also bear upon the projection mechanisms that are foundational to poll-based predictions. Three principal projection mechanisms are employed when polls render vote share predictions. The first of these involves projecting findings derived from a sample of respondents onto a larger population, represented by voters on election day. While random sampling, wherein each member of the target population has the same probability of inclusion, represents the ideal basis for the sample-to-population

[494] Gelman and King, pp. 409 – 410.

projection, polling organisations are often forced to engage in departures from this ideal due to issues of non-response and the difficulty of isolating a fully representative sampling frame.[495] Given these issues, polling organisations are often required to engage in post-survey alterations, adjusting responses on the basis of weighting systems to better align with the characteristics of the voting population.[496]

The second projection mechanism concerns establishing likely rates of voter turnout.[497] In addition to ensuring that predictions are based on a large, representative sample of data, polling organisations must ensure that predicted vote share distributions are representative of those individuals within the population who are likely to vote on election day.[498] As the population of individuals *likely* to vote differs from the population of individuals *able* to vote,[499] this process involves identifying the subset of respondents within a sample – itself a subset of the target population – who are likely to vote on election day and, therefore, identifying those stated voting intentions that are likely to be impactful.[500] Approaches taken to identifying likely voters typically rely on using survey questions to establish respondents' past voting behaviour and projecting this onto future elections.[501]

The third projection mechanism involves translating respondents' reported future voting behaviour into estimates of actual voting intention. Though this issue possesses similarities with the projection of likely turnout – reported voting intention is, after all, of no consequence if the individual reporting it does not vote – it chiefly reflects whether reported preferences translate into real-world behaviours on election day. The projection of responses onto future

[495] Tudor and Wall, p. 6.
[496] Ibid.
[497] Hillygus, 2011; Traugott and Tucker, 1984
[498] Rentsch, Schaffner, and Gross, p. 786.
[499] Ibid.
[500] Ibid.
[501] Murray, Riley, and Scime, p. 162.

voting behaviour is complicated by the presence of undecided voters,[502] social desirability bias,[503] and item nonresponse.[504] Approaches to projection often attempt to overcome these issues by relying on differing survey modes,[505] the identification of implicit attitudes and preferences,[506] or establishing respondents' confidence in their stated intentions.[507]

Differences in the composition of electoral characteristics can be expected to impact upon each of these three projection mechanisms. Both the sample-to-population and vote likelihood projections are based on identifying the likely voting population on election day. Here, 'likely' is the operative term, as the voting population on election day is unlikely to exactly mirror the enfranchised population,[508] even in elections conducted under compulsory voting laws.[509] Nevertheless, the process of identifying likely voters in election encompassing compulsory voting is necessarily easier than it is in other states, as the set of likely voters is equal to the enfranchised population. While the process of identifying likely voters varies between polling organisations,[510] respondents' answers to questions regarding their past voting behaviour are often used as a heuristic.[511] Basing the likelihood of future behaviour on the nature of past behaviour requires strong continuity between cases. If this continuity is not present, then two cases are unlikely to resemble one another, undermining the ability of the past to be reliably

[502] Stephen Fisher and Rosalind Shorrocks, 'Collective Failure? Lessons from Combining Forecasts for the UK's Referendum on EU Membership', *Journal of Elections, Public Opinion and Parties*, 28.1 (2018), 59 – 77 (p. 59).
[503] Jazmin Brown-Iannuzzi. Maxine Najle, and Will Gervais, 'The Illusion of Political Tolerance: Social Desirability and Self-reported Voting Preferences', *Social Psychological and Personality Science*, 10.3 (2019), 364 – 373 (p. 364).
[504] Ibid.
[505] Pei-shan Liao, 'Social Desirability Bias and Mode Effects in the Case of Voting Behaviour', *Bulletin de Methodologie Sociologique*, 132.1 (2016), 73 – 83 (pp. 74 – 75).
[506] Luciano Arcuri and others, 'Predicting the Vote: Implicit Attitudes as Predictors of the Future Behaviour of Decided and Undecided Voters', *Political Psychology*, 29.3 (2008), 369 – 387 (p. 370).
[507] Jonathan Mellon, *All You Have to Do is Ask: Measuring Uncertainty in Vote Intention* (2017), <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2958302> [accessed 11/07/2022].
[508] Aina Gallego, 'Understanding Unequal Turnout: Education and Voting in Comparative Perspective', *Electoral Studies*, 29.2 (2010), 239 – 247 (p. 239).
[509] Louth and Hill, p .26.
[510] Rentsch, Schaffner, and Gross, p. 786.
[511] Brian Duff and others, 'Good Excuses: Understanding Who Votes with an Improved Turnout Question', *Public Opinion Quarterly*, 71.1 (2007), 67 – 90 (p. 68).

projected onto the future. A core element of electoral heterogeneity concerns the level of turnout between cases. Differences in the levels of turnout between two elections, especially if substantial and asymmetrically distributed across demographics, have the potential to invalidate inferences based on past behaviour, disrupting pollsters' turnout and voting population projection mechanisms, and increasing the likelihood of prediction error.

Equally, differences in partisan loyalty between elections have the potential to impact upon the turnout-centric projection mechanisms used by polls to render predictions. As the strength of partisan loyalty acts as a determinant of levels of participation within an election,[512] sharp differences in its strength between elections has the potential to profoundly impact turnout levels, undermining the validity of turnout projections, and increasing the likelihood of polling error due to the ensuing mismatch between the compositions of the expected and actual voting populations.

The ability to project voters' responses onto future voting behaviour is likely to be affected by the nature and fluidity of electoral environments. This can be understood through the lens of partisanship. The projection of voting intention from the fieldwork dates over which a poll is conducted on to election day is likely to be more straightforward and less prone to unexpected deviation in electoral environments characterised by rigid, wide-scale partisanship. In such environments, a greater proportion of voters will be loyal to political parties and, therefore, unlikely to change their voting intention between the date at which they are polled and the date they cast their vote at the ballot box. By contrast, in elections characterised by a lower degree of strong partisan loyalty, fewer individuals will be devotedly loyal to political parties and, therefore, a greater proportion of voters has the potential to change their voting intention

[512] Rau, p. 1021.

between the date at which they are polled and election day, making it more difficult to reliably project voter sentiment across this timespan.

Differences in the policy positions taken by parties and candidates between elections can also be expected to influence the presence of social desirability bias, complicating the projection of reported voting intention onto future voting behaviour, and affecting the likelihood of polling error. In an election contested by two uncontroversial candidates adopting conventional policy positions, respondents to pre-election polls are less likely to falsify their responses due to the pressures of social desirability bias, as neither candidate is perceived to be socially undesirable. Due to this, the process of projecting reported voting intention onto future behaviour is more straightforward, decreasing the likelihood of polling error. On the other hand, if one or more of the candidates or parties contesting an election espouses controversial policies, or is deemed socially undesirable, respondents may be more inclined to lie about their voting intentions, rendering the process of projecting this reported intention onto future behaviour more difficult, increasing the potential for polling error.[513]

The impact of respondents lying about their intention to support controversial or socially undesirable candidates or parties has been identified as a potential driver of polling error in past elections.[514] Though it remains disputed,[515] a degree of support has been found to support the contention that 'shy Trump voters' – those individuals who lied about, or were otherwise disinclined to reveal, their intention to vote for Donald Trump when polled due to his controversial nature – were a source of polling error in the 2016 US presidential election.[516]

[513] Noelle-Neumann, p. 307.
[514] Prosser and Mellon, p. 769.
[515] Alexander Coppock, 'Did Shy Trump Supporters Bias the 2016 Polls? Evidence from a Nationally-representative List Experiment', *Statistics, Politics and Policy*, 8.1 (2017), 29-40 (p. 29).
[516] Kevin H. Wozniak, Brian R. Calfano, and Kevin M. Drakulich, 'A "Ferguson Effect" on 2016 Presidential Vote Preference? Findings from a Framing Experiment Examining "Shy Voters" and Cues Related to Policing and Social Unrest', *Social Science Quarterly*, 100.4 (2019), 1023-1038 (p. 1023).

Similarly, 'shy Tories' – respondents who lied about or otherwise failed to disclose their intention to support the Conservative Party when polled due to the perceived societal stigma surrounding it – were identified as a central driver of polling error.[517] Again, while disagreement about its impact exists,[518] the suspected impact of respondents lying about their intention to vote for the Conservative Party on polling error was such that it catalysed post-election analyses dedicated to its interrogation.[519]

In cases such as these, respondents lying about their intended voting behaviour results in polls systematically understating the likely vote share of controversial or socially undesirable candidates or parties,[520] increasing the likelihood that their vote share estimates exhibit systematic error come election day.

In addition to the impact of individual differences in characteristics between elections, inter-related groupings of characteristics are also likely to affect polling error when considered in tandem. For example, together, the number of parties or candidates contesting an election, their respective policy positions, and the strength of partisan loyalty in the electorate combine to affect the likelihood of voter defection and, therefore, the extent to which respondents' stated voting intention can be projected unchanged onto election day.[521] The effects of these variables are, however, variously (non-)compensatory. That is, the effect associated with each of the variables individually can be expected to be amplified or mitigated by the presence of the others. In an election characterised by weaker partisan loyalty, voter defections are likely to be more common, complicating the projection of reported voting intention onto election day behaviour and increasing the likelihood of polling error. These difficulties of projection are

[517] Prosser and Mellon, p. 769.
[518] Mellon and Prosser, p. 661.
[519] Crewe, pp. 341 – 360.
[520] Prosser and Mellon, p. 769.
[521] Paul S. Herrnson and James M. Curry, 'Issue Voting and Partisan Defections in Congressional Elections', *Legislative Studies Quarterly*, 36.2 (2011), 281 – 307 (p. 281).

amplified when an election is conducted in an environment of weak partisanship is also contested by a larger number of parties, as already disloyal voters are presented with a greater number of avenues down which to defect. However, if these parties present disparate policy agendas, defections are less likely, as voters are able to find common ground with fewer parties. This will, in turn, reduce the difficulty of projecting reported voting intention onto election day behaviour, decreasing the likelihood of polling error.

Further impactful, (non-)compensatory combinations of election-level variables can be imagined. The impact of a significant difference in turnout is likely to have a more pronounced impact on polling error if an election is characterised by weak partisan loyalty, as incoming voters are likely to be more widely (and less predictably) distributed, increasingly the likelihood of misattribution and, therefore, error. Conversely, large differences in turnout are likely to be less impactful in the presence of strong partisan loyalty, as additional voters are more likely to distribute themselves predictability along partisan lines, decreasing the chances of misattribution and reducing the likelihood of polling error.

Many of the prospectively impactful electoral characteristics outlined within this sub-section have been present in past instances of polling failure. If those electoral characteristics proposed as determinants of polling error within this chapter had not been witnessed occurring alongside past instances of misprediction, their impact would lack a degree of substantive plausibility. As such, that notable past polling misses have occurred in elections incorporating prospectively impactful characteristics reifies the existence of these characteristics as plausible sources of error. In the following sub-section, I outline a series of past polling failures that have occurred in the presence of prospectively impactful electoral characteristics and expound the way in which these characteristics may have contributed to or accounted for the error exhibited.

The Substantive Plausibility of Electoral Characteristics as Sources of Polling Error

My proposed theory has substantive implications for past instances of polling error. In this sub-section, I reference a selection of prominent polling failures addressed in the literature review that lend themselves to election-level understandings. The purpose of doing so is twofold. In the first instance, through this process, I demonstrate that the theoretical contention that electoral characteristics possess the potential to affect polling error is afforded suggestive support by the real-world circumstances in which poll-based mispredictions have occurred. In the second, by identifying the scope of the applicability of election-level understandings to past polling failures, I illustrate the practical need for a new, election-orientated theory of polling error.

While the poll-level approach adopted in assessments of the 1948 US presidential polling failure was undeniably thorough, the absence of election-level considerations is problematic, as they supplement its main findings. The Social Science Research Council's conclusion that the misallocation of undecided voters was to blame for the failure lends itself to an election-level interpretation. The misallocation of undecided voters is only sufficient to bring about a polling miss when the number of undecided voters is suitably large to tip the balance of power in an election. The 1948 US presidential election was considerably closer than preceding contests.[522] In closer elections, indecision is more likely to have a significant impact, as the number of undecided voters required to affect the outcome is smaller, rendering issues surrounding their allocation more pronounced. As such, the closer electoral margins that characterised the election stand as a plausible contributing factor to the polling miss.

[522] Britannica, *United States Presidential Election Results* (2022), <https://www.britannica.com/topic/United-States-Presidential-Election-Results-1788863> [accessed 16/07/2022].

Additionally, undecided voters are more likely to exist in large enough numbers to alter the substantive results of elections in contests involving a larger number of parties or candidates, as indecision amongst voters is more likely in the presence of a greater range of choices. The 1948 presidential election saw the most significant third-party candidacy since 1924 in the form of the Dixiecrat, Strom Thurmond.[523] Given the increased prominence of a third-party candidate, voters were presented with a greater number of viable candidates from which to choose than in the typical partisan duopoly that dominates US elections. The greater number of viable choices has the potential to have heightened indecision amongst the electorate, increasing the number of undecided voters, and exacerbating problems surrounding their allocation.

The election-level factors affecting undecided voter allocation in 1948 also influence the subject of the Social Science Research Council’s second conclusion: the presence of a late swing in voting intention. Unless their reluctance is the result of social desirability bias, those who state that they are undecided when polled are more likely to make up their mind later in the campaign than those whose state a clear preference. The presence of a greater number of undecided voters therefore increases the likelihood of a late shift in voting intention. As the larger number of viable candidates contesting the election may plausibly have increased indecision amongst the electorate it may, by proxy, have increased the likelihood of a perceived late swing in voting intention, further confounding polling predictions.

In addition to having the potential to amplify poll-level errors, the characteristics of the 1948 US presidential election presented an environment more conducive to polling error than previous contests. Beyond its potential to increase indecision amongst the electorate, the

[523] Nadine Cohodas, *Strom Thurmond and the Politics of Southern Change*, (Mercer University Press, 1994) pp. 1 – 524; Fred E. Haynes, ‘The Significance of the Latest Third-party Movement’, *The Mississippi Valley Historical Review*, 12.2 (1925), 177 – 186 (pp. 177 – 186).

prominence of the third-party candidacy of Strom Thurmond, in certain states at least, increased the fragmentation of the vote and necessitated the accurate prediction of three vote shares. Not only does heightened fragmentation increase the likelihood of error in vote share predictions, given the increased complexity it introduces to the process of vote allocation, but the need to predict support for a significant and emergent third-party candidacy presents further issues. Not only does the emergence of an impactful third party stand to disrupt the practices underpinning polling predictions in a system previously predicated on near-universal two-party competition, but the prediction of the vote share gained by new and emergent parties has been found to be significantly more difficult than it is for their established counterparts.[524] As such, the increased level of multi-party competition in the 1948 US presidential election stands as a plausible contributing factor to the poll-based misprediction, especially given the direction of its failure, and the fact that Thurmond won in states that had typically trended Democratic.

The 1948 US presidential election also exhibited the lowest level of turnout in a presidential election since 1924.[525] This low level of turnout has the potential not only to have confounded the turnout projection mechanisms used by polling organisations, but also to have altered the nature of the voting population sufficiently to undermine the representativeness of sampling procedures.

Although the election-level approach to misprediction was not adopted in assessments of the 1948 polling failure, electoral characteristics were plausible contributory factors to the misprediction. Their relevance to several of the main conclusions drawn in the wake of the 1948 misprediction, along with their potential to have created an electoral environment more

[524] Castillo-Manzano, López-Valpuesta, and Pozo-Barajas, p. 81.

[525] The American Presidency Project, *Voter Turnout in Presidential Elections* (2020), <https://www.presidency.ucsb.edu/statistics/data/voter-turnout-in-presidential-elections> [accessed 30/06/2020].

conducive to error than previous contests, speaks to the problematic nature of their exclusion in contemporary analyses and the importance of their recognition moving forward.

Though the recognition of low turnout as a potential source of polling error in the 1970 UK general election was encouraging, the exclusion of other plausible election-level considerations, such as the closeness of the race, was problematic given the nature of the election itself. The fact that the percentage point difference between the leading candidates in the election was within the margin of error stated by polls – a margin often as high as six percentage points[526] – the closeness of the election existed as an eminently plausible source of error. That is, the electoral environment was such that routine variation in polling estimates that would have been less consequential in a differently composed environment was sufficient to bring about substantive misprediction

That assessments of polling in the February 1974 UK general election did not pursue election-level enquiries is surprising, as due to the success of parties beyond the traditional British duopoly, the failure of polls lent itself to an election-level understanding. The election of February 1974 saw the highest level of meaningful multi-party competition in the United Kingdom since the conclusion of the Second World War.[527] Not only did the Liberal and Scottish National parties almost triple and double their vote shares, respectively,[528] but minor parties collectively won a record-equalling total of 37 seats.[529] The weakening of the previous Labour-Conservative duopoly rendered the election a substantially different environment in which to conduct polls than preceding contests. The presence of a greater number of

[526] Koff, p. 471.
[527] UK Political Info, *1974 February General Election Results Summary* (2022), <https://www.ukpolitical.info/1974Feb.htm> [accessed 16/07/2022].
[528] UK Political Info, *1974 February General Election Results Summary* (2020), <http://www.ukpolitical.info/1974Feb.htm> [accessed 03/07/2020]; UK Political Info, *1970 General Election Results Summary* (2020), <http://www.ukpolitical.info/1970.htm> [accessed 03/07/2020].
[529] Crewe, Särlvik, and Alt, p. 131.

meaningful parties increased the fragmentation of both the vote and seat share distributions. Significant increases in vote fragmentation resulted in erstwhile minor parties becoming important actors, necessitating the accurate prediction of their performances. As this creates the need for a greater number of accurate, per-party predictions to correctly call the overall outcome of an election, it increases the likelihood of misprediction. Given that the misprediction of 1974 was one of seats rather than plurality vote shares, the sudden emergence of minor parties as meaningful seat-winners, further complicated electoral calculus, increasing the likelihood of error.

The record performance of smaller parties in the February 1974 UK general election also speaks to an environment of diminished or re-aligned partisanship. While the deterioration of partisanship along traditional two-party lines within the UK was recognised in wider literature,[530] its impact on the performance of polls during the election was not included in post-mortem assessments. Not only does the deterioration of traditional party loyalties speak to the rise in prominence of smaller parties within the election, but also to an increasingly fluid electoral environment, in which voting intentions cannot necessarily be projected unchanged onto election day due to the increased risk of defections, late swings, and indecision. Due to this, the diminished partisanship that characterised the United Kingdom in 1974 speaks to a challenging electoral environment in which to render predictions that lends itself far more readily to error than previous contests.

Though cited as the cause of many past mispredictions, the failure of polls to detect late swing in French voting intention during the 1978 general election is particularly interesting, as it can be supplemented by an election-level understanding. Despite polls unanimously predicting a

[530] Ibid.

comfortable victory for the Socialist Party one week prior to election day,[531] there was purportedly a shift to the right amongst centrists who had been tempted by the Socialist Party, but whose support had proven fickle.[532] Last minute shifts in voting intention such as this can be viewed through an election-level lens as a function of weak partisanship. In an electoral environment characterised by low levels of partisanship, there is a greater likelihood of a proportion of the electorate sufficiently large to alter its outcome engaging in last-minute decision-making, as strong loyalties are not in place to act as heuristics. By contrast, in an environment characterised by high levels of partisanship, voters' decision-making will predominantly follow staunch party allegiances and will be firmer in nature, reducing the likely of voters changing their mind later in the campaign. As such, the fluid loyalties within the electorate may have plausibly contributed to an electoral environment in which polling error was more likely.

The late swing in voting intention from Carter (a Democrat) to Reagan (a Republican) in the 1980 US presidential election also speaks to an atmosphere of weak partisanship within the election. This notion is supported by the fact that the election saw the most significant third-party candidacies since 1968.[533] The presence of such sizeable third-party support indicates that strong partisan loyalty for the two main parties was diminished in the 1980 contest, increasing the likelihood of shifts in support, such as those observed in the closing stages of the campaign.[534] Given the nature of the late shifts in support attributed to the polling failure,

[531] Wright., p. 24.
[532] Ibid., p. 41.
[533] Howard J. Gold, 'Third Party Voting in Presidential Elections: A Study of Perot, Anderson, and Wallace', *Political Research Quarterly*, 48.4 (1995), 751 – 773 (p. 752).
[534] M. K. Collins, 'The Effect of the Iranian Hostage Crisis on the 1980 Presidential Election', *Tenor of Our Times*, 2.6 (2013), 28 – 35 (p. 34); James Glen Stovall, 'Incumbency and News Coverage of the 1980 Presidential Election Campaign', *The Western Political Quarterly*, 37.4 (1984), 621 – 631 (p. 629); Lee Sigelman and Pamela Johnston Conover, 'The Dynamics of Presidential Support during International Conflict Situations: The Iranian Hostage Crisis', *Political Behaviour*, 3.4 (1981), 303 – 318 (p. 305).

it is not unreasonable to suggest that the failure of poll-based predictions in the 1980 US presidential election was driven by an electoral environment of weakened partisanship.

Beyond conclusions surrounding partisanship voiced in assessments of polling error in the 1982 US mid-terms, election-level factors lend themselves to the poll-level conclusions provided in the literature, most notably issues surrounding question comprehension. It is plausible that issues of question comprehension were not so much a poll-level failing – that is, a mechanistic failure of wording or presentation – but were an artefact of the low-information electoral environments that characterise mid-term elections.[535] In such low information environments, it is difficult for voters to form opinions on the basis of real-time information given its scarcity.[536] Polling questions designed to elicit the rationale and likelihood of voter decision-making, no matter how well designed, are therefore unlikely to successfully capture information, as its presence is diminished by the electoral environment.

Despite the recognition of turnout as a source of error, the absence of further election-level factors in post-election analyses of polling in the 1992 UK general election is problematic due to their plausible impact on the misprediction. Most significantly, the election occurred at the tail end of a steady decline in partisan loyalty within the United Kingdom, exhibiting the lowest levels of very strong partisanship since 1964.[537] It stands to reason that this nadir in partisan loyalty within the UK created a fluid electoral environment that was unlike those preceding it, leading to predictive difficulties. Not only would this decrease the degree to which projection mechanisms based on past behaviour could be used to render effective predictions, but it increases the likelihood of sizeable shifts in voting intention sufficient to confound polling accuracy. By way of reinforcement, comparing the vote share distribution to that of 1987

[535] Jackson, p. 385.
[536] Ibid.
[537] Bobby Duffy and others, *Divided Britain? Polarisation and Fragmentation Trends in the UK*, (London: The Policy Institute, 2019), p. 51.

indicates that the Conservative Party likely gained disaffected Liberal Democrats from the previous election,[538] indicating a shift in support between differing parties.

The characteristics possessed by the 1997 French legislative election lend its associated polling error to election-level understandings. The contest itself was a snap election, as President Chirac dissolved the National Assembly prematurely.[539] Though the length of the campaign was not significantly different from the preceding contest as a result of this,[540] the time for which the impending election was known to the voting public was.[541] As such, voters were forced to engage in election-related decision-making suddenly, in the absence of significant prior notice, and without a lengthy period of time over which their preferences could mature. Consequently, not only would this make indecision amongst the electorate more likely, especially if they were polled earlier in the election cycle, but it would also increase the presence of late decision-making. As an undetected late swing in voting intention was amongst the principal causes identified for the polling miss,[542] it is not unreasonable to suggest that this situation came about due to the constellation of electoral characteristics comprised by the election itself.

The 1997 legislative election also occurred during a period in which the strength of partisan identification was decreasing in France,[543] rendering political allegiances within the electorate more fluid than in past contests and increasing the likelihood of voters moving between opposing parties. As the election saw a significant swing in support towards the Socialist Party

[538] UK Political Info, *1987 General Election Results Summary* (2020), <http://www.ukpolitical.info/1987.htm> [accessed 08/07/2020]; UK Political Info, *1992 General Election Results Summary* (2020), <http://www.ukpolitical.info/1992.htm> [accessed 08/07/2020].
[539] Ibid., p. 71.
[540] Inter-parliamentary Union, *France Parliamentary Chamber: Assemblée Nationale Elections Held in 1997* (1997), <http://archive.ipu.org/parline-e/reports/arc/2113_97.htm> [accessed 13/07/2020].
[541] Ibid.
[542] Hainsworth, p. 73.
[543] Sally Marthaler, 'La course au centre: Policy Convergence and Partisanship in France, 1981 – 2002', *French Politics, Culture and Society*, 28.2 (2010), 75 – 95 (p. 78).

at the expense of the conservative and centre-right Rally for the Republic and Union for French Democracy parties,[544] weak partisanship also serves as plausible determinant of the late swing in voting intention that confounded polls.

The failure of polls to predict the 1998 Quebec general election in Canada lends itself to an election-level approach concerning an environment of weak or realigning partisan loyalty. Quebec has a long history of fluid partisanship.[545] When this is coupled with the large upswing in support for the Action Démocratique de Quebec party in 1998,[546] it is not unreasonable to suggest that a sufficient number of voters could have moved from Parti Québécois to Action Démocratique de Quebec – both parties in support of secession, the key issue in Quebecoise politics[547] – to alter the election outcome and confound the polls. While the rise of Action Démocratique de Quebec at the expense of Parti Québécois was noted,[548] it was not directly attributed to the polling miss.

While issues of sampling could undoubtedly result in overestimation, partisan fluidity also stands as a logical election-level contributor to the overestimation of the performance of Fianna Fáil by polls in the Irish general election of 2002. Though many voters abandoned Fine Gael in favour of Fianna Fáil, the minor parties saw considerable gains in terms of seats during the election.[549] As the polling failure surrounded the misprediction of a Fianna Fáil seat majority,

544 Hainsworth, p. 71.
545 Roger Gibbins, 'Another New West: Environmentalism and the New Policy Agenda', in *How Ottawa Spends 1991-92: The Politics of Fragmentation*, ed. by Frances Abele, (Ottawa: Carleton University Press, 1991), p. 110.
546 Élections Québec, *General Elections November 30, 1998 – December 14 (Masson)* (2020), <https://www.electionsquebec.qc.ca/english/provincial/election-results/general-elections.php?e=18&s=2#s> [accessed 10/07/2020]; Élections Québec, *General Elections September 12, 1994 – October 24 (Saint-Jean)* (2020),<https://www.electionsquebec.qc.ca/english/provincial/election-results/general-elections.php?e=27&s=2#s> [accessed 10/07/2020].
547 John Meisel, 'Unresolved Ambiguity: Quebec After the Election of 1998', *Government and Opposition*, 34.3 (1999), 333 – 351 (pp. 336 – 337).
548 Allan, O'Reilly, and Vengroff, p. 517.
549 Fiachra Kennedy, 'The 2002 General Election in Ireland', *Irish Political Studies*, 17.2 (2002), 95 – 106 (pp. 102 – 103).

this shift is significant and not without precedent. Since 1990, partisanship has been particularly low in Irish elections,[550] increasing the likelihood of shifts in support between parties. Given that minor parties gained four times more seats than the final seat deficit that prevented Fianna Fáil from attaining a majority in the Dáil,[551] it is not unreasonable to suggest that the fluid nature of partisanship within the election, and the attendant likelihood of voters to defect to minor parties, may have existed a determinant of polling error.

The error exhibited by polls in the French presidential election of 2002 also lends itself to an election-level understanding. The rise of National Front in the election came at the expense of Socialist Party,[552] with polls overestimating the likely performance of the latter. As these two parties stood as disparate entities, often occupying polar policy positions,[553] this speaks to an alignment shift in the loyalties held by the electorate. The 2002 presidential election occurred at the nadir of a marked decline in partisanship and the perceived importance of left- and right-wing indicators amongst the electorate.[554] As loyalties within the electorate were at their most fluid, with voters caring little for the ideological orientation of parties, this would increase the likelihood of voters altering their support in ways that would have otherwise been considered unthinkable. As such, it is stands to reason that the unprecedented environment of voter fluidity may have contributed to the polling error witnessed in the French presidential election of 2002.

In the same year, turnout was higher in the Hungarian general election than any past contest held within the country.[555] Moreover, in comparison to the preceding election in 1998, it saw

[550] Gary Murphy and Theresa Reidy, 'Presidential Elections in Ireland: From Partisan Predictability to the End of Loyalty', *Irish Political Studies*, 27.4 (2012), 615 – 634 (p. 615).
[551] Kennedy, p. 103.
[552] Raymond Kuhn, 'The French Presidential and Parliamentary Elections, 2002', *Representation* 39.1 (2002), 44 – 56 (p. 44).
[553] Michael Laver, Kenneth Benoit, and Nicolas Suager, 'Policy Competition in the 2002 French Legislative and Presidential Elections', *European Journal of Political Research*, 45.4 (2006), 667 – 697 (p. 675).
[554] Marthaler, p. 78.
[555] Benoit, p. 119.

significantly increased levels of strong partisan loyalty amongst the electorate.[556] As the Socialist Party and Fidesz had entrenched themselves as the uncontested champions of the Hungarian left and right, respectively,[557] strong partisan support was overwhelmingly subject to a two-party split. As issues of likely voter modelling and non-response were central to poll-level assessments of the failure,[558] it is plausible that the unprecedented levels of turnout and partisanship conspired to bring about a voting population on election day that was sufficiently different to projections and ultimately confounded pre-election polls.

Though the assessment of the two-week moratorium in the 2006 Italian election was encouraging for the advancement of election-level enquiry, the lack of further investigation into electoral factors is problematic, as they offer both a deeper understanding of the effect of the moratorium and plausible alternative sources of polling error. Since the success of Berlusconi and the Forza Italia movement in 1994, Italian politics has been characterised by increasing personalisation, a process through which political candidates replace political parties as the object of support.[559] As the personalisation of a political system increases, the prevalence of traditional partisan loyalty decreases.[560] Given that the personalisation of Italian politics was more prominent in the 2006 campaign than any previous contest,[561] it follows that partisan loyalties would have been at their lowest. As Berlusconi engaged in a last-minute surge of personalised campaigning,[562] it is not unreasonable to suggest that the diminished level of

[556] Zsolt Enyedi and Gábor Tóka, 'The Only Game in Town: Party Politics in Hungary', in *Party Politics in New Democracies*, ed. by Paul Webb and Stephen White, (Oxford: Oxford University Press, 2007), p. 154.
[557] Ibid., p. 155.
[558] Bodor, pp. 450 – 452.
[559] Donatella Campus and Gianfranco Pasquino, 'Leadership in Italy: The Changing Role of Leaders in Elections and in Government', *Journal of Contemporary European Studies*, 14.1 (2006), 25 – 40 (p. 25).
[560] Gideon Rahat and Tamir Sheafer, 'The Personalization(s) of Politics: Israel, 1949 – 2003', *Political Communications*, 24.1 (2007), 65 – 80 (p. 65).
[561] Donatella Campus, 'The 2006 Election: More Than Ever, a Berlusconi-centred Campaign', *Journal of Modern Italian Studies*, 11.4 (2006), 516 – 531 (p. 516).
[562] Ibid., pp.526 – 529.

partisanship was such that a sufficiently large number of voters switched allegiances to confound the polls, especially in light of the moratorium on of late campaign polling.

That a wide array of past polling failures have occurred in electoral environments that lend themselves to election-level understandings of error speaks to the substantive plausibility of a connection between electoral characteristics and the propensity for poll-based misprediction. However, these instances of co-occurrence may be circumstantial and do not ipso facto mean that the composition of the electoral environments observed served as a determinant of the polling error exhibited. However, they do make real their potential to have done so.

As an election-level understanding of polling error on the basis of the characteristics held by elections and their changeable nature between cases is both theoretically motivated and broadly applicable to past instances of misprediction, speaks to the need for epistemological and ontological re-orientation within the study of polling error to better reflect its likely determinants. However, before such theoretical re-orientation can occur, the validity and utility of an election-level understanding to polling error must be established empirically.

When considered together, the changeable nature of the characteristics possessed by elections, the theoretically expected effects of these changes on polling error, and the substantive plausibility of these effects provided by real-world examples of misprediction allow me to draw the first hypothesis with which to test the validity of the election-level theory of polling error outlined in this chapter:

> *H1: Membership within different elections will affect the degree to which polls exhibit error*

If this hypothesis holds, the error exhibited by pre-election polls will exist, at least partially, as a function of the election in which they are conducted and will, therefore, vary between contests. It would be remiss, however, to suggest that polling error can only vary between elections. It will be clear to most keen observers of pre-election polling that error also varies on the basis of the country in which a poll is conducted and the polling organisation conducting it. Therefore, for my hypothesis to truly hold, polling error must not only vary between elections, but this variance must be robust to the presence of country- and organisation-level controls. To test this, in the subsequent chapter I decompose the variance associated with pre-election polling error across a global dataset of elections using a novel multi-level model.

Chapter 4 – Assessing the Importance of Election-level Differences for Polling Error: Decomposing Error Variance Using a Novel Multi-level Model

> *"It is on relatively high levels of abstraction . . . that two different things may be evaluated, spoken of, or dealt with as though they were identical ... what is important is that we realise that [impactful] differences exist".*[563]
>
> \- Wendell Johnson (1946)

Though differences between elections, and the expectation that these differences will be impactful for polling error, can be identified a priori, their relevance and the extent of their impact cannot. In this chapter, I address my second research question and establish the empirical validity of an election-focused ontology of polling error. To achieve this, I demonstrate the importance of election-level differences for polling inaccuracy by analysing their effect on error variation. This chapter also serves to frame the second major contribution of this thesis: the most expansive polling dataset collected to date within political science.

The chapter is divided into five sections. In the first section, I address the novel dataset used within this thesis and unpack the multi-level nature of polling error. To do this, I outline the process through which my dataset was created and provide a novel four-level nested and partially crossed model to represent its fundamentally multi-level nature. I contend that a multi-level structure comprising four levels is the most appropriate approach to analysing polling error and the sources of its variance and substantiate this contention by comparing my four-level model to existing models of polling error. In addition to this, I outline a multi-levelling modelling approach to analysing the importance of the four grouping levels in which sources of polling error variance are housed.

[563] Wendell Johnson, *People in Quandaries: The Semantics of Personal Adjustment*, (New York: Harper and Row, 1946), p. 179.

In the second section, I establish the basis for my inferential analysis. I identify that the measure of interest to assess the impact of election-level differences on polling error is the intra-class correlation coefficient (ICC). I describe the ways in which I measure the ICC for both my continuous and binary measures of polling error, acknowledging that a range of approaches exist through which this can be achieved. I also outline the approaches taken to estimating the parameters of the models from which the ICC will be derived. I provide a comprehensive breakdown of the principal approaches taken to achieving this in multi-level models and identify the most dependable, or preferred, approaches.

In the third section, I describe the ways in which I measure polling error. I put forward eight approaches to measurement, five of which are derived from the literature, with the remaining three existing as novel operationalisations of my own design. These novel operationalisations stand as a secondary contribution of this thesis. In total, my measurement strategies are designed to capture both the random and systematic elements of distributive inaccuracy, as well as the binary and continuous conceptualisations of bounded and substantive inaccuracy outlined in the previous chapter.

In the fourth section, I provide descriptive analysis of the nature and extent of the polling error captured by my dataset. I begin by establishing that membership within election-level groupings is a statistically significant determinant of my measures of polling error. I then visualise the variance of my measures of error between elections. Through this, I demonstrate that levels of observed error vary between elections, lending suggestive evidence to the hypothesis that membership within different elections affects the degree to which polls exhibit error. While I contend that these election-level differences warrant further investigation, I remain cognisant of the presence of potentially confounding factors. I address these through an exploration of the importance of country- and pollster-level differences for polling error, concluding that they must be controlled for in later assessments of the impact of election-level differences.

In the final section, I use the ICC values derived from my models to demonstrate that the election-level is a consistently important driver of polling error variance, even when the country- and pollster-levels are controlled for. Importantly, by decomposing polling error variance across a range of differentially specified models, I robustly demonstrate that the importance of the election-level is not simply an artefact of model specification or case selection.

On the basis of my analysis, I conclude that I can confidently reject the null hypothesis that no relationship exists between polling error and election-level differences. As such, I find support for my alternative hypothesis, H1, that membership within different elections will affect the degree to which polls exhibit error. I argue that this conclusion bears further investigation into the specific election-level differences that drive the variable presence of polling error.

4.1: Polling Dataset, Hierarchically Nested Error, and Multi-level Modelling

Sample Selection and Rationale

To facilitate the analysis within this thesis, I employ a novel polling dataset capturing 11,832 voting intention polls in 497 general elections across 83 countries. My dataset contains polls conducted within the official campaign period of my studied elections. The selection of this sample of polls is motivated by two concerns. The first is the degree to which voting intention polls can reasonably be understood to be predictive of voting behaviour on election day. Within wider polling research, it is well-established that only polls conducted in reasonable proximity to election day provide meaningful information on the likely behaviour of voters on election day and are, therefore, predictive of it.[564] This concerns the progression of voters' preferences and decision-making which are found to slowly crystallise over the course of an election cycle, with crystallisation increasing as election day approaches.[565] Indeed, as elections draw nearer, voters are enlightened as to their partisan identities, informing their vote choice.[566] As such, voters' intended and actual voting behaviour converge over the course of election cycles.[567] This pattern of enlightenment and crystallisation has been identified over a wide range of contests.[568]

At the most basic level, for voters' preferences to crystallise, they must have a target. While in some cases the candidates and parties contesting elections are known considerably in advance, in others their identities and extent are not known until closer to election day.[569] It is unreasonable to expect voters to be able to engage in informed decision-making regarding their intended vote in the absence of a clear array of candidates to choose form. Indeed, candidate evaluation has been found to bear closely on voter decision-making and, therefore, their eventual voting behaviour.[570] In the absence of candidates and parties that are known to be standing in a given election, voters will be making decisions between unknown choices and will therefore provide responses to polls that do not and, importantly, cannot meaningfully or reliably reflect their eventual voting behaviour come election day. As such, polls conducted in this environment, such as those fielded far in advance of election day or before candidates have

[564] Jennings, Lewis-Beck, and Wlezien, p. 949; Tudor and Wall, p. 1; Sturgis and others, p. 765; Sohlberg and Branham, p. 6; Shirani-Mehr and others, p. 5; Jennings and Wlezien, p. 3.
[565] Erikson, Panagopoulos, and Wlezien, pp. 482-483; Gelman and King, p. 409.
[566] Gelman and King, pp. 409 – 410.
[567] Steven E. Finkel, 'Re-examining the Minimal Effects Model in Recent Presidential Elections', *Journal of Politics*, 55 (1993), 1 – 21 (p. 2).
[568] Robert Anderson, James Tilly, and Anthony Heath, 'Political Knowledge and Enlightened Preferences: Party Choice through the Electoral Cycle', *British Journal of Political Science*, 35 (2005), 285-302 (p. 285).
[569] Randall E. Adkins and Andrew J. Dowdle, 'The Money Primary: What Influences the Outcome of Pre-primary Presidential Nomination Fundraising?', *Presidential Studies Quarterly*, 32.2 (2002), 256-275 (p. 264).
[570] Abramowitz, p. 979.

been formally selected by parties, do not provide information on voting intention that can reasonably be used to predict electoral outcomes.

By contrast, election campaigns, contested by known parties and candidates, present environments in which voters can be asked to decide between an array of identifiable choices. In this environment, voters become increasingly informed as to the issue positioning of candidates and parties.[571] Indeed, election campaigns themselves 'deliver the fundamentals' to voters, insofar as they bring the political and economic landscape of an election into clearer and more immediate view.[572] This enables voters to begin aligning this landscape with their own political preferences, allowing their decision-making and voting intention to crystallise. The onset of campaigns also leads voters to view events through their own particular partisan lens,[573] bringing about behaviours, views, and levels of support that may not have been evident outside of the campaign period. As they bring candidates, partisan sentiment, and political preferences into focus, voters are far more able to provide responses to polls during the campaign that meaningfully reflect their intended voting behaviour than they are outside of it.

Beyond this, in countries that do not operate under fixed election calendars, individuals polled outside of campaign periods will not be aware when the next election is being held and will, therefore, not reasonably be able to articulate meaningful preferences regarding it. As such, in these cases, only polls conducted within the official campaign period can be taken to be representative of sentiment that meaningfully relates to future electoral behaviour. Indeed, when coupled with the lack of information surrounding parties and candidates, the inclusion of

---

[571] Erikson, Panagopoulos, and Wlezien, p. 483.

[572] Kevin Arceneaux, 'Do Campaigns Help Voters Learn? A Cross-national Analysis', *British Journal of Political Science*, 36 (2005), 159-173 (p. 159); Lynn Vavreck, *The Message Matters: The Economy and Presidential Campaigns*, (New Jersey: Princeton University Press, 2009), pp. 2-4.

[573] Angus Campbell and others, *The American Voter* (New York: Wiley, 1960), p. 77; Larry Bartels, 'Partisanship and Voting Behaviour, 1952-1996', *American Journal of Political Science*, 44 (2000), 35-50 (p. 35).

polls from outside of election campaigns is likely to introduce significant noise into a dataset and no useful signal.

To further underscore the importance of using in-campaign polls, even in cases where candidates and parties can be identified significantly ahead of an election, they may be subject to change over a long enough time horizon and therefore not represent the choice put to voters on election day. To illustrate this, consider the British electoral cycle of 1979 – 1983. Polls taken across the entirety of the cycle will capture a different set of parties and candidates. In the early stages of the cycle, before the resignation of James Callaghan as Labour leader in 1979 and his replacement by Michael Foot,[574] polls put to voters would elicit preferences that centred on candidates who ultimately would not factor into the contest on election day. Equally, polls conducted early in the election cycle, prior to the rise of the Social Democratic Party,[575] would ask voters to choose between an incorrect array of parties. In cases such as this, polls conducted significantly in advance of an election will not provide information that can reasonably be used to predict election results.

This issue is further underscored by the progression of US presidential elections. In US presidential election cycles, candidates announce their intention to run for office significantly in advance of election day, with individual announcements occurring progressively over a period of weeks and months.[576] This pool of candidates contests primaries, with general election candidates nominated in national party conventions later in the election cycle.[577] With this in mind, not only does the pool of candidates change throughout the early stages of the

[574] John Baylis, *British Defence Policy* (London: Palgrave Macmillan, 1989), p. 92.
[575] Robert Ford and Matthew Goodwin, *Revolt on the Right: Explaining Support for the Radical Right in Britain* (London: Routledge, 2014), p. 1.
[576] Huyen Le and others, 'Bumps and Bruises: Mining Presidential Campaign Announcements on Twitter', *Proceedings of the 28th ACM Conferences on Hypertext and Social Media*, 1 (2017), 215-224 (p. 215).
[577] Shaun Bowler and David M. Farrell, 'The Study of Election Campaigning' in *Electoral Strategies and Political Marketing* ed. by Shaun Bowler and David M. Farrell, (London: Palgrave Macmillan, 1992), 1-23 (p. 12).

election cycle, but many of these candidates do not ultimately bear on the outcome of the presidential election itself. This presents problems for polls conducted significantly in advance of election day. Polls conducted in the early stages of US presidential election cycles will not only capture the extent of voter support for a range of candidates that is not necessarily reflective of the choice they will face on election day, but the popularity of the candidates elicited by these polls – even those candidates that successfully make the general election ballot – may provide inferences that cannot be usefully mapped onto the electoral behaviour of voters. For example, polls conducted in the early stages of the 2008 presidential election did not adequately capture the popularity and eventual electoral success of Barack Obama.[578]

The ability of polls to correctly capture the conditions at play on election day speaks to the second motivation for the selection of in-campaign polls. I confine my dataset to in-campaign polls, as election-level characteristics are often not fixed significantly in advance of election day. Characteristics such as the number of parties or candidates contesting an election, the ideological distance between these candidates, and the level of partisan loyalty will vary over long time horizons, often only stabilising later in an election cycle.[579] Given this, election-level characteristics do not apply equally to polls across the election cycle, precluding even-handed analysis. By way of an example, consider again the 1983 UK general election. The emergence of the Social Democratic Party over the course of the election cycle necessarily altered the number of parties contesting the election.[580] Polls conducted prior to its emergence therefore capture a constellation of parties that is not reflective of those that ultimately contested the election. With the emergence of a new party, these polls also necessarily fail to capture levels of partisan loyalty that are reflective of those at play within the eventual election. Finally, when

[578] Le and others, p. 215.
[579] Gelman and King, p. 409.
[580] Ford and Goodwin, p. 1.

combined with the change in Labour leadership from the centre-left orientation of Callaghan to the most considerably left-wing focus of Foot,[581] polls conducted early in the election cycle will encompass parties that are differently orientated in terms of ideology when compared to those that ultimately contest the election.

As election-level characteristics are, necessarily, set on a per-election basis, the inclusion of polls from early in an election cycle presents significant problems. Due to the changeability of election-level characteristics across long time horizons, those characteristics that eventually come to represent a given election do not apply to all polls conducted across an election cycle. As such, the election-level variables that characterise a given election do not bear on the error presented by all polls that ostensibly relate to it. A given set of electoral characteristics can only be said to stand as drivers of error in those polls to which they reasonably apply. A period of time across which polls are subjected to the same set of election-level characteristics – a set of characteristics that is also representative of the environment come election day – must therefore be identified.

As shown by the earlier example and past research,[582] election-level variables only begin to stabilise later in electoral cycles. These characteristics often only reach levels that are representative of the environment on election day during the official campaign. For example, the range of candidates and parties contesting an election only comes into focus after their official nomination or acceptance,[583] processes which often herald the beginning of an official election campaign.[584] Voters possess inadequate information regarding, and exposure to, candidates prior to election campaigns to allow their opinions regarding them to crystallise.[585]

[581] Baylis, p. 92.
[582] Gelman and King, pp. 409 – 410.
[583] Gideon Rahat, 'Candidate Selection: The Choice Before the Choice', *Journal of Democracy*, 18.1 (2007), 157-170 (p. 157).
[584] Bowler and Farrell, p. 12.
[585] Gelman and King, p. 409.

Similarly, campaigns enlighten voters as to their partisan identities.[586] Consequently, the degree of partisan loyalty they exhibit – and, ultimately, the levels of partisanship that drive their behaviour at the ballot box – cannot be adequately captured outside of the campaign period. Given this, election campaigns present environments in which polls are subject to election-level characteristics that have matured and crystallised adequately to be representative of the environment faced by voters on election day. The process of crystallisation that campaigns bring about also ensures that polls are acted upon by a common set of electoral characteristics. To this end, I limit my dataset to in-campaign polls to ensure that they capture, and are acted upon by, a common set of election-level characteristics that is sufficiently representative of the factors at play on election day to facilitate meaningful and defensible analysis of their impact on polling error.

Confining the sample of polls used within my dataset to those conducted within election campaigns therefore ensures two things. First, it ensures that the polls used for analysis present representations of voting intention that can reasonably be considered to be representative of future voting behaviour. Second, it ensures that polls are affected by a constellation of election-level characteristics that is representative of the election to which they relate. It also allows this thesis to meaningfully contribute to the literature in which it sits, as using a sample of in-campaign polls ensures that analysis is conducted within a similar scope to existing and established studies.

Establishing the beginning of election campaigns, and therefore identifying the point after which a poll can be considered 'in-campaign', does not lend itself to a universal approach. The beginning of the official election campaign period is signalled differently across countries and systems. In parliamentary systems that adhere to the Westminster model, election campaigns

[586] Ibid., p. 410.

can broadly be said to begin after the dissolution of parliament.[587] However, while this holds in certain parliamentary systems,[588] in others, the official campaign period begins after the issuing of an electoral writ.[589] In presidential systems incorporating primaries or two-round systems, election campaigns formerly begin after the selection of candidates.[590] Many countries also hold elections according to fixed election schedules, with elections and campaigns occurring on dates prescribed by law.[591] I take care to account for these differences when identifying the official campaign periods of my studied elections.

To provide examples of the manner in which campaign start dates are determined across my studied elections, in the case of the UK, I take the official election campaign period to begin after the dissolution of parliament in Westminster, in a manner similar to Sanders,[592] Similarly, for the Republic of Ireland, I take the campaign period of my studied elections to begin after the dissolution of the Dáil Éireann.[593] In a similar vein, I take elections in Pakistan to begin

[587] Bowler and Farrell, pp. 11 – 12.
[588] UK Parliament, *Dissolution of Parliament* (2023), <https://www.parliament.uk/about/how/elections-and-voting/general/dissolution/> [accessed 04/08/2023]; Irish Statute Book, *Electoral Act, 1992* (2023), <https://www.irishstatutebook.ie/eli/1992/act/23/section/39/enacted/en/html> [accessed 04/08/2023]; ABC News, *Pakistan election campaign begins* (2013), <https://www.abc.net.au/news/programs/the-world/2013-03-27/pakistan-election-campaign-begins/4598540> [accessed 04/08/2023].
[589] Australian Electoral Commission, *2022 Federal Election Timetable* (2023), <https://www.aec.gov.au/Elections/federal_elections/2022/timetable.htm> [accessed 04/08/2023]; Elections Canada, *The Writ of Election* (2023), <https://www.elections.ca/content.aspx?section=vot&dir=bkg&document=writ&lang=e> [accessed 04/08/2023].
[590] United States Government, *Presidential Election* Process (2022), <https://www.usa.gov/election> [accessed 04/08/2023]; Daily Nation, *Presidential Candidates Present Papers to IEBC* (2013), <https://archive.ph/2013.02.20-235205/http:/elections.nation.co.ke/news/Presidential-candidates-present-papers-to-IEBC-/-/1631868/1674424/-/ty88lrz/-/index.html> [accessed 04/08/2023]; All Africa, *Kenya: Campaign Period to Officially Start on May 29 as IEBC Gazettes Election Data* (2022), <https://allafrica.com/stories/202201240205.html> [accessed 04/08/2023].
[591] Bowler and Farrell, p. 12.
[592] David Sanders, 'Pre-election polling in Britain, 1950 – 1997', *Electoral* Studies, 22, 1-20 (p. 1).
[593] Jane Suiter, 'The Irish Dáil Election 2007', *Irish Political Studies*, 23.1 (2008), 99-110 (pp. 100-101); Michael Gallagher, *Irish Elections 1948-77: Results and Analysis* (Oxford: Routledge, 2009), p. 3.

after the dissolution of its national assembly.[594] In the case of Australia and Canada, election campaigns are taken to begin on the date at which formal electoral writs are issued.[595]

In the same manner as Bowler and Farrell,[596] for US presidential elections, I take the general election campaign period to begin after the candidates contesting the election are formalised by the national conventions of the Democratic and Republican parties. Similarly, I take campaign periods in Kenyan presidential elections to begin after the list of candidates has been approved and published by the Independent Electoral and Boundaries Commission.[597] In the case of presidential elections with two-round systems, the campaign period for the second round is taken as the length of time from the conclusion of the first round to polling day on the second. In those cases where the specific mechanism for triggering the official campaign period is unclear, start dates are taken from scholarship on the nature and duration of election campaigns,[598] governmental accounts of past elections,[599] and contemporary media coverage

[594] Imrana Begum, 'General Election in Pakistan: A Critical Study', *FWU Journal of Social* Sciences, 16.3 (2022), 132 – 143 (p. 133); Dawn, *National Assembly stands dissolved as second successive democratic government completes five-year term* (2018), <https://www.dawn.com/news/amp/1411167> [accessed 04/08/2023].
[595] Elisabeth Gidengil and others, 'Priming and campaign context: Evidence from recent Canadian elections' in *Do Political Campaigns Matter? Campaign Effects in Elections and Referendums* ed. by David M. Farrell and Rudiger Schmitt-Beck, (London: Routledge, 2002), 76-92 (p. 77); Graeme Orr and George Willians, 'Electoral Challenges: Judicial Review of Parliamentary Elections in Australia', *Syndey Law Review,* 23 (2001), 54 – 94 (p. 58).
[596] Bowler and Farrell, p. 12.
[597] All Africa, *Kenya: Campaign Period to Officially Start on May 29 as IEBC Gazettes Election Data* (2022), <https://allafrica.com/stories/202201240205.html> [accessed 04/08/2023].
[598] David Day, *John Curtin: A Life* (Sydney: Harper Collins, 1999), p. 508; Natasha Lindstaedt, *Democratic Decay and Authoritarian Resurgence* (Bristol: Bristol University Press, 2001), pp. 278 – 280; Frank Esser, 'Dimensions of Political News Cultures: Sound Bite and Image Bite News in France, Germany, Great Britain and the United States', *The International Journal of Press/Politics*, 13.4 (2008), 401 – 428 (p. 412); Ivana Feric and Vesna Lamza Posavec, 'Opinion Polls, Voters' Intentions and Expectations in the 2011 Croatian Parliamentary Elections', *European Quarterly of Political Attitudes and Mentalities*, 2.4 (2013), 4-15 (p. 7).
[599] Parliament of Canada, *Length of Federal Elections* (2023), <https://web.archive.org/web/20150924131725/http:/www.parl.gc.ca/Parlinfo/Compilations/ElectionsAndRidings/LengthCampaigns.aspx> [accessed 04/08/2023]; Australian Electoral Commission, *Federal Elections* (2023), <https://www.aec.gov.au/Elections/federal_elections/> [accessed 04/08/2023]; Australian Electoral Commission, *Election dates 1901 – present* (2023), <https://www.aec.gov.au/Elections/federal_elections/election-dates.htm> [accessed 04/08/2023]; US Department of Justice, *CRS Report for Congress: Croatia 2003 Elections and New Government* (2004), <https://www.justice.gov/file/199811/download> [accessed 04/08/2023].

of my studied contests.[600] Ultimately, the official campaign periods within my dataset have a maximum extent of 150 days in the case of legislative elections and 90 days in the case of presidential contests, meaning that, even at their most extreme, they still sit within the timeframes across which polls are considered meaningfully predictive of electoral outcomes.[601]

In addition to being conducted during election campaigns, the polls within my dataset relate to global general elections held between 1936 and 2020. I define general elections as those which involve the entirety of the enfranchised population of a country, rather than a geographically bounded subset. As such, I do not include local or regional elections within my data. As explained in the literature review, scientific pre-election polling as it is understood today emerged in the USA prior to the Second World War during the 1936 presidential election.[602] 1936 therefore serves as the starting point of the dataset. Other analyses of pre-election polling have often limited their scope to elections that occurred after the conclusion of the Second World War in 1945 due in large part to the cancellation or suspension of elections by many warring countries during the conflict.[603] However, non-European democracies – most notably

[600] ABC News, *Lube, legs and lies: Lacklustre start to federal election campaign sees politicians resort to personal attacks* (2019), <https://www.abc.net.au/news/2019-04-12/federal-election-2019-day-one-taxes-legs-lubricant/10996876> [accessed 04/08/2023]; News Room 1015, *Federal Election 2016: Start of Week 2 of the 8 Week Campaign* (2016), <https://1015fm.com.au/2016/05/federal-election-2016-start-of-week-2-of-the-8-week-campaign/> [accessed 04/08/2023]; ABC News, *Australian election campaign begins* (2013), <https://www.abc.net.au/news/programs/the-world/2013-08-05/australian-election-campaign-begins/4866762> [accessed 04/08/2023]; News 24, *Kibaki: I deserve another term* (2007), <https://web.archive.org/web/20090112031044/http:/www.news24.com/News24/Africa/News/0,,2-11-1447_2193347,00.html> [accessed 04/08/2023]; Asharq Al-Aswat, *Campaigning begins for Tunisia's parliamentary elections* (2014), <https://web.archive.org/web/20141008171651/http:/www.aawsat.net/2014/10/article55337236> [accessed 04/08/2023]; Kyiv Post, *Ukrainian parliament reduces presidential campaign to 90 days* (2009), <https://archive.kyivpost.com/article/content/ukraine-politics/ukrainian-parliament-reduces-presidential-campaign-45841.html> [accessed 04/08/2023]; SE Times, *Parties Jockey for support well ahead of Serbia's elections* (2011), https://web.archive.org/web/20111017132048/http:/setimes.com/cocoon/setimes/xhtml/en_GB/features/setimes/features/2011/08/09/feature-04 [accessed 04/08/2023].

[601] Jennings, Lewis-Beck, and Wlezien, p. 960.

[602] Igo, p. 109.

[603] Helmut Norpoth, 'To Change or Not to Change Horses: The World War II Elections', *Presidential Studies Quarterly*, 42.2 (2012), 324 – 342 (p. 324).

the USA, Canada, and Australia – held wartime elections in which pre-election polling was conducted.[604] Therefore, while 1945 serves as a logical starting point for studies into elections and polling within Europe due to the wartime disruption of electoral timetables, it is less applicable to studies which are global in nature such as that which I conduct within this thesis.

Though it is global in nature, presently existing as the only polling dataset to encompass countries from each populated continent of the world, care was taken to ensure that those countries and elections included within the dataset would be capable of providing meaningful and comparable data. To facilitate this, I established two necessary conditions for inclusion:

(1) *Countries must hold meaningfully contested, democratic elections.*

(2) *Elections must possess pre-election polls that focus on vote share projections.*

To satisfy the first necessary condition, a country must earn a score of ≥6 within the Polity V dataset in an election year for that election to be included within the dataset. This score indicates that a country is conducting free and fair elections at that time which yield reliable results.[605] This condition does not, necessarily, reflect the quality of the polls conducted for an election, but rather the reliability of the results against which their accuracy is judged. If these results cannot be relied upon, then neither can the outcome of calculations incorporating them, most notably measures of polling error. As such, ensuring the reliability of electoral results is crucial for the validity of later analytical outputs.

To satisfy the second necessary condition, organisations conducting voting intention polling for a given election must produce and release vote share projections on the basis of their

[604] Richard Moe, *Roosevelt's Second Act: The Election of 1940 and the Politics of War*, (New York: Oxford University Press), p. xiv; David M. Jordan, *FDR, Dewey, and the Election of 1944*, (Bloomington: Indiana University Press, 2011), pp. 1 – 2; Lionel H. Laing, 'The Pattern of Canadian Politics: The Elections of 1945', *American Political Science Review*, 4 (1946), 760 – 765 (p. 760); Murray Goot, 'Labor's 1943 Landslide: Political Market Research, Evatt, and the Public Opinion Polls', *Labour History*, 106 (2014), 149 – 166 (p. 149).

[605] Centre for Systematic Peace, *Polity5: Regime Authority Characteristics and Transitions Datasets* (2018), <https://www.systemicpeace.org/inscrdata.html> [accessed 21/01/2022].

fieldwork rather than projected seat shares. This is to ensure the tractability and comparability of measures of polling error across studied elections. Not only do the mechanisms for projecting vote share to seat share vary between electoral systems and countries, but they also vary within those countries that have undergone changes in electoral system during the period 1936 – 2020, such as New Zealand, Armenia, and Mongolia.[606]

The requirement for the findings of polls to be represented as vote share percentages results in the exclusion of certain established democracies, most notably India and Israel, altogether, as these countries principally produce seat share predictions from pre-election polling, rather than the more widely employed vote share predictions.[607] While the reasoning behind this focus on seat shares chiefly rests on media pressures for more specific predictions and the attendant re-orientation of pre-election polling after the 1977 Knesset election in Israel,[608] and the fact that each seat in Indian legislative elections represents a unique, presidential-style election necessitating its own differently specified approach to prediction,[609] a full exploration of the idiosyncrasies of these issues is beyond the scope of this thesis.

The polls used to create my novel polling dataset were gathered from a variety of online and offline sources. I gathered pre-election polls from a range of publicly accessible polling datasets including, but not limited to, those curated by Jennings and Wlezien,[610] Wells,[611] the

[606] Keith Jackson and Alan McRobie, *New Zealand Adopts Proportional Representation: Accident? Design? Evolution?*, (New York: Routledge, 1998), pp. 1 – 2; Julian Dierkes, 'Mongolia in 2016', *Asian Survey*, 57.1 (2017), 128 – 134 (p. 129).
[607] Roberto Cerina and Raymond Duck, 'Polling India via Regression and Post-stratification of Non-probability Online Sample', *PLoS ONE*, 16.11 (2021), 1 – 34 (p. 1); Shamir, p. 62.
[608] Shamir, p. 64.
[609] C. S. Krishna, *C. S. Krishna: The Problem with Pre-poll Surveys* (2014), <https://www.business-standard.com/article/opinion/c-s-krishna-the-problem-with-pre-poll-surveys-114051000919_1.html> [accessed 27/05/2022].
[610] Jennings and Wlezien, *Replication Data for: Election Polling Errors Across Time and Space*.
[611] Anthony Wells, *UK Polling Report: Survey and Polling News from YouGov's Anthony Wells* (2019), <https://ukpollingreport.co.uk/> [accessed 23/12/2020].

Guardian,[612] Pack,[613] FiveThirtyEight,[614] and the Financial Times.[615] I also gathered polling results from online data published by polling organisations including, but not limited to, Gallup,[616] YouGov,[617] IPSOS,[618] Harris,[619] Kantar,[620] and Roy Morgan.[621] These sources were supplemented with keyword searches for online media coverage of global general elections from September 1936 – 2020, existing aggregations of national polling data,[622] and election-specific literature, such as government reports, private post-mortem analyses, and the Nuffield Election Studies.[623]

Through the process of assembling my dataset, it became increasingly apparent that polling error was housed within four distinct grouping levels: the poll, pollster, country, and election levels. In the following sub-section, I expand on the nature of these grouping levels and unpack the intuitively multi-level nature of polling error, establishing its importance for understanding sources of inaccuracy.

---

612 The Guardian and ICM, *All Guardian/ICM Poll Results* (2019), https://docs.google.com/spreadsheets/d/1oHcxlAbkTJmqfOxYQM22cvjjjRf5pETIF30x7L-qybc/edit#gid=0 [accessed 19 January 2019].

613 Mark Pack, *Pollbase: Opinion Polls Database from 1943-Today* (2019), <https://www.markpack.org.uk/opinion-polls/> [accessed 23/12/2020].

614 FiveThirtyEight, *Latest Polls* (2020), <https://projects.fivethirtyeight.com/polls/president-general/> [accessed 23 December 2020].

615 The Financial Times, *UK General Election Poll Tracker* (2019), < https://www.ft.com/content/263615ca-d873-11e9-8f9b-77216ebe1f17> [accessed 23/12/2020].

616 George H. Gallup, *The Gallup Poll: Public Opinion 1935 – 1971. Volume One: 1935 – 1948*, (Random House: New York, 1972), pp. 31 – 249.

617 YouGov, *Political Tracker Archive* (2022), <https://yougov.co.uk/topics/politics/explore/topic/Political_tracker_archive> [accessed 12/09/2022].

618 Ipsos, *Political Monitor Archive* (2022), <https://www.ipsos.com/en-uk/political-monitor-archive> [accessed 12/09/2022].

619 The Harris Poll, *Poll Archive* (2022), <https://harvardharrispoll.com/> [accessed 12/09/2022].

620 Kantar, *Kantar Public UK Polling Archive* (2022), <https://www.kantar.com/expertise/policy-society/kantar-public-uk-polling-archive> [accessed 12/09/2022].

621 Curia, *Archives: Roy Morgan* (2022), <https://www.curia.co.nz/company/roy-morgan/> [accessed 12/09/2022].

622 MIT Election Data and Science Lab, *Data* (2022), <https://electionlab.mit.edu/data> [accessed 12/09/2022]; Opinium, *Political Polling* (2022), <https://www.opinium.com/resource-center/> [accessed 12/09/2022].

623 David Butler and Richard Rose, *The British General Election of 1959*, (London: Palgrave Macmillan, 1960), pp. 10 – 15; David Butler, Anthony King, and Fintan Hoey, *The British General Election of 1966* (London: Palgrave Macmillan, 1966), pp. 31 – 33; David Butler and Dennis Kavanagh, *The British General Election of 2001* (London: Palgrave Macmillan, 2002), pp. 277 – 282.

The Multi-level Nature of Polling Error: Unpacking a Novel Four-level Approach

Multi-level structures comprise multiple units of analysis that are ordered hierarchically.[624] In this hierarchy, units of analysis are grouped at differing levels of abstraction. Those units grouped at lower levels of abstraction are nested within, or crossed between, those grouped at higher levels of abstraction.[625] Nesting occurs when each individual unit of analysis at a given level belongs to a single unit in another level, while crossing occurs when units of analysis in a given level belong to more than one unit in another.[626]

Polling error is an intuitively multi-level phenomenon. The assertion that polls and their attendant errors are housed within and informed by a multi-level structure is not a new idea. Research has been conducted to analyse polling error using multi-level decomposition techniques,[627] while organisations dedicated to predicting election results have relied on multi-level techniques for years, albeit often implicitly. In this sub-section, I put forward a novel multi-level understanding of sources of polling error. Where my approach differentiates itself from past multi-level approaches to understanding polling error is in its structure. While previous multi-level models of polling and its associated error have employed no more than three levels, with these levels taken to adopt a fully nested structure,[628] I contend that a four-level nested and partially crossed structure better represents the multi-level reality of pre-election polling and its sources of error. This structure is displayed in Figure 7.

[624] Marco R. Steenbergen and Bradford S. Jones, 'Modelling Multilevel Data Structures', *American Journal of Political Science*, 46.1 (2002), 218 – 237 (p. 219).
[625] Ibid., p. 218.
[626] Emmeke Aarts et al., 'A Solution to Dependency: Using Multilevel Analysis to Accommodate Nested Data', *Nature Neuroscience*, 17.4 (2014), 491 – 496 (p. 491).
[627] Tudor and Wall, p. 12.
[628] Ibid.

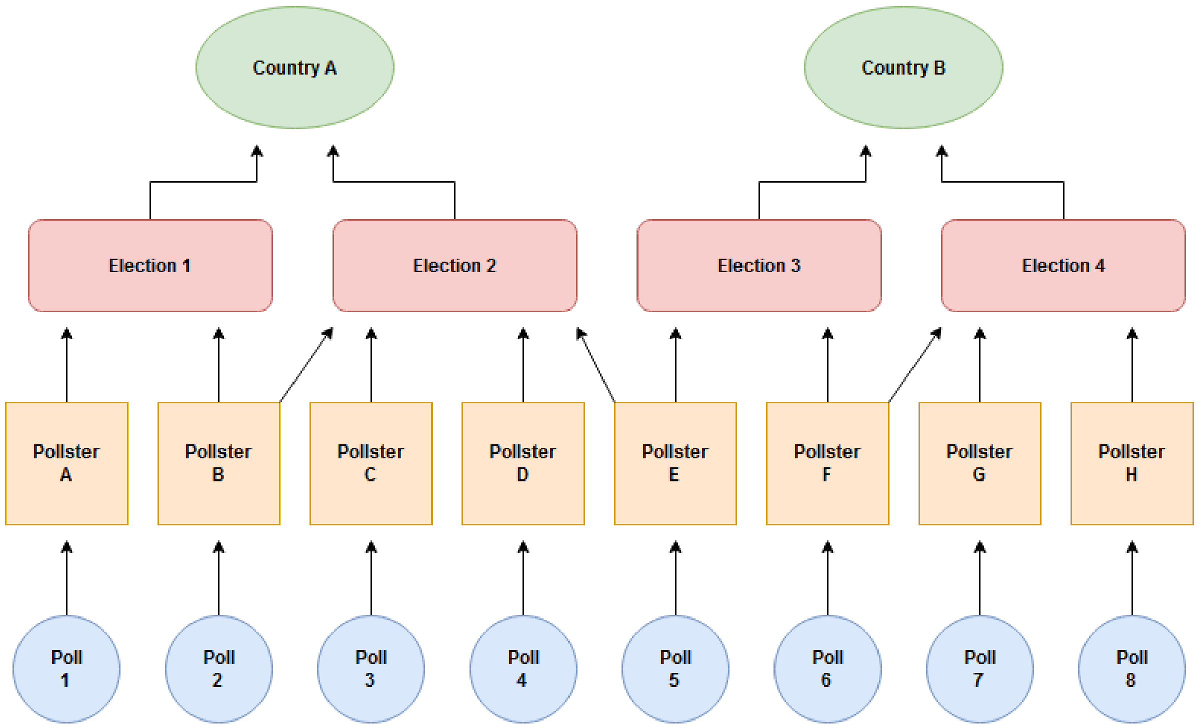

Figure 7: The four-level hierarchical structure of sources of polling error. Interpreted from bottom to top, individual polls (blue) are nested within individual polling organisations (orange), polling organisations are partially crossed between elections (red), and elections are fully nested within individual countries (green).

My multi-level structure of sources of polling error displayed in Figure 7 contains four distinct levels. The first of these levels is the poll level which exists at the bottom of the hierarchical structure. The poll level comprises individual pre-election polls and their measurements of voting intention. As unpacked in the literature review, polls are well understood to possess characteristics that are variously conducive to polling error. The presence of the poll level within my four-level structure is designed to capture these sources of error and their effect on polling inaccuracy.

Individual pre-election polls are conducted by individual polling organisations. Each poll is therefore nested within a single polling organisation and cannot belong to more than one organisation.[629] In this way, polls are grouped within given polling organisations. This

[629] Aarts et al., p. 491.

relationship is captured by my multi-level structure and denoted by the individual arrows connecting each poll to a single polling organisation within Figure 7.

As polls are nested within polling organisations, these organisations sit at the second level of the hierarchy. I refer to this grouping level as the pollster level. The pollster level comprises individual polling organisations and their associated methods and practices. From house effects and issues of post-survey weightings,[630] to partisan leanings and concerns surrounding political sponsorship,[631] the nature and actions of polling organisations are well understood to bear upon polling error. The presence of the pollster level within my four-level structure is designed to capture these sources of error and their effect on polling inaccuracy.

While individual polls are nested in a given polling organisation, and are therefore affected by the characteristics specific to that organisation, different polls are often nested within differing organisations. As such, they are subject to different organisational characteristics that can be expected to differentially affect the degree to which they exhibit error. The presence of the pollster level within my multi-level structure also serves to capture the effect of these differences on polling error and, therefore, the impact of the polling organisation in which polls are nested on their propensity for inaccuracy.

While most polling organisations conduct polls across multiple elections, certain polling organisations only conduct polls in relation to individual elections (within my data, examples include iPoll in Ghana and Kult in Albania). As polling organisations may either be exclusively grouped within individual elections or exist across multiple contests they are most appropriately understood as being partially crossed with elections.[632] This association is captured by the third level of my multi-level structure of polling error, referred to as the election

[630] Jackman, p. 500; Pickup and Johnson, pp. 272 – 284; Bergman and Holmquist, p. 307.
[631] Shamir, p.62.
[632] Ibid.

level, and denoted by the arrows variously connecting polling organisations to one or more elections.

As polling organisations conduct polls within different elections, these polls are necessarily affected by different constellations of electoral characteristics that arise from the heterogeneity of elections as phenomena outlined in the previous chapter. Within my multi-level structure, the election level is designed to capture the effect of these differences on polling error and, therefore, the impact of the election in which polls are conducted on their propensity to exhibit inaccuracy.

As the elections contained within my dataset are general elections, they cannot relate to multiple countries. Here, a country is defined as a sovereign state. So, while a general election may span multiple nations within a state – such as Wales, Scotland, and Northern Ireland within UK general elections – the election still only pertains to one sovereign state, or country. This stands in contrast to elections that span multiple countries, such as EU parliamentary elections, which are not addressed within this thesis. Given the focus on general elections, individual elections are therefore taken to occur in individual countries. From this, elections are most appropriately understood to be fully nested within given countries. This grouping arrangement is captured by the fourth level of my multi-level structure, referred to as the country level, and represented by the arrows connecting elections to individual countries.

The countries in which elections are held necessarily comprise differing constellations of characteristics. They differ to varying extents in terms of culture, population composition, affluence, electoral system, and governmental structure amongst many other variables. The impact of differences between countries on the propensity for polls to exhibit error has been recognised within both academic and industry-adjacent literature.[633] The inclusion of the country grouping level within my multi-level structure is designed to capture the effect of differences between the countries in which the elections that polls attempt to predict are held.

Each of the four levels within my multi-level structure is important to fully understanding and decomposing sources of polling error, as each level comprises variables that can be expected to bear upon the inaccuracy of polls. Without their inclusion, key contributing factors to polling error will be missing from analysis. However, no previous assessment of polling error has adopted a comparable four-level structure. To demonstrate why my four-level approach is vital to properly understanding sources of polling error – and, more specifically, the importance of the election level itself – I compare it to existing models, highlighting their drawbacks and identifying how my model corrects for them.

While multi-level approaches have been employed to better understand and model polling error, the models employed are often reductive. The most straightforward approach adopted in the emerging field of polling error decomposition is two-level in nature. Studies adopting this approach directly investigate the impact of election-level characteristics on polling error absent consideration for other grouping factors.[634] This understanding of the nature of polling error can be understood as a nested, two-level model as represented below in Figure 8.

[633] Tudor and Wall, p. 1; Sohlberg and Branham, p. 8; Mellon and Prosser, p. 662; Durand, 'The Polls of the 2007 French Presidential Campaign', pp. 275 – 298; Jon Puleston, *Are We Getting Worse at Political Polling?* (2017), <https://shop.esomar.org/uploads/public/events-and-awards/events/2017/congress/documents/ESOMAR-Congress-2017_185446_61_Puleston.pdf> [accessed 16/09/2022].

[634] Jennings and Wlezien, p. 280; Tudor, p. 41.

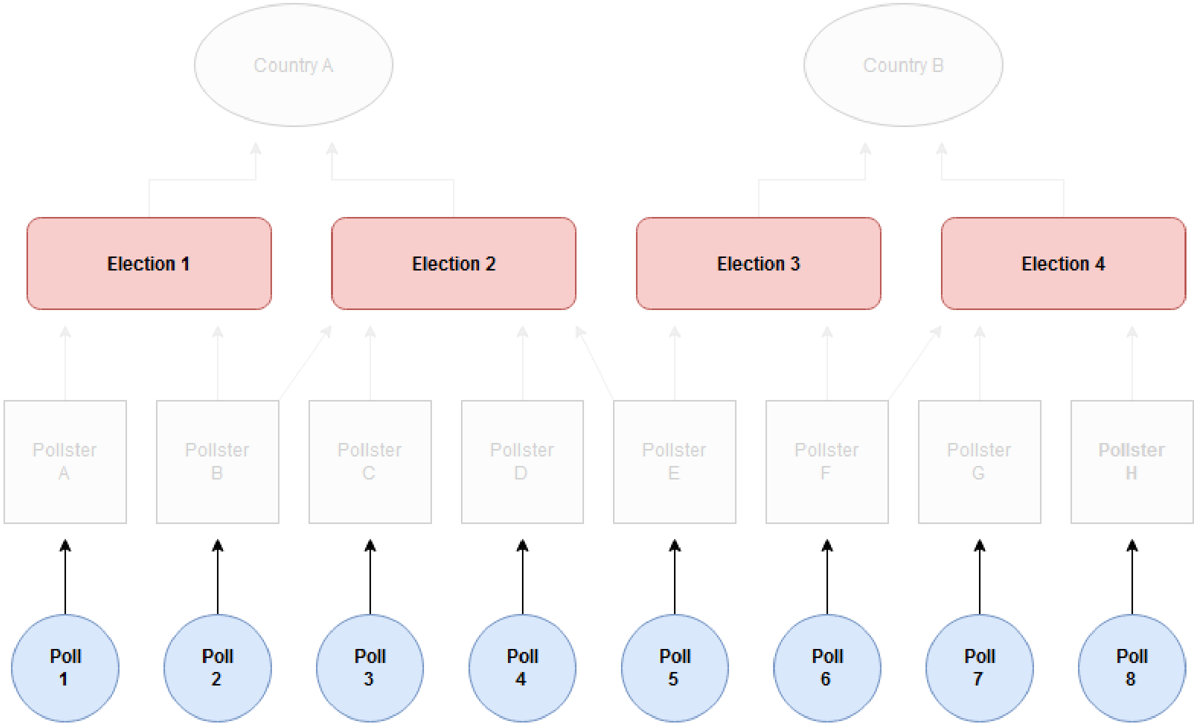

Figure 8: The two-level understanding of sources of polling error that has emerged within the literature, with polls directly nested within elections. This approach does not account for either the pollster- or country-level grouping factors (greyed out within the figure).

As is clear in Figure 8, under the two-level understanding, polls are taken to be directly nested within elections which exist as the highest level of abstraction within the data structure. Both the pollster and country grouping levels that house sources of polling error are excluded under this structure. Adopting a two-level data structure with polls nested directly within elections presents a series of advantages. Most prominently, when used in investigations into the impact of election-level characteristics on polling error,[635] it presents the clearest relationship between polls and the grouping factor of interest. It also requires the least complex modelling approach

[635] Ibid.

given the lack of multiple group membership on the part of the units, or cross-classification,[636] resulting in less intense computational requirements.[637]

Despite the direct nature of the relationships it comprises and its relative computational efficiency, adopting a two-level approach to understanding polling error presents a series of critical drawbacks. As can be seen in Figure 8, adopting a two-level approach to understanding sources of polling error neglects the nesting of polls within polling organisations and the nesting of elections within countries. Any analysis of the impact of the election-level grouping on polling error using this model will therefore be conducted absent factors contained within these grouping levels.

The exclusion of the country and pollster grouping levels within two-level models excludes two sets of variables that are likely to bear upon polling error. The pollster level incorporates variables such as the differing partisan leanings of polling organisations, the impact of house effects, and the effect of political sponsorship, all of which are well understood to affect polling error.[638] To omit such variables from the study of polling error is to fail to capture important sources of its variation. The failure to capture differences between polling organisations in two-level models of polling error is remedied by the inclusion of the pollster-level in my four-level structure.

Differences between countries are widely understood to be of consequence for polling error.[639] These differences manifest themselves in a variety of ways which can be expected to bear upon the accuracy of polls. Certain countries enforce polling moratoriums banning pre-election

[636] Martin Krzywinski, Naomi Altman, and Paul Blainey, 'Nested Designs', *Nature Methods*, 11 (2014), 977 – 978 (p. 977).
[637] David Melamed and Mike Vuolo, 'Assessing Differences Between Nested and Cross-classified Hierarchical Models', *Sociological Methodology*, 49.1 (2019), 220 – 257 (pp. 226 – 227).
[638] Jackman, p. 500; Pickup and Johnson, pp. 272 – 284; Bergman and Holmquist, p. 307; Shamir, p. 62.
[639] Tudor and Wall, p. 1; Sohlberg and Branham, p. 8; Mellon and Prosser, p. 662; Durand, 'The Polls of the 2007 French Presidential Campaign', pp. 275 – 298.

polling for a period of time prior to election day.[640] This makes it more difficult for pre-election polls to capture late swings in voter decision-making, leading to increased error likelihood. Countries also possess different electoral systems which may differentially lend themselves to the accurate prediction of outcomes. The varying levels of inequality and development within countries, captured by indices such as GINI and GDP, may also impact the ability of polls to render accurate predictions. Greater inequality may lead to difficulties in reaching lower income individuals, so too could heightened levels of poverty, as individuals may be less readily contactable (or uncontactable) via certain polling mediums. For reasons such as these, the exclusion of differences between countries within two-level understandings of polling error results in their failure to capture potentially impactful differences between countries. This deficiency is remedied within my four-level model of polling error through the inclusion of the country grouping level.

While the country grouping level is absent in two-level understandings of polling error, it has been incorporated into three-level models. Presently, only one such model has been used within academic research.[641] This understanding takes polls to be nested within individual elections and these elections to be nested in turn within individual countries. This three-level structure is represented below in Figure 11.

[640] Lynda Lee Kaid and Jesper Strömbäck, "Election News Coverage Around the World: A Comparative Perspective", *The Handbook of Election News Coverage Around the World* (New York: Routledge, 2009), pp. 441 – 452.

[641] Tudor and Wall, p. 12.

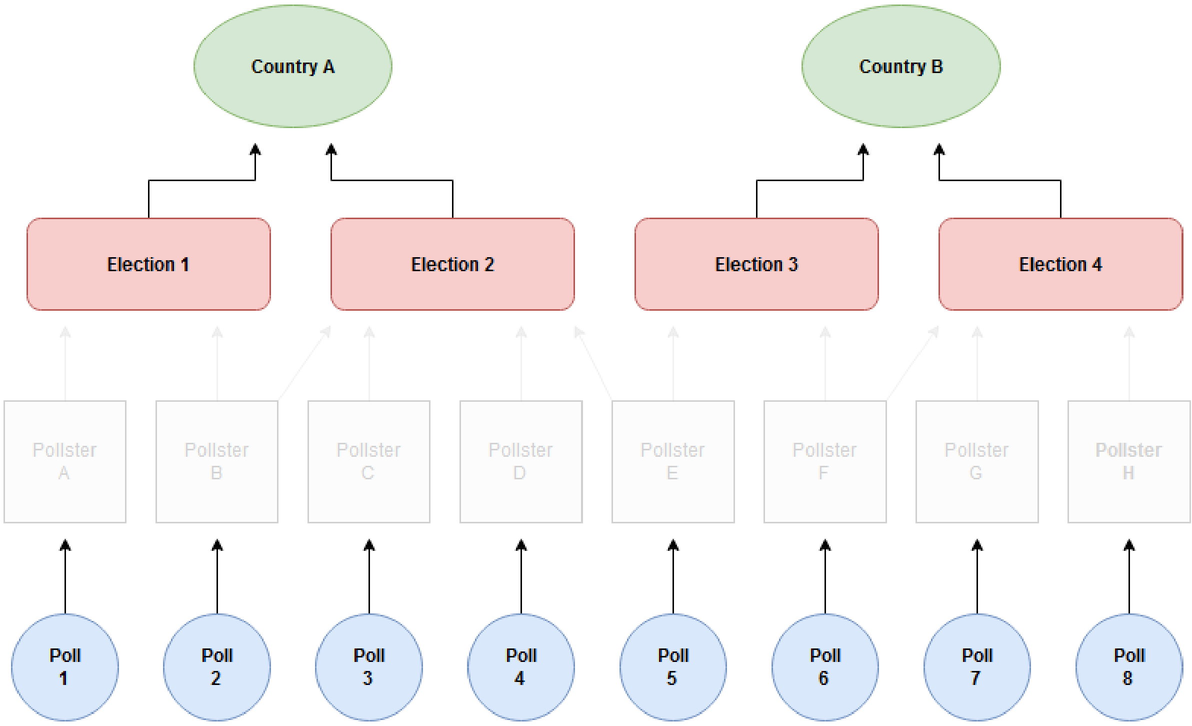

Figure 9: The three-level understanding of sources of polling error employed by Tudor and Wall with polls nested within elections which are themselves nested within countries.[642] Though more comprehensive than two-level models, the three-level model nevertheless excludes the pollster grouping level (greyed out within the figure) and its associated relationship with both the poll and election levels.

As is clear from Figure 9, a three-level understanding of sources of polling error presents a more complete picture than two-levels models, excluding only the differences between polling organisations captured by the pollster grouping level. The principal benefit of this model over its two-level counterpart is that it allows the effect of the country level on polling error to be controlled for in analysis. This ensures that any observed deviation in error between elections is not simply an artefact of those elections having been conducted in different countries. Therefore, it leads to more reliable measurements of the effect of election-level differences on polling error. It has the additional benefit of remaining a fully nested model with no instances

[642] Ibid.

of cross-classification allowing for the same, straightforward calculative approach as its two-level nested counterpart.

Despite its analytical advantages, adding a third level centring on country-level differences to the data structure still fails to fully capture the multi-level nature of polling within the real world. While polls are conducted in relation to individual elections, they are done so by individual polling organisations. From Figure 9, it is clear that the three-level understanding excludes the polling organisations in which polls rest and is therefore incapable of accounting for the impact on polling error that arises from them.

By excluding the pollster level, three-level models run the risk of misattributing a portion polling error to differences between elections – that is, differences between the characteristics that elections comprise – which may in fact exist as an artefact of the differing polling organisations that are housed within the elections themselves. In this way, the absence of a fourth level capturing these differences undermines the degree to which election-level findings from three-level models are representative of reality.

Differences between the polling organisations in which polls are nested can be expected to affect their propensity for misprediction for a variety of reasons. Different polling organisations employ different methods, such as varying sampling procedures,[643] differing survey modes,[644] and alternate weighting schemes.[645] As outlined in the literature review, these differences affect the quality of the predictions that they make and, in so doing, introduce the likelihood of differences in polling error arising between polling organisations. As such, failing to include

[643] D. Stephen Voss, Andrew Gelman, and Gary King, 'A Review: Pre-election Survey Methodology: Details from Eight Polling Organisations, 1988 and 1992', *The Public Opinion Quarterly*, 59.1 (1995), 98 – 132 (p. 101).
[644] Durand and Johnson, p. 183.
[645] British Polling Council, *Poll Methodology, Weighting, and Adjustment Systems* (2013), https://www.britishpollingcouncil.org/wordpress/wp-content/uploads/2013/11/andrew-cooper.pdf> [accessed 7 February 2022].

pollster grouping level from multi-level structures of polling error is to exclude a potentially impactful source of polling error which may, in turn, impact upon the importance of other grouping levels in analysis.

Multi-level understandings of sources of polling error incorporating pollster-level differences have been employed outside of academia in the models employed by organisations that forecast national elections. The organisation that most clearly employs a multi-level approach to polling error that incorporates pollster-level differences is the polling aggregator *FiveThirtyEight* in their poll-based forecasts of US elections. The models employed by *FiveThirtyEight* account for pollster-level differences, as well as limited election-level characteristics specific to the contest being forecast.[646]

The approach adopted by *FiveThirtyEight* can be conceived of as a three-level model of sources of polling error centring on individual elections. This structure is represented below in Figure 10. Though it contains three clear grouping levels, these levels are less intricately connected than other three-level representations of sources of polling error given its focus on individual elections. This necessitates a directly nested relationship between the pollster and election levels within the model, precluding partial crossing between contests.

[646] FiveThirtyEight, *How FiveThirtyEight's House, Senate and Governor Models Work*; Nate Silver, *How Our Primary Model Works* (2020), <https://fivethirtyeight.com/features/how-fivethirtyeight-2020-primary-model-works/> [accessed 8 February 2022]; Nate Silver, *How FiveThirtyEight's House Model Works* (2018), <https://fivethirtyeight.com/features/2018-house-forecast-methodology/> [accessed 8 February 2022]; Nate Silver, *How The FiveThirtyEight Senate Forecast Model Works* (2014), <https://fivethirtyeight.com/features/how-the-fivethirtyeight-senate-forecast-model-works/> [accessed 8 February 2022].

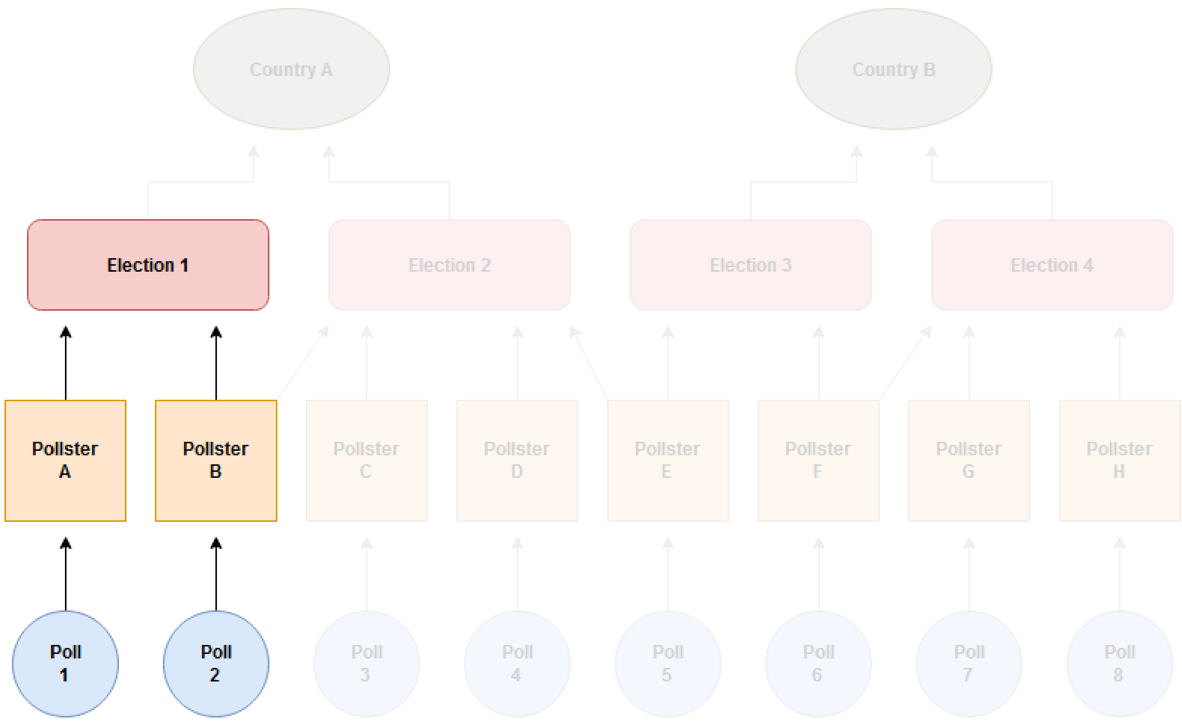

Figure 10: The three-level understanding of sources of polling error employed by the polling aggregator, *FiveThirtyEight*. Polls are taken to be nested within polling organisations which are, in turn, taken to be nested within a given election.

Organisations rendering national elections forecasts focus on individual elections in individual countries. In the case of *FiveThirtyEight*, forecasts only focus on individual elections within the USA. As shown in Figure 10, this means that their approach to sources of polling error does not include country-level differences, as the country in which the polls they address are nested does not change. While the model includes election-level factors, such as current approval ratings and the presence of scandals,[647] it does not include differences *between* election-level groupings. Given its focus on individual elections, it is unable to do so.

The inability to directly capture differences between elections has ramifications for assessments of the effect of pollster-level differences. The three-level model employed by *FiveThirtyEight* assesses the performance of pollster over time, positively weighting those

[647] Nate Silver, *How FiveThirtyEight's House, Senate, and Governor Models Work* (2014), <https://fivethirtyeight.com/features/how-the-fivethirtyeight-senate-forecast-model-works/> [accessed 16/09/2022].

pollsters with records of good performance and negatively weighting those that exhibit poor performance.[648] This process is based on the assumption of path dependency, as it holds that the likely performance of polling organisations in a given election is directly informed by their performance in past contests. Not only does this path dependency not necessarily hold in forecasting,[649] but it fails to account for the heterogeneity of elections and the likelihood of different constellations of factors bearing on the performance of polls between contests, affecting the error they present.

While it remains an artefact of their focus on singular elections, the nesting of polling organisations within individual elections within election-specific three-level models also presents an issue, as it is unrepresentative of their reality. Indeed, polling organisations possess a complex relationship with both the election and country grouping levels within the multi-level structure of polling error. Larger polling organisations, such as Gallup or YouGov, are often international in scope and have extensive histories within the industry. Therefore, not only do they relate to multiple countries, but they also relate to multiple elections due to their longevity. Given this, they are crossed between them.[650] However, certain, smaller polling organisations only exist within individual countries and only pertain to individual elections. This is evident within my data in relation to polling organisations such as iPoll in Ghana or Kult in Albania. In light of this, polling organisations are better understood as being partially crossed between elections,[651] with some relating to individual contests and others relating to multiple contests.

---

[648] Ibid.
[649] Spyros Makridakis, Robin M. Hogarth, and Anil Gaba, 'Why Forecasts Fail. What to Do Instead', *MIT Sloan Management Review*, 51.2 (2010), 83 – 90 (p. 84).
[650] Krzywinski, Altman, and Blainey, p. 977
[651] Ibid.

The distinction between whether polling organisations are nested within or partially crossed between elections is import for their inclusion within multi-level modelling in later analysis. Though this distinction is an element of the data, rather than a modelling decision, the data can nevertheless be modelled incorrectly. Indeed, for the same data, fully nested and cross-classified models produce different parameter estimates.[652] The errant use of fully nested models in lieu of crossed models also leads to an increased risk in type 1 error and wrongly rejecting a correct null hypothesis.[653] Therefore, if a cross-classified model is misrepresented as a nested model within analysis, or vice versa, the researcher risks producing unreliable results.

Though each of the two- and three-level models addressed present advantages and are undoubtedly useful tools for the interrogation of polling error, they each possess drawbacks which make them suboptimal for its multi-level decomposition. In order to achieve the most reliable and justifiable measure of the effect of election-level differences on polling error, it is necessary not only to combine the attributes of the lower-level models addressed in this section, but also to correctly identify the partially crossed relationship between polling organisations and elections. I achieve this through the four-level nested and partially crossed model introduced at the beginning of this section.

Adopting a sequential approach to model complexity – layering levels into analysis incrementally from two to four – allows for a stepwise approach to variance decomposition and the analysis of the importance of the grouping levels within my multi-level structure of polling error both individually and in tandem. I adopt this approach in later analysis, to assess the variance accounted for by each level of my model. However, the election-level ICC values

[652] Melamed and Vuolo, p. 230; Holger Schielzeth and Shinichi Nakagawa, 'Nested by Design: Model Fitting and Interpretation in a Mixed Model Era', *Methods in Ecology and Evolution*, 4.1 (2012), 14 – 24 (p. 14).
[653] Ibid., p. 235.

produced by the final nested and partially crossed four-level model will stand as the most reliable, as all grouping levels of interest are being controlled for.

To represent the hierarchical structure of polling error, my data is necessarily multi-level in nature, comprising variables from each of its four levels. The multi-level nature of my data plays an important role in dictating the appropriate approach to later analysis. In the following sub-section, I address the manner in which it does this and identify appropriate approaches to analysing multi-level data.

From Multi-level Polling Error to Multi-level Modelling

The multi-level nature of polling error necessitates a multi-level approach to its analysis that goes beyond standard multiple regression. In standard multi-variate regression, individual observations are assumed to be independent,[654] rendering it inappropriate for group-level analyses. In data sets with pronounced hierarchical clustering, decomposition on the basis of linear regression leads to heightened type 1 error and therefore an increased likelihood of rejecting a true null hypothesis.[655] It also leads to artificially negative bias in standard error estimation, leading to unrepresentatively narrow confidence intervals around point estimates.[656] Therefore, standard, multi-variate regression techniques are insufficient and inappropriate for multi-level variance decomposition.

To successfully incorporate the clustered, hierarchical nature of polling error in analysis, a multi-level approach to analysis must be adopted.[657] Analysis of multi-level data is typically achieved through multi-level modelling.[658] As I am interested in the amount of polling error

[654] Stephen Gorard, 'What is Multi-level Modelling For?', *British Journal of Educational Studies*, 51.1 (2003), 46 – 63 (p. 49).
[655] P Clarke, 'When Can Group Level Clustering Be Ignored? Multilevel Models Versus Single-level Models with Sparse Data', *Journal of Epidemial Community Health*, 62 (2008), 752 – 758 (p. 752).
[656] Ibid.
[657] Gorard, p. 49.
[658] Ibid.

variation attributable to membership within different elections, I employ null multi-level models, often referred to as unconditional means or variance components models.[659] While the impact of membership within different elections on polling error could be assessed straightforwardly using a two-level null model, to do so would be to ignore the effect of the additional grouping levels identified within my multi-level data structure. To accommodate these additional levels, I employ a four-level null model as outlined in equation 2 where Y_{ijkl} is the observed error for poll *i* conducted by pollster *j* for election *k* in country *l*, β_0 is the mean error across all grouping levels, υ_l is the effect of country *l*, τ_{kl} is the effect of elections nested in countries, λ_{jkl} is the effect of polling organisations partially crossed between elections which are nested within countries, and ϵ_{ijkl} is the poll-level residual error term.[660]

$$Y_{ijkl} = \beta_0 + \upsilon_l + \tau_{kl} + \lambda_{jkl} + \epsilon_{ijkl} \qquad (2)$$

In null multi-level models, the effect of a given grouping factor on an outcome refers to the difference in its mean value relative to others.[661] In the case of equation 2, the effect of membership within a given country on polling error, υ_l, represents the difference between the mean error associated with country *l* and the overall mean error across all grouping levels. Countries with high values of υ_l tend, on average, to produce polls with higher error, while though with lower values tend to produce less erroneous polls. The effect of membership within elections nested within countries, τ_{kl}, is the difference between the mean error associated with election *k* and that of country *l*. Similarly, elections with higher values of τ_{kl} tend to produce polls with higher error, while those with lower scores are on average more likely to produce polls that present lower error. The effect of membership within polling organisations partially

[659] Donald E. Stokes, ‘A Variance Components Model of Political Effects’, *Mathematical Applications in Political Science*, 1.1 (1965), 61 – 85 (p. 61); Stephen Raudenbush and Anthony Bryk, *Hierarchical Linear Models: Applications and Data Analysis Methods*, (London: Sage, 2002), p. 228.
[660] Raudenbush and Bryk, p. 229.
[661] Ibid.

crossed between elections, λ_{jkl}, is the difference between the mean error associated with pollster j and that associated with election k. As such, organisations with higher values of λ_{jkl} tend to give rise to polls that present higher error, while those with lower values produce polls that exhibit lower error on average. Finally, the observed value of Y_{ijkl} for a given poll, i, represents the difference between the error exhibited by an individual poll and the mean score associated with the polling organisation in which it is nested. This difference therefore represents the effect of factors at the poll-level on polling error.

Residuals at all levels of null multi-level models are taken to be normally distributed with a mean of zero and individual variance terms.[662] Variance is partitioned on the basis of the departure of group-level means from the overall mean.[663] In the case of my example, variance is partitioned into four components, as four group-level mean terms exist within equation 2. The decomposition of these variance components allows for the overall effect of membership within given grouping levels to be established. In the following section, I outline the various approaches used to decompose multi-level variance across models concerned with both continuous and binary outcome variables.

4.2: Approaches to Decomposing Multi-level Variance Components

Multi-level Variance Decomposition for Continuous Outcome Variables

To decompose the variance components associated with multi-level models, the key measure is the intra-class correlation coefficient (ICC). The ICC measures the proportion of total variance in the data that is accounted for by a given grouping level.[664] As such, calculating the

[662] Bristol University LEMMA, *Comparing Groups Using Multilevel Modelling* (2022), <https://www.cmm.bris.ac.uk/lemma/mod/lesson/view.php?id=276&pageid=336&startlastseen=no> [accessed 12/08/2022].

[663] Ibid.

[664] Andrew Gelman and Jennifer Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*, (Cambridge: Cambridge University Press), p. 448.

ICC stands as the best method for establishing the degree of error variance account for by the election level and, therefore, to test the expectation that membership within different elections will affect the degree to which polls exhibit error, as established in hypothesis 1.

For two-level models concerned with continuous outcome variables, the ICC is calculated as shown in equation 3, where $\sigma^2_{\text{between}}$ represents the variance accounted for by differences between your groupings of interest and σ^2_{within} represents the remaining within-group variance.[665]

$$\text{ICC}_{\text{continuous}} = \frac{\sigma^2_{\text{between}}}{\left(\sigma^2_{\text{within}} + \sigma^2_{\text{between}}\right)} \tag{3}$$

While the between- and with-group nomenclature is intuitive in two-level data structure, as they possess only one grouping of interest with the remainder of observations neatly and clearly resting within these groups, it is less useful for structures with >2 levels. Though within-group variance refers to the variance associated with all remaining levels contained within the grouping level of interest,[666] this is neither immediately nor intuitively clear. Indeed, as examples often only relate to two-level models,[667] this nomenclature may be, wrongly, taken to refer only to the level of data immediately nested within the grouping level of interest.

The lack of clarity presented by the two-level nomenclature can be remedied by displaying the ICC calculation for an individual grouping level as a simple proportion of the total variance accounted for by all levels within a multi-level data structure. This is shown in equation 4 with

[665] Carly A. Bobak, Paul J. Barr, and A. James O'Malley, 'Estimation of an Inter-Rater Intra-Class Correlation Coefficient That Overcomes Common Assumption Violations in the Assessment of Health Measurement', *BMC Medical Research Methodology*, 18.1 (2018), 93 – 114 (p. 96).
[666] David Liljequist, 'Intraclass Correlation – A Discussion and Demonstration of Basic Features', *PLoS One*, 14.7 (2019), 1 – 35 (p. 1).
[667] Henry Goldstein, *Multilevel Statistical Models: Fourth Edition*, (New Jersey: John Wiley and Sons), p. 24.

reference to the ICC associated with the third grouping level in a four-level data structure, such as that employed within this thesis.

$$\mathrm{ICC_{continuous}} = \frac{\sigma^2_{\mathrm{level\ 3}}}{(\sigma^2_{\mathrm{level1}} + \sigma^2_{\mathrm{level2}} + \sigma^2_{\mathrm{level3}} + \sigma^2_{\mathrm{level4}})} \quad (4)$$

This can be generalised into the form:

$$\mathrm{ICC_{continuous}} = \frac{\sigma^2_x}{\sum_{\mathrm{i=1}}^{\mathrm{n}} \sigma^2_{\mathrm{i}}} \quad (5)$$

Where σ^2_x represents the variance associated with the grouping level of interest and $\sum_{i=1}^{n} \sigma_i^2$ represents the sum of the variances across all levels within the multi-level model. This generalised approach to calculation can be applied to variance decomposition in *n*-level multi-level models with any number of grouping factors.

While equation 4 and its generalisation are commonly used to decompose variance,[668] it is also possible to estimate ICC values from the variance components of one-way ANOVAs given their fundamental association using the measures ICC1,[669] eta-squared,[670] and omega-squared.[671] As these measures rely on one-way ANOVAs, they can only be used to decompose variance in relation to one grouping factor.[672] This means that they can only be used as robustness checks for the ICC values calculated in relation to my two-level continuous measures of polling inaccuracy, as they cannot accommodate the additional grouping factors

[668] Paul D. Bliese, 'Within-Group Agreement, Non-Independence, and Reliability', in *Multilevel Theory, Research, and Methods in Organizations*, ed. by Katherine J. Klein and Steve W. J. Kozlowski, (San Francisco: Jossey-Bass, 2000), pp. 349 – 381 (p. 355).
[669] Gwowen Shieh, 'A Comparison of Two Indices for the Intraclass Correlation Coefficient', *Behavioural Research Methods*, 44 (2012), 1212 – 1223 (p. 1213).
[670] Shieh, p. 1214; Bliese, p. 356.
[671] 671 Casper Albers and Daniel Lakens, 'When Power Analyses Based on Pilot Data are Biased: Inaccurate Effect Size Estimators and Follow-up Bias, *Journal of Experimental Social Psychology*, 74 (2018), 187 – 195 (p. 190).
[672] Amanda Ross and Victor L. Willson, *Basic and Advanced Statistical Tests*, (Boston: Sense Publishers, 2017), p. 21.

necessary for three- and four-level models. These robustness checks, as well as the way in which the additional approaches to variance partitioning are calculated, are contained within Appendix A.

In addition to the range of measures applicable to decomposing variance in multi-level models, there exists a large array of approaches to estimating the model parameters on which these measures rely. These methods of estimation present various advantages, with some being more widely used and dependable than others. In the next sub-section, I present these prospective modelling strategies, outlining their benefits and drawbacks.

Approaches to Parameter Estimation for Continuous Models

To derive the parameters necessary to calculate my measure of interest, the ICC, I employ a series of estimation techniques. Conventional analyses of variance (ANOVAs) stand as the most common forms of variance decomposition in relation to continuous outcome variables.[673] As they categorise data into groupings of interest, they allow for the most straightforward acquisition of the between- and within-group variance measurements required by ICC calculations.[674] Importantly, ANOVAs directly provide the model parameters necessary for the calculation of ICC1, η^2, and ω^2. Despite this, the generalised calculation of the ICC is applicable to a range of estimation strategies which I now address.

In addition to ANOVAs, another common approach to estimating the model parameters necessary for ICC calculation is the use of linear multi-level models.[675] Linear multi-level models are hierarchical models concerned with a continuous outcome variable.[676] One of the

[673] Matthew E. Wolak, Daphne J. Fairbairn, and Yale R. Paulsen, 'Guidelines for Estimating Repeatability', *Methods in Ecology and Ecology and Evolution*, 3.1 (2012), 129 – 137 (p. 132); Bliese, p. 355.
[674] Zhaoxia Yu and others, 'Beyond T Test and ANOVA: Applications of Mixed-Effects Models for More Rigorous Statistical Analysis in Neuroscience Research', *Neuron*, 110.1 (2022), 21 – 35 (p. 25).
[675] Ibid., p. 28.
[676] Ibid., p. 32.

most widely used methods for estimating parameters in models containing continuous outcome variables is maximum likelihood estimation (MLE).[677] MLE seeks to find the set of model parameters which make the observed data most probable.[678] To achieve this, it iteratively fits probability distributions to the observed data and determines the conditional probability of observing the given data.[679] The conditional probability of observing data given a certain set of parameters is known as the likelihood function,[680] and it is this which MLE seeks to maximise.

Though approaches involving MLE perform well when used on large datasets, especially those with a large group *n*, their estimates of model parameters are negatively biased when applied to datasets with a small group *n*.[681] As such, they can be expected to perform well in macroscopic analyses of my dataset, and therefore stands as my preferred estimative approach when assessing the dataset as a whole. Despite this, MLE may perform less well in country-wise error decomposition due to the reduction in sample size.

Alongside MLE, Restricted maximum likelihood estimation (RMLE) exists as the other principal form of likelihood-based parameter estimation for linear multi-level models. RMLE operates in much the same way as MLE, but estimates likelihood functions using fewer parameters.[682] In reducing the number of parameters addressed, RMLE reduces influence of potential nuisance parameters, that is parameters that are not of immediate importance to estimating the likelihood function.[683] This allows it to outperform MLE in small *n* datasets, as

[677] Joop J. Hox, Mirjam Moerbeek, and Rens van de Schoot, *Multilevel Analysis: Techniques and Applications*, (New York: Routledge, 2017), p. 297.
[678] Richard J. Rossi, *Mathematical Statistics: An Introduction to Likelihood Based Inference*, (Hoboken: John Wiley and Sons, 2018), p. 226.
[679] Ibid.
[680] Ibid.
[681] Yahia El-Horbaty and Eman Hanafy, ‘Some Estimation Methods and Their Assessment in Multilevel Models: A Review’, *Biostatistics and Biometrics Open Access Journal*, 5 (2018), 1 – 8 (p. 3).
[682] Ibid.
[683] Ibid.

it produces less biased results.[684] As such, it presents a more viable approach to calculating the ICC on a country-wise basis, given the relatively small group *n* presented by each case. However, on average, the parameter estimates provided by RMLE present higher standard errors, affecting their quality.[685]

Some of the drawbacks presented by frequentist estimation strategies are corrected for within their Bayesian counterparts. The most popular Bayesian method of parameter estimation for multi-level models is Markov chain Monte Carlo (MCMC) estimation. Bayesian MCMC corrects for the frequentist estimative issues surrounding small group *n*, leading to improve parameter estimation.[686] Given this, Bayesian MCMC modelling stands as my preferred form of estimation for country-wise ICC calculations which are presented in Appendix A5.

Bayesian estimation rests on setting priors before engaging in analysis.[687] This involves establishing prior expectations for the distributions of the parameters within the model. In multi-level models, this process focuses on hyperparameters.[688] While model parameters are quantities of interest that can be estimated from data, such as beta coefficients in traditional regression,[689] hyperparameters are values external to a model that cannot be estimated from data and are often used to aid in the estimation of model parameters.[690] The optimal hyperparameter values for a given model are often not known a priori.[691] As such, they are

[684] Hox, Moerbeek, and van de Schoot, p. 298.
[685] Goldstein, p. 24.
[686] Bradley Carlin and Thomas Louis, 'Identifying Prior Distributions that Produce Specific Decisions with Application to Monitoring Clinical Trials', in *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*, ed. by John Donald, Kathryn Chaloner, and Arnold Zellner, (New York: John Wiley and Sons, 1996), p. 497; Raudenbush and Bryk, p. 13.
[687] Ibid.
[688] Andrew Gelman, 'Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper)', *Bayesian Analysis*, 1 (2006), 515 – 534 (p. 516).
[689] Ibid.
[690] Max Kuhn and Kjell Johnson, *Applied Predictive Modelling*, (New York: Springer, 2013), pp. 64 – 65.
[691] Ibid.

typically set using heuristics or established best practices. For multi-level models, the principal hyperparameters are the variances of the residual error terms.[692]

The variance of the residual error terms happens to be the focus of analysis within this chapter. To illustrate why this is the case, equations 6 and 7 address the foundation and components of a simple null two-level model which is illustrative of those used in later analysis. A standard null regression model is simply one which does not contain any predictor variables. As such, it only contains a y-intercept and a residual error term, as shown in equation 6.

$$\mathrm{Y}_i = \beta_0 + e_i \tag{6}$$

A null model will simply create a line of best fit with a slope of zero that represents the mean value of Y, as this is the best prediction possible in the absence of predictor variables. In such a model, residuals therefore represent the distance between this mean line and each data point. The variance in this model is then simply the square of these distances from the mean.

A two-level null model is similarly specified as shown in equation 7. Predictor variables are once again absent, such that the y-intercept, β_0, again represents the mean of the Y value of interest. However, here it represents the overall mean of the Y value across all groups. While a standard regression model only has one set of residuals, a two-level model possesses two: the group-level residuals, u_j, and the individual-level residuals, e_{ij}.

$$\mathrm{Y}_{ij} = \beta_0 + u_j + e_{ij} \tag{7}$$

The group-level residuals represent the difference between the overall mean of Y and the means of the j level two groupings within the model. The individual-level residuals represent the

[692] Gelman, p. 516.

difference between the Y value associated with the i^{th} individual observation and the mean of the level two grouping in which that observation rests.

As these residuals represent differences between values and their respective means, their squares denote variance. The squared group-level residual terms therefore represent the variance between groups, while the squared individual-level residuals represent the variation within groups. These between- and within-group variances can be substituted into our earlier ICC calculation as shown in equation 8. Therefore, the analysis within this chapter relates directly to the variance of the group- and individual-level residual error terms within its multi-level models.

$$\mathrm{ICC} = \frac{\sigma^2{}_{\mathrm{between}}}{(\sigma^2{}_{\mathrm{within}} + \sigma^2{}_{\mathrm{between}})} = \frac{\sigma^2{}_{\mathrm{level2}}}{(\sigma^2{}_{\mathrm{level1}} + \sigma^2{}_{\mathrm{level2}})} \tag{8}$$

As they are the hyperparameters of interest, the variance of the residual error terms must be given their own prior distributions.[693] As variance cannot take on negative values,[694] a sensible prior distribution would be one that terminates at zero and therefore precludes negative values. To this end, half-Cauchy priors are recommended for the hierarchical variance hyperparameters in multi-level models.[695] The half-Cauchy prior distribution terminates at zero, and therefore precludes negative values, as it represents the right half of a Cauchy distribution symmetric about a mean of zero. Half student-t priors are also commonly used in multi-level modelling.[696] Within the brms R package, which I use to run my Bayesian multi-level models, the default half student-t priors are restricted to be non-negative, making them

[693] Gelman, p. 516.
[694] Ibid.
[695] Gelman, p. 520; Nicholas G. Polson and James G. Scott, 'On the Half-Cauchy Prior for a Global Scale Parameter', *Bayesian Analysis* 7.4 (2012), 887 – 902 (p. 896).
[696] Nathan P. Lemoine, 'Moving Beyond Noninformative Priors: Why and How to Choose Weakly Informative Priors in Bayesian Analysis', *Oikos*, 128 (2019), 912 – 928 (p. 915).

applicable to variance hyperparameters.[697] These possess benefits over their half-Cauchy counterpart, most notably the facilitation of more reliable model convergence.[698]

Given its improved performance and widespread use, a half student-t prior distribution that terminates at zero stands as my preferred prior. However, to demonstrate the robustness of the findings presented by my Bayesian models, and to illustrate that they are not simply artefacts of a given prior distribution, I also run them using half-Cauchy priors. These robustness checks are contained within Appendix A.

The approaches to ICC calculation and model parameter estimation used for my continuous variables are not directly to my binary measures of polling error. In the following sub-section, I outline the ways in which they must be adapted or replaced to facilitate the decomposition of the variance presented by my binary variables.

Multi-level Variance Decomposition for Binary Outcome Variables

To accommodate my binary measures of polling error, approaches to ICC calculation that are able to accommodate their differing variance structures are required. The earlier equations for ICC calculation are not applicable to models concerning binary outcome variables. This is due to their underlying logistic distribution.[699] Given this distribution, extracting the variance components required by the ICC calculation is a more involved process. While this can be

---

[697] Paul-Christian Bürkner, *Package 'brms'* (2021), <https://cran.r-project.org/web/packages/brms/brms.pdf> [accessed 10 February 2022].

[698] Paul-Christian Bürkner, 'brms: An R Package for Bayesian Multilevel Models Using Stan', *Journal of Statistical Software*, 80 (2017), 1 – 28 (p. 11).

[699] Shinichi Nakagawa, Paul C. D. Johnson, and Holger Schielzeth, 'The Coefficient of Determination R^2 and Intra-class Correlation Coefficient from Generalized Linear Mixed-Effects Models Revisited and Expanded', *Journal of the Royal Society Interface*, 14 (2017), 134 (p. 134).

achieved through a variety of methods, such as model linearisation and simulation,[700] the approach most applicable to my binary variables is the latent variable approach.[701]

The latent variable approach is applied when a binary measure represents a discretisation of an underlying continuous threshold variable.[702] Both of my binary measures of polling error – correctly predicting the party or candidate with the largest vote share and the presence of significant bias – can be conceived of as discretisations based on an underlying continuous variable crossing a given threshold. In the case of the presence of significant bias, a coding of 1 is given if the continuous value associated with leading party bias rests outside of the threshold set by a 95% confidence interval, while a coding of 0 is given otherwise. As such, it clearly represents a discretisation of an underlying continuous threshold variable.

While the presence of significant bias presents an absolute threshold over and under which its binary coding is determined, the classification of whether a poll correctly predicts the party or candidate in receipt of the largest share of the vote is based on relative thresholds. That is, the threshold over which the vote share gained by a party or candidate can be said to be the largest is relative to the accomplishments of the other parties or candidates contesting an election. As such, it varies between contests. Nevertheless, if the predicted vote share of a party or candidate exceeds the threshold set by the next most successful competitor, then they are predicted by a poll to possess the largest vote share. If the share of the vote received by this same party or candidate on election day again exceeds the threshold set by the next most successful competitor, then they have succeeded in acquiring the largest share of the vote, as predicted, meriting a binary classification of 1. If these conditions are not met, then a classification of 0 is given. Therefore, the binary determination of whether a poll correctly predicts the party of

[700] Harvey Goldstein, William Browne, and Jon Rasbash, 'Partitioning Variation in Multilevel Models', *Understanding Statistics*, 1.4 (2002), 223 – 231 (pp. 226 – 227).
[701] Nakagawa, Johnson, and Schielzeth, p. 134.
[702] Ibid.

candidate in receipt of the largest vote share also represents a discretisation of an underlying continuous threshold variable, albeit relative.

Under the latent variable approach to ICC calculation, level one variance is fixed at $\frac{\pi^2}{3}$.[703] By way of an example, the ICC calculation for the third level of a model concerned with a binary outcome variable based on a continuous threshold variable would be calculated as shown in equation 9.

$$\mathrm{ICC}_{\mathrm{binary}} = \frac{\sigma^2_{\mathrm{level3}}}{\left[\left(\frac{\pi^2}{3}\right) + \sigma^2_{\mathrm{level2}} + \sigma^2_{\mathrm{level3}} + \sigma^2_{\mathrm{level4}}\right]} \tag{9}$$

This can be generalised into the form:

$$\mathrm{ICC}_{\mathrm{binary}} = \frac{\sigma^2_{x}}{\left[\left(\frac{\pi^2}{3}\right) + \sum_{\mathrm{i}=2}^{\mathrm{n}} \sigma^2_{\mathrm{i}}\right]} \tag{10}$$

Where σ_x^2 represents the variance associated with the grouping level of interest and $\sum_{\mathrm{i}=2}^{\mathrm{n}} \sigma^2_{\mathrm{i}}$ represents the sum of the variances associated with the remaining levels within the model excluding the first. This calculative approach can be used to decompose variance in n-level multi-level models with any number of grouping factors concerning binary outcome variables.

In much the same way as the calculation of ICC values differs for models concerning binary outcome variables, so too do approaches to the estimation of the variance components that they require. In the following sub-section, I address approaches taken to model the variance of my binary measures of polling inaccuracy.

[703] Ibid.

Approaches to Parameter Estimation for Binary Models

Methods for estimating model parameters in multi-level models concerned with binary outcome variables come in two principal forms: frequentist likelihood-based approaches and Bayesian Markov Chain Monte Carlo (MCMC) approaches.[704] Likelihood-based approaches estimate model parameters on the basis of log likelihoods.[705] While maximum likelihood and restricted maximum likelihood estimation can be used to find the maximally probable likelihood function for parameter estimation in linear multi-level models, no standard solution or expression exists for calculating the equivalent maximal log likelihood function for multi-level models concerning non-normal outcome variables.[706] However, alternative approximation strategies exist that are based on quadrature methods.

Quadrature methods are approaches to the approximation of the definite integral of a function – in this case the log likelihood function – using numerical integration.[707] This approximates the unknown definite integral of a function using the weighted sum of a number of equally spaced points sampled from the function itself.[708] These points are known as quadrature points and serve to distinguish the two mostly commonly used quadrature methods in multi-level modelling: Laplace approximation and adaptive Gauss-Hermite quadrature (AGHQ).[709]

Both Laplace approximation and AGHQ use numerical integration to approximate the optimal shape of the conditional probability distribution, or likelihood function, for a given set of

[704] Sophia Rabe-Hesketh, Anders Skrondal, and Andrew Pickles, ‘Reliable Estimation of Generalized Linear Mixed Models Using Adaptive Quadrature’, *The Stata Journal*, 2 (2002), 1 – 21 (p. 2); Daniel McNeish, ‘Estimation Methods for Mixed Logistics Models with Few Clusters’, *Multivariate Behavioral Research*, 51 (2016), 790 – 804 (p. 793).
[705] Rabe-Hesketh, Skrondal, and Pickles, p. 3; McNeish, p. 794.
[706] Gary King and Langche Zeng, ‘Logistic Regression in Rare Events Data’, *Political Analysis*. 9.2 (2001), 137 – 163 (p. 141).
[707] Rabe-Hesketh, Skrondal, and Pickles, p. 4.
[708] Ibid.
[709] Ibid.

observed data.[710] Through this process, they also approximate the parameters that best represent this data.[711] While Laplace approximation uses only one quadrature point in its approximation of likelihood functions, AGHQ uses a greater number, providing it with the ability to produce more accurate approximations.[712]

Despite the potential improvements to accuracy offered by AGHQ, it possesses a series of significant shortcomings. The additional accuracy offered by its larger number of quadrature points is only apparent in multi-level analyses involving small group *n*, rendering it negligible in larger studies.[713] To compound this, AGHQ approaches fail to converge when applied to multi-level models with >2 levels. This is because the likelihood function of two-level models can be represented as the product of one-dimensional integrals, while models with >2 levels cannot, precluding the use of AGHQ.[714] Given these issues, the improved accuracy offered by AGHQ will not be apparent when applied to my dataset given its large group *n* and it cannot be used to approximate likelihood functions for my three- and four-level models, rendering it less useful than Laplace approximation.

The definite integral of the log likelihood function can also be approximated using Bayesian Markov Chain Monte Carlo methods (MCMC).[715] MCMC approaches operate in much the same way as the quadrature methods discussed earlier, but instead of sampling evenly spaced values from the integrand, they approximate integrals using a series of random values.[716] While MCMC numerical integration performs more effectively when approximating high-

[710] Ibid.
[711] Ibid.
[712] Silvia Bianconcini, Silvia Cagnone, and Dimitris Rizopoulos, 'Approximate Likelihood Inference in Generalized Linear Latent Variable Models Based on the Dimension-wise Quadrature', *Electronic Journal of Statistics*, 11 (2017), 4404 – 4423 (p. 4405).
[713] Ibid.
[714] Helen Ogden, *Fitting GLMMs with glmmsr* (2018), <https://cran.r-project.org/web/packages/glmmsr/vignettes/glmmsr-vignette.pdf> [accessed 17 February 2022].
[715] William H. Press and others, *Numerical Recipes: The Art of Scientific Computing*, (Cambridge: Cambridge University Press, 2007), p. 398
[716] Ibid.

dimensional integrals containing a large number of variables,[717] it presents issues concerning sampling error and is often considered a method of last resort.[718]

As I will be applying numerical integration procedures to null multi-level models which, by definition, have a minimal number of dimensions, the benefits of Bayesian MCMC integration will not be realised in relation to my data. When this is combined with the problems presented by AGHQ in relation to three- and four-level models, Laplace approximation stands as my preferred method of parameter estimation for multi-level models concerning a binary outcome variable. I do, however, use AGHQ and Bayesian MCMC integration as robustness checks where appropriate. Such checks may prove useful as, in spite of its applicability across increasingly highly dimensional multi-level data structures, the singular quadrature point approach taken by Laplace approximation can make approximating complex integrals challenging given its simplicity, especially in relation to binary outcome variables.[719]

Interpreting ICC Values in Later Analysis

With the methods of parameter estimation established, I move to display the ICC estimates that result from them. However, to allow this to be done productively, I first address the manner in which they are to be interpreted. ICC values are best understood as normalised percentages. A value of 0 indicates that zero percent of the variance associated with a given outcome variable results from its membership within a certain grouping level. Conversely, a value of 1 indicates that one hundred percent of the variance associated with an outcome variable is the result of its membership within a given grouping level.[720] As such, the higher the ICC value, the greater

[717] Ibid.
[718] Mario J. Miranda and Paul R. Fackler, *Lecture Notes in Computational Economic Dynamics* (1997), <http://fmwww.bc.edu/ec-p/software/Miranda/chapt6.pdf> [accessed 11 February 2022].
[719] Marinela Capanu, Mithat Gonen, and Colin B. Begg, 'An Assessment of Estimation Methods for Generalized Linear Mixed Models with Binary Outcomes', *Stat Med*, 32.26 (2013), 1 – 24 (p. 1).
[720] Nicolas Sommet and Davide Morselli, 'Keep Calm and Learn Multilevel Logistic Modeling: A Simplified Three-step Procedure Using Stata, R, Mplus, and SPSS', *International Review of Social Psychology*, 30 (2017), 203 – 218 (p. 204).

the amount of variance explained by the grouping level. The percentage value over which multi-level modelling is deemed justifiable is five percent, represented by an ICC value of 0.05.[721]

To account for the uncertainty surrounding variance point estimates, I calculate standard errors along with 95% confidence and credibility intervals using a range of methods. For ICC1 estimates, I use Searle's exact confidence limit equation due to its applicability to unbalanced data.[722] For my null multi-level models, I use the logit transformation and delta method,[723] while values are directly calculated from model outputs for Bayesian, η^2, and ω^2 estimates.

I decompose the variance components associated with null multi-level models across the three conceptualisations of polling error outlined in the previous chapter: distributive, bounded, and substantive error. However, before this can be achieved, each of these conceptualisations must be operationalised for use within analysis. In the following section, I provide the manner in which each I measure each of my outlined conceptualisations of error.

4.3: Operationalising Polling Error

To employ the distributive, bounded, and substantive conceptualisations of polling error outlined in the previous chapter analytically, they must be operationalised. In the following section, I outline the approaches I take to measuring each of these conceptualisations of polling error. Overall, I present eight approaches to measuring polling error inclusive of measures of bias. The five approaches taken to measuring distributive inaccuracy are derived from the literature, while the remaining three approaches to measuring bounded and substantive

[721] Tenko Raykov, 'Intraclass Correlation Coefficients in Hierarchical Designs: Evaluation Using Latent Variable Modelling', *Structural Equation Modelling*, 18 (2011), 73 – 90 (p. 81).
[722] J. D. Thomas and R. A. Hulquist, 'Interval Estimation for the Unbalanced Case of the One-way Random Effects Model', *Annals of Statistics*, 6 (1978), 582 – 587 (p. 584).
[723] Nakagawa, Johnson, and Schielzeth, p. 134.

inaccuracy stand as novel operationalisations of my own creation. The formulae used for measuring each of my outcome variables is presented in bold, with those calculations leading up to or informing their measurement remaining in normal typeface. I begin by outlining my approaches to measuring distributive inaccuracy, as it lends itself most straightforwardly to operationalisation.

Measuring Distributive Inaccuracy

Distributive inaccuracy lends itself most straightforwardly to operationalisation due to the existence of several commonly used approaches to its measurement. Of the measures devised by Mosteller and the Social Science Research Council in 1949,[724] Measures 3 and 5 have come to see near ubiquitous use within the discipline since their endorsement by Mitofsky in 1998.[725] These can most straightforwardly be understood as the average difference between predicted and actual party vote shares, and the difference between predicted and actual margins of victory, respectively.

The average difference between predicted and actual party votes shares is represented by mean absolute error (MAE). MAE concerns the average, absolute percentage point difference between predicted and actual vote shares and is calculated as shown in equation 11, where Poll_i represents the predicted vote share of a party, Vote_i represents the actual percentage vote share received by that party on election day, and n represents the total number of parties contesting an election.[726]

[724] Mosteller and others, pp. 54 – 55.
[725] Warren J. Mitofsky, 'Was 1996 a Worse Year for Polls than 1948?', *Public Opinion Quarterly*, 62 (1998), 230 – 249 (pp. 230 – 249); Traugott, 'The Accuracy of the National Pre-election Polls', pp. 645 – 648; Costas Panagopoulos, Kyle Endres, and Aaron C. Weinschenk, 'Pre-election Poll Accuracy and Bias in the 2016 U.S. General Elections', *Journal of Elections, Public Opinion and Parties*, 28.2 (2018), 157 – 172 (p. 160); Callegaro and Gasperoni, p. 158; Crewe, 'The Opinion Polls: The Election They Got (Almost) Right', pp. 684 – 698; McElroy and Marsh, pp. 159 – 176.
[726] Mosteller and others, pp. 54 – 55.

$$\mathbf{MAE} = \frac{\sum_{\mathbf{i=1}}^{\mathbf{n}} |(\mathbf{Poll_i} - \mathbf{Vote_i})|}{\boldsymbol{n}} \tag{11}$$

The difference between the predicted and actual margin of victory concerns the difference between two differences. Specifically, it measures the difference between the predicted vote share margin between the top two parties or candidates within an election and their actual margin on election day.[727] I refer to this measure as the difference in margin (DIM) and calculate it as shown in equation 12, where A_p and B_p represent the predicted vote shares of the two leading parties or candidates in an election, whilst A_v and B_v represent the actual vote shares that they receive on election day. Taking the absolute values of these differences is common,[728] as it allows the difference in margin to focus on random error through the preclusion of directionality.

$$\mathbf{DIM} = |(\mathbf{A_p} - \mathbf{B_p}) - (\mathbf{A_v} - \mathbf{B_v})| \tag{12}$$

MAE and DIM serve as the first operationalisations of distributive polling (in)accuracy that I adopt within this thesis. However, despite their ubiquitous use, they still possess shortcomings. MAE is incapable of accounting for the directionality of error, and therefore bias, while the difference in the margin of victory is unable to account for any more than the two leading parties or candidates within an election. Though such an approach befits strong two-party competition, in which third-parties are of minimal consequence, it imposes a false dichotomy on meaningfully multi-party elections. This results in the exclusion of other vote share estimations from the measurement of inaccuracy. However, the two leading parties in an election are the most likely to be instrumental in government formation, either individually or

[727] Ibid.

[728] Michael W. Traugott, 'Assessing Poll Performance in the 2000 Campaign', *Public Opinion Quarterly*, 65.3 (2001), 389 – 419 (p. 393); Durand and others, 'Report of the WAPOR Committee', p. 25.

in coalition. As such, correctly predicting the margin between them is arguably more important than accurately predicting the vote shares received by marginal parties. Nevertheless, to assess polling (in)accuracy in totality, the predicted vote shares for all parties need to be considered.

The shortcomings surrounding directionality can partially be remedied by Measure A. Devised by Martin, Traugott, and Kennedy, Measure A takes the natural logarithm of the predicted vote share margin between the two leading parties or candidates divided by their actual margin on election day.[729] Despite using absolute values, it focuses on the degree to which polls over- or under-estimate the vote shares of these parties and is therefore more intuitively understood as a measurement of leading party bias (LPB). It is calculated as shown in equation 13, where p_1 and p_2 represent the predicted vote shares of the two leading parties or candidates, and v_1 and v_2 represent the shares of the vote that they receive on election day. Absolute values are used within the equation, as taking the natural logarithm of negative values yields imaginary numbers.

$$\mathbf{LPB} = \mathbf{ln}\left[\left(\frac{\mathbf{p_1}}{\mathbf{p_2}}\right) - \left(\frac{\mathbf{v_1}}{\mathbf{v_2}}\right)\right] \tag{13}$$

Given its focus on the two leading parties or candidates in an election, the measurement of leading party bias suffers from the same issue of omitting smaller parties or candidates. Despite this, it is capable of capturing bias present within vote share estimates. A value of zero represents perfect agreement between a poll and the election result.[730] The greater the distance between the value of LPB and zero, the less accurate a poll is considered to be. If the value is negative, a poll has overestimated the first candidate or party relative to the final election results. Conversely, if the value is positive, it has overestimated the second party or

[729] Martin, Traugott, and Kennedy, p. 352.
[730] Ibid., p. 351.

candidate.[731] Though originally devised for use in US elections which are dominated by two-party competition, rendering third party candidacies largely insignificant (though notable historical exceptions exist),[732] leading party bias can be measured in elections dominated by two coalitions,[733] as well as those containing multi-party contests.[734] Due to the range of election types encompassed by my dataset, I calculate leading party bias in relation to the two leading parties in each contest in a manner similar to Wright, Farrar, and Russell.[735]

While the raw value of LPB represents the absolute degree to which polls under- or over-estimated the vote shares of parties or candidates, some divergence from a value of zero is expected due to sampling error.[736] Under this understanding, only instances of significant over- or under-estimation are taken to be representative of true bias. To establish a threshold over which deviance is considered bias, I calculate a 95% confidence interval around zero.[737]

The process for calculating the confidence interval using the variance equation for leading party bias is outlined in equations 14, 15, and 16. Equation 14 represents the method used to calculate the variance of leading party bias, where n represents the total number of respondents in a poll, p_1 represents the proportion of respondents supporting the first party or candidate within a poll, and p_2 represents the proportion supporting the second.[738] Equation 16 represents the calculation of the standard error of the estimates of leading party bias (LPB). The standard error of the mean is given by dividing the standard deviation of the sample by the square root of the number of observations. As polls are assessed individually, the number of observations

[731] Ibid.
[732] Steven J. Rosenstone, Roy L. Behr, and Edward H. Lazarus, *Third Parties in America: Citizen Response to Major Party Failure*, (Princeton: Princeton University Press, 1984), p. 4.
[733] Callegaro and Gasperoni, p. 154.
[734] Wright, Farrar, and Russell, p. 116.
[735] Ibid.
[736] Ibid.
[737] Ibid.
[738] Ibid.

takes a value of one, rendering the standard error equal to the standard deviation, which is simply the square root of the variance given by equation 14. Therefore, the square root of the variance is used in lieu of the standard error to calculate the 95% confidence interval in equation 16.

$$\mathrm{Var(LPB)} = 1/(\mathrm{n} * \mathrm{p}_1 * \mathrm{p}_2) \tag{14}$$

$$\mathrm{SE(LPB)} \ = \ \sigma/\sqrt{\mathrm{n}} \ = \ \sqrt{\mathrm{Var(LPB)}} \tag{15}$$

$$95\%\ \mathrm{CI} = 0\ \pm 1.96\left(\sqrt{\mathrm{Var(LPB)}}\right) \tag{16}$$

Any deviation from zero that exceeds the boundaries of the 95% confidence interval results in a poll being considered biased, whilst deviation within the interval is accounted for by sampling error and is not seen to represent bias. This results in a binary operationalisation of leading party bias in which the presence of significant bias is dichotomised. I refer to this additional binary measurement as significant bias.

The 95% confidence interval surrounding each value of leading party bias is itself a set of values, bounded by the upper and lower limits of the interval. As such, if the value taken by leading party bias is an element of this set, then its deviation from a value of zero can considered reasonable, and it cannot be said to represent significant bias. However, if the value is not an element of this set, then its deviation from zero exceeds reasonable tolerances, and it can be said to be representative of significant bias. Such a condition would generate a significantly biased poll (SBP). This is formalised in equation 17 where 1 denotes the presence of significant bias and 0 represents the absence of significant bias.

$$\mathbf{SBP} = \begin{cases} \mathbf{0\ if\ LPB} \in \mathbf{95\%\ CI} \\ \mathbf{1\ if\ LPB} \notin \mathbf{95\%\ CI} \end{cases} \tag{17}$$

To disentangle leading party bias from its two-party focus, Arzheimer and Evans devised A_i' as a measure of polling bias.[739] A_i' decomposes polling bias on a per-party basis and is therefore more intuitively understood as party bias. It is calculated using the log odds ratio shown in equation 18, in which p_i represents the normalised percentage of support received by a party in a poll and v_i represents the normalised percentage of the vote they receive on election day.[740]

$$\text{Party bias} = \ln\left[\left(\frac{p_i}{(1-p_i)}\right) \Big/ \left(\frac{v_i}{(1-v_i)}\right)\right] \tag{18}$$

In much the same way as leading party bias, positive values for per party bias indicate that a poll has overestimated the vote share of a party, negative values indicate underestimation, and a value of zero represents perfect agreement between prediction and outcome.[741] Given that party bias provides under- and over-estimation measures on a per-party basis, it provides multiple measurements per poll. Cases in which multiple values of the same error measure are associated with individual polls are not conducive to decomposing error within the multi-level structures outlined earlier in this chapter. This is because the atomic, or first, level of these models – the level on which the nesting structure is built – comprises polls as its individual observational units. Through this, it measures differences between polls conceived of as singular, unitary entities. Fragmenting these entities into a series of party-specific measures not only unnecessarily complicates the nesting structure (as differences between parties are implicitly contained within the variable coverage of polls), but also does not allow for the explicit analysis of between-poll differences. The importance of this will become clearer later in this chapter when I introduce the intra-class correlation coefficient and its analogues.

---

[739] Kai Arzheimer and Jocelyn Evans, 'A New Multinomial Accuracy Measure for Polling Bias', *Political Analysis*, 22 (2014), 31 – 44 (p. 33); Kai Arzheimer and Jocelyn Evans, 'Estimating Polling Accuracy in Multiparty Elections Using Surveybias', *The Stata Journal*, 16.1 (2016), 139 – 158 (p. 141).
[740] Arzheimer and Evans, 'A New Multinomial Accuracy Measure', p. 33.
[741] Arzheimer and Evans, 'Estimating Polling Accuracy', p. 141.

Given the differences surrounding the use of a per-party measure of polling bias, Arzheimer and Evans' Measure B provides a single, aggregate measure of party bias per poll.[742] As shown in equation 19, it takes the average of the per-party bias measurements given by part bias across the number of parties covered by a poll, *n*.[743] As it provides an average measure of bias across parties, it is more intuitively understood as average per-party bias (APB). If all parties contesting an election are covered in a poll, thereby allowing it to account for 100% of the vote share, the per-party bias values cancel out in the aggregate as the extent of any over-estimation would directly mirror that of under-estimation. To prevent this, the absolute value of APB is often taken. However, polls rarely account for every single party contesting an election, especially in highly fragmented multi-party systems. Moreover, capturing directionality – the extent of over- or under-estimation – is key to the measurement of bias. Taking the absolute value removes the ability to do this and therefore prevents directional bias from being represented. Though including signed values in the calculation of APB is recognised to provide an overly optimistic measure of polling bias,[744] it is nevertheless the only way to capture the directionality of bias in the aggregate. As such, I calculate the signed average party bias as show in equation 19.

$$\mathbf{APB} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \boldsymbol{PB}_{\mathrm{i}}}{\boldsymbol{n}} \tag{19}$$

In tandem, the outlined measures of distributive inaccuracy allow for the measurement of both of its key components: random and systematic error. Indeed, they are often used in conjunction

[742] Arzheimer and Evans, 'A New Multinomial Accuracy Measure', p. 36.
[743] Ibid.
[744] Ibid.

for just this purpose.[745] With the operationalisation of distributive inaccuracy established, I move to discuss the measurement strategies for bounded inaccuracy.

Measuring Bounded Inaccuracy

While a range of existing operationalisations exist for distributive polling inaccuracy the same cannot be said for bounded inaccuracy. As such, a novel approach to its measurement is needed. Fundamentally, bounded inaccuracy is based on the margin of error of polls (MOE). This is commonly calculated with 95% confidence as shown in equation 20, where n represents the sample size of a poll and θ is a constant with a value of 0.5.[746]

$$\mathrm{MOE} = \pm 1.96\sqrt{\frac{\theta(1-\theta)}{n}} \tag{20}$$

Equation 20 produces a normalised percentage value which I use to produce a range of possible values surrounding the vote share point estimates provided by polls. These ranges can be conceived of as sets of possible vote shares surrounding the point estimates provided for each party by a poll. Each poll produces its own sets on the basis of its predicted vote share point estimates and sample size. For a typical pre-election poll with a sample of 1000 respondents, the margin of error surrounding a given vote share point estimate is $\pm$3%.[747] The set containing the range of possible values within this margin of error (P) can be represented as shown in equation 21, with vote shares ranging from three percentage points below prediction p to three percentage points above.

[745] Christopher T. Stout and Reuben Kline, 'I'm Not Voting for Her: Polling Discrepancies and Female Candidates', *Political Behaviour*, 33 (2011), 479 – 503 (p. 488); Panagopoulos, 'Pre-election Poll Accuracy in the 2008 General Elections', p. 899.
[746] John Morgan and Philip C. Stocken, 'Information Aggregation in Polls', *American Economic Review*, 98.3 (2008), 864 – 896 (p. 875).
[747] Anthony Wells, *Understanding Margin of Error* (2011), <https://yougov.co.uk/topics/politics/articles-reports/2011/11/21/understanding-margin-error> [accessed 01/09/2022].

$$P = \{p - 3 \ \dots \ p + 3\} \tag{21}$$

As discussed in the previous chapter, bounded inaccuracy can be conceived of in two forms: binary and continuous. The binary conceptualisation of bounded inaccuracy simply concerns whether the vote shares received by parties on election day sit within the set of values surrounding the relevant vote share point estimates offered by polls. For any given poll, if the vote share received by a party on election day is contained within the set of values surrounding their predicted vote share as defined by the margin of error, then that poll is boundedly accurate. If their vote share is not contained within this set, then the poll is boundedly inaccurate. Formally, this is represented by equation 22, where a poll is accurate (and represented by a value of 1) if the vote share of a party, *V*, is an element of set of values defined by its margin of error, *P*. If *V* is not an element of *P*, then a poll is incorrect and represented by a value of 0.

$$BA = \begin{cases} 1 \text{ if } V \in P \\ 0 \text{ if } V \notin P \end{cases} \tag{22}$$

Importantly, pre-election polls render multiple predicted vote shares. The number of vote shares predicted can be no fewer than two, as elections require contestation between at least two parties or candidates to occur, and, while some polls tend not to address smaller parties, has no strictly defined upper limit. Each of these predicted vote shares will have a set of possible values surrounding it as defined by its margin of error. As the values associated with the binary operationalisation of bounded inaccuracy are assigned per party, each of these sets of values will then be compared to the vote shares received by their respective parties on election day to establish whether the predictions they represent were boundedly accurate. Each poll is therefore assigned as many values for BA as parties it addresses. Therefore, BA presents a similar problem to A'_{i}, insofar as it provides multiple values of the same error measure per poll, making error decomposition using this metric problematic. Again, this will be expanded upon in the subsequent section. Fortunately, the bounded conceptualisation of polling inaccuracy lends itself to a continuous operationalisation which can be aggregated on a per-poll basis. The first step in this process is outlined in equation 23 where *V* represents the vote share received by the party on election day, $\overline{\lim}\,\mathrm{MOE}$ represents the upper bound of the margin of error surrounding the predicted vote share of this party, and $\underline{\lim}\,\mathrm{MOE}$ represents its lower bound.

$$\text{If BA} = 0 \;\rightarrow \mathrm{BI} = \begin{cases} V - \overline{\lim}\,\mathrm{MOE} \;\text{ if } V > \overline{\lim}\,\mathrm{MOE} \\ \left|V - \underline{\lim}\,\mathrm{MOE}\right| \text{ if } V < \underline{\lim}\,\mathrm{MOE} \end{cases} \tag{23}$$

The presence of continuous bounded inaccuracy is strictly conditional upon a poll being boundedly inaccurate in a binary sense. If this condition is met, the degree of continuous bounded inaccuracy is measured as the extent to which the vote share of a party lies outside the margin of error of a poll. If the vote share received by a party on election day lies above the upper limit of the margin of error surrounding the predicted vote share provided by a poll, then this upper limit is subtracted from the vote share to establish the degree to which the poll was boundedly inaccurate. If the vote share is smaller than the lower limit of the margin of error, the absolute difference is taken to avoid negative values. The degree of bounded inaccuracy exhibited by a poll is measured on a per-party basis. Therefore, each poll possesses as many values as it addresses parties.

As the degree of bounded inaccuracy provides multiple error values for each poll, it is incommensurate with the election-level decomposition of polling error. As decomposition rests on the clear distinction between within-group and between-group variances, these groups and their nested nature must be represented unidimensionally within the data. In the two-level model that most straightforwardly represents the focus of my thesis, between-group error variance is that which occurs between elections, while within-group variance is that which occurs within elections and, therefore, between those polls nested within these elections. If each poll is assigned multiple values for the same measure of error, this measure ceases to uniquely vary between polls and, instead, becomes a measure which varies both within *and* between polls. This complicates the hierarchical data structure on which the decomposition rests.

The problem presented by individual measures of per-party bounded can be remedied through aggregation. This is shown in equation 24 which takes the average of the per-party, continuous bounded inaccuracy values associated with a poll.

$$\mathbf{ABI} = \frac{\sum_{\mathrm{i=1}}^{\mathbf{n}} \boldsymbol{BI}_{\mathrm{i}}}{\boldsymbol{n}} \tag{24}$$

The outcome of this aggregation will be directly affected by the manner in which average bounded inaccuracy is coded. Empty data cells will not contribute towards *n*, but zeroes will. Therefore, it is possible to calculate two forms of average bounded in accuracy. In the first (ABI 1), parties for which a poll does not exhibit bounded inaccuracy are excluded from the calculation. In the second (ABI 2), these parties are coded as having zero continuous bounded inaccuracy and are therefore included within the denominator of equation 24. As these calculative differences stand to alter the impact of average bounded, I include both approaches in later analysis.

Measuring Substantive Inaccuracy

The accuracy of polls' ability to correctly call substantive political outcomes is the least straightforward measure to operationalise in a way that is internationally tractable. As discussed in the previous chapter, the manner in which the substantive, politically relevant outcome of elections is determined varies between systems and, therefore, countries. It also necessarily varies within countries over time if they are subject to changing electoral systems.

So, devising a measure to translate predicted vote shares into politically relevant seat configurations in a manner that is both tractable internationally and over time is not feasible.

The difficulty of devising a singular, suitably tractable measure of substantive inaccuracy does not mean that its underlying properties cannot be measured. In the previous chapter, I posited that the substantive (in)accuracy of polls within can be gleaned from whether they successfully predict the party or candidate with the largest share of the vote. In my analysis, I refer to this measure as 'largest vote recipient correct' (LVRC). It is calculated in relation to whether a poll correctly identifies the party or candidate that receives the largest share of the vote. This can be operationalised straightforwardly as a simple binary. If the party or candidate in receipt of the largest share of the vote on election day is the same that predicted by a poll, then that poll is said to be correct and coded as 1. If this is not the case, then a poll is considered incorrect and coded as 0. This is formalised in equation 25 below.

$$\mathbf{LVRC} = \begin{cases} \mathbf{0\ if\ predicted\ recipient} \neq \mathbf{actual\ recipient} \\ \mathbf{1\ if\ predicted\ recipient} = \mathbf{actual\ recipient} \end{cases} \tag{25}$$

Inter-relation Between Measures: Ex-ante Expectations

Despite being conceptually distinct, the measures of polling inaccuracy operationalised within this chapter can be expected to have a degree of practical inter-relation. This inter-relation will be particularly pronounced for those measures with calculative overlap or a common focus. Most clearly, MAE can be expected to have considerable overlap with the variants of ABI. As MAE represents the average of all unsigned deviations between predicted and actual vote shares, a poll with a higher MAE will have a larger average difference between these values. The larger this difference, the more likely it is that the vote share received by a party or candidate on election day sits outside of the margin of error associated with a poll. This increases the likelihood of bounded inaccuracy for all parties addressed by a poll and therefore increases the probability of higher average bounded inaccuracy across these parties.

A high DIM value implies poor prediction of the margin between the leading parties or candidates in an election. This misprediction may concern the over- or under-estimation of the vote share of the leading party or candidate, the party or candidate in second place, or both. As the vote share received by all parties must sum to 100 – that is, all valid, unspoiled ballots will be distributed across all competitors in each election – any instance of the over- or under-estimation of the vote share received by one competitor will result in other, cumulatively proportionate mispredictions elsewhere. So, the higher the DIM value, the larger the misprediction to be distributed across other competitors is. The larger this misprediction, the greater the average error of a poll can be expected to be. In this way, DIM and MAE can be expected to be related.

Larger APB values are representative of a greater number of meaningful over- and under-estimations of vote shares. The larger this value, the more likely a poll is to meaningfully mispredict the vote share of each party or candidate. The increased likelihood of per-contestant error lends itself to the expectation that, when aggregated, these errors would lead to a higher MAE value. As such, APB and MAE can be expected to be related.

My two binary measures of polling inaccuracy, LVRC and SBP, can also be expected to be related. Both concern the vote shares of the leading parties or candidates within an election. A SBP value of 1 is predicated on the presence of statistically significant over-or under-estimation surrounding the vote share estimate of the leading party or candidate. If a poll significantly over- or under-estimates the vote share of the leading party or candidate, it is less likely to correctly call the recipient of the largest vote share, as a discrepancy exists between their prediction and the reality of electoral returns. As such, SBP and LVRC can be expected to be negatively related, with a value of 1 in the former being more likely to yield a value of 0 in the latter.

Given their inherent interconnectedness, the two conceptualisations of average bounded inaccuracy (ABI 1 and ABI 2) will also possess a high degree of association. Though these relationships can be expected on the basis of theoretical interplay between my measures of error, it is important to put them to the test in order to establish whether they are borne out within my data.

Testing Expectations

To test the extent of inter-relation between my outcome variables, I use correlational measures. The most commonly used measure of association between continuous variables is Pearson's *r* correlation coefficient.[748] This measures the degree of linear dependence between two variables.[749] A correlation coefficient of 0 represents perfect independence, while values of 1 and -1 represent perfect positive dependence and perfect negative dependence, respectively.[750] In either case, one variable can be used to perfectly predict the other.

In the case of my outcome variables, such a value would be problematic if it arises between two variables measuring the same conceptualisation of polling error. In such a case, the inclusion of both outcome variables within later analysis would be redundant, as their measurements would not be meaningfully distinct. The Pearson's correlations for my continuous measures of polling inaccuracy are displayed in Table 1.

[748] Haldun Akoglu, 'User's Guide to Correlation Coefficients', *Turkish Journal of Emergency Medicine*, 18.3 (2018), 91 – 93 (p. 91).
[749] Gábor J, Székely and Maria L. Rizzo, 'Brownian Distance Covariance', *The Annals of Applied Statistics*, 3.4 (2009), 1236 – 1265 (p. 1237).
[750] Akoglu, p. 91.

Table 1: Pearson's r correlation coefficients for continuous measures of polling inaccuracy.

Error Measures	DIM	LPB	MAE	APB	ABI 1	ABI 2
DIM	1*					
LPB	0.255*	1*				
MAE	0.498*	0.227*	1*			
APB	-0.030*	-0.192*	-0.142*	1*		
ABI 1	0.514*	0.277*	0.769*	-0.165*	1*	
AB1 2	0.596*	0.294*	0.910*	-0.215*	0.872*	1*

* = significant to p ≤ 0.05
DIM = difference in margin, LPB = leading party bias, MAE = mean absolute error,
APB = average party bias, ABI = average bounded inaccuracy.

The strength of association between my continuous measures of polling inaccuracy varies from the weak -0.192 to the very strong 0.910. All correlations presented in Table 1 are significant to 95% and are therefore reliably different from 0. Fortunately, no measures of the same conceptualisation of polling error possess strong correlations. As such, each separate operationalisation is suitably distinct from every other, allowing them all to be justifiably included as discrete measures.

Expectedly, the two conceptualisations of average bounded inaccuracy (ABI 1 and ABI 2) display a high degree of inter-relation with a correlation coefficient of 0.872. Given that they represent coding variations of the same underlying measure, this strong correlation is not a concern.

Strong correlations are also present between mean absolute error (MAE) and both ABI 1 and ABI 2. These sit at 0.769 and 0.910, respectively. This bears out the ex-ante expectation of a strong relationship between these two outcome variables. While this expectation is borne out, the other expected relationships are not present within my dataset. APB possesses only a weak negative relationship with MAE (-0.142), while DIM and MAE possess a middling positive association with a correlation of 0.498. Though this positive relationship was not unexpected, its relative weakness was.

Pearson's *r* correlations are not appropriate for testing the inter-relation between my binary measures of polling inaccuracy, as they are not continuous variables. To measure the degree of dependence between my binary error measures, I employ the phi coefficient. The phi coefficient is the most commonly used measure of the strength of inter-relation between binary variables.[751] It is directly comparable to the Pearson's *r* correlation coefficient, as it is measured on the same scale from -1 to 1.[752] These values are also interpreted in the same manner, with 0 representing complete independence, -1 representing a perfect negative association, and 1 representing a perfect positive association.

Using the phi coefficient to assess the strength of inter-relation between my binary measures of polling inaccuracy, LVRC and SBP, produces a value of -0.043. This indicates that a negligibly small relationship exists between them. While this follows the expectation of a negative relationship between the two variables, its magnitude is lower than expected. This is perhaps understandable, as the variables measure distinct conceptualisations of polling inaccuracy, but is nevertheless unexpected.

With the inter-relationship between my outcome variables established, I move to establish the importance of election-level differences for polling error. I begin this process by providing descriptive analysis illustrating the election-by-election distribution of polling error across my

[751] M. De Cáceres, X. Font, and F. Oliva, 'Assessing Species Diagnostic Value in Large Data Sets: A Comparison Between Phi-coefficient and Ochiai Index', *Journal of Vegetation Science*, 19.6 (2008), 779 – 788 (p. 781).
[752] Miquel de Cáceres and Pierre Legendre, 'Associations Between Species and Groups of Sites: Indices and Statistical Inference', *Ecology*, 90.12 (2009), 3566 – 3574 (p. 3568).

measures. This serves to provide an intuitive demonstration of the variation of error across elections and, in so doing, establishes the basis for further investigation into the effect and importance of these differences.

4.4: Descriptive Analysis: Establishing the Basis for Election-level Investigation

The Significance of Membership Within Election-level Groupings

If differences between elections exist as important drivers of polling error, then membership within different elections ought to affect the prominence of this error. To test whether this is the case for my data, I run statistical tests to establish whether membership within different elections results in statistically significant differences between the mean values of my continuous polling error measures. If these statistically significant differences exist, then further analysis into the impact of the election level on polling error is warranted.

The most common approach to this is the use of classical one-way ANOVAs.[753] The use of classical one-way ANOVAS is based on the assumption of equal variance across groups. If variance is not equal across groups, then alternative testing strategies are required.[754] To establish whether classical one-way ANOVAs are suitable for my data, I subject it to tests of homoscedasticity which assess the equality of variance across election-level groupings. A range of tests exists which assess the null hypothesis that variance is homogenously distributed across grouping levels. However, these tests rest on a series of assumptions concerning the underlying data.[755] The first of these concerns the distribution taken by the data.

[753] Hae-Young Kim, 'Analysis of Variance (ANOVA): Comparing Means of More Than Two Groups', *Restorative Dentistry & Endodontics*, 39.1 (2014), 74 – 77 (p. 74).
[754] Ibid.
[755] Ibid.

The manner in which data is distributed dictates the broad form of statistical test that is appropriate to use. Two principal forms of test exist: parametric and non-parametric. Crucial to the selection of either parametric or non-parametric tests is whether the underlying data is normally distributed.[756] Parametric tests assume a normal distribution, while non-parametric tests do not.[757] To aid in choosing a test, I plot the distributions of my continuous measures of polling inaccuracy in Figure 11 below.

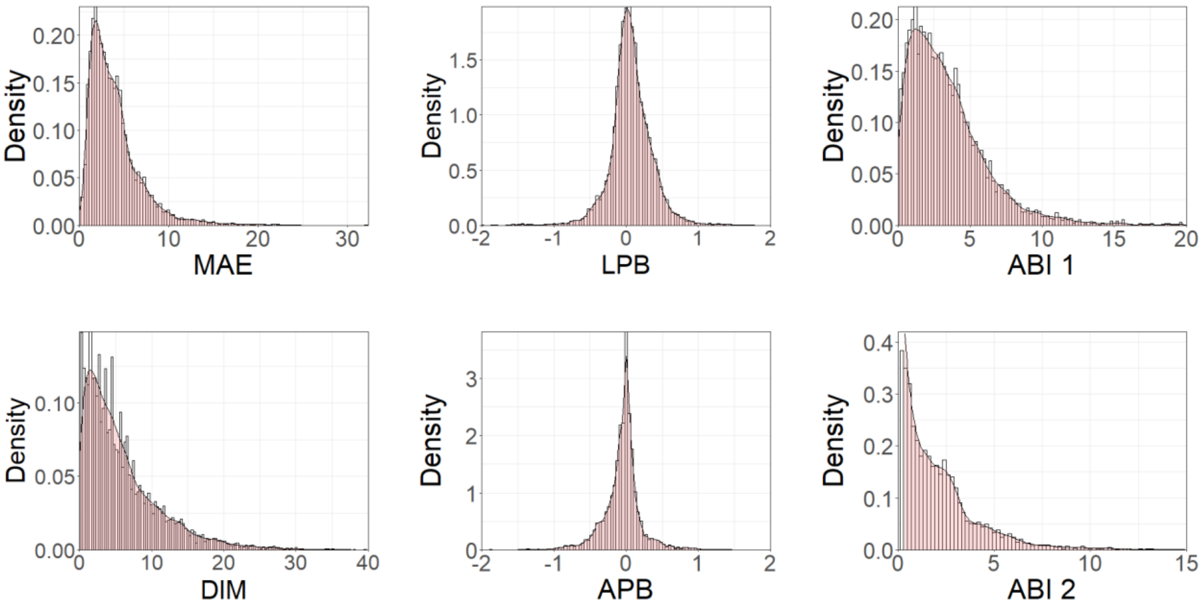

Figure 11: Histograms and overlayed density plots displaying the distribution of each continuous operationalisation of polling inaccuracy. MAE = mean absolute error, DIM = difference in margin, LPB = leading party bias, APB = average party bias, ABI 1 = first measure of average bounded inaccuracy, ABI 2 = second measure of average bounded inaccuracy.

From Figure 11, it is clear that while LPB is largely normally distributed about a mean of zero, and APB presents a slightly left-skewed, approximately normal distribution with high kurtosis, the other measures of polling inaccuracy present right-skewed non-normal distributions. Therefore, parametric tests of homoscedasticity cannot be applied across my continuous

[756] Amandeep Kaur and Robin Kumar, 'Comparative Analysis of Parametric and Non-parametric Tests', *Journal of Computer and Mathematical Sciences*, 6.6 (2015), 336 – 342 (p. 337).
[757] Ibid.

measures of polling inaccuracy. Instead, a non-parametric test of homoscedasticity is required. Of the available tests, the Fligner-Killeen test is the most robust to strong departures from the normal distribution, making it the most suitable for my measures of continuous polling inaccuracy.[758] Fligner-Killeen tests assess the null hypothesis that all group variances are homogenous.[759] In relation to my data, these tests assess whether the variance associated with my continuous measures of polling inaccuracy meaningfully differs between election-level groupings. The results of running these tests across my continuous measures of polling inaccuracy are displayed in Table 2 below. Throughout this chapter, and the remainder of the thesis, I deem statistics to be significant if they present a p-value of ≤ 0.05 and therefore have no more than a 5% probability of resulting from chance. This significance is denoted by a single asterisk.

Table 2: Fligner-Killeen tests of variance homogeneity across all continuous operationalisations of polling inaccuracy complete with chi-squared statistics values and p-values.

Error Measure	**χ^2 statistic**
MAE	3078*
DIM	3209*
LPB	3820*
APB	4795*
ABI 1	1633*
ABI 2	3121*

*significant at $p \leq 0.05$
MAE = mean absolute error, DIM = difference in margin, LPB = leading party bias,
APB = average party bias, ABI = average bounded inaccuracy.

758 Michael A. Fligner and Timothy J. Killeen, 'Distribution-Free Two-sample Tests for Scale', *Journal of the American Statistical Association*, 71.353 (1976), 210 – 213 (p. 210).
759 Ibid.

Table 2 reveals that, for all my continuous operationalisations of polling inaccuracy, Fligner-Killeen tests return chi-squared values that are statistically significant to 95% with p-values of ≤ 0.05. As such, the null hypothesis that variance is homogenous between election-level groupings within my dataset can be confidently rejected, and I conclude that variance is unequally distributed between groups within my data. Though this is to be expected a priori, as membership within different elections ought to differentially affect error variance due to their differing compositions, it means that I am unable to reliably use classical one-way ANOVAs to assess the impact of election-level differences on polling inaccuracy within my data. I must therefore employ an alternative to the one-way ANOVA that is robust to heterogenous group variance.

While many alternative tests exist in cases of heterogenous group variance,[760] given that my continuous variables do not uniformly present normal distributions, it is necessary that I choose a non-parametric test. The most common non-parametric alternative to the one-way ANOVA is the Kruskal-Wallis test.[761] This tests whether statistically significant differences exist between election-level groupings for each of my continuous measures of polling inaccuracy. The results of running Kruskal-Wallis tests across these measures are displayed in Table 3 below.

[760] Andrew J. Tomarken and Ronald C. Serlin, 'Comparison of ANOVA Alternatives Under Variance Heterogeneity and Specific Noncentrality Structures', *Psychological Bulletin*, 99.1 (1986), 90 – 99 (p. 90).
[761] M. Kraska-Miller, *Nonparametric Statistics for Social and Behavioral Sciences*, (Boca Raton: CRC Press, 2014), p. 123.

Table 3: Results of Kruskal-Wallis tests across all continuous operationalisations of polling inaccuracy complete with chi-squared statistics and p-values.

Error Measure	χ^2 statistic
MAE	5870*
DIM	4447*
LPB	5476*
APB	7329*
ABI 1	3249*
AB1 2	4689*

*significant at $p \leq 0.05$
MAE = mean absolute error, DIM = difference in margin, LPB = leading party bias, APB = average party bias, ABI = average bounded inaccuracy.

From the values displayed within Table 3, membership within different elections can be seen to produce statistically significantly different error values. This provides suggestive evidence that election-level differences serve as statistically significant drivers of my continuous measures of polling inaccuracy, lending suggestive evidence to my first hypothesis that elections vary in their ability to be accurately predicted.

To subject my dichotomised measures of polling inaccuracy to a similar analysis, I employ chi-squared tests of independence. In the context of this chapter, they are better understood as assessments of association, as they test the null hypothesis that variables are independent of one another and therefore not meaningfully associated.[762] I use these chi-squared tests to establish whether the election in which polls are conducted is meaningfully associated with the levels of error presented by my dichotomised measures of polling inaccuracy. The output of these tests is displayed in Table 4 below.

[762] Mary L. McHugh, 'The Chi-square Test of Independence', *Biochemia Medica*, 23.2 (2013), 143 – 149 (p. 143).

Table 4: The results of chi-squared tests of independence across all dichotomised measures of polling inaccuracy complete with chi-squared statistics and p-values.

Error Measure	χ^2 statistic
LVRC	6481*
SBP	2988*

*significant at $p \leq 0.05$
LVRC = largest vote recipient correct, SBP = significantly biased poll

From the figures presented in Table 4, it is clear that the election in which polls are conducted is statistically significantly associated with my binary measures of polling inaccuracy. It is significant to 95% in both cases with a p-values of ≤ 0.05. When these chi-squared results are considered alongside the results of the Kruskal-Wallis tests, it is clear that election-level differences are significant drivers of all of my measures of polling inaccuracy. As such, they merit further investigation.

While statistical tests indicate that the extent of prediction error exhibited by polls is driven by elections and therefore differs between them, they do not provide an intuitive sense of the degree to which this occurs. In the following sub-section, I visualise the difference in average error presented by polls across my studied elections to get a better sense of the impact of election-level differences on their accuracy.

Visualising the Effect of Election-level Differences on Polling Error

Figure 12 presents the average distributive inaccuracy exhibited by polls in my studied elections from 1936 to 2020. From the top pane of the figure, it is immediately clear that the average presented by polls differ considerably between my studied elections. Average MAE values vary from single-digit errors in the case of certain elections, to over twenty-point errors in the case of others. While the increased dispersion of average MAE in elections conducted after the year 2000 leads to differences between elections appearing more pronounced during this period, substantial and consistent differences in the average MAE exhibited by polls are apparent between elections across the full 84-year span of my dataset.

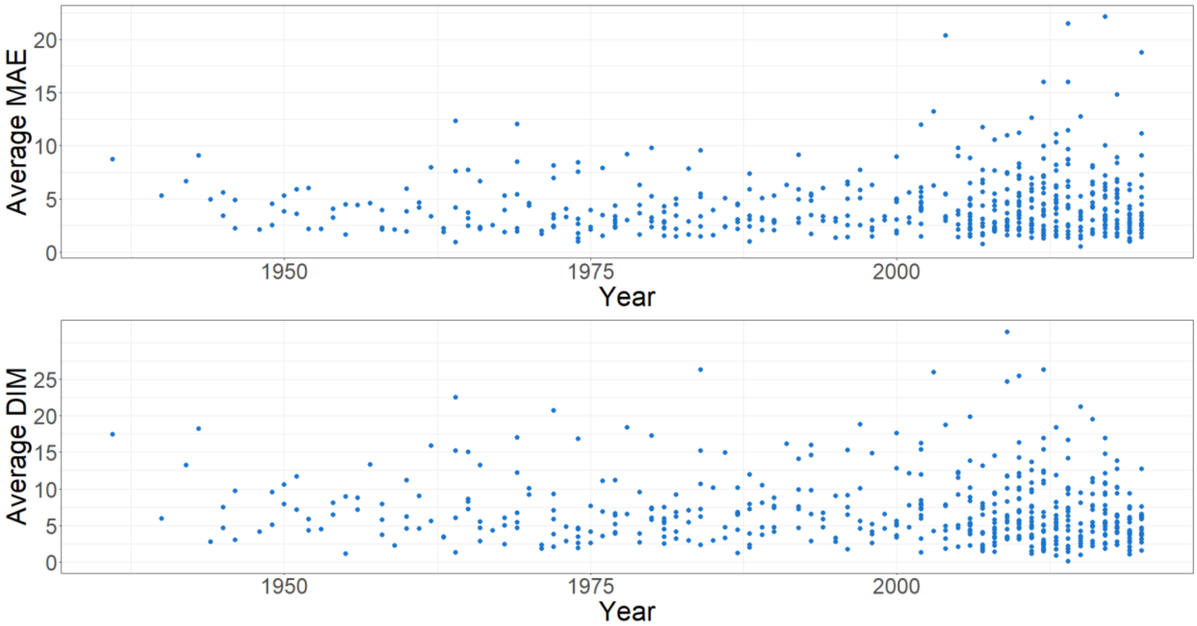

Figure 12: Average values for mean absolute error (MAE) and the difference in margin (DIM) exhibited by polls across all studied elections ordered chronologically. Each dot within the figure represents an individual election.

When the bottom pane of Figure 12 is considered, similar large-scale variation is evident in average DIM across elections, with values ranging from single-digit errors to errors of over forty points. Indeed, the evolution of average DIM values is broadly similar to that of MAE which is to be expected as they are both measures of the same distributive conceptualisation of polling inaccuracy and are therefore capturing similar forms of error. Barring the presence of single outlying value, the dispersion of average DIM values exhibited by polls across my studied elections remains similar over their 84-year span, indicating sustained and substantial differences in polling error between elections over time. That both measures of distributive polling inaccuracy vary between elections lends the first suggestive empirical evidence to the contention that differences between elections affect the propensity of polls to exhibit error.

Figure 13 displays the values associated with my measures of average bounded polling error (ABI 1 and ABI 2) over time with each point representing an individual election. It is clear that polls conducted for some elections exhibit high average ABI 1 values, while those conducted for others exhibit low values. Even with the more generous operationalisation of ABI 2, the same widespread variation is present, with the average bounded inaccuracy of polls clearly varying between elections.

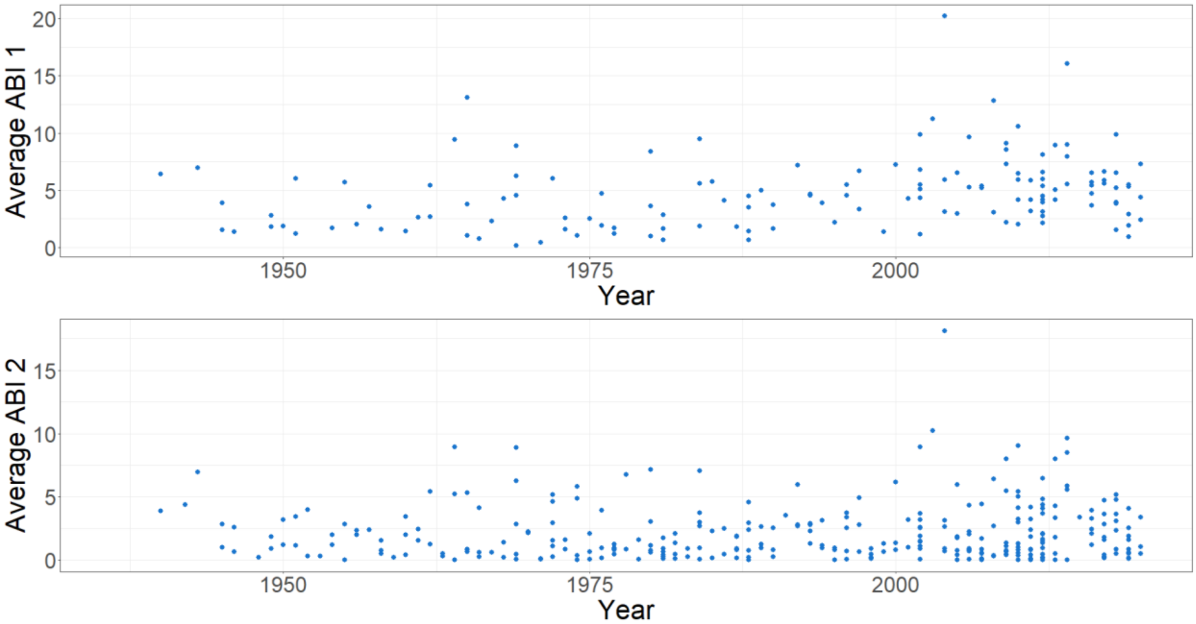

Figure 13: The mean values of both measures of average bounded inaccuracy (ABI 1 and ABI 2) exhibited by polls across all studied elections ordered chronologically. Each dot in the figure represents an individual election.

From Figure 13, it is equally clear that the average values of both ABI 1 and ABI 2 exhibited by polls across my studied elections present considerable dispersion over time. This indicates that differences in bounded polling error between elections are sustained, rather than existing as artefacts of a given time period or cluster of elections. That both measures of bounded polling error vary meaningfully and consistently between elections lends further suggestive evidence to the notion that election-level differences impact upon polling error.

To assess the variability of my measures of polling bias (LPB and APB) in a similar manner, I plot their average prominence across my studied elections in Figure 14 below. From the top pane of the figure, it is clear that the average LPB values presented by polls varies between my studied elections. Not only do they vary in magnitude, but they also vary directionally. In some elections, the average LPB exhibited by polls assumes positive values and therefore represents significant over-estimation, while in others it assumes negative values, indicating substantial under-estimation. That average LPB values vary to the extent of presenting directional differences between elections lends suggestive evidence to the assertion that differences between elections exist as key drivers of the extent and nature of polling bias.

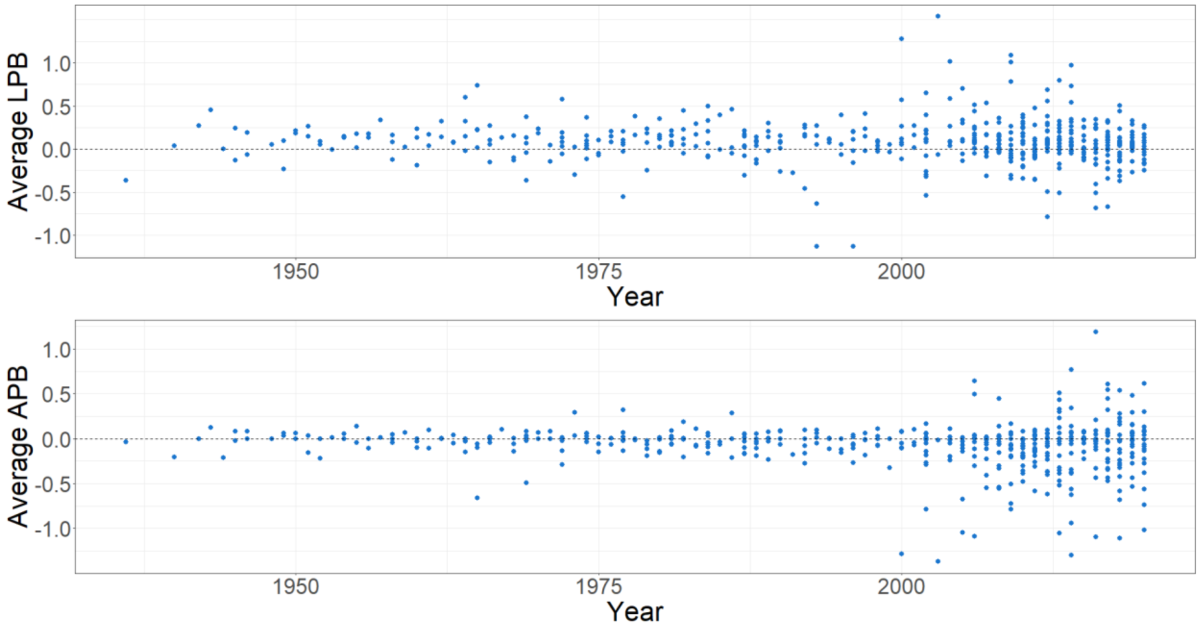

Figure 14: Average values for the leading party bias (LPB) and average per-party bias (APB) exhibited by polls across all studied elections ordered chronologically. Each dot in the figure represents an individual election.

The same behaviour is evident across measurements of average APB displayed in the lower pane of the figure, albeit to a somewhat lesser extent. While average LPB exhibits consistently high variation between elections over time, the dispersion of average APB values is comparatively muted prior to the year 2000. Nevertheless, average APB values consistently vary in both magnitude and direction between elections over the span of my data.

When considered alongside earlier findings concerning measures of distributive and bounded inaccuracy, the variability of the extent and nature of polling bias between elections lends further suggestive evidence to the contention that election-level differences bear upon polling inaccuracy to a meaningful degree.

Visualising the variability of my binary measures of polling inaccuracy across elections tells a more extreme version of the same story. To allow for clearer visualisation, the values associated with each binary variable have been transformed into percentages. These percentages represent the proportion of positive binary outcomes (coded as 1) exhibited by polls per election and are presented as density plots in Figure 15. In the case of SBP, positive values represent polls that exhibit statistically significant bias, while positive values in the case of LVRC indicate that polls correctly predicted the largest vote share recipient in a given election.

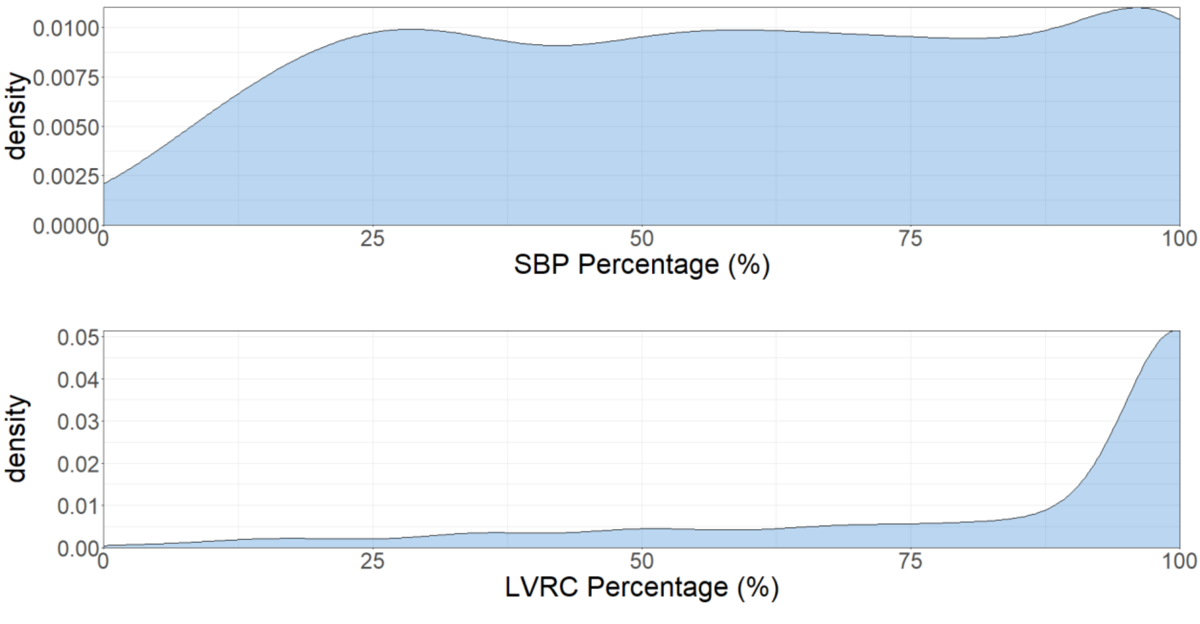

Figure 15: The density of positive values for SBP and LVRC across all polls contained within my dataset. SBP = significantly biased poll, LVRC = largest vote share recipient correct.

From Figure 15, it is clear that both SBP and LVRC exhibit substantial variability across my studied elections. The density of SBP percentages indicates that the degree to which polls exhibit statistically significant bias varies considerably between elections. At its most extreme, all polls associated with certain elections exhibit statistically significant levels of bias, while polls associated with other elections exhibit no bias of this nature at all. Beyond these extreme differences, more granular variation is apparent. The shape of the density plot associated with SBP indicates that the proportion of polls presenting statistically significant bias across my studied elections varies over the full range of possible values. The shape of the plot is also such that the majority of SBP proportions occur frequently within the data, with certain values possessing similar densities and, therefore, frequencies of occurrence. When this is considered alongside the extreme range of differences in SBP between elections, the degree to which polls exhibit statistically significant bias can be said to not only vary considerably between elections, but to do so frequently.

The density plot associated with LVRC lends itself to similar conclusions. From the density of proportions displayed, it is clear that in certain elections all polls correctly predict the recipient of the largest vote share, while in others all polls fail to do so. This serves as a clear and extreme example of the variability of LVRC values between elections. Though elections are most frequently characterised by polls that universally correctly predict the largest vote share recipient – as denoted by the density peak at 100% – my studied contests nevertheless present LVRC error proportions that occupy the full range of possible values. While the frequency of these values may be diminished, their presence further indicates variability in substantive polling error between elections.

From the analysis of descriptive plots, it is clear that each of my measures of distributive, bounded, and substantive polling error, as well as polling bias, varies between elections, lending suggestive evidence to the notion that election-level differences bear upon the inaccuracy of polls. While this is encouraging for the validity of hypothesis one, it is insufficient evidence in isolation. Though the observed variation in polling error between elections warrants further investigation, potentially confounding variables contained in alterative grouping levels must be identified before this can be conducted meaningfully.

Visualising the Need to Control for Country- and Pollster-level Differences

Given the multi-level nature of polling error, simply assessing the impact of election-level differences in isolation is insufficient to establish their impact. Polls are also subject to sources of error contained within other grouping levels, most notably the country and pollster levels. To establish the impact of the differences housed at these levels and the importance of controlling for them in later analysis, Figures 16 and 17 display the extent of average polling error across the different countries and polling organisations contained within my dataset.

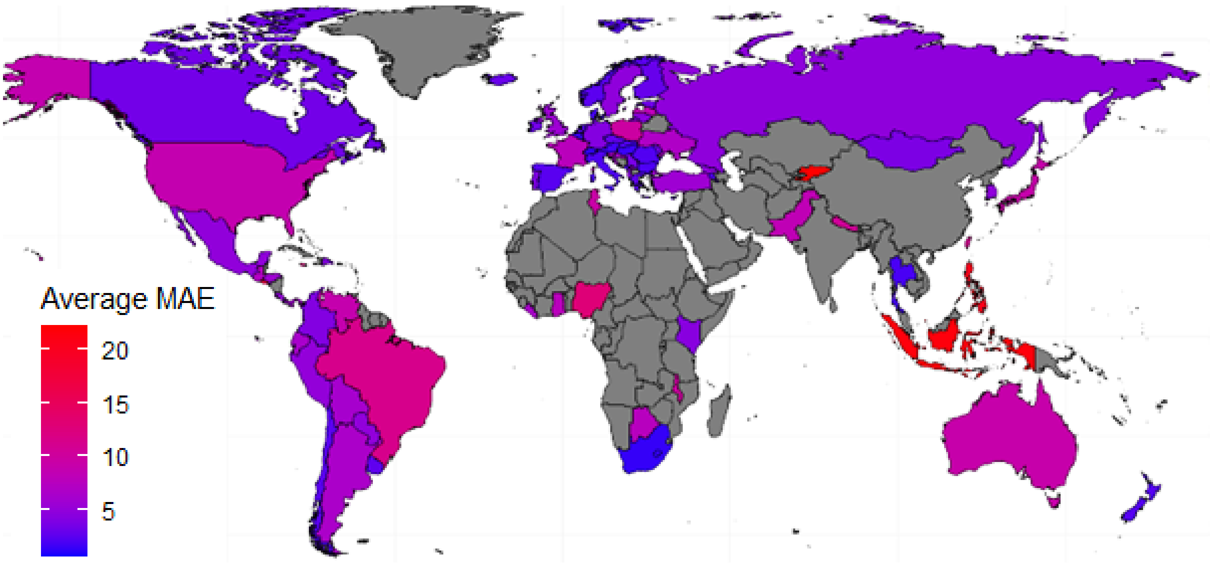

Figure 16: The average polling error, represented as mean MAE, across all countries within my dataset. The extent of error is displayed using a graduated colour scale from low (blue) to high (red), with grey shading representing those countries not included in my dataset.

From Figure 16, it is clear that the average mean absolute error (MAE) exhibited by polls varies considerably across countries. While certain countries present average MAE values of below five percentage points, others present errors upwards of twenty points. Polls conducted in Kyrgyzstan and Indonesia present the largest average MAE values, while polls conducted in

European countries generally present the lowest levels of average MAE. Though this would imply that countries in Europe and Asia present the lowest and highest errors on average, respectively, it is notable that the average error of polls varies both between and within continents. Though variation in error is most pronounced between countries housed within Asia, it is nevertheless apparent across countries contained within all six populated continents of the world. The error exhibited by polls also presents regional trends within continents. For example, divisions in average MAE can be seen between southern and northern Europe, south-east and north-east Asia, as well as southern and northern Africa, broadly defined. That polling error not only varies between countries and continents, but also between regions within continents, speaks to the need to control for country-, continent-, and sub-continental differences when assessing drivers of polling error variance.

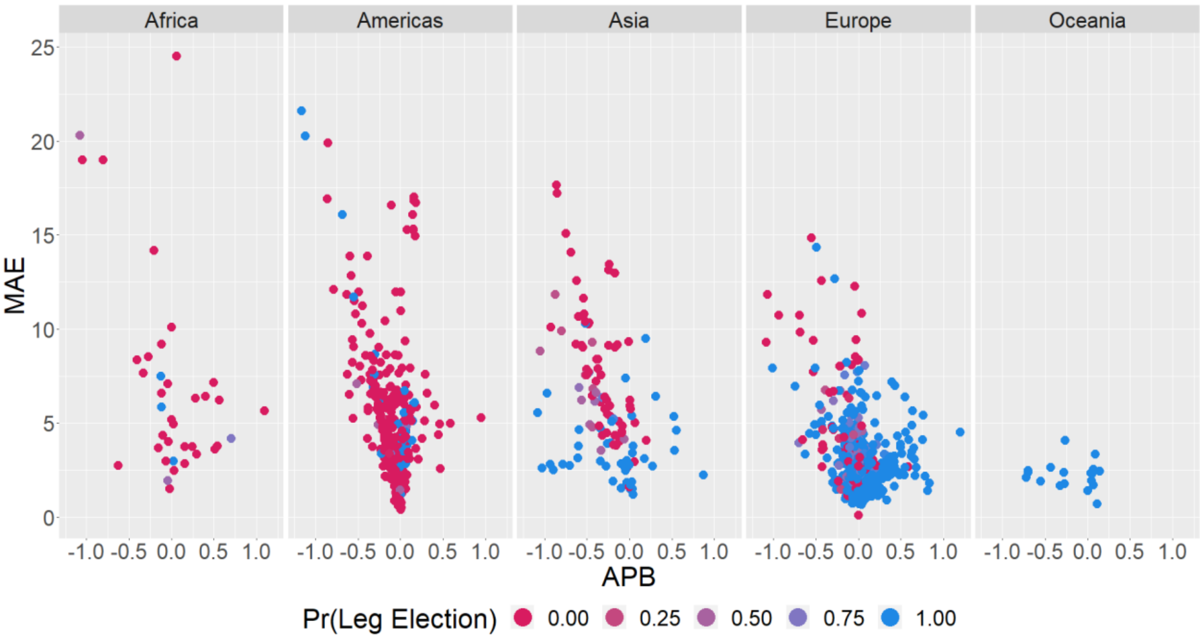

Figure 17: The average mean absolute error (MAE) and average per-party bias (APB) across all polling organisations within my dataset broken down by continent. Each point within the figure represents an individual polling organisation. Organisations are shaded according to the proportion of their polls that focus on legislative elections to capture differences in focus.

From Figure 17, it is clear that average extent of both polling error and bias varies considerably between polling organisations. This variance is most prominent across polling organisations conducting polls in Africa and least pronounced in those operating in Oceania. Nevertheless, the error and bias presented by organisations based in the remaining continental groupings also exhibit considerable variation. In Asia, polling organisations that principally conduct polls in relation to presidential elections exhibit considerably higher polling error and bias on average than those that focus on legislative contests. This trend is less apparent in other continental groupings, though polling organisations that focus on presidential contests do present a greater range of average error and bias within Africa. That the extent of polling error and bias varies between polling organisations speaks to the need to control for pollster-level differences when assessing polling error variance. The difference in the extent of error and bias presented by polling organisations across countries also further underscores the importance of controlling for country-level differences when decomposing polling error variance.

With the need to control for pollster- and country-level differences established, I move to analyse the effect of election-level differences on polling error variance. In the first instance, I decompose variance across two-level models, capturing the effect of election- and poll-level differences. I then decompose variance across three and four level models to account for the effect of pollster- and country-level differences.

4.5: Measuring the Effect of Election-level Differences on Polling Error Variance

ICC Estimates of Election-level Variance from Two-level Models

To establish the amount of polling error variance accounted for by the election level, I run a series of null multi-level models. Table 3 presents the results from my preferred maximum-likelihood based estimative approaches. If differences between elections are meaningful drivers of polling error, then I would expect a non-trivial proportion of variance in my measures of polling inaccuracy to be accounted for by the election level. While the minimum threshold for considering such group-level effects is 5%,[763] implicit within my hypothesis is the expectation that this value will be higher.

To begin, I present the ICC estimates born of my two-level frequentist models of polling inaccuracy. This allows for the most direct analysis of the impact of the election level differences, albeit absent country- and pollster-level controls. The results derived from these models are presented in Table 5.

Table 5: ICC estimates for continuous measures from maximum likelihood two-level models across whole dataset including standard errors (SE) and 95% confidence intervals.

	MAE	**DIM**	**LPB**	**APB**	**ABI 1**	**ABI 2**
MLE						
ICC	0.63	0.51	0.47	0.67	0.61	0.72
SE	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
95% CI	0.60 - 0.66	0.47 - 0.55	0.43 - 0.51	0.63 - 0.70	0.57 - 0.65	0.68 - 0.75
RMLE						
ICC	0.64	0.51	0.47	0.67	0.61	0.72
SE	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
95% CI	0.60 - 0.67	0.47 - 0.55	0.43 - 0.51	0.63 - 0.70	0.57 - 0.65	0.69 - 0.75

DIM = difference in margin, LPB = leading party bias, MAE = mean absolute error,
APB = average party bias, ABI = average bounded inaccuracy.

The ICC values displayed in Table 5 range from 0.47 in the case of LPB to 0.72 in the case of ABI 2. When the 95% confidence intervals are considered, this range extends from 0.43 to 0.75. This illustrates that, for my measures of continuous accuracy, election-level differences

[763] Raykov, p. 81.

account for between 43% and 75% of polling error variance. Both MLE and RMLE estimation strategies return almost identical results when taken to the second significant figure. Importantly, all ICC estimates sit substantially above the 5% threshold justifying the investigation of group-level effects. When this is considered in tandem with the sizable proportion of error accounted for by the election-level, this provides supportive evidence to the hypothesis that election-level differences are impactful for polling error.

To ensure that the results reported in Table 5 are robust and not simply artefacts of given estimative procedures, Table A1 in Appendix A contains a series of robustness checks. These checks report ICC values from the additional frequentist estimative procedures outlined earlier in this chapter (ICC1, Eta-, and Omega-squared). The values range from 0.35 to 0.67, indicating that election-level differences account for between 35% and 67% of variance across my continuous measures of polling inaccuracy. This range broadly agrees with the results reported in Table 5 and, again, all values rest considerably above the 5% threshold justifying the interrogation of group-level effects.

When the results reported in Table 5 are considered in tandem with the robustness checks, it is clear that election-level differences are impactful drivers of the variance exhibited by each of my measures of polling inaccuracy. However, their importance varies between conceptualisations of inaccuracy. ICC values range from 0.39 to 0.70 across my measures of distributive inaccuracy, while they range from 0.44 to 0.75 across my measures of bounded inaccuracy. While both ranges indicate that election-level differences account for a substantial portion of the observed variance in my measures of polling error, these differences are slightly more important for measures of bounded inaccuracy.

In sum, the ICC results calculated across my frequentist models of continuous inaccuracy lend further support to the hypothesis that election-level differences matter for polling error and, therefore, elections vary in their ability to be accurately predicted. They also indicate that the substantive findings in support of this hypothesis are robust across a variety of frequentist estimative techniques.

To test whether this holds across other approaches, Table 6 presents ICC values calculated in relation to my continuous measures of polling inaccuracy from models estimated using Bayesian MCMC. The models displayed within the table use my preferred half student-t priors and stand in agreement with their frequentist counterparts. ICC values range from 0.46 in the case of LPB to 0.72 in the case of ABI 2. When 95% confidence intervals are considered, this range extends from 0.43 to 0.74. This indicates that between 43% and 74% of the variance exhibited by my measures of polling inaccuracy is the result of election-level differences.

Table 6: ICC estimates for continuous measures from Bayesian MCMC two-level models across whole dataset including standard errors (SE) and 95% confidence intervals.

	MAE	**DIM**	**LPB**	**APB**	**ABI 1**	**ABI 2**
Half Student-t Priors						
ICC	0.64	0.51	0.46	0.67	0.61	0.72
SE	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)
95% CI	0.61 - 0.66	0.48 - 0.54	0.43 - 0.50	0.65 - 0.70	0.58 - 0.64	0.69 - 0.74

DIM = difference in margin, LPB = leading party bias, MAE = mean absolute error,
APB = average party bias, ABI = average bounded inaccuracy.

The range of ICC values displayed within Table 6 is strikingly similar to that provided by frequentist maximal likelihood models and agrees substantively with that provided by frequentist robustness checks. The values also agree that election-level differences exist as more impactful drivers of variance across my measures of bounded inaccuracy than my measures of distributive inaccuracy. Inclusive of 95% confidence intervals, ICC values for the former range from 0.58 to 0.74, while values for the latter range from 0.43 to 0.70. This presents a similar story to that told by my frequentist models. Importantly, all reported values sit significantly above the 5% threshold justifying the assessment of group-level effects, further undergirding the importance of adopting a multi-level approach to sources of polling error.

The ICC values provided in Table 6 again provide supportive evidence for the hypothesis that election-level differences are impactful for polling error and that, therefore, elections vary in their ability to be accurately predicted. It also suggests that the findings derived from earlier models were not simply an artefact of frequentist approaches to parameter estimation. However, in the same manner as earlier models, to ensure that the results calculated from the Bayesian MCMC model in Table 6 are robust and not simply an artefact of a given prior specification, I run a series of robustness checks contained in Table A2 of Appendix A.

The robustness checks for my Bayesian MCMC model calculate ICC values on the basis of additional model specifications using half-Cauchy priors. The range of values provided by these robustness checks is identical to that displayed in Table 6, ranging from 0.46 in the case of LPB, to 0.72 in the case of ABI 2. This indicates that election-level differences account for between 46% and 72% of the variance exhibited by my continuous measures of polling inaccuracy. Importantly, this agreement between models demonstrates that my findings are not simply an artefact of given prior distributions, but are in fact robust across a range of specifications.

The ICC values displayed in Table 6 along with their attendant robustness checks provide further evidence in support of the hypothesis that election-level differences are impactful for polling error and, therefore, that elections vary in their ability to be accurately predicted. For my continuous measures of polling inaccuracy, the values that support this hypothesis are robust across a variety of both frequentist and Bayesian approaches to parameter estimation, demonstrating that they are not simply artefacts of any given modelling approach.

While the ICC figures provided so far suggest that my hypothesis holds in the case of my continuous measures of polling inaccuracy, I move to analyse whether this is the case for my binary measures. Table 7 displays the results of calculating ICC values from two-level models estimated using my preferred approach, Laplace approximation across my binary measures of polling inaccuracy, correctly calling the recipient of the largest vote share (LVRC) and whether a poll is significantly biased (SBP).

Table 7: ICC estimates for binary measures from two-level models using Laplace approximation across whole dataset including standard errors (SE) and 95% confidence intervals.

	LVRC	**SBP**
Laplace Approximation		
ICC	0.86	0.43
SE	(0.02)	(0.03)
95% CI	0.81 – 0.89	0.38 – 0.49

LVRC = leading vote recipient correct; SBP = significantly biased poll

From Table 7, it is immediately clear that election-level differences are substantially more important for LVRC than SBP. Election-level differences account for 86% of the variance associated with LVRC as opposed to 43% of the observed variance in SBP. The ICC value of 0.43 associated with SBP sits within the range of values associated with distributive inaccuracy reported earlier within the chapter (0.39 to 0.70), rendering it unsurprising. However, the value associated with LVRC, 0.86, indicates that election-level differences are more impactful drivers of substantive inaccuracy than my other conceptualisations of polling inaccuracy.

Both ICC values reported in Table 7 indicate that election-level differences are important determinants of the variance associated with my binary measures of polling inaccuracy. Importantly, when 95% confidence intervals are considered, the reported ICC values sit a considerable distance from the 5% threshold justifying the assessment of group-level effects. This further affirms the appropriateness of a multi-level approach to understanding sources of polling inaccuracy.

To ensure that the ICC values reported for my binary measures of polling error are not simply an artefact of Laplace approximation, Table A3 in Appendix A contains robustness checks which calculate the amount of variance accounted for by the election-level in models estimated using adaptive Gauss-Hermite quadrature. These values agree with those reported in Table 7, with election-level differences accounting for 88% of the variance associated with LVRC and 44% of the variance associated with SBP. This corroborates the assertion that election-level differences serve as more pronounced drivers of substantive polling inaccuracy than my other conceptualisations. It also affirms that the variance in SBP accounted for by election-level differences is substantial and within the range of values associated with other measures of distributive inaccuracy. Again, in all cases the ICC values reported sit significantly above the 5% threshold for justifying the assessment of group-level effects, further underscoring the appropriateness of treating sources of polling error as multi-level in nature.

When considered together, the results presented in Table 7 and Appendix A support the hypothesis that election-level differences stand as significant drivers of measures of polling inaccuracy, and therefore, that elections vary in their ability to be accurately predicted. However, these results are only derived from frequentist models. To ensure that the impact of election-level differences on my binary measures of polling inaccuracy is robust to different analytical approaches, I move to calculate ICC values from Bayesian MCMC models. To achieve this, Table 8 displays ICC estimates for my binary measures of polling inaccuracy taken from Bayesian MCMC models using my preferred half student-t priors.

Table 8: ICC estimates for binary measures from Bayesian MCMC two-level models run the across whole dataset including standard errors (SE) and 95% confidence intervals.

	LVRC	**SBP**
Half Student-t Priors		
ICC	0.84	0.45
SE	(0.02)	(0.03)
95% CI	0.81 – 0.88	0.39 – 0.50

LVRC = largest vote recipient correct; SBP = significantly biased poll

The ICC values calculated from the Bayesian MCMC models agree substantively with their frequentist counterparts. Election-level differences account for 84% of the observed variance in LVRC, while they account for 45% of the variance in SBP. These values further indicate that election-level differences are more impactful drivers of variance for substantive polling inaccuracy than for my other conceptualisations. The ICC value associated with SBP again rests within the range associated with other measures of distributive inaccuracy (39% to 70%). The robustness checks for this model conducted in Table A4 of Appendix A using half-Cauchy priors produce similar results, with election-level differences accounting for 85% of the variance exhibited by LVRC and 44% of that associated with SBP. This indicates that the findings displayed in Table 8 are robust across a range of prior specifications.

When considered alongside their robustness checks, the ICC values reported in Table 8 provide further evidence for the importance of election-level differences as important drivers of the variance exhibited by both LVRC and SBP. This, in turn, supports the hypothesis that election-level differences are significant drivers of polling inaccuracy and, therefore, that elections vary in their ability to be accurately predicted.

The ICC values contained within Tables 5 through 8 along with their attendant robustness checks in Appendix A provide evidence to suggest that election-level differences exist as significant drivers of polling inaccuracy. However, as these values are calculated using two-level models, the magnitude of the percentages reported may not be truly representative of reality given the absence of controls for confounding grouping levels. To test the impact of these confounding levels on election-level ICC scores, I move to assess three-level models with elections nested within countries.

ICC Estimates of Election-level Variance from Three-level Models

To test whether the country in which elections, and therefore polls, take place affects the variance exhibited by my measures of polling inaccuracy, I calculate ICC values across a series of three-level models with elections nested within countries. This allows for country-level effects to be controlled for, providing a more representative measurement of the variance accounted for by the election-level alone. Table 9 displays the election-level ICC values calculated from three-level models based on maximum likelihood estimation across all continuous measures of polling inaccuracy when the impact of country-level differences on polling error variance is controlled for.

Table 9: Election-level ICC estimates for continuous measures from three-level models using maximum likelihood estimation (MLE) and restricted maximum likelihood estimation (RMLE) run across the whole dataset including standard errors (SE) and 95% confidence intervals.

	MAE	**DIM**	**LPB**	**APB**	**ABI 1**	**ABI 2**
MLE						
ICC	0.27	0.41	0.34	0.24	0.25	0.22
SE	(0.03)	(0.02)	(0.02)	(0.03)	(0.04)	(0.03)
95% CI	0.21 - 0.33	0.37 - 0.46	0.29 - 0.39	0.18 - 0.29	0.18 - 0.32	0.18 - 0.27
RMLE						
ICC	0.27	0.42	0.34	0.24	0.25	0.22
SE	(0.03)	(0.02)	(0.02)	(0.03)	(0.04)	(0.03)
95% CI	0.20 - 0.32	0.37 - 0.46	0.29 - 0.38	0.18 - 0.29	0.18 - 0.31	0.18 - 0.26

DIM = difference in margin, LPB = leading party bias, MAE = mean absolute error,
APB = average party bias, ABI = average bounded inaccuracy.

From Table 9, it is clear that controlling for the country-level decreases the amount of variance exhibited by my measures of polling error that can be attributed to election-level differences. When compared to earlier two-level models, controlling for country-level differences reduces election-level ICC values by between 0.10 in the case of DIM and 0.50 in the case of ABI 2. As such, controlling for country-level differences differentially reduces the magnitude of election-level ICC values across my continuous measures of polling inaccuracy. These reductions are accounted for in Table A5 of Appendix A. After accounting for the impact of the country level, election-level differences still account for between 22% and 42% of the variance exhibited by my continuous measures of polling inaccuracy. When 95% confidence intervals are considered, this range extends from lows of 18% in the case of ABI 1, ABI 2, and APB to a high of 46% in the case of DIM. These figures remain considerably higher than the 5% threshold justifying the assessment of grouping levels.

Together, the ICC values reported in Table 9 demonstrate that election-level differences remain important drivers of variance across my measures of polling inaccuracy even when country-level effects are controlled for. This provides further support for the hypothesis that election-level differences are impactful for polling error variance and, therefore, that elections vary in their ability to be accurately predicted.

To establish whether the findings derived from frequentist models hold across different estimative techniques, Table 10 displays ICC values calculated from models estimated using Bayesian MCMC. These models use my preferred student-t priors and again display the percentage of variance displayed by my continuous measures of polling error that is accounted for by election-level differences when the country level is controlled for.

Table 10: Election-level ICC estimates for continuous measures from Bayesian MCMC three-level models across whole dataset including standard errors (SE) and 95% confidence intervals.

	MAE	**DIM**	**LPB**	**APB**	**ABI 1**	**ABI 2**
Half Student-t Priors						
ICC	0.27	0.41	0.33	0.24	0.25	0.22
SE	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
95% CI	0.25 - 0.29	0.39 - 0.43	0.31 - 0.35	0.21 - 0.26	0.22 - 0.28	0.19 - 0.25

DIM = difference in margin, LPB = leading party bias, MAE = mean absolute error,
APB = average party bias, ABI = average bounded inaccuracy.

Much like earlier frequentist models, the ICC values in Table 10 indicate that controlling for country-level effects reduces the proportion of the variance observed across my continuous measures of polling inaccuracy that can be attributed to election-level differences. Compared to earlier two-level models, ICC values are reduced by between 0.10 in the case of DIM and 0.50 in the case of ABI 2. This indicates that controlling for country-level differences reduces the amount of variance across my measures that can be attributed to election-level differences by between 10% and 50%. The range of these reductions agrees substantively with earlier frequentist models, with DIM and ABI 2 again occupying its extremes.

Despite the reductions that result from controlling for country-level differences, Table 10 indicates that election-level differences still account for between 22% and 41% of the variance observed across my continuous measures of polling inaccuracy. All of the percentages reported within the table remain considerably above the 5% threshold, even when 95% confidence intervals are considered, justifying the assessment of group-level effects. These values lend further support to the hypothesis that election-level differences are impactful drivers of polling error and, therefore, that elections vary in their ability to be accurately predicted.

The ICC values reported in Table 10 are robust across a range of prior specifications. Table A6 in Appendix A displays the ICC values calculated from Bayesian MCMC models using alternative half-Cauchy priors. These values are identical to those calculated from models using half student-t priors, indicating that differential prior specifications do not affect results.

When considered together, the ICC values calculated from my frequentist and Bayesian three-level models demonstrate that election-level differences remain substantial drivers of variance in my continuous measures of polling inaccuracy even when country-level effects are controlled for. This conclusion has been shown to be robust across a range of model specifications and is therefore not simply an artefact of a given estimative approach or prior distribution.

To assess whether election-level differences remain impactful drivers of my binary measures of polling inaccuracy when controlling for country-level differences, I run a series of three-level models using my preferred estimative approaches. I begin with a frequentist three-level model estimated using Laplace approximation which is displayed in Table 11.

Table 11: Election-level ICC estimates for binary measures from three-level models using Laplace Approximation across whole dataset including standard errors (SE) and 95% confidence intervals.

	LVRC	SBP
Laplace Approximation		
ICC	0.75	0.34
SE	(0.02)	(0.01)
95% CI	0.71 – 0.77	0.32 – 0.37

LVRC = leading vote recipient correct; SBP = significantly biased poll

The ICC values calculated from this model indicate that election-level differences account for 75% of the variance exhibited by LVRC and 34% of the variance exhibited by SBP. Compared to earlier two-level models, controlling for country-level effects reduces the amount of variance that can be attributed to election-level differences by 11% in the case of LVRC and 9% in the case of SBP. Nevertheless, the election-level still accounts for a significant proportion of the variance observed across my binary measures of polling inaccuracy, with values sitting comfortably above the 5% threshold meriting the investigation of group-level effects.

Even when the 95% confidence intervals are considered, the ICC values presented in Table 11 indicate that differences between elections serve as substantial drivers of my binary measures of polling error. This lends further support to the hypothesis that election-level differences exist as significant drivers of polling inaccuracy and, therefore, that elections vary in their ability to be accurately predicted.

To establish whether the conclusions drawn from models using Laplace approximation are robust across different estimative strategies, Table 12 displays ICC values calculated from a three-level model estimated using Bayesian MCMC with my preferred half student-t priors. These values agree substantively with those reported in earlier frequentist models, indicating the election-level differences account for 69% of the variance observed in LVRC and 35% of the variance observed in SBP when country-level differences are controlled for. When additional prior specifications are considered (see: Table A8 of Appendix A), election-level differences are seen to be responsible for 68% of the variance associated with LVRC and, again, 35% of the variance associated with SBP, indicating that the findings displayed in Table 12 are robust across different priors.

Table 12: Election-level ICC estimates for binary measures of polling error calculated from Bayesian MCMC three-level models across whole dataset including standard errors (SE) and 95% confidence intervals.

	LVRC	**SBP**
Half Student-t Priors		
ICC	0.69	0.35
SE	(0.02)	(0.01)
95% CI	0.65 - 0.71	0.32 - 0.37

LVRC = largest vote recipient correct; SBP = significantly biased poll.

When the ICC values displayed in Table 12 are considered alongside their robustness checks, it is clear that election-level differences still account for a substantial proportion of the variance observed across my measures of binary polling inaccuracy even when country-level effects are controlled for. This lends further support to the hypothesis that election-level differences exist as significant drivers of polling inaccuracy and, therefore, that elections vary in their ability to be accurately predicted.

In total, the findings displayed in Tables 9 through 12 demonstrate that election-level differences are significant determinants of the variance displayed by both my continuous and binary measures of polling error even when country-level differences are controlled for. This conclusion is robust across both frequentist and Bayesian estimative techniques, as well as a variety of prior specifications.

ICC Estimates of Election-level Variance from Four-level Models

The final test of the significance of election-level differences as drivers of polling error variance comes from decomposing variance in relation to four-level models. These models allow for both the country- and pollster-levels to be accounted for, providing the most dependable and representative measure of the importance of the election level. To analyse the variance accounted for by election-level differences across four-level models, I proceed in the same manner as previous sub-sections, beginning with frequentist maximum likelihood models of my continuous measures of polling inaccuracy displayed in Table 13.

Table 13: Election-level ICC estimates for continuous measures of polling error calculated from maximum likelihood four-level models across whole dataset including standard errors (SE) and 95% confidence intervals.

	MAE	**DIM**	**LPB**	**APB**	**ABI 1**	**ABI 2**
MLE						
ICC	0.26	0.38	0.32	0.21	0.25	0.22
SE	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
95% CI	0.24 - 0.28	0.37 - 0.39	0.31 - 0.33	0.20 - 0.23	0.22 - 0.28	0.18 - 0.25
RMLE						
ICC	0.26	0.38	0.32	0.21	0.25	0.21
SE	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
95% CI	0.24 - 0.28	0.37 - 0.39	0.31 - 0.33	0.20 - 0.23	0.22 - 0.28	0.17 - 0.24

DIM = difference in margin, LPB = leading party bias, MAE = mean absolute error,
APB = average party bias, ABI = average bounded inaccuracy.

The ICC values displayed in Table 13 range from 0.21 in the case of APB and ABI 2 to 0.38 in the case of DIM. This indicates that election-level differences account for between 21% and 38% of the variance observed across my continuous measures of polling inaccuracy, even when both the country- and pollster-levels are accounted for. When 95% confidence intervals are considered, this range extends from 17% to 39%. As these values sit significantly above the 5% threshold for considering group-level effects, it is clear that election-level differences remain important drivers of error variance even when contained within a four-level structure of competing variance parameters and are worthy of group-level study.

When the ICC values calculated from four-level models are compared to their earlier three-level counterparts, controlling for pollster-level differences can be seen to diminish the impact of both the election and the country levels. The inclusion of pollster-level differences sees small reductions in the amount of observed variance in polling error attributable to election-level differences, ranging from a 1% decrease in the case of MAE to 4% in the case of DIM when estimated using RMLE. Similarly, the inclusion of pollster-level differences reduces the amount of variance accounted for by the country level, ranging from 9% in the case of MAE to 2% in the case of LPB. This suggests that a small portion of variance in polling error brought about by differences between pollsters was errantly attributed to election- and country-level differences. This is explored further in Table A10 of Appendix A.

Overall, the ICC estimates displayed in Table 13 indicate that election-level differences still account for a significant proportion of the variance observed across my continuous measures even when the country- and pollster-levels are controlled for. This lends yet more support to the hypothesis that election-level differences exist as significant drivers of polling error and, therefore, that elections vary in their ability to be accurately predicted.

To establish whether these findings are robust, or merely an artefact of frequentist maximum likelihood estimation, I calculate ICC values from four-level models estimated using Bayesian MCMC with my preferred half student-t priors. These values are presented in Table 14 and range from 0.21 in the case of ABI 2 and 0.38 in the case of DIM. This indicates that election-level differences account for between 21% and 38% of the variance observed across my continuous measures of polling inaccuracy when country- and pollster-level differences are controlled for. When 95% confidence intervals are accounted for, this range extends from 18% to 39%. These percentages agree substantively with those reported from four-level frequentist models, indicating that the findings are robust across estimative approaches.

Table 14: Election-level ICC estimates for continuous measures of polling error from Bayesian MCMC four-level models across whole dataset including standard errors (SE) and 95% confidence intervals.

	MAE	**DIM**	**LPB**	**APB**	**ABI 1**	**ABI 2**
Half Student-t Priors						
ICC	0.26	0.38	0.31	0.23	0.25	0.21
SE	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
95% CI	0.24 - 0.29	0.37 - 0.39	0.30 - 0.33	0.21 - 0.25	0.22 - 0.28	0.18 - 0.25

DIM = difference in margin, LPB = leading party bias, MAE = mean absolute error,
APB = average party bias, ABI = average bounded inaccuracy.

The test whether the findings presented in Table 14 are robust across different prior distributions, Table A11 of Appendix A displays ICC values calculated from four-level Bayesian MCMC models estimated using alternative half-Cauchy priors. These values almost exactly mirror those calculated from models using half student-t priors, varying by only 1% in the case of ABI 1. This demonstrates that the findings in Table 14 not only agree with earlier frequentist models, but are also robust across alternative Bayesian prior specifications.

When compared to three-level models, controlling for the additional pollster-level yields small reductions in the amount of observed variance in my measures of polling error attributable to election-level differences. These reductions range from 1% in the case of MAE and APB, to 3% in the case of DIM. However, not all measures of polling inaccuracy are affected by the inclusion of pollster-level differences, with the ICC value associated with APB remaining unchanged. The variance attributable to both country- and pollster-level differences within four-level Bayesian MCMC models is explored further in Table A12 of Appendix A.

Irrespective of the manner in which variance is apportioned between levels, all election-level ICC values displayed in Table 14 sit substantially above the 5% threshold justifying the investigation of group-level effects. When considered alongside the ICC values calculated from earlier frequentist maximum likelihood models, they demonstrate that election-level differences exist as significant drivers of the variance observed across my continuous measures of polling inaccuracy even when country- and pollster-level differences are controlled for. This lends further support to the hypothesis that elections vary in their ability to be accurately predicted.

To test whether these conclusions hold across my binary measures of polling inaccuracy, I calculate ICC values for them across a range of equivalent four-level models. The first of these uses Laplace approximation and its results are displayed in Table 15. The ICC value associated with SBP indicates that 32% of its variance can be attributed to election-level differences and sits within the range of values calculated from earlier three-level models, agreeing with them substantively.

Table 15: Election-level ICC estimates for binary measures of polling error calculated from four-level models using Laplace Approximation across whole dataset including standard errors (SE) and 95% confidence intervals.

	LVRC	**SBP**
Laplace Approximation		
ICC	0.75	0.31
SE	(0.04)	(0.02)
95% CI	0.68 – 0.82	0.28 – 0.34

LVRC = leading vote recipient correct; SBP = significantly biased poll.

In the case of SBP, the ICC value of 0.31 indicates that election-level differences account for 31% of its observed variance within four-level models. From this, it is clear that the inclusion of the pollster level brings about a small reduction in the observed impact of election-level differences on the variance exhibited by SBP, down from 34% in earlier three-level models. This is in keeping with expected reductions brought about by the inclusion of pollster-level controls, but demonstrates that election-level differences remain a substantial driver of variance in SBP, even within four-level models.

By contrast, while the ICC value of 0.75 associated with LVRC indicates that election-level differences also remain significant determinants of its variance within four-level models, this value remains unchanged from earlier three-level models, suggesting that the inclusion of the pollster level does not alter the proportion of variance in LVRC attributable to election-level differences. However, the elevated standard error and wider confidence interval associated with the ICC estimate for LVRC may suggest a degree of imprecision which is not unprecedented when using Laplace approximation for more intricate estimative tasks. This suspicion extends to the country- and pollster-level values displayed in Table A13 of Appendix A. As I identified earlier in this chapter, it may be that the integral of the four-level model is too complex for a more simplistic estimative procedure such as Laplace approximation, especially in relation to binary outcome variables. Due to this, the validity of the ICC value calculated in relation to LVRC needs to be checked. This can be achieved using Bayesian MCMC modelling.[764]

Table 16 displays ICC values from four-level models estimated using Bayesian MCMC with half student-t priors. These values apply to both LVRC and SBP and can be used to check the validity of those values calculated from models using Laplace approximation. The ICC value calculated for SBP, 0.32, indicates that 32% of its variance can be attributed to election-level differences. This agrees with the estimate calculated from the model using Laplace approximation. However, the ICC estimate for LVRC does not. At 0.68, it sits below the value calculated from Laplace approximation and is more in keeping with the reductions caused by the layering of controls from two- to three-level models. As such, while it agrees with the substantive conclusion of the Laplace model, it presents a more dependable and consistent estimate of the impact of election-level differences on the variance exhibited by LVRC in four-level models.

Table 16: Election-level ICC estimates for binary measures of polling error calculated from additional Bayesian MCMC four-level models across whole dataset including standard errors (SE) and 95% confidence intervals.

	LVRC	**SBP**
Half Student-t Priors		
ICC	0.68	0.32
SE	(0.02)	(0.01)
95% CI	0.64 – 0.71	0.30 – 0.33

LVRC = leading vote recipient correct; SBP = significantly biased poll

[764] Capanu, Gonen, and Begg, p. 1.

To establish whether the findings displayed in Table 16 are robust across alternative model specifications, Table A14 in Appendix A calculates ICC values for my measures of binary polling accuracy using alternative, half-Cauchy priors. These values agree substantively with those calculated from models using half student-t priors, varying by 1% in the case of LVRC and remaining identical in the case of SBP. This demonstrates that the findings displayed in Table 16 are robust across a range of models and are not simply artefacts of a given estimative procedure. Additionally, Table A15 in Appendix A decomposes the impact of the pollster and country levels within four-level Bayesian models, finding that country-level differences are impactful drivers of observed variance in both LVRC and SBP, while pollster-level differences only stand as meaningful drivers of variance in SBP, having little impact on LVRC.

When considered together, the ICC values calculated from my four-level models indicate that the election-level exists as a significant driver of variance across both my continuous and binary measures of polling inaccuracy even when country- and pollster-level differences are controlled for. This lends yet more support to the hypothesis that election-level differences impact upon polling accuracy and, therefore, that elections vary in their ability to be accurately predicted.

Validating a New Ontology: The Importance of Election-level Differences for Polling Error

Within this chapter I have iteratively and robustly demonstrated one key finding: differences between elections matter for polling error. For my continuous measures of error, the amount of variance accounted for by election-level differences ranged from 43% to 75% in two-level models, 18% to 46% in three-level models, and 17% to 39% in four-level models. The decreasing proportion of error accounted for by the election-level as model complexity increases is indicative of the impact of controlling for country-level effects in the case of three-level models and both country- and pollster-level effects in the case of four-level models. This iterative reduction across models indicates that both country- and pollster-level differences stand as important drivers of polling error. However, even when these additional grouping levels are controlled for, election-level differences remain substantial determinants of polling error variance, lending support to the hypothesis that elections vary in their ability to be accurately predicted.

A similar progression was evident across my binary measures of error, with election-level differences accounting for between 38% and 89% of variance when calculated from two-level models, 32% and 77% when calculated from three-level models, and 28% to 71% when calculated from four-level models (excluding unreliable models estimated using Laplace approximation). While the proportion of variance again decreases as additional grouping-levels are controlled for, election-level differences remain substantial determinants of polling error variance even within four-level models. Indeed, election-level differences account for a larger proportion of variance across my binary measures of error than their continuous counterparts, but remain substantial drivers of variance in both cases.

The importance of election-level differences as drivers of polling error remains evident when their effect is broken down across my three conceptualisations of error. For measures of distributive inaccuracy, election-level differences account for between 43% and 70% of variance when calculated across two-level models, 18% to 46% across three-level models, and 24% to 39% across four-level models. When the upper limits of these ranges are considered, this again represents a general decrease as additional grouping-levels are controlled for. Nevertheless, election-level differences account for substantial proportions of the variance observed across my measures of distributive polling error even when these confounding grouping factors are iteratively accounted for.

A similar progression is apparent across my measures of bounded inaccuracy, with the impact of election-level differences ranging from 57% to 75% in the case of two-level models, 18% to 32% in the case of three-level models, and 17% to 28% in the case of four-level models. As before, while the effect of election-level differences was found to be very pronounced in two-level models, it reduced dramatically when country-level differences were controlled for, and fell again when pollster-level differences were controlled for. Nevertheless, election-level differences remained important drivers of error, accounting for proportions of variance considerably above the 5% threshold justifying the interrogation of grouping levels.

Election-level differences have the most pronounced impact on my measures of substantive polling inaccuracy. While the proportion of error variance accounted for by election-level differences is still subject to iterative reduction between models, ranging from 81% to 89% in two-level models, 65% to 77% in three-level models, and 64% to 71% in four-level models, it remains consistently higher than the proportion of variance accounted for by the election-level across other conceptualisations of polling error.

The fact that election-level differences consistently account for substantial proportions of variance across all measures of polling error over a wide range of models controlling for the effect of additional grouping levels, lends robust support to my first hypothesis that polling error varies as a function of the election in which polls are conducted. Moreover, it lends empirical support to the new election-level ontology of polling error that I have put forward within this thesis. The usefulness of this new ontology merits further investigation, most notably into those election-level differences responsible for affecting polling error which I unpack in the following chapter.

Chapter 5 – Which Differences Make a Difference? Identifying Plausible Election-level Predictors of Polling Error Variance

"It is a test of true theories not only to account for but to predict phenomena".[765]

- William Whewell (1840)

That the election in which polls are conducted affects the variation of the error they present speaks to the importance of differences between elections as drivers of variable polling inaccuracy. Though the decomposition of the variance parameters in multi-level models allows for a macroscopic appreciation of the importance of election-level differences as drivers of variance in polling error, it is unable to identify which differences between elections possess the most prominent impact on this variance. In this chapter I identify a series of intuitive election-level variables that can be expected to affect polling error variation and unpack their likely impact across my distributive, bounded, and substantive conceptualisations of polling inaccuracy, as well as polling bias. To enable this, I break the chapter down into three sections.

In the first section, I outline the theory underpinning the expectation that electoral characteristics ought to be predictive of polling error variation. I contend that as differences between elections account for a substantial degree of variation in the error they present, they lend themselves to the prediction of the presence of this variation. To investigate this contention, I identify thirteen election-level characteristics that vary between contests and describe how they can be expected to affect my distributive, bounded, and substantive conceptualisations of polling error, as well as polling bias.

In the second section, I hold that assessing the impact of electoral characteristics in isolation fails to adequately capture their relationship with polling error variance. I hold that several of

[765] William Whewell, The *Philosophy of the Inductive Sciences*, (London: John W. Parker Publishing, 1840), p. 46.

my identified election-level variables can be expected to interact with one another, altering the way they impact variation in polling error. I outline a series of expected two- and three-way interactions and describe their expected impact across my three conceptualisations of polling inaccuracy, along with polling bias.

In the third section, I recognise the importance of controlling for variables within the other grouping levels of sources of polling error. I identify a range of variables at the poll, pollster, and country levels and outline their expected effect on polling error variance. I explain their use as controls in later prediction models and describe the importance of their inclusion to ensure the robustness of election-level findings.

5.1: Identifying Likely Election-level Predictors of Polling Error Variance

The previous chapter demonstrated that membership within different elections exists as a meaningful driver of polling error variation. As elections are compositionally heterogenous, membership within different elections entails membership within differently composed phenomena. That membership within differently composed elections affects error variance speaks to the importance of differences between elections as drivers of this variance. Given that these differences drive error variation, and therefore make it more or less likely, it stands to reason that they can be used to predict its extent.

In this section, I identify thirteen electoral characteristics that can be expected to differ between contests and bear upon polling error variation. The characteristics that I address are the closeness of margin of victory within a contest, the effective number of elective parties (ENEP) contesting an election, ENEP change between elections, whether a contest is a second-round run-off election, whether an election is legislative or presidential, the level of voter turnout, turnout change between elections, differences in the level of strong partisan loyalty, the extent of late decision-making within the electorate, the electoral system, instances of change between electoral systems, whether an election is scheduled or snap, the extent of ideological differences between the parties or candidates contesting an election, and differences in the number of registered voters between elections.

While the thirteen variables addressed in this chapter do not comprise the universe of all possible differences between elections, I contend that they are the most intuitive examples of differences that can be expected to be impactful drivers of variance in polling error. In what follows, I move through each of my thirteen election-level variables and address the manner in which they can be expected to increase or decrease the likelihood of each of my conceptualisations of polling error and therefore affect their variance across elections. I begin by addressing the closeness of the margin of victory within my studied elections.

Closeness of the Margin of Victory

The first intuitive election-level variable that can be expected to affect polling error variation is the closeness of a given electoral contest. The closeness of an election refers to the margin between the vote shares received by the two leading parties or candidates on election day. While an average closeness between leading parties or candidates could be taken from polls throughout an election campaign to measure this margin, such an approach may not be indicative of the true margin between vote shares, as the predictions offered by polls over the course of a campaign may be systematically wrong. As such, the true closeness or marginality of an election can only be known after the fact from electoral returns. Due to this, I measure it as the absolute percentage point difference between the parties or candidates in receipt of the largest and second largest shares of the vote in each of my studied elections.

The expected effect of the closeness of an election on the likelihood of polling error variation differs across my distributive, bounded, and substantive conceptualisations, bearing most closely on substantive polling error. If the race in a given election is close, the difference between the percentage vote shares of the leading parties and candidates will be slim. As such, small errors – even expected errors within the margin of error – are more likely to result in a poll incorrectly predicting the substantive victor of an election. This makes polls more likely to incorrectly predict the party or candidate in receipt of the largest share of the vote, therefore increasing the probability of substantive error.

The closeness of an election can also be expected to effect distributive, bounded, and substantive polling error. This expectation is based on the satisfaction of one necessary condition: that the close margin observed in election returns is indicative of a consistently close race during the campaign and not simply the tightening of vote shares in the closing stages of a campaign due to last-minute shifts in voting intention. Close elections such as this lend themselves to increased substantive polling error, as the slim margin between candidates and parties is such that small polling errors – even those that fall within the margin of error associated with them – are sufficient to cause a poll to identify an incorrect winner. Recognising the heightened potential for substantive error, during such campaigns, pollsters are more likely to devote a greater number of resources to ensuring the accuracy of their polls in an attempt to avoid it, especially given the primacy afforded to substantive accuracy in post-election assessments of polling error conducted by the media. This increased level of resourcing lends itself to decreased distributive polling error, as a greater focus is placed on accurate vote share distributions. This greater focus on distributive accuracy also lends itself to decreased bounded polling error, as polls are less likely to exhibit errors that exceed the bounds of their margins of error.

As the difference between predicted and actual vote shares can be expected to be lower, close elections also decrease the likelihood of bounded inaccuracy, as errors are less likely to be of a magnitude sufficient to exceed stated margins of error. Of course, as shown in the predictive failures addressed in the literature review, it is possible that polls are systematically incorrect and fail to adequately track campaign dynamics. In such elections, polls would exhibit high levels of distributive inaccuracy, increasing the likelihood of bounded inaccuracy. It is also possible that a narrow margin between leading parties or candidates observed after the fact is not indicative of a close campaign, but rather of a radical last-minute narrowing of vote share differences caused by a scandal, voters 'returning home' to historical partisan preferences in the closing days of an election,[766] or a late swing in voting intention.

Despite these possibilities, it is not unreasonable to expect the close margin of victory in an election to be indicative of a consistently contest. Election campaigns often fail to significantly affect voting intention, instead serving to guide events down a path pre-determined by fundamental factors and the priming of existing partisan sentiment within the electorate.[767] While the voting intention of a subset of swing or 'floating' voters may be affected by electoral campaigns,[768] these individuals represent a minority of voters. While these voters will necessarily be sufficient to alter outcomes in closely contested elections, in others they will not be significant in the determination of the outcome.[769] Moreover, partisan loyalty often renders the voting intention of the majority of electorates deterministic.[770] When coupled with the fact that scandals of the scale necessary to radically overhaul these loyalties and alter the course of

[766] Richard Herr, 'Partisan Chickens Coming Home to Roost in the 2010 Tasmanian Election: Consequences of the 1998 Reduction in Size of the Parliament, *Public Administration Today*, 22 (2010), 33 – 35 (p. 33); Michael Henderson, 'Finding the Way Home: The Dynamics of Partisan Support in Presidential Campaigns', *Political Behavior*, 37 (2015), 889 – 910 (p. 889).
[767] Gelman and King, pp. 434 – 435; Bernard Berelson, Paul Lazarsfeld, and William McPhee, *Voting: A Study of Opinion Formation in a Presidential Election*, (Chicago: University of Chicago Press, 1954), pp. 15 – 19.
[768] Kenneth F. Greene, 'Campaign Effects and the Elusive Swing Voter in Modern Machine Politics', *Comparative Political Studies*, 54.1 (2021), 77 – 109 (p. 77).
[769] James E. Campbell, 'Do Swing Voters Swing Elections?', in *The Swing Voter in American Politics*, ed. by William G. Mayer (Washington D.C.: Brookings Institution Press, 2007), p. 130.
[770] Rau, p. 1021.

elections are rare – though not unheard of[771] – the large-scale movement of voters across party lines is not generally likely within any given contest.

For the closeness of the margin of victory to matter for polling error variance across my studied elections, it must meaningfully differ between them. Figure 18 displays the differences in the margin of victory across all 497 elections within my dataset. These differences are broken down over time and across continents.

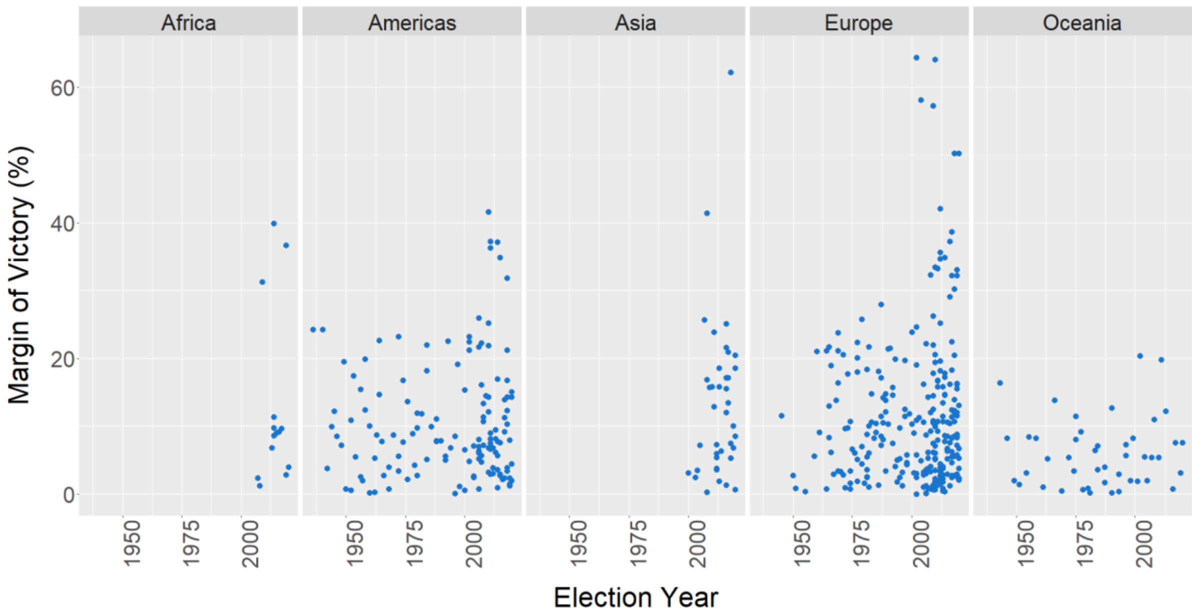

Figure 18: The margin of victory across each of my studied elections from 1936 – 2020, broken down by continent. Each point within the figure represents an individual election.

From Figure 18, it is clear that the margin of victory differs considerably between elections. Though a degree of clustering is present in values associated with elections held after the year 2000 in the Americas and Europe, values are generally well dispersed, indicating consistent differences in the margin of victory. Not only do these differences occur consistently, but they are often substantial in nature. Indeed, the margin of victory in elections conducted across Asia and Europe varies from single digits, representing closely fought elections, to upwards of sixty

[771] Laura Stoker, 'Judging Presidential Character: The Demise of Gary Hart', *Political Behaviour*, 15.2 (1993), 193 – 223 (p. 193); Dennis Halcoussis, Anton D. Lowenberg, and G. Michael Phillips, 'An Empirical Test of the Comey Effect on the 2016 Presidential Election', *Social Science Quarterly*, 101.1 (2019), 161 – 171 (p. 161).

percentage points, indicating landslide victories. Though the range of differences exhibited by elections held in other continents is less extreme, it is nevertheless pronounced. The margin of victory in elections held in Africa and Americas ranges over a span of forty percentage points, while the margins in Oceanian elections range over twenty percentage points. That differences in the margin of victory between elections are consistent and often sizable lends support to their ability to affect polling error variation across these contests.

The Effective Number of Elective Parties

In addition to the margin of victory associated with elections, I include the effective number of elective parties (ENEP) present in each contest as an election-level predictor of polling error variance. ENEP serves as a measure of the number of political parties contesting an election weighted by the size of the shares of vote they receive. As such, it is a measure of the number of impactful parties contesting an election.[772] To calculate ENEP, I use the approach devised by Laakso and Taagepera shown in equation 26, where n represents the number of parties in receipt of at least 1% of the vote in an election and V_{i}^2 represents the square of the normalised percentage vote share received by each party.[773]

$$\mathrm{ENEP} = \frac{1}{\sum_{\mathrm{i}=1}^{n} V_{\mathrm{i}}^2} \tag{26}$$

Within this thesis, I take ENEP to apply parties in the case of my studied legislative elections and candidates in the case of my studied presidential elections. For the purposes of their substitution in equation 26, candidates contest presidential elections as discrete, vote-gaining entities in much the same way as parties contest legislative elections. As such, they are

[772] G. V. Golosov, ‘The Effective Number of Parties: A New Approach’, *Party Politics*, 16.2 (2010), 171 – 192 (p. 171).
[773] Markku Laakso and Rein Taagepera, ‘Effective Number of Parties: A Measure with Application to West Europe’, *Comparative Political Studies*, 1979 (12.1), 3 – 27 (p. 4).

interchangeable within the equation, allowing it to be interpreted identically. Due to this interchangeability, ENEP is commonly used to calculate the effective number of parties in wider polling literature.[774]

The effective number of elective parties (ENEP) contesting an election can be expected to differentially affect the presence of my three conceptualisations of polling error and, therefore, their variation between contests. It can be expected to bear most directly on distributive polling error. The higher the ENEP value associated with an election, the greater the degree to which the overall vote is fragmented, as it is distributed between a greater number of parties or candidates. Given the greater amount of fragmentation, correctly predicting the vote shares received by parties is likely to be more difficult in elections characterised by a higher effective number of parties. As such, high levels of ENEP can be expected to be predictive of distributive polling error. The increased likelihood of distributive inaccuracy also makes instances of bounded inaccuracy more likely, as the propensity of polls to exhibit vote share errors beyond their margins of error is increased.

In addition to distributive and bounded inaccuracy, ENEP can also be expected to be predictive of substantive inaccuracy, especially when considered in concert with the closeness of the contest. In elections contested by a larger number of effective parties, a greater number of actors capable of syphoning meaningful proportions of the vote from leading parties are present. In contests characterised by close competition, this process may be sufficient to alter the party in receipt of the largest share of the vote, thereby affecting the ability of polls to correctly identify them, leading to substantive error. However, in elections not characterised

[774] Mark P. Jones, 'Electoral Laws and the Effective Number of Candidates in Presidential Elections', *The Journal of Politics*, 61.1 (1999), 171 – 184 (p. 175).

by close competition between leading parties and candidates, ENEP is less likely to bring about substantive polling error.

For the ENEP values associated with elections to exist as plausible drivers of polling error variance across my studied elections, they must meaningfully differ between contests. Figure 19 displays the differences in the ENEP values associated with my 497 studied elections. These differences are broken down over time and across continents.

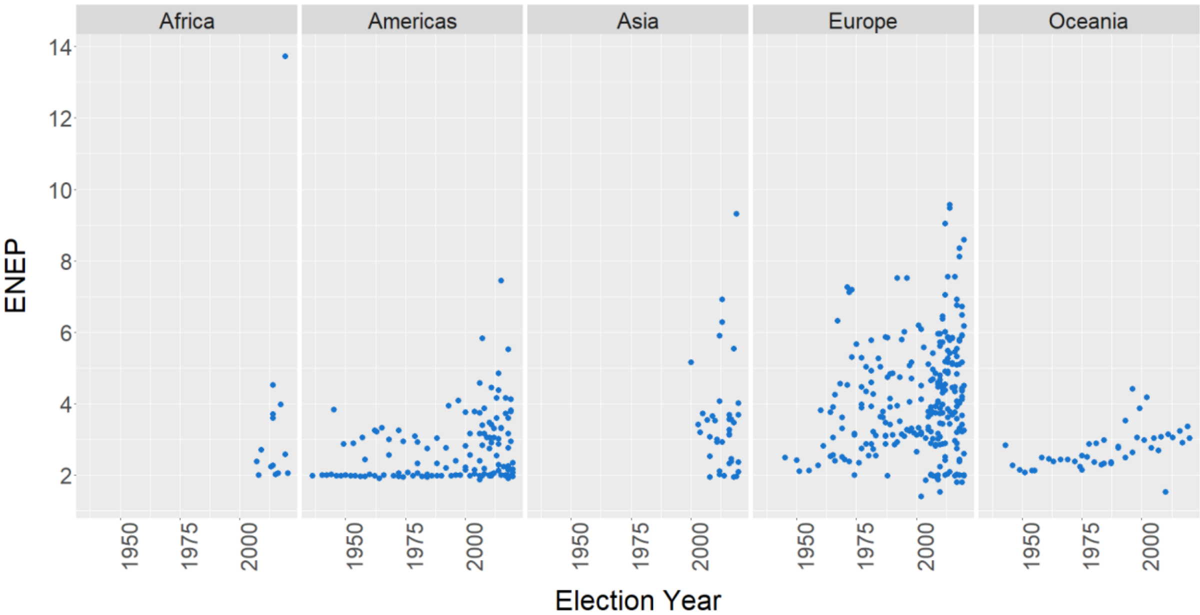

Figure 19: The ENEP values associated with each of my studied elections from 1936 – 2020, broken down by continent. Each point within the figure represents an individual election.

From Figure 19, it is clear that the effective number of elective parties varies considerably across elections. Though continental trends are visible, such as the clustering of consistently low ENEP values in the Americas that results from the bipartisan nature of US elections or the notable increase in Oceanian ENEP values after the adoption of proportional representation in New Zealand in 1993, consistent differences in ENEP between elections are visible across all continental groupings. Even the exclusion of clear outliers within continents, such as the ENEP value of 13.7 presented by the 2019 Tunisian presidential election and the value of 9.3 associated with the 2019 Indonesian legislative election, does not detract from the consistency of inter-election differences in ENEP. Across all continental groupings, the magnitude of differences in ENEP values between elections is also notable. Excluding outliers, ENEP values can be seen to range from ~2 to ~4 in both African and Oceanian elections, 2 to ~7 in the case of elections in the Americas and Asia, and < 2 to ~10 in the case of European elections. That ENEP differs to such an extent between elections lends support to the contention that it exists as a plausible driver of polling error variance across contests.

ENEP Change Between Elections

While differences in the raw number of effective electoral parties contesting an election can be expected to create environments that are variously conducive to polling error, so too can the extent of change in ENEP values between elections. Large changes in the effective number of electoral parties contesting one election to the next create issues for the effective poll-based prediction of vote share distributions. If the number of effective elective parties increases significantly between elections, a greater number of parties or candidates receive a meaningful share of the vote. This indicates that substantial proportions of the electorate altered their voting behaviour between elections, moving to support new, erstwhile under-represented parties or candidates. By contrast, a significant decrease in the number of effective elective parties between elections characterises a polar shift in voting behaviour, with voters moving away from a diverse range of smaller parties and candidates to instead coalesce around fewer, larger parties or candidates.

In both cases, decision-making within the electorate has diverged considerably from past behaviour. This has the potential to confound the likely voter models employed by polling organisations, as past behaviour no longer serves as a reliable indicator of future voting intention. It may also symbolise changing or diminishing partisan loyalties within the electorate, further impacting the degree to which past behaviour is indicative of future actions.

As changes in ENEP values between elections represent disruptive instances of discontinuity in voting behaviour, the extent of these changes between elections bears closely on the ability of polls to accurately predict vote share distributions, as they undermine likely voter models that principally rely on path dependency between past actions and future behaviours.[775]

Changes in ENEP between elections can be expected to affect the prominence of each of my conceptualisations of polling error, though bear most closely upon distributive error. As difficulties surrounding the projection of past voting behaviour onto future voting intention undermine the likely turnout models on which polls rest, they make accurately predicting vote share distributions more difficult, increasing the likelihood of distributive error. An increased likelihood of distributive error also lends itself to greater bounded polling error, as polls are more likely to exhibit errors sufficient to exceed their stated margins of error. If the shift in voters' support for parties or candidates is considerable between elections, represented by large changes in ENEP, polls are at risk of fundamentally misattributing their voting intention on the basis of past behaviour, increasing the likelihood of substantive error.

For changes in ENEP between elections to affect the variation of my measures of polling error, they must be sufficiently large and variable between cases. To investigate whether this is the case, Figure 20 displays the magnitude of changes in ENEP between my 497 studied elections broken down over time and across continents.

[775] G. R. Murray, C. Riley, and A. Scime, 'Pre-election Polling: Identifying Likely Voters Using Iterative Expert Data Mining', *Public Opinion Quarterly*, 73.1 (2009), 159 – 171 (p. 162).

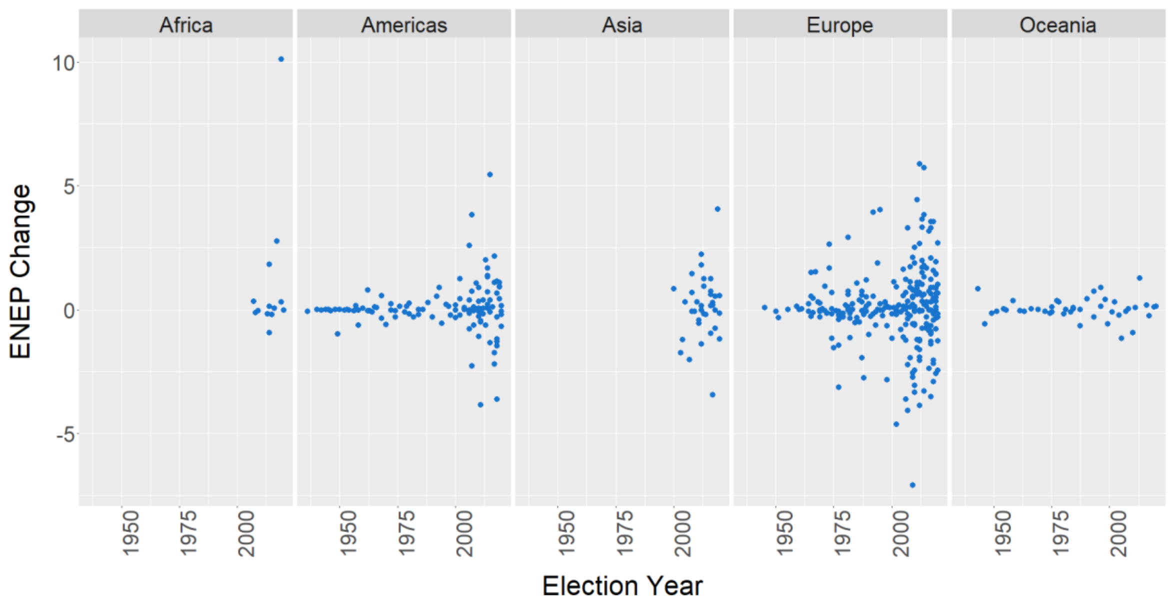

Figure 20: The extent of ENEP change between my studied elections from 1936 – 2020, broken down by continent. Each point within the figure represents an individual election.

From the figure, changes in ENEP between elections are immediately apparent across all five continental groupings and generally increase in magnitude over time. Substantial ENEP changes between elections are more consistently evident over time in Africa, Asia and Europe relative to other continents. While considerable changes in ENEP between elections are evident in the Americas after 1950, elections conducted prior to 1950 exhibit only one instance of significant ENEP change between contests. This is an artefact of the dominance of bipartisan politics in the USA during this period, which was only disrupted by the brief rise of Strom Thurmond as a third-party candidate in 1948.[776] The extreme range of changes in ENEP between elections conducted in Africa is driven by the outlying Tunisian presidential election of 2019. However, even when this outlier is ignored, the changes in ENEP between the remaining elections still present notable variation. Oceanian elections generally display lower levels of ENEP change between elections, with changes increasing after the adoption of proportional representation by New Zealand in 1993. That changes in the ENEP values

[776] Cohodas, p. 5.

associated with elections are consistently evident lends support to their plausibility as drivers of polling error variation.

Second-round Presidential Run-off Elections

The expected impact on polling error variation underlying the effective number of parties can be extended to apply to the type of election under investigation. While the effective number of parties present in presidential and legislative elections are positively correlated when considered cumulatively,[777] this is not the case in presidential elections involving second round run-offs, as they are contested by a reduced pool of candidates.

Second round run-offs occur in presidential elections held under majoritarian or plurality threshold systems when no single candidate receives a majority vote share or a share of the vote over a prescribed margin.[778] The second round of these presidential elections is exclusively contested by those two candidates who received the largest and second largest shares of the vote in the first round. Necessarily, this reduces both the effective and physical number of candidates contesting the election. Given the reduction in the numbers of candidates vying for a share of the vote, the round of presidential elections can be expected to affect the prominence of polling error.

Second round run-off elections can be expected to bear most directly on distributive polling error. On one hand, as the overall vote is less fragmented in second round run-off elections due to the reduced number of candidates, the task of apportioning vote shares is, at least

[777] Gary W. Cox, *Making Votes Count: Strategic Coordination in the World's Electoral Systems*, (New York: Cambridge University Press, 1997), pp. 203 – 222; Mark P. Jones, *Electoral Laws and the Survival of Presidential Democracies* (Notre Dame: University of Notre Dame Press, 1995), pp. 75 – 77; Scott Mainwaring and Matthew Soberg Shugart, 'Conclusion: Presidentialism and the Party System', in *Presidentialism and Democracy in Latin America*, ed. by Scott Mainwaring and Matthew Soberg Shugart (New York: Cambridge University Press, 1997), pp. 394 – 437.
[778] Sarah Birch, 'Two-round Electoral Systems and Democracy', *Comparative Political Studies*, 36.3 (2003), 319 – 344 (p. 321).

theoretically, more straightforward, lending itself to reduced distributive error. This reduction in error surrounding predicted vote shares also makes bounded inaccuracy less likely, as differences between predicted and actual vote shares are less likely to be large enough to exceed the bounds set by the margins of error surrounding polling estimates. However, on the other hand, the reduction in the number of candidates contesting second-round elections presents similar problems to sharp reductions in the effective number of parties between contests. That is, the voting intention of individuals who previously supported candidates that are no longer on the ballot may be more difficult to predict, especially if significant differences exist between the candidates contesting a second-round election and the candidates who are no longer present. This may result in heightened distributive polling error, which lends itself to an increased chance of bounded error.

Not only do second round run-off elections encompass similar issues to significant drops in ENEP between contests, but they also mirror the difficulties presented by changes in turnout levels. This is because voter turnout is typically different in run-off elections relative to first round contests.[779] Differences in turnout between first and second round elections complicate the extent to which vote shares from first round contests can be mapped on to run-offs, leading to the potential for distributive polling error. The issues presented by changes in turnout between first and second round elections is further complicated by the fact that the directionality of these changes is often inconsistent.[780] Given the potential for differences in the direction of turnout changes *between* second round elections, they present the potential to confound the expected turnout models on which the estimated vote share distributions provided by polls rest, again raising the potential for distributive polling error.

[779] Stephen G. Wright, 'Voter Turnout in Runoff Elections', *The Journal of Politics*, 51.2 (1989), 385-396 (p. 385).
[780] Statista, *Voter turnout in the presidential elections in France between 1965 and 2022, by round* (2023), <https://www.statista.com/statistics/1068866/participation-rate-voter-turnout-presidential-elections-france/> [accessed 16/08/2023].

The reduced number of candidates contesting the second round of presidential elections can also be expected to affect the ability of polls to correctly predict the recipient of the largest share of the vote and, therefore, bear upon my substantive conceptualisation of polling error. As second round run offs are contested by the two leadings candidates, polls are faced with fewer prospective largest vote share recipients. This makes the process of correctly identifying the recipient both far more straightforward and likely, as even random selection would yield a success rate of 50%. To this end, polls conducted for second round run offs in presidential elections can be expected to present reduced substantive error.

Table 17 captures the number of second-round presidential run offs within my dataset. Given that only 18 countries in my dataset have presidential elections which incorporate second round run-offs, and these run-offs are not always mandatory if a candidate reaches a vote share threshold in the first round, second round run-offs comprise a minority of the elections within my dataset. Though their effect may be small in the aggregate, as they only comprise ~8% of studied elections, the 39 second round run-offs captured by my dataset make it possible to investigate the expectation that they possess characteristics which affect polling error variation.

Table 17: The number of first- and second-round elections within my dataset along with the percentage of elections within the dataset that they represent.

Election Round	Number	Percentage
Round 1	458	92.1%
Round 2	39	7.9%

Election Type: Legislative vs. Presidential

The type of election being contested can also be expected to impact the degree to which polling error varies. In legislative elections, decision-making within the electorate stabilises sooner than in presidential contests.[781] Given that voting intentions exhibit greater stability over longer time horizons in legislative elections, fluctuations within the electorate that could confound polling predictions are less likely and polls have a greater amount of time to hone their predictions and correctly identify the less dynamic intentions of the electorate. From this, it ought to be easier for polls to correctly predict the vote shares received by parties and candidates in legislative elections. Therefore, distributive, bounded, and substantive error can be expected to be lower in legislative contests than their presidential counterparts.

The presidential or legislative focus of an election also affects the degree to which voters turn out. While presidential elections often occur in strong party systems where capturing the nomination or endorsement of a political party is an important – and sometimes necessary[782] – element of electoral success,[783] they are nevertheless more candidate-focused in nature. By contrast, though candidate evaluations factor into voter decision-making in legislative elections,[784] they are more party-focused in nature than presidential contests. Candidate-focused elections exhibit lower turnout than party-focused elections, even when controlling for contextual differences.[785] Low turnout levels have been widely blamed for polling failures by pollsters and academics alike,[786] with polling organisations facing issues of accurately representing the voting population in those contests in which expected voters fail to turnout to expected levels. As unexpected abstention is more probable in low turnout elections, presidential elections present reduced electorates that are more likely to confound the expected

[781] Jennings and Wlezien, 'Election Polling Errors Across Time and Space', p. 279.
[782] Bowler and Farrell, p. 12.
[783] Pedro C. Magalhães, 'What Are (Semi)Presidential Elections About? A Case Study of the Portuguese 2006 Elections', *Journal of Elections, Public Opinion and Parties*, 17.3 (2007), 263-291 (p. 268).
[784] Katjana Gattermann and Claes H. De Vreese, 'The role of candidate evaluations in the 2014 European Parliament elections: Towards the personalization of voting behaviour?', *European Union Politics*, 18.3 (2017), 447-468 (p. 447).
[785] Peter Söderlund, 'Candidate-centred Electoral Systems and Voter Turnout', *West European Politics*, 40.3 (2017), 516 – 533 (p. 516).
[786] Daoust, p. 739.

voter models employed by polls. In this sense, polling error can be expected to be higher in presidential elections than their legislative counterparts.

Voters behave in less predictable ways in candidate-centric elections than they do in contests centred on parties.[787] In these elections, their voting behaviour is less predictable due to the weakened nature of partisan cues.[788] Therefore, the behaviour of voters can be expected to be more fluid in presidential contests, with decision-making less firmly anchored to past behaviours or political loyalties. This fluidity has the potential to affect the reliability with which past voting behaviour and present voting intention can be projected onto the future, undermining the likely voter projection mechanisms underpinning pre-election polls, and increasing the likelihood that they present distributive error.

Table 18 displays the proportion of legislative and presidential elections within my dataset. Two thirds of the elections it captures are legislative contests, while the remaining third are presidential. Though my dataset contains a greater number of legislative elections, a sufficiently large number of presidential elections is present to ensure their inclusion in the training and testing subsets central to *k*-fold cross validation.

Table 18: The number of legislative and presidential elections contained within my dataset, along with the percentage of all elections that they comprise.

Election Type	Number	Percentage
Legislative	332	66.8%
Presidential	165	33.2%

[787] Sergiu Gherghina and Mihail Chiru, 'Determinants of Legislative Voting Loyalty Under Different Electoral Systems: Evidence from Romania', *International Political Science Review*, 35.5 (2014), 523 – 541 (p. 531).
[788] Ibid.

Differing Levels of Voter Turnout

As was made clear in the literature review, voter turnout levels have regularly been asserted as the cause of past polling failures. Though previous scholarship has not fully elaborated the theoretical mechanisms underpinning this assertion,[789] the importance of turnout levels for polling error variation can be surmised from the role of likely voter models in pre-election polling. In elections characterised by higher levels of turnout, polling organisations are less reliant on the use of likely voter models to estimate the eventual electorate, as a larger proportion of individuals polled ahead of election day go on to vote.[790] Moreover, according to the law of dispersion,[791] in higher turnout elections, participation is more equally distributed across demographic groups, resulting in a more representative electorate.[792] In high turnout systems, such as those that implement compulsory voting laws, the turnout models undergirding polls necessarily require less manipulation as the extent of turnout in any given election is more clear cut. However, high levels of voter turnout may also amplify issues surrounding portions of the population who are difficult to poll. In compulsory voting systems, there will necessarily be groups of citizens within the populace – and therefore within the population of likely voters – who are more difficult to poll than others for reasons such as low response rates and high levels of refusal.[793] These groups will represent a substantial proportion of the electorate who are likely to vote but for whom polls face difficulty in rendering accurate

[789] Daoust, p. 740.
[790] Sohlberg and Branham, p. 3.
[791] Herbert Tingsten, *Political Behaviour: Studies in Electoral Statistics* (London: P.S. King and Son, 1937), p. 7.
[792] Mikael Persson, Maria Solevid and Richard Öhrvall, 'Voter Turnout and Political Equality: Testing the 'Law of Dispersion' in a Swedish Natural Experiment', *Politics*, 33.3 (2013), 172 – 184 (p. 172).
[793] Ronald R. Rindfuss and others, 'Do low survey response rates bias results? Evidence of Japan', *Demographic Research* 32 (2015), 797-828 (p. 797); Emilia Peytcheva and Robert M. Groves, 'Using Variation in Response Rates of Demographic Subgroups as Evidence of Nonresponse Bias in Survey Estimates', *Journal of Official Statistics*, 25.2 (2009), 193-201 (p. 193).

predictions. While demographic and response rate adjustments can be employed to mitigate these issues,[794] widespread issues of nonresponse have proven difficult to resolve.[795]

The use of likely voter models to project turnout levels is problematic as they often introduce artificial volatility into the electorate that is not borne out on election day, resulting in misleading projections.[796] As such, the reduced reliance on the use of likely voter models in higher turnout elections can be expected to benefit the representativeness of polling estimates, reducing the likelihood of distributive error. This in turn reduces the likelihood of bounded error, as polls are less likely to exhibit sufficient error to exceed their stated margins of error. However, it is important to recognise the potential for high turnout to exaggerate issues that result from low response rates and high response refusals, leading to large proportions of likely voters whose voting intention is difficult to predict, increasing the potential for the vote share predictions provided by polls to exhibit distributive error and thereby increasing the likelihood of bounded inaccuracy.

By contrast, in elections characterised by lower levels of turnout, polling organisations are increasingly reliant on the use of likely voter models to identify the eventual population of voters. Despite their issues, the use of likely voter models in such elections is nevertheless more useful than disregarding them altogether.[797] However, the vagaries of likely voter models, along with their potentially unrepresentative insistence on path dependence between contests are such that increased reliance upon them raises the potential for misleading polling estimates and therefore increased distributive error. This also increases the probability of bounded polling

[794] Eric L. Dey, 'Working with Low Survey Response Rates: The Efficacy of Weighting Adjustments', *Research in Higher Education*, 38 (1997), 215-227 (p. 215).
[795] Clifford F. Thies, 'Polls and elections: The Chicago record poll and the election of 1896', *Presidential Studies Quarterly*, 48.1 (2018), 127-138 (p. 128).
[796] Robert Erikson, Costas Panagopoulos, and Christopher Wlezien, 'Likely (and Unlikely) Voters and the Assessment of Campaign Dynamics', *Public Opinion Quarterly*, 68.4 (2004), 588 – 601 (p. 588).
[797] Desart and Holbrook, p. 435.

error, as increased distributive error increases the likelihood of polls exhibiting levels of error sufficient to exceed the bounds of their margins of error.

For differences in turnout levels between elections to affect polling error variation, they must be sufficiently large and variable between cases. Figure 21 displays the magnitude of differences in turnout across my 497 studied elections broken down over time and across continents.

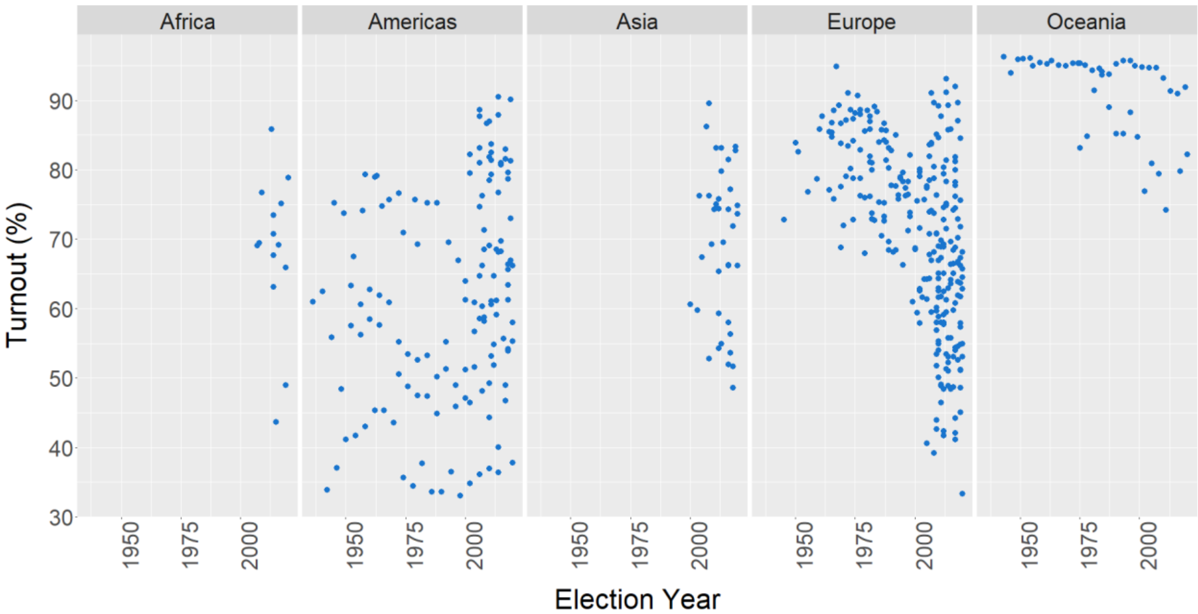

Figure 21: The levels of voter turnout across my studied elections from 1936 – 2020, broken down by continent. Each point within the figure represents an individual election.

From Figure 21, it is clear that levels of voter turnout vary considerably between elections over time across all five continental groupings. Differences in turnout are diminished between elections conducted in Oceania principally due to the compulsory voting laws within Australia.[798] However, even in the presence of these laws, levels of voter turnout can still be seen to vary between elections. Variance in voter turnout between elections is far more considerable in elections conducted across other continental groupings. While variance in turnout between elections is consistently high over time in Africa, the Americas, and Asia, it

[798] Louth and Hill, p. 25.

increases considerably in Europe after the year 2000. This exists as an artefact of the dissolution of the Soviet Union and subsequent democratisation of much of Eastern Europe which lead to pre-election polling being conducted in a greater number of countries, many of which exhibited and continue to exhibit volatility in political engagement.[799] That levels of voter turnout vary consistently between my studied elections speaks to its plausibility as a driver of polling error variance.

The Extent of Turnout Change Between Elections

While differences in the raw level of voter turnout can be expected to establish electoral environments that are more or less conducive to polling error, its effect has rarely been evidenced by analysts,[800] and its impact has been questioned in relation to error concerning the margin between leading parties.[801] Rather than focusing on the differences in the absolute level of turnout in a given election, the importance of voter turnout for polling error variance can be framed in terms of the extent of turnout change between elections. I calculate the change in turnout between contests as the signed percentage difference between successive elections of the same type within the same country. The likely effect of this change on polling error can be understood in terms of the likely voter models and weighting strategies employed by polling organisations. As individuals often misreport their intention to vote,[802] likely voters are often identified using composite survey indices. Amongst other indicators, such as interest in politics and knowledge of voting locations,[803] likely voter models largely rest on information regarding

[799] Derek S. Hutcheson and Elena A. Korosteleva, 'Patterns of Participation in Post-Soviet Politics', *Comparative European Politics*, 4 (2006), 23 – 46 (p. 23).
[800] Mellon and Prosser, p. 663.
[801] Daoust, p. 740.
[802] S. Ansolabehere and E. Hersh, 'Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate', *Political Analysis*, 20.4 (2012), 437 – 459 (p. 440); R. Bernstein, A. Chadha, and R. Montjoy, 'Overreporting Voting: Why it Happens and Why it Matters', *Public Opinion Quarterly*, 65.1 (2001), 22 – 44 (p. 23).
[803] P. Freedman and K. Goldstein, 'Building a Probable Electorate from Pre-election Polls: A Two-stage Approach', *Public Opinion Quarterly*, 60.4 (1996), 574 – 587 (p. 577); Rentsch, Schaffner, and Gross, (p. 786).

the past voting behaviour of individuals.[804] This assumption renders these models susceptible to changes in levels of turnout between contests, especially when these changes are sizeable, as they undermine the extent to which past voting behaviour can be accurately mapped on to future electoral outcomes.

Using past voting history as an indicator of future behaviour is predicated on the assumption of continuity between cases. If this continuity is not present, and voters either fail to turn out as readily as they did or alternatively turn out in far greater numbers than expected, likely voter models can be confounded.[805] Given that the successful identification of likely voters and, by extension, the eventual voting population on election day is integral to successfully predicting election returns, factors that increase the difficulty of achieving this can be expected to increase the likelihood of each of my measures polling error.

The degree to which turnout change can be expected to bear upon my measures of polling error is contingent on the magnitude and composition of the changes observed between elections. Large-scale changes in the extent of turnout between elections can be expected to bear most closely upon substantive polling error. Significant shifts in turnout patterns between elections have the potential to severely undermine likely voter models, such that projections based on past behaviour no longer provide useful guides to future voting intention. If these shifts represent marked departures from the demographic composition of past electorates, then they also have the potential to undermine the weighting procedures employed by polling organisations to ensure that the responses drawn from a sample of individuals by polls reflect the likely voting population. Such shifts have the potential to confound polling estimates and result in widespread substantive error.

[804] Murray, Riley, and Scime, p. 162.
[805] Kennedy and others, p. 5.

Small changes in turnout between elections also have the potential to bear on substantive polling error, especially if they represent instances in which historically apathetic demographics turn out to an unexpected degree, while erstwhile engaged demographics turn out to a reduced extent. This was made apparent in the 2016 US presidential election when a pronounced change in the composition of turnout was sufficient to undermine weighting procedures based on education,[806] resulting in widespread substantive error despite a relatively small change in turnout from the preceding contest.

Large-scale changes in turnout between elections also bear directly on measures of polling bias. Severe unforeseen shifts in turnout patterns between contests have the potential to result in the systematic over- or under-estimation of candidates or parties by polls, especially in the presence of herding in the polling industry. This was again made apparent in the 2016 US presidential election, where differences in the nature and composition of voter turnout from the 2012 contest, along with polls herding around the conventional wisdom that Hillary Clinton would win, resulted in the systematic underestimation of the share of the vote received by Donald Trump.[807]

By contrast, less severe changes in turnout between elections are more likely to bear upon distributive polling error. While small changes in the level and composition of voter turnout are unlikely to be sufficient to bring about widespread substantive misprediction in elections that are not characterised by extremely close margins, they are likely to impact the accuracy of the likely turnout projection mechanisms underpinning polls, increasing their probability of presenting distributive error. In turn, increased levels of distributive error make instances of

[806] Ibid., pp. 4 – 5.
[807] Ibid.

bounded inaccuracy more likely, as polls present greater levels of error, increasing the likelihood of breaching the bounds set by their margins of error.

For the expectations concerning the impact of turnout change on my measures of polling error to be borne out, the instances of turnout change between my studied elections must be suitably large and varied. To determine whether this is the case, Figure 22 displays the extent of turnout change between subsequent elections held in the same country from 1936 – 2020.

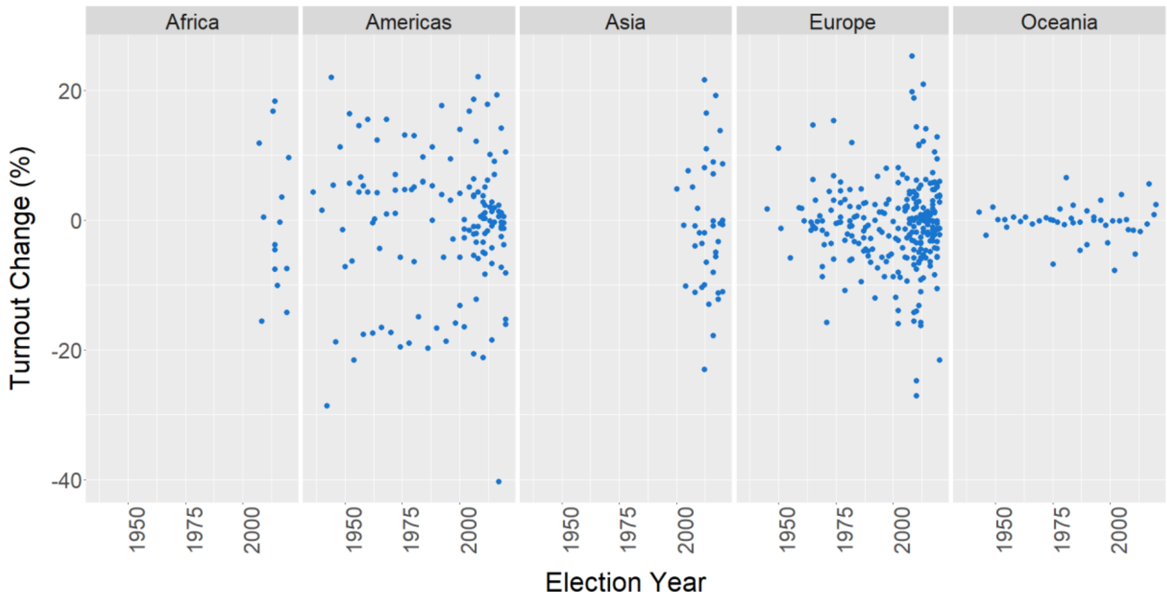

Figure 22: The change in turnout between subsequent elections from 1936 – 2020 broken down by continent. Each point within the figure represents an individual election.

From Figure 22, it is clear that the extent of turnout change between subsequent elections varies over time across all five continental groupings. Changes in the extent of turnout between elections present similar levels of dispersion across the Americas, Asia, and Europe, barring the presence of one notable outlier. This indicates that differences in turnout vary consistently and often considerably between elections conduced within these continents. Dispersion is lower in the case of elections held in Oceania, principally due to the compulsory voting laws present in Australia.[808] However, despite this lower dispersion, changes in turnout nevertheless vary to a notable extent between Oceanian elections. That changes in turnout between elections vary consistently between elections lends plausibility to their ability to exist as drivers of polling error variance across cases.

The Strength of Partisan Loyalty Within the Electorate

Across all forms of election, the strength of partisan loyalty among the electorate is an ever-present and changeable factor. In elections characterised by stronger partisan loyalty, the voting intention of the electorate can be expected to be more deterministic insofar as a greater proportion of voters will reliably vote for the party or candidate to which they feel loyal. Even if voters espouse indecision or dissention when polled, they are nevertheless more likely to 'come home', that is vote in-line with partisan expectations, if they identify as having strong loyalties,[809] especially if exposed to events that activate this loyalty during the campaign.[810] Therefore, it ought to be easier for polls to correctly predict vote shares in elections characterised by strong partisan loyalty, reducing the likelihood of distributive, bounded, and substantive error.

Figure 23 displays the strength of partisan loyalty in each of my studied elections for which data was available. I measure the presence of strong partisan loyalty within my studied elections as the proportion of respondents to pre-election surveys who state that they either fell very close to or feel very strong support for a given party. As its impact on my measures of polling error is based on its relative presence or absence in a given election, for its expected effect to be present within my data, its levels must vary between the elections it contains.

[808] Louth and Hill, p. 25.
[809] Henderson, p. 889.
[810] D. Sunshine Hillygus and Simon Jackman, 'Voter Decision Making in Election 2000: Campaign Effects, Partisan Activation, and the Clinton Legacy', *American Journal of Political Science*, 47.4 (2003), 583 – 596 (p. 583).

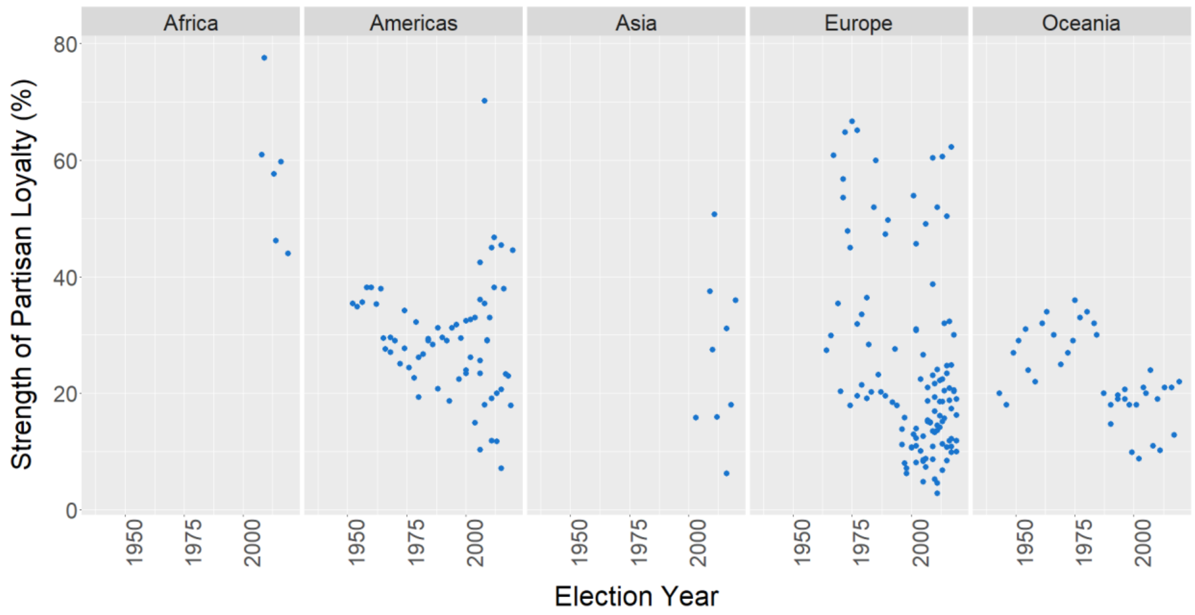

Figure 23: The strength of partisan loyalty within the electorate across elections 1950 – 2020. Each point in the figure represents an individual election.

From Figure 23, the extent of partisan loyalty across my studied elections can be seen to vary from nearly 80% of the electorate feeling strong loyalty to a given party or candidate, to less than 10%. Between these extremes, the strength of partisan loyalty varies considerably and consistently on an election-by-election basis across elections held across all studied areas. That the strength of partisan loyalty varies across my studied elections lends plausibility to its ability to serve as a driver of polling error variance across cases.

The Degree of Late Decision-making Within the Electorate

The relative presence or absence of strong partisan loyalties within the electorate is related to another changeable election characteristic: late decision-making. In elections characterised by stronger partisanship, a greater proportion of the electorate can be expected to settle on their intended voting behaviour sooner in the campaign, as their votes are rendered largely deterministic by their stated loyalty to a given party, hastening their decision-making. By contrast, elections involving weaker partisan loyalties are more likely to contain greater levels of late decision-making, as individuals must settle on their intended voting behaviour in the absence of partisan heuristics.

Late decision-making within the electorate can be expected to affect my measures of polling error. Blamed for many past predictive failures,[811] sudden, last-minute shifts in voting intention have the capacity to not only confound the likely voter models employed by polling organisations, leading to increased distributive error, but to entirely upend the trends observed over the course of a campaign, resulting in unforeseen election results and bringing about large-scale substantive error.

Pre-election polls are likely to struggle to capture last-minute shifts in voting intention for two key reasons. The first is that pre-election polls are lagged representations and projections of public sentiment. They take place over set fieldwork periods, typically ranging from a matter of days to over a week,[812] which undermines their ability to rapidly capture and represent sudden shifts in sentiment, especially during the final days of a campaign. The second concerns the presence of polling moratoriums. Certain countries enforce bans on pre-election polling for a set number of days prior to an election.[813] This prevents polls from being conducted and released during this period, preventing them from capturing last minute changes in public sentiment. I address the presence of moratoriums later when focusing on country-level predictor variables.

While late decision-making bears most closely on measures of substantive polling error, it can also be expected to affect measures of bias. Last minute shifts in voting intention, especially if substantial and unidirectional increase the likelihood of systematic over- or under-estimation

[811] Robinson, p. 141; Wright, p. 41.
[812] Callegaro and Gasperoni, p. 155; FiveThirtyEight, *Latest Polls* (2022), <https://projects.fivethirtyeight.com/polls/president-general/> [accessed 15/03/2022]; YouGov, *Politics & Current Affairs* (2022), <https://yougov.co.uk/topics/politics/survey-results> [accessed 15/03/2022].
[813] Bale, p. 15.

of vote shares on the part of polls, especially in the presence of herding. These systematic errors lend themselves to perceptions of bias and will affect measures associated with it.

I measure late decision-making within an election as the proportion of individuals who stated that their voting intention crystallised on election day or late in the election campaign when responding to post-election surveys. As its measurement relies on post-election surveys, the extent of late decision-making in an election can only be known after the fact. To identify the proportion of late decision-makers in my studied elections, I consulted the global election surveys conducted by the Comparative Study of Electoral Systems (CSES) given their uniformity and comparability.[814] However, the CSES surveys only span the period 1996 – 2021 and therefore do not provide data for my studied elections that sit outside of this timeframe.

To remedy the limitations of the CSES surveys and gather data on the extent of late decision-making amongst the electorate in my studied elections that sit outside of its scope, I gathered additional data from a range of sources. These encompassed regional political surveys, including the Afro Barometer,[815] Asian Barometer,[816] Latino Barometer,[817] and Global Barometer surveys.[818] I also consulted individual election studies conducted by the Making Electoral Democracy Work project to further capture the extent of late decision-making amongst the electorate in those elections and countries not contained within the CSES data.[819]

[814] Comparative Study of Electoral Systems, *Download Data and Documentation* (2021), <https://cses.org/data-download/download-data-documentation/> [accessed 05/03/2022].
[815] Afro Barometer, *Data* (2022), <https://afrobarometer.org/data> [accessed 05/03/2022].
[816] Asian Barometer, *Data Release* (2022), <http://www.asianbarometer.org/data/data-release> [accessed 05/03/2022].
[817] Latino Barometer, *Data* (2022), <https://www.latinobarometro.org/latContents.jsp> [accessed 05/03/2022].
[818] Global Barometer Surveys, *Surveys* (2022), <https://www.globalbarometer.net/survey_do> [accessed 05/03/2022].
[819] Damien Bol and others, *MEDW 2014 Belgian National Election Study* (2017), <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/7GA3IT> [accessed 05/03/2022]; Ignacio Lago and others, *MEDW 2016 Spanish National Election Study* (2017), <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/XHBLOT> [accessed 05/03/2022].

In addition to these studies, I also drew on national election studies to ensure that I was able to gather late decision-making data for as many of my studied elections as possible.[820] Ultimately, through this process, I obtained late decision-making data for 97 elections in 15 countries. I consulted the same sources to identify the extent of strong partisan loyalty amongst the electorates in my studied elections, obtaining values for 257 elections across 50 countries.

While pre-election surveys containing questions concerning the presence of partisan loyalties within the electorate are relatively common, post-election survey containing questions probing the timing of respondents' decision-making are far less widespread. Therefore, while I was able to gather considerable data regarding partisanship, the data I was able to collate on decision-making is, by comparison, considerably reduced, limited primarily to those countries with well-established national election studies. To the best of my knowledge, data pertaining

[820] Australian Election Study, *Interactive Charts* (2020), <https://australianelectionstudy.org/interactive-charts/> [accessed 05/03/2022]; Queen's University, *Canadian Opinion Research Archive* (2022), <https://www.queensu.ca/cora/data-holdings> [accessed 05/03/2022]; AUTNES, *Austrian National Election Study* (2019), <https://www.autnes.at/en/page/3/?a=register&c=index&siteid=1> [accessed 05/03/022]; Mark Swyngedouw, *Belgian General Election Study 2007* (2008), <https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:64978> [accessed 05/03/2022]; Rigsarkivet, *The Danish Election Survey 1971 – 2019* (2022), <https://www.sa.dk/en/the-danish-election-survey-1971-2019/> [accessed 05/03/2022]; German Longitudinal Election Study, *Download Data* (2022), <https://gles-en.eu/download-data/> [accessed 05/03/2022]; T. W. G. van der Meer, H. van der Kolk and R. Rekker, *Dutch Parliamentary Election Study 2017 (DPES/NKO 2017)* (2017), <https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:101156> [accessed 05/03/2022]; New Zealand Election Study, *About the New Zealand Election Study* (2020), <http://www.nzes.org/> [accessed 05/03/2022]; Sami Borg and Kimmo Grönlund, *FSD2653 Finnish National Election Study 2011* (2011), <https://services.fsd.tuni.fi/catalogue/FSD2653?lang=en&study_language=en> [accessed 05/03/2022]; SSJDA Direct, *Social Science Japan Data Archive* (2022), <https://ssjda.iss.u-tokyo.ac.jp/Direct/> [accessed 05/03/2022]; Norwegian Center for Research Data, *National Election Surveys* (2022), <https://www.nsd.no/nsddata/serier/norske_valgundersokelser_eng.html> [accessed 05/03/2022]; South African National Election Study, *Comparative National Elections Project, South Africa 2015* (2020), <https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/603> [accessed 05/03/2022]; Swedish National Data Service, *Swedish National Election Studies (SNES)* (2022), <https://snd.gu.se/en/catalogue/collection/snes> [accessed 05/03/2022]; British Election Study, *Data* (2022), <https://www.britishelectionstudy.com/data/> [accessed 05/03/2022]; American National Election Studies, *Data Center* (2022), <https://electionstudies.org/data-center/> [accessed 05/03/2022]; Italian National Election Studies, *Data Request* (2022), <http://www.itanes.org/dati/#> [accessed 05/03/2022]; SWISSUbase, *Swiss Election Study (Selects) 2019* (2022), <https://www.swissubase.ch/en/catalogue/studies/13846/16586/overview> [accessed 05/03/2022]; Rafael Oganesyan, *Back to the Future? The 2018 Parliamentary Elections and the Armenian Voter* (2018), <https://evnreport.com/elections/back-to-the-future-the-2018-parliamentary-elections-and-the-armenian-voter/?fbclid=IwAR0xmG4C82glJpoDUe6ltAvC9eP54_8T4bLoSq6CluaxpJW7WQOx6QYHSY8> [accessed 05/03/2022].

to the remaining elections and countries either does not exist or is not available to researchers, even by request. Even with these limitations, to the best of my knowledge this thesis contains the most complete account of levels of partisanship and late decision-making across global elections currently present within scholarship.

For late decision-making amongst the electorate to stand as a plausible predictor of polling error variance, it must vary between cases. Figure 24 displays the extent of late decision-making in elections from 1950 – 2020 for all elections for which data was available broken down by continent. Each point in the figure represents an individual election.

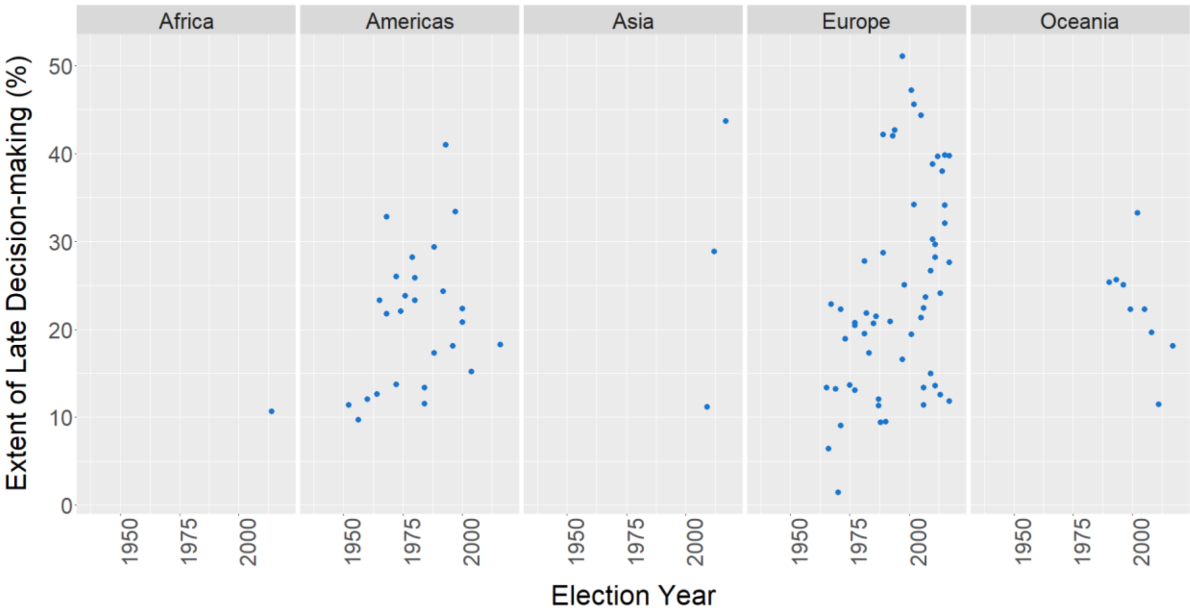

Figure 24: The extent of late decision-making in elections from 1950 - 2020. Each point represents an individual election.

From Figure 24, it is clear that the extent of late decision-making within the electorate varies consistently and often considerably between elections conducted in the Americas, Asia, Europe, and Oceania. As data on late decision-making in Africa was only available for one election, I am unable to comment on the nature of its variance between cases. That the extent of late decision-making varies notably between elections conducted in all other continental groupings lends plausibility to its ability to drive variance in polling error between cases.

Changes in Electoral System Between Contests

While the type of electoral system under which elections are held is a country-level consideration, as it principally differs between countries rather than individual elections, instances of electoral system change and their effects exist as temporally isolated, election-specific events. For polling organisations operating within a country, continuity in electoral system between contests allows approaches to prediction to be refined over time. Lessons learned in previous contests can be directly applied to future contests, as the rules of the game will not have fundamentally changed. Given the cycle of predictive post-mortems conducted in the wake of elections outlined in the literature review, it could be argued that polling organisations are reliant on this continuity for the calibration of their methods. A change in electoral system would profoundly alter the context in which an election is conducted, altering the relative prominence or absence of characteristics such as partisanship and vote fragmentation. Significant changes to these factors would undermine the utility of previous best practices established under the old system and would also reduce the ability of polling organisations to rely on measures of past voting behaviour, as this behaviour would exist as an artefact of the nature of the previous system. To this end, changes in electoral systems between elections can be expected to make it more difficult for polling organisations to accurately predict outcomes, as they will need to adjust to the new systemic reality.

Changes in electoral system between elections can be expected to affect measures of distributive polling error due to its potential to affect the fragmentation of the vote. The number of parties contesting an election is, at least partially, a function of the electoral system in which it is taking place.[821] Certain electoral systems, such as those based on single member district

[821] Rein Taagepera, 'The Number of Parties as a Function of Heterogeneity and Electoral System', *Comparative Political Studies*, 32.5 (1999), 531 – 548 (p. 531); Robert G. Moser, 'Electoral Systems and the Number of Parties in Post-Communist States', *World Politics*, 51.3 (1999), 359 – 384 (p. 359).

plurality rules, permit the existence of fewer effective parties.[822] By contrast, proportionally representative electoral systems with multi-member districts permit a greater number of effective parties.[823] Differences in the number of effective parties within electoral systems necessitate a differing number of vote share predictions on the part of polls. A larger number of predictions provides a greater number of opportunities to present distributive error, while a smaller number of predictions provides fewer, making distributive error less likely.

Beyond vote fragmentation, changes in electoral system between contests have the potential to profoundly impact voter behaviour. Most notably, different electoral systems differentially affect the likelihood of tactical voting.[824] Given the changes in voter behaviour that they facilitate, differences in electoral systems between contests stand to undermine the ability of polls to base vote share estimates on past voting behaviour. This in turn undermines the likely voter models on which polling estimates rests, increasing the likelihood of distributive error, as voters can no longer be relied upon to act and vote as they once did.

A change in electoral system between contests can also be expected be bear upon substantive polling error. As different electoral systems encompass differing approaches to transforming vote shares into seat shares in legislative elections,[825] or establishing the relationship between popular vote shares and electoral success in presidential elections,[826] they fundamentally alter the criteria by which substantive electoral outcomes are determined. In transitioning from contests conducted in one electoral system to another, pre-election polls must therefore adjust to these new processes. Adjusting from one process of votes-to-seats transformation, or one

[822] Rein Taagepera and Matthew Soberg Shugart, *Seats and Votes: The Effects and Determinants of Electoral Systems*, (New Haven: Yale University Press, 1989), pp. 84 – 85.
[823] Ibid., p. 87.
[824] Michael Gallagher, 'Electoral Systems and Voting Behaviour', in *Development in West European Politics*, ed. by Martin Rhodes, Paul Heywood, and Vincent Wright (New York: St. Martin's Press, 1997), p. 114.
[825] Nils-Christian Bormann and Matt Golder, 'Democratic Electoral Systems Around the World, 1946 – 2011', *Electoral Studies*, 32 (2013), 360 – 369 (pp. 361 – 363).
[826] Ibid., pp. 367 – 368.

understanding of the relationship between vote shares and success in executive elections, to another may increase the likelihood of substantive error, as past methods require revision to suit the new electoral landscape. As these methods will not have been tested in the context of the new electoral system in a given country, they present the potential for large-scale predictive error due to a lack of calibration over time, increasing the likelihood of substantive misprediction.

As changes in electoral system are typically accompanied by changes in the behaviour of voters,[827] pre-election polls are not only faced with the challenge of adapting to new approaches to transforming raw votes into substantive political outcomes, but are less able to rely on projection mechanisms that rely on the behaviour of voters in past elections, such as likely turnout models.[828] Even if previous patterns of voting behaviour remain the same after a change in electoral system, allowing polls to rely on past methods of projecting voting intention, they are likely to result in notably different substantive outcomes due to the existence of new rules governing their transformation.[829] Given the inability to reliably project past behaviour onto future voting intention and the fundamental change in the mechanisms governing the transformation of votes into political outcomes, pre-election polls can be expected to be more likely to present substantive prediction error in instances of electoral system change.

As the expected effect of electoral system change on polling error hinges on the difficulties of adjusting to a new reality, I capture instances of system change between successive elections within the same country. Importantly, I take electoral system change to refer to instances of change from one democratic electoral system to another, as opposed to instances in which countries transition away from non-democratic political regimes towards democratic electoral

[827] Gallagher, p. 114.
[828] Rentsch, Schaffner, and Gross, p. 786.
[829] Ibid.

systems. As such, I do not include instances in which autocratic regimes with little or no mechanisms for popular voting transition into democracies, such as the transition experienced by states in Eastern Europe after the collapse of the Soviet Union or states in Northern Africa after the Arab Spring. While these events undoubtedly constitute changes in electoral system, elections held within autocratic states are rarely meaningfully, or trustworthily, contested, and pre-election polling is often not conducted. Given this, little reliable data is available to gauge the impact of such changes on polling error, rendering their inclusion of little use.

A range of instances of electoral system change occurred from 1936 to 2020. My dataset captures those changes in electoral systems that occur between elections in which meaningful pre-election polling takes places and, therefore, can reasonably be expected to bear on the efficacy of this polling. While other instances of electoral system change exist within my studied time period, these are variously unsuitable or inapplicable for inclusion within my dataset. In the case of electoral changes such as the 1962 shift in France away from the indirect election of presidents via an electoral college to their direct election by popular vote,[830] as well as those catalysed by the establishment of new democracies in post-Soviet states after the collapse of the Soviet Union and its satellite states, as well as the emergence of newly-democratised states in Africa and Latin America, these changes are unsuitable for inclusion within my dataset as they did not occur between direct elections that were subjected to meaningful polling. Rather, they represent the beginning of direct elections that can be meaningfully polled and therefore cannot be said to bear on polling error between contests. As such, they cannot reasonably be included as predictors of polling inaccuracy.

[830] Christine Fauvelle-Aymar, Michael S. Lewis-Beck, and Richard Nadeau, 'French Electoral Reform and the Abstention Rate', *Parliamentary Affairs*, 64.1 (2011), 45-60 (p. 45).

By contrast, other instances of electoral change, such as the 1959 overhaul of the Icelandic electoral system,[831] hypothetically lend themselves to occurring between direct democratic elections capable of being polled, but do not bear on polls captured by my data set. While a range of these cases exist,[832] they often occurred at times in which polling had not spread to the states in questions or apply to cases for which polling data is not readily available. As such, these instances of electoral change are inapplicable for inclusion as predictors of polling error.

In total, my dataset captures 20 instances of election system change that occurred between elections for which I have polling data. These changes occurred across a wide range of countries across my studied time period. France briefly changed from a majoritarian two-round system to proportional representation in 1986,[833] only to return to the two-round system in 1988.[834] Italy too has engaged in several instances of electoral system change. In 1993, it underwent electoral system change moving away from proportional representation towards a majoritarian-leaning mixed member electoral system.[835] Italy engaged in electoral reform again in 2005, moving away from its mixed member system to an adjusted form of proportional representation;[836] and again in 2015, after the Constitutional Court ruled that adjustments to the previously implemented proportional system were unconstitutional, leading to its

[831] The New York Times, *Iceland to Adopt Electoral Reform* (1959), <https://www.nytimes.com/1959/04/12/archives/iceland-to-adopt-electoral-reform.html> [accessed 24/08/2023].

[832] Alan Renwick, 'Electoral Reform in Europe since 1945', *West European Politics*, 34.3 (2011), 456-477 (pp. 466-469).

[833] Andrew Knapp, 'Proportional but Bipolar: France's Electoral System in 1986', *West European Politics*, 10.1 (1987), 89-114 (p. 89).

[834] Fauvelle-Aymar, Lewis-Beck, and Nadeau, p. 45.

[835] Richard S. Katz, 'Reforming the Italian Electoral Law, 1993', in *Mixed-Member Electoral Systems: The Best of Both Worlds*, ed. by Matthew Soberg Shugart and Martin P. Wattenberg, (Oxford: Oxford University Press, 2003), 96-112 (p. 96).

[836] Alan Renwick, Chris Hanretty, and David Hine, 'Partisan Self-interest and Electoral Reform: The New Italian Electoral Law of 2005', *Electoral Studies*, 28.3 (2009), 437-447 (p. 437).

replacement with de-facto proportional representation;[837] and *again* in 2017 when this proportionally representative system was replaced by a mixed electoral system.[838]

Further electoral system change was evident across the world during this timeframe. In 1994, the electoral system in Japan changed from centring on the single non-transferable vote and become a mixed member majoritarian system.[839] Two years later, New Zealand adopted a new electoral system in its 1996 legislative election, following its transition from the use of single member district plurality rules to mixed member proportional representation.[840] In 2008, Romania briefly adopted a parallel voting system before returning to proportional representation in 2012.[841] In the same year, Taiwan changed from an electoral system centring on the single non-transferable vote to the use of parallel voting.[842] Hungary also changed its electoral system in 2011, altering the nature of the mixed-member electoral system originally implemented in 1989 to prioritise single-member districts.[843] Electoral system change was also evident in Mongolia in its 2012 legislative election, following its change from a plurality system to a mixed system,[844] and again in 2016 following its change from a mixed system back to a plurality-based system.[845] At the same time, electoral change was evident in Greece in the

[837] Alessandro Chiaramonte, 'The Unfinished Story of Electoral Reforms in Italy', *Contemporary Italian Politics*, 7.1 (2015), 10-26 (p. 10).
[838] Alessandro Chiaramonte and Roberto D'Alimonte, 'The New Italian Electoral System and its Effects on Strategic Coordination and Disproportionality', *Italian Political Science*, 13.1 (2018), 8-18 (p. 8).
[839] Steven R. Reed and Michael F. Thies, 'The Causes of Electoral Reform in Japan', in *Mixed-Member Electoral Systems: The Best of Both Worlds*, ed. by Matthew Soberg Shugart and Martin P. Wattenberg, (Oxford: Oxford University Press, 2003), 152-172 (p. 152).
[840] Jack Vowles, 'Electoral System Change, Generations, Competitiveness and Turnout in New Zealand, 1963 – 2005', *British Journal of Political Science*, 40.4 (2010), 875 – 895 (p. 875).
[841] Aurelian Giugal and others, 'Reforming an Electoral System – An Experiment That Failed: Romania 2008-2012', *Representation*, 56.1 (2020), 111-126 (p. 111).
[842] Hans Stockton, 'How Rules Matter: Electoral Reform in Taiwan', *Social Science Quarterly*, 91.1 (2010), 21-41 (p. 21).
[843] Zoltán Kovács and Gyorgy Vida, 'Geography of the New Electoral System and Changing Voting Patterns in Hungary', *Acta Geobalcanica*, 1.2 (2015), 55-64 (p. 58).
[844] Li Narangoa, 'Mongolia in 2011: Resources Bring Friends and Wealth', *Asian Survey*, 52.1 (2012), 81 – 87 (p. 87).
[845] Sergey Radchenko and Mendee Jargalsaikhan, 'Mongolia in the 2016-17 Electoral Cycle: The Blessings of Patronage', *Asian Survey*, 57.6 (2017), 1032 – 1057 (p. 1043).

elections of June 2012 and September 2015 following changes from open list proportional representation to closed list proportional representation – changes mandated by Greek electoral law due to their proximity to previous contests[846] – and its subsequent return to open list PR in the regularly scheduled elections of January 2015 and 2019, respectively.[847] Following this, South Korea moved from parallel voting to proportional representation in 2019.[848] In the same year, Thailand moved from the use of parallel voting to a mixed single vote system.[849] Finally, in 2020 the electoral system of Ukraine also changed from parallel voting to proportional representation.[850] Each of these instances of electoral system change is captured within my dataset as they plausibly bear on the error exhibited by the polls housed within it.

While changes in electoral system within established democracy are evidently rare, this rarity may serve to make their occurrence more impactful. As they do not occur frequently, they do not represent events which polling organisations are necessarily prepared for, nor equipped on the basis of past experience within a country to deal with. However, changes in electoral system rarely occur without forewarning in the form of proposed legislation and governmental debates.[851] So, it is unlikely that polling organisations would be entirely caught off guard by these changes. However, any alterations to the methods and approaches taken to polling in light of this knowledge will nevertheless remain untested until the first election held under the new

[846] Maciej A. Górecki and Michal Pierzgalski, 'Legislated Candidate Quotas and Women's Descriptive Representation in Preferential Voting Systems', *European Journal of Political Research*, 61 (2022), 154 – 174 (p. 171).
[847] Ibid.
[848] Lee Minji, *National Assembly Passes Electoral Reform Amid Opposition Lawmakers' Protest* (2019), <https://en.yna.co.kr/view/AEN20191227003854315> [accessed 24/08/2023].
[849] Siripan Nogsuan Sawasdee, 'Electoral Integrity and the Repercussions of Institutional Manipulations: The 2019 General Election in Thailand', *Asian Journal of Comparative Politics*, 5.1 (2020), 52-68 (p. 55).
[850] Interfax, *Electoral Code Becomes Effective in Ukraine* (2020), <https://en.interfax.com.ua/news/general/633561.html> [accessed 24/08/2023].
[851] Bjørn Høyland and Martin G. Søyland, 'Electoral Reform and Parliamentary Debates', *Legislative Studies Quarterly*, 44.4 (2019), 593 – 615 (p. 593); David M. Farrell, Jane Suiter, and Clodagh Harris, 'The Challenge of Reforming a "Voter-friendly" Electoral System: The Debates Over Ireland's Single Transferable Vote', *Irish Political Studies*, 32.2 (2017), 293 – 310 (p. 293).

electoral system. As such, instances of electoral system change remain tests for poll-level adjustments that are fraught with the potential for error.

Scheduled vs. Snap Elections

Beyond their focus and the rules that govern their outcomes, elections take two broad forms: scheduled and snap. Scheduled elections are those which occur at regular intervals prescribed by electoral law, while snap elections are those which are called sooner than expected. As displayed in Table 19, my dataset captures 87 snap and 410 regularly scheduled elections. From this, it is clear that snap elections occur less frequently than their scheduled counterparts, but nevertheless occur to a degree sufficient for analysis in later models.

Table 19: The number of scheduled and snap elections within my dataset along with the proportion of elections they represent.

Election Type	Number	Percentage
Snap	87	17.5%
Scheduled	410	82.5%

Whether an election occurs at its regularly scheduled point in time or is called earlier than expected is likely to bear on the presence of polling error. Regularly scheduled elections are typically characterised by longer campaign periods than snap elections.[852] As one of the primary functions of election campaigns is to persuade individuals to vote for a given party or candidate,[853] and voters become increasingly aware of their voting preferences over the course of campaigns,[854] longer campaigns provide the electorate a longer time horizon over which to

[852] Rune Karlson, Georg Lutz, and Patrik Ohberg, 'Candidate Campaigns in Comparative Perspective', in *Parliamentary Candidates Between Voters and Parties: A Comparative Perspective*, ed. by Lieven de Winter, Rune Karlsen, and Hermann Schmitt (Oxford: Routledge, 2021), p. 82.
[853] Thomas M. Holbrook, *Do Campaigns Matter?*, (Thousand Oaks: Sage, 1996), p. 26.
[854] Gelman and King, p. 409.

crystallise their voting intention. This crystallisation lends itself to more stable voting intention which can be projected onto future voting behaviour with greater confidence. As poll-based predictions of vote share distributions rest on a series of projection mechanisms, this leads to the expectation of reduced distributive error and, in turn, a reduced likelihood of polls exhibiting error sufficient to exceed the bounds set by their associated margins of error.

By contrast, snap elections provide considerably reduced notice of their occurrence, as they are often called as the result of votes of no confidence or to press perceived political advantages.[855] They are therefore typically characterised by reduced campaign periods. This provides less time for voting intentions amongst the electorate to stabilise, increasing the likelihood of volatility and increasing the difficulty of accurately prediction vote share distributions, as voting intentions are more likely to change between the fieldwork dates of polls and election day. This lends itself to the expectation of increased distributive error on the part of polls which, in turn, increase the likelihood that they present errors of sufficient size to exceed their stated margins of error, leading to bounded inaccuracy.

Snap elections also bear upon the likelihood of polls presenting substantive misprediction. The volatility associated with voting attention in snap elections and the reduced time in which electoral decision-making can occur lends itself to increased late decision-making amongst the electorate. If this late decision-making occurs on a sufficiently large scale and sees voters predominantly opt for one party or candidate over others, it lends itself to substantial error in vote share predictions, increasing the likelihood of substantive misprediction. This is expectation is heightened if late decision-making manifests as late swing, defined as a shift in

[855] Jason Roy and Christopher Alcantara, ‘The Election Timing Advantage: Empirical Fact or Fiction?’, *Electoral Studies*, 31.4 (2012), 774 – 781 (p. 774); Jih-wen Lin, ‘How Are the Power of the President Decided? Vote Trading in the Making of Taiwan’s Semi-presidential Constitution’, *International Political Science Review*, 38.5 (2016), 659 – 672 (p. 663); Jakob-Moritz Eberl, Lena Maria Huber, and Carolina Plescia, ‘A Tale of Firsts: the 2019 Austrian Snap Election’, *West European Politics*, 43.6 (2020), 1350 – 1363 (p. 1353).

voting intention that occurs after the final poll for an election is conducted,[856] thereby eluding polling predictions.

The greater volatility of voting intentions in snap elections undermines the ability to which they can be accurately projected onto election day. As the predictions rendered by pre-election polls rest on a series of projection mechanisms, this decreases the accuracy with which they can call the overall winner of an election, leading to the expectation of increased substantive error.

Difference in Ideological Distance Between Parties/Candidates

The ideological distance between parties and voters exists as a key driver of voter decision-making.[857] As such, similarities and differences between the political alignments of the parties or candidates contesting elections can be expected to affect polling error due to their influence on the ease of decision-making amongst the electorate. In elections characterised by clear ideological differences between parties or candidates, the decision-making process faced by voters ought to be more straightforward, as the parties and candidates possess distinct identities which will differentially align with the existing ideological preferences of the electorate with little to no overlap. In elections characterised by close ideological similarity between parties or candidates, voters can be expected to be faced by a more challenging decision-making process, as the ideological positions of those contesting the election will be less distinct and may overlap, reducing the likelihood of different parties or candidates neatly aligning with existing ideological presences within the electorate. This in turn is likely to lead to increased late

[856] Roger Mortimore and others, 'BPC/MRS Enquiry into Election Polling 2015: Ipsos Mori Response and Perspective', *International Journal of Market Research*, 59.3 (2017), 285 – 300 (p. 286).
[857] Pedro Riera and Francisco Cantú, 'Electoral Systems and Ideological Voting', *European Political Science Review*, (2022), 1 – 19 (p. 16).

decision-making within an election, making it more difficult for polls to render accurate vote share predictions, leading to increased distributive inaccuracy.

Equally, the adoption of positions deemed radical by the electorate may give voters pause for thought, resulting in hesitancy and an increase in late decision-making. In any given election, the median voter typically rests towards the centre of the left-right ideological continuum.[858] While voters in certain countries are more inclined to lean to either the left or right, the median voter always exists a considerable distance from the ideological extremes.[859] As such, in elections characterised by ideologically disparate contestants – and therefore candidates or parties more likely to sit away from the ideological centre – a greater degree of hesitancy amongst voters can be expected, as voters feel more ideologically distant from parties or candidates. This hesitancy lends itself to greater late decision-making within an election, increasing the likelihood of late swings, and leading to the potential for increased distributive polling error. If hesitancy is widespread, and late swings are both extensive and unidirectional, then it also increases the likelihood of substantive misprediction on the part of polls.

Ideological misalignment between voters and parties or candidates also increases the likelihood of abstention,[860] thereby affecting turnout and potentially confounding the voter turnout models employed by pollsters. Difficulties in correctly identifying likely voters lend themselves not only to increased distributive error, due to the greater difficulty predicting vote share distributions, but also to increased substantive error. If the abstention confounding likely voter models is widespread and is more prevalent in certain socio-political groupings than others, it

[858] Jan-Emmanuel De Neve, 'The Median Voter Data Set: Voter Preferences Across 50 Democracies', *Electoral Studies*, 30.4 (2011), 865 – 871 (pp. 866 – 867); Heemin Kim and Richard C. Fording, 'Voter Ideology in Western Democracies: An Update', *European Journal of Political Research*, 42 (2003), 95 – 105 (p. 99).
[859] Ibid.
[860] James Adams, Jay Dow, and Samuel Merill III, 'The Political Consequences of Alienation-based and Indifference-based Voter Abstention: Applications to Presidential Elections', *Political Behavior,* 28.1 (2006), 65 – 86 (p. 66).

may increase the likelihood of substantive polling error by affecting the substantive outcome of elections.

For the expected effect of the ideological positions of parties and candidates to affect my measures of polling error, the differences present within my studied elections must be sizeable enough to generate a response from the electorate. I measure the difference in the ideological position of parties and candidates in a given election as the standard deviation of their left-right alignment scores as provided by the 1999 – 2019 Chapel Hill expert survey data.[861] Taking the standard deviation provides a measure of the variability in left- and right-wing alignment between the parties or candidates contesting an election. Elections characterised by a lower standard deviation are contested by parties or candidates of more similar political alignments. Conversely, elections characterised by higher standard deviations are contested by parties or candidates with more variable, and therefore disparate, political alignments. As the data from which these standard deviations are calculated only covers the period 1999 – 2019, I only possess data on the left-right alignment of the parties and candidates for a subset of my studied elections.

Figure 25 displays the left-right standard deviation of political parties and candidates in those elections for which data was available. It is immediately apparent that data on the ideological difference between parties and candidates gathered from Chapel Hill expert surveys was only available for elections conducted in Europe. This presents a notable limitation that future work should attempt to overcome. Nevertheless, data on the left-right standard deviation of political parties and candidates across European elections provides an insight into its variability between cases and, therefore, its plausibility as a driver of polling error.

[861] Seth Jolly and others, 'Chapel Hill Survey Trend File, 1999 – 2019', *Electoral Studies*, 75 (2022), 102420 (p. 102420).

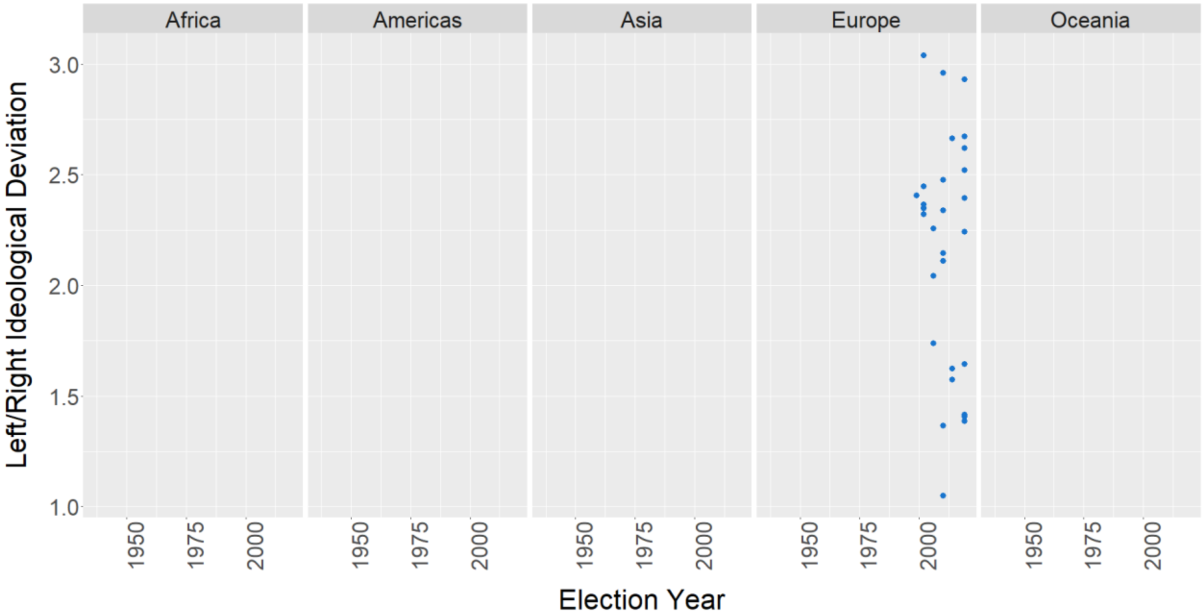

Figure 25: The standard deviation of the ideological alignments of parties and candidates across those elections for which data was available.

Despite data on the ideological difference between parties and candidates only being available for elections conducted in Europe, it is clear from Figure 25 that across these elections, the ideological distance between parties and candidates varies considerably, indicating that their nature and polarity differs across contests. That the ideological differences between parties and candidates differ notably between elections lends plausibility to their ability to serve as drivers of variance in polling error across cases.

Differences in the Number of Registered Voters

Differences in the number of registered voters between elections can be expected to impact my measures of polling error due to their ability to alter the composition of the voting population. While not all registered voters actually vote,[862] even in countries with compulsory voting,[863] the act of registering to vote betrays an interest in electoral participation.[864] Though the

[862] Craig Allen Smith, *Presidential Campaign Communication: The Quest for the White House*, (Malden: Polity Press, 2010), p. 5.
[863] Louth and Hill, p. 32.
[864] David W. Nickerson, 'Do Voter Registration Drives Increase Participation? For Whom and When?', *The Journal of Politics*, 77.1 (2014), 88 – 101 (p. 88).

presence of compulsory voter registration in countries complicates the relationship between the act of registering and voting likelihood, registration nevertheless serves as a prominent driver of voting behaviour.[865] Pronounced changes in the number of registered voters therefore represent a likely influx of new voters. As they have not been previously registered, these voters do not have an established voting history on which to base expectations of voting behaviour. As likely voter models depend on past behaviours to establish the likelihood of individuals voting in an election,[866] the presence of a large number of individuals with no prior voting history has the potential to undermine them. This in turn makes the accurate prediction of vote share distributions more difficult, leading to the expectation of increased distributive polling error. In presenting increased distributive error, polls are not only more likely to exhibit errors sufficient to exceed their stated margins of error, but are also more likely to offer substantively incorrect predictions, especially in close contests.

Though demographic data on registered voters is difficult to obtain, or simply non-existent, across my studied countries due to the widespread practice of respondent anonymisation within pre-election polling,[867] large changes in the number of registered voters bring with them an increased likelihood of significant shifts in the demographic composition of those individuals registered and, therefore, likely to vote. This also has the potential to undermine the likely voter models on which poll-based predictions rest, especially if the demographic composition of the electorate changes significantly between elections – as was reported to have happened during the purported ‘youthquake’ surrounding the British general election of 2017[868] – or otherwise inactive demographics suddenly decide to vote to a greater extent, as was seen in Donald

[865] Robert S. Erikson, ‘Why Do People Vote? Because they are Registered’, *American Politics Research*, 9.3 (1981), 259 – 276 (p. 259).
[866] Rentsch, Schaffner, and Gross, p. 783.
[867] Walden, p. 184.
[868] Christopher Prosser and others, ‘Tremors but no Youthquake: Measuring Changes in the Age and Turnout Gradients at the 2015 and 2017 British General Elections’, *Electoral Studies*, 64 (2020), 102129 (p. 102129).

Trump's mobilisation of former non-voters in the 2016 US presidential election.[869] As before, compromised likely voter models increase the likelihood of polls exhibiting distributive, bounded, and substantive error.

Registered voter figures are taken from the International IDEA voter turnout dataset.[870] While the dataset contains occasional inconsistencies that have been recognised in previous scholarship,[871] it remains the only comprehensive, global dataset on voter registration presently available. Figure 26 displays the number of registered voters across elections from 1936 – 2020 broken down by continent.

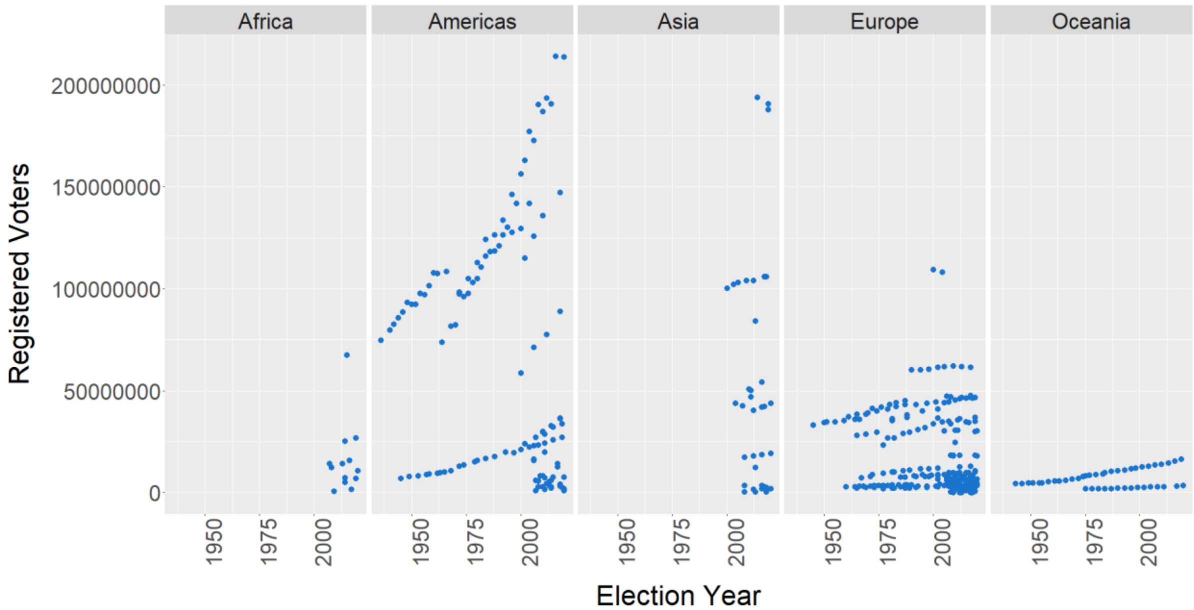

Figure 26: The number of registered voters across my studied elections from 1936 to 2020 broken down by continent. Each point within the figure represents an individual election.

Figure 26 displays clear differences in the number of registered voters between elections over time. In most countries, a steady, approximately linear increase in the number of registered

[869] Ron Johnston and others, 'Was the 2016 United States' presidential contest a deviating election? Continuity and change in the electoral map – or "Plus ça change, plus ç'est la mème géographie"', *Journal of Elections, Public Opinion and Parties*, 27.4 (2017), 369 – 388 (p. 386).

[870] International IDEA, *Voter Turnout Database* (2022), <https://www.idea.int/data-tools/world-view/40> [accessed 17/08/2022].

[871] Jonathan Mellon and others, 'Aggregate Turnout is Mismeasured' (2018), available at SSRN: https://ssrn.com/abstract=3098436 [accessed 23/04/2022].

voters proportionate to the increase in population over time is to be expected. Successive points representing this trend are consistently visible across most of continental groupings. Even in cases when the number of registered voters has declined over time, often as a result of aging populations such as that displayed by Japan,[872] or is more sporadic in nature, as appears to be the case in African elections (itself likely an artefact of data quality issues)[873], differences in the number of registered voters are clearly evident between elections. While certain progressions of points within Figure 26 illustrate substantial changes in the number of registered voters between cases, others illustrate more gradual change. Despite this, differences in voter registration between contests remain consistent. The consistent occurrence of differences in the number of registered voters between elections coupled with the presence of pronounced differences between certain elections lends plausibility to their existence as drivers of polling error between cases.

While differences in electoral characteristics between cases are necessary to allow them to serve as drivers of polling error variance, for these differences to be useful as discrete predictors of error variance, the variables on which they are based must be statistically independent. In the following sub-section, I explore the degree to which my stated election-level variables present statistical (in)dependence by exploring the extent to which they are correlated with one another.

Pearson's Correlations Between Election-level Predictor Variables

Varying degrees of inter-connectedness can be expected to exist between my election-level predictor variables. For example, strong partisan loyalty in the electorate would be expected to

[872] Yasuo Takao, 'Aging and Political Participation in Japan: The Dankai Generation in a Political Swing', *Asian Survey*, 49.5 (2009), 852 – 872 (p. 855).
[873] Astrid Evrensel, *Voter Registration in Africa: A Comparative Analysis*, (Johannesburg: EISA Publishing, 2010), p. 190.

be connected to the level of voter turnout, with voters who feel stronger political loyalties being more likely to vote on election day. Equally, the extent of late decision-making amongst the electorate can be expected to be positively related to the effective number of elective parties contesting an election, as a greater array of choices lends itself to increased indecision.

Due the inter-relation present between my election-level predictor variables, they can be expected to exhibit a degree of correlation. Strong correlations between variables run the risk of introducing multicollinearity due to statistical dependence.[874] While multicollinearity does not necessarily affect the ability of models to accurately predict outcomes, it does obscure the contribution of individual predictor variables.[875] As such, the presence of multicollinearity within my data would undermine my ability to effectively analyse the usefulness of election-level variables as predictors of polling error.

To ensure that no instances of multi-collinearity exist between my election-level predictor variables, I present their degree of association below in Figure 27. The association between my variables is represented using linear, pairwise correlations calculated using Pearson's rho. Strong positive correlations are presented in blue, while strong negative correlations are displayed in red. Correlation coefficients are also provided for more granular analysis of associations between variables.

[874] Daniel J. Mundfrom, Michelle DePoy Smith, and Lisa W. Kay, 'The Effect of Multicollinearity on Prediction in Regression Models', *General Linear Model Journal*, 44.1 (2018), 24 – 28 (p. 24).
[875] Ibid.

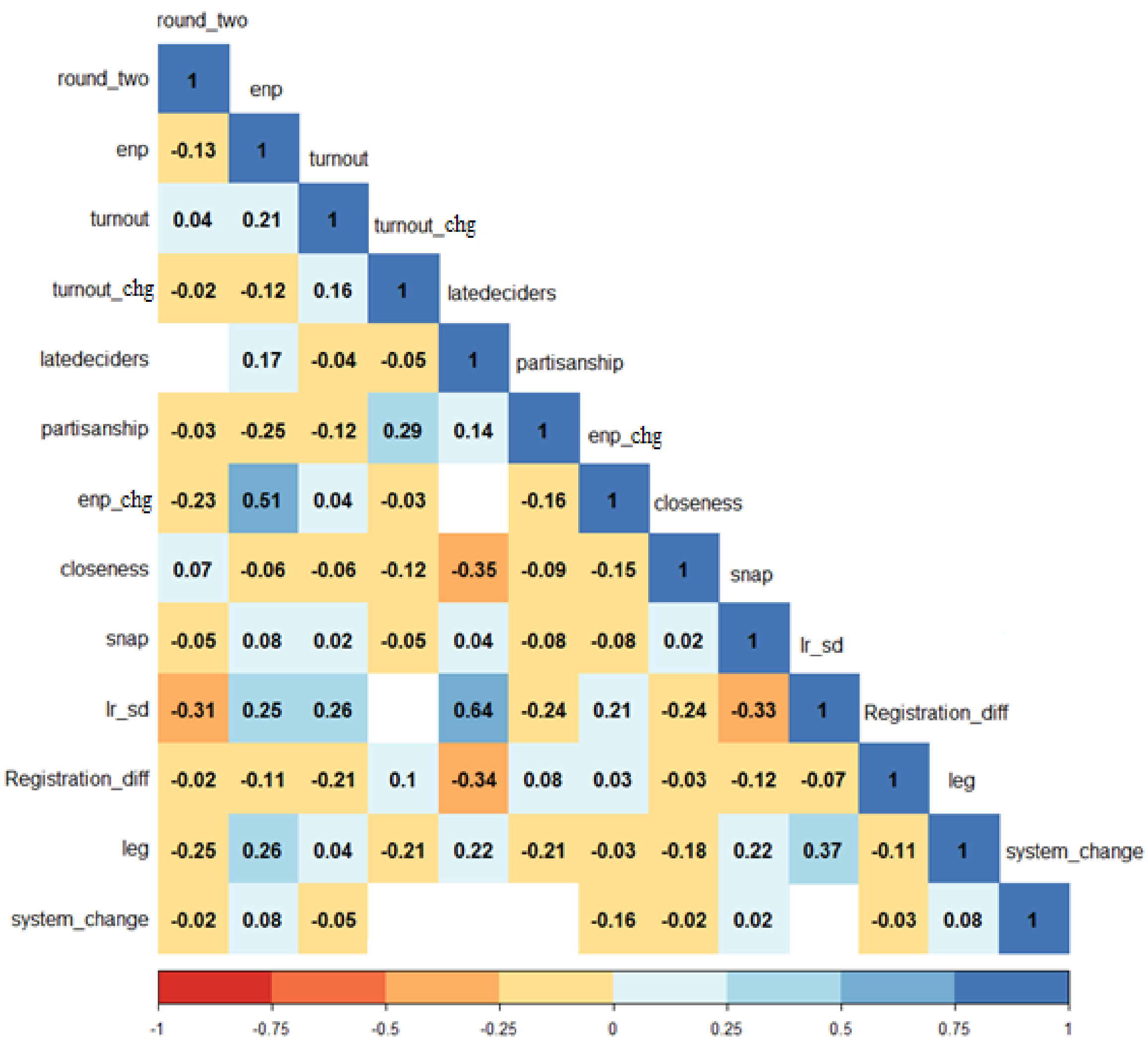

Figure 27: Pearson's correlations between election-level predictor variables. The strength of correlations is displayed both numerically and on a graduated colour scale displayed at the bottom of the figure. Strong positive correlations are displayed in blue, while strong negative correlations are displayed in red. White cells indicate that insufficient information was available to draw a meaningful correlation.

From the figure, it is clear that the correlations between my election-level predictor variables vary considerably in both strength and direction. Assessing the strength of correlations allows for the detection of multicollinearity which undermines the degree to which inferences can be drawn from analysis due to strong inter-relation between variables. A correlation coefficient of ≥0.80 is taken to indicate the presence of multicollinearity.[876] Fortunately, none of the

[876] Noora Shrestha, 'Detecting Multicollinearity in Regression Analysis', *American Journal of Applied Mathematics and Statistics*, 8.2 (2020), 39 – 42 (p. 40).

correlations between my variables are strong enough to denote the presence of multicollinearity. The strongest positive correlations exist between the left-right standard deviation of parties (lr_sd) and the presence of late deciders in the electorate (0.64), and between ENEP and ENEP change (0.51). By contrast, the strongest negative correlations exist between late decision-making and both the closeness in winning margin (-0.35) and the difference in registered voters between elections (-0.34).

Each of these more pronounced correlations is either in-line with theoretical expectations or is, at the very least, understandable. Late decision-making can be expected to increase in the presence of greater ideological extremes amongst the parties or candidates contesting an election, given the centrist tendencies of the median voter.[877] The positive relationship between ENEP and ENEP change is understandable given their common focus on the effective number of electoral parties within elections. Similarly, the negative relationship between late decision-making and both the closeness in winning margin and the difference in registered voters are intelligible. A situation can be imagined in which the extent of late decision-making amongst the electorate increases and, providing that these late deciders break in the same or similar directions, the closeness of a given election decreases. Equally, as the number of registered voters increases positively between elections, it is not unreasonable to expect the extent of late decision-making to decrease, especially in systems without compulsory registration. This is because registered voters can be expected to be motivated and are therefore more likely to have decided for whom they are going to vote prior to election day. Indeed, this decision may be the driving force behind their choice to register, as strong party preference has been linked to rates of electoral participation.[878]

[877] De Neve, pp. 866 – 867; Kim and Fording, p. 9.
[878] Rau, p. 1021.

5.2: The Expected Impact of Election-level Differences in Interaction

Though electoral characteristics can be expected to bear upon polling error individually, it may be that the impact of a given electoral characteristic on polling error varies according to the value taken by another. That is, that the impact of a characteristic is enhanced or moderated by another such that the interaction between them warrants analysis.[879] Building on the earlier discussion of the likely (non-)compensatory effect of electoral characteristics within the theory chapter, I isolate likely interactions between my specified election-level variables and identify their expected impact on my measures of polling error. What follows is not intended to be an exhaustive list of all possible interactions between electoral characteristics. Rather, it is simply a plausible array of interactions suitable for exploration in later analysis. I limit my focus to likely two- and three-way interactions, reserving higher order interactions for exploration in later models.

Among my specified election-level variables, a two-way interaction between turnout change and ENEP change between elections can be expected. Specifically, the impact of turnout change on polling error is likely to be more pronounced in elections characterised by large changes in ENEP. Significant changes in ENEP between elections indicate that a meaningfully different number of parties received notable shares of the vote, suggesting that voters either moved to or away from parties between elections to an impactful extent. Turnout change in these circumstances may be driven by the emergence of new parties attracting previously apathetic voters, or by voters moving to support previously marginal parties.[880] Here, anticipating the shift of existing voters to a previously marginal party may prove a difficult task

[879] Ulf Andersson, Alvaro Cuervo-Cazurra, and Bo Bernhard Nielsen, 'From the Editors: Explaining Interaction Effects Within and Across Levels of Analysis', *Journal of International Business Studies*, 45 (2014), 1063 – 1071 (p. 1064).

[880] Tim Immerzeel and Mark Pickup, 'Populist radical right parties mobilizing 'the people'? The role of populist radical right success in voter turnout', *Electoral Studies*, 40 (2015), 347-360 (p. 347).

for polls, especially if this party is considered to be radical in nature, as individuals may be less inclined to reveal their intention to vote for them when polled.[881] Moreover, in the case of predicting the performance of newly emergent parties, little precedent exists for gauging their likely support before the fact, rendering predictions of their eventual vote share more difficult than established parties for which substantial historical voting data exists. The difficulty in anticipating support for newly emergent parties is compounded if their eventual support base comprises previously apathetic voters who have not turned out in previous elections due to a feeling unrepresented by the previous array of choices.

Equally, turnout changes that occur alongside substantial shifts in ENEP may be motivated by the decline, ideological reorientation, or collapse of established political parties, such as the Socialist Party in France,[882] which result in them garnering a notably smaller share of the vote or, potentially, not garnering a meaningful share of the vote at all. This comes about as voters no longer turn out to vote for parties that they previously supported due to their absence or refuse to vote for existing parties due to disagreements with their new ideological position. The problem that this presents for polling is two-fold. In this situation, polls are faced with the need to establish whether voters who previously supported parties that have since collapsed or re-orientated themselves ideologically remain likely to turn out on election day. It may be that voters who find themselves without a party, or are disaffected with their previous party, will turn out in support of another. However, these voters may elect not to turn out and instead opt to remain home on election day. In the first instance, polls are faced with the difficult task of reallocating previously engaged voters to different parties in the newly composed electoral environment, increasing the likelihood for inaccuracy in the estimated vote shares provided by

[881] Noelle-Neumann, p. 307.
[882] Ben Clift and Sean McDaniel, 'Is this crisis of French socialism different? Hollande, the rise of Macron, and the reconfiguration of the left in the 2017 presidential and parliamentary elections', *Modern & Contemporary France*, 25.4 (2017), 403-415 (p. 403).

polls, leading to the expectation of increased error. In the second, polls must contend with a proportion of previous voters no longer turning out. While, if these individuals can be identified, this removes the need for their re-allocation to other parties, concentrated demobilisation such as this fundamentally alters the composition of the population of voters who turn out on election day which has the potential to confound the likely voter models on which poll-based predictions rest, further increasing the likelihood of polling error.

By contrast, elections characterised by smaller changes in ENEP are less likely to capture the emergence of new parties, shifts to support previously marginal parties, or the collapse or re-orientation of established parties. To this end, they are less likely to bring about those circumstances that amplify the impact of turnout change on polling error. Nevertheless, it is important to recognise that it also possible for turnout change to bear on polling error when ENEP values remain similar between cases. In these circumstances, it may be that established parties have ideologically realigned to a degree and accrued or lost a small number of votes that is not sufficient to significantly alter measures of ENEP, but which may prove to be important in close contests and lead to polls incorrectly calling the winner of contests if they go unforeseen or occur late within the campaign, resulting in substantive error.

A two-way interaction between turnout change and the absolute ENEP value associated with elections can also be expected due to likelihood of party emergence in different systems. Proportionally representative electoral systems typically result in elections that encompass higher ENEP values.[883] Elections encompassing high ENEP values, such as those conducted in proportionally representative election systems, are more conducive to party emergence than those with lower levels of ENEP, such those conducted under majoritarian systems.[884] As such,

[883] Andreas Ladner and Henry Milner, 'Do voters turn out more under proportional than majoritarian systems? The evidence from Swiss communal elections', *Electoral Studies*, 18.2 (1999), 235-250 (p. 235).
[884] Raimondas Ibenskas and Allan Sikk, 'Patterns of party change in Central and Eastern Europe, 1990-2015', *Party Politics*, 23.1 (2017), 43-54 (p. 50).

elections that exhibit higher ENEP values are more likely to bring about the previously identified issues associated with party emergence and turnout change that bear upon the accuracy of polls than those that exhibit lower ENEP values, resulting in the expectation of more pronounced polling error.

A two-way interaction between the turnout level in an election and its ENEP value can also be anticipated. Instances of voters switching parties between elections are more pronounced in high ENEP environments, such as those brought about by proportional representation, than in elections characterised by lower ENEP values, such as those conduced in majoritarian systems.[885] For voters to switch parties from one election to the next, they must necessarily have turned out in support of a party in the first contest. As such, the difficulties posed by party switching do not centre on turnout *change* between elections, but rather concern the re-allocation of *existing* turnout from the preceding election to the present contest.

In high ENEP elections, the re-allocation of previous turnout to the election of interest is affected by the added issue of the volatility introduced by prevalent party switching. This impacts the manner in which turnout is distributed across contests, increasing the likelihood of differences between elections, even if they are similarly composed. To this end, turnout levels could remain constant between contests and capture the same demographic composition, yet still manifest as notably different vote share distributions come election day. To a degree, this can be expected in all elections. If the same set of voters turns out, they won't necessarily vote in the same manner from one election to the next. However, due to the heightened potential for party switching, these differences will likely be more pronounced in high ENEP environments. As such, while turnout levels can still be expected to bear on polling error in the manner identified earlier in the thesis, the heightened degree of volatility in the distribution of turnout

[885] André Blais and others, 'Assessing the psychological and mechanical impact of electoral rules: A quasi-experiment', *Electoral Studies*, 31 (2012), 829-837 (p. 831).

introduced by the increased prevalence of party switching in high ENEP elections can be expected to make the process of identifying likely voting behaviour more difficult as it undermines the degree to which the past relationship between turnout patterns and vote share distributions can be used to inform current expectations, increasing the potential for polling error.

The two-way interactions between turnout and ENEP is likely to be mitigated by partisanship, leading to a three-way interaction. The prevalence of party switching between elections is likely to be contingent on the extent of strong partisan loyalty amongst the electorate. On the basis of their loyalty to a given party, partisans are far less likely to defect to another party between elections than non-partisans. To this end, strong partisanship can be expected to constrain the extent of likely party switching, reducing the degree of volatility it introduces to turnout distribution between cases. This will allow the past relationships between voter turnout and the distribution of vote shares to be used more reliably to inform current expectations, reducing the likelihood of polls rendering errant vote share estimates.

Similarly, the two-way interaction between turnout change and ENEP change also has the potential to be mitigated by partisanship, though this expectation requires a greater degree of unpacking. Partisans are well-understood to be more likely to turn out than non-partisans,[886] leading to the expectation that non-voters are more likely to be non-partisans. Turnout change between elections implies a change in the proportion of non-voters. The proportion of non-voters either decreases as a result of increased mobilisation, or increases as a result of demobilisation. With this in mind, in cases where changes in ENEP are driven by the emergence of a new party, associated turnout change is likely to capture the mobilisation of previous non-voters in support of this party. In strong partisan environments, a greater

[886] Rau, p. 1021; Bartels, p. 37.

proportion of the electorate can be expected to be partisans loyal to established political parties, reducing the proportion of individuals who are non-partisans and, therefore, likely non-voters. This reduces the extent to which an emerging party is able to mobilise non-voters, constraining the share of the vote they are able to garner. Ultimately, this limits the impact of the turnout change that the emergence of a new party may catalyse on vote share distributions, reducing the degree to which it is likely to drive an increase in polling error.

The impact of differences in turnout between elections can also be expected to be related to the margin of victory enjoyed by the winning party or candidate. As such, a two-way interaction between them can be expected. Differences in turnout are more likely to be impactful in elections characterised by smaller margins of victory as the difference in the size of voting populations that they entail is more likely to affect substantive electoral outcomes given the slim margins by which results are determined. This increased likelihood leads to the expectation that differences in turnout are more likely to impact upon the ability of polls to predict substantive electoral outcomes in closer contests. By contrast, differences in turnout are less likely to be impactful in elections characterised by wider margins of victory, as the attendant differences in the size of voting populations has a reduced chance of meaningfully altering the substantive outcome of these contests and reducing the chance of substantive polling error.

A two-way interaction between turnout change and partisanship can also be anticipated. High levels of strong partisan loyalty amongst the electorate constrain the degree to which votes will be distributed amongst parties or candidates, as the majority of votes will be cast for those parties to whom the electorate is loyal. As such, in elections characterised by strong partisan loyalty, differences in turnout between elections, and therefore differences in the size of the voting population, are less likely to confound polling predictions given the increased predictability of the manner in which these differences will impact upon voting behaviour. This reduces the likelihood of distributive polling error and, therefore, the probability of polls exhibiting both bounded and substantive error. However, in low partisanship environments, the dispersion of differences in turnout will be less deterministic and is more likely to affect a greater number of vote shares in less predictable ways, increasing the difficulty of accurately predicting vote share distributions. This increases the likelihood of distributive polling error which lends itself to more widespread bounded error and increases the probability of polls presenting substantively incorrect predictions.

Lower levels of partisan loyalty are also more likely to increase the impact of a large number of effective elective parties (ENEP) on polling error, leading to a two-way interaction between these variables. In elections characterised concurrently by larger ENEP values and diminished partisan loyalty amongst the electorate, voters are not only presented with a greater number of potentially viable choices, but can also be expected to be more open in their decision-making, leading to a greater and less predictable dispersion of votes. This increases the difficulty of accurately predicting vote share distributions, increasing the likelihood of distributive polling error and, in turn, both bounded and substantive polling error.

Late decision-making amongst the electorate possesses a potential interaction with the level of turnout in a given election. Disagreement exists as to the traits possessed by late deciders.[887] While this is understandable, as late deciders represent a heterogenous group of voters,[888] it does not lend itself to a clear-cut understanding of the manner in which turnout and late decision-making are most likely to interact. Nevertheless, suggestions within the literature lend

[887] Brox and Giammo, p. 333.
[888] Luigi Ceccarini and Ilvo Diamanti, 'The Election Campaign and the Last-minute Deciders', *Contemporary Italian Politics*, 5.2 (2013), 130-148 (p. 134).

themselves to a range of scenarios in which turnout can be expected to bear on late decision-making, permitting the expectation of an interaction between the two variables.

High turnout elections are often indicative of heightened enthusiasm within the electorate,[889] while late decision-making has been linked to a lack of political interest and, by extension, a lack of enthusiasm at the prospect on engaging in elections.[890] As such, the increase in enthusiasm marked by high turnout has the potential to reduce the extent of late decision-making, as a smaller proportion of individuals will feel sufficiently unenthused as to only decide on their voting behaviour towards the end of the campaign. The diminished scope of late decision-making reduces the potential for sizeable swings in voting intention late in the campaign, making polling error less likely.

By way of contrast, low turnout in an election speaks more readily to a widespread lack of interest and enthusiasm amongst the electorate. To this end, low turnout elections are more likely to be representative of a generally disaffected electorate who are unlikely to have been motivated by political campaigning to a meaningful extent. Under these circumstances, it may be that a greater number of individuals ultimately decide to vote later in the campaign out of a sense of civic duty,[891] rather than earlier in the campaign out of enthusiasm for the candidates or parties at play. This would lead to a greater proportion of voters engaging in late decision-making, increasing the potential for late swings in voting intention, and lending itself to the possibility of heightened polling error.

[889] Nicholas A. Valentino and others, 'Election Night's Alright for Fighting: The Role of Emotions in Political Participation', *The Journal of Politics*, 73.1 (2011), 156-170 (p. 156).
[890] J. David Gopoian and Sissie Hadjiharalambous, 'Late-deciding voters in presidential elections', *Political Behavior*, 16 (1994), 55-78 (pp. 74-76).
[891] André Blais and Christopher H. Achen, 'Civic Duty and Voter Turnout', *Political Behaviour*, 41 (2019), 473-497 (p. 473).

The impact of late decision-making is also plausibly dependent on the effective number of elective parties contesting an election (ENEP). In elections characteristics by higher ENEP values, a greater number of parties or candidates exist for the votes cast by late deciders to be dispersed across. Therefore, their impact has the potential to be more diffuse than lower ENEP environments and less impactful on individual party or candidate vote shares. This diminished effect lends itself to the expectation of reduced distributive polling error. As it is necessary for their occurrence, diminished distributive inaccuracy also reduces the likelihood of both bounded and substantive error.

The diffusion of votes brought about by differences in ENEP can also be expected to affect the degree to which turnout is likely to impact polling error. Higher turnout elections that are contested by a larger number of effective parties will be characterised by a greater number of individuals engaging in complex decision-making between a range of viable voting choices. As the projection mechanisms undergirding polling predictions seek to accurately map these decision-making processes, a higher turnout election characterised by a large number of parties makes this process more difficult, as it necessitates the accurate prediction of increasingly diffuse decision-making over a greater number of voters. By the same token, the difference in ENEP between elections can also be expected to affect the degree to which turnout affects polling error. The diffusion of turnout across a larger number of effective parties lends itself to increased distributive error, as vote share distributions are likely to be more difficult to accuracy predict. Given the interconnection between my conceptualisations of polling error, this also increases the likelihood of polls exhibiting bounded inaccuracy, as they are more likely to present errors sufficient to exceed their margins of error, as well as substantive inaccuracy, as distributive inaccuracy is a necessary condition for its occurrence.

The two-way interactions between turnout and both ENEP and ENEP change can be expected to be affected by the presence of partisanship, leading to a likely three-way interaction. This expectation is based on the degree to which partisan loyalties constrain the fragmentation of voting behaviour. In elections characterised by a high degree of partisan loyalty, voters will coalesce around those parties to which they are loyal, reducing the fragmentation of voting behaviour. As such, high levels of partisan loyalty may mitigate the effect of increases in ENEP, as the introduction of newly viable parties is less likely to affect the behaviour of voters loyal to their competitors. This can be expected to reduce the difficulty of accurately predicting vote share distributions, leading to reduced distributive error. Given the interconnection between measures, a reduction in distributive inaccuracy can also be expected to reduce the likelihood of polls exhibiting both bounded and substantive inaccuracy.

On the other hand, low levels of partisan loyalty increase the likelihood of voter fragmentation. This increases the probability of a larger number of effective parties, or a significant difference in ENEP between elections, increasing the difficulty of correctly assigning voters, especially in a high turnout environment, as a greater number of voters exist to engage in fragmentary voting behaviour. Under these conditions, a higher degree of distributive error is to be expected, given the increased difficulty of accurately predicting vote share distribution. Increased distributive error not only makes polls more likely to exceed their margins of error, thereby presenting bounded inaccuracy, but also increases the likelihood that they present substantive mispredictions, especially in closely contested elections.

The impact of the level of turnout in an election on polling error can be expected to vary on the basis of the strength of partisan loyalty within the electorate, leading to a two-way interaction. A high turnout election characterised by low partisan loyalty is likely to be more conducive to distributive polling error than a low turnout election characterised by high partisan loyalty, as a greater number of voters will be engaging in unpredictable decision-making in the absence of partisan heuristics, making the accurate prediction of vote share distributions more difficult. This, in turn, also increases the likelihood of both bounded and substantive inaccuracy.

Election-level variables can be expected to differentially affect measures of my three conceptualisations of polling error, as well as polling bias. While interactions between them can be expected to impact upon distributive, bounded, and substantive inaccuracy, their anticipated effect in isolation often bears more closely to one conceptualisation of error than others. In the following sub-section, I unpack the degree to which the identified election-level variables are likely to affect my different conceptualisations of polling error, as well as measures of polling bias.

The Differential Impact of Electoral Characteristics Across Conceptualisations of Error

From the exploration of likely additive and interactive effects, election-level variables can be expected to aid in the prediction of distributive, bounded, and substantive polling error, as well as polling bias. While all of my conceptualisations of polling error are likely to be impacted by my stated election-level variables, the extent of this impact is likely to differ between conceptualisations. For example, several bear more closely on substantive error than the other conceptualisations. The margin of victory associated with a given election – and therefore the closeness of the competition between parties and candidates – is likely to have a far more pronounced effect on substantive error than other conceptualisations, so too are large-scale differences in turnout levels between elections and substantial degree of late decision-making amongst the electorate. By contrast, the impact of election-level variables on polling bias may be tempered by the heightened importance of predictor variables housed in the pollster-level of my model, as the degree to which polls exhibit bias has been found to be an artefact of the partisan leaning and house effects associated with the polling organisations that conduct them,[892] along with any sponsorship a given organisation may be in receipt of.[893]

The interconnection between my conceptualisations of polling error further illuminates their likely relationship with election-level predictors. The presence of distributive polling error can be expected to be positively correlated with bounded error. That is, as the presence of distributive error increases, so too does the likelihood of polls exhibiting error sufficient to exceed the bounds established by their margins of error. Equally, if a poll exhibits bounded inaccuracy, it necessarily also exhibits distributive error in excess of the bounds set by its margin of error. Given this interconnection, election-level variables that are predictive of distributive error ought also to be predictive of bounded error and vice versa.

While the connection between distributive and bounded polling error is likely to be reciprocal, the same cannot be said of their connection with substantive error. Distributive error stands as a necessary condition for substantive error, as the vote share predictions offered by polls must be erroneous in order to result in substantive misprediction, and large-scale instances of distributive polling error increase the likelihood of substantive misprediction. However, instances of substantive misprediction do not necessarily entail large-scale distributive errors. For example, one poll may correctly predict the winner of an election but overestimate their vote share by fifteen points and another may incorrectly predict the winner of election, but only underestimate their share of the vote by two points. The first poll is substantively correct, but presents pronounced distributive error, while the second is substantively incorrect yet present minimal distributive error. As such, while distributive error increases the likelihood of

[892] Pickup and Johnson, pp. 272 – 284; Bergman and Holmquist, p. 307; Shamir, p.62; Jackman, p. 500.
[893] Shamir p. 62.

substantive misprediction, the presence of substantive inaccuracy does not necessarily increase the likelihood of meaningful distributive inaccuracy.

A similar relationship can be expected between bounded and substantive inaccuracy. While bounded inaccuracy increases the likelihood of substantive misprediction, as it represents instances in which polls exhibit sufficiently large distributive errors to breach their margins of error, the presence of substantive inaccuracy does not necessarily entail bounded inaccuracy. A poll could present distributive error within its margin of error and still fail to correctly call the winner of an election, especially in close contests. Equally, a poll could present error far in excess of its margin of error and still correctly predict the winner of an election, especially in cases where vote shares are overestimated. As such, while the presence of bounded inaccuracy increases the likelihood of substantive misprediction, instances of substantive inaccuracy do not necessarily entail distributive errors sufficient to cause bounded error.

Given their inter-relation, all election-level variables and interactions that can be expected to meaningfully bear on distributive and bounded inaccuracy can also be expected to affect substantive inaccuracy to a substantial degree. However, not all election-level variables and interactions that meaningfully bear on substantive inaccuracy can be expected to significantly affect distributive and bounded inaccuracy. As a greater number of factors can be expected to meaningfully bear upon substantive inaccuracy, election-level variables and interactions ought therefore to be more predictive of it than other conceptualisations of error.

The expectation that election-level variables will be most predictive of variance in substantive polling error is not only motivated by theoretically driven relationships and the interplay between my measures of error, but is also empirically grounded. The ICC findings presented in the previous chapter indicate that election-level differences account for a greater proportion of observed variance in substantive polling error than alternative conceptualisations. Variables housed at the election level are therefore more impactful for variance in substantive polling error and stand to be more useful predictors of it.

On the basis of the expected relationships between election-level variables and polling error, along with the interplay between my measures of polling inaccuracy, the likely importance of variables housed in alternative grouping levels, and the empirical findings presented in the previous chapter, I draw my second hypothesis:

> *H2: Election-level variables will aid models in predicting polling error variance both additively and interactively, proving most useful in the case of substantive error and least useful in the case of bias.*

In order to meaningfully test the utility of election-level variables as predictors of polling error, it is necessary to control from predictor variable housed in the additional grouping levels of its four-level structure. In the following section, I outline a series of predictor variables from the poll, pollster, and country grouping levels to serve as controls in later analysis.

5.3: Controlling for Predictor Variables from Additional Grouping Levels

The election-level differences identified in the previous section are not the only variables of importance when considering polling error. As polling error is a multi-level phenomenon, factors from additional grouping levels are likely to bear upon it. Beyond the election level, three other groups of variables can be expected to affect polling error. These exist at the poll, pollster, and country levels housed within the four-level structure of sources of polling error (see: Figure 28 below).

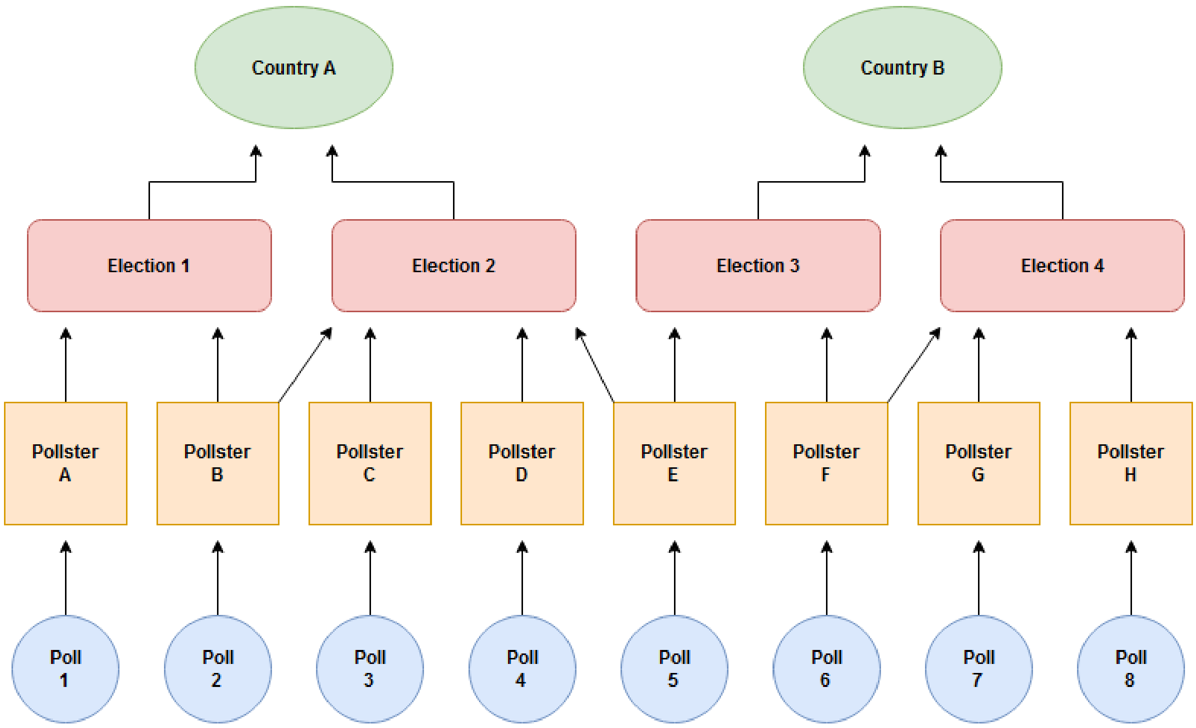

Figure 28: The four levels at which sources of polling error are clustered, as represented by earlier multi-level models. Error is clustered at the country level (green), the election level (red), the pollster level (orange), and the poll level (blue).

In the previous chapter, differences associated with the poll, pollster, and country grouping levels were found to account for non-trivial amounts of variance in polling error. While election-level differences remain the primary focus of this thesis, it would be remiss not to account for the predictive utility of the variables contained within these levels, as to omit them would be to inadequately represent the multi-level nature of polling error. Table 20 presents a range of variables from each of these additional grouping levels gathered for use as controls in later prediction models along with statistics describing their nature. In the following sub-sections, I outline the rationale behind the selection of each of these control variables.

Table 20: Control variables housed in the poll-, pollster-, and country-level groupings of sources of pre-election polling error. Continuous and binary variables are provided along with their minimum (Min.) and maximum (Max.) values, means (μ), and standard deviations (σ). For binary and categorical variables, the number of categories (Cats.) is also provided.

Variable	Type	μ	σ	Min.	Max.	Cats.
Poll-level Controls						
Sample size	Continuous	1871	4039	100	152,640	-
Days to election	Continuous	48	40	0	150	-
Pollster-level Controls						
Organisation	Categorical	-	-	-	-	1,033
Occurrences	Continuous	33.18	61.33	1	436	-
Country-level Controls						
Country	Categorical	-	-	-	-	83
Region	Categorical	-	-	-	-	5
Subregion	Categorical	-	-	-	-	18
Regime type	Categorical	-	-	-	-	3
Electoral system	Categorical	-	-	-	-	6
Moratorium	Binary	0.38	0.49	0	1	2
Compulsory vote	Binary	0.12	0.33	0	1	2
GINI	Continuous	34.78	6.75	20.50	63.00	-
GDP	Continuous	31.44	17.36	1.53	116.65	-

Poll-level Control Variables

From the research addressed in the literature review, it is clear that a range of poll-level factors bear closely upon polling error. I focus on two prominent poll-level factors that can be expected to differ between polls and affect the degree to which they are likely to exhibit predictive error. These characteristics are the sample size of respondents used by a poll given its influence on their margin of error;[894] and the proximity of a poll to election day, as the predictive accuracy of polls generally increases as the time to election day decreases.[895]

The poll-level factors selected for inclusion within this thesis do not represent the universe of all poll-level variables that could plausibly affect polling error. Indeed, other characteristics associated with polls can also be expected to affect their levels of displayed error. These included the day of the week and time of year that respondents are contacted due to its potential impact on response rates and, therefore, its ability to bring about non-response bias if the voting intention of those who fail to respond to polls differs systematically from the voting intention of those who do;[896] as well as the impact of the idiosyncrasies of individual interviewers (in the case of face-to-face and live telephone polls).[897] While the day of the week and time of year that respondents are contacted does affect response rates,[898] lower response rates have not been found to systematically reduce polling accuracy.[899]

[894] Lau, p. 5.
[895] Jennings, Lewis-Beck, and Wlezien, p. 949.
[896] Glen Shinn, Matt Baker, and Gary Briers, 'Response Patterns: Effect of Day of Receipt of an E-mailed Survey Instrument on Response Rate, Response Time, and Response Quality', *Journal of Extension*, 45.2 (2007), 1 – 7 (p.1).
[897] Brady T. West and Annelies G. Blom, 'Explaining Interviewer Effects: A Research Synthesis', *Journal of Survey Statistics and Methodology*, 5.2 (2017), 175 – 211 (p. 175).
[898] Herbert H. Blumberg, Carolyn Fuller, and A. Paul Hare, 'Response Rates in Postal Surveys', *Public Opinion Quarterly*, 38.1 (1974), 113 -123 (p. 113); Rindfuss, p. 818.
[899] Scott Keeter and others, 'Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey', *Public Opinion Quarterly*, 70.5 (2006), 759 – 779 (p. 759); Robert Groves and Emilia Peytcheva, 'The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-analysis', *Public Opinion Quarterly*, 72.2 (2008), 167 – 189 (p. 167).

A prominent, additional poll-level characteristic that may be expected to affect polling error is survey mode. While the scale of the effect of survey modes on polling error is not yet fully understood,[900] they nevertheless stand as a potential source of systematic polling error.[901] Indeed, impacts on the quality and representativeness of responses received across differing survey modes have been observed,[902] undermining polling accuracy. While the inclusion of mode would be a useful poll-level control for inclusion within later analysis, data regarding it is surprisingly difficult to acquire, especially to the extent required for meaningful inclusion within a large, heterogenous dataset. As such, I do not include it as a poll-level control within models, as its inclusion would significantly decrease the sample of polls available for use in analysis, which present pronounced problems for prediction models. Future studies ought to attempt to incorporate survey mode in further multi-level analyses of polling error. However, for this to be achieved, a pronounced increase in data accessibility will be required.

Additional characteristics that inform the vote share estimates provided by polls can also be expected to affect the error that they present. These include the manner in which non-response is handled, the treatment of non-voters in estimates, and demographic weighting procedures. Though they factor into the estimates provided by polls, such characteristics exist as decisions made by the polling organisations who conduct pre-election polls and specify their methodologies. Due to this, they are better understood as members of the pollster-level grouping of sources of polling error. It is to this grouping level and the control variables housed within it that I turn my attention in the following sub-section.

[900] Mellon and Prosser, p. 20.
[901] Jennings and Wlezien, 'Election Polling Errors Across Time and Space', p. 283.
[902] L. Chang and J. A. Krosnick, 'National Surveys via RDD Telephone Interviewing Versus the Internet: Comparing Sample Representativeness and Response Quality', *Public Opinion Quarterly*, 73.4 (2009), 641 – 678 (p. 641); D. Heerwegh, 'Mode Differences Between Face-to-Face and Web Surveys: An Experimental Investigation of Data Quality and Social Desirability Effects', *International Journal of Public Opinion Research*, 21.1 (2009), 111 – 121 (p. 111).

Pollster-level Control Variables

Differences between polling organisations have long been understood to affect the degree to which pre-election polls are likely to exhibit error. The impact of the many differences in methodology that exist between polling houses have been extensively recognised in past literature.[903] Despite the recognition of their importance, specific information on the methods employed by polling organisations is often unavailable, with past researchers having decried the lack of transparency in this area.[904] Given the difficulty of isolating specific methodological differences between polling organisations, a more general approach to capturing house effects is required.

To capture the impact of house effects on polling error, I employ a categorical variable encompassing different polling organisations. Each polling organisation is assigned its own category to capture the unique constellation of methodological choices, adjustments, and survey techniques it comprises. As the methodological decisions of polling organisations are often not publicly disclosed, this is intended to serve as umbrella variable to capture the myriad differences between polling houses that may go unrecognised in – or be unrecognisable by – more specific measures.

I also measure the number of repeated occurrences of a given polling organisation in a given country to capture the expectation of variable performances between houses. Different polling houses may possess histories of good performance in the context of a given country and, therefore, would be expected to be more likely perform well in future elections within that country relative to other organisations. Though this relationship is not deterministic, as

[903] Pickup and others, p. 209; Christopher Wlezien and Robert S. Erikson, 'The Horse Race: What Polls Reveal as the Election Campaign Unfolds', *International Journal of Public Opinion Research*, 19.1 (2007), 74 – 88 (p. 74); Christopher Wlezien, 'Presidential Election Polls in 2000: A Study in Dynamics', *Presidential Studies Quarterly*, 33 (2003), 172 – 187 (p. 176); Prosser and Mellon, p. 776; Andrew Gelman p. 69.
[904] Stirton, pp. 310 – 313.

different elections present different challenges, and reliance on the existence of path dependencies between the past and future in prediction has long been warned against,[905] its expectation can nevertheless be justified. As polling organisations are often sponsored by media organisations to produce and publish polls,[906] repeated polls conducted by a given organisation in a given country represents continued investment in that organisation and a degree of trust. The greater the number of occurrences of polling organisation in a country, the more it has been invested in and the more it has been trusted to conduct polls. As such, it can be expected to have performed well enough in past elections to justify this investment and have refined its methods for the context of that country over time. This can be expected to lead to improved performance and a reduced likelihood of polling error. By contrast, organisations with fewer occurrences in a given country can be expected to perform less impressively, presenting greater polling error.

Country-level Control Variables

The country in which an election is conducted bears closely on its progression and the nature of its outcome. Factors including its electoral system, political regime, and rules concerning the publication of polling results all serve to affect the framework within which it is conducted, its evolution over time, and the degree to which this evolution can accurately be charted. These factors conspire to make the country in which an election is held an important determinant of electoral outcomes. In so doing, they render country-level effects a plausible source of polling error.

As with previous grouping levels, the impact of country-level differences can be intuited from the varying amounts of polling error variance accounted for by membership within different

[905] Makridakis, Hogarth, and Gaba, p. 84.
[906] Shamir, p. 62.

countries shown in the previous chapter. I take these country-level differences to be any characteristic possessed by a country which varies, or has the potential to vary, between cases. I outline nine country-level characteristics which are likely to affect polling error and are included in later prediction models as controls. While these do not represent the universe of all possible differences between countries which are important for polling error, I contend that they are the most prominent. I separate these characteristics into two categories: microscopic and macroscopic. Microscopic characteristics are those which vary between each country within my dataset. Together, they capture a constellation of traits unique to each country across each of my studied election years. The microscopic characteristics I include in later control models are: the gross domestic product (GDP) and Gini coefficient (GINI) of a country during an election year due to their effect on turnout levels,[907] the presence of a polling moratorium in a country prior to election day due to their impact on the ability of polls to detect last-minute shifts in voting intention,[908] and whether a country employs compulsory voting due to its impact on turnout and the quality of voter decision-making.[909]

By contrast, macroscopic country-level characteristics capture traits which either characterise several countries within my dataset simultaneously or exist as umbrella variables to capture any other meaningful country-level differences missed by the identified microscopic characteristics. The macroscopic characteristics I identify are: the type of electoral system

[907] Christopher J. Anderson and Pablo Beramendi, 'Income, Inequality, and Electoral Participation', in *Democracy, Inequality, and Representation in Comparative Perspective*, ed. by Pablo Beramendi and Christopher J. Anderson, (New York: Russell Sage Foundation, 2008), pp. 279 – 280; Frederick Solt, 'Economic Inequality and Democratic Political Engagement', *American Journal of Political Science*, 52.1 (2008), 48 – 60 (p. 48); Frederick Solt, 'Does Economic Inequality Depress Electoral Participation? Testing the Schattschneider Hypothesis', *Political Behaviour*, 32 (2010), 285 – 301 (p. 285); Robert Goodin and John Dryzek, 'Rational Participation: The Politics of Relative Power', *British Journal of Political Science*, 10.3 (1980), 273 – 292 (p. 284); [907] Henry Brady, Sidney Verba, and Kay Lehman Schlozman, 'Beyond SES: A Resources Model of Political Participation', *American Political Science Review*, 89.2 (1995), 271 – 294 (p. 271); Allan Meltzer and Scott Richard, 'A Rational Theory of the Size of Government', *Journal of Political Economy*, 89.5 (1981), 914 – 927 (p. 925).

[908] Tse-min Lin and Brian Roberts, 'Markets and Politics: The 2000 Taiwanese Presidential Election', in *Topics in Analytical Political Economy*, ed. by Melvin Hinich and William A. Barnett, (Amsterdam: Elsevier, 2007), p. 144.

[909] Peter John Loewen, Henry Milner, and Bruce M. Hicks, 'Does Compulsory Voting Lead to More Informed and Engaged Citizens? An Experimental Test', *Canadian Journal of Political Science*, 41.3 (2008), 655 – 672 (p. 666).

employed within a country due to its effect on partisan loyalty and ENEP;[910] the type of political regime within a country given its impact on both ENEP and turnout;[911] the region and subregion in which a country exists to capture international differences in electoral behaviour, voting culture, and polling practices which exist between areas of the world;[912] and the country itself to serve as an umbrella variable to capture any remaining intranational differences not included within microscopic traits.

Each of these plausible predictors of polling error from the poll, pollster, and country grouping levels is included in later prediction models to ensure the robustness of election-level findings. In the following chapter, I unpack the modelling procedures and accuracy measures used to test the ability of election-level characteristics to aid in the prediction of polling error.

[910] Paul R. Abramson and others, 'Comparing Strategic Voting Under FPTP and PR', *Comparative Political Studies*, 43.1 (2010), 61 – 90 (p. 65); Jennings and Wlezien, p. 225.
[911] Margit Tavits, 'Direct Presidential Elections and Turnout in Parliamentary Contests', *Political Research Quarterly*, 62.1 (2009), 42 – 54 (p. 42).
[912] Michael Marsh, 'Electoral Context', *Electoral Studies*, 21.2 (2022), 207 – 217 (p. 207).

Chapter 6 – From Predictors to Predictions: Assessing the Ability of Election-level Differences to Predict Variation in Polling Error

"Prediction, not narration, is the real test of our understanding of the world".[913]

- Nassim Nicholas Taleb (2007)

In this chapter I address my third research question and explore the degree to which electoral characteristics are able to aid in the prediction of polling error variance. To address this question, I break the chapter down into three sections. In the first section, I outline the modelling procedures and test statistics used to assess the ability of election-level variables to predict polling error. I establish the regression-based nature of my prediction models and explain that they take two forms: additive and interactive. I then unpack the manner in which their performance is measured, identifying that root mean square error (RMSE) is used in the case of models concerning continuous measures of polling error, while the proportion of correct classifications is used for models concerning dichotomous measures. Finally, I address the way in which I approach missing data within my analysis, settling on complete case analysis as the most appropriate method.

In the second section, I provide the outputs of my prediction models. I begin by presenting the outputs of additive prediction models containing only election-level variables. To illustrate the nature of my models and the approach taken to their analysis, I unpack the findings derived from models tasked with the prediction of mean absolute error (MAE), finding election-level variables to be useful in its prediction. I then explore the usefulness of my election-level variables across models tasked with the prediction of my additional measures of distributive, bounded, and substantive inaccuracy, as well as polling bias. Due to the spatial limitations of

[913] Nassim Nicholas Taleb, *The Black Swan: The Impact of the Highly Improbable*, (New York: Random House, 2007), p. 133.

this thesis, this exploration concerns the aggregate findings derived from these models. The individual findings on which these aggregations are based are presented in Appendix B. Through the exploration of aggregate findings, I find initial support for my second hypothesis, as election-level variables improve the ability of models to predict all forms of polling error, doing so to a greater extent in relation to substantive error, and to a diminished extent in relation to bias.

Following my exploration of additive models, I present the outputs from interactive prediction models containing only election-level variables. In keeping with the previous sub-section, I unpack findings relating to MAE, identifying that two- and three-way interactions between these variables largely aid models in its prediction, though do not do so as consistently as the main effects associated with individual election-level variables. I then explore the aggregate ability of these interactions to predict my additional distributive, bounded, and substantive measures of polling error, as well as measures of polling bias. The individual outputs from which these aggregations are derived can be found in Appendix B. Through exploring these aggregate outputs, I find further support for my second hypothesis, as election-level interactions improve the ability of models to predict all forms of polling error, proving most useful in the case of substantive error and least useful in the case of bias.

To ensure that the findings presented by my additive and interactive election-level only models are robust, I conclude the second section of this chapter by presenting the outputs of prediction models containing control variables from the poll, pollster, and country grouping levels. I find that the ability of election-level variables to aid in the prediction of polling error is robust to the presence of controls across each of my measures of distributive, bounded, and substantive polling error, as well as measures of polling bias.

In the third and final section, I explore the impact of alternative modelling approaches on my findings and address emergent themes identified in earlier analysis. I recognise that my main prediction models assume a linear relationship between election-level variables, present a degree of overfitting, and limit interactivity to two- and three-level interactions. To assess whether these limitations substantively affect my findings, I employ additional prediction models using random forest ensemble techniques to permit non-linearity, lasso regression to reduce overfitting through feature selection, and fully interactive regression to permit large-scale interactivity between election-level predictors. I find that election-level variables universally improve the ability of these models to predict polling error, even in the presence of controls, with alternate model specifications generally offering improved performance relative to earlier models based on linear and logistic regression. From this, I conclude that findings concerning the usefulness of election-level variables as predictors of polling error are robust to alternate modelling strategies, though future work is needed to identify optimal prediction model specifications.

I conclude the third section by exploring the impact of differently composed subsets of data on my prediction model findings. I note that the usefulness of election-level variables as predictors of polling error varied across the subsets of data used in my main analyses. I recognise that, in the four-level structure of polling error, election-level factors are nested within countries and, therefore, vary between countries. I therefore isolate a series of theoretically motivated, country-based subsets of polls and test the ability of election-level variables to predict polling error within them. I find that election-level variables improve the ability of models to predict polling error irrespective of the subset of data provided to them, though the degree to which they do so varies. From this, I conclude that the importance of election-level variables as predictors of polling error is robust to differently composed subsets of data, though future scholarship ought to further explore those contexts in which they are more or less useful.

6.1: Modelling Procedures, Test Statistics, and Dealing with Missing Data

To gauge how useful my election-level variables are for predicting variation in polling error, I begin by running a series of stepwise regression models. I employ two forms of stepwise regression-based prediction model: additive and interactive. In additive models, I layer individual election-level variables iteratively, measuring their impact on the ability of the model to accurately predict polling error. In interactive models, I include all main effects associated with individual election-level variables and layer individual two- and three-way interactions on top of these, measuring their effect on the predictive accuracy of the model.

The prediction models I employ within this thesis are principally based on variants of linear regression. The use of linear regression is motivated by the theoretical expectation of broadly linear relationships between electoral characteristics and my measures of polling error established in chapter three. For my continuous measures of polling error, predictions are based on multiple linear regression using ordinary least squares (OLS), while for my binary measures of error, predictions are rendered using generalised linear models with a logit link function. The use of classic parametric statistical methods, such as OLS, is based on a series of assumptions that are rarely satisfied by real-world data.[914] Indeed, many of my measures of polling error do not perfectly conform to the assumption of normally distributed data. The violation of such assumptions can result in inaccurate p-values, coefficients, and confidence intervals in traditional statistical inference.[915] However, as my models concern predictive accuracy, and are therefore not concerned with drawing inferences about the nature of statistical relationships,[916] their ability to be meaningfully interpreted in the presence of assumption violations is more robust. Nevertheless, to ensure the reliability of findings drawn

[914] David M, Erceg-Hurn and Vikki M. Mirosevich, 'Modern Robust Statistical Methods: An Easy Way to Maximize the Accuracy and Power of Your Research', *American Psychologist*, 63.7 (2008), 591 – 601 (p. 591).
[915] Ibid.
[916] Leo Breiman, 'Statistical Modeling: The Two Cultures', *Statistical Science*, 16.3 (2001), 199 – 231 (p. 199).

from parametric models, I employ a series of non-parametric models later in the chapter as a robustness check.

The additive linear models used to render predictions across my principal models are based on equation 27,[917] where predicted values for a measure of polling error, $\hat{y}$, are modelled as a linear function of the additive sum of the individual effects, $\hat{\beta}_1$ ... $\hat{\beta}_n$, of predictor variables x_1 ... x_n plus some unknown error $\in$. In these models, the effect of each predictor variable is independent of all other predictors.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_n x_n + \in \quad (27)$$

In my additive linear models, ordinary least squared regression is used to identify a linear relationship between my continuous measures of polling error and a range of predictor variables within a subset of training data. This relationship is then applied to unseen, out-of-sample data to produce predicted values of these measures of polling error on the basis of a new set of predictor values. I measure the accuracy of my prediction models in terms of their out-of-sample performance. That is, I measure their ability to accurately predict unseen polling error values. This involves splitting my data into training and testing subsets. Models are trained on training subsets and the relationships observed within these subsets are then applied to unseen data within testing subsets. Model performance concerns how well the relationships observed within the training data generalise to unseen test data. This process seeks to replicate the real-world application of prediction models which use known data to predict unknown future outcomes.

[917] Rob J. Hyndman and George Anthanasopoulos, *Forecasting: Principles and Practice*, (Melbourne: Otexts, 2018), p. 169.

In train/test splitting, the majority of data is held back to train a model which is then tested on an unseen minority of data.[918] The performance of a model measured in relation to an individual train/test split may simply be an artefact of the composition of these subsets and, therefore, may not be truly representative of the out-of-sample performance of a prediction model.[919] To account for this, predictive accuracy is ideally calculated across a series of train/test splits and averaged.[920] Taking an average mitigates the idiosyncrasies of individual splits and provides a more reliable measure of model accuracy.[921]

A range of resampling approaches exist to provide aggregate measures of prediction error.[922] Of these approaches, k-fold cross validation and its repeated variant have been found to perform better than alternatives.[923] K-fold cross validation splits a dataset into *k* non-overlapping, roughly equal subsets or folds and serves as a form of train/test splitting[924] The value of *k* is typically set at 5 or 10.[925] Each of the *k* folds is iteratively held back as a test set to gauge the accuracy of models trained on the remaining *k-1* folds.[926] Through this process, all available data is used in both training models and testing their performance, providing a representative measure of predictive performance that is not unduly influenced by any given train/test split.

While simple k-fold cross validation provides a more dependable measure of predictive accuracy than an individual train/test split, its estimates often exhibit high variance due to the

---

918 Borislava Vrigazova, 'The Proportion for Splitting Data into Training and Test Set for the Bootstrap in Classification Problems', *Business Systems Research*, 12.1 (2021), 228 – 242 (p. 228).
919 Avrim Blum, Adam Kalai, and John Langford, 'Beating the Hold Out: Bounds for K-fold and Progressive Cross-validation', *Proceedings of the 12th Annual Conference on Learning Theory*, 1 (1999), 1 – 6 (p. 1).
920 Ibid.
921 Ibid.
922 Gaoxia Jiang and Wenjian Wang, 'Error Estimation Based on Variance Analysis of K-fold Cross-validation', *Pattern Recognition* 69 (2017), 94 – 106 (p. 94).
923 Ibid.
924 Kuhn and Johnson, p. 69.
925 Ibid., p. 70.
926 Ibid., pp. 69 – 70.

differing composition of the folds.[927] This variance can be reduced by increasing the number of folds used in cross-validation.[928] However, as the number of folds used in k-fold cross validation grows larger, so too does the difference in size between the training and test subsets, increasing the bias presented by estimates.[929]

To resolve the bias-variance trade-off, repeated k-fold cross validation can be used.[930] Rather than increasing *k* and introducing bias, this approach holds its value constant and repeats cross-validation across *k* newly partitioned folds *n* times.[931] Through this repetition, it stabilises the aggregate error estimate, mitigating issues of variance.[932]

In light of the benefits of aggregation, I employ repeated k-fold cross validation when measuring the accuracy of my prediction models and the impact of predictor variables upon it. Specifically, I employ repeated 10-fold cross validation. In repeated 10-fold cross validation, data is divided into 10 roughly equal sets with each set iteratively held back to test a model trained on the remaining nine sets. The out-of-sample predictive performance of the model is averaged over these 10 sets to provide an aggregate measure of accuracy. This process is visualised in Figure 29 below.

[927] Matthew Tuson and others, 'Predicting Future Geographic Hotspots of Potentially Preventable Hospitalisations Using All Subset Model Selection and Repeated k-fold cross validation', *International Journal of Environmental Research and Public Health*, 18 (2021), 1 – 21 (p. 6).
[928] Ibid.
[929] Kuhn and Johnson, p. 70.
[930] Ji-Hyun. Kim, 'Estimating Classification Error Rate: Repeated Cross-Validation, Repeated Hold-out and Bootstrap', *Computation Statistics and Data Analysis*, 53.11 (2009), 3735 – 3745 (p. 3735).
[931] Kuhn and Johnson, p. 70.
[932] Jiang and Wang, p. 95.

Figure 29: Illustration of the repeated 10-fold cross validation process taken from Sontakke and others.[933] The data set is divided into 10 roughly equal folds with each fold iteratively held back to test a model trained on the remaining 9 folds. This process is repeated 10 times with each instance of cross validation producing a measure of predictive error ($E_{1\dots10}$). The sum of these error values is then averaged to provide an aggregate measure of predictive accuracy (E).

The key test statistic for assessing the ability of election-level variables to predict variation in my continuous measures of polling inaccuracy is root mean square error (RMSE). RMSE measures the degree to which predicted values deviate from observed results within a model by taking the square root of the variance of the residuals.[934] It is therefore a measure of the degree to which predicted values fit observations and, as such, assesses the degree to which variance in observed values is accounted for by models.[935] RMSE is calculated as shown in

[933] Sumedh Sontakke and others, ‘Classification of Cardiotocography Signals Using Machine Learning’, in *Intelligent Systems and Applications*, ed. By Kohei Arai, Supriya Kapoor, and Rahul Bhatia, (Cham: Springer, 2018), p. 442.
[934] Dulakshi Karunasingha, ‘Root Mean Square Error or Mean Absolute Error? Use Their Ratio as Well’, *Information Sciences*, 585 (2022), 609 – 629 (p. 610); Murodjon Sultanov and others, ‘Modelling End-of-season Soil Salinity in Irrigated Agriculture Through Multi-Temporal Optical Remote Sensing, Environmental Parameters, and In Situ Information’, *Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 86.5 (2018), 221 – 233 (p. 227).
[935] Micah Russell and others, ‘Toward a Novel Laser-based Approach for Estimating Snow Interception’, *Remote Sensing*, 12 (2020), 1146 – 1157 (p. 1151); Sultanov and others, p. 227.

equation 28 by taking the square root of the average of the sum of squared differences between predicted values of an outcome variable, $\hat{y}_i$ ... $\hat{y}_n$, and its observed values, y_i ... y_n.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{28}$$

High levels of RMSE represent larger distances between predicted and observed values, representing worse predictive accuracy and poorer explanation of variance, while lower levels of RMSE represent smaller distances between predictions and observations, indicating improved predictive accuracy and stronger explanation of variance. Reductions in RMSE values therefore represented improvements in the ability of a model to accurately predict error values and, consequently, better explain their variance. I measure RMSE values to three significant figures so as not to mask small improvements in the ability of models to accurately predict polling error variance.

To assess the utility of my predictor variables in models concerning variance across my binary measures of polling error, an alternative approach is required. While my predictions concerned with continuous outcome variables are derived form of regression-based models, those concerned with binary outcome variables are derived from classification models. Rather than seeking to predict a continuous value, they seek to correctly classify a binary outcome.[936] Accordingly, accuracy is measured in terms of the proportion of correct classifications made by a model.[937] The greater the correct classification rate, the more accurately a set of predicted points resembles observed values and, therefore, accounts for its variance. I therefore measure the ability of election-level variables to predict variation in my binary measures of polling error

[936] Vrigazova, p. 228.
[937] Ibid.

by assessing their impact on the rate of correct classifications made by the models in which they are included.

I use repeated 10-fold cross validation to generate aggregate measures of both the RMSE and the proportion of correct classifications presented by my prediction models. Through this process, the accuracy of each individual model specification, and therefore the impact of each individual variable or group of variables, is measured as an average across 100 instances of cross validation.

If my second hypothesis holds and election-level variables prove useful for predicting polling error, their inclusion within prediction models will result in reduced RMSE and an increased proportion of correct classifications, respectively. To test my hypothesis, I begin by assessing the impact of my election-level variables on predictions of polling error in isolation. In the following subsection, I describe the format of my election-level prediction models and the nature of the variables included within them before decomposing their impact on predictive accuracy.

Balancing Unbalanced Data for use in Classification Models

While repeated k-fold cross validation alone is sufficient to produce reliable measures of out-of-sample accuracy for predictions concerning my continuous measures of polling error, my binary measures require additional data processing. Prediction models concerning my binary measures of polling error, SBP and LVRC, are better understood as classification models based on logistic regression. A degree of imbalance between categories is likely in all real-world classification problems, as the distribution of categorical outcomes is rarely symmetric. Imbalance presents problems for prediction models.[938] Imbalanced data is a problem when rendering predictions via classification models, as most classification algorithms are based on

[938] Alberto Fernández and others, *Learning from Imbalanced Data Sets* (Cham: Springer, 2018), p. 16.

the assumption of balanced data,[939] with imbalanced data severely affecting their performance.[940] Optimal classification performance requires balanced classes,[941] with imbalanced data biasing prediction models towards the majority class.[942] In the case of SBP, the majority class is 1, indicating that a poll was significantly biased. The algorithm will therefore be biased towards predicting the presence of error. This undermines the validity of measures of their accuracy, such as correct classification rate.[943]

As prediction algorithms, classifiers principally focus on relationships associated with instances of the majority class in cases of imbalanced data.[944] For my two binary measures of polling error, this disproportionality presents differing problems. For SBP, the imbalance favours positive values. As such, the algorithm will learn disproportionately more about instances of biased polling than instances of unbiased polling. For LVRC, the imbalance overwhelmingly favours negative values. This leads the algorithm to learn a considerable amount about the absence of error and comparatively far too little about the presence of error. As the aim is to predict the presence of error, this is problematic. In both cases, the prediction algorithm will be biased, albeit in opposing directions, undermining the degree to which measures of accuracy can be trusted.[945]

In addition to issues concerning the reliability of predictions, applying *k*-fold cross validation to severely imbalanced data is not a sensible analytical approach. In cases of severe imbalance, instances of one category to be classified (the majority class) considerably outnumber those of

[939] Pradeep Kumar and others, 'Classification of Imbalanced Data: Review of Methods and Applications', *IOP Conference Series: Materials Science and Engineering*, 1099 (2021), 1 – 20 (p. 3); Wonjae Lee and Kangwon Seo, 'Downsampling for Binary Classification with a Highly Imbalanced Dataset Using Active Learning', *Big Data Research*, 28 (2022), 1 – 19 (p. 1).
[940] Joonho Gong and Hyunjoong Kim, 'RHSBoost: Improving Classification Performance in Imbalance Data', *Computational Statistics and Data Analysis*, 111 (2017), 1 – 13 (p. 3); Lee and Seo, p. 1.
[941] Lee and Seo, p. 2.
[942] Gong and Kim, p. 2.
[943] Ibid., pp. 1 – 2.
[944] Gong and Hyunjoong, p. 2.
[945] Ibid.

the other (the minority class). Given that *k*-fold cross validation is based on splitting data into training and testing subsets, this means that certain subsets of data may only contain elements of the majority class, given its overwhelming presence in the data. This would lead to the outcome variable existing as a constant in these subsets, undermining the ability of models to predict it due to its lack of variance. Indeed, in such situations, an intercept-only predictive model would perfectly predict *y* for every given *x*, as taking the mean of a constant yields the constant itself.

To correct for the imbalance within my data, I use downsampling. Though several approaches exist,[946] downsampling ensures that predictions are run across real-world datapoints, as opposed to the duplicated or artificially generated datapoints born of alternative approaches (see, for example, upscaling).[947] In downsampling, instances of the majority class (in this case values of 1) are randomly dropped from the dataset until their number is equal to that of the minority class.[948] This results in a balanced dataset. While downsampling necessarily results in a degree of information loss due to the omission of datapoints,[949] the oversampling present in alternative approaches has been found to worsen the performance of classifiers.[950] As downsampling has not been found to present these performance drawbacks, acquitting itself well across performance metrics,[951] I use it to deal with the imbalance present in my data.

946 Andrea Dal Pozzolo, Olivier Caelen, and Gianluca Bontempi, 'Comparison of Balancing Techniques for Unbalanced Dataset', *Mach. Learn. Gr. Univ. Libr, Bruxelles Belgium*, 16.1 (2010), 732 – 735 (p. 732).
947 Ibid.
948 Max Kuhn, 'Building Predictive Models in R Using the Caret Package', *Journal of Statistical Software*, 28.5 (2008), 1 – 26 (p. 14).
949 Seba Susan and Amitesh Kumar, 'The Balancing Trick: Optimized Sampling of Imbalanced Datasets – A Brief Survey of the Recent State of the Art', *Engineering Reports*, 3.4 (2021), 1 – 24 (p. 8).
950 Ibid.
951 Shivani Tyagi and Sangeeta Mittal, 'Sampling Approaches for Imbalanced Classification Problem in Machine Learning', in *Proceedings of ICRIC 2019,* ed. by Pradeep Kumar Singh and others (Cham: Springer, 2020), pp. 209 – 221; Chris Drummond and Robert C. Holte, 'Class Imbalance and Cost Sensitivity: Why Under-sampling Beats Oversampling', *Workshop on Learning from Imbalanced Datasets*, 11 (2003), 1 – 8 (p. 8).

While a single downsampling pass results in a balanced dataset, it does not necessarily result in a dataset that is predictively useful. As instances of the majority class are dropped, a singular downsampled dataset is unlikely to be representative of the relationship present in the larger dataset from which it was drawn.[952] As understanding these relationships rests at the core of machine learning-based prediction, this is not ideal, but can be mitigated. Given that the process of dropping majority class values from the data is random, creating multiple downsampled subsets yields differently composed subsets of data. These subsets will therefore represent the relationship between the majority class values and the outcome variable differently. Aggregating across a large number of such subsets reduces information loss,[953] allowing a greater number of instances of the majority class to be included and, in so doing, more closely representing the relationships present in the original, larger dataset. Ideally, all values of the majority class will be represented at least once across the aggregated subsets to ensure that predictions are informed by the full range of original information. While creating multiple subsets of randomly differing composition will result in the over-representation of some values relative to the parent dataset, as their composition is determined at random, the process will not be biased towards any given value or set of values, as all values possess the same probability of inclusion in any given subset.

The process required to determine the number of downsampled subsets necessary to account for each value in the majority class at least once is as variation of the coupon collector's problem from probability theory with group draws of fixed size,[954] often referred to as the

[952] Che Ngufor and Janusz Wojtusiak, 'Learning from Large Distributed Data: A Scaling Down Sampling Scheme for Efficient Data Processing', *International Journal of Machine Learning and Computing*, 4.3 (2013), 216 (p. 216).
[953] X. Y. Liu, J. Wu, and Z. H. Zhou, 'Exploratory Undersampling for Class-Imbalance Learning', *IEEE Transactions on Systems, Man, and Cybernetics*, 39.2 (2009), 539 – 550 (pp. 545 – 546).
[954] Wolfgang Stadje, 'The Collector's Problem with Group Drawings', *Advances in Applied Probability*, 22.4 (1990), 866 – 882 (p. 868); Marco Ferrante and Monica Saltalamacchia, 'The Coupon Collector's Problem', *Material Matemátics*, 2.35 (2014), 1 – 35 (p. 18).

sticker collector's problem.[955] Downsampling is typically conceived of as a process that drops at random *n* instances of the majority class, where *n* is equal to the difference in size between the majority and minority classes.[956] However, as the purpose of downsampling is to reduce the number of instances in the majority class in an unbiased manner such that it is equal in size to the minority class,[957] it is functionally equivalent to conceive of it as a process of random value retention. Under this understanding, *n* values from the majority class are retained at random, where *n* is equal to the number of values in the minority class. As each value in the majority class is unique and therefore only occurs once, values must be removed from the population once sampled to prevent unrepresentative duplication within the subset. Given this, producing a downsampled subset is tantamount to randomly drawing a sample of fixed size *n* from the majority class without replacement where *n* is the size of the minority class.

Under the understanding of downsampling as a process of random value retention, I use the Monte Carlo simulation approach employed by Diniz et al. to determine the average number of subsets needed to account for each majority class value at least once.[958] This approach involves repeatedly drawing samples of a fixed size without replacement from a population and noting the values drawn in each pass. Samples are taken until all discrete values in the population have been drawn at least once. This process is repeated *n* times and the average number of samples required is reported.[959] In my analysis, I take the average of 1,000 iterations of this resampling process due to its widespread use as a reliable statistical benchmark.[960] I use

[955] Márcio Diniz and others, 'The Sticker Collector's Problem', *The College Mathematics Journal*, 47.4 (2016), 255 – 263 (p. 255).
[956] Sofia Visa and Anca Ralescu, 'Issues in Mining Imbalanced Data Sets – A Review Paper', *Proceedings of the Sixteenth Midwest Artificial Intelligence and Cognitive Science Conference*, 1 (2005), 1 – 7 (p. 2).
[957] Ibid.
[958] Diniz and others., p. 256.
[959] Ibid.
[960] Peter C. Austin and Jack. V. Tu, 'Bootstrap Methods for Developing Predictive Models', *The American Statistician*, 58.2 (2004), 131 – 137 (p. 133).

each downsampled subset to perform 100 out-of-sample predictions using repeated k-fold cross validation.

Dealing With Missing Data in Prediction Models

Missing values present problems for predictive modelling which vary according to the nature of their missingness.[961] Values can either be missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR).[962] For data to be MCAR, the cause of its missingness must be unrelated to information contained within a dataset. The probability of missingness is therefore the same for all units of analysis.[963] Conversely, for data to be MAR, its missingness must be directly related to information present within a dataset.[964] To this end, MAR is a confusing misnomer, as values are not missing in a random manner, but missingness is rather distributed in a deliberate manner that is dependent on existing data. Finally, for data to be MNAR, its missingness must not only be a function of data contained within the dataset, but also unmeasured factors exogenous to the data.[965]

Given the completely random and unconnected nature of their missingness, MCAR values can simply be removed without biasing results.[966] Values that are MAR or MNAR present issues for statistical models, given the non-random nature of their missingness and its causes.[967] If MAR or MNAR values are present within data, they must be dealt with prior to running

[961] Sarah Fletcher Mercaldo and Jeffrey D. Blume, 'Missing Data and Prediction: The Pattern Submodel', *Biostatistics*, 236 – 252 (p. 236).
[962] Donald. B. Rubin, 'Inference and Missing Data', *Biometrika*, 63.3 (1976), 581 – 590 (p. 581); Andrew Gelman and Jennifer Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*, (Cambridge: Cambridge University Press, 2006), pp. 530 – 531.
[963] Gelman and Hill, p. 530.
[964] Ibid.
[965] Joost R. van Ginkel and others, 'Rebutting Existing Misconceptions About Multiple Imputation as a Method for Handling Missing Data', *Journal of Personality Assessment*, 120.3 (2020), 297 – 308 (p. 298).
[966] Gelman and Hill, p. 530.
[967] Ibid.

statistical models in order to ensure not only that these models will function, but that produce the most reliable outputs possible given the data provided.

Within my polling data, three key predictor variables present missing values: partisanship, late decision-making, and the left/right standard deviation of the political position of parties/candidates. The missing values associated these variables exist as a function of a variety of variables that are variously endogenous and exogenous to my data. Of the variables observed within my data, the missingness in my predictors is partially a function of the country identifier variable within my dataset, as values are present for certain countries and absent for others. This is an artefact of the nature of pre-election polling in given countries and the variable presence of questions concerning the strength of partisan loyalty and the timing of voter decision-making. Partisanship is also associated with levels of voter turnout, with a greater number of voters turning out in elections characterised by stronger partisan loyalty within the electorate,[968] while late decision-making can be conceived of a function of both the effective number of parties in an election due to its impact on the range of choices presented to the electorate, as well as snap elections due to their shorter campaign lengths.

Of those variables not observed within my data, the missing values are also partially a function of time, as values associated with them have been captured in some years but not in others. For example, in British Election Study data, questions regarding partisan loyalty were not asked prior to 1970.[969] Supplemental cross-national election studies are often sporadic in their geographical coverage over survey waves.[970] This leads to further intermittence in values over time. Similarly, the Chapel Hill Expert Survey, which exists as the source of my left/right

[968] Rau, p. 1021.

[969] British Election Study, *1963 – 1970 Political Change in Britain* (2022), <https://www.britishelectionstudy.com/data-object/1963-1970-political-change-in-britain/> [accessed 16/08/2022].

[970] CSES, *Election Studies* (2022), <https://cses.org/data-download/download-data-documentation/election-studies/> [accessed 17/08/2022].

standard deviation measurements, does not provide data prior to 1999 and only provides data in waves once every four years.[971] Time is not substantively captured within my dataset, as there is little evidence to suggest that polling error exists as a function of it.[972]

Beyond time, the missingness in my data is also a function of a range of further variables that are exogenous to my data. Broadly, values for partisanship and late decision-making are more likely to be missing in those countries with less well-funded, less established polling industries or those countries not the subject of national or cross-national election studies. Variables capturing these characteristics are not present within my dataset, as information pertaining to them is either difficult to obtain (such as funding figures) or not immediately pertinent to polling error (such as the development of polling industries, as it is a function of time).

More specifically, whether electoral candidates have previously engaged in party switching have been found to affect the strength of partisan loyalty,[973] as has whether a legislature is primarily composed of newer or seasoned legislators.[974] Perceptions of corruption and retrospective analyses of political performance on the part of voters have also been found to determine levels of partisan loyalty.[975] None of these variables are present within my dataset. Similarly, late decision-making amongst the electorate is a function of the variable issue- or candidate-orientated nature of voters.[976] The sex and age of voters has also been identified as determinants of late decision-making,[977] along with levels of political disaffection amongst the

[971] Chapel Hill Expert Survey, *1999-2019 Chapel Hill Expert Survey (CHES) Trend File* (2022), <https://www.chesdata.eu/1999-2019chestrend> [accessed 17/06/2022].
[972] Jennings and Wlezien, 'Election Polling Errors Across Time and Space', p. 280.
[973] Gherghina and Chiru, p. 536.
[974] Ibid.
[975] Mihail Chiru and Sergiu Gherghina, 'When Vote Loyalty Fails: Party Performance and Corruption in Bulgaria and Romania', *European Political Science Review*, 4.1 (2012), 29 – 49 (p. 49).
[976] Jan Eric Blumenstiel and Thomas Plischke, 'Changing Motivations, Time of the Voting Decision, and Short-term Volatility – The Dynamics of Voter Heterogeneity', *Electoral Studies*, 37 (2015), 28 – 40 (p. 28).
[977] Willocq, p. 53.

electorate.[978] As was the case with partisanship, these determinants are not present within my dataset.

While the determinants of partisan loyalty and late decision-making are variously endogenous and exogenous to my data, the determinants of the left/right standard deviation of the political position of parties/candidates are wholly exogenous to it. While the likelihood of large-scale ideological differences between contestants may be considered a function of time, with the major political parties in countries positioning themselves at various distances from one another across the years to attract voters or conform to the prevailing political zeitgeist, the difference in their political positions is principally a function of their policies. The individual policy positions of the parties or candidates contesting my studied elections are contained within my dataset, as their aggregate scores are taken directly from the Chapel Hill Expert Survey.[979]

Given that the missing values within my predictor variables exist as a function of both variables endogenous and exogenous to my dataset, they are MNAR, as the missingness cannot be modelled solely as a function of observed data. As the missing values associated with my predictors are MNAR, they are non-ignorable and need to be dealt with before the variables can be used within prediction models. While a range of methods exist to deal with missing data,[980] the most prominent approaches are complete case analysis (CCA) and multiple imputation (MI).[981] CCA involves listwise deletion where cases are dropped from the dataset if they possess a missing value for a specified variable.[982] By comparison, MI is a Monte Carlo technique in which *n* values are simulated on the basis of existing relationships between the

[978] Ibid., p. 62.
[979] Jolly and others, p. 102420.
[980] S. W. J. Nijman and others, 'Missing Data is Poorly Handled and Reported in Prediction Model Studies Using Machine Learning: A Literature Review', *Journal of Clinical Epidemiology*, 142 (2022), 218 – 229 (p. 218).
[981] Ibid.
[982] Ibid.

variable presenting missingness and other predictors in the dataset.[983] These values are then aggregated to provide estimates to replace missing entries.[984]

Given that CCA omits individual observations and can result in the removal of entire cases in instance of widespread missingness, it results in models being run over a reduced subset of data.[985] This process of reduction has the potential to bias results due to the omission of data.[986] However, CCA is capable of producing unbiased results when removing MAR or MNAR values in the case of regression-based models if the probability of a case being complete (i.e., missing no values) is conditionally independent of the outcome variable, y, given the inclusion of the set of predictors, $\{x_1 \dots x_n\}$, responsible for rendering values MAR or MNAR within models.[987] This is formalised in equation 29.

$$\text{CCA} = \text{unbiased if P(complete case)} \perp\!\!\!\perp y \mid \{x_1 \dots x_n\} \qquad (29)$$

Though my principal prediction models are regression-based, it is necessarily impossible to include the full set of predictor variables responsible for rendering values that are MNAR within model specifications, as many of these variables are exogenous to the dataset. As such, using CCA to analyse subsets of data has the potential to bias results.

While MI is capable of producing less biased results than CCA in the case of data that is MNAR,[988] this is far from certain, with results derived using MI often being found to be more

983 Joseph L. Schafer, 'Multiple Imputation: A Primer', *Statistical Methods in Medical Research*, 8 (1999), 3 – 15 (p. 3).
984 Patrick Royston, 'Multiple Imputation of Missing Values', *The Stata Journal*, 4.3 (2004), 227 – 241 (p. 228).
985 Ibid.
986 Jonathan Sterne and others, 'Multiple Imputation for Missing Data in Epidemiological and Clinical Research: Potential and Pitfalls', *The British Medical Journal*, 338 (2009), 1 (p. 1); Ofer Harel and others, 'Multiple Imputation for Incomplete Epidemiologic Studies', *American Journal of Epidemiology*, 187.3 (2018), 576 – 584 (p. 578).
987 Rachael A. Hughes and others, 'Accounting for Missing Data in Statistical Analyses: Multiple Imputation is Not Always the Answer', *International Journal of Epidemiology*, 48.4 (2019), 1294 – 1304 (p. 1294); Gelman and Hill, p. 530.
988 van Ginkel and others, p. 302.

biased than those derived from CCA when data is MNAR.[989] In either case, both approaches result in bias when dealing with data that is MNAR.[990] As both CCA and MI produce biased results when data is MNAR,[991] the missingness mechanism alone is insufficient to determine how best to deal with missing data. Ultimately, no solution works perfectly in the case of data that is MNAR.[992] As such, a context-specific assessment of the nature of missing values and their ability to be defensibly estimated on the basis of observed data is required.[993] Undertaking such an assessment of my data reveals that many of the assumptions underpinning MI do not hold and that the nature and distribution of missing values within it precludes their resolution using estimation-based techniques.

Estimation techniques, such as MI, seek to fill in missing values on the basis of existing relationships within data.[994] This process is predicated on the assumption that these relationships serve as a suitable basis on which to estimate missing values. In the case of my data, this assumption does not hold. The problem is two-fold. The first issue concerns the distribution of the missing values within my data. Values for partisanship, late decision-making, and the left/right standard deviation of political parties/candidates are missing entirely for certain countries, though are available for others. As MI uses the relationship between observed instances of an incomplete variable and other covariates within the dataset to fill in its missing values,[995] the determination of this relationship will necessarily draw on data from countries in which values for the incomplete variables are available. The application of this relationship to produce estimates to replace missing values is therefore predicted on the

[989] Thomas B. Pepinsky, 'A Note on Listwise Deletion Versus Multiple Imputation', *Political Analysis*, 26.4 (2018), 480 – 488 (p. 480).
[990] Ibid.
[991] Ibid.
[992] Kristin L. Sainani, 'Dealing with Missing Data', *PM&R*, 7.9 (2015), 990 – 994 (p. 990).
[993] Ibid.
[994] Gelman and Hill, p. 530.
[995] Ibid.

determinants of the variables that present missingness remaining constant, or at least closely comparable, between cases. This assumption must hold for imputed values to be dependable.

A brief exploration of my data shows that the assumed cross-comparability of relationships does not hold. To illuminate this finding, I address the lack of cross-case consistency in theoretically motivated relationships between incomplete variables and covariates within my data, beginning with the expected relationship between partisanship and turnout. Though higher levels of turnout may be considered a predictor of stronger partisan loyalty amongst electorates,[996] this relationship is inconsistent between countries. The inconsistent relationship between these variables is displayed in Table 21.

Table 21: A comparison of the average ENEP and extent of late decision-making across those countries within my dataset for which data was jointly available.

Country	Average Strong Partisan Loyalty (%)	Average Voter Turnout (%)
Austria	16.53	76.04
Belgium	26.01	89.85
Brazil	28.32	81.47
Czech Rep.	8.79	66.80
Dominican Rep.	70.26	56.11
Finland	14.55	70.01
Kenya	57.69	79.56
Norway	54.04	78.92
Poland	4.69	54.23
Portugal	9.61	60.06
Romania	21.58	47.92
South Africa	44.84	68.87
Switzerland	11.15	47.59
UK	20.78	69.39

996 Rau, p. 1021.

From Table 21, it is clear that no consistent relationships exists between average strong partisan loyalty and average voter turnout. Certain countries present low levels of strong partisan loyalty, yet exhibit high levels of voter turnout. However, others present high levels of strong partisan loyalty and relatively low levels of turnout. The variability in the relationship between partisanship and turnout is such that there is insufficient consistency across cases for multiple imputation to generate reliable values for missing data in the case of one country on the basis of observed relationships in others.

Similarly, late decision-making amongst the electorate may be considered a function of the number of parties contesting an election and the tyranny of choice accompanying it. However, no consistent relationship exists between these variables across countries. To illustrate this, Table 22 displays the ENEP and late decision-making across a range of countries that serve to demonstrate that lack of consistent cross-country relationships.

Table 22: A comparison of the average ENEP and extent of late decision-making across those countries within my dataset for which data was jointly available.

Country	Average ENEP	Average Late Decision-making (%)
Belgium	6.48	30.15
Canada	3.51	27.18
Denmark	5.34	21.26
Finland	4.32	14.25
Germany	4.00	13.04
Japan	4.27	22.11
Netherlands	4.91	31.75
New Zealand	3.25	20.48
Norway	4.72	25.92
South Africa	2.92	10.70
Spain	4.23	29.70
Sweden	5.17	24.77
Switzerland	6.28	39.86
UK	3.29	26.84
USA	2.10	18.10

From Table 22, it is clear that, on average, widespread late decision-making is accompanied by a large number of effective electoral parties in certain countries but is accompanied by a low number of parties in others. Equally, low levels of late decision-making are accompanied by a low number of effective electoral parties in certain countries and a relatively high number of parties in others. Given the lack of a consistent relationship across cases, it would be inappropriate to impute missing values for one country on the basis of the values presented by another.

Late decision-making may also be considered a function of partisan loyalty, existing to a greater extent in elections characterised by fewer partisan loyalists.[997] Table 23 displays the relationship between the average proportion of strong partisan loyalty in the electorate and the average extent to which voters exhibit late decision-making across a range of countries.

Table 23: A comparison of the average strength of partisan loyalty and extent of late decision-making across those countries within my dataset for which data was jointly available.

Country	Average Strong Partisan Loyalty (%)	Average Late Decision-making (%)
Belgium	26.01	30.15
Canada	19.24	27.18
Denmark	48.92	21.26
Finland	14.55	14.25
Germany	11.88	13.04
Japan	27.75	22.11
Netherlands	24.75	31.75
New Zealand	12.95	20.48
Norway	50.04	25.92
South Africa	44.84	10.70
Spain	20.17	29.70
UK	20.78	26.84
USA	35.80	18.10

[997] Rudiger Schmitt-Beck and Julia Partheymuller, 'Why Voters Decide Late: A Simultaneous Test of Old and New Hypotheses at the 2005 and 2009 German Federal Elections', *German Politics*, 21.3 (2012), 299 – 316 (p. 299).

From Table 23, no clear cross-case relationship is present. On average, low levels of late decision-making are accompanied by strong partisanship in certain countries and weak partisanship in others. Equally, high levels of late decision-making are accompanied by weak partisanship in certain countries, such as Switzerland, and strong partisanship in others.

Exploring these theoretically motivated examples suggests that the generalisable, cross-case relationships required for the imputed values generated by MI to be dependable are not present within my data. This renders the estimation of missing values in one country on the basis of trends observed in another problematic.

The second issue precluding the reasonable use of MI concerns its reliance on predictors that solely exist within the dataset itself. For imputed values to be useful, the relationships on which they are based must be well-suited to estimating them.[998] As these relationships are based on the variables present in a dataset, their value is contingent on their estimative capacity. Not only are theoretically motivated relationships within my dataset inconsistent between countries, but my dataset does not capture a variety of important predictors for partisanship, late decision-making, and the left/right standard deviation of political parties/candidates. After all, this was not its purpose. Due to this, many important determinants of missing values are exogenous to my data and therefore unavailable for use in MI. As such, my data does not contain a sufficient array of covariates to defensibly estimate the missing values within it.

Given that theoretically motivated relationships are not generalisable across cases and that important covariates are not present, MI is not well suited to resolving the missingness within my data. As the missing values cannot reasonably be estimated on the basis of the data provided, I use CCA within later prediction models that include data relating to partisanship, late decision-making, and the left/right standard deviation of the political positions of

[998] Gelman and Hill, p. 539.

parties/candidates. While I remain cognisant of its potential to introduce bias, the impact of this bias will vary. Those models using all observations within the dataset will not suffer from bias, while those concerning incomplete variables will present bias contingent on the extent of the missingness associated with these variables.

To explore the impact of any potential bias introduced by CCA, section 5 of this chapter decomposes the effect of running prediction models across different subsets of countries and Appendix B5 further discusses the effect of differing data compositions on predictions. The appendix also demonstrates that resolving the presented issues of missingness is not simply a case of gathering more data, as the required data simply does not exist. Indeed, these issues of missingness as well as issues of data homogeneity faced later in the thesis are a product of the current state of global polling and are insoluble at this present moment in time.

Before moving on, it is important to note that while CCA informs the analysis within this chapter and its associated appendix, no missingness is present in any of my outcome variables, nor is it present within the grouping-level indicators used in the previous chapter. As such, it was not necessary to contend with missing values in previous analyses.

6.2: Prediction Model Outputs

To initially assess the impact of each of my election-level predictor variables, I run a series of additive prediction models based on multiple linear regression. Though these models do not account for any possible interactions between variables, or the mitigating impact of control variables from different grouping-levels, they nevertheless provide an initial indication of the predictive usefulness of individual election-level variables. As such, these additive models serve as the first step towards establishing the importance of election-level variables as predictors of polling error.

The election-level predictor variables contained within my models can be split into two distinct sets: those which can be known ahead of election day, and those which can only be established after the event. I label these categories as ex-ante and ex-post, respectively. Of my 13 specified election-level variables, turnout, the change in turnout between contests, the margin of victory, ENEP, and ENEP change can only be known after an election has been contested and are therefore ex-post in nature. All remaining election-level variables can be known before an election has concluded, rendering them ex-ante.

While I contend that both sets of election-level variables are of likely importance for the prediction of polling error, the impact that they have upon prediction models must be considered differently. As ex-ante variables can be known ahead of time, they can be used to predict the likelihood of error variance before an election occurs. In this sense, they can be used in a traditionally forward-facing predictive manner. However, as ex-post variables can only be known after the fact, they can only be used to understand sources of error variance after an election has concluded. Given their different interpretations and potential future utility, I separate ex-ante and ex-post predictor variables within my models.

I include my election-level variables within prediction models in a stepwise manner and measure the iterative change in prediction error associated with their inclusion. This allows the effect of each variable to be understood clearly in isolation. I also measure the cumulative change in prediction error associated with my election-level variables relative to null models. This provides a measure of the aggregate benefit of their inclusion relative to a model containing no predictor variables.

Table 24 displays the average RMSE values associated with predictions of MAE made using iteratively specified, additive prediction models containing election-level variables. Each specification represents the additive inclusion of an additional election-level variable. To avoid the loss of data and the potential introduction of bias, all observations are included within the models presented in the table. This was achieved by dropping the variables concerning partisanship, late decision-making, and left/right standard of political parties/candidates. These variables are included in subsequent tables featuring models based on subsets of data.

Table 24: Average RMSE values for MAE calculated from repeated 10-fold cross validation across stepwise, additive prediction models based on linear regression. All ex-ante and ex-post election-level variables are included iteratively and models are based on all available data (n = 11,832).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
MAE ~ 1 (Null)	2.905	-	-
Ex-ante Variables			
+ snap	2.903	-0.002	-0.002
+ election type	2.823	-0.080	-0.082
+ round two	2.823	0.000†	-0.082†
+ system change	2.813	-0.010	-0.092
+ registration difference	2.811	-0.002	-0.094
Ex-post Variables			
+ turnout	2.796	-0.015	-0.109
+ turnout change	2.785	-0.011	-0.120
+ ENEP	2.767	-0.018	-0.138
+ ENEP change	2.766	-0.001	-0.139
+ margin of victory	2.758	-0.008	-0.147

† Including the round two variable resulted in a negligible reduction in RMSE that could not be displayed to three significant figures.

From Table 24, it is clear that when compared to the null model, the inclusion of each of my election-level variables results in a reduction in average RMSE and, therefore, an improvement in the ability of the model to accurately predict values of MAE. However, the reduction in average RMSE associated with the round two variable is minimal. In total, the inclusion of all election-level variables results in a 5.06% reduction in average RMSE when compared to the null model.

Cumulatively, ex-ante election-level variables can be seen to be more useful in aiding predictions of MAE than ex-post variables, reducing average RMSE by a significantly larger amount. However, the majority of this reduction is accounted for by including differences in election type (legislative vs. presidential). By comparison, ex-post variables more consistently brought about impactful reductions in average RMSE.

Of the election-level predictors used in Table 24, accounting for the type of election being polled (legislative or presidential) produced the largest individual reduction in average RMSE, with the system change between elections, the turnout and turnout change associated with an election, and the effective number of electoral parties also producing comparatively sizeable reductions in prediction error. This indicates that they bear most closely on the variance in the MAE exhibited by polls, with MAE more likely to notably vary between different types of election, elections conducted under new electoral systems, and elections exhibiting different levels of voter turnout and ENEP relative to previous contests.

From the findings displayed in Table 24, at no point does the inclusion of election-level variables negatively affect predictive accuracy, with all models performing better than the null. This suggests that all studied election-level variables bear on the degree to which MAE varies, albeit to varying degrees, as each improves the ability of the model to produce predictions that better match observed data.

Table 25 displays the results of additive election-level prediction models for MAE drawing on polls conducted for the subset of contests for which the strength of partisan loyalty within the electorate could be calculated. It is clear that the inclusion of election-level variables is useful for the accurate prediction of MAE values, with all variables accounting for reductions in average RMSE that are meaningful to three significant figures, other than whether an election was snap or scheduled in nature which brought about a smaller reduction. Overall, the inclusion of all election-level variables results in a 3.98% reduction in average RMSE when compared to the null model.

Table 25: Average RMSE values for MAE calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the strength of partisan loyalty within the electorate could be calculated (n = 9,115).

Model	Average RMSE	Iterative Change in Average RMSE	Change in RMSE Compared to Null
MAE ~ 1 (Null)	2.613	-	-
Ex-ante Variables			
+ snap	2.613†	0.000†	0.000†
+ election type	2.588	-0.025	-0.025
+ round two	2.587	-0.001	-0.026
+ system change	2.572	-0.015	-0.041
+ registration difference	2.570	-0.002	-0.043
+ partisanship	2.568	-0.002	-0.045
Ex-post Variables			
+ turnout	2.553	-0.015	-0.060
+ turnout change	2.548	-0.005	-0.065
+ ENEP	2.538	-0.010	-0.075
+ ENEP change	2.537	-0.001	-0.076
+ margin of victory	2.509	-0.028	-0.104

† Including the snap election variable resulted in a 0.0003 reduction in average RMSE.

Compared to previous models encompassing all cases, the round two variable brings about a reduction in RMSE that can be detected to three significant figures, while the snap election variable does not. This indicates that the importance of the former as a predictor of MAE is heightened for the subset of elections in my data for which measures of partisanship are available, while the importance of the latter is diminished. Both election type and the presence of system change between elections bring about sizeable reductions in RMSE and therefore remain meaningful predictors of MAE. Similarly, the effective number of electoral parties contesting an election (ENEP) also retains its predictive utility. However, interestingly, the margin of victory serves as a stronger predictor of MAE in the subset of elections for which partisanship figures were available than it does across my dataset as a whole.

Of the election-level variables included in Table 25, the margin of victory in an election and differences in the type of election being polled (legislative vs. presidential) account for the largest reductions in average RMSE. This suggests that differences in margin of victory and election type bear most acutely on variance in MAE values across those elections for which partisanship figures were available, as their inclusion proves most beneficial in allowing the model to accurately predict the dispersion of observed data points. Importantly, the inclusion of partisanship within models improved predictive accuracy, indicating that it bears upon the variance exhibited by MAE.

When considered cumulatively, ex-post variables account for a greater proportion of error reduction than their ex-ante counterparts, indicating that they are cumulatively more useful for the prediction of out-of-sample. This suggests that, across those elections contained within the subset of data concerning partisanship, election-level variables that can be known after election day bear most closely on the variance displayed by MAE. Nevertheless, election-level variables that can be known ahead of election day still improve the ability of the model to predict error and are, therefore, still useful.

Table 26 displays the results of additive election-level prediction models run across polls conducted for the subset of elections for which the extent of late decision-making in the electorate could be calculated. This further limits the number of polls available for analysis, reducing *n* to 3,285.

Table 26: Average RMSE values for MAE calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the extent of late decision-making within the electorate could be calculated (n = 3,285).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
MAE ~ 1 (Null)	2.531	-	-
Ex-ante Variables†			
+ snap	2.530	-0.001	-0.001
+ election type	2.529	-0.001	-0.002
+ system change	2.536	-0.003	-0.005
+ registration difference	2.518	-0.018	-0.021
Ex-post Variables			
+ turnout	2.488	-0.030	-0.051
+ turnout change	2.486	-0.002	-0.053
+ ENEP	2.453	-0.033	-0.086
+ ENEP change	2.432	-0.021	-0.107
+ margin of victory	2.422	-0.010	-0.117
+ late deciders	2.403	-0.019	-0.136

† No round two presidential elections were present in the subset, so the variable was removed from analysis.

As with the other additive prediction models for MAE, the inclusion of election-level predictors can be seen to improve model performance. Overall, the inclusion of all specified election-level variables results in a 5.37% reduction in average RMSE. This suggests that, cumulatively, the variables contained within the table bear upon the degree to which MAE values vary, as they improve the ability of the prediction model to accurately predict the dispersion of observed values.

Unpacking the effect of individual election-level variables within the table, it is clear that the effective number of electoral parties (ENEP) accounts for the largest reduction in average RMSE, followed by turnout, and ENEP change between elections. As such, these election-level variables improve the accuracy of the model to the greatest extent and therefore bear most closely on the variation of MAE across the subset of elections for which late decision-making could be calculated. Importantly, the extent of late decision-making amongst the electorate is found to improve the accuracy of the prediction model, reducing its average RMSE. Consequently, it can be said to bear upon the observed variance of MAE, as it improves the ability of the model to predict the position of its values.

Though all election-level predictors contribute individually to the improvement of predictive accuracy, the reduction in average RMSE accounted for by election type is almost negligibly small in cases for which late decision-making could be measured. This represents a change when compared to similar models run across differing sets of data, as the election type variable had previously aided in the reduction of average RMSE more considerably. The changeable impact of election-level predictors between subsets speaks to their variable importance as predictors of error variance across differing contests. Here, the reduced predictive benefit of the election type variable indicates that it is less useful in the prediction of variance in observed MAE values across those elections for which late decision-making data was available than those for which the strength of partisanship could be calculated or the set of all studied elections.

By contrast, in a similar vein to earlier models tasked with rendering predictions across alternative sets of data, turnout remains a useful predictor of MAE. Indeed, to a greater or lesser extent, all election-level variables identified as useful predictors in previous models also prove useful in the output displayed in Table 26. This suggests that, while the degree may vary, the usefulness of election-level variables as predictors of MAE remains a consistent theme across differently composed subsets of data. This is encouraging, as it indicates that their predictive utility, and therefore the degree to which they can be seen to bear upon the variance exhibited by MAE, is not simply an artefact of a given set of data.

When considered cumulatively, the ex-post variables presented in Table 26 account for a noticeably larger reduction in average RMSE than their ex-ante counterparts. This suggests that election-level variables that can only be known after election day may be more useful predictors of error variance than those that can be known ahead of time. However, ex-ante variables remain usual predictors of the variance in MAE exhibited by polls in their own right.

Table 27: Average RMSE values for MAE calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the left/right standard deviation of the political position of parties and candidates could be calculated (n = 939).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
MAE ~ 1 (Null)	1.909	-	-
Ex-ante Variables†			
+ snap	1.904	-0.005	-0.005
+ election type	1.880	-0.024	-0.029
+ round two	1.873	-0.007	-0.036
+ registration difference	1.873	0.000‡	-0.036‡
+ left/right std. dev.	1.869	-0.004	-0.040
Ex-post Variables			
+ turnout	1.868	-0.001	-0.041
+ turnout change	1.866	-0.002	-0.043
+ ENEP	1.854	-0.012	-0.055
+ ENEP change	1.854‡	0.000‡	-0.055‡
+ margin of victory	1.837	-0.017	-0.072

† No instances of system change between elections were present in the subset, so the variable was removed from analysis.
‡ Including the registration difference and ENEP change variables resulted in negligible reductions in RMSE that could not be detected to three significant figures.

Table 27 displays results from additive election-level prediction models tasked with predicting out-of-sample MAE across polls conducted for elections in which the left/right standard deviation of the political parties or candidates contesting them could be calculated. The results indicate that the inclusion election-level variables generally improves the ability of models to accurately predict the variance exhibited by MAE across these cases, with most variables accounting for a notable reduction in average RMSE. Overall, the inclusion of election-level predictors resulted in an 3.77% reduction in average RMSE when compared to the null model.

Of the election-level variables that aid the prediction of variance in MAE, election type, the margin of victory, and the effective number of elective parties contesting an election (ENEP) account for the largest individual reduction in RMSE. This indicates that, across those cases in which the left/right standard deviation of the political positions of parties and candidates could be calculated, these election-level differences bear upon the variance exhibited by MAE more acutely than the other studied variables, as they allow the model to account for the greatest proportion of unexplained variance and, therefore, better predict the position of observed MAE values.

The disappearance of the predictive utility of certain variables, such as registration difference, and the reduction in the importance of turnout as a predictor suggests that their impact was driven by cases no longer included within the subset of data for which the left/right standard deviation of the political positions of parties and candidates could be calculated. This underscores that, while the general importance of election-level differences as predictors of MAE persists across subsets, the specific variables of importance and the degree to which they are important varies. Despite this, the prominence of certain variables as predictors of polling error persists across subsets. Most notably, the margin of victory and ENEP consistently account for relatively sizeable reductions in average RMSE across differing subsets of elections.

Though diminished relative to the most prominent election-level predictors of MAE, the left/right standard deviation of the political parties or candidates contesting an election can also be seen to be useful in the prediction of MAE, accounting for a modest reduction in average RMSE. This suggests that it serves as a driver of the variation exhibited by MAE across the set of cases addressed, as it improves the ability of models to better predict the position of observed data.

Table 28 displays the results of MAE prediction models run across those cases for which the strength of partisan loyalty, the extent of late decision-making amongst the electorate, and the left/right standard deviation of political parties or candidates could be jointly measured. The inclusion of all three of these variables necessarily reduced the size of a subset of polling data available for analysis (n = 293). It also reduced the range of variables available for assessment. No snap elections, differences in election type, or round two presidential elections existed within the subset, so these variables could not be included within prediction models. Additionally, there was insufficient variance in the ENEP change between elections for it to be included as a predictor variable.

Table 28: Average RMSE values for MAE calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the strength of partisan loyalty and extent of late decision-making within the electorate could be calculated, as well as the left/right standard deviation of the political position of parties and candidates (n = 293).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
MAE ~ 1 (Null)	1.630	-	-
+ left/right std. dev.	1.550	-0.080	-0.080
+ partisanship	1.536	-0.014	-0.094
+ late deciders	1.305	-0.231	-0.325

Given the relatively small size of the subset, predictive models began to overfit the data when they contained 5 or more predictors. These over-specified models then began to perform poorly in out-of-sample testing as a result. Owing to this, a parsimonious approach to modelling was necessary to produce useful predictions. The relative paucity of additional predictor variables motivated my decision to focus on three variables: the left/right standard deviation of parties/candidates, the strength of partisan loyalty within the electorate, and the proportion of late deciders within a given contest. This subset of polling data represents the universe of cases for which this information is jointly available, providing the only opportunity to explore their concurrent predictive impact. Table 28 presents the results of this exploration.

From Table 28, it is clear that the left/right standard deviation of parties, the strength of partisan loyalty amongst the electorate, as well as the extent of late decision-making in an election each serve as impactful predictors of variance in the MAE exhibited by polls in the cases for which they are jointly available. Cumulatively, their inclusion within prediction models results in a 19.94% decrease in RMSE, though this is principally driven by the extent of late decision-making within elections, which accounts for by far the largest reduction in prediction error. This suggests that, across elections in which information pertaining to the strength of partisan loyalty, the extent of late decision-making, and the distribution of party and candidate alignments is available, late decision-making within the electorate serves as the principal drivers of observed variance in MAE, allowing models to more accurately predict its dispersion. Nevertheless, both partisanship and the distribution of party and candidate alignments remain useful predictors of MAE, both contributing meaningful reductions to the average RMSE exhibited by the model.

The individual reductions in RMSE brought about by the variables within Table 28, along with their sizeable cumulative reduction, lends additional weight to the contention that election-level characteristics are useful predictors of polling error variance. However, it further suggests that the subset of data used to render predictions is a factor in determining the extent to which individual election-level variables are considered to be predictively useful. This phenomenon arises continually across the additional prediction models displayed in Appendix B and is addressed in greater detail later in the thesis.

Importantly, given the limited size of the sample of data used by models in Table 28, the findings displayed should be treated with caution. Combined data on partisanship, the left/right standard deviation of the policy positions of parties, and the extent of late decision-making amongst the electorate is only available for seven elections. This serves to caution against the over-interpretation of the perceived predictive usefulness of variables. Nevertheless, the findings of this exploratory analysis are encouraging and further analysis of the usefulness of these variables as predictors in combination may be of interest to future scholarship, should the collection of the data on which their calculation rests become more widely adopted.

The findings that result from running a similar suite of additive prediction models across my additional measures of polling error are displayed in Appendix B across Tables B1 through B35. These analyses were consigned to the appendix due to the spatial limitations placed on this thesis. In what follows, I present the aggregate findings that result from these additional additive prediction models, as well as aggregating findings related to predictions of variance in MAE.

In order to directly compare findings across models, a degree of transformation is required. As RMSE assumes the same units as the outcome variable it is used to assess predictions of,[999] values cannot be directly compared between models concerning outcome variables that are measured using different units. RMSE is also scale dependent and, therefore, cannot be directly compared across models concerned with outcome variables measured on different scales.[1000] Several solutions to the scale dependency and unit assumption of RMSE have been

[999] Nada R. Sanders, 'Measuring Forecast Accuracy: Some Practical Suggestions', *Production and Inventory Management Journal*, 38.1 (1997), 43 – 46 (p. 45).
[1000] Rob J. Hyndman and Anne B. Koehler, 'Another Look at Measures of Forecast Accuracy', *International Journal of Forecasting*, 22.4 (2006), 679 – 688 (p. 682).

proposed.[1001] Most prominent amongst these is the normalisation of RMSE values.[1002] Various approaches exist to the normalisation of RMSE but, despite its prominence, no consensus exists within the literature as to which approach is most appropriate. Each approach possesses its own unique shortcomings,[1003] with performance varying according to the nature of the variables to which normalised RMSE is applied, leaving the appropriateness of many forms of normalisation in doubt.

As my measures of polling error are not measured on the same scale and often assume different units, I must transform my raw RMSE values to allow for cross-measure comparison. Given the lack of consensus surrounding normalised RMSE and the significant shortcomings inherent within the various approaches to it, I opt for a different approach to facilitating comparison. As RMSE is scale-dependent, the errors it calculates exist on the same scale as the outcome variable.[1004] The issue of scale dependence is that identical RMSE values cannot be compared across scales, as they lend themselves to different interpretations. That is, a given change in RMSE may be more consequential in prediction models relating to one outcome variable than another depending on the scales on which these variables are measured. A simple solution to this problem is to present the changes in RMSE associated with the inclusion of predictor variables within models as proportions of the total error they represent.

Representing prediction error as a proportion is unit free and therefore not bound by the scale on which an outcome variable is measured.[1005] Using proportions in this way also allows findings to be sensitive to the scales assumed by different outcome variables. Identical changes in RMSE are no longer incorrectly taken to be equally impactful across differently scaled

[1001] Maxim Vladimirovich Shcherbakov and others, 'A Survey of Forecast Error Measures', *World Applied Sciences Journal*, 24 (2013), 171 – 176 (p. 174).
[1002] Ibid.
[1003] Ibid.
[1004] Hyndman and Anthanasopoulos, p. 46.
[1005] Ibid.

variables, as the magnitude of these changes is represented relative to the nature and scope of the scale on which they are measured. This allows for the meaningful and representative comparison of the impact of changes in RMSE across models. The use of proportions to represent changes in RMSE also provides the added benefit of enabling the clear comparison of findings between regression and classification models.

Table 29: The average percentage point reduction in RMSE associated with each election-level variable across all additive prediction models for all continuous measures of polling error.

Variable	MAE (%)	DIM (%)	LPB (%)	APB (%)	ABI 1 (%)	ABI 2 (%)
Snap	0.10	0.15	0.44	1.16	0.19	0.22
Election type	1.25	0.15	0.66	1.47	0.90	0.93
Round two	0.20	0.98	0.00	0.00	0.00	0.05
System change	0.35	0.18	0.35	0.37	0.32	0.42
Registration difference	0.29	0.09	0.65	0.84	0.04	0.00
Partisanship	0.49	1.30	0.26	1.27	1.05	1.29
Left/right std. dev.	1.73	1.68	1.99	0.37	0.26	3.05
Turnout	0.58	0.40	0.00	0.00	0.32	0.63
Turnout change	0.19	0.26	0.00	1.66	0.07	0.09
ENEP	0.73	0.60	0.34	2.33	0.22	0.44
ENEP change	0.31	0.37	0.33	1.19	0.40	1.01
Margin of victory	0.66	0.86	2.48	0.37	1.23	1.70
Late decision-making	7.46	7.30	0.71	0.00	7.07	6.26

Table 29 displays the average percentage point reduction in RMSE associated with each election-level predictor variable across each of my continuous measures of distributive and bounded polling error, as well as polling bias. From the percentages, it is clear that election-level variables almost universally improve the ability of models to accurately predict the presence of each of my measures of polling error, though improvements vary in both extent and consistency.

For the mean absolute error (MAE) exhibited by polls, all specified election-level variables produce reductions in the RMSE of prediction models, improving their ability to accurately predict it. Of the variables addressed, the presence of late decision-making in the electorate, the left/right standard deviation of the political position of parties contesting elections, and the type of election under consideration (legislative or presidential) are the most predictive of variance in MAE. That these variables produce the most substantial reductions in prediction error indicates that they account for the largest portion of unexplained variance in observed MAE values, suggesting that differences in election type, differing proportions of late decision-makers between, and differences in the ideological alignment of the parties and candidates contesting them serve as notable drivers of the degree to which the MAE exhibited by polls varies.

Beyond these variables, the effective number of elective parties contesting an election (ENEP), the margin of victory within an election, the strength of partisan loyalty amongst the electorate, and the level of turnout were each found to produce notable reductions in RMSE, thereby improving predictions of MAE and better accounting for its variance. While each of the remaining election-level variables also improve the accuracy of predictions of the MAE exhibited by polls, the extent to which they do so is comparatively diminished.

The findings derived from models tasked with predicting the difference in margin (DIM) between polling predictions and actual electoral returns tell a similar story. Each of the specified election-level variables aids in the prediction of DIM by reducing average RMSE.

The presence of late decision-making amongst the electorate again produces by far the largest improvement to predictive accuracy, followed by the left-right standard deviation between the political positions of parties. This indicates that the DIM exhibited by polls also varies most notably across elections that exhibit differences in the extent of late decision-making amongst the electorate and the ideological distance between competing parties and candidates. Such commonalities with MAE are to be expected, as they are both measures of the same distributive conceptualisation of polling error.

Additional electoral characteristics produced notable reductions in RMSE for predictions of DIM. In some cases, these reductions represent a departure from the findings drawn from models centring on MAE. While levels of turnout and the margin of victory in an election again resulted in clear improvements in the ability of models to accurately predict DIM, so too did the extent of partisanship within the electorate and whether a contest was a round two presidential election. This suggests that these variables serve as meaningful predictors of the variance exhibited by DIM. While both partisanship and the second-round election binary were useful predictors in the case of MAE, their usefulness was comparatively diminished. As such, while commonalities exist between conceptions of error, the usefulness of electoral characteristics as predictors of polling error can be seen to vary depending on how polling error is measured.

Despite these differences, all election-level variables again produced notable percentage point reductions in RMSE. This indicates that election-level variables bear on the degree to which the DIM exhibited by polls varies, as their inclusion allows models to more accurately predict values and, in so doing, better account for observed variance. When considered together, the results presented by models centred on MAE and DIM suggest that the studied election-level variables serve as universally useful predictors of distributive polling inaccuracy.

While the inclusion of electoral characteristics in prediction models results in consistent improvements to predictive accuracy in the case of my measures of distributive accuracy, results from models concerned with predicting measures of polling bias (LPB and APB) produces less consistent results. Ultimately, the diminished usefulness of election-level variables in the case of polling bias is in line with theoretical expectations.

In the case of leading party bias (LPB), while the majority of election-level variables result in reductions in average RMSE, therefore improving the ability of models to accurately predict LPB values, they do so less consistently than earlier models centred on distributive inaccuracy. Indeed, three electoral characteristics – the second-round election binary, turnout levels, and turnout change between contests – do not improve the predictive ability of models. Of those electoral characteristics that improve the ability of models to predict LPB values, the margin of victory in an election and the ideological position of parties and candidates account for the largest decreases in average RMSE, followed by the presence of late decision-making amongst the electorate and the type of election in question (legislative or presidential). This indicates that these are the election-level variables that bear most acutely on the variance exhibited by LPB, as they provide the largest improvement to the accuracy of models, thereby allowing them to successfully account for a greater amount of observed variance.

The ability of election-level variables to aid in the prediction of the average per-party bias (APB) is also characterised by inconsistency. Again, three variables do not produce reductions in the average RMSE presented by models and, therefore, do not improve their ability to predict APB. In the same vein as LPB, both the second-round election binary and the level of turnout in elections did not reduce average RMSE, suggesting a degree of commonality in those variables that do not serve as useful predictors of polling bias. However, unlike LPB, the extent of late decision-making is not predictively useful in the case of APB, indicating that the impact of election-level variables again varies across measurement approaches.

Of those electoral characteristics that yielded improvements to predictive accuracy, the margin of victory in an election, along with the left/right standard deviation of the ideological positions of parties, stand as the most prominent contributors to predictive accuracy. This indicates that the variance observed in APB values is most acutely driven by differences the margin of victory exhibited by elections and changes in the ideological positioning of parties and candidates between contests than the other election-level variables included in modelling. Despite this, barring the three variables that were not predictively useful, all remaining variables aided in the prediction of APB, suggesting that election-level characteristics exist as broadly useful predictors of the average party bias exhibited by polls.

While it is in line with theoretical expectations, the diminished predictive impact of electoral characteristics in the case of measures of bias may be an artefact of their low standard deviation about their respective means and, therefore, of the way in which they are measured. Values of LPB and APB are centred about a mean of zero and are measured on granular scales such that their values are not widely dispersed about the mean, leading to low standard deviation. The low standard deviation exhibited by LPB and APB is such that the predictive performance of the null model is inflated relative to other measures of polling error. Given that the null model operates in the absence of predictor variables, the most likely value assumed by any outcome variable measurement – and therefore the best prediction of it – is simply the mean of all outcome variable values. As the values assumed by LPB and APB do not deviate from the mean to a great extent, the predictions offered by the null model are likely to present reduced error when compared to other, more widely dispersed measures of polling inaccuracy. The increased performance of the null model leaves less room for predictive improvement through the addition of predictor variables, resulting in diminished changes to model accuracy on their inclusion.

The inclusion of electoral characteristics within models tasked with the prediction of both measures of the average bounded inaccuracy of polls (ABI 1 and ABI 2) yield a range of similar findings. This is unsurprising, as both measures exist as alternative operationalisations of the same underlying conception of bounded polling error. Across both measures, the inclusion of most of the specified electoral characteristics improves the ability of models to accurately predict error values by reducing average RMSE. Notably, the accuracy of predictions of both ABI 1 and ABI 2 is increased most prominently by the extent of late decision-making within the electorate, with models concerning both measures of bounded inaccuracy also benefitting to similar extents from the inclusion of the extent of partisanship amongst the electorate, differences in election type, and the presence of system change between contests. This suggests that these variables stand as the most important and consistent election-level drivers of variance in bounded polling inaccuracy.

Despite these similarities, differences in the predictive utility of election-level variables are visible between measures of bounded inaccuracy. While the second-round election binary improves predictions of ABI 2, it is not predictively useful in the case of ABI 1. Likewise, while the difference in registered voters between contests improves the predictive accuracy of models concerned with ABI 1, it is of no benefit to models concerned with ABI 2. Indeed, differences in the importance of predictors between measures of bounded polling error is visible across almost all specified electoral characteristics, with notable differences occurring in the case of the ideological position of parties, ENEP change between contests, and the level of turnout in an election. This again underscores the impact of the manner in which polling error is measured on the degree to which electoral characteristics are deemed to be predictive of it. Nevertheless, election-level variables remain broadly useful predictors of measures of bounded inaccuracy irrespective of the approach taken to its measurement.

When considered cumulatively, the findings displayed in Table 29 illuminate those electoral characteristics that are most generally useful for the prediction of polling error variance. The left/right standard deviation of the political positions of parties contesting elections offers broad improvements to the predictive accuracy of models across all continuous measures of polling inaccuracy. The impact of the margin of victory in an election on the predictive accuracy of models also transcends conceptualisations of polling inaccuracy. Other electoral characteristics bear more closely upon individual conceptualisations. Late decision-making offers the most prominent predictive improvement to models concerned with measures of distributive and bounded polling error, while the snap election variable bears most directly on the accuracy of models concerned with predicting polling bias. This suggests that, while election-level variables stand as broadly useful predictors of polling error, the extent of the usefulness of individual predictors varies between conceptualisations of error.

Table 30 explores the average utility of election-level variables as predictors of my binary measures of polling error. As predictions of binary outcomes are better understood as attempts at classification,[1006] the impact of each of my election-level variables is measured in terms of their average percentage point impact on the correct classification rate of models. For election-level variables to be considered useful predictors of the variance associated with binary measures of polling error, they must necessarily increase the proportion of correct classifications associated with a model.

[1006] Graham Elliot and Robert P. Lieli, 'Predicting Binary Outcomes', *Journal of Econometrics*, 174.1 (2013), 15 – 26 (p. 24).

Table 30: The average percentage improvement in classification accuracy associated with each election-level variable across all additive prediction models.

Variable	SBP Correct Classification Change (%)	LVRC Correct Classification Change (%)
Snap	1.22	0.63
Election type	3.02	7.90
Round two	-0.36	0.67
System change	1.57	0.63
Registration difference	0.14	-0.08
Partisanship	7.68	33.39
Left/right std. dev.	2.36	9.07
Turnout	1.62	2.22
Turnout change	3.90	1.51
ENEP	2.62	1.19
ENEP change	0.40	2.20
Margin of victory	-0.34	0.78
Late deciders	1.39	0.57

From Table 30, it is clear that all specified election-level variables bar two improve the ability of models to correctly classify instances of significantly biased polling (SBP), while all but one aid in the correct classification of whether polls predict the largest vote share recipient in an election correctly (LVRC). This indicates that election-level variables, broadly conceived, stand as useful predictors of both significant polling bias and substantive polling inaccuracy.

For classifications of SBP, the strength of partisan loyalty amongst the electorate can be seen to account for the largest improvement in correct classification of all specified election-level variables, followed by the change in turnout between elections, and the type of election being polled. These findings suggest that observed variance in SBP is driven more prominently by these election-level factors than the other variables under consideration. Beyond these, all election-level variables other than the second-round binary and the margin of victory improved the correct classification rate of the model, improving predictions of polling error.

For classifications of LVRC, partisanship again produces the most prominent individual improvement in predictive accuracy, yielding by far the largest improvement to the rate of correct classification. Differences in election type also brought about a notable improvement to predictive accuracy, so too does the ideological position of the parties and candidates. From this, variance in the substantive error exhibited by polls can be said to be more closely affected by these election-level differences than by the other variables included within classification models.

Though similarities exist between those election-level variables that are useful for the classification of SBP and LVRC, notable differences also exist. For example, while the second-round election binary is a useful predictor of LVRC, it negatively impacts the ability of models to correctly classify instances of SBP. Similarly, while differences in the number of registered voters between elections actively reduce the accuracy of models predicting the presence of LVRC, they prove useful in the prediction of SBP. Such differences are to be expected, however, as though both SBP and LVRC are binary operationalisations of polling error, they measure profoundly different conceptions of error.

While previous tables decompose the individual predictive utility of my 13 specified electoral characteristics, Table 31 explores their impact when considered in combination with one another. To do so, it presents the average degree to which the inclusion of election-level variables within models improves their ability to predict polling error variation. Ranges are supplied alongside these averages to better understand the nature of improvements across models.

Table 31: The average improvement to predictive accuracy yielded by optimal election-level model specifications across all subsets of data for all measures of polling error. Averages are accompanied by the range of improvements provided by these models.

Measure	Average Additive Improvement (%)	Additive Improvement Range (%)
MAE	7.62	3.77 – 19.94
DIM	4.69	2.17 – 9.62
LPB	3.47	0.77 – 6.62
APB	5.30	3.57 – 9.97
ABI 1	5.27	1.04 – 15.37
ABI 2	7.41	2.81 – 19.49
SBP	14.23	8.11 – 19.08
LVRC	26.29	8.75 – 45.87

From Table 31, it is clear that, on average, the inclusion of election-level variables improves the ability of models to accurately predict each of my measures of polling accuracy, thereby reducing the disparity between predicted and observed values, improving their ability to account for observed variance. However, the range of improvements across models varies considerably. Across all models, the inclusion of election-level variables improves predictions of polling error by an average of ~3.5% to ~26%, depending on the measure of error addressed. Election-level variables are most useful in accurately predicting substantive polling error (LVRC), improving predictive accuracy by an average of 26.29% and a maximum of 45.87%. By contrast, they are least useful in the prediction of leading party bias (LPB), improving predictions by an average of 3.47% and a maximum of 6.62%. These findings are in keeping with theoretical expectations, with election-level variables proving most useful as predictors of substantive polling inaccuracy and least useful as predictors of a measure of polling bias. However, election-level predictors prove more useful than expected in the case of models concerning the SBP and APB measures of polling bias, often providing larger percentage point improvements than in models centring on measures of distributive inaccuracy (MAE and DIM).

Across all subsets of data, the ability of election-level variables to aid in the prediction of polling error ranges from 0.77% to 45.87%, implying not only that their importance as predictors varies across measures of polling error, but also that it varies considerably on the basis of the composition of the data used within models. This implication is explored further later within this chapter. Even accounting for variation between models, it is clear that the inclusion of electoral characteristics within models aids their ability to accurately predict polling error values, albeit to varying degrees.

While the results of additive prediction models serve to illustrate the improvements in understanding polling error variance that come from adopting an election-level approach, in the following sub-section, I increase the complexity of model specifications to better capture the likely real-world behaviour of electoral characteristics as predictors of polling error variance.

Interactive Prediction Models

To explore the relationship between polling error and the election-level interactions outlined in the previous section, I run a series of analyses layering two- and three-way interactions into existing additive prediction models. For the purposes of comparison, these interactive prediction models are run across the same (sub)sets of data used within additive analyses and use an identical repeated 10-fold cross validation procedure. In the case of each (sub)set, only those interactions relevant to the variables it contains are included within prediction models.

Interactive linear models build on their additive counterparts and adhere to theoretically motivated variants of the approach outlined in equation 30, incorporating isolated two- and three-way interactions additively in a way that layers them on to existing main effects.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4 + \hat{\beta}_5 x_1 x_2 + \hat{\beta}_6 x_3 x_4 + \hat{\beta}_7 x_1 x_2 x_3 + \in \quad (30)$$

The inclusion of isolated two- and three-way interactions in my principal interactive prediction models, rather than iteratively interacting all predictor variables, is theoretically motivated decision. While several plausible two-way interactions exist between my election-level predictor variables, with a number of three-way interactions expected in the presence of measurements of partisanship, it proved difficult to unpack higher order interactions in a way that made the assessment of their individual impact theoretically justifiable. Notably, I only include interactions between my election-level variables, as these remain the principal focus of the thesis.

The utility of each interaction is assessed in relation to its ability to improve the accuracy of predictions of polling error relative not only to a null model, but also to relevant additive models from earlier analysis. Predictive improvement is again measured in terms of the reduction to average RMSE in the case of continuous measures of error and improvements to the correct classification rate in the case of binary measures. Table 32 displays the utility of a range of two-way interactions in the prediction of MAE across my full dataset of polls.

Table 32: Average RMSE values for MAE calculated from repeated 10-fold cross validation across interactive linear regression models iteratively including all election-level variables. Values are calculated for all data (n = 11,832).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
MAE ~ 1 (Null)	2.905	-	-
MAE ~ additive variables	2.758	-0.147	-0.147
+ turnout change × margin of victory	2.740	-0.018	-0.165
+ turnout change × ENEP	2.740†	0.000†	-0.165†
+ turnout change × ENEP change	2.739	-0.001	-0.166
+ ENEP × turnout	2.737	-0.002	-0.168
+ turnout × ENEP change	2.737	0.000	-0.168

† Including the two-way interaction between turnout change and ENEP resulted in a small 0.0002 reduction in average RMSE.

From the results presented in Table 32, it is clear that many of the interactions between election-level variables improve the ability of the model to accurately predict MAE. However, the contribution of two-way interactions to model accuracy is far less consistent than the additive inclusion of individual election-level variables, with two interactions failing to meaningfully improve predictive performance. Nevertheless, overall, the inclusion of interactions leads to a further 0.021 reduction in average RMSE over solely additive models, representing an additional 14% improvement in predictive accuracy.

Of the interactions contained within Table 32, the interaction between turnout change and margin of victory stands as the most significant contributors to predictive accuracy. This indicates that the propensity for MAE values to vary is driven more directly by these two-way interactions than the others included within the model. While the two-way interactions between turnout change and ENEP change, and ENEP and turnout bring about reductions in average RMSE, respectively, the degree to which they do so is comparatively diminished.

Table 33: Repeated 10-fold Cross validated RMSE values for MAE calculated from interactive linear regression models. Values are calculated from the subset of data for which partisanship values were available (n = 9,115).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
MAE ~ 1 (Null)	2.613	-	-
MAE ~ additive variables	2.509	-0.104	-0.104
+ turnout change × margin of victory	2.477	-0.032	-0.136
+ turnout change × ENEP	2.476	-0.001	-0.137
+ turnout change × ENEP change	2.470	-0.006	-0.143
+ ENEP × turnout	2.468	-0.002	-0.145
+ turnout × ENEP change	2.465	-0.003	-0.148
+ ENEP × partisanship	2.465	0.000	-0.148
+ turnout change × partisanship	2.465	0.000	-0.148
+ turnout × partisanship	2.463	-0.002	-0.150
+ turnout change × ENEP × partisanship	2.457	-0.006	-0.156
+ turnout change × ENEP change × partisanship	2.454	-0.003	-0.159
+ ENEP × turnout × partisanship	2.440	-0.014	-0.173
+ turnout × ENEP change × partisanship	2.430	-0.010	-0.183

Table 33 displays the degree to which two- and three-way interactions aid in the prediction of MAE across the subset of polls conducted for elections in which the strength of partisan loyalty amongst the electorate could be calculated. Overall, it is clear that the inclusion of election-level interactions improves the ability of the model to accurately predict MAE. Cumulatively, the interactions included the table result in a further 0.079 reduction in average RMSE over solely additive models, representing an additional 75.96% improvement in predictive accuracy.

Of the interactions included in Table 33, the two-way interaction between turnout change and the margin of victory in an election again provides by far the largest individual decrease in average RMSE and, therefore, proves the most useful in accurately predicting MAE, followed by the three-way interaction between ENEP, turnout, and partisanship. This suggests that the variance in MAE exhibited by these polls is more acutely driven by these interactions than others, as they allow the model to account for a greater amount of erstwhile unexplained variance.

Though the impact of the remaining interactions displayed in the table is comparatively diminished, all but two interactions positively improve the ability of the model to accurately predict MAE values. Not only does this further suggest that interactions between election-level variables serve as useful predictors of distributive polling error, but it also indicates that the predictive utility of interactions varies across subsets of data, as a far greater number of interactions yield meaningful reductions in average RMSE across the subset of elections for which partisanship data could be gathered as opposed to the larger dataset as a whole.

Table 34 displays the degree to which interactions between election-level variables aid in the prediction of MAE across the subset of polls conducted for elections in which the extent of late decision-making in the electorate could be calculated. While the majority of interactions improve the ability of models to accurately predict MAE, the degree to which they reduce average RMSE varies considerably. When considered in tandem, the inclusion of election-level interactions results in an additional reduction in average RMSE of 0.092 over solely additive models, representing a 67.65% improvement in predictive accuracy.

Table 34: Repeated 10-fold cross validated RMSE values for MAE calculated from interactive linear regression models. Models draw on the subset of data for which the extent of late decision-making within the electorate could be established (n = 3,285).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
MAE ~ 1 (Null)	2.531	-	-
MAE ~ additive variables	2.403	-0.136	-0.136
+ turnout change × margin of victory	2.384	-0.019	-0.155
+ turnout change × ENEP	2.384	0.000	-0.155
+ turnout change × ENEP change	2.367	-0.017	-0.172
+ ENEP × turnout	2.347	-0.020	-0.192
+ turnout × ENEP change	2.347	0.000	-0.192
+ late deciders × ENEP	2.318	-0.029	-0.221
+ late deciders × turnout	2.313	-0.005	-0.226
+ late deciders × turnout change	2.311	-0.002	-0.228

Though six of the eight interactions included within the prediction models displayed in Table 34 yield reductions in average RMSE, the degree to which they improve predictive accuracy varies considerably. The most impactful interaction is that between the extent of late decision-making within the electorate and the effective number of parties contesting an election, while the least impactful interaction was that between the extent of late decision-making and the magnitude of turnout change. This indicates that variance in the MAE values exhibited by polls is most acutely affected by the intersection of late decision-making and ENEP within elections and least affected by the inter-relation between late decision-making and turnout shifts between elections.

From the assessment of the model outputs presented in Tables 32 through 34, it is clear that including election-level interactions within prediction models generally increases the accuracy with which they are able to predict the MAE values exhibited by polls. Indeed, collectively, their inclusion yields improvements in predictive accuracy when compared to solely additive models that range from 14% to ~76% depending on the subset of data addressed and the range of interactions used. This not only indicates their usefulness as predictors of MAE but, due to the reduction in RMSE associated with them, suggests that they bear upon the degree to which MAE varies, allowing its drivers to be better understood.

Table 35: The average percentage point improvement to predictive accuracy associated with each two- and three-way election-level interaction in prediction models across all continuous measures of polling error relative to additive models.

Interaction	MAE (%)	DIM (%)	LPB (%)	APB (%)	ABI 1 (%)	ABI 2 (%)
turnout change × margin of victory	18.98	9.13	4.18	2.56	25.24	23.32
turnout change × ENEP	0.37	8.73	20.91	6.22	6.12	2.72
turnout change × ENEP change	6.32	21.98	10.32	4.73	6.47	15.72
ENEP × turnout	6.00	13.09	0.00	14.87	1.74	3.38
turnout × ENEP change	0.96	2.19	1.31	18.52	2.67	1.35
ENEP × partisanship	0.00	0.74	1.67	0.00	0.00	0.90
turnout change × partisanship	0.00	0.00	0.00	2.31	0.11	0.00
turnout × partisanship	1.92	2.21	0.00	0.00	3.41	3.60
late deciders × ENEP	21.32	52.56	11.77	22.22	25.00	33.70
late deciders × turnout	3.68	25.00	11.77	0.00	1.19	1.09
late deciders × turnout change	1.47	4.49	0.00	6.67	0.00	0.00
turnout change × ENEP × partisanship	5.78	5.88	0.00	0.00	12.50	8.11
turn. chg. × ENEP chg. × partisanship	2.89	15.44	0.00	7.69	1.14	0.00
ENEP × turnout × partisanship	13.46	6.62	16.67	0.00	25.00	12.61
turnout × ENEP change × partisanship	9.62	12.50	0.00	2.31	26.14	20.72

To assess whether similar conclusions can be drawn in relation to my other measures of polling error, the findings that result from the inclusion of election-level interactions in models tasked with their prediction are presented in Appendix B across Tables B36 through B56. The average improvement to predictive accuracy that these interactions offer over additive models across all continuous measures of polling error is displayed above in Table 35. Improvements are again represented as percentages to facilitate cross measure comparison.

From Table 35, it is clear that the inclusion of election-level interactions within prediction models widely improves the ability of models to accurately predict continuous measures of polling error. However, the average impact of individual interactions varies considerably across measures. Broadly, the inclusion of interactions is most beneficial for predictions of measures of distributive polling error (MAE and DIM), as a greater number of interactions improve predictive accuracy when compared to other conceptions. The inclusion of interactions is also widely beneficial for the prediction of measures of bounded polling error (ABI 1 and ABI 2), with the majority of two- and three-way interactions improving model accuracy on average. By contrast, the inclusion of interactions is least beneficial for predictions of polling bias (most notably LPB), as fewer interactions bring about improvements to model accuracy. This suggests that, on average, election-level interactions are most useful for the prediction of variance in distributive polling error and least useful in the prediction of variance in polling bias, while remaining beneficial to predictions of bounded error. Importantly, predictions of all measures of polling accuracy are made more accurate by at least some of the stated two- and three-way interactions, affirming their general usefulness as predictors of polling error variance.

For predictions of MAE, all election-level interactions bar two reduce average RMSE and, therefore, improve the ability of models to predict observations and better account for their variance. This suggests that interactions between election-level variables stand as broadly meaningful drivers of MAE. Of those interactions that are impactful, the two-way interactions between late decision-making within the electorate and the effective number of parties contesting an election, as well as the interaction between turnout change and the margin of victory, yield by far the largest individual improvements to predictive accuracy. This suggests that these interactions bear more closely on the variance of MAE than others.

The inclusion of election-level interactions also leads to wide-ranging improvements in the ability of models to accurately predict DIM, with all but one of the specified interactions bringing about reductions in average RMSE. The two-way interaction between late decision-making amongst the electorate and the effective number of electoral parties contesting an election (ENEP) again accounts for the largest improvement to predictive accuracy. Given its prominence as a predictor of MAE, this suggests that it stands as a significant driver of distributive polling error.

While further commonalities with MAE exist, such as the continued importance of the interaction between turnout and the margin of victory, notable differences are also present. The two-way interaction between late decision-making and the level of turnout in an election serves as a substantially more important predictor of DIM than MAE, so too does the interaction between ENEP and turnout. Broadly, interactions including partisanship serve as more useful predictors of DIM than MAE, especially in three-way interactions with ENEP, turnout, and their associated changes between elections. That many of the most impactful interactions for the prediction of DIM centre on the interaction of turnout and the effective number of parties speaks to their particular importance as predictors of its variance.

While interactions between election-level variables often improves the ability of models to predict LPB, their ability to do so is substantially less consistent. Of the 15 interactions addressed, seven failed to bring about improvements to predictive accuracy. This suggests that the usefulness of election-level interactions as predictors of LPB is diminished relative to other measures of polling error. Nevertheless, as the majority of election-level interactions bring about improvements to predictive accuracy, they remain useful and informative drivers of leading party bias.

Of the election-level interactions that improve predictions of LPB, the two-way interaction between turnout change and ENEP produces the most pronounced improvement to predictive accuracy. This suggests that variance in the LPB values exhibited by polls is most acutely driven by the inter-connection between these variables than the other interactions addressed. Interestingly, the predictive improvement associated with the interaction between turnout change and ENEP is far greater in the case of LPB than any other measure of polling error. This suggests that the inter-relationship between these variables bears more closely on leading party bias, and therefore stands as a more important driver of it, than my other measures of error.

Beyond the two-way interaction between turnout change and ENEP, the ability of models to predict LPB is also positively affected by the interaction between late decision-making and ENEP which remains a consistently useful interaction. Additionally, the interaction between ENEP and partisanship produces a larger improvement in the ability of models to predict LPB than other measures of polling error. Indeed, the relationship between levels of strong partisan sentiment and both turnout levels and ENEP appears particularly useful for the prediction of LPB, with both two- and three-way interactions between these variables yielding larger improvements to predictive accuracy than observed in models concerning other measures of polling error.

Perhaps unsurprisingly considering their effect on predictions of leading party bias (LPB), the inclusion of election-level interactions also has a diminished impact on predictive accuracy in models tasked with the prediction of average party bias (APB) when considered relative to other measures. Five of the fifteen interactions included within modelling failed to improve predictive accuracy, indicating that interactions between election-level variables are less consistently useful as predictors of APB than measures of either distributive or bounded polling error, with only models centred on LPB being positively affected by fewer interactions.

Despite this, the majority of interactions between election-level variables improve the ability of models to accurately predict APB values. This suggests that, while several interactions are of no importance, election-level interactions stand as generally useful drivers of the variance exhibited by APB. Of those interactions that yielded improvements in predictive accuracy, the two-way interaction between late decision-making and ENEP produces by far the largest increase in accuracy, further underscoring its predictive usefulness. This is followed by the interaction between turnout levels and ENEP change between elections, which yields a far larger improvement to predictive accuracy than can be seen in models concerned with other measures of polling error. This suggests that the intersection of these variables is particularly useful for the prediction of APB.

The predictive impact of election-level interactions is improved in relation to measures of bounded polling error (ABI 1 and ABI 2), with all but two interactions improving predictions of ABI 1 and all but three improving predictions of ABI 2. This indicates that a greater number of the stated interactions improve the ability of models to accurately predict measures of bounded polling error than measures of polling bias and, more generally, that election-level interactions stand as important drivers of bounded polling error, allowing models to better account for its variance.

Perhaps unsurprisingly given their common conceptual focus, models tasked with the prediction of ABI 1 and ABI 2 chiefly benefit from the same set of interactions. The three-way interaction between turnout levels, ENEP change, and the strength of partisanship yields the most pronounced improvement to predictive accuracy in the case of ABI 1 and proves useful in the case of ABI 2. Similarly, the two-way interaction between late decision-making and ENEP brings about the largest improvement in models concerned with ABI 2 and yields a substantial improvement in models concerned with ABI 1. The continued improvements to predictive accuracy associated with the interaction between late decision-making and ENEP – improvements that prove consistent across all measures of polling error – also serve to underscore the particular predictive usefulness of this interaction.

While commonalities exist between results born of models concerned with ABI 1 and ABI 2, some differences are also present. The two-way interaction between turnout change and ENEP change proves more than twice as useful as a predictor of ABI 2 than ABI 1, while certain interactions that prove useful in the case of ABI 1 do not yield improvements in the case of ABI 2 and vice versa. This suggests that the predictive importance of election-level interactions varies across approaches to measuring the same conceptualisation of polling error.

In total, it is clear that two- and three-way interactions between election-level variables stand as generally useful predictors across all of my continuous measures of polling error. To establish the extent to which election-level interactions stand as useful predictors of my binary measures of polling error, Table 36 displays the average predictive improvement over purely additive models associated with each interaction across models concerned with LVRC and SBP. In keeping with earlier findings, the majority of election-level interactions improve the predictive ability of models, increasing their ability to correctly classify instances of LVRC and SBP. This suggests that election-level interactions also stand as useful predictors of binary measures of polling error.

Table 36: The average change in the percentage of correct classifications of SBP and LVRC values associated with each two- and three-way interaction across all election-level classification models relative to additive models.

Interaction	SBP Correct Classification Change (%)	LVRC Correct Classification Change (%)
turnout change × margin of victory	3.38	5.17
turnout change × ENEP	3.09	-0.90
turnout change × ENEP change	2.58	2.74
ENEP × turnout	-1.72	18.39
turnout × ENEP change	-0.30	3.19
ENEP × partisanship	0.27	0.21
turnout change × partisanship	1.59	4.67
turnout × partisanship	0.79	7.86
late deciders × ENEP	2.35	11.79
late deciders × turnout	4.16	0.34
late deciders × turnout change	0.54	11.96
turnout change × ENEP × partisanship	4.33	2.19
turnout change × ENEP change × partisanship	11.74	0.57
ENEP × turnout × partisanship	14.37	2.27
turnout × ENEP change × partisanship	1.32	0.28

From the results presented in Table 36, it is clear that while the majority of election-level interactions improve the performance of classification models, not all interactions aid in the prediction of my binary measures of polling error. Indeed, the inclusion of some interactions actually decreases the accuracy of predictions. In the case of SBP, two-way interactions between turnout about both ENEP and ENEP change reduce correct classification rate, while in the case of LVRC, only the interaction between turnout change and ENEP reduces predictive accuracy. This suggests that, when considered in the aggregate, these interactions do not serve as useful predictors of SBP and LVRC.

Despite this, the majority of interactions within the table prove predictively useful. Models tasked with the prediction of SBP benefit chiefly from the inclusion of three-way interactions between partisanship, ENEP, turnout, and their associated changes between elections, as these interactions account for the largest individual improvements to rates of correct classification. Beyond this, the two-way interaction between late decision-making and turnout also yields a notable improvement to predictive accuracy, so too do the interactions between turnout change and both ENEP and margin of victory, respectively.

In the case of LVRC, a greater number of election-level interactions can be seen to positively affect the correct classification rate of models. Interactions can generally be seen to produce larger improvement to predictive accuracy than those observed in models concerned with SBP, suggesting that interactions serve as more useful predictors of LVRC than SBP. Of the interactions addressed, the two-way interaction between ENEP and turnout levels produces the largest improvement to correct classification rate. The fact that this interaction negatively affects the ability of models to correctly classify instances of SBP lends further support to the assertion that the importance of election-level interactions as predictors of polling error varies between conceptualisations.

While the contribution of individual election-level interactions paints a varied picture of their importance, when considered in the aggregate, the predictive improvement that they offer over purely additive models becomes clear. The average predictive improvement provided by election-level interactions over purely additive model specifications is displayed in Table 37.

Table 37: The average percentage point improvement to predictive accuracy yielded by the including election-level interactions within prediction models across all subsets of data relative to additive models. Averages are accompanied by the range of improvements provided by these specifications across all models.

Measure†	**Average Predictive Improvement from Interactions (%)**	**Improvement Range (%)**
MAE	52.63	14.29 – 67.65
DIM	96.96	40.00 – 158.97
LPB	46.59	14.29 – 66.66
APB	50.62	10.00 – 111.10
ABI 1	71.91	29.41 – 137.50
ABI 2	58.79	10.68 – 58.79
SBP	24.69	15.04 – 34.67
LVRC	43.81	11.20 – 69.83

† MAE = mean absolute error, DIM = difference in margin, LPB = leading party bias, APB = average per-party bias, ABI 1 = first measure of average bounded inaccuracy, ABI 2 = second measures of average bounded inaccuracy, SBP = significantly biased poll, LVRC = largest vote share recipient correct.

As focus of these averages is the degree to which the inclusion of interactions is able improve upon additive models, I only include those interactions deemed predictively useful in earlier analysis. That is, averages are calculated across optimally specified interactive models for each subset of data. The degree to which each specified election-level interaction is predictively useful across my additional measures of polling error and each subset of data is displayed in Appendix B across Tables B36 through B56.

From the averages displayed in Table 37, it is clear that including election-level interactions in predictions models improves their ability to accurately predict all measures of polling error. However, the extent to which they do so varies. The inclusion of interactions proves the most useful in predictions of DIM, resulting in an average improvement of ~97% over solely additive models. By contrast, their inclusion proves least useful in predictions of SBP, yielding an average improvement of ~25% over solely additive models. When the range of improvements across subsets is considered, the inclusion of election-level interactions can be seen to bring about predictive improvements of between 10% and ~159% over purely additive models. This underscores their usefulness as predictors and drivers of polling error, though the extent of their usefulness varies considerably on the basis of the subset of data used in predictive modelling. The extent of the impact of differing subsets of data is explored later in the thesis.

On average, the inclusion of interactions between election-level variables proves most useful in the case of models concerned with measures of distributive polling error, producing an average predictive improvement of ~75%. The inclusion of interactions proves least useful in the case of models tasked with predicting measures of polling bias, bringing about an average improvement of ~40%. The variable impact of election-level interactions across conceptualisations further affirms that, while interactions are again universally beneficial to the predictive accuracy of models, the manner in which polling error is conceived bears on the extent to which interactions between electoral characteristics improve model performance.

Though additive and interactive election-level models suggest that electoral characteristics are useful predictors of polling error, both in isolation and in interaction with one another, it is important to establish whether the findings they present are robust to the inclusion of control variables. In the following sub-section, I outline the process of including control variables within my prediction models and analyse the outputs that result from their inclusion.

The Robustness of Prediction Model Findings to Controls

While interactive models better represent the real-world relationship between election-level variables and polling error, they still fall short of representing the multi-level nature of the error itself. As outlined in chapter four, polling error can also be expected to be affected by factors contained within the poll, pollster, and country grouping levels. To establish whether the substantive findings presented by my interactive election level models are robust to such a multi-level specification, I run a series of models containing control variables from the additional grouping levels of interest. The inclusion of control variables in this manner presents problems for predictive modelling that arise from the nature of polling data at large. In the following subsection, I outline the processes necessary to prepare the control variables for inclusion within my prediction models and the constraints placed on analysis of polling error by the present state of global data.

Rank Deficiency and the Need to Recategorise Control Variables

While continuous control variables could be included in prediction models without alteration, categorical variables must be represented numerically before they can be used effectively within models.[1007] A wide range of approaches exist to representing categorical data numerically.[1008] However, many of these are inappropriate for use within my prediction models, as they presume ordinal or numeric relationships between categories which do not

[1007] Hussain Alkharusi, 'Categorical Variables in Regression Analysis: A Comparison of Dummy and Effect Coding', *International Journal of Education*, 4.2 (2012), 202 – 210 (p. 203).
[1008] Kedar Potdar, Taher S. Pardawala, and Chinmay D. Pai, 'A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers', *International Journal of Computer Applications*, 175.4 (2017), 7 -9 (pp. 7 – 8); John T. Hancock and Taghi M. Khoshogftaar, 'Survey on Categorical Data for Neural Networks', *Journal of Big Data*, 7.1 (2020), 1 – 41 (pp. 11 – 38).

exist,[1009] insert random noise that is unrepresentative of the underlying data,[1010] or recode the variable in relation to a referent or exemplar category.[1011] Of the range of possibilities, I represent my categorical variables using binary dummies. This involves representing each category with a binary variable indicating its presence or absence.[1012] Though the cardinality of my categorical variables varies, creating binary dummies to represent each category nevertheless increases the dimensionality of my data and subsequent models,[1013] leading to reduced computational efficiency. However, it does not alter or incorrectly represent the underlying categories, nor does it alter the intended focus of the variables themselves. Due to this, it is the most appropriate transformation for use on my categorical predictor variables.

While it is the most appropriate form of transformation for my categorical variables, the use of binary dummy coding in prediction models is complicated by the use of repeated k-fold cross validation and issues of rank deficiency. Rank deficiency is principally caused by instances of linear dependence between predictor variables or high dimensional data.[1014] While rank deficiency is not a problem unique to prediction models based on k-fold cross validation, the process does render its occurrence more likely. As k-fold cross validation splits a dataset into *k* subsets, it necessarily divides the observations associated with predictor variables *k* times. Within my data, several binary dummies are created from categorical variables with high cardinality. That is, they contain a large number of categories. The cardinality of these variables is such that observations are fragmented across a large number of categories. Producing dummy variables from such fragmentary data results in a substantial number of sparsely populated

[1009] Potdar and Pardawala, p. 7.
[1010] Kevin E. O'Grady and Deborah R. Medoff, 'Categorical Variables in Multiple Regression: Some Cautions', *Multivariate Behavioural Research*, 23.2 (1988), 243 – 260 (p. 244).
[1011] Alkharusi, p. 206.
[1012] Ibid., p. 203.
[1013] Ivan Lopez-Arevalo and others, 'A Memory-efficient Encoding Method for Processing Mixed-type Data on Machine Learning', *Entropy*, 22.12 (2020), 1 – 21 (p. 4).
[1014] Athanassios Kondylis, Ali S. Hadi, and Mark Werner, 'The BACON Approach for Rank-deficient Data', *Statistics in the Twenty-First Century*, 8.3 (2012), 359 – 379 (p. 359).

binaries, containing a larger number of zeroes than ones. When these binary dummies are split into subsets to facilitate k-fold cross validation, the sparsity of positive values causes them to exist as constants with a value of zero in one or more subsets.

The existence of a column of zeroes in a matrix of predictor variables leads to rank deficiency in predictions. This occurs due to the linear dependency. Two vectors of predictor variables are linearly dependent if they are scalar multiples of one another and any set containing a vector of zeroes exhibits linear dependence.[1015] This follows logically, as any set of vectors (or matrix) containing a column of zeroes will necessarily exhibit linear dependence, as all other columns exist as scaler multiples of zero of it, leading to rank deficiency.

To account for the problems of constancy within subsets caused by sparsely populated binaries and avoid rank deficient predictions, it was necessary to recategorise my categorical variables by placing their values into fewer, larger bins. To achieve this, I principally employed an empirically motivated approach to recategorisation. This involved iteratively placing those categories with the fewest occurrences into a 'rare' category until all categories contained sufficient occurrences to be distributed reliably across subsets, preventing rank deficiency. The threshold for inclusion within the 'rare' category varied on the basis of the number of parameters contained within my models and the size of the datasets they relied upon.

Given the large number of under-represented subregions within global pre-election polling, adopting an empirically motivated approach to their recategorisation resulted in the majority of subregions being categorised as 'other' due to their relative scarcity within the data. This led to reductive trinary and binary categorisations that homogenised subregions and represented them in a way that was neither theoretically justifiable, nor representative of the function of the original categorical variable. Importantly, the reductive categorisations were not predictively

[1015] Dan Margalit and Joseph Rabinoff, *Interactive Linear Algebra*, (Atlanta: Georgia Institute of Technology, 2017), p. 69.

useful, as their homogeneity and low cardinality was such that they were providing information already captured (with improved granularity) by the region variable. As such, to be meaningfully included within models, the transformation of subregions required an alternative approach.

Of the global subregions designated by the UN Statistics Division, four possess significantly longer polling histories than all others. These subregions are North America, Western Europe, Northern Europe, and Australia and New Zealand. Across these subregions within the subset of data used for control models, pre-election polling data is available as far back as 1936 in the case of North America, with widespread, publicly available polling data available for all other regions no later than 1960. By comparison, widespread polling data across the remaining subregions is only available after 1996. Given the additional time over which polling has been widely conducted in the four identified subregions, they can be expected to account for a significantly greater number of pre-election polls than the remaining subregions. This expectation is borne out in the subset of my data for which all control variables are available, as these subregions account for 72% of the pre-election polls it contains.

While ideally my categorical variables would have been included unchanged within my prediction models to capture the discrete effect of all categories that they comprise, rank deficiency and the need for recategorisation is an unavoidable artefact of the present state of pre-election polling. At this moment in time, comparatively few polling organisations operate on a reliably cross-national, or cross-regional, basis, with many organisations existing as country- or region-specific operations. Organisations with a limited geographical scope will necessarily factor into the polling of fewer elections in fewer countries. This decreases their likelihood of being included consistently within the training and test subsets required for assessing out-of-sample prediction accuracy, undermining their ability to be used in prediction models individually without introducing rank deficiency.

As the world is only one generation removed from the fall of the Soviet Union and a little over a decade removed from the Arab Spring, global politics is presently characterised by a large number of relatively young democracies. Pre-election polling within these young democracies is in its nascent stages and necessarily spans a shorter amount of time than the polling conducted in older, more established democracies. Therefore, polling data is asymmetrically available across democracies, with an abundance of information available for Western states and a comparative paucity of information available for most others. This asymmetry renders difficult the reliable inclusion of less extensively polled countries within the training and test subsets required for assessing out-of-sample prediction accuracy, undermining their ability to be used in prediction models individually without introducing rank deficiency.

Dealing with these problems of variable scarcity is not a matter of simply gathering more data, as additional data does not currently exist. As such, at this present moment in time, the four-level models proposed within this chapter, inclusive of the necessary instances of re-categorisation they contain, exist as the most comprehensive and appropriate models for predicting polling error permitted by the limitations of existing polling data. As the availability of polling data both broadens and deepens over time with the growth of new democracies and the expansion of polling organisations, the impact of control variables – especially those housed within the pollster- and country-levels – will be able to be explored in an unaltered fashion. However, the present nature of polling data is such that it necessitates the homogenisation of smaller polling organisations and less extensively polled democracies to avoid issues of rank deficiency. While their individual inclusion in models containing fewer parameters (a number of parameters equal to or less than the number of datapoints relating to them) would certainly be possible, such models would fail to capture the multi-level nature of polling error and would therefore be under-specified.

In addition to the need to re-categorise certain variables, the inclusion of controls within prediction models reduces the amount of data available for use within them. This reduction results from the fact that the full range of specified control variables is only jointly available for a subset of polls within my dataset. Ultimately, the inclusion of the full range of poll-, pollster-, and country-level control variables within models reduces the sample size of data available for use within them to 13,189. The model outputs that result from this reduced sample size, along with the recategorized control variables, are presented in the following subsection.

Prediction Models Incorporating Control Variables

In this subsection, I provide the results of prediction models incorporating poll-, pollster-, and country-level control variables. The purpose of including these control variables is to assess whether election-level predictor variables remain useful in the prediction of polling error in their presence. Given this, I include my specified control variables in models prior to layering in election-level variables. I begin by presenting the results of models tasked with the prediction of mean absolute error (MAE). I then move to present the average results of models tasked with the prediction of my additional measures of polling error, along with MAE, with individual model outputs for additional measures provided in Appendix B across Tables B57 through B84.

Table 38 presents the findings that result from layering control variables into a model tasked with the prediction of MAE using polls for which the full range of controls were available. It is clear that even in the presence of controls, election-level variables improve the ability of the model to accurately predict MAE, both additively and interactively. Additively, election-level variables account for a 0.055 reduction in average RMSE and in interaction they cumulatively account for a 0.036 reduction in average RMSE.

Table 38: K-fold cross validated RMSE values for MAE calculated using the subset of data for which all control variables were available (n = 5,432).

Model	Average RMSE	ΔRMSE	ΔRMSE Compared to Null
MAE ~ 1 (Null)	2.646	-	-
+ poll-level controls	2.621	-0.025	-0.025
+ pollster-level controls	2.608	-0.013	-0.038
+ country-level controls	2.226	-0.382	-0.420
+ additive election-level variables†	2.171	-0.055	-0.475
+ turnout change × margin of victory	2.162	-0.009	-0.484
+ turnout change × ENEP	2.149	-0.013	-0.497
+ turnout change × ENEP change	2.134	-0.015	-0.512
+ ENEP × turnout	2.134	0.000	-0.512
+ turnout × ENEP change	2.135	0.001	-0.511

† There was an insufficient number of presidential elections within the subset of data for the election type and round two variables to be included within the model, as their scarcity resulted in rank deficient predictions due to the presence of constants in one or more of the test/train splits generated by the k-fold cross validation process.

In keeping with earlier models, election-level variables prove most predictively useful when included in the model additively, but remain impactful when included interactively. Despite this, the predictive utility of election-level variables is reduced in comparison to earlier models, both additively and interactively. The reduced impact of including election-level variables additively is understandable, given the predictive improvements associated with controls and the order in which variables are added to the model. Control variables are included before election-level variables within the model to assess the ability of electoral characteristics to improve predictive accuracy in their presence. As each set of control variables allows the model to more accurately predict MAE, thereby reducing average RMSE, there is reduced scope for election-level variables to further improve accuracy. Beyond this, a proportion of the variance explained by control variables may have been misattributed to election-level predictors in previous models, further accounting for the reduction in their predictive utility.

In the presence of controls, all bar two of the interactions included in Table 38 remain predictively useful. This indicates that, broadly, election-level interactions remain useful predictors of MAE in the presence of controls. However, the predictively utility of interactions is reduced relative to the additive inclusion of election-level variables. The increased impact of main effects relative to higher order effects can be understood through the hierarchical ordering principle. On the basis of the hierarchical ordering principle, main effects tend to be larger than two-way interactions, which in turn tend to bring about larger effects than three-way interactios.[1016] As such, the diminished of interactions relative to main effects is understandable.

The diminished impact of interactions may also be the result of overfitting, given the point at which they are included within the model. Overfitting concerns the degradation in the ability of prediction models to successfully generalise relationships to out-of-sample data due to containing too many features.[1017] The high number of features results in the identification of highly complex and idiosyncratic relationships within training data that do not exist to the same extent in testing data, resulting in diminished out-of-sample predictive performance.[1018] That these relationships do not exist to the same degree – or sometimes at all – in testing data indicates that they exist either by chance or as an artefact of noise within the training data.[1019]

[1016] Xiang Li, Nandan Sudarsanam, and Daniel D. Frey, ‘Regularities in Data from Factorial Experiments’, *Complexity*, 11.5 (2006), 32 – 45 (p. 34).
[1017] Douglas M. Hawkins, ‘The Problem of Overfitting’, *Journal of Chemical Information and Computer Sciences*, 44 (2004), 1 – 12 (p. 2).
[1018] Xue Ying, ‘An Overview of Overfitting and its Solutions’, *Journal of Physics: Conf. Series*, 1168 (2019), 1 – 7 (p. 1).
[1019] Cullen Schaffer, ‘Overfitting Avoidance as Bias’, *Machine Learning*, 10 (1993), 153 – 178 (p. 153).

While the presence of overfitting may not be immediately apparent in Table 38 – after all, the interactions deemed to be predictively unhelpful do share a common component – the models presented are considerably more highly specified than earlier models due to the presence of controls. Importantly, they are run across a diminished subset of data and therefore entail smaller training and testing sets than previous models. As overfitting is more likely to occur in highly specified prediction models using small training sets,[1020] it is not unreasonable to suggest that it may be responsible for the declining predictive performance presented by the final two model specifications in Table 38. The suggestion of overfitting is provided substantiation by findings relating to MAE displayed later in this chapter, as well as those centring on other continuous measures of polling error presented in Appendix B. Despite this, though the position of the reduction in predictive accuracy in the table lends itself to overfitting, it may simply indicate that interactions containing turnout levels are not useful predictors of MAE in the presence of controls.

Despite the reduction in their predictive impact, election-level variables still improve the ability of the model to accurately predict MAE in the presence of control variables from the poll-, pollster-, and country-level groupings in models. This suggests that substantive conclusions drawn regarding the usefulness of election-level variables as predictors are robust to the presence of controls. To assess whether additional election-level variables and interactions remain predictively useful in the presence on controls, Table 39 presents the reduction in

[1020] Bingzhe Wu and others, 'Reducing Overfitting in Deep Convolutional Neural Networks Using Redundancy Regularizer', in *Artificial Neural Networks and Machine Learning – ICANN 2017*, ed. by Alessandra Lintas and others, (Cham: Springer, 2017), p. 49; Raymond Liu and Duncan F. Gillies, 'Overfitting in Linear Feature Extraction for Classification of High-dimensional Image Data', *Pattern Recognition*, 53 (2016), 73 – 86 (p. 76).

average RMSE associated with election-level characteristics across those elections for which partisanship data could be gathered alongside controls.

Table 39: RMSE values for MAE calculated from prediction models using repeated 10-fold cross-validation across the subset of data for which all control variables were available alongside partisanship figures (n = 4,384).

Model	Average RMSE	RMSE Change	RMSE Change Relative to Null
MAE ~ 1 (Null)	2.390	-	-
+ poll-level controls	2.354	-0.036	-0.036
+ pollster-level controls	2.339	-0.015	-0.051
+ country-level controls†	1.987	-0.352	-0.403
+ additive election-level variables‡	1.918	-0.069	-0.472
+ turnout change × margin of victory	1.896	-0.022	-0.494
+ turnout change × ENEP	1.896	0.000	-0.494
+ turnout change × ENEP change	1.895	-0.001	-0.495
+ ENEP × turnout	1.894	-0.001	-0.496
+ turnout × ENEP change	1.880	-0.014	-0.510
+ ENEP × partisanship	1.879	-0.001	-0.511
+ turnout change × partisanship	1.880	0.001	-0.510
+ turnout × partisanship	1.880	0.000	-0.510
+ turnout change × ENEP × partisanship	1.881	0.001	-0.509
+ turnout chg. × ENEP chg. × partisanship	1.881	0.000	-0.509
+ ENEP × turnout × partisanship	1.874	-0.007	-0.516
+ turnout × ENEP chg. × partisanship	1.871	-0.003	-0.519

† Given the composition of the subset of data containing partisanship measurements, it was not possible to include the electoral system variable within country-level controls, as the distribution of system types was such that even reducing it to a binary variable failed to resolve issues of rank deficiency.
‡ There was an insufficient number of presidential elections within the original subset of data for the election type and round two variables to be included within the model.

The results provided in Table 39 tell a similar story to those presented by previous prediction models incorporating controls. The inclusion of election-level variables, both additively and interactively, remains useful in the prediction of MAE. This suggests that conclusions regarding their predictive utility drawn from earlier models are robust to the inclusion of control variables. Additively, election-level variables inclusive of partisanship reduce average RMSE by 0.069, while two- and three-way interactions between election-level variables collectively reduce average RMSE by 0.047. As such, election-level variables again prove more predictively useful when included in models additively, rather than interactively. However, the impact of both forms of inclusion is diminished relative to earlier models absent controls.

Despite their broad usefulness as predictors, several interactions displayed in Table 39 do not retain their predictive utility in the presence of controls. The two-way interactions between turnout change no longer brings about a reduction in average RMSE, nor does the interaction between turnout change and ENEP. Similarly, the three-way interaction between turnout change, ENEP change, and partisanship also no longer yields a predictive improvement. This suggests the that the predictive utility presented by these interactions in earlier modelling is not robust to the presence of controls and is better attributed to variables present within the control groupings. This illustrates the importance of subjecting models to controls in order to avoid errantly assigning importance to relationships that are not robust to their presence. Nevertheless, the majority of interactions displayed within the table remain predictively useful in the presence of controls, suggesting that earlier conclusions surrounding their utility are substantively robust.

Table 40 displays the results of running models tasked with the prediction of MAE across the subset of data for which control variables and the extent of late decision-making within the electorate were jointly available. From the table, it is clear that including election-level variables additively within the model again improves the ability of models to accurately predict MAE, even in the presence of controls, yielding a reduction of 0.007 in average RMSE.

However, ostensibly, election-level interactions cannot be seen to improve the predictive accuracy of the model.

Table 40: K-fold cross validated RMSE values for MAE including election-level interactions and based on the subset of data for which all control variables were available alongside late decision-making figures (n = 1,557).

Model	Average RMSE	RMSE Change	RMSE Change Relative to Null
MAE ~ 1 (Null)	1.699	-	-
+ poll-level controls	1.530	-0.169	-0.169
+ pollster-level controls	1.506	-0.024	-0.193
+ country-level controls†	1.463	-0.043	-0.236
+ additive election-level variables‡	1.456	-0.007	-0.243
+ turnout change × margin of victory	1.456	0.000	-0.243
+ turnout change × ENEP	1.456	0.000	-0.243
+ turnout change × ENEP change	1.456	0.000	-0.243
+ ENEP × turnout	1.456	0.000	-0.243
+ late deciders × ENEP	1.456	0.000	-0.243
+ late deciders × turnout	1.456	0.000	-0.243
+ late deciders × turnout change	1.456	0.000	-0.243

† Given the composition of the subset of data containing late decision-making measurements, it was not possible to include the electoral system or polling moratorium variables within country-level controls, as even reducing system to a binary variable failed to resolve issues of rank deficiency.
‡ There was an insufficient number of presidential elections within the subset of data for the election type and round two variables to be included within the model, as their scarcity resulted in rank deficient predictions due to the presence of constants in one or more of the train/test splits generated by k-fold cross validation. Similarly, there were no instances of system change within the subset, precluding its inclusion within models.

Two opposing conclusions can be drawn from the inability of election-level interactions to improve predictive accuracy in the model displayed in Table 40. The first is that interactions simply do not serve as useful predictors of MAE across the subset of data for which late decision-making data and control variables are available, with earlier reductions to average RMSE not being robust to controls. The second is that the lack of predictive improvement is the result of overfitting, given the substantially diminished subset of data used and the high number of features contained within the model. Table 41 explores these conclusions by assessing the predictive utility of election-level interactions across models incorporating a reduced number of features.

Table 41: Unpacking the effect of overfitting on the predictive utility of election-level interactions in models tasked with predicting MAE values across the subset of data for which all control variables were available alongside late decision-making figures (n = 1,557).

Model	Average RMSE	RMSE Change	RMSE Change Relative to Null
MAE ~ 1 (Null)	1.699	-	-
+ poll-level controls	1.530	-0.169	-0.169
+ pollster-level controls	1.506	-0.024	-0.193
+ country-level controls†	1.463	-0.043	-0.236
+ turnout change × margin of victory	1.463	0.000	-0.236
+ turnout change × ENEP	1.465	0.002	-0.234
+ turnout change × ENEP change	1.457	-0.008	-0.242
+ ENEP × turnout	1.445	-0.002	-0.244
+ late deciders × ENEP	1.432	-0.013	-0.257
+ late deciders × turnout	1.432	0.000	-0.257
+ late deciders × turnout change	1.432	0.000	-0.257

† Given the composition of the subset of data containing late decision-making measurements, it was not possible to include the electoral system or polling moratorium variables within country-level controls, as even reducing system to a binary variable failed to resolve issues of rank deficiency.

The model displayed in Table 41 minimises the potential for overfitting by only including control variables and election-level interactions. By doing so, it reduces the number of features in the model by the maximum extent possible while preserving the ability to test the robustness of interactions to controls. Results indicate that three of the interactions included within models – those between turnout change and ENEP change, ENEP and turnout, and late decision-making and ENEP – yield improvements to predictive accuracy in the presence of controls when the number of features used for modelling is reduced. This suggests that their inability to improve predictive accuracy in earlier modelling was an artefact of overfitting, rather than reflective of their lack of predictive utility. However, the other interactions in question failed to bring about improvements to predictive accuracy, even in models with reduced features. This indicates that their ability to improve predictions of MAE is not robust to the presence of controls.

The findings displayed in Table 41 indicate that several election-level interactions improve the ability of models to predict MAE values across the subset of data for which late decision-making data is available, even in the presence of controls. However, it is clear that the extent of their ability to do so is obscured by overfitting in highly specified models. Due to the impact of overfitting, models containing a reduced feature set are included in addition to fully specified models in the outputs presented for my additional measures of polling error in Appendix B.

The results derived from the additional models in Appendix B tell a similar story to those seen in earlier models concerned with MAE. While election-level interactions often fail improve predictive accuracy in highly specified models, their predictive utility is evident in models run across a reduced feature set. Indeed, the predictive utility of many election-level interactions can be seen to be robust to the inclusion of control variables across all measures of polling error. The greatest number of interactions retain their predictive utility in the case of LVRC, while the least number of interactions remain predictively useful in the case of LPB. This suggests that the predictive utility of election-level interactions across the subset of data containing late decision-making figures is most robust to controls in the case of substantive inaccuracy and least robust in the case of polling bias. More broadly, it further underscores the role of overfitting in obscuring the impact of erstwhile predictively useful interactions. The impact of overfitting is further explored later in this thesis using alternative modelling approaches incorporating automatic feature selection.

The degree to which all findings related to my other measures of polling error are robust to the presence of controls is presented in Appendix B across Tables B57 through B84. The average reduction in RMSE associated with the inclusion of election-level variables, both additively and interactively, in these control models, as well as those concerning MAE, is displayed below in Table 42. As before, reductions are presented as proportions to enable cross-measure comparison.

Table 42: The average reduction in RMSE that results from including election-level variables both additively and interactively in models concerning continuous measures of polling error that contain poll-, pollster-, and country-level controls.

Error Measure†	**Average RMSE Reduction in the Presence of Controls (%)**
MAE	20.78
DIM	123.69
LPB	50.23
APB	33.29
ABI 1	17.53
ABI 2	18.71

†MAE = mean absolute error, DIM = difference in margin, LPB = leading party bias, APB = average per-party bias, ABI 1 = average bounded inaccuracy measure 1, ABI 2 = average bounded inaccuracy measure 2.

From Table 42, it is clear that, on average, the inclusion of election-level variables, both additively and interactively, improves the ability of models to predict all continuous measures of polling error in the presence of controls by reducing the average RMSE they exhibit. This indicates that the findings derived from earlier models are substantively robust, as the inclusion of control variables does not eliminate the predictive utility of election-level variables, and lends support to the core contention of my second hypothesis. While the presence of controls does diminish the extent to which election-level variables prove predictively useful relative to earlier models, they still improve the prediction of continuous measures of polling error by at least ~18%.

Election-level variables can be seen to be most useful as predictors of DIM and least useful as predictors of ABI 1. However, it is immediately clear that the decrease in average RMSE associated with DIM is anomalous when compared to other measures. The substantial decrease in average RMSE associated with DIM is largely driven by the two-way interaction between turnout change and margin of victory in the reduced model run across the subset of data containing late decision-making data (displayed in Table B60 of Appendix B). This suggests that this interaction is a particularly pronounced predictor of DIM in those elections contained within this subset. Excluding this finding lowers the reduction in average RMSE associated with DIM to 63.75%, bringing it more closely in line with other measures. Despite this reduction, the substantive conclusion remains the same: in the presence of controls, election-level variables are more useful as predictors of DIM than other continuous measures of polling error.

When considered cumulatively, in the presence of controls, election-level variables are most useful in predicting measure of distributive polling error (MAE and DIM), even when excluding the anomalous finding related to DIM. By contrast, in control models, election-level variables are least useful in predicting measures of bounded polling error (ABI 1 and ABI 2). While the ability of election-level variables to improve predictions of all conceptualisations of polling error offers encouraging support for my second hypothesis, the increased utility of election-level variables as predictors of polling bias (LPB and APB) when compared to bounded inaccuracy (ABI 1 and ABI 2) runs counter to theoretical expectations. That election-level variables prove more useful as predictors of measures of polling bias in the presence of controls than expected a priori suggests that the degree of bias exhibited by polls is not simply a factor of partisan biases or over-correction and the pollster level and is, in fact, affected profoundly by differences in electoral characteristics.

Table 43 presents the average improvement to correct classification rates associated with including election-level variables, both additively and interactively, in prediction models concerning binary measures of polling error that contain control variables. The model outputs used to calculate these averages are presented in Appendix B.

Table 43: The average improvement in predictive accuracy gained by including election-level variables additively and interactively in models containing poll-, pollster-, and country-level controls.

Error Measure†	Average Classification Improvement in the Presence of Controls (%)
SBP	8.62
LVRC	8.45

†SBP = significantly biased poll,
LVRC = largest vote share recipient correct

From Table 43, it is clear that, on average, election-level variables improve the ability of models to predict substantive polling error (LVRC) and binary measures of polling bias (SBP), even in the presence of controls. Interestingly, on average, the improvement associated with including election-level variables in models tasked with predicting the presence of SBP is fractionally larger than that associated with models tasked with predicting LVRC. While this finding runs counter to theoretical expectations, that election-level variables prove predictively useful in the presence of controls from different grouping levels indicates that earlier findings were substantively robust and lends support to the main thrust of my second hypothesis.

While Tables 42 and 43 assess the cumulative impact of election-level characteristics and interactions in prediction models containing controls across individual measures of polling error, Table 44 unpacks their individual predictive utility across my three conceptualisations of polling error, and polling bias.

Table 44: The average predictive benefit associated with including election-level variables both additively and interactively in prediction models containing controls across measures of distributive, bounded, and substantive polling error, as well as polling bias.

Polling Error Type	**Average Additive Improvement (%)**	**Additive Improvement Range (%)**	**Average Interactive Improvement (%)**	**Interactive Improvement Range (%)**
Distributive	25.54	2.97 – 73.30	44.35	6.44 – 222.49
Bounded	8.33	4.79 – 12.56	9.55	3.95 – 20.96
Substantive	5.08	4.17 – 6.16	4.40	2.05 – 7.74
Bias	20.89	0.83 – 59.09	11.06	0.69 – 26.76

Table 44 indicates that the inclusion of electoral characteristics into models, both additively and interactively, improves their ability to predict all forms of polling error in the presence of controls. When included additively, election-level variables improve predictions of measures of distributive polling error more acutely than other conceptualisations. By contrast, on average, the additive inclusion of election-level variables improves predictions of substantive polling error to the lowest extent. While this runs counter to theoretical expectations, when the additive improvement ranges are considered, the lowest extent of improvements associated with polling bias is far smaller than the lowest extent of other ranges. As such, while election-level variables bring about a sizeable improvement to predictions of polling bias on average, the range associated with polling bias also houses the smallest individual improvement across all conceptualisations of error.

In the same vein as including election-level variables additively, the inclusion of election-level interactions also improves the ability of models to accurately predict all of my conceptualisations of error in the presence of controls. Election-level interactions yield the largest predictive improvement in models concerned with measures of distributive accuracy. However, it is immediately apparent that the uppermost extent of the improvement range associated with distributive polling error is anomalous relative to other conceptualisations. This anomaly is again driven by the outsized impact of the two-way interaction between turnout change and margin of victory in the reduced model used to predict DIM across the subset of data containing the extent of late decision-making amongst the electorate (presented in Table B60 of Appendix B). Excluding this finding lowers the uppermost end of the improvement range to 21.44 and lowers the average improvement associated with interactions to 8.72. This alters substantive conclusions as, while election-level interactions remain useful predictors of distributive polling error in the presence of controls, after the exclusion of the anomalous result, they prove most useful as predictors of polling bias on average, running counter to theoretical expectations. Despite this, as is the case in additive findings, the lowest extent of the interactive improvement range associated with polling bias remains substantially lower than other conceptualisations of error, indicating that it again encompasses the lowest individual improvement in predictive accuracy.

From Table 44, it is clear that, on average, the inclusion of election-level interactions brings about the smallest predictive improvement to models concerned with substantive polling error. Though this runs counter to theoretical expectations, the findings presented in the table provide strong supportive evidence for the main contention of my second hypothesis, namely that election-level variables will prove useful as predictors of polling error both additively and interactively.

Despite their universal utility, across most conceptualisations of polling error, the predictive improvement yielded by two- and three-way interactions is diminished relative to the inclusion of individual election-level variables additively. It is clear from the outputs presented in Appendix B that the usefulness of interactions as predictors of my conceptualisations of error ultimately varies between the subsets of data used for modelling. This phenomenon was evident across previous model outputs and is explored in greater depth in the following section. More broadly, the generally diminished impact of two- and three-way interactions in the presence of controls is to be expected. Not only does the hierarchical ordering principle lead to the expectation that interactions will have a reduced effect relative to main effects,[1021] but the inclusion of sets of relevant predictor variables from other grouping levels as controls – each of which improves the ability of models to predict polling error in its own right – is sufficient to occlude the small improvements yielded by two- and three-way interactions.

Models inclusive of controls also include a far greater number of variables than earlier, election-level only models. As such, they are more likely to violate the modelling ideal of parsimony, increasing the likelihood of overfitting. Given that election-level variables are layered into the model after controls, they are more likely to be responsible for violations of parsimony and overfitting, leading to degraded predictive performance. Future work ought to be conducted to establish whether the ordering of predictor variables affects model outputs. Such re-ordering is unlikely to alter substantive conclusions – namely, that election-level variables are useful predictors of polling error – but may alter findings in relation to the usefulness of individual two- and three-way interactions between electoral characteristics.

It would be remiss to address control models without unpacking the effect of the control variables themselves. Across all measures of polling error, control variables housed within the

[1021] Li, p. 34.

poll-, pollster-, and country-level grouping clusters improve the predictive ability of models, as was expected a priori. Given their nature as controls, secondary to the focus of the thesis, variables from these additional grouping levels were homogenised and layered into models as discrete sets (see: Section B6 of Appendix B). As such, the impact of individual control variables cannot be unpacked. However, the performance of each set of controls can nevertheless be evaluated. Considered broadly, each set of control variables proved predictively beneficial across all models and all measures of polling error. Across measures of distributive (MAE and DIM), bounded (ABI 1 and ABI 2), and substantive polling error (LVRC), country-level variables proved the most predictively useful set of controls. This trend was also observed across most models concerned with leading party bias (LPB) and average party bias (APB). Despite this, the set of poll-level variables proved the most predictively useful control set in models tasked with classifying significantly biased polls (SBP). This suggests that while country-level variables generally stand as the most predictively useful controls, poll-level controls bear more closely on instances of significant polling bias.

That each set of control variables proved predictively useful in its own right not only affirms the importance of their inclusion in models to evaluate the robustness of election-level findings, but also speaks to the importance of adopting a multi-level approach to understanding and predicting polling error. To exclude the poll, pollster, and country levels from models of polling error is to fundamentally misrepresent the nature of its sources and to lose a wide range of predictively valuable indicators. Future work ought therefore to adopt a multi-level approach to polling error, incorporating the four levels elaborated in this thesis, and seek to fully explore the utility of predictor variables beyond the election-level grouping cluster.

To bring my findings together, in the following sub-section I summarise the results presented across additive, interactive, and control-based prediction models. In doing so, I provide a concise answer to my third research question, outline support for my second hypothesis, and identify those electoral characteristics most useful in predicting my measures of polling error and, therefore, accounting for the variance they present. I also identify trends and limitations within my main analysis for exploration in the following section.

The Usefulness of Electoral Characteristics as Predictors of Polling Error Variance

Through the analysis contained in this chapter, I have demonstrated that electoral characteristics stand as useful predictors of variance in pre-election polling error. Additive models demonstrate the importance of individual electoral characteristics as predictors of polling error. The findings that result from these models indicate that electoral characteristics improve the ability of models to predict polling error by an average of 26.29% in the case of substantive error, 7.67% in the case of bias, 6.34% in the case of bounded error, and 6.16% in the case of distributive error. These findings indicate that election-level variables meaningfully improve the ability of models to accurately account for observed variance across all of my measures of polling error by reducing the distance between predict values and observables. This supports the substantive contention put forward in my second hypothesis that, additively, election-level variables would prove predictively useful.

When unpacked, the findings that result from additive models also support the secondary contention in hypothesis two that election-level variables would be the most useful in relation to predictions of substantive polling error. However, the findings suggest that election-level variables are more predictive of polling bias in the aggregate than expected. This finding is largely driven by my measure of significantly biased polls (SBP), as models perform more poorly in relation to other measures of bias (LPB and APB). On average, election-level variables prove less useful as predictors of LPB than all measures distributive, bounded, and substantive polling error. However, they prove to be more useful predictors of APB than both DIM and ABI 1. As such, while some evidence exists in support of the expectation that election-level variables would prove less useful as predictors of polling bias than of other conceptualisations of error, findings are dependent on the manner in which bias is measured.

Several election-level variables cross the boundaries between conceptualisations of polling error, proving useful predictors of bounded and distributive polling error, as well as polling bas. The strength of partisan loyalty amongst the electorate serves as a useful predictor of all continuous measures of polling error, allowing models to more accurately account for their variance by reducing the distance between predicted values and observations. From this, differences in the strength of partisan loyalty between elections can be understood to exist as important drivers of the variance exhibited by continuous measures of polling error writ large.

The ideological distance between the parties or candidates contesting an election also universally aids models in predicting measures of distributive and bounded polling error, as well as polling bias. This suggests that differences in the ideological distance between parties and candidates between elections bears closely on the variance exhibited by continuous measures of polling error and indicates that they serve as wide-ranging drivers of polling error. Similarly, both ENEP levels and ENEP change between contests also aid models to accurately predict all continuous measures of polling error. This suggests that the extent of distributive and bounded polling error, as well as polling bias, is driven by these variables.

The margin of victory associated with an election also improves the ability of models to accurately predict all measures of continuous polling error. This indicates that distributive polling error, bounded polling error, and polling bias are all driven by differences in the margin of victory in elections. In the same vein, both election type and system change also improve the predictive performance of models across all continuous measures of polling error. This suggests that differences in election type (legislative vs presidential) and instances of electoral system change between contests bear upon the extent to which the predicted vote share distributions offered by polls diverge from results, the degree to which the prediction error they present violates the boundaries set by the margin of error, and their propensity to exhibit bias.

Though a high degree of overlap exists with regards to the predictive utility of election-level variables across continuous measures of polling error, certain variables are especially useful for the prediction of individual conceptualisations or measures of polling error. Turnout is a particularly useful predictor of measures of both bounded and distributive polling error, but does not benefit models concerned with polling bias. This suggests that the observed variance in measures of distributive and bounded polling error shares a common driver. This is unsurprising given the theoretical inter-connection between the two conceptualisations, with heightened levels of distributive error directly increasing the extent of bounded error, leading to the expectation that they ought to vary in tandem.

By comparison, turnout change between elections proves considerably more useful as a predictor of APB than other continuous measures of polling error. This indicates that, while its predictive impact is generally diffuse across measures, turnout change is a particularly pronounced driver of the average per-party bias exhibited by polls. In the same vein, the round two election binary produces notable predictive improvements in the case of measures of distributive polling error, but does not have a meaningful impact on models concerned with bounded polling error or polling bias. This suggests that it bears more closely on the extent to which polls present errant vote share distributions and misrepresent the margin between leading candidates than other continuous measures of error.

Across models concerned with continuous measures of polling error, the extent of late decision-making amongst the electorate brings about by far the largest improvements to predictive accuracy. However, it proves considerably more useful as a predictor of distributive and bounder polling error, yielding considerably diminished improvements in the case of polling bias. This indicates that, while late decision-making stands as a prominent driver of variance in MAE, DIM, and ABI variants, it is far less prominent as a driver of LPB and APB. Indeed, in the case of APB, it fails to improve the predictive accuracy of models.

Equally, the type of election being polled (legislative or presidential) is more important for predictions of bounded polling error, allowing models to account for a greater amount of its observed variance than other conceptualisations of polling error. This indicates that the extent of variation in the bounded error exhibited by polls is more closely driven by differences in election type than other conceptualisations of error.

Across models concerned with binary measures of polling error, the majority of election-level variables stand as useful predictors of both SBP and LVRC, improving the performance of classification models. This underscores the ability of election-level variables to transcend conceptual boundaries and prove useful as predictors of differing conceptualisations of polling error. Despite this, certain election-level variables prove more useful as predictors of substantive polling error than significant polling bias and vice versa. Both the margin of victory in an election and the round two election binary prove useful as predictors of substantive polling error, but impede the ability of models to correctly classify instances of significant polling bias. Conversely, differences in the number of registered voters between elections exists as a useful predictor of significant polling bias, but undermines the performance of models tasked with classifying instances of substantive polling error.

While my additive prediction models demonstrate the usefulness of electoral characteristics as predictors of polling error in isolation, my interactive models show that they are also predictively useful when considered in interaction with one another. In comparison to additive models, the inclusion of election-level interactions improves the average predictive accuracy of models by ~75% in the case of distributive polling error, ~65% in the case of bounded error, 43.81% in the case of substantive error, and 40.63% in the case of polling bias. This indicates that, while the inclusion of election-level interactions substantially improves the ability of models to correctly predict all conceptualisation of polling error, they prove most useful as predictors of distributive polling error and least useful as predictors of polling bias.

The aggregate findings provided by my interactive prediction models provide substantive support for my second hypothesis, insofar as they demonstrate that election-level interactions improve the ability of models to predict polling error variance. They also provide support for the secondary contention that election-level interactions would prove least useful as predictors of polling bias. However, they do not support the contention that they would be most useful in predictions of substantive error. Rather, they indicate that election-level interactions are most useful as predictors of distributive polling error. Future scholarship ought to further investigate the relationship between election-level interactions and measures of polling error to unpack and better understand the nature of these relationships.

In addition to aggregate findings, my interactive prediction models also identify a series of specific two- and three-way interactions between electoral characteristics that are particularly useful in predicting polling error variance. Much as was the case with additive models, many interactions prove predictively useful across all continuous measures of polling error. The two-way interaction between turnout change and margin of victory improves the ability of models to correctly predict all continuous measures of polling error, but proves most useful as a predictor of bounded polling error. This indicates that, while the interaction between these election-level variables affects all forms of continuous polling error, it most prominently drives variance in average bounded inaccuracy.

Of the universally beneficial interactions, the two-way interaction between late decision-making and ENEP yields the largest improvements to predictive accuracy, but proves most beneficial to predictions of distributive and bounded polling error. The tendency of those interactions that improve predictions of all continuous forms of polling error to bear more closely on certain conceptualisations and measures than others is widely evident. The two-way interaction between turnout change and ENEP universally improves the ability of models to accurately predict continuous measures of polling error. However, it proves considerably more useful as a predictor of LPB than other measures. Similarly, the interaction between turnout change and ENEP change also serves to universally improve predictive accuracy, but proves most useful as a predictor of DIM. Finally, the interaction between turnout and ENEP change also enables models to more accurately predict all continuous measures of polling error, but most acutely improves predictions of APB.

While some interactions universally improve the ability of models to predict continuous measures of polling error, others do not. Rather, they only improve predictions of given conceptualisations of polling error. The two-way interaction between turnout and partisanship and the three-way interaction between turnout change, ENEP change, and partisanship prove predictively useful for measures of distributive and bounded error, but do not improve the ability of models to predict polling bias. Equally, the interaction between late decision-making and turnout changes is predictively useful in models concerned with distributive polling error and polling bias, but does not aid those tasked with predicting bounded inaccuracy. The three-way interaction between turnout, ENEP change, and partisanship is most useful as a predictor of bounded polling error, while the two-way interaction between turnout change and partisanship aids models in predicting bounded polling error and polling bias, but is not useful as a predictor of distributive inaccuracy.

The results from models tasked with predicting continuous measures of polling error indicate that election-level interactions stand as useful predictors of polling error. This provides support for the substantive contention of my second hypothesis. However, it is clear that while many interactions improve the ability of models to predict all continuous measures of polling error, the benefit associated with others is more specific. Each conceptualisation of polling error has election-level interactions that bear more closely upon it than others. While a greater number of interactions prove more useful as predictors of distributive and bounded error than polling bias, each conceptualisation of error present interactions that do not prove predictively useful. As such, while some evidence exists that election-level interactions prove least useful as predictors of polling bias, findings are far from conclusive.

The impact of election-level interactions on the ability of models to correctly classify binary measures of polling error tells a similar story. Much like models concerned with continuous polling inaccuracy, the majority of election-level interactions universally improve the ability of models to correctly classify measures of binary polling error. However, differences in predictive usefulness are present. The two-way interaction between ENEP and turnout proves most useful for the classification of LVRC, but undermines the ability of models to correctly classify instances of SBP. Similarly, the three-way interaction between ENEP, turnout, and partisanship provides the largest improvement to the ability of models to classify SBP, but presents a much-reduced improvement in the case of LVRC.

Differences in predictively utility are evident across a range of interactions. While the interactions between late decision-making and both ENEP and turnout change are useful predictors of both binary measures of polling error, they prove most beneficial for predictions of LVRC. By contrast, the three-way interaction between turnout change, ENEP change, and partisan also improves the ability of models to accurately predict both binary measures of polling error, but most acutely benefits predictions of SBP. The interaction between turnout change and ENEP also proves more beneficial in the case of SBP, as it undermines the accuracy of models tasked with correctly classifying instances of LVRC. Despite this, a greater number of election-level interactions prove useful as predictors of LVRC than SBP, though the difference is small.

Overall, the impact of election-level interactions on the ability of models to correctly classify binary measures of polling error lends support to the substantive contention of my second hypothesis, insofar as interactions improve model accuracy. Much like earlier models, findings also offer a degree of support to the secondary contention that election-level interactions will prove less useful as predictors of polling bias, as a greater number of interactions prove predictively useful in the case of LVRC. However, the average impact of interactions runs counter to this expectation, albeit marginally.

In models containing controls, the additive inclusion of election-level variables universally improves the ability of models to predict variance across all measures of polling inaccuracy. However, their inclusion proves most beneficial for predictions of variance in distributive polling error and least useful in the prediction of variance in substantive polling error. This runs counter to expectations and speaks to a diminished relationship between substantive polling error and election-level variables. Future research ought therefore to further unpack the relationship between electoral characteristics and substantive polling error to better understand its nature. Despite this, the fact that the additive inclusion of election-level variables improves the ability of models to accurately predict all measures of polling error indicates that findings derived from earlier models are substantively robust and lends further support to the main thrust of my second hypothesis.

While electoral characteristics universally improve predictive accuracy when inputted into models additively, the inclusion of controls affects the improvement associated with two- and three-way interactions between them. In the presence of controls, election-level interactions prove most useful as predictors of distributive polling error on average and least useful as predictors of substantive error. This mirrors the findings derived from including election-level variables into control models additively and suggests that election-level variables, both additively and interactively, bear more closely on measures of distributive polling error than others. It also underscores the diminished relationship between election-level variables and substantive polling error.

The finding that, in the presence of controls, election-level interactions are most useful as predictors of measures of distributive inaccuracy is largely driven by the outsized impact of the two-way interaction between turnout change and margin of victory on predictions of DIM across the subset of elections for which late decision-making data could be gathered. While this may simply indicate that this interaction serves as a particularly pronounced driver of polling error in polls conducted for these elections, its removal as an anomaly alters substantive conclusions. In the absence of this effect, the inclusion of election-level interactions in control models proves most predictive of measures of polling bias on average, again running counter to theoretical expectations. Nevertheless, the improvements associated with predictions of polling bias present the largest range and its lowest extent displays by far the smallest individual improvement of all measures addressed. As such, while on average the inclusion of election-level interactions in the presence of controls improves predictions of polling bias to the greatest degree, the extent of these improvements is highly variable across subsets, encompassing the smallest individual improvement.

The variable extent of the relationship between election-level interactions and measures of polling bias is further evident in the consistency of findings. In the presence of controls, election-level interactions more consistently improve predictions of measures of distributive and bounded polling error than polling bias. That predictions of polling bias present greater average improvements in spite of this inconsistency suggests that measures of polling bias are principally driven by specific interactions, or that average findings are chiefly driven the predictive impact of interactions in given subsets or in relation to given approaches to measurement. This inconsistency supports the wide range associated with improvements to predictions of polling bias and further underscores the variable relationship between election-level interactions and measures of polling bias.

While these findings suggest that, across certain subsets of elections, election-level interactions improve predictions of polling bias to a lesser extent than other conceptualisations and that, in the presence of controls, individual interactions bear more consistently on measures other than bias, election-level variables nevertheless bear more closely on polling bias than would be expected a priori. Future research should therefore also investigate the relationship between election-level variables and polling bias, as findings suggest that the degree of bias exhibited by polls is not simply a factor of partisan biases on the part of pollsters or an artefact of over-correction in response to past misses but is, in fact, affected by differences in electoral characteristics and interactions between them to a considerable, albeit variable and often inconsistent, extent.

Across models containing control variables, it is also clear that the inclusion of controls generally diminishes the extent to which election-level interactions aid the predictive performance of models. The diminished performance of interactions is to be expected on the basis of the hierarchical ordering principle,[1022] and may be exacerbated by overfitting as election-level interactions are added last within my model specifications. In the case of control models, this means that they are included after a considerable number of additional variables, resulting in the risk of overfitting, especially in diminished subsets of data. Overfitting of this kind was found to be particularly pronounced in control models relating to the subset of data

[1022] Li, p. 34.

including late decision-making figures. I explore the potential for mitigating the effect of overfitting on findings using alternative estimative approaches in the following section.

Despite the variable findings concerning polling bias and the generally diminished nature of interactions, the outputs that result from including election-level interactions in models containing controls support the substantive core of my second hypothesis, insofar as they demonstrate that interactions between electoral characteristics serve as useful predictors of polling error, even in the presence of variables from other grouping levels.

Though election-level variables stand as useful predictors of all measures of polling error across my main analyses, both additively and interactively, the models on which these analyses are based all rest on linear assumptions. In addition to this, a series of themes emerge across their outputs that are worthy of further investigation. To ensure the robustness of my findings, in the following section I explore alternative modelling strategies and unpack two key themes in my analysis: suspected overfitting and the variability of the predictive utility of election-level variables across differently composed subsets of data.

6.3: Exploring Different Modelling Approaches and Emergent Themes

The models used to test the ability of election-level variables to predict variance in polling error within my main analysis assumed the presence of a linear relationship and did not allow for higher order interactions between predictors. While these choices were theoretically motivated and proved predictively useful, their exclusive use begs the question whether alternative modelling approaches that allow for non-linearity and complex interactivity may prove beneficial.

Precedent for expecting non-linear and complexly interactive relationships between election-level variables and polling error can be found in works that suggest elections as chaotic phenomena.[1023] The variables comprised by chaotic phenomena are not only highly interconnected, but also evolve together in a non-linear manner.[1024] Given the suggestion of chaos at the election-level, models that allow for non-linearity and greater interactive complexity may better represent the relationship between election-level variables and polling error variation, proving predictively beneficial.

To assess the benefits of non-linear modelling, I employ random forests which permit non-linear relationships between election-level variables and polling error.[1025] Random forests are an application of ensemble-based prediction, which involves the use of multiple models in tandem to predict a given outcome.[1026] These models take the form of decision trees that are based on samples drawn using bootstrapped aggregation, with each tree assuming a random selection of model features.[1027] Regression-based random forests return the average of the predicted values across these trees, while classification-based forests output the class identifier selected by the majority of trees.[1028]

To explore the utility of models permitting higher order interactions, I employ fully interactive linear regressions and generalised linear models. These models permit interactions between all election-level predictor variables to accommodate the potential for complex interactivity between the electoral characteristics beyond that theorised within this thesis and explored in earlier analysis.

[1023] Scott de Marchi, 'Adaptive Models and Electoral Instability', *Journal of Theoretical Politics*, 11.3 (1999), 393 – 419 (p. 399); Jens Koed Madsen, *The Psychology of Micro-targeted Election Campaigns*, (Oxford: Palgrave Macmillan, 2019), p. 284.
[1024] Ibid.
[1025] Lidia Auret and Chris Aldrich, 'Interpretation of Nonlinear Relationships Between Process Variables by Use of Random Forests', *Minerals Engineering*, 35 (2012), 27 – 42 (p. 28).
[1026] Khaled Fawagreh, Mohamed Medhat Gaber, and Eyad Elyan, 'Random Forests: From Early Developments to Recent Advancements', *Systems Science and Control Engineering*, 2 (2014), 602 – 609 (p. 603).
[1027] Ibid.
[1028] Tin Kam Ho, 'Random Decision Forests', *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, 1 (1995), 278 – 282 (pp. 278 – 282).

The modelling principle of parsimony holds that models ought to contain only those elements necessary for successfully modelling the outcome of interest and nothing more.[1029] Models that are not parsimonious, and therefore contain more elements than is necessary, run the risk of overfitting.[1030] Signs of overfitting were evident across my earlier prediction models, most notably those including control variables. While a range of approaches to dealing with overfitting exist, a prominent method aimed at its mitigation is feature selection.[1031] Feature selection distinguishes between those model features that are useful for modelling an outcome of interest and those that are not on the basis of some empirical criterion.[1032] As such, it presents an intuitive approach to reducing overfitting as, through the removal of less useful features, it brings models closer to the parsimonious ideal.

I employ Least Absolute Shrinkage and Selection Operator (LASSO) regression to address perceived issues of overfitting in relation to my continuous measures of polling error as it engages in feature selection. Predictor variables are retained within, or excluded from, LASSO regression models on the basis of their ability to meaningfully improve predictive accuracy.[1033] This is achieved through shrinkage which results in the coefficients associated with minimally useful or redundant predictors tending towards zero, resulting in their removal from consideration.[1034] In this way, extraneous features are removed from models, aiding in the prevention of overfitting.

To explore the impact of these alternate modelling strategies, I begin by providing the raw RMSE reductions associated with the inclusion of election-level variables within these models

[1029] Hawkins, p. 1.
[1030] Ibid.
[1031] Ying, p .1.
[1032] Ibid.
[1033] J. Ranstam and J. A. Cook, 'LASSO Regression', *Journal of British Surgery*, 105.10 (2018), 1348 – 1349 (p. 1348).
[1034] Fan Li, Yiming Yang, and Eric P. Xing, 'From Lasso Regression to Feature Vector Machine', *Advances in Neural Information Processing Systems*, 18 (2005), 1 – 8 (p. 1).

to allow for the assessment of their variable effect within each of my continuous measures of error. I then provide the percentage point improvements to RMSE they yield relative to earlier linear models to allow for the comparison of effects between measures.

Table 45 displays the reduction in RMSE associated with the inclusion of election-level variables in random forest, fully interactive, and LASSO regression-based models alongside controls. Models are run using all available data, with the analysis of differences between subsets of data reserved for the following sub-section.

Table 45: The change in root mean square error (ΔRMSE) associated with including election-level variables both additively and interactively alongside controls within additional modelling approaches.

Model	MAE ΔRMSE	DIM ΔRMSE	LPB ΔRMSE	APB ΔRMSE	ABI 1 ΔRMSE	ABI 2 ΔRMSE
Fully interactive	-0.093	-0.165	-0.010	-0.010	-0.041	-0.059
Random forest	-0.116	-0.164	-0.010	-0.011	-0.045	-0.066
LASSO regression	-0.094	-0.163	-0.010	-0.009	-0.041	-0.059

MAE = mean absolute error, DIM = difference in margin, LPB = leading party bias, APB = average party bias, ABI 1 = first measure of average bounded inaccuracy, ABI 2 = second measure of average bounded inaccuracy

From Table 45, it is clear that including election-level variables in alternatively specified models leads to universal reductions in RMSE, even in the presence of controls. Their inclusion therefore improves the ability of models to accurately predict the variance exhibited by my continuous measures of polling error. This indicates that my earlier findings are substantively robust and not simply an artefact of linear modelling.

Unpacking the impact of the alternative modelling strategies within each measure reveals that random forest modelling generally produces the most pronounced reductions in RMSE. When considered relative to the fully interactive and LASSO models, random forests produce larger reductions in RMSE in the case of MAE, APB, ABI 1, and ABI 2. As such, they not only produce the largest reduction in RMSE for the majority of my continuous measures of polling error, but do so across measures of distributive and bounded error, as well as bias. The outputs displayed in Table 45 are derived from random forest models run with out-of-the-box hyperparameters. Therefore, hyperparameter tuning may improve performance further and exists as a promising avenue of future research.

While random forests generally produce the largest reductions in RMSE, fully interactive modelling outperforms other modelling strategies in the case of DIM. Though LASSO regression generally produces smaller improvements than random forest modelling, it nevertheless proves more useful than fully interactive modelling in the case of MAE. This suggests that the impact of alternative modelling strategies varies across measures of polling error. This variable performance indicates that the functions of the additional modelling techniques – non-linearity, complex interactivity, and overfitting – are differentially important across measures of polling error. This speaks to fundamental differences in the nature of the relationship between election-level variables and different forms of polling error which bears further exploration in future scholarship.

Table 46 presents the average percentage point reduction in RMSE associated with the inclusion of election-level variables in alternative models alongside controls relative to earlier linear models. This allows any improvements associated with additional modelling strategies to be assessed and facilitates cross-measure comparison due to the absence of scale dependency. It also enables the exploration of small-scale improvements between modelling approaches that could not be displayed to three significant figures in the previous table.

Table 46: The increased reduction in average RMSE associated with the inclusion of election-level variables in alternative models containing controls across measures of distributive error, bounded error, and polling bias measured relative to earlier linear models.

Measure	Fully Interactive Model (%)	Random Forest (%)	LASSO Regression (%)
MAE	1.63	27.54	2.80
DIM	1.80	1.46	0.30
ABI 1	0.34	8.78	1.07
ABI 2	1.20	13.73	1.99
LPB	1.16	2.33	0.39
APB	5.64	16.93	1.57

From Table 46 it is clear that, relative to linear models, alternative model specifications universally improve the degree to which election-level variables are predictive of variance across my continuous measures of polling error. In the case of fully interactive and random forest models, this indicates that the relationship between election-level variables and polling error is better represented by non-linearity and complex interactivity than it is linearity. In the case of LASSO regression, improved performance suggests that the optimal specification of election-level predictor variable is more parsimonious than the maximal specifications used in earlier modelling. Both sets of improvements call not only for greater exploration in the nature of the relationship between election-level variables and their effect on polling error, but also into the optimal configuration of predictor variables. Future research would do well to pursue both of these avenues.

The extent of the predictive improvements offered by alternative modelling approaches varies across measures of polling error. Fully interactive regression models produce the largest percentage point reduction in RMSE in relation to predictions of the average per-party bias displayed by polls (APB) and offers the lowest predictive improvement with regards to the first measure of bounded inaccuracy (ABI 1). This suggests that the extent of complex interactivity between election-level variables is greater in the case of polling bias and bears least closely on their propensity to breach the bounds set by the margin of error. More broadly, it indicates that the degree to which higher order interactions between election-level variables affect polling error varies across measures. Future scholarship ought to further explore the variable nature of complex interactivity between election-level variables to unpack both its nature and the impact it has on polling error.

Random forest models produce by far the largest improvements in predictive accuracy and do so over five of my six continuous measures of polling error. The magnitude of these improvements suggests that the connection between election-level variables and polling error is better represented as a non-linear relationship. However, while non-linear modelling universally improves upon multiple linear regression, it does not always outperform alternative methods, as indicated by the superior performance of fully interactive modelling in the case of DIM. Nevertheless, the improvements associated with random forest modelling are so pronounced that future research would be remiss not to employ them and fully explore the utility of non-linear modelling for the prediction of polling error variation using election-level variables.

The variable improvements associated with LASSO regression indicate that measures of polling error are differentially affected by overfitting, or at least variously benefit from parsimonious modelling. This suggests that the optimal configuration of election-level predictors varies across measures of polling error. While the findings from the linear models in my main analysis support this, as election-level variables were found to be variously useful across measures of error, future research ought to seek to identify the optimal configuration of election-level predictors for each measure to better understand their relationship with polling error.

While both random forest and fully interactive models can be used in relation to binary measures of polling error, lasso regression cannot as it stands as variant of linear regression and therefore centres on predicting a continuous outcome variable. To overcome this, I employ a generalised linear model than engages in stepwise feature selection on the basis of the Akaike information criterion associated with each predictor variable. As the Akaike information criterion measures the goodness-of-fit of a given model,[1035] its change in relation to the addition of each election-level predictor variable represents the degree to which it affects the ability of the model to accurately account for the dispersion of observed data points. The inclusion or exclusion of variables on the basis of the Akaike information criterion is therefore functionally equivalent to LASSO regression, insofar as it engages in feature selection motivated by model improvement.

Though they differ to earlier linear models, the additional models used to assess my binary measures of polling error nevertheless draw on the same process of downsampling derived from the sticker collector's problem outlined earlier in the thesis. In the case of models concerning SBP, 4,069 values are drawn from a population of 4,317 at random without replacement. Over 1000 Monte Carlo simulations, it required on average 3.6 draws for each discrete value in the majority class to be drawn at least once. As such, I round to the nearest appropriate integer and run models across 4 downsampled subsets of data. Running the same procedure for LVRC required 64 draws, necessitating 64 subsets.

Table 47 displays the degree to which including election-level predictor variables in the presence of controls, both additively and interactively, affects the ability of alternatively

[1035] S. A. Abu Baker and others, 'Modelling Loss Data Using Composite Models', *Insurance: Mathematics and Economics*, 61 (2015), 146 – 154 (p. 148).

specified models to correctly classify the presence of significantly biased polls (SBP) and polls that correctly predict the recipient of the largest vote share (LVRC).

Table 47: The improvement to correct classification rates that results from including election-level variables alongside controls in alterative modelling approaches.

Model	SBP Classification Improvement (%)	LVRC Classification Improvement (%)
Fully interactive model	12.93	8.29
Random forest	12.42	8.39
Stepwise GLM	11.67	8.90

The findings in Table 47 indicate that the inclusion of election-level variables in alternatively specified models universally improves their ability to correctly classify instances of polling error, even in the presence of controls, thereby better accounting for the variance they present. This further suggests that my earlier findings regarding their usefulness are substantively robust and not simply an artefact of generalised linear modelling or a given set of predictor variables.

The predictive benefit of alternative modelling strategies varies across my binary measures of polling error. In the case correctly classifying significantly biased polls (SBP), fully interactive regression modelling provides the largest predictive benefit. That encompassing complex interactivity again proves most beneficial for measures of polling bias suggests that higher order interactions between election-level variables bear more closely on it than other conceptualisations of polling error. Future research ought therefore to seek to identify these higher order interactions to unpack and better understand their relationship with the bias presented by pre-election polls.

By contrast, in the case of classifying instances in which polls correctly predict the recipient of the largest vote share in an election (LVRC), the use of stepwise generalised linear modelling with feature selection proved more beneficial than alternative modelling approaches. This indicates that models tasked with the prediction of LVRC benefit more greatly from parsimonious modelling approaches than from models permitting complex interactivity or non-linearity. It also suggests that the prediction of variance in LVRC is marginally more accurate under assumptions of linearity that models embracing non-linearity which is encouraging for the representativeness of the linear relationships proposed between substantive polling error and election-level variables earlier in this thesis.

To explore the degree to which alternative modelling techniques improve on multiple linear regression, Table 48 presents the change in correct classification rates associated with fully interactive, random forest, and stepwise GLMs relative to earlier linear regression models. Findings relate to the ability of models to correctly classify significantly biased polls (SBP) and polls that correctly predict the recipient of the largest vote share in a given election (LVRC).

Table 48: The change in correct classification rates that result from the inclusion of election-level variables alongside controls in alternative models relative to earlier, linear models.

Measure	Fully Interactive Model Correct Classification Change (%)	Random Forest Correct Classification Change (%)	Stepwise GLM Correct Classification Change (%)
SBP	2.80	2.29	1.54
LVRC	0.08	0.18	0.69

From Table 48 it is immediately apparent that, while all additional modelling approaches yield improvements in predictive accuracy over earlier linear regression models, these improvements are often quite small, especially in the case of LVRC. Despite this, that election-level variables improve the ability of alternative models to correctly classify instances of binary polling error indicates that my earlier findings are substantively robust, as improvements are not simply an artefact of multiple linear regression.

Of the improvements to predictions of SBP, fully interactive modelling proves more useful than other approaches. This speaks to earlier improvements in the prediction of polling bias through models permitting interaction between election-level variables. These findings further suggest that future research ought to investigate the connection between polling bias and higher order interactions between election-level variables to better understand its nature and predictive utility.

In the case of LVRC, the use of stepwise generalised linear models yields a larger improvement to correct classification rates than other approaches, albeit minor. This suggests that, of the alternative approaches, classifications of LVRC benefit more acutely from stepwise feature selection. This underscores the exploratory findings presented earlier in the thesis identifying the impact of overfitting on predictions in highly specified models (such as the control models used here). Overall, the minor improvements to the accuracy of predictions of LVRC offered by alternative modelling techniques is understandable in light of the strong performance of earlier linear models. In all cases, four-level models based on multiple linear regression present correct classification rates of between 86% and 93%, leaving little room for predictive improvement. For a full account of the accuracy of these models, see: Appendix B, Tables B81 through B84.

Ultimately, that alternative modelling strategies were unable to produce substantial improvements to predictive accuracy over earlier linear models speaks to the applicability of multiple linear regression to the relationship between binary measures of polling error and election-level variables. Nevertheless, that alternative modelling approaches do bring about small improvements over linear modelling suggests that future scholarship ought to engage further with approaches embracing non-linearity, stepwise feature selection, and complex interactivity when exploring the relationship between binary polling error and election-level variables.

In addition to the linear assumptions underpinning earlier models and the suspected instances of overfitting they present, another trend is evident in my analysis: findings vary on the basis of the subset of data used to calculate them. In the following sub-section, I propose a rationale for this variability and demonstrate that my substantive findings are robust across a range of differently composed subsets of data.

Exploring the Effect of Data Composition on Predictive Performance

It is clear from the results presented within this chapter and throughout Appendix B that the impact of election-level variables on the accuracy with which models are able to predict variance in polling error differs on the basis of the subsets of data they draw upon. The subsets of data used for modelling the predictive utility of election-level variables in this thesis vary according to the presence of measurements of specified electoral characteristics. As measurements of certain electoral characteristics – most notably partisanship, late decision-making, and the left/right ideological deviation of parties and candidates – are available for certain election but not for others, these subsets necessarily vary in terms of the elections they comprise.

The multi-level nature of polling error can be used as a theoretical framework with which to group elections and test the effect of variably composed subsets of data on the predictive utility of election-level variables. In the multi-level structure of sources of polling error, elections occur within individual countries and are therefore nested within them. This nesting arrangement is emphasised in Figure 30. As elections are nested within individual countries, different countries necessarily contain differing subsets of elections. Data can therefore be broken down on a country-level basis to provide differently composed subsets of elections.

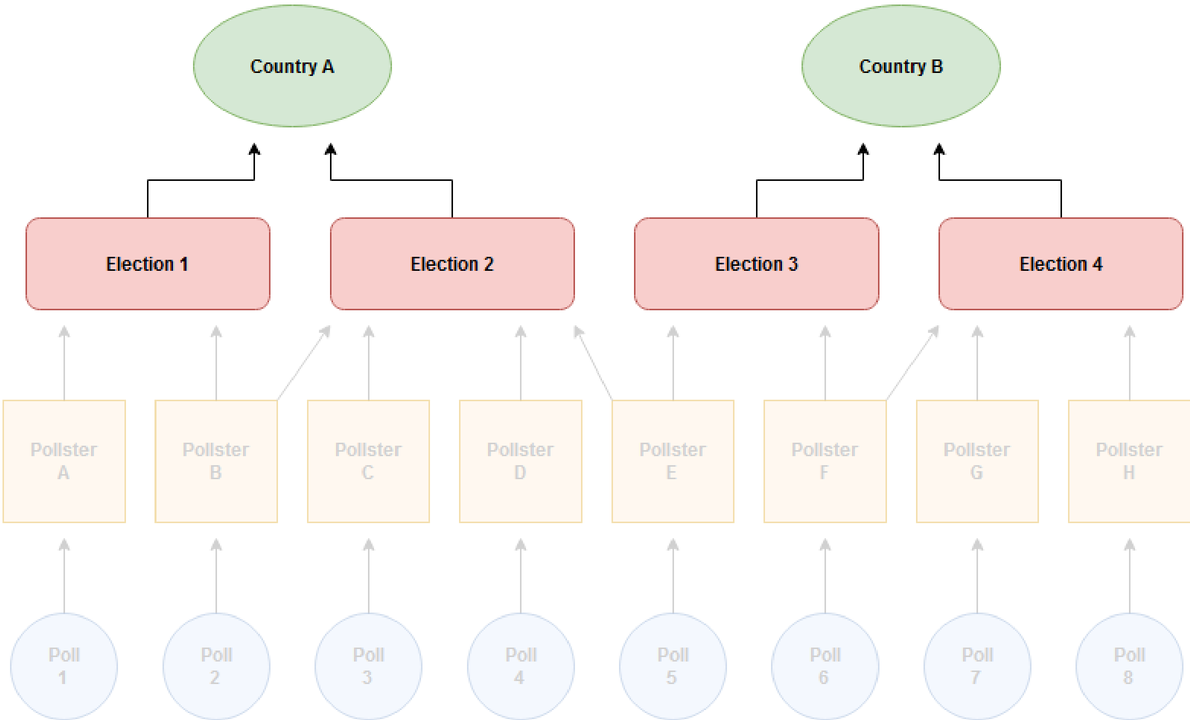

Figure 30: The multi-level structure of sources of polling error, emphasising the nested relationship between the election and country grouping levels.

The creation of subsets of elections on the basis of the countries in which they occur is not only motivated by the multi-level nature of polling error, but also by the variation of election-level ICC findings between countries displayed in Section A5 of Appendix A. That the variance in polling error accounted for by election-level differences varies internationally speaks to their varying predictive utility across countries. As such, not only do subsets of elections based on the countries in which they occur offer a platform on which to investigate the variation of findings between differently composed subsets, but differences in the impact of election-level variables between countries offer a plausible explanation for it.

To ensure that country-based subsets of elections are of sufficient size to accommodate the training and test splits inherent within repeated *k*-fold cross validation, I divide elections across characteristically similar clusters of countries. Principally, I divide elections into subsets on the basis of the continent in which they occur to capture macroscopic similarities between geographically proximal units. However, I also partition elections into country-based subsets centred on poll-related commonalities that can be expected to bear upon error variance.

Given the need for a sufficient number of datapoints to avoid issues of rank deficiency in prediction models, I am only able to separate out elections that occur in Europe, Asia, and the Americas. Collectively, my dataset contains too few polls relating to Africa and Oceania for elections occurring in these continents to addressed as discrete sets. In addition to these continent-based groupings, I gather together elections that take place in countries that previously existed as either members of the Soviet Union or broader Warsaw Pact. I do so to capture relatively young democracies within my dataset where pre-election polling – itself contingent on democratic elections – can be expected to have a shorter history and, as such, be less developed. This relative lack of development lends itself to the expectation of heightened error variance between cases, as polls have had less time, and fewer elections, over which to calibrate their accuracy.

I calculate the average predictive improvement associated with the inclusion of election-level variables, both additively and interactively, across each subset of elections in the presence of controls. For this to be tractable across newly composed, country-based subsets of elections, concessions must be made, as certain election-level variables are not available for all elections in all countries. Due to this, a selection of variables that is present across all cases must be used. To achieve this, I use those election-level variables available across my full dataset and, therefore, across all countries.

Table 49 displays the average improvement in predictive accuracy associated with the inclusion of the selected set of election-level variables, both additively and interactively, in models tasked with the prediction of continuous measures of polling error calculated across varying subsets of elections. Improvements are displayed in the table as raw changes in RMSE. This decision is motivated by the fact that differences in the importance of election-level variables as predictors of polling error variance between subsets occurred across models concerned with the same measures of inaccuracy. As such, analysis concerns the comparison of variation *within* measures of polling error, rather than between them, permitting the unaltered use of RMSE. Values are derived from models using multiple linear regression to allow the earlier variability between subsets associated with these models to be directly unpacked.

Table 49: The change in average root mean square error (ΔRMSE) compared to a null model associated with the inclusion of election-level variables and interactions in models tasked with the prediction of continuous measures of polling error across a range of subsets of elections. All findings are averages drawn from out-of-sample predictions across repeated 10-fold cross validation using fully specified linear regression models

Subset Composition	MAE ΔRMSE	DIM ΔRMSE	LPB ΔRMSE	APB ΔRMSE	ABI 1 ΔRMSE	ABI 2 ΔRMSE
Asian states†	-0.964	-0.107	-0.003	-0.017	-0.888	-0.887
European states	-0.113	-0.250	-0.025	-0.020	-0.045	-0.051
States in the Americas†	-0.159	-0.879	-0.018	-0.025	-0.276	-0.159
Ex-Warsaw Pact and Soviet states†	-0.618	-0.868	-0.011	-0.029	-0.291	-0.447

† No instances of system change exist between the elections conducted in the Americas within my dataset, no round two presidential elections occurred within Asia, and no instances of system change occurred across my set of elections in ex-Warsaw Pact and Soviet states. As such, these variables were removed in models addressing these subsets to avoid issues of rank deficiency.

From Table 49, it is clear that the inclusion of election-level variables universally reduces the average RMSE exhibited by models focusing on each continuous measure of polling error. However, the degree to which they do so varies across differently composed subsets of data. This lends suggestive evidence to the contention that the variation in the predictive utility of election-level variables across subsets observed in earlier analysis exists as an artefact of the countries they comprise. This speaks to the variable importance of election-level characteristics as predictors of polling error variance across countries. This is corroborated by wider research,[1036] as well as the ICC findings presented in Section A5 of Appendix A.

The findings presented in Table 49 allow the differing utility of election-level variables across country-level groupings to be unpacked. Across three of the six measures of polling error addressed, election-level variables are most useful as predictors of variance in Asian states. While this finding may be driven by the high levels of error presented by polls conducted for Indonesian elections within my dataset (visualised in the world map presented in chapter 4), it nevertheless suggests that the error exhibited by polls conducted in Asia varies more considerably on the basis of differences between elections than polls conducted for countries in other continents.

By comparison, election-level differences are generally the least useful predictors of polling error variance in European states. This suggests that election-level variables, and therefore differences between elections, bear less closely on the error exhibited by polls in European states than those conducted in other continents. As most European countries encompass parliamentary systems and employ variants of proportional representation, this may suggest that these two variables combine to improve polling error and improve accuracy relative to other contexts.

The reduced impact of election-level differences on polling error in European states may further indicate that parliamentary systems and proportional representation interact with the degree to which democracies are established to allow polling organisations to attain higher levels of accuracy across elections with greater ease, as democratic regimes are generally, though not universally, more well-established in Europe than the other continents or sets of

[1036] Tudor and Wall, p. 15.

elections addressed. Ultimately, further study is needed to unpack the factors that drive the diminished impact of election-level variables on polling error variance across European states relative to those in other continents.

The findings in Table 49 also allow the potential existence of a Red Queen problem in polling – that is, the need to work harder over time to maintain stability in polling error – to be explored. As identified earlier in the thesis, it may be that the stasis seen in polling error since 1936 is the result of polling occurring in an ever-increasing number of countries over time. The difficulty of conducting polls in these emerging contexts is suggested by the increased variance in polling error over time which surfaces in line with their emergence. Further to this, it may be that polling error has fared better over time in areas with a longer polling history, such as Europe, but that this has been occluded by an ever-growing global dataset.

The RMSE values presented across differently composed subsets of countries provide a degree of support for both the Red Queen problem and the potential for a different evolution in polling error over time within Europe. Election-level variables account for greater reductions in RMSE across elections conducted in Asia and the Americas, as well as in ex-Warsaw Pact and Soviet states, than those conducted in Europe. This suggests that differences between elections bear more closely on the error exhibited by polls conducted in these states than those conducted in Europe. This lends itself to the expectation of higher cross-election volatility in terms of polling error in these cases. As polling in states in Asia and the Americas, as well as ex-Warsaw Pact and Soviet states often, though not universally, came about after states in Europe, it may be that this cross-case volatility accounts for the increase in polling error variance observed over time. If this is the case, then as polling has emerged in these states, it has become more difficult for polling organisations globally to converge on high levels of predictive accuracy over time. That the average error presented by polls has remained largely static despite this increasing difficulty would therefore indicate the existence of a Red Queen problem, as mitigating this difficulty and maintaining consistent performance would necessarily entail great effort on the part of pollsters, with the effort necessary to maintain performance levels increasing in tandem with the emergence of difficult cases.

By contrast, the findings pertaining to Europe suggest diminished cross-election volatility in polling error, as the inclusion of election-level differences bears less closely on polling error variance than other subsets. Reduced cross-election volatility of this kind indicates that convergence on a higher level of accuracy over time ought to be easier in these cases than those characterised by more considerable cross-case variability. However, if this is the case, it has likely been occluded in my global dataset by the high variance and volatility introduced by emerging states over time. Future longitudinal study is therefore needed to understand and unpack the region-specific evolution of polling error over time and its impact on aggregate trends.

Table 50: The improvement in correct classification rates compared to a null model associated with the inclusion of election-level variables and interactions in models tasked with the prediction of binary measures of polling error across a range of subsets of elections. All findings represent averages across out-of-sample predictions from repeated 10-fold cross validation across a varying number of balanced datasets.

Subset Composition	Average SBP Classification Improvement (%)	Average LVRC Classification Improvement (%)
Asian states†	20.72%	32.80%
European states	13.04%	19.28%
States in the Americas†	16.96%	33.93%
Ex-Warsaw Pact and Soviet states†	21.36%	25.81%

† No instances of system change exist between the elections conducted in the Americas within my dataset and no round two presidential elections occurred within Asia. As such, these variables were removed in models addressing these subsets to avoid issues of rank deficiency.

To establish the impact of different subsets of countries on findings related to my binary measures of polling error, Table 50 presents the average change in the correct classification rate of models associated with the inclusion of election-level variables across different country-based subsets. Findings relate to the ability for models to correct classify significantly biased polls (SBP) and polls that correctly identify the recipient of the largest vote share in a given election (LVRC).

The findings in Table 50 demonstrate that election-level variables consistently improve the ability of models to correctly classify polling error, and therefore better predict its variance, across different country-based subsets. However, the degree to which they do so varies. This further underlines the variable utility of election-level variables as predictors of polling error across differing subsets of countries. Despite this variability, election-level variables nevertheless remain useful predictors of polling error across all subsets of countries.

In terms of significant polling bias (SBP), election-level variables most acutely benefit correct classification rates in ex-Warsaw Pact and Soviet states. By contrast, they are least useful as predictors of SBP in European states. That these findings differ from earlier, continuous measures of polling bias suggests that the usefulness of election-level variables as predictors of polling error not only differs between countries, but also varies across operationalisations of the same forms of polling error.

Election-level variables are most useful for the correct classification of polls that correctly predict the largest vote share recipient (LVRC) in states in the Americas and least useful in the case of European states. That election-level variables again bear upon polling error in European states to the lowest extent – and have done across all conceptualisations of polling error – further underscores the diminished level of cross-election volatility in these cases and suggests that it is robust across all forms of error.

Further (sub-)regional decomposition of the importance of election-level variables as predictors of polling error variance would be interesting, but results in too few datapoints per subset to avoid issues of rank deficiency. For example, given that it principally encompasses two polled states, the Oceanian region provided insufficient data to use in k-fold cross validation across highly specified models. The small number of countries also leads to lack of variation in certain indicators, leading to further rank deficiency due to presence of constants. As more polling data becomes available in the future, scholars ought to further unpack the country-based variability of the importance of election-level characteristics as predictors of polling error variance.

In the following sub-section, I summarise how the findings that result from the exploration of additional model specifications bears upon my main analyses. I also unpack what these findings tell us about the relationship between election-level variables and polling error, and the avenues for future research they present.

What Does the Exploration of Additional Models and Emergent Themes Tell Us?

That election-level variables remain useful predictors of polling error variance in alternatively specified models across all measures of polling error indicates that the substantive findings of these thesis are robust and not simply an artefact of a given set of modelling choices. Differences between elections consistently improve the ability of models to predict variance in polling error, even accounting for control variables from alternative grouping levels and differently composed subsets of data.

While the results presented by alternative modelling strategies are encouraging in their support for the main findings of this thesis, they cast further light on the nature of the relationship between election-level variables and polling error. Though the linear assumptions underpinning models in the main analysis of this thesis proved predictively useful, the improvements offered by models permitting non-linearity and complex interactivity suggest that the relationship between election-level variables and polling error may be more intricate and interconnected than straightforward linearity. Future scholarship would do well to further investigate the nature of this relationship and identify mechanisms that lend themselves to non-linear and highly interactive understandings.

The results from LASSO regressions and stepwise GLMs indicate that the predictive accuracy of models generally benefits from feature selection. This lends support to the notion that the observed degradation in highly specified models within Appendix B is overfitting and not, necessarily, a lack of importance of the part of election-level interactions added later in the modelling process. However, it may simply be that these interactions are not useful predictors of polling error variance. Further work is needed to establish their importance using parsimonious models and ought to be a priority for subsequent research.

Unpacking the impact of election-level variables differs across differently composed subsets of data indicates that the importance of election-level differences as drivers of polling error variance varies between countries. Though wider research has identified this trend,[1037] and the effect of a range of country-level differences is captured within my four-level models, future research ought nevertheless to decompose why this variability occurs. Within this thesis, I have proposed several theoretical mechanisms that would serve as a grounding for such an investigation and offer a foundation for future work to build on.

Overall, the findings presented by this thesis present a great many avenues for future research and stand to alter the way in which polling error variance is studied. In the following chapter, I conclude by relating the results of this thesis back to my three research questions, unpacking the practical utility of adopting an election-level understanding of polling error variance, and more fully exploring the directions available for future research.

[1037] Tudor and Wall, p. 15.

Chapter 7 – A Foundation to Build Upon: Towards Ontological and Epistemological Re-orientation in the Study of Pre-election Polling Error

"[When] the profession can no longer evade anomalies that subvert the existing tradition ... then begin the [analyses] that lead the profession at last to a new set of commitments, a new basis for the practice of science".[1038]

- Thomas Kuhn (1962)

In assessing the actions that underpin good prediction, Tetlock holds that one must not only get it right but also think in the right way.[1039] As this thesis has made clear, pollsters do not always get it right and, due to the lack of improvement in polling error over the past nine decades, do not necessarily think the right way when assessing instances of misprediction. In this thesis, I have proposed and tested a new approach to understanding polling error that recognises the importance of electoral heterogeneity. I have identified election-level differences as theoretically plausible drivers of polling inaccuracy and situated them within a novel four-level model of sources of polling error. Through the multi-level decomposition of variance terms within this model, I have demonstrated that differences between elections account for a substantial portion of past polling error variance. By analysing the out-of-sample performance of an exhaustive range of prediction models, I have also robustly shown that the inclusion of election-level variables allows models to more accurately predict polling error variance, even in the presence of controls.

In this concluding chapter, I summarise the findings of this thesis, unpack their importance and practical utility, and identify avenues for future research. I begin by relating the findings of this thesis back to its stated research questions, identifying how each has been answered in turn.

[1038] Kuhn, p. 5.

[1039] Philip E. Tetlock, *Expert Political Judgement: How Good Is It? How Can We Know?*, (Princeton: Princeton University Press, 2005), pp. 10 – 19.

Empirical Findings and Answers to Research Questions

In this thesis, I set out to answer three research questions concerning the drivers of variance in pre-election polling error. By identifying that the dominant poll-level understanding of polling error was incommensurate with its evolution over time, I held that sources of error beyond the mechanisms inherent within pre-election polls must necessarily bear on their accuracy. I contended that differences between the elections that polls seek to predict stand as plausible drivers of the error they present. To investigate this contention, I established my first research question:

> $\mathbf{RQ_1}$**:** To what degree can variance in polling error be expected to be a function of differences between elections?

To answer this question, I developed a novel conceptual framework through which to view elections as intuitively heterogenous phenomena. Within this framework, I held elections to possess a core of common characteristics that vary in magnitude between cases. To establish the generalisable expectation of variance between cases, I demonstrated that the likelihood of two elections possessing constellations of characteristics with identical magnitudes is vanishingly slim. I identified that the differing magnitude of characteristics between elections affects their predictability as phenomena by altering the degree to which they can be considered clock- or cloud-like on the Popperian predictive continuum. Through a discussion of the predictability of clock- and cloud-like phenomena, I established the potential for differences in the magnitude of characteristics held by elections to affect the degree to which they lend themselves to accurate prediction and, therefore, the degree of polling error likely to be associated with them.

I also identified differences between elections as likely drivers of polling error by unpacking their effect on the projection mechanisms undergirding poll-based predictions. I recognised that the predicted vote share distributions provided by polls rest on three principal projections: the sample to population projection, the vote likelihood projection, and the projection of survey responses onto future voting behaviour. I held that the differences between elections that form the core of their heterogeneity as phenomena affect the degree to which these projection mechanisms can be relied upon. In so doing, I posited that differences between elections variously lend themselves to polling error.

To make real the potential for differences between elections to bear upon polling error, I provided a series of real-world examples. By examining the characteristics possessed by past elections that were the focus of poll-based mispredictions, I offered a series of election-level explanations for the error exhibited by polls. That past instances of polling error lend themselves to being understood through the lens of election-level differences afforded substantive plausibility to the contention that differences between elections have the potential to serve as drivers of polling error.

By establishing the heterogeneous nature of elections of phenomena, the impact of this heterogeneity on the predictability of elections as phenomena and the reliability of the projection mechanisms central to polling, as well as providing real-world examples of the likely impact of election-level differences, I demonstrated that polling error can defensibly and intuitively be understood as a function of differences between elections. In so doing, I answered my first research question and devised my first testable hypothesis:

> *H1: Membership within different elections will affect the degree to which polls exhibit error*

While polling error could be expected to vary as a function of election-level differences a priori, the empirical validity of this theory needed to be established in relation to real-world data. To this end, the creation of my first hypothesis provided a testable foundation on which to interrogate my second research question:

> $\mathbf{RQ_2}$**:** To what extent can the expectation that variance in polling error exists as a function of differences between electoral characteristics be validated empirically?

To answer this question, I devised a novel four-level model through which to understand sources of polling error. I posited that sources of polling error exist in four distinct, interconnected grouping levels: the poll, pollster, election, and country levels. By decomposing previous frameworks for understanding sources of polling error, I held that my four-level approach better represents its reality than approaches comprising fewer, less intricately connected dimensions. Through this, I demonstrated that in order to accurately assess the impact of election-level differences on polling error, it is necessary to control for influential variables housed within the poll, pollster, and country grouping levels.

Given the four-level nature of sources of polling error, I identified multi-level variance decomposition as the most appropriate approach to establishing the degree to which polling error varies as a function of election-level differences. To achieve this, I estimated the intra-class correlation coefficient (ICC) associated with the election-level grouping to decompose the proportion of polling error variance attributable to differences between elections. I decomposed polling error variance across the most expansive polling dataset gathered to date, capturing 11,832 in-campaign polls conducted in 497 elections across 83 countries.

Through estimating election-level ICC values across two-, three-, and four-level multi-level models, I demonstrated that differences between elections account for a substantial proportion of variance in distributive, bounded, and substantive polling error, even in the presence of controls. To ensure the robustness of these findings, I provided a series of additional variance estimates derived from alternative approaches to variance decomposition and models using differing prior specifications and methods of parameter estimation that agreed substantively with my main findings.

By decomposing the proportion of variance in polling error associated with differences between elections across my dataset, I found strong support for my first hypothesis, concluding that membership within different elections serves a prominent driver of the error exhibited by polls. Through doing so, I provided a clear answer to my second research question, establishing that the theoretical expectations concerning the impact of election-level differences on polling error could be verified empirically.

While multi-level variance decomposition was capable of identifying that differences between elections serve as drivers of polling error variance, it was incapable of identifying those specific differences that mattered. To identify these differences, I turned to prediction. As polling error varies as a function of differences between elections, I contended that these differences ought to be predictive of variance. This contention led to my third research question:

> $\mathbf{RQ_3}$**:** To what degree can differences in electoral characteristics aid in the prediction of polling error variance?

To answer this question, I identified a series of differences between elections that could plausibly serve as predictors of polling error variance and unpacked the mechanisms through which they could be expected to do so. I established that election-level variables are likely to be predictive of polling error variance both individually and in interaction with one another. Through unpacking the expected impact of election-level variables on my measures of polling error both individually and interactively, I devised my second testable hypothesis as a vehicle for answering $\mathrm{RQ_3}$:

H2: Election-level variables will aid models in predicting polling error variance both additively and interactively, proving most useful in the case of substantive error and least useful in the case of bias.

To test this hypothesis, I employed two forms of prediction model: additive and interactive. Additive prediction models incorporated election-level variables in a stepwise manner to assess their individual impact on predictive accuracy. By contrast, interactive prediction models introduced two- and three-way interactions between election-level variables to assess their impact in concert with one another. These models were based on linear assumptions, motivated by the expectation of linear relationships between election-level variables and measures of polling error.

By measuring the degree to which the inclusion of election-level variables changed the root mean square error (RMSE) and correct classification rate of prediction models, I was able to establish the extent to which they were able to accurately predict polling error and, therefore, allow models to more accurately account for its variance. To ensure that my findings were robust and not simply an artefact of a given training and test split, I employed repeated 10-fold cross validation. The findings that resulted from this process provided strong substantive support for my second hypothesis, as election-level variables were found to consistently aid models in predicting polling error variance both individually and in interaction with one another. Election-level variables remained predictively useful in models containing controls, indicating that this substantive finding is robust to their presence.

While I found strong support for the substantive claim put forward in my second hypothesis – that election-level variables would aid models in predicting polling error – I found less consistent support for the specific contention that they would prove most useful in predictions of substantive polling error and least useful in predictions of bias. In additive models, the inclusion of election-level variables was found to yield the largest improvement to predictions of substantive polling error. However, while election-level variables were found to be least predictive of leading party bias, their performance was notably improved in relation to both average party bias and significant polling bias. Relatedly, while the two- and three-way interactions between election-level variables housed within interactive models were found to be least predictive of significant polling bias, they were not found to provide the greatest benefit to predictions of substantive polling error. Instead, election-level interactions proved most useful as predictors of distributive polling error.

Including controls within prediction models altered the degree to which election-level variables were considered predictively useful. In the presence of controls, the inclusion of individual election-level variables was found to be most useful in the prediction of measures of distributive polling error and least useful in the prediction of substantive polling error. These findings stand in opposition to the theoretical expectations established earlier in the thesis and, as such, warrant further investigation in subsequent research. Within control models, the usefulness of election-level interactions also subverts expectations, as they prove most useful in the prediction of distributive polling error and least useful in the prediction of substantive inaccuracy. Nevertheless, all model outputs supported the core hypothesis that the inclusion of election-level variables would improve the ability of models to predict polling error variance.

To ensure that the usefulness of election-level variables as predictors of polling error was not an artefact of linear modelling choices and specified interactions, I employed a range of additional prediction models that allowed for non-linearity and complex, higher order interactivity. Across these models, election-level variables were found to consistently improve the accuracy of predictions across all measures of polling error. This indicates that the substantive findings of this thesis are robust to alternative modelling specifications.

Through these analyses, I answered my third research question by demonstrating that election variables consistently aid models in the prediction of polling error when included individually and often yield smaller improvements when considered interactively. These findings are not only robust to the presence of controls, but are also robust to differing model specifications.

When its findings are considered collectively, this thesis has demonstrated that differences between elections exist as theoretically motivated and empirically impactful drivers of polling error variance. It has also shown that a range of election-level differences serve to improve the ability of models to predict polling error variance, even in the presence of controls. In the following sub-section, I establish why these findings are important and how they can be employed practically.

The Theoretical and Practical Importance of This Research

The importance of the work presented in this thesis is both theoretical and practical. Its theoretical importance is rooted in the puzzle at the heart of pre-election polling elaborated at its beginning. As average polling error has not meaningfully reduced over the past nine decades despite a continual process of methodological revision at the poll level, it is clear that the dominant poll-level approach to understanding polling error is insufficient to fully capture the factors that drive it. To remedy this, this thesis provides a novel four-level framework to better understand the nature of drivers of polling error.

The four-level model put forward in this thesis provides an expanded theoretical framework for conceiving of factors that bear upon polling error. When compared to the dominant poll-level understanding of polling error and existing multi-level approaches to its decomposition, the four-level model better represents the reality of sources of polling error and allows for an improved understanding of their nature and interconnection. Through the empirical validity afforded to it by control models, the four-level models put forward in this thesis also makes clear the need to approach polling error as a multi-level problem, rather than a unidimensional issue centred on polling methodology.

Of the four levels within its multi-level model of polling error, this thesis emphasises the importance of the election level. In so doing, it provides the first conceptualisation of elections as heterogenous phenomena that can be expected to bear upon polling error. By robustly demonstrating that the differences between elections serve as drivers of polling error variance, it provides empirical validation to this theoretical expectation and illuminates the need to account for election-level differences in assessments of polling error.

By framing polling error as a four-level phenomenon and demonstrating both the theoretical and empirical importance of electoral heterogeneity as a driver of its variance, this thesis makes clear the need for ontological and epistemological re-orientation in the study of polling error towards an understanding that not only embraces electoral heterogeneity, but also accepts the fundamentally multi-level nature of polling error.

In addition to its theoretical importance, the work contained within this thesis is practically important for two key reasons. The first concerns the information contained within its findings. The results provided by the prediction models run within this thesis allow for a better understanding of those circumstances in which polling error is likely to vary. Specifically, the outputs of prediction models identify those election-level differences that bear most closely on the observed variance of polling error and, therefore, isolate those differences between elections that are likely to cause polling error to vary. This not only allows for a deeper understanding of why polling error was found to vary between past elections, but also lends itself to a degree of forewarning as to whether the error exhibited by polls in a given election will vary relative to the previous contest on the basis of differences between them as phenomena. However, it must be noted that the nature of the findings in this thesis is such that they are incapable of indicating the direction that error will vary in relative to the previous contest, merely that it is more likely to vary in the presence of certain differences than others.

The second element of the practical importance of the work contained within this thesis concerns the manner in which it can be used moving forward. As its findings identify election-level differences that affect the degree to which polling error varies, this thesis necessarily isolates electoral factors that bear upon the accuracy of the predicted vote share distributions offered by polls. These factors could therefore be used to inform these predictions. This is afforded plausibility by the impact of ex-ante election-level variables than can be known in advance of election day on polling error variance. However, it is important to note that the findings in this thesis do not improve the accuracy of poll-based predictions in and of themselves, nor are they able to do so without additional work. Specifically, the directionality of the variance in polling error brought about by differences between elections must necessarily be established before they can be used to meaningfully inform or improve poll-based predictions.

The need to establish the direction in which polling error varies in the presence of election-level differences stands as a clear avenue for future research that arises from the work in this thesis. In the following sub-section, I elaborate on the need for this research and identify further avenues of enquiry that future research ought to pursue on the basis of this thesis.

Avenues for Future Research

As this thesis demonstrates the theoretical and empirical impact of election-level differences as drivers of polling error variance, future studies of polling error ought therefore to account for differences between elections when assessing it. Given the generalisable importance of election-level differences across countries, the inclusion of differences between elections in the study of polling error variance ought to become universal and be adopted irrespective of the cases addressed by future research.

Though the election-level differences addressed in this thesis represent those differences that most intuitively bear upon polling error variance, they do not represent the universe of all possible differences between elections that may affect the degree to which polling error varies. This is made clear by the disconnect between the total variance in polling error attributable to differences between elections as established by multi-level variance decomposition and the improvements in the prediction of polling error variance that result from my studied election-level variables. Future research would do well to identify further election-level variables – and interactions between these variables – that bear upon polling error variance to better understand the extent and importance of election-level differences as drivers of polling inaccuracy.

Due to its focus on intra-class correlation coefficients and changes in RMSE between prediction models, this thesis does not turn its attention to the regression coefficients that the regression-based models from which these values are derived have the potential to provide. This speaks to its most prominent limitation: that it does not unpack the extent and direction of the variance in polling error brought about by election-level differences that it identifies. Knowledge of the direction in which polling error varies as a result of differences between elections and the extent to which it does so would not only allow for a deeper understanding of the impact of election-level differences, but would also allow those differences that can be known ahead of election day to inform vote share predictions. Accounting for the impact of election-level differences ahead of time may aid in reducing polling error, or better understanding the likelihood of poll-based predictions proving erroneous. Future research should therefore strive to establish the extent and direction of the variance in polling error attributable to election-level differences.

Moving forward, research ought to adopt a multi-level approach to understanding and decomposing polling error, appropriating and building upon the four-level model proposed within this thesis. Not only does the four-level model offer the most comprehensive account of sources of polling error to date, but it does so in a manner that is more representative of the reality of pre-election polling error than existing approaches. However, as this thesis principally focuses on the election-level, work remains to be done unpacking and exploring the nature and effect of the remaining three levels. This is especially true of the country level grouping of variables within the model. While the variables contained within the poll and pollster levels have been the subject of a great deal of academic literature, as made clear within Chapter 2, country-level differences remain comparatively underexplored. As they serve as important control variables and directly affect the degree to which election-level differences bear upon polling error variance, as demonstrated in Section 6.3, future work ought to further unpack the nature and effect of country-level differences on polling error to better understand their importance.

When building on the multi-level approach to understanding polling inaccuracy presented in this thesis, future work would do well to incorporate the impact of time, especially when further unpacking the impact of pollster-level differences on polling error. While the four-level model presented within this thesis captures differences between polling organisations, it does so somewhat homogenously. As each polling organisation is identified using the same category across all elections for which it conducts polls, the current multi-level structure fails to account for changes in the methodologies adopted by pollsters over time, as they are presently homogenised. That is, they are treated as capturing the same constellation of factors from one election to the next. Due to this, any variance in polling error that arises from changes in the methodologies employed by pollsters over time may be mistakenly attributed to other factors or grouping levels within the current model. Therefore, it may be that pollster-level differences stand as more impactful drivers of polling error than is recognised within the analysis presented by this thesis. Future work ought therefore to attempt to incorporate time into the manner in which polling organisations are included within subsequent multi-level models.

The findings presented in this thesis indicate that observed variance in polling bias is more closely affected by election-level differences than would be expected a priori. As polling bias is traditionally thought of as a pollster-level phenomenon born of partisan biases or over-correction on the basis of past mistakes,[1040] the fact that election-level differences bear closely upon it – often more closely than other conceptualisations of polling inaccuracy – runs counter to expectations. Future scholarship ought to establish why this is the case and further explore the relationship between variance in the bias exhibited by polls and differences between the elections they attempt to predict.

Though the linear assumptions underpinning the main prediction models in this thesis proved predictively useful, modelling approaches permitting non-linearity and complex interactivity were better able to account for observed variance in polling error, resulting in more accurate predictions. Future research ought to therefore move beyond linear modelling towards further exploration of non-linearity and higher order interactions to better understand the relationship between polling error variance and differences between elections that better captures their reality.

In a similar vein, the findings that result from LASSO regression and stepwise GLMs indicate that understanding variance in polling error as a function of election-level differences may benefit from more parsimonious modelling than that contained with the highly specified models presented in this thesis. As such, future research should seek to establish the optimal model specifications for accounting for polling error variance and, in so doing, better identify

[1040] Pickup and Johnston, pp. 272 – 284; Bergman and Holmquist, p. 307; Jackman, p. 500.

those election-level differences that provide meaningful improvements to the predictive accuracy of models. Though LASSO models and stepwise GLMs may lend themselves to the atheoretical identification of optimal model specifications through the iterative assessment of model fit statistics, it would be more beneficial for future work to further develop and test theories regarding the relationship between election-level variables and measures of polling error to identify those that provide the greatest predictive utility.

As the availability of global pre-election polling data grows in both breadth and depth, future work will be burdened to a lesser extent by rank deficiency and able to engage in more expansive models that include a greater range of variables housed within the four grouping levels of sources of polling error. Though the growth of polling data will necessarily be a slow process, once its scope increases sufficiently, future works will be better able to investigate the impact of those variables at the country and election levels that present data limitations prevented this thesis from unpacking without transformation.

Much research into the relationship between polling error variance and differences between elections therefore remains to be done. However, through its demonstration of the theoretical and empirical benefits of ontological and epistemological re-orientation in the study of pre-election polling error towards a recognition of the importance of electoral heterogeneity, this thesis provides the foundation on which this research can build.

Appendix A: Additional Calculative Techniques and Robustness Checks to Support the Inferential Analysis in Chapter 4

This appendix supplements the inferential analysis in the fourth chapter of my thesis. It is split into five sections, A1 through A5. Section A1 contains the additional approaches to variance decomposition for two-level models used within later robustness checks and describes the manner in which they are calculated. These additional approaches are ICC1, eta-squared, and omega-squared. Section A2 contains robustness checks for the ICC estimates provided for both my continuous and binary measures of polling inaccuracy across two-level models. Section A3 contains similar robustness checks for my three-level models. In addition to these checks, the section contains the country-level ICC values for each of my measures of polling inaccuracy to explore the impact of differences between countries. Section A4 provides robustness checks for my four-level models. These checks are accompanied by both the country- and pollster-level ICC values to further explore the impact of differences between these grouping levels within my fully specified model. Finally, Section A5 decomposes election-level ICC values across individual countries to illustrate the variable importance of election-level differences as drivers of polling error variance between countries. These ICC values are estimated via Bayesian MCMC using preferred half student-t priors and control for the impact of pollster-level differences between countries.

A1: Alternative Approaches to Partitioning Group-level Variance

While the ICC equations outlined within the main body of the thesis are the most commonly used approaches to partitioning group-level variance in two-level models, alternative approaches exist which I use within this appendix as robustness checks. ICC1 is calculated using the variance components of a one-way ANOVA as shown in equation A1, where $MS_{between}$ is the mean square value between the groupings of interest, MS_{within} is the total mean square value between all levels in the model, and *k* represents the number of groupings of interest.[1041]

$$ICC1 = \frac{MS_{between} - MS_{within}}{MS_{between} + [(k-1) * MS_{within}]} \tag{A1}$$

To ensure that its findings are robust, ICC1 is often accompanied by eta-squared (η^2).[1042] Akin to ICC1, η^2 estimates the variance explained by the grouping level using the variance components of a one-way ANOVA. It is calculated as shown in equation A2, where $SS_{between}$ represents the sum of squares between groupings and SS_{total} is the total sum of squares.

$$\eta^2 = \frac{SS_{between}}{SS_{total}} \tag{A2}$$

Despite its common use as a robustness check, η^2 often produces variance estimates that are positively biased.[1043] Though this bias decreases as the size of datasets increases,[1044] making it more reliable in large *n* studies such as mine, η^2 can be accompanied by omega-squared (ω^2) which is adjusted down to account for any positive bias. ω^2 is calculated as shown in equation A3, where $SS_{between}$ represents the sum of squares between groupings of interest, SS_{total} represents the total sum of squares in the model, $DF_{between}$ represents the degrees of freedom between groupings, and MS_{within} represents the mean sum of squares within all model grouping levels.[1045]

$$\omega^2 = \frac{SS_{between} - (DF_{between} * MS_{within})}{SS_{total} + MS_{within}} \tag{A3}$$

1041 Bliese, p. 355.
1042 Shieh, p. 1214; Bliese, p. 356.
1043 Daniel Lakens, 'Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for T-tests and ANOVAs', *Frontiers in Psychology*, 4 (2013), 863 (p. 863).
1044 Albers and Lakens, p. 190.
1045 Ibid.

To ensure that the findings presented within my analysis are robust to alternative estimative techniques, I use these additional equations to calculate the proportion of polling error accounted for by the election level within my data. As each additional approach relies on the variance components of a one-way ANOVA, they can only accommodate one grouping factor. Therefore, they can only be used in relation to two-level models, as they cannot accommodate the additional grouping factors required for three- and four-level models.

A2: Additional ICC Estimates for Two-level Models

To ensure that the findings reported in the main body of the thesis are robust to alternative calculative techniques, Table A1 displays ICC estimates derived from two-level models for my continuous measures of polling inaccuracy using the additional approaches to variance decomposition outlined above.

The ICC values generated using these alternative techniques agree with the substantive findings of the maximum-likelihood approaches reported within the chapter. Values range from 0.35 to 0.62, indicating that 35% to 62% of variance in my continues measures of polling error is due to election-level differences. When 95% confidence intervals are taken into consideration, this range extends to 32% to 65%. Importantly, none of the ICC estimates drop below the 5% threshold justifying the use of multi-level approaches to analysis. These ranges broadly agree with the equivalent range of 39% to 79% reported within the main body of the thesis, indicating its robustness to differing estimative techniques. This lends further evidence to the contention that election-level differences matter for polling inaccuracy and, therefore, the hypothesis that elections vary in their ability to be accurately predicted.

Table A1: ICC estimates for continuous measures of polling inaccuracy from additional frequentist two-level models including standard errors (SE) and 95% confidence intervals.

	MAE	DIM	LPB	APB	ABI 1	ABI 2
ICC1						
ICC	0.52	0.43	0.40	0.63	0.49	0.60
SE	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
95% CI	0.49 - 0.56	0.40 - 0.47	0.37 - 0.43	0.60 - 0.66	0.45 - 0.53	0.57 - 0.64
$\boldsymbol{\eta^2}$						
ICC	0.54	0.46	0.42	0.64	0.51	0.62
SE	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
95% CI	0.51 - 0.56	0.42 - 0.49	0.38 - 0.46	0.61 - 0.67	0.47 -0.55	0.53 - 0.60
$\boldsymbol{\omega^2}$						
ICC	0.52	0.43	0.39	0.62	0.49	0.60
SE	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
95% CI	0.49 - 0.54	0.39 - 0.47	0.35 - 0.43	0.60 - 0.64	0.44 - 0.54	0.57 - 0.63

MAE = mean absolute error, DIM = difference in margin, LPB = leading party bias, APB = average party bias, ABI = average bounded inaccuracy.

Much like the values reported within the main body of the thesis, Table A1 demonstrates that election-level differences exist as important drivers of polling error variance. They also agree that election-level differences are least important for the variance of LPB, accounting for between 39% and 42%, and most important for ABI 2, accounting for between 60% and 62%. The presented values also concur that election-level differences are slightly more important for measures of bounded inaccuracy than distributive inaccuracy. My measures of bounded inaccuracy present an average ICC value of 0.55, while the measures of distributive inaccuracy present an average of 0.50. This suggests that election-level differences account for an average of 55% of the observed variance of my measures of bounded inaccuracy, and an average of 50% of the variance of my measures of distributive inaccuracy. Importantly, inclusive of 95% confidence intervals, each model presents values considerably above the 5% threshold justifying the assessment of group-level effects, undergirding the importance of a multi-level approach to polling error.

To test whether the conclusions drawn from my frequentist models hold across other techniques, Table A2 presents further ICC calculations from additional prior specifications of the Bayesian MCMC models for my continuous measures of polling inaccuracy. These additional priors encompass half-Cauchy prior distributions and have the additional purpose of testing the robustness of my Bayesian findings to alternative prior specifications.

Table A2: ICC estimates for continuous measures from additional Bayesian MCMC two-level models across whole dataset including standard errors (SE) and 95% confidence intervals.

	MAE	**DIM**	**LPB**	**APB**	**ABI 1**	**ABI 2**
Half-Cauchy Priors						
ICC	0.64	0.51	0.46	0.67	0.61	0.72
SE	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)
95% CI	0.61 - 0.66	0.48 - 0.54	0.43 - 0.50	0.65 - 0.70	0.58 - 0.64	0.69 - 0.74

MAE = mean absolute error, DIM = difference in margin, LPB = leading party bias, APB = average party bias, ABI = average bounded inaccuracy.

The ICC values calculated using half-Cauchy priors in Table A2 present an identical range to those reported by models using half student-t priors in the main body of the text. Inclusive of 95% confidence intervals, election-level differences can be seen to account for between 43% and 74% of the variance exhibited by my continuous measures of polling inaccuracy. This agrees with the findings reported in the main body of the thesis and lends further support to the hypothesis that election-level differences are impactful drivers of polling error and, therefore, that elections vary in their ability to be accurately predicted.

Table A3 contains robustness checks for the ICC values presented by my two-level models of binary polling inaccuracy estimated using Laplace approximation. These checks employ adaptive Gauss-Hermite quadrature as the chief alternative to the Laplace approximation reported in the main body of the thesis in relation to two-level models.

Table A3: ICC estimates for binary measures from two-level models using adaptive Gauss-Hermite quadrature (AGHQ) across whole dataset including standard errors (SE) and 95% confidence intervals.

	LVRC	**SBP**
AGHQ		
ICC	0.88	0.44
SE	(0.02)	(0.03)
95% CI	0.84 – 0.91	0.39 – 0.50

LVRC = leading vote share recipient correct;
SBP = significantly biased poll

The values displayed in Table A3 strongly support those reported within the main body of the thesis, only deviating by two percentage points in the case of LVRC and one percentage point in the case of SBP. Election-level differences again account for the majority of variance associated with LVRC (88%), further indicating that they stand as more impactful drivers of substantive polling inaccuracy than my other conceptualisations. These differences also account or 44% of the variance associated with SBP. This not only indicates that election-level differences exist as a significant driver of its variance, but that the degree of variance that it accounts for sits firmly within the range of values associated with other measures of distributive inaccuracy (39% to 70%).

When taken alongside the ICC values reported from models using Laplace approximation within the main body of the thesis, the contents of Table A3 not only demonstrate that the findings are robust across a series of frequentist estimative techniques, but also lend further support to the hypothesis that election-level differences matter for polling error variance and, therefore, that elections differ in their ability to be accurately predicted.

Table A4 displays ICC values derived from two-level Bayesian MCMC models using alternative half-Cauchy priors. These values serve to test the robustness of the figures derived from the model using half student-t priors reported within the main body of the thesis.

Table A4: Election-level ICC estimates for binary measures of polling error calculated from Bayesian MCMC two-level models across whole dataset using alternative half-Cauchy priors. ICC values are supplemented by standard errors (SE) and 95% confidence intervals.

	LVRC	**SBP**
Half- Cauchy Priors		
ICC	0.85	0.44
SE	(0.02)	(0.03)
95% CI	0.81 – 0.88	0.39 – 0.50

DIM = difference in margin, LPB = leading party bias, MAE = mean absolute error, APB = average party bias, ABI = average bounded inaccuracy.

The ICC values calculated using half-Cauchy priors in Table A4 closely resemble those reported in the main body of the thesis. Election-level differences account for 85% of the variance in LVRC and 44% of the variance of SBP. The replication of these values across prior specifications demonstrates the robustness of the findings reported in the main body of the thesis. It also lends further support to the hypothesis that election-level differences are impactful drivers of polling error and, therefore, that elections vary in their ability to be accurately predicted.

When considered collectively, the robustness checks contained with Section A2 demonstrate that, substantively, the ICC values calculated from two-level models in the main body of the thesis hold across a range of estimative techniques and prior specifications. This provides evidence that the conclusions drawn from these two-level models are not simply born of given modelling choices or approaches to variance partitioning.

A3: Additional ICC Estimates for Three-level Models

Table A5 contains estimates of the variance attributable to country-level differences in the maximum likelihood three-level models reported within the main body of the thesis. Though not the main focus of the thesis, these values are nevertheless important to understanding why the election-level ICC values from three-level models take on the values that they do and to underscore the importance of including a country level in multi-level approaches to sources of polling error.

From Table A5, it is clear that the impact of country-level differences varies substantially between my measures of polling error. ICC values range from 0.10 in the case of DIM to 0.60 in the case of ABI 2, indicating that country-level differences account for between 10% and 60% of the observed variance in my continuous measures of polling error. The importance of country-level differences varies significantly between those measures concerned with the leading parties or candidates in an election (DIM and LPB) and those measures which represent aggregations of error across a greater number of competitors (MAE, APB, and ABI variants).

Table A5: Country-level ICC estimates for continuous measures from maximum likelihood three-level models across whole dataset including standard errors (SE) and 95% confidence intervals.

	MAE	DIM	LPB	APB	ABI 1	ABI 2
MLE						
ICC	0.44	0.10	0.16	0.51	0.46	0.60
SE	(0.06)	(0.03)	(0.04)	(0.06)	(0.07)	(0.06)
95% CI	0.33 - 0.56	0.05 - 0.18	0.10 - 0.25	0.40 - 0.61	0.33 - 0.59	0.48 - 0.71
RMLE						
ICC	0.45	0.10	0.16	0.51	0.46	0.60
SE	(0.06)	(0.03)	(0.04)	(0.06)	(0.07)	(0.06)
95% CI	0.34 - 0.56	0.06 - 0.19	0.10 - 0.25	0.41 - 0.62	0.33 - 0.60	0.48 - 0.71

MAE = mean absolute error, DIM = difference in margin, LPB = leading party bias,
APB = average party bias, ABI = average bounded inaccuracy.

Post hoc explanations for the division in ICC values displayed in Table A5 can be posited. The measures of polling inaccuracy which are least affected by country-level differences solely concern the two leading parties or candidates within an election. All elections, irrespective of the country in which they are nested, will be contested by at least two parties or candidates, regardless of differences in their electoral system. Therefore, all elections will possess values for DIM and LPB that are not significantly impacted by otherwise meaningful country-level differences, such as the number of parties permitted by electoral systems. This lends itself to a diminished understanding of the role of country-level differences for polling error variance. By contrast, the measures concerning aggregations will vary on the basis of the number of parties contesting an election and, therefore, the number of error values created. As this number will vary by electoral system, and therefore by country, the variance associated with these measures can be expected to be more significantly affected by country-level differences.

While these country-level differences are interesting, they are not, in isolation, the focus of this thesis. The country-level ICC values associated with my measures of polling inaccuracy only matter insofar as they affect the proportion of variance attributable to the election-level. This is why the individual country-level values have been relegated to this appendix and why further discussion of the magnitude of differences in country-level ICC values is not present. I concede that these differences are interesting and, perhaps, unexpected given their severity. However, beyond the post hoc explanations offered, further investigation is beyond the scope of this thesis and exists as an avenue for future research.

Table A6 presents additional election-level ICC values for my range of continuous error measures calculated from Bayesian MCMC models using half-Cauchy priors. These values represent the proportion of variance in these measures attributable to election-level differences when country-level differences are controlled for.

Table A6: Election-level ICC estimates for continuous measures of polling inaccuracy from additional Bayesian MCMC three-level models across whole dataset including standard errors (SE) and 95% confidence intervals.

	MAE	**DIM**	**LPB**	**APB**	**ABI 1**	**ABI 2**
Half-Cauchy Priors						
ICC	0.27	0.41	0.33	0.24	0.25	0.22
SE	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
95% CI	0.24 - 0.29	0.39 - 0.42	0.32 - 0.34	0.21 - 0.25	0.22 - 0.28	0.19 - 0.25

MAE = mean absolute error, DIM = difference in margin, LPB = leading party bias,
APB = average party bias, ABI = average bounded inaccuracy.

The ICC estimates presented in Table A6 are identical to those reported from models using half student-t priors within the main body of the thesis. This indicates that these findings and the conclusions based upon them are robust to differing prior distributions and not simply artefacts of a given model specification. As such, from Bayesian MCMC models, election-level differences can be confidently said to account for between 22% and 41% of the observed variance across my continuous measures of polling inaccuracy when country-level effects are controlled for.

The variance accounted for by country-level differences responsible for the reductions in Table A6 are presented in Table A7 as ICC values. When 95% confidence intervals are considered, these values indicate that country-level differences account for between 11% and 68% of the variance observed across my continuous measures of polling inaccuracy. As such, it would be remiss of me to conclude that only election-level differences are impactful for polling error. Indeed, across each of my measures of error, the proportion of variance accounted for by country-level differences maps closely, though not exactly, to the reduction in variance explained by election-level differences. Importantly, most reductions sit within the 95% confidence intervals reported by models. Reductions in the error variance explained by election-level differences are to be expected, as reductions such as this are the very point of including controls within models, and the inexact nature of their mapping is indicative of alternative sources of error variance beyond the scope of the model. It is important to note that while through my models I hope to represent sources of polling error as best as possible, I do not claim to capture all possible sources and, therefore, account for 100% of variance.

Table A7: Country-level ICC estimates for continuous measures from Bayesian MCMC three-level models across whole dataset including standard errors (SE) and 95% confidence intervals.

	MAE	**DIM**	**LPB**	**APB**	**ABI 1**	**ABI 2**
Half Student-t Priors						
ICC	0.45	0.11	0.15	0.52	0.46	0.60
SE	(0.05)	(0.03)	(0.03)	(0.04)	(0.05)	(0.04)
95% CI	0.35 - 0.53	0.05 - 0.17	0.11 - 0.24	0.44 - 0.61	0.36 - 0.57	0.51 - 0.68
Half-Cauchy Priors						
ICC	0.45	0.11	0.15	0.52	0.46	0.61
SE	(0.05)	(0.03)	(0.03)	(0.04)	(0.06)	(0.04)
95% CI	0.36 - 0.54	0.06 - 0.17	0.11 - 0.22	0.44 - 0.61	0.35 - 0.57	0.52 - 0.68

MAE = mean absolute error, DIM = difference in margin, LPB = leading party bias, APB = average party bias, ABI = average bounded inaccuracy.

The ICC values presented in Table A7 display the same strong variability exhibited by earlier country-level estimates. This indicates that, for certain error measures, country-level differences are significant determinants of variance, but for others their effect is far smaller. Despite this, no values fall below the 5% threshold justifying the investigation of group level effects, even when 95% confidence intervals are considered. This indicates that country-level differences remain and important and impactful driver of error variance across my continuous measures of polling inaccuracy.

The ICC values displayed in Table A8 serve as robustness checks for those calculated from Bayesian MCMC three-level models using half student-t priors, as they are derived from models using half-Cauchy priors. The ICC values calculated from the model using half-Cauchy priors indicate that 68% of the variance observed in LVRC and 35% of the variance observed in SBP can be attributed to election-level differences. These values agree substantively with those reported in the main body of the thesis. As such, the reported findings are robust across differing prior specifications.

Table A8: Election-level ICC estimates for binary measures of polling error calculated from Bayesian MCMC three-level models across whole dataset using alternative half-Cauchy priors. ICC values are supplemented by standard errors (SE) and 95% confidence intervals.

	LVRC	**SBP**
Half- Cauchy Priors		
ICC	0.68	0.35
SE	(0.02)	(0.02)
95% CI	0.63 – 0.71	0.33 – 0.39

LVRC = largest vote share recipient correct, SBP = significantly biased poll.

Table A9 displays the country-level ICC estimates associated with the three-level model reported in the main body of thesis estimated using Bayesian MCMC. The reported ICC values indicate that country-level differences exist as important additional determinants of variance across my binary measures of polling error, reducing the variance attributable to election-level differences. The variance attributed to country-level differences maps closely, though not directly on to the reduction seen in the variance accounted for by the election-level, suggesting the presence of sources of variance beyond the scope of the model. This said, all reductions rest within the reported 95% confidence intervals.

Table A9: Country-level ICC estimates for binary measures of polling inaccuracy calculated from Bayesian MCMC three-level models across whole dataset including standard errors (SE) and 95% confidence intervals.

	LVRC	**SBP**
Half Student-t Priors		
ICC	0.16	0.11
SE	(0.04)	(0.04)
95% CI	0.08 – 0.24	0.04 – 0.19
Half-Cauchy Priors		
ICC	0.17	0.11
SE	(0.04)	(0.04)
95% CI	0.09 – 0.26	0.04 – 0.19

LVRC = largest vote share recipient correct; SBP = significantly biased poll.

A4: Additional ICC Estimates for Four-level Models

Table A10 contains ICC values representing the proportion of variance across my continuous measures of polling inaccuracy accounted for by the country- and pollster-level differences. These values indicate that between 10% and 55% of variance across my continuous measures can be attributed to country-level differences, while 8% to 18% can be attributed to pollster-level differences. When 95% confidence intervals are considered, both country- and pollster-level differences consistently account for at least the 5% of variance necessary to justify their discrete, group-level investigation

Table A10: Country- and Pollster-level ICC estimates for continuous measures of polling error calculated from maximum likelihood (MLE) and restricted maximum likelihood (RMLE) four-level models across whole dataset including standard errors (SE) and 95% confidence intervals.

	MAE	**DIM**	**LPB**	**APB**	**ABI 1**	**ABI 2**
Country-level ICC Estimates						
MLE						
ICC	0.35	0.10	0.13	0.44	0.42	0.55
SE	(0.04)	(0.03)	(0.03)	(0.03)	(0.05)	(0.04)
95% CI	0.27 - 0.44	0.05 - 0.15	0.08 - 0.20	0.40 - 0.51	0.32 - 0.52	0.46 - 0.63
RMLE						
ICC	0.36	0.10	0.14	0.45	0.43	0.55
SE	(0.04)	(0.03)	(0.03)	(0.02)	(0.05)	(0.04)
95% CI	0.27 - 0.44	0.05 - 0.15	0.08 - 0.20	0.40 - 0.51	0.32 - 0.52	0.46 - 0.63
Pollster-level ICC Estimates						
MLE						
ICC	0.18	0.08	0.11	0.16	0.06	0.09
SE	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
95% CI	0.16 - 0.19	0.07 - 0.09	0.10 - 0.13	0.14 - 0.17	0.05 - 0.07	0.07 - 0.09
RMLE						
ICC	0.18	0.08	0.11	0.16	0.06	0.08
SE	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
95% CI	0.16 - 0.19	0.07 - 0.09	0.10 - 0.13	0.14 - 0.17	0.05 - 0.07	0.07 - 0.09

MAE = mean absolute error, DIM = difference in margin, LPB = leading party bias, APB = average party bias, ABI = average bounded inaccuracy.

Table A11 presents robustness checks for the ICC values calculated from four-level Bayesian MCMC models within the main body of the thesis. These models are estimated using alternative half-Cauchy priors. The values displayed within the table range from 21% in the case of ABI 2 to 38% in the case of DIM. When 95% confidence intervals are considered, this range extends from 18% to 39%. These values agree substantively with those reported from Bayesian MCMC models with half student-t priors in the main body of the thesis, varying by only 1% in the case of ABI 1. The similarity between ICC values across different prior specifications demonstrates that the findings presented within the thesis are robust across differing estimative approaches and are therefore not simply the artefact of a given modelling approach.

Table A11: Election-level ICC estimates for continuous measures of polling inaccuracy from additional Bayesian MCMC four-level models across whole dataset including standard errors (SE) and 95% confidence intervals.

	MAE	**DIM**	**LPB**	**APB**	**ABI 1**	**ABI 2**
Half-Cauchy Priors						
ICC	0.26	0.38	0.31	0.23	0.24	0.21
SE	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
95% CI	0.23 - 0.28	0.37 - 0.39	0.30 - 0.33	0.21 - 0.25	0.21 - 0.27	0.18 - 0.24

MAE = mean absolute error, DIM = difference in margin, LPB = leading party bias, APB = average party bias, ABI = average bounded inaccuracy.

Table A12 contains robustness ICC values associated with the country- and pollster-levels calculated from my four-level Bayesian MCMC models. From these values, it is clear that the prominence of the country-level is diminished within four-level models when compared to earlier three-level models. Indeed, country-level differences are largely negligible determinants of error variance in the case of DIM and LPB, with their prominence being reduced by an average of 13%. Across these same measures, pollster-level differences account for an average of 15% of observed variance. Given the lack of significant movement in relation to election-level ICC values, this indicates that much of variance attributed to country-level differences within three-level models was in fact the product of different pollsters contained within different countries.

Table A12: Country- and Pollster-level ICC estimates for continuous measures of polling error calculated from additional Bayesian MCCM four-level models using Half-Cauchy priors across whole dataset including standard errors (SE) and 95% confidence intervals.

	MAE	**DIM**	**LPB**	**APB**	**ABI 1**	**ABI 2**
			Country-level ICC Estimates			
Half-Cauchy Priors						
ICC	0.24	0.03	0.01	0.28	0.31	0.35
SE	(0.04)	(0.02)	(0.01)	(0.04)	(0.05)	(0.05)
95% CI	0.16 - 0.32	0.00 - 0.09	0.00 - 0.04	0.19 - 0.36	0.20 - 0.40	0.23 - 0.44
			Pollster-level ICC Estimates			
Half-Cauchy Priors						
ICC	0.20	0.15	0.11	0.15	0.13	0.13
SE	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
95% CI	0.17 - 0.22	0.14 - 0.16	0.09 - 0.13	0.13 - 0.17	0.12 - 0.15	0.11 - 0.17

MAE = mean absolute error, DIM = difference in margin, LPB = leading party bias, APB = average party bias, ABI = average bounded inaccuracy.

Table A13 presents the country- and pollster-level ICC values calculated from the four-level models using Laplace approximation. These estimates share the same potential problems as those presented within the main body of the thesis. Given the complexity of the four-level models and the comparative difficulty of approximating the integral of their likelihood functions, their estimation may be too difficult for simple techniques like Laplace approximation.[1046] With this in mind, the values displayed in Table A13 are to be treated with suspicion. Though they indicate that both country- and pollster-level differences are largely negligible determinants of variance across my binary measures of polling error, I cannot draw any conclusions from the values presented until their validity has been checked. This can be achieved through Bayesian MCMC approaches which are far more capable of handling complex modelling structures.[1047]

Table A13: Country- and Pollster-level ICC estimates for binary measures of polling error calculated using Laplace approximation from four-level models covering whole dataset including standard errors (SE) and 95% confidence intervals.

	LVRC	**SBP**
Country-level ICCs		
Laplace Approximation		
ICC	0.00	0.05
SE	(0.01)	(0.01)
95% CI	0.00 - 0.03	0.03 - 0.07
Pollster-level ICCs		
Laplace Approximation		
ICC	0.01	0.12
SE	(0.01)	(0.02)
95% CI	0.00 - 0.04	0.09 - 0.15

MAE = mean absolute error, DIM = difference in margin, LPB = leading party bias, APB = average party bias, ABI = average bounded inaccuracy.

[1046] Capanu, Gonen, and Begg, p. 1.
[1047] Press and others, p. 398.

Table A14 represents the final robustness check for chapter 4 of my thesis. It calculates election-level ICC values for my binary measures of polling inaccuracy from models estimating using Bayesian MCMC across additional half-Cauchy priors. For LVRC, these values range from 0.65 to 0.72, indicating that between 65% and 72% of the variance observed in LVRC can be attributed to election-level differences. This range agrees substantively with that reported in the main body of the thesis, indicating that the reported findings are robust across differing prior specifications.

Table A14: Election-level ICC estimates for binary measures of polling error calculated from Bayesian MCMC four-level models across whole dataset using alternative half-Cauchy priors. ICC values are supplemented by standard errors (SE) and 95% confidence intervals.

	LVRC	**SBP**
Half- Cauchy Priors		
ICC	0.69	0.32
SE	(0.02)	(0.02)
95% CI	0.65 – 0.72	0.29 – 0.35

MAE = mean absolute error, DIM = difference in margin, LPB = leading party bias, APB = average party bias, ABI = average bounded inaccuracy

The same conclusion can be drawn in relation to SBP. The ICC values displayed in Table A14 range from 0.29 to 0.35, indicating that between 29% and 35% of the variance observed in SBP can be attributed to election-level differences. Again, this range agrees substantively with that reported within the main body of the thesis, indicating that this finding is also robust to alternative estimative approaches.

Table A15: Country- and pollster-level ICC estimates for binary measures error calculated from additional Bayesian MCMC four-level models across whole dataset including standard errors (SE) and 95% confidence intervals.

	LVRC	SBP
Country-level ICCs		
Half Student-t Priors		
ICC	0.14	0.11
SE	(0.04)	(0.04)
95% CI	0.06 – 0.22	0.05 – 0.19
Half- Cauchy Priors		
ICC	0.13	0.12
SE	(0.04)	(0.04)
95% CI	0.06 – 0.22	0.05 – 0.19
Pollster-level ICCs		
Half Student-t Priors		
ICC	0.02	0.12
SE	(0.01)	(0.01)
95% CI	0.02 – 0.03	0.10 – 0.14
Half-Cauchy Priors		
ICC	0.03	0.12
SE	(0.01)	(0.01)
95% CI	0.01 – 0.03	0.10 – 0.14

MAE = mean absolute error, DIM = difference in margin, LPB = leading party bias, APB = average party bias, ABI = average bounded inaccuracy

Table A15 unpacks the impact of pollster and country level differences withing four-level models estimated using Bayesian MCMC. Immediately, it is clear that previous ICC values calculated using Laplace approximation were not entirely representative. While pollster-level differences remain negligible determinants of variance in LVRC, accounting for between 1% and 3%, country-level differences can be seen to be impactful drivers of its variance, accounting for between 6% and 22%. In the case of SBP, both the country and pollster levels stand as important drivers of observed variance. When 95% confidence intervals are considered, country-level differences accounting for between 5% and 19% of observed variance in SBP, while pollster-level differences account for between 10% and 14%.

A5: The Variation of Election-level ICC Values Across a Subset of Countries

Table A16: ICC values associated with the election-level across those countries for which at least 400 polls were available. Estimates are derived from models using MCMC with weakly informative student-t priors and concern continuous measures of polling error.

Country	MAE†	95% CI	DIM	95% CI	LPB	95% CI	APB	95% CI	ABI 1	95% CI	ABI 2	95% CI
AUS	0.44 (0.12)	0.27-0.75	0.39 (0.07)	0.25-0.53	0.20 (0.07)	0.09-0.35	0.34 (0.05)	0.25-0.43	0.29 (0.08)	0.10-0.47	0.64 (0.07)	0.48-0.76
CAN	0.54 (0.07)	0.40-0.67	0.62 (0.07)	0.48-0.74	0.54 (0.08)	0.35-0.68	0.28 (0.08)	0.12-0.44	0.42 (0.08)	0.27-0.57	0.46 (0.07)	0.32-0.61
FRA	0.81 (0.05)	0.70-0.89	0.78 (0.05)	0.68-0.86	0.83 (0.04)	0.74-0.90	0.88 (0.04)	0.76-0.93	0.84 (0.05)	0.73-0.91	0.84 (0.04)	0.74-0.90
GER	0.37 (0.08)	0.22-0.55	0.33 (0.07)	0.20-0.49	0.37 (0.08)	0.24-0.55	0.53 (0.09)	0.34-0.68	0.34 (0.08)	0.19-0.52	0.34 (0.08)	0.20-0.52
UK	0.39 (0.07)	0.26-0.54	0.40 (0.07)	0.27-0.54	0.22 (0.06)	0.10-0.32	0.46 (0.07)	0.33-0.61	0.24 (0.11)	0.07-0.49	0.43 (0.11)	0.23-0.68
USA	0.50 (0.04)	0.42-0.58	0.56 (0.04)	0.47-0.64	0.58 (0.05)	0.46-0.67	0.50 (0.05)	0.43-0.61	0.43 (0.05)	0.34-0.52	0.50 (0.05)	0.41-0.59

† Standard errors are provided below each ICC estimate in parentheses.

The findings displayed in Table A16 underscore conclusions drawn in the main body of the thesis, insofar as they indicate that the importance of election-level differences as drivers of polling error variation varies across countries. This can be seen in the differences in the ICC values associated with each country contained within the table. The findings also underscore another theme that emerges within the main analysis of this thesis: the importance of election-level differences also varies between measures of polling error. This is evident in the differences between the ICC values associated with each measure of error *within* the countries contained in the table.

Appendix B: Additional Figures and Predictive Models to Support the Inferential Analysis in Chapter 5

In this appendix, I provide additional models and discussion to elaborate the analysis present in chapter 5. The appendix is split into six sections. Sections B1 through B4 provide additional additive prediction models for measures of distributive polling error, bounded error, polling bias, and substantive error, respectively. Section B5 presents additional interactive prediction models for all measures of polling error other than MAE (as it is displayed in the main body of the thesis). Finally, Section B6 contains additional robustness checks for all measures of polling error, excluding MAE, to ensure that earlier findings produced by election-level only prediction models are robust to the presence of control variables from additional grouping clusters within the four-level structure of polling error.

B1: Additional Additive Prediction Models for Measures of Distributive Polling Error

This section contains additional additive prediction models for my remaining measure of distributive polling error, the difference in margin exhibited by a poll (DIM). Models are run across my full polling dataset, as well as the subsets of data for which partisanship, late decision-making, and the left-right standard deviation of political parties is available both individually and jointly. The models find that election-level variables stand as useful predictors of DIM, reducing the average RMSE exhibited by prediction models. However, the extent to which election-level variable prove predictively useful varies on the basis of the subset of data used for modelling. The sections that follow decompose the predictive utility of election-level variables across my measures of bounded inaccuracy, bias, and substantive inaccuracy.

Additive Prediction models for Difference in Margin (DIM)

Table B1: Average RMSE values for DIM calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated from all available data (n = 11,832).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
DIM ~ 1 (Null)	5.519	-	-
Ex-ante Variables			
+ snap	5.514	-0.005	-0.005
+ election type	5.513	-0.001	-0.006
+ round two	5.513†	0.000†	-0.006†
+ system change	5.503	-0.010	-0.016
+ registration difference	5.503†	0.000†	-0.016†
Ex-post Variables			
+ turnout	5.496	-0.007	-0.023
+ turnout change	5.478	-0.018	-0.041
+ ENEP	5.462	-0.016	-0.057
+ ENEP change	5.445	-0.017	-0.074
+ margin of victory	5.399	-0.046	-0.120

† Including the round two and registration difference predictor variables yielded negligible reductions in average RMSE that could not be displayed to three significant figures.

Table B2: Average RMSE values for DIM calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the strength of partisan loyalty within the electorate could be calculated (n = 9,115).

Model	**Average RMSE**	**Iterative Change in RMSE**	**Change in RMSE Compared to Null**
DIM ~ 1 (Null)	5.322	-	-
Ex-ante Variables			
+ snap	5.317	-0.005	-0.005
+ election type	5.311	-0.006	-0.011
+ round two	5.311†	0.000†	-0.011†
+ system change	5.295	-0.016	-0.027
+ registration difference	5.295†	0.000†	-0.027†
+ partisanship	5.294	-0.001	-0.028
Ex-post Variables			
+ turnout	5.293	-0.001	-0.029
+ turnout change	5.273	-0.020	-0.049
+ ENEP	5.256	-0.017	-0.066
+ ENEP change	5.236	-0.020	-0.086
+ margin of victory	5.186	-0.050	-0.136

† Including the round two and registration difference predictor variables resulted in negligible reductions in average RMSE that could not be displayed to three significant figures.

Table B3: Average RMSE values for DIM calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the extent of late decision-making within the electorate could be calculated (n = 3,285).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
DIM ~ 1 (Null)	5.905	-	-
Ex-ante Variables†			
+ snap	5.897	-0.008	-0.008
+ election type	5.884	-0.013	-0.021
+ system change	5.881	-0.003	-0.024
+ registration difference	5.876	-0.005	-0.029
Ex-post Variables			
+ turnout	5.870	-0.006	-0.035
+ turnout change	5.869	-0.001	-0.036
+ ENEP	5.826	-0.043	-0.079
+ ENEP change	5.794	-0.032	-0.111
+ margin of victory	5.754	-0.040	-0.151
+ late deciders	5.749	-0.005	-0.156

† No round two presidential elections were present within the subset of data used for modelling, so the variable was removed from consideration.

Table B4: Average RMSE values for DIM calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the left/right standard deviation of the political position of parties and candidates could be calculated (n = 939).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
DIM ~ 1 (Null)	5.105	-	-
Ex-ante Variables†			
+ snap	5.090	-0.015	-0.015
+ election type	5.078	-0.012	-0.027
+ round two	5.028	-0.050	-0.077
+ registration difference	5.031	0.003	-0.074
+ left/right std. dev.	4.978	-0.053	-0.127
Ex-post Variables			
+ turnout	4.909	-0.069	-0.196
+ turnout change	4.892	-0.017	-0.213
+ ENEP	4.838	-0.054	-0.267
+ ENEP change	4.825	-0.013	-0.280
+ margin of victory	4.774	-0.051	-0.331

† No instances of system change were present between elections within the subset of data used for modelling, so the variable was removed from consideration.

Table B5: Average RMSE values for DIM calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the strength of partisan loyalty and extent of late decision-making within the electorate could be calculated, as well as the left/right standard deviation of the political position of parties and candidates (n = 293).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
DIM ~ 1 (Null)	6.235	-	-
+ left/right std. dev.	6.091	-0.144	-0.144
+ partisanship	5.931	-0.160	-0.304
+ late deciders	5.027	-0.904	-0.600

B2: Additional Additive Prediction Models for Measures of Bounded Polling Error

Additive Prediction Models for the Average Bounded Inaccuracy 1 (ABI 1)

Table B6: Average RMSE values for ABI 1 calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated from all available data (n = 7,403).

Model	Average RMSE	Iterative Change in Average RMSE	Change in RMSE Compared to Null
ABI 1 ~ 1 (Null)	2.810	-	-
Ex-ante Variables			
+ snap	2.806	-0.004	-0.004
+ election type	2.766	-0.040	-0.044
+ round two	2.766†	0.000†	-0.044†
+ system change	2.760	-0.006	-0.050
+ registration difference	2.759	-0.001	-0.051
Ex-post Variables			
+ turnout	2.755	-0.004	-0.055
+ turnout change	2.751	-0.004	-0.059
+ ENEP	2.744	-0.007	-0.064
+ ENEP change	2.743	-0.001	-0.065
+ margin of victory	2.723	-0.020	-0.085

† Including the round two predictor variable yielded a negligible reduction in average RMSE that could not be displayed to three significant figures.

Table B7: Average RMSE values for ABI 1 calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the strength of partisan loyalty within the electorate could be calculated (n = 5,909).

Model	Average RMSE	Iterative Change in Average RMSE	Change in RMSE Compared to Null
ABI 1 ~ 1 (Null)	2.648	-	-
Ex-ante Variables			
+ snap	2.642	-0.006	-0.006
+ election type	2.632	-0.010	-0.016
+ round two	2.632†	0.000†	-0.016†
+ system change	2.621	-0.011	-0.027
+ registration difference	2.620	-0.001	-0.028
+ partisanship	2.619	-0.001	-0.029
Ex-post Variables			
+ turnout	2.619†	0.000†	-0.029†
+ turnout change	2.618	-0.001	-0.030
+ ENEP	2.610	-0.008	-0.038
+ ENEP change	2.610†	0.000†	-0.038†
+ margin of victory	2.560	-0.050	-0.088

† Including the round two, turnout, and ENEP change predictor variables yielded a negligible reduction in average RMSE that could not be displayed to three significant figures.

Table B8: Average RMSE values for ABI 1 calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the extent of late decision-making within the electorate could be calculated (n = 2,488).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
ABI 1 ~ 1 (Null)	2.340	-	-
Ex-ante Variables†			
+ snap	2.340‡	0.000‡	0.000‡
+ election type	2.340‡	0.000‡	0.000‡
+ registration difference	2.339	-0.001	-0.001
Ex-post Variables			
+ turnout	2.322	-0.017	-0.018
+ turnout change	2.321	-0.001	-0.019
+ ENEP	2.317	-0.004	-0.023
+ ENEP change	2.308	-0.009	-0.032
+ margin of victory	2.283	-0.025	-0.057
+ late deciders	2.256	-0.027	-0.084

† No second-round presidential elections were present within the subset, so the variable was removed from consideration. Additionally, too few instances of system change were present to be used within the model.

‡ Including the snap and election type predictor variables yielded negligible reductions in average RMSE that could not be displayed to three significant figures.

Table B9: Average RMSE values for ABI 1 calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the left/right standard deviation of the political position of parties and candidates could be calculated (n = 708).

Model	Average RMSE	Change in Average RMSE	Change in RMSE Relative to Null
ABI 1 ~ 1 (Null)	2.220	-	-
Ex-ante Variables†			
+ registration difference	2.221	0.001	0.001
+ left/right std. dev.	2.217	-0.004	-0.003
Ex-post Variables			
+ turnout	2.215	-0.002	-0.005
+ turnout change	2.217	0.002	-0.003
+ ENEP	2.214	-0.003	-0.006
+ ENEP change	2.197	-0.017	-0.023
+ margin of victory	2.202	0.005	-0.018

† No instances of system change were present within the subset of data used for modelling, so this predictor was removed from consideration. Additionally, two few instances of snap and second round elections existed within the subset of data to permit their use in predictive modelling as they failed to provide meaningful variation. Finally, electoral type was largely homogenous within the subset of data used, rendering the election type variable unsuitable as a predictor variable. These issues can be attributed to the small size of the subset of data used.

Table B10: Average RMSE values for ABI 1 calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the strength of partisan loyalty and extent of late decision-making within the electorate could be calculated, as well as the left/right standard deviation of the political position of parties and candidates (n = 270).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
ABI 1 ~ 1 (Null)	2.095	-	-
+ left/right std. dev.	2.088	-0.007	-0.007
+ partisanship	2.045	-0.043	-0.050
+ late deciders	1.773	-0.272	-0.322

Additive Prediction Models for Second Measure of Average Bounded Inaccuracy (ABI 2)

Table B11: Average RMSE values for ABI 2 calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated from all available data (n = 8,754).

Model	Average RMSE	Iterative Change in Average RMSE	Change in RMSE Compared to Null
ABI 2 ~ 1 (Null)	2.270	-	-
Ex-ante Variables			
+ snap	2.265	-0.005	-0.005
+ election type	2.223	-0.042	-0.047
+ round two	2.222	-0.001	-0.048
+ system change	2.215	-0.007	-0.055
+ registration difference	2.215†	0.000†	-0.055†
Ex-post Variables			
+ turnout	2.207	-0.008	-0.063
+ turnout change	2.204	-0.003	-0.066
+ ENEP	2.192	-0.012	-0.078
+ ENEP change	2.192†	0.000†	-0.078†
+ margin of victory	2.167	-0.025	-0.103

† Including the registration difference and ENEP change predictor variables yielded negligible reductions in average RMSE which could not be displayed to three significant figures.

Table B12: Average RMSE values for ABI 2 calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the strength of partisan loyalty within the electorate could be calculated (n = 6,950).

Model	Average RMSE	Iterative Change in Average RMSE	Change in RMSE Compared to Null
ABI 2 ~ 1 (Null)	2.121	-	-
Ex-ante Variables			
+ snap	2.114	-0.007	-0.007
+ election type	2.097	-0.017	-0.024
+ round two	2.096	-0.001	-0.025
+ system change	2.085	-0.011	-0.036
+ registration difference	2.085†	0.000†	-0.036†
+ partisanship	2.083	-0.002	-0.038
Ex-post Variables			
+ turnout	2.080	-0.003	-0.041
+ turnout change	2.079	-0.001	-0.042
+ ENEP	2.072	-0.007	-0.049
+ ENEP change	2.072†	0.000†	-0.049†
+ margin of victory	2.010	-0.062	-0.111

† Including the registration difference and ENEP change predictor variables resulted in negligible reductions in average RMSE that could not be displayed to three significant figures.

Table B13: Average RMSE values for ABI 2 calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the extent of late decision-making within the electorate could be calculated (n = 2,658).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
ABI 2 ~ 1 (Null)	1.841	-	-
Ex-ante Variables†			
+ snap	1.839	-0.002	-0.002
+ election type	1.840	0.001	-0.001
+ registration difference	1.840‡	0.000‡	-0.001‡
Ex-post Variables			
+ turnout	1.814	-0.026	-0.027
+ turnout change	1.814‡	0.000‡	-0.027‡
+ ENEP	1.808	-0.006	-0.033
+ ENEP change	1.781	-0.027	-0.060
+ margin of victory	1.761	-0.020	-0.080
+ late deciders	1.749	-0.012	-0.092

† No second-round presidential elections were present within the subset, so the variable was removed from consideration. Moreover, too few instances of system change existed within the subset to be included within the prediction model.

‡ Including the registration difference and turnout change predictor variables resulted in negligible reductions in average RMSE which could not be displayed to three significant figures.

Table B14: Average RMSE values for ABI 2 calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the left/right standard deviation of the political position of parties and candidates could be calculated (n = 813).

Model	**Average RMSE**	**Iterative Change in RMSE**	**Change in RMSE Compared to Null**
ABI 2 ~ 1 (Null)	1.457	-	-
Ex-ante Variables			
+ snap	1.458	0.001	0.001
+ election type	1.443	-0.015	-0.014
+ round two	1.437	-0.006	-0.020
+ registration difference	1.438	0.001	-0.019
+ left/right std. dev.	1.424	-0.014	-0.033
Ex-post Variables			
+ turnout	1.425	0.001	-0.032
+ turnout change	1.426	0.001	-0.031
+ ENEP	1.418	-0.008	-0.039
+ ENEP change	1.410	-0.008	-0.047
+ margin of victory	1.416	0.006	-0.041

Table B15: Average RMSE values for ABI 2 calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the strength of partisan loyalty and extent of late decision-making within the electorate could be calculated, as well as the left/right standard deviation of the political position of parties and candidates (n = 283).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
ABI 2 ~ 1 (Null)	1.324	-	-
+ left/right std. dev.	1.256	-0.068	-0.068
+ partisanship	1.223	-0.033	-0.101
+ late deciders	1.066	-0.157	-0.258

B3: Additional Additive Prediction Models for Measures of Polling Bias

Additive Prediction Models for Leading Party Bias (LPB)

Table B16: Average RMSE values for LPB calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated from all available data (n = 11,832).

Model	Average RMSE	Iterative Change in Average RMSE	Change in RMSE Compared to Null
LPB ~ 1 (Null)	0.302	-	-
Ex-ante Variables			
+ snap	0.301	-0.001	-0.001
+ election type	0.301†	0.000†	-0.001†
+ round two	0.301†	0.000†	-0.001†
+ system change	0.301†	0.000†	-0.001†
+ registration difference	0.301†	0.000†	-0.001†
Ex-post Variables			
+ turnout	0.301†	0.000†	-0.001†
+ turnout change	0.301†	0.000†	-0.001†
+ ENEP	0.300	-0.001	-0.002
+ ENEP change	0.299	-0.001	-0.003
+ margin of victory	0.295	-0.004	-0.007

† Including the election type, round two, system change, registration difference, turnout, and turnout change variables resulted in small reductions in average RMSE that could not be displayed to three significant figures. The prevalence of small-scale changes to average RMSE is likely an artefact of the granular scale on which LPB is measured. It may, however, simply be that measures of bias are less well predicted by the selected range of predictor variables than other forms of polling error. This contention is explored within the main body of the text.

Table B17: Average RMSE values for LPB calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the strength of partisan loyalty within the electorate could be calculated (n = 9,115).

Model	Average RMSE	Iterative Change in Average RMSE	Change in RMSE Compared to Null
LPB ~ 1 (Null)	0.283	-	-
Ex-ante Variables			
+ snap	0.282	-0.001	-0.001
+ election type	0.282†	0.000†	-0.001†
+ round two	0.282†	0.000†	-0.001†
+ system change	0.281	-0.001	-0.002
+ registration difference	0.281†	0.000†	-0.002†
+ partisanship	0.281†	0.000†	-0.002†
Ex-post Variables			
+ turnout	0.281†	0.000†	-0.002†
+ turnout change	0.281†	0.000†	-0.002†
+ ENEP	0.280	-0.001	-0.003
+ ENEP change	0.280†	0.000†	-0.003†
+ margin of victory	0.277	-0.003	-0.006

† Including the election type, round two, registration difference, partisanship, turnout, turnout change, ENEP, and ENEP change variables resulted in small reductions reduction in average RMSE which could not be displayed to three significant figures. The prevalence of small-scale changes to average RMSE may be an artefact of the granular scale on which LPB is measured. It may, however, simply be that measures of bias are less well predicted by the selected range of predictor variables than other forms of polling error. This contention is explored within the main body of the text.

Table B18: Average RMSE values for LPB calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the extent of late decision-making within the electorate could be calculated (n = 3,285).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
LPB ~ 1 (Null)	0.309	-	-
Ex-ante Variables†			
+ snap	0.307	-0.002	-0.002
+ election type	0.307‡	0.000‡	-0.002‡
+ system change	0.307‡	0.000‡	-0.002‡
+ registration difference	0.305	-0.002	-0.004
Ex-post Variables			
+ turnout	0.305‡	0.000‡	-0.004‡
+ turnout change	0.305‡	0.000‡	-0.004‡
+ ENEP	0.305‡	0.000‡	-0.004‡
+ ENEP change	0.304	-0.001	-0.005
+ margin of victory	0.293	-0.011	-0.016
+ late deciders	0.292	-0.001	-0.017

† No round two presidential elections were present within the subset of data, so the variable was removed from consideration.

‡ Including the election type, system change, turnout, turnout change, and ENEP variables resulted in small reductions reduction in average RMSE which could not be displayed to three significant figures. The prevalence of small-scale changes to average RMSE may be an artefact of the granular scale on which LPB is measured. It may, however, simply be that measures of bias are less well predicted by the selected range of predictor variables than other forms of polling error. This contention is explored within the main body of the text.

Table B19: Average RMSE values for LPB calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the left/right standard deviation of the political position of parties and candidates could be calculated (n = 939).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
LPB ~ 1 (Null)	0.302	-	-
Ex-ante Variables†			
+ snap	0.302‡	0.000‡	0.000‡
+ election type	0.300	-0.002	-0.002
+ round two	0.300‡	0.000‡	-0.002‡
+ registration difference	0.300‡	0.000‡	-0.002‡
+ left/right std. dev.	0.294	-0.006	-0.008
Ex-post Variables			
+ turnout	0.294‡	0.000‡	-0.008‡
+ turnout change	0.294‡	0.000‡	-0.008‡
+ ENEP	0.294‡	0.000‡	-0.008‡
+ ENEP change	0.294‡	0.000‡	-0.008‡
+ margin of victory	0.282	-0.012	-0.020

† No instances of electoral system change between elections were present within the subset of data, so the variable was removed from consideration.

‡ Including the snap, round two, registration difference, turnout, turnout change, ENEP, and ENEP change variables resulted in small reductions reduction in average RMSE which could not be displayed to three significant figures. The prevalence of small-scale changes to average RMSE may be an artefact of the granular scale on which LPB is measured. It may, however, simply be that measures of bias are less well predicted by the selected range of predictor variables than other forms of polling error. This contention is explored within the main body of the text.

Table B20: Average RMSE values for LPB calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the strength of partisan loyalty and extent of late decision-making within the electorate could be calculated, as well as the left/right standard deviation of the political position of parties and candidates (n = 293).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
LPB ~ 1 (Null)	0.388	-	-
+ left/right std. dev.	0.389	0.001	0.001
+ late deciders	0.386	-0.003	-0.002
+ partisanship	0.365	-0.001	-0.003

Additive Prediction Models for Average Party Bias (APB)

Table B21: Average RMSE values for APB calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated from all available data (n = 11,721).

Model	Average RMSE	Iterative Change in Average RMSE	Change in RMSE Compared to Null
APB ~ 1 (Null)	0.280	-	-
Ex-ante Variables			
+ snap	0.278	-0.002	-0.002
+ election type	0.272	-0.006	-0.008
+ round two	0.272†	0.000†	-0.008†
+ system change	0.271	-0.001	-0.009
+ registration difference	0.271†	0.000†	-0.009†
Ex-post Variables			
+ turnout	0.271†	0.000†	-0.009†
+ turnout change	0.271†	0.000†	-0.009†
+ ENEP	0.271†	0.000†	-0.009†
+ ENEP change	0.270	-0.001	-0.010
+ margin of victory	0.270†	0.000†	-0.010†

† Including the round two, registration difference, turnout, turnout change, ENEP, and margin of victory predictor variables yielded small reductions in average RMSE which could not be displayed to three significant figures. The prevalence of these small-scale changes could again be an artefact of the relatively granular scale on which APB is measured. It may, however, simply be that measures of bias are less well predicted by the selected range of predictor variables than other forms of polling error. This contention is explored within the main body of the text.

Table B22: Average RMSE values for APB calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the strength of partisan loyalty within the electorate could be calculated (n = 9,050).

Model	Average RMSE	Iterative Change in Average RMSE	Change in RMSE Compared to Null
APB ~ 1 (Null)	0.268	-	-
Ex-ante Variables			
+ snap	0.265	-0.003	-0.003
+ election type	0.259	-0.006	-0.009
+ round two	0.259†	0.000†	-0.009†
+ system change	0.258	-0.001	-0.010
+ registration difference	0.258†	0.000†	-0.010†
+ partisanship	0.257	-0.001	-0.011
Ex-post Variables			
+ turnout	0.257†	0.000†	-0.011†
+ turnout change	0.257†	0.000†	-0.011†
+ ENEP	0.257†	0.000†	-0.011†
+ ENEP change	0.256	-0.001	-0.012
+ margin of victory	0.255	-0.001	-0.013

† Including the round two, registration difference, turnout, turnout change, and ENEP predictor variables yielded small reductions in average RMSE which could not be displayed to three significant figures. The prevalence of these small-scale changes could again be an artefact of the relatively granular scale on which APB is measured. It may, however, simply be that measures of bias are less well predicted by the selected range of predictor variables than other forms of polling error. This contention is explored within the main body of the text.

Table B23: Average RMSE values for APB calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the extent of late decision-making within the electorate could be calculated (n = 3,285).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
APB ~ 1 (Null)	0.238	-	-
Ex-ante Variables†			
+ snap	0.238‡	0.000‡	0.000‡
+ election type	0.236	-0.002	-0.002
+ system change	0.236‡	0.000‡	-0.002‡
+ registration difference	0.234	-0.002	-0.004
Ex-post Variables			
+ turnout	0.234‡	0.000‡	-0.004‡
+ turnout change	0.234‡	0.000‡	-0.004‡
+ ENEP	0.234‡	0.000‡	-0.004‡
+ ENEP change	0.230	-0.004	-0.008
+ margin of victory	0.229	-0.001	-0.009
+ late deciders	0.229‡	0.000‡	-0.009‡

† No second-round presidential elections were present in the subset, so the variable was removed from consideration.
‡ Including the snap, system change, turnout, turnout change, and late decision-making predictor variables yielded small reductions in average RMSE which could not be displayed to three significant figures. The prevalence of these small-scale changes could again be an artefact of the relatively granular scale on which APB is measured. It may, however, simply be that measures of bias are less well predicted by the selected range of predictor variables than other forms of polling error. This contention is explored within the main body of the text.

Table B24: Average RMSE values for APB calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the left/right standard deviation of the political position of parties and candidates could be calculated (n = 939).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
APB ~ 1 (Null)	0.301	-	-
Ex-ante Variables			
+ snap	0.296	-0.005	-0.005
+ election type	0.294	-0.002	-0.007
+ round two	0.294†	0.000†	-0.007†
+ registration difference	0.294†	0.000†	-0.007†
+ left/right std. dev.	0.291	-0.003	-0.010
Ex-post Variables			
+ turnout	0.291†	0.000†	-0.010†
+ turnout change	0.286	-0.005	-0.015
+ ENEP	0.279	-0.007	-0.022
+ ENEP change	0.272	-0.007	-0.029
+ margin of victory	0.271	-0.001	-0.030

† Including the round two, registration difference, and turnout predictor variables yielded small reductions in average RMSE which could not be displayed to three significant figures. The prevalence of these small-scale changes could again be an artefact of the relatively granular scale on which APB is measured. It may, however, simply be that measures of bias are less well predicted by the selected range of predictor variables than other forms of polling error. This contention is explored within the main body of the text.

Table B25: Average RMSE values for APB calculated from repeated k-fold cross validation across stepwise, additive linear regression models iteratively including all election-level variables. Values are calculated for elections for which the strength of partisan loyalty and extent of late decision-making within the electorate could be calculated, as well as the left/right standard deviation of the political position of parties and candidates (n = 293).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
APB ~ 1 (Null)	0.092	-	-
+ left/right std. dev.	0.090	-0.002	-0.002
+ partisanship	0.088	-0.002	-0.004
+ late deciders	0.088	0.000	-0.004

Additive Prediction Models for Significantly Biased Poll (SBP)

Table B26: The average proportion of correct classifications for SBP calculated from 10 iterations of down-sampled, repeated 10-fold cross validation across stepwise, additive logistic regression models. Accuracy is calculated as an average of averages from repeated k-fold cross validation across 4 balanced, down-sampled data frames of all studied elections (n = 4 × 8,138).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
SBP ~ 1 (Null)	49.99	-	-
Ex-ante Variables			
+ snap	51.88	1.89	1.89
+ election type	53.96	2.08	2.08
+ round two	53.71	-0.25	1.83
+ system change	55.19	1.48	3.31
+ registration difference	55.25	0.06	3.37
Ex-post Variables			
+ turnout	55.97	0.72	4.09
+ turnout change	55.44	-0.53	3.56
+ ENEP	59.29	3.85	7.41
+ ENEP change	59.99	0.70	8.11
+ margin of victory	59.66	-0.33	7.78

Following the process of downsampling described in chapter 5, for the model displayed in Table B26, 4,069 majority class values are retained at random in every subset from a total population of 4,317. From Monte Carlo simulations, it takes on average 3.6 random samples of 4,069 values without replacement to account for all discrete instances of the majority class the population at least once. As such, I round to the nearest appropriate integer and analyse SBP across 4 downsampled subsets and take the average of reported values to ensure that findings are as closely representative of the relationships present in the larger dataset from which the subsets were drawn. This process is applied to all following tables concerning binary measures of polling error, with the number of downsampled subsets used displayed in table captions.

Table B27: The average proportion of correct classifications for SBP calculated from repeated 10-fold cross validation across 4 balanced, down-sampled data frames of elections for which the strength of partisan loyalty amongst the electorate was known (n = 4 × 6,658).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
SBP ~ 1 (Null)	49.99	-	-
Ex-ante Variables			
+ snap	52.19	2.20	2.20
+ election type	56.33	4.14	6.34
+ round two	56.15	-0.18	6.16
+ system change	57.81	1.66	7.82
+ registration difference	58.27	0.46	8.28
+ partisanship	57.68	-0.59	7.69
Ex-post Variables			
+ turnout	58.53	0.85	8.54
+ turnout change	59.90	1.37	9.91
+ ENEP	60.41	0.51	10.42
+ ENEP change	61.93	1.52	11.94
+ margin of victory	61.32	-0.61	11.33

Table B28: The average proportion of correct classifications for SBP calculated from repeated 10-fold cross validation across 5 balanced, down-sampled data frames of elections for which the extent of late decision-making amongst the electorate was known (n = 5 × 2,482).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
SBP ~ 1 (Null)	49.96	-	-
Ex-ante Variables†			
+ snap	49.22	-0.74	-0.74
+ election type	56.19	6.97	6.23
+ registration difference	56.21	0.02	6.25
Ex-post Variables			
+ turnout	56.99	0.78	7.03
+ turnout change	64.95	7.96	14.99
+ ENEP	66.96	2.01	17.00
+ ENEP change	66.22	-0.74	16.26
+ margin of victory	66.91	0.69	16.95
+ late deciders	66.55	-0.36	16.59

† There was an insufficient number of round two presidential elections and instances of system change between elections within the subset to avoid rank deficiency, so these variables were dropped from consideration.

Table B29: The average proportion of correct classifications for SBP calculated from repeated k-fold cross validation across 5 balanced, down-sampled data frames of elections for which left/right standard deviation of the political position of parties/candidates could be calculated (n = 5 × 696).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
SBP ~ 1 (Null)	49.77	-	-
Ex-ante Variables†			
+ snap	51.31	1.54	1.54
+ election type	50.18	-1.13	0.41
+ round two	49.53	-0.65	-0.24
+ registration difference	49.56	0.03	-0.21
+ left/right std. dev.	54.27	4.71	4.50
Ex-post Variables			
+ turnout	58.41	4.14	8.64
+ turnout change	65.20	6.79	10.93
+ ENEP	69.30	4.10	15.03
+ ENEP change	69.40	0.10	15.13
+ margin of victory	68.29	-1.11	14.02

† Too few instances of system change were present in the subset to allow for its inclusion within prediction models due to issues of rank deficiency. The variable was therefore removed from consideration.

Table B30: The average proportion of correct classifications for SBP calculated from 10 iterations of repeated 10-fold cross validation across 7 balanced, down-sampled data frames of elections for which the strength of partisan loyalty and extent of late decision-making within the electorate could be calculated, as well as the left/right standard deviation of the political position of parties and candidates (n = 7 × 222).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
SBP ~ 1 (Null)	49.57	-	-
+ partisanship	65.51	15.94	15.94
+ left/right std. dev.	65.51	0.00	15.94
+ late deciders	68.65	3.14	19.08

B4: Additional Additive Prediction Models for Measures of Substantive Polling Error

Additive Prediction Models for Largest Vote Recipient Correct (LVRC)

Table B31: The average proportion of correct classifications for LVRC calculated from 10 iterations of down-sampled, repeated k-fold cross validation across stepwise, additive logistic regression models. Accuracy is calculated as an average of averages from repeated k-fold cross validation across 64 balanced, down-sampled data frames of all studied elections (n = 64 × 2,972).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
LVRC ~ 1 (Null)	49.19	-	-
Ex-ante Variables			
+ round two	50.46	1.27	1.27
+ system change	50.66	0.20	1.47
+ election type	56.66	6.00	7.47
+ snap	56.88	0.22	7.69
+ registration difference	56.77	-0.11	7.58
Ex-post Variables			
+ turnout	57.81	1.04	8.62
+ turnout change	57.94	0.13	8.75
+ ENEP	56.52	-1.42	7.33
+ ENEP change	56.69	0.17	7.50
+ margin of victory	56.60	-0.09	7.41

Table B32: The average proportion of correct classifications for LVRC calculated from 10 iterations of down-sampled, repeated k-fold cross validation across stepwise, additive logistic regression models. Accuracy is calculated as an average of averages from repeated k-fold cross validation across 69 balanced, down-sampled data frames of the elections for which the strength of partisan loyalty amongst the electorate was known (n = 69 × 2,108).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
LVRC ~ 1 (Null)	49.88	-	-
Ex-ante Variables			
+ round two	49.94	0.06	0.06
+ system change	51.00	1.06	1.12
+ election type	59.79	8.79	9.91
+ snap	60.01	0.22	10.13
+ registration difference	60.42	0.31	10.44
+ partisanship	63.14	2.72	13.16
Ex-post Variables			
+ turnout	63.99	0.85	14.01
+ turnout change	62.32	-1.67	12.34
+ ENEP	62.98	0.66	13.00
+ ENEP change	65.92	2.94	15.94
+ margin of victory	66.21	0.29	16.23

Table B33: The average proportion of correct classifications for LVRC calculated from 10 iterations of down-sampled, repeated k-fold cross validation across stepwise, additive logistic regression models. Accuracy is calculated as an average of averages from repeated k-fold cross validation across 64 balanced, down-sampled data frames of the elections for which the degree of late decision-making amongst the electorate was known (n = 64 × 744).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
LVRC ~ 1 (Null)	49.79	-	-
Ex-ante Variables†			
+ election type	58.87	9.08	9.08
+ snap	59.95	1.08	10.16
+ registration difference	60.78	0.83	10.99
Ex-post Variables			
+ turnout	63.62	2.84	13.83
+ turnout change	65.34	1.72	15.55
+ ENEP	66.33	0.99	16.54
+ ENEP change	67.64	1.31	17.85
+ margin of victory	67.89	0.25	18.10
+ late deciders	67.52	-0.37	17.73

† No instances of second-round presidential elections or electoral system change between elections were present within the subset, so these variables were removed from consideration.

Table B34: The average proportion of correct classifications for LVRC calculated from 10 iterations of down-sampled, repeated k-fold cross validation across stepwise, additive logistic regression models. Accuracy is calculated as an average of averages from repeated k-fold cross validation across 36 balanced, down-sampled data frames of the elections for which the left/right standard deviation of the political position of parties/candidates was known (n = 36 × 296).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
LVRC ~ 1 (Null)	49.38	-	-
Ex-ante Variables†			
+ election type	57.11	7.73	7.73
+ snap	58.10	0.99	8.72
+ registration difference	56.77	-1.33	7.39
+ left/right std. dev.	70.28	13.51	20.90
Ex-post Variables			
+ turnout	74.43	4.15	25.05
+ turnout change	80.29	5.86	30.91
+ ENEP	84.80	4.51	35.42
+ ENEP change	89.17	4.37	39.79
+ margin of victory	91.87	2.70	42.49

† No instances of second-round presidential elections or electoral system change between elections were present within the subset, so these variables were removed from consideration.

Table B35: The average proportion of correct classifications for LVRC calculated from 10 iterations of down-sampled, repeated k-fold cross validation across stepwise, additive logistic regression models. Accuracy is calculated as an average of averages from repeated k-fold cross validation across 21 balanced, down-sampled data frames of the elections which the strength of partisan loyalty and extent of late decision-making within the electorate could be calculated, as well as the left/right standard deviation of the political position of parties and candidates (n = 21 × 122).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
LVRC ~ 1 (Null)	49.23	-	-
+ left/right std. dev.	60.21	10.98	10.98
+ partisanship	93.60	33.39	44.37
+ late deciders	95.10	1.50	45.87

B5: Additional Interactive Prediction Models for All Error Measures

Interactive Election-level Prediction Models for Difference in Margin (DIM)

Table B36: Repeated 10-fold cross validated RMSE values for DIM calculated from interactive linear regression models using election-level variables. Models draw on all available polling data (n = 11,832).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
DIM ~ 1 (Null)	5.519	-	-
DIM ~ additive variables	5.399	-0.120	-0.120
+ turnout change × margin of victory	5.389	-0.010	-0.130
+ turnout change × ENEP	5.382	-0.007	-0.137
+ turnout change × ENEP change	5.359	-0.023	-0.160
+ ENEP × turnout	5.358	-0.001	-0.161
+ turnout × ENEP change	5.351	-0.007	-0.168

Table B37: K-fold cross validated RMSE values for DIM calculated from interactive linear prediction models. Values are calculated from the subset of data for which partisanship values were available (n = 9,115).

Model	**Average RMSE**	**Iterative Change in RMSE**	**Change in RMSE Compared to Null**
DIM ~ 1 (Null)	5.322	-	-
DIM ~ additive variables	5.186	-0.136	-0.136
+ turnout change × margin of victory	5.167	-0.019	-0.155
+ turnout change × ENEP	5.155	-0.012	-0.167
+ turnout change × ENEP change	5.128	-0.027	-0.194
+ ENEP × turnout	5.121	-0.007	-0.201
+ turnout × ENEP change	5.120	-0.001	-0.202
+ ENEP × partisanship	5.119	-0.001	-0.203
+ turnout change × partisanship	5.119	0.000	-0.203
+ turnout × partisanship	5.116	-0.003	-0.206
+ turnout change × ENEP × partisanship	5.108	-0.008	-0.214
+ turnout change × ENEP change × partisanship	5.087	-0.021	-0.235
+ ENEP × turnout × partisanship	5.078	-0.009	-0.244
+ turnout × ENEP change × partisanship	5.061	-0.017	-0.261

Table B38: K-fold cross validated RMSE values for DIM calculated from interactive linear prediction models. Values are calculated from the subset of data for which the extent of late decision-making within the electorate could be established (n = 3,285).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
DIM ~ 1 (Null)	5.905	-	-
DIM ~ additive variables	5.749	-0.156	-0.156
+ turnout change × margin of victory	5.741	-0.008	-0.164
+ turnout change × ENEP	5.723	-0.018	-0.182
+ turnout change × ENEP change	5.681	-0.042	-0.224
+ ENEP × turnout	5.629	-0.052	-0.276
+ turnout × ENEP change	5.629	0.000	-0.276
+ late deciders × ENEP	5.547	-0.082	-0.358
+ late deciders × turnout	5.508	-0.039	-0.397
+ late deciders × turnout change	5.501	-0.007	-0.404

Interactive Prediction Models for Leading Party Bias (LPB)

Table B39: K-fold cross validated RMSE values for LPB calculated from interactive linear prediction models. Values are calculated from all available data (n = 11,832).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
LPB ~ 1 (Null)	0.302	-	-
LPB ~ additive variables	0.295	-0.007	-0.007
+ turnout change × margin of victory	0.295	0.000	-0.007
+ turnout change × ENEP	0.295	0.000	-0.007
+ turnout change × ENEP change	0.294	-0.001	-0.008
+ ENEP × turnout	0.294	0.000	-0.008
+ turnout × ENEP change	0.294	0.000	-0.008

Table B40: K-fold cross validated RMSE values for LPB calculated from interactive linear prediction models. Values are calculated from the subset of data for which partisanship values were available (n = 9,115).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
LPB ~ 1 (Null)	0.283	-	-
LPB ~ additive variables	0.277	-0.006	-0.006
+ turnout change × margin of victory	0.277†	0.000†	-0.006†
+ turnout change × ENEP	0.275	-0.002	-0.008
+ turnout change × ENEP change	0.274	-0.001	-0.009
+ ENEP × turnout	0.274	0.000	-0.009
+ turnout × ENEP change	0.274†	0.000†	-0.009†
+ ENEP × partisanship	0.274†	0.000†	-0.009†
+ turnout change × partisanship	0.274	0.000	-0.009
+ turnout × partisanship	0.274	0.000	-0.009
+ turnout change × ENEP × partisanship	0.274	0.000	-0.009
+ turnout change × ENEP change × partisanship	0.274	0.000	-0.009
+ ENEP × turnout × partisanship	0.273	-0.001	-0.010
+ turnout × ENEP change × partisanship	0.273	0.000	-0.010

† Including the two-way interaction between turnout change and the margin of victory in an election yielded a small 0.0004 reduction in average RMSE. Similarly, including the two-way interaction between turnout and ENEP change resulted in a 0.0002 reduction in average RMSE, while the interaction between ENEP and partisanship yielded a 0.0001 reduction.

Table B41: K-fold cross validated RMSE values for LPB calculated from interactive linear prediction models. Values are calculated from the subset of data for which the extent of late decision-making within the electorate could be established (n = 3,285).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
LPB ~ 1 (Null)	0.309	-	-
LPB ~ additive variables	0.292	-0.017	-0.017
+ turnout change × margin of victory	0.291	-0.001	-0.018
+ turnout change × ENEP	0.286	-0.005	-0.023
+ turnout change × ENEP change	0.286	0.000	-0.023
+ ENEP × turnout	0.286	0.000	-0.023
+ turnout × ENEP change	0.286†	0.000†	-0.023†
+ late deciders × ENEP	0.284	-0.002	-0.025
+ late deciders × turnout	0.282	-0.002	-0.027
+ late deciders × turnout change	0.282	0.000	-0.027

† Including the two-way interaction between turnout and ENEP change resulted in a small 0.0001 reduction in average RMSE.

Interactive Prediction Models for Average Party Bias (APB)

Table B42: K-fold cross validated RMSE values for APB calculated from interactive linear prediction models. Values are calculated from all available data (n = 11,721).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
APB ~ 1 (Null)	0.280	-	-
APB ~ additive variables	0.270	-0.010	-0.010
+ turnout change × margin of victory	0.270	0.000	-0.010
+ turnout change × ENEP	0.270†	0.000†	-0.010†
+ turnout change × ENEP change	0.269	-0.001	-0.011
+ ENEP × turnout	0.269†	0.000†	-0.011†
+ turnout × ENEP change	0.269	0.000	-0.011

† Including the two-way interaction between turnout change and ENEP resulted in a small 0.0006 reduction in average RMSE, while including the interaction between ENEP and turnout resulted in a 0.0007 reduction.

Table B43: K-fold cross validated RMSE values for APB calculated from interactive linear prediction models. Values are calculated from the subset of data for which partisanship values were available (n = 9,050).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
APB ~ 1 (Null)	0.268	-	-
APB ~ additive variables	0.255	-0.013	-0.013
+ turnout change × margin of victory	0.254	-0.001	-0.014
+ turnout change × ENEP	0.254†	0.000†	-0.014†
+ turnout change × ENEP change	0.254†	0.000†	-0.014†
+ ENEP × turnout	0.252	-0.002	-0.016
+ turnout × ENEP change	0.252	0.000	-0.016
+ ENEP × partisanship	0.252	0.000	-0.016
+ turnout change × partisanship	0.252†	0.000†	-0.016†
+ turnout × partisanship	0.252	0.000	-0.016
+ turnout change × ENEP × partisanship	0.252	0.000	-0.016
+ turnout change × ENEP change × partisanship	0.251	-0.001	-0.017
+ ENEP × turnout × partisanship	0.251	0.000	-0.017
+ turnout × ENEP change × partisanship	0.251†	0.000†	-0.017†

† Including the two-way interaction between turnout change and ENP yielded a small reduction in average RMSE of 0.0002 Including the interactions between turnout change and ENEP change brought about a 0.0004 reduction. Including the interaction between turnout change and partisanship yielded a 0.0003 reduction. Including the three-way interaction between turnout, ENEP change, and partisanship resulted in a 0.0003 reduction.

Table B44: K-fold cross validated RMSE values for APB calculated from interactive linear prediction models. Values are calculated from the subset of data for which the extent of late decision-making within the electorate could be established (n = 3,285).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
APB ~ 1 (Null)	0.238	-	-
APB ~ additive variables	0.229	-0.009	-0.009
+ turnout change × margin of victory	0.229	0.000	-0.009
+ turnout change × ENEP	0.228	-0.001	-0.010
+ turnout change × ENEP change	0.228†	0.000†	-0.010†
+ ENEP × turnout	0.226	-0.002	-0.012
+ turnout × ENEP change	0.221	-0.005	-0.017
+ late deciders × ENEP	0.219	-0.002	-0.019
+ late deciders × turnout	0.219	0.000	-0.019
+ late deciders × turnout change	0.219†	0.000†	-0.019†

† Including the two-way interaction between turnout change and ENEP change yielded a small 0.0001 reduction in average RMSE, while including the interaction between late deciders and turnout change brought about a 0.0006 reduction.

Interactive Prediction Models for Average Bounded Inaccuracy 1 (ABI 1)

Table B45: Average RMSE values for ABI 1 calculated from interactive linear prediction models using repeated 10-fold cross validation. RMSE values represent the average performance of 100 out-of-sample predictions and are based on all available data (n = 7,403).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
ABI 1 ~ 1 (Null)	2.810	-	-
ABI 1 ~ additive variables	2.723	-0.085	-0.085
+ turnout change × margin of victory	2.705	-0.018	-0.103
+ turnout change × ENEP	2.702	-0.003	-0.106
+ turnout change × ENEP change	2.700	-0.002	-0.108
+ ENEP × turnout	2.698	-0.002	-0.110
+ turnout × ENEP change	2.698	0.000	-0.110

Table B46: Average RMSE values for ABI 1 calculated from interactive linear prediction models using repeated 10-fold cross validation. RMSE values represent the average performance of 100 out-of-sample predictions and are based on the subset of cases for which partisanship values were available (n = 5,909).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
ABI 1 ~ 1 (Null)	2.648	-	-
ABI 1 ~ additive variables	2.560	-0.088	-0.088
+ turnout change × margin of victory	2.534	-0.026	-0.114
+ turnout change × ENEP	2.522	-0.012	-0.126
+ turnout change × ENEP change	2.507	-0.015	-0.141
+ ENEP × turnout	2.505	-0.002	-0.143
+ turnout × ENEP change	2.499	-0.006	-0.149
+ ENEP × partisanship	2.499	0.000	-0.149
+ turnout change × partisanship	2.499†	0.000†	-0.149†
+ turnout × partisanship	2.496	-0.003	-0.152
+ turnout change × ENEP × partisanship	2.485	-0.011	-0.163
+ turnout change × ENEP change × partisanship	2.484	-0.001	-0.164
+ ENEP × turnout × partisanship	2.462	-0.022	-0.186
+ turnout × ENEP change × partisanship	2.439	-0.023	-0.209

† Including the two-way interaction between turnout change and partisanship yielded a small 0.0001 reduction in average RMSE.

Table B47: Average RMSE values for ABI 1 calculated from interactive linear prediction models using repeated 10-fold cross validation. RMSE values represent the average performance of 100 out-of-sample predictions and are based the subset of cases for which the extent of late decision-making within the electorate could be established (n = 2,488).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
ABI 1 ~ 1 (Null)	2.340	-	-
ABI 1 ~ additive variables	2.256	-0.084	-0.084
+ turnout change × margin of victory	2.229	-0.021	-0.105
+ turnout change × ENEP	2.230	0.001	-0.104
+ turnout change × ENEP change	2.230	0.000	-0.104
+ ENEP × turnout	2.230†	0.000†	-0.104†
+ turnout × ENEP change	2.231	0.001	-0.103
+ late deciders × ENEP	2.210	-0.021	-0.124
+ late deciders × turnout	2.209	-0.001	-0.125
+ late deciders × turnout change	2.209	0.000	-0.125

† Including the two-way interaction between ENP and turnout yielded a small 0.0005 reduction in average RMSE.

Interactive Prediction Models for Average Bounded Inaccuracy 2 (ABI 2)

Table B48: Average RMSE values for ABI 2 calculated from interactive linear prediction models using repeated 10-fold cross validation. RMSE values represent the average performance of 100 out-of-sample predictions and are based on all available data (n = 8,754).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
ABI 2 ~ 1 (Null)	2.270	-	-
ABI 2 ~ additive variables	2.167	-0.103	-0.103
+ turnout change × margin of victory	2.156	-0.011	-0.114
+ turnout change × ENEP	2.156†	0.000†	-0.114†
+ turnout change × ENEP change	2.156	0.000	-0.114
+ ENEP × turnout	2.156	0.000	-0.114
+ turnout × ENEP change	2.156	0.000	-0.114

† Including the two-way interaction between turnout change and ENEP yielded a small reduction in average RMSE of 0.0004.

Table B49: Average RMSE values for ABI 2 calculated from interactive linear prediction models using repeated 10-fold cross validation. RMSE values represent the average performance of 100 out-of-sample predictions and are based on the subset of cases for which partisanship values were available (n = 6,950).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
ABI 2 ~ 1 (Null)	2.121	-	-
ABI 2 ~ additive variables	2.010	-0.111	-0.111
+ turnout change × margin of victory	1.984	-0.026	-0.137
+ turnout change × ENEP	1.979	-0.005	-0.142
+ turnout change × ENEP change	1.974	-0.005	-0.147
+ ENEP × turnout	1.974†	0.000†	-0.147†
+ turnout × ENEP change	1.970	-0.004	-0.151
+ ENEP × partisanship	1.969	-0.001	-0.152
+ turnout change × partisanship	1.969	0.000	-0.152
+ turnout × partisanship	1.965	-0.004	-0.156
+ turnout change × ENEP × partisanship	1.956	-0.009	-0.165
+ turnout change × ENEP change × partisanship	1.956	0.000	-0.165
+ ENEP × turnout × partisanship	1.942	-0.014	-0.179
+ turnout × ENEP change × partisanship	1.919	-0.023	-0.202

† Including the two-way interaction between turnout and ENEP change resulted in a small 0.0004 reduction in average RMSE.

Table B50: Average RMSE values for ABI 2 calculated from interactive linear prediction models using repeated 10-fold cross validation. RMSE values represent the average performance of 100 out-of-sample predictions and are based on the subset of cases for which the extent of late decision-making within the electorate could be established (n = 2,658).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
ABI 2 ~ 1 (Null)	1.841	-	-
ABI 2 ~ additive variables	1.749	-0.092	-0.092
+ turnout change × margin of victory	1.716	-0.033	-0.125
+ turnout change × ENEP	1.713	-0.003	-0.128
+ turnout change × ENEP change	1.711	-0.002	-0.130
+ ENEP × turnout	1.702	-0.009	-0.139
+ turnout × ENEP change	1.702†	0.000†	-0.139†
+ late deciders × ENEP	1.671	-0.031	-0.170
+ late deciders × turnout	1.672	0.001	-0.169
+ late deciders × turnout change	1.672	0.000	-0.169

† Including the two-way interaction between turnout and ENP difference yielded a small 0.0004 reduction in average RMSE.

Interactive Classification Models for Significantly Biased Poll (SBP)

Table B51: The average proportion of correct classifications for SBP calculated from 10 iterations of down-sampled, repeated k-fold cross validation across stepwise, additive logistic regression models. Accuracy is calculated as an average of averages from repeated k-fold cross validation across 4 balanced, down-sampled data frames of all studied elections (n = 4 × 8,138).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
SBP ~ 1 (Null)	49.99	-	-
SBP ~ additive variables	59.66	7.78	7.78
+ turnout change × margin of victory	60.28	0.62	8.40
+ turnout change × ENEP	60.41	0.13	8.53
+ turnout change × ENEP change	60.83	0.42	8.95
+ ENEP × turnout	59.63	-1.20	7.75
+ turnout × ENEP change	59.68	0.05	7.80

Table B52: The average proportion of correct classifications for SBP calculated from 10 iterations of down-sampled, repeated k-fold cross validation across stepwise, additive logistic regression models. Accuracy is calculated as an average of averages from repeated k-fold cross validation across 4 balanced, down-sampled data frames of all elections for which the strength of partisan loyalty amongst the electorate was known (n = 4 × 6,658).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
SBP ~ 1 (Null)	49.99	-	-
SBP ~ additive variables	61.32	11.33	11.33
+ turnout change × margin of victory	61.36	0.04	11.37
+ turnout change × ENEP	61.47	0.11	11.48
+ turnout change × ENEP change	61.70	0.23	11.71
+ ENEP × turnout	61.83	0.13	11.84
+ turnout × ENEP change	61.65	-0.18	11.66
+ ENEP × partisanship	61.68	0.03	11.69
+ turnout change × partisanship	61.86	0.18	11.87
+ turnout × partisanship	61.95	0.09	11.96
+ turnout change × ENEP × partisanship	62.44	0.49	12.45
+ turnout change × ENEP change × partisanship	63.77	1.33	13.78
+ ENEP × turnout × partisanship	65.40	1.63	15.41
+ turnout × ENEP change × partisanship	65.25	-0.15	15.26

Table B53: The average proportion of correct classifications for SBP calculated from 10 iterations of down-sampled, repeated k-fold cross validation across stepwise, additive logistic regression models. Accuracy is calculated as an average of averages from repeated k-fold cross validation across 5 balanced, down-sampled data frames of all elections in which the extent of late decision-making within the electorate was known (n = 5 × 2,482).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
SBP ~ 1 (Null)	49.96	-	-
SBP ~ additive variables	66.55	16.59	16.59
+ turnout change × margin of victory	66.85	0.30	16.89
+ turnout change × ENEP	67.95	1.10	17.99
+ turnout change × ENEP change	67.90	-0.05	17.94
+ ENEP × turnout	69.41	1.51	19.45
+ turnout × ENEP change	69.42	0.01	19.46
+ late deciders × ENEP	69.81	0.39	19.85
+ late deciders × turnout	70.50	0.69	20.54
+ late deciders × turnout change	70.59	0.09	20.63

Interactive Classification Models for Largest Vote Recipient Correct (LVRC)

Table B54: Average proportion of correct classifications for LVRC calculated from interactive logistic classification models. Accuracy is calculated as an average of averages from repeated k-fold cross validation across 64 balanced, down-sampled data frames of all studied elections (n = 64 × 2,972).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
LVRC ~ 1 (Null)	49.19	-	-
LVRC ~ additive variables	56.60	7.41	7.41
+ turnout change × margin of victory	56.84	0.24	7.65
+ turnout change × ENEP	56.18	-0.66	6.99
+ turnout change × ENEP change	56.65	0.47	7.46
+ ENEP × turnout	57.43	0.78	8.24
+ turnout × ENEP change	57.17	-0.26	7.98

Table B55: Average proportion of correct classifications for LVRC calculated from interactive classification models based on multiple logistic regression. Accuracy is calculated as an average of averages from 10 iterations of repeated 10-fold cross validation across 69 balanced, down-sampled data frames of those elections for which the strength of partisan loyalty amongst the electorate was known (n = 69 × 2,108).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
LVRC ~ 1 (Null)	49.88	-	-
LVRC ~ additive variables	64.11	14.13	14.13
+ turnout change × margin of victory	65.97	1.86	15.99
+ turnout change × ENEP	66.65	0.68	16.67
+ turnout change × ENEP change	66.89	0.24	16.91
+ ENEP × turnout	67.42	0.53	17.44
+ turnout × ENEP change	68.68	1.26	18.70
+ ENEP × partisanship	68.71	0.03	18.73
+ turnout change × partisanship	69.37	0.66	19.39
+ turnout × partisanship	70.48	1.11	20.50
+ turnout change × ENEP × partisanship	70.79	0.31	20.81
+ turnout change × ENEP change × partisanship	70.87	0.08	20.89
+ ENEP × turnout × partisanship	71.19	0.32	21.21
+ turnout × ENEP change × partisanship	71.23	0.04	21.25

Table B56: Average proportion of correct classifications for LVRC calculated from interactive classification models based on multiple logistic regression. Accuracy is calculated as an average of averages from 10 repetitions of repeated 10-fold cross validation across 64 balanced, down-sampled data frames of those elections for which the extent of late decision-making within the electorate was known (n = 64 × 744).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
LVRC ~ 1 (Null)	49.79	-	-
LVRC ~ additive variables	67.52	17.73	17.73
+ turnout change × margin of victory	67.36	-0.16	17.57
+ turnout change × ENEP	67.61	0.25	17.82
+ turnout change × ENEP change	67.64	0.03	17.85
+ ENEP × turnout	74.89	7.25	25.10
+ turnout × ENEP change	75.63	0.74	25.84
+ late deciders × ENEP	77.72	2.09	27.93
+ late deciders × turnout	77.78	0.06	27.99
+ late deciders × turnout change	79.90	2.12	30.11

B6: The Robustness of Additional Interactive Prediction Models to Controls

In this section, I provide the outputs from prediction models containing control variables across my additional measures of distributive, bounded, and substantive polling error, as well as polling bias. Overall, the outputs demonstrate that election-level variables improve the ability of models to predict all measures of polling error, even in the presence of controls. They also demonstrate the predictive utility of variables from the poll, pollster, and country grouping levels, underscoring the benefit of adopting a four-level approach to decomposing polling error.

The Robustness of DIM Findings to Controls

Table B57: K-fold Cross validated RMSE values for DIM calculated from the subset of data for which all control variables were available (n = 5,432).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
DIM ~ 1 (Null)	4.856	-	-
+ poll-level controls	4.850	-0.006	-0.006
+ pollster-level controls	4.836	-0.014	-0.020
+ country-level controls	4.410	-0.426	-0.466
+ additive election-level variables	4.286	-0.124	-0.590
+ turnout change × margin of victory	4.274	-0.012	-0.602
+ turnout change × ENEP	4.269	-0.005	-0.607
+ turnout change × ENEP change	4.267	-0.002	-0.609
+ ENEP × turnout	4.269	0.002	-0.607
+ turnout × ENEP change	4.248	-0.021	-0.628

Table B58: K-fold cross validated RMSE values for DIM calculated the subset of data for which all control variables were available alongside partisanship figures (n = 4,384).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
DIM ~ 1 (Null)	4.498	-	-
+ poll-level controls	4.494	-0.004	-0.004
+ pollster-level controls	4.475	-0.019	-0.023
+ country-level controls	4.138	-0.337	-0.360
+ additive election-level variables	3.874	-0.264	-0.624
+ turnout change × margin of victory	3.872	-0.002	-0.626
+ turnout change × ENEP	3.870	-0.002	-0.628
+ turnout change × ENEP change	3.861	-0.009	-0.637
+ ENEP × turnout	3.862	0.001	-0.636
+ turnout × ENEP change	3.827	-0.035	-0.671
+ ENEP × partisanship	3.828	0.001	-0.670
+ turnout change × partisanship	3.829	0.001	-0.669
+ turnout × partisanship	3.823	-0.006	-0.675
+ turnout change × ENEP × partisanship	3.803	-0.020	-0.695
+ turnout change × ENEP change × partisanship	3.804	0.001	-0.694
+ ENEP × turnout × partisanship	3.805	0.001	-0.693
+ turnout × ENEP change × partisanship	3.807	0.002	-0.691

Table B59: K-fold cross validated RMSE values for DIM that result from the inclusion of election-level variables both additively and interactively across the subset of data for which all control variables were available alongside late decision-making figures (n = 1,557).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
DIM ~ 1 (Null)	3.792	-	-
+ poll-level controls	3.783	-0.009	-0.009
+ pollster-level controls	3.706	-0.077	-0.086
+ country-level controls	3.623	-0.083	-0.169
+ additive election-level variables†	3.589	-0.034	-0.203
+ turnout change × margin of victory	3.589	0.000	-0.203
+ turnout change × ENEP	3.589	0.000	-0.203
+ turnout change × ENEP change	3.589	0.000	-0.203
+ ENEP × turnout	3.589	0.000	-0.203
+ late deciders × ENEP	3.589	0.000	-0.203
+ late deciders × turnout	3.589	0.000	-0.203
+ late deciders × turnout change	3.589	0.000	-0.203

† There was an insufficient number of presidential elections within the subset of data for the election type and round two variables to be included within the model, as their scarcity resulted in rank deficient predictions

Table B60: K-fold cross validated RMSE values for DIM that result from the inclusion of election-level interactions across the subset of data for which all control variables were available alongside late decision-making figures (n = 1,557). The model is run across a reduced feature set to explore prospective overfitting in the more complex specification.

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
DIM ~ 1 (Null)	3.792	-	-
+ poll-level controls	3.783	-0.009	-0.009
+ pollster-level controls	3.706	-0.077	-0.086
+ country-level controls	3.623	-0.083	-0.169
+ turnout change × margin of victory	3.334	-0.289	-0.458
+ turnout change × ENEP	3.342	0.008	-0.450
+ turnout change × ENEP change	3.332	-0.010	-0.460
+ ENEP × turnout	3.244	-0.088	-0.548
+ late deciders × ENEP	3.239	-0.005	-0.553
+ late deciders × turnout	3.248	0.009	-0.544
+ late deciders × turnout change	3.247	-0.001	-0.545

† There was an insufficient number of presidential elections within the subset of data for the election type and round two variables to be included within the model, as their scarcity resulted in rank deficient predictions

The Robustness of LPB Findings to Controls

Table B61: K-fold cross validated RMSE values for LPB calculated from the subset of data for which all control variables were available (n = 5,432).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
LPB ~ 1 (Null)	0.272	-	-
+ poll-level controls	0.271	-0.001	-0.001
+ pollster-level controls	0.265	-0.006	-0.007
+ country-level controls	0.253	-0.012	-0.019
+ additive election-level variables	0.243	-0.010	-0.029
+ turnout change × margin of victory	0.243	0.000	-0.029
+ turnout change × ENEP	0.243	0.000	-0.029
+ turnout change × ENEP change	0.243†	0.000†	-0.029†
+ ENEP × turnout	0.243	0.000	-0.029
+ turnout × ENEP change	0.243	0.000	-0.029

† Including the two-way interaction between turnout change and ENEP change resulted in a small 0.0002 reduction in average RMSE.

Table B62: K-fold cross validated RMSE values for LPB calculated the subset of data for which all control variables were available alongside partisanship figures (n = 4,384).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
LPB ~ 1 (Null)	0.256	-	-
+ poll-level controls	0.255	-0.001	-0.001
+ pollster-level controls	0.246	-0.009	-0.010
+ country-level controls†	0.234	-0.012	-0.022
+ additive election-level variables	0.221	-0.013	-0.035
+ turnout change × margin of victory	0.221	0.000	-0.035
+ turnout change × ENEP	0.220	-0.001	-0.036
+ turnout change × ENEP change	0.220	0.000	-0.036
+ ENEP × turnout	0.220‡	0.000‡	-0.036‡
+ turnout × ENEP change	0.220	0.000	-0.036
+ ENEP × partisanship	0.219	-0.001	-0.037
+ turnout change × partisanship	0.219	0.000	-0.037
+ turnout × partisanship	0.218	-0.001	-0.038
+ turnout change × ENEP × partisanship	0.218‡	0.000‡	-0.038‡
+ turnout change × ENEP change × partisanship	0.217	-0.001	-0.039
+ ENEP × turnout × partisanship	0.217‡	0.000‡	-0.039‡
+ turnout × ENEP change × partisanship	0.217	0.000	-0.039

† Given the composition of the subset of data containing partisanship measurements, it was not possible to include the electoral system variable within country-level controls, as even reducing it to a binary variable failed to resolve issues of rank deficiency.
‡ Including the interaction between ENEP and turnout yielded a 0.0005 reduction in RMSE, while including the interaction between turnout change, ENEP, and partisanship yielded a 0.0004 reduction, and the interaction between ENEP, turnout, and partisanship yielded a reduction of 0.0004.

Table B63: K-fold cross validated RMSE values for LPB that result from the inclusion of election-level variables both additively and interactively across the subset of data for which all control variables were available alongside late decision-making figures (n = 1,557).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
LPB ~ 1 (Null)	0.258	-	-
+ poll-level controls	0.256	-0.002	-0.002
+ pollster-level controls	0.227	-0.029	-0.031
+ country-level controls†	0.205	-0.022	-0.053
+ additive election-level variables‡	0.200	-0.005	-0.058
+ turnout change × margin of victory	0.200	0.000	-0.058
+ turnout change × ENEP	0.200	0.000	-0.058
+ turnout change × ENEP change	0.200	0.000	-0.058
+ ENEP × turnout	0.200	0.000	-0.058
+ late deciders × ENEP	0.200	0.000	-0.058
+ late deciders × turnout	0.200	0.000	-0.058
+ late deciders × turnout change	0.200	0.000	-0.058

† Given the composition of the subset of data containing late decision-making measurements, it was not possible to include the electoral system or polling moratorium variables within country-level controls, as even reducing them to binary variables failed to resolve issues of rank deficiency.
‡ There was an insufficient number of presidential elections within the subset of data for the election type and round two variables to be included within the model, as their scarcity resulted in rank deficient predictions due to

Table B64: K-fold cross validated RMSE values for LPB that result from the inclusion of election-level interactions across the subset of data for which all control variables were available alongside late decision-making figures (n = 1,557). The model is run across a reduced feature set to explore prospective overfitting in the more complex specification.

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
LPB ~ 1 (Null)	0.258	-	-
+ poll-level controls	0.256	-0.002	-0.002
+ pollster-level controls	0.227	-0.029	-0.031
+ country-level controls†	0.205	-0.022	-0.053
+ turnout change × margin of victory	0.205	0.000	-0.053
+ turnout change × ENEP	0.200	-0.005	-0.058
+ turnout change × ENEP change	0.199	-0.001	-0.059
+ ENEP × turnout	0.199	0.000	-0.059
+ late deciders × ENEP	0.199	0.000	-0.059
+ late deciders × turnout	0.200	0.001	-0.058
+ late deciders × turnout change	0.199	-0.001	-0.059

† Given the composition of the subset of data containing late decision-making measurements, it was not possible to include the electoral system or polling moratorium variables within country-level controls, as even reducing them to binary variables failed to resolve issues of rank deficiency.

‡ There was an insufficient number of presidential elections within the subset of data for the election type and round two variables to be included within the model, as their scarcity resulted in rank deficient predictions due to

The Robustness of APB Findings to Controls

Table B65: K-fold Cross validated RMSE values for APB calculated from the subset of data for which all control variables were available (n = 5,432).

Model	**Average RMSE**	**Iterative Change in RMSE**	**Change in RMSE Compared to Null**
APB ~ 1 (Null)	0.277	-	-
+ poll-level controls	0.266	-0.011	-0.011
+ pollster-level controls	0.264	-0.002	-0.013
+ country-level controls	0.235	-0.029	-0.042
+ additive election-level variables	0.230	-0.005	-0.047
+ turnout change × margin of victory	0.230	0.000	-0.047
+ turnout change × ENEP	0.227	-0.003	-0.050
+ turnout change × ENEP change	0.226	-0.001	-0.051
+ ENEP × turnout	0.226	0.000	-0.051
+ turnout × ENEP change	0.226	0.000	-0.051

Table B66: K-fold cross validated RMSE values for APB calculated the subset of data for which all control variables were available alongside partisanship figures (n = 4,384).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
APB ~ 1 (Null)	0.277	-	-
+ poll-level controls	0.261	-0.016	-0.016
+ pollster-level controls	0.259	-0.002	-0.018
+ country-level controls†	0.222	-0.037	-0.055
+ additive election-level variables	0.205	-0.017	-0.072
+ turnout change × margin of victory	0.204	-0.001	-0.073
+ turnout change × ENEP	0.203	-0.001	-0.074
+ turnout change × ENEP change	0.202	-0.001	-0.075
+ ENEP × turnout	0.202	0.000	-0.075
+ turnout × ENEP change	0.200	-0.002	-0.077
+ ENEP × partisanship	0.200	0.000	-0.077
+ turnout change × partisanship	0.200	0.000	-0.077
+ turnout × partisanship	0.200	0.000	-0.077
+ turnout change × ENEP × partisanship	0.200	0.000	-0.077
+ turnout change × ENEP change × partisanship	0.200	0.000	-0.077
+ ENEP × turnout × partisanship	0.199	-0.001	-0.078
+ turnout × ENEP change × partisanship	0.199	0.000	-0.078

† Given the composition of the subset of data containing partisanship measurements, it was not possible to include the electoral system variable within country-level controls, as even reducing it to a binary variable failed to resolve issues of rank deficiency.

Table B67: K-fold cross validated RMSE values for APB that result from the inclusion of election-level variables both additively and interactively across the subset of data for which all control variables were available alongside late decision-making figures (n = 1,557).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
APB ~ 1 (Null)	0.263	-	-
+ poll-level controls	0.244	-0.019	-0.019
+ pollster-level controls	0.222	-0.022	-0.041
+ country-level controls†	0.192	-0.030	-0.071
+ additive election-level variables‡	0.182	-0.010	-0.081
+ turnout change × margin of victory	0.182	0.000	-0.081
+ turnout change × ENEP	0.182	0.000	-0.081
+ turnout change × ENEP change	0.182	0.000	-0.081
+ ENEP × turnout	0.182	0.000	-0.081
+ late deciders × ENEP	0.182	0.000	-0.081
+ late deciders × turnout	0.182	0.000	-0.081
+ late deciders × turnout change	0.182	0.000	-0.081

† Given the composition of the subset of data containing late decision-making measurements, it was not possible to include the electoral system or polling moratorium variables within country-level controls, as even reducing it to a binary variable failed to resolve issues of rank deficiency.
‡ There was an insufficient number of presidential elections within the subset of data for the election type and round two variables to be included within the model, as their scarcity resulted in rank deficient predictions due to

Table B68: K-fold cross validated RMSE values for APB that result from the inclusion of election-level interactions across the subset of data for which all control variables were available alongside late decision-making figures (n = 1,557). The model is run across a reduced feature set to explore prospective overfitting in the more complex specification.

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
APB ~ 1 (Null)	0.263	-	-
+ poll-level controls	0.244	-0.019	-0.019
+ pollster-level controls	0.222	-0.022	-0.041
+ country-level controls†	0.192	-0.030	-0.071
+ turnout change × margin of victory	0.186	-0.006	-0.077
+ turnout change × ENEP	0.185	-0.001	-0.078
+ turnout change × ENEP change	0.185	0.000	-0.078
+ ENEP × turnout	0.173	-0.012	-0.090
+ late deciders × ENEP	0.173	0.000	-0.090
+ late deciders × turnout	0.182	0.005	-0.085
+ late deciders × turnout change	0.180	-0.002	-0.087

† Given the composition of the subset of data containing late decision-making measurements, it was not possible to include the electoral system or polling moratorium variables within country-level controls, as even reducing it to a binary variable failed to resolve issues of rank deficiency.
‡ There was an insufficient number of presidential elections within the subset of data for the election type and round two variables to be included within the model, as their scarcity resulted in rank deficient predictions due to

The Robustness of ABI 1 Findings to Controls

Table B69: K-fold Cross validated RMSE values for ABI 1 calculated from the subset of data for which all control variables were available (n = 4,706).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
ABI 1 ~ 1 (Null)	2.708	-	-
+ poll-level controls	2.672	-0.036	-0.036
+ pollster-level controls	2.662	-0.010	-0.046
+ country-level controls	2.249	-0.413	-0.459
+ additive election-level variables	2.227	-0.022	-0.481
+ turnout change × margin of victory	2.222	-0.005	-0.486
+ turnout change × ENEP	2.212	-0.010	-0.496
+ turnout change × ENEP change	2.208	-0.004	-0.500
+ ENEP × turnout	2.208	0.000	-0.500
+ turnout × ENEP change	2.208	0.000	-0.500

Table B70: K-fold cross validated RMSE values for ABI 1 calculated the subset of data for which all control variables were available alongside partisanship figures (n = 3,787).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
ABI 1 ~ 1 (Null)	2.386	-	-
+ poll-level controls	2.326	-0.060	-0.060
+ pollster-level controls	2.311	-0.015	-0.075
+ country-level controls†	1.964	-0.347	-0.422
+ additive election-level variables‡	1.911	-0.053	-0.475
+ turnout change × margin of victory	1.901	-0.010	-0.485
+ turnout change × ENEP	1.902	0.001	-0.484
+ turnout change × ENEP change	1.903	0.001	-0.483
+ ENEP × turnout	1.901	-0.002	-0.485
+ turnout × ENEP change	1.888	-0.013	-0.498
+ ENEP × partisanship	1.889	0.001	-0.497
+ turnout change × partisanship	1.891	0.002	-0.495
+ turnout × partisanship	1.891	0.000	-0.495
+ turnout change × ENEP × partisanship	1.889	-0.002	-0.497
+ turnout change × ENEP change × partisanship	1.890	0.001	-0.496
+ ENEP × turnout × partisanship	1.890	0.000	-0.496
+ turnout × ENEP change × partisanship	1.890	0.000	-0.496

† Given the composition of the subset of data containing partisanship measurements, it was not possible to include the electoral system variable within country-level controls, as even reducing it to a binary variable failed to resolve issues of rank deficiency. Similarly, too few instances of compulsory voting were present to allow for its inclusion. The scarcity of these values resulted in rank deficient predictions is due to the presence of constants in one or more of the train/test splits generated by the repeated k-fold cross validation process.

Table B71: K-fold cross validated RMSE values for ABI 1 that result from the inclusion of election-level variables both additively and interactively across the subset of data for which all control variables were available alongside late decision-making figures (n = 1,492).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
ABI 1 ~ 1 (Null)	1.779	-	-
+ poll-level controls	1.523	-0.256	-0.256
+ pollster-level controls	1.511	-0.012	-0.268
+ country-level controls†	1.488	-0.023	-0.291
+ additive election-level variables‡	1.474	-0.014	-0.305
+ turnout change × margin of victory	1.473	-0.001	-0.306
+ turnout change × ENEP	1.473	0.000	-0.306
+ turnout change × ENEP change	1.473	0.000	-0.306
+ ENEP × turnout	1.473	0.000	-0.306
+ late deciders × ENEP	1.473	0.000	-0.306
+ late deciders × turnout	1.473	0.000	-0.306
+ late deciders × turnout change	1.473	0.000	-0.306

† Given the composition of the subset of data containing late decision-making measurements, it was not possible to include the electoral system or polling moratorium variables within country-level controls, as even reducing it to a binary variable failed to resolve issues of rank deficiency.
‡ There was an insufficient number of presidential elections within the subset of data for the election type and round two variables to be included within the model, as their scarcity resulted in rank deficient predictions.

Table B72: K-fold cross validated RMSE values for ABI 1 that result from the inclusion of election-level interactions across the subset of data for which all control variables were available alongside late decision-making figures (n = 1,492). The model is run across a reduced feature set to explore prospective overfitting in the more complex specification.

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
ABI 1 ~ 1 (Null)	1.779	-	-
+ poll-level controls	1.523	-0.256	-0.256
+ pollster-level controls	1.511	-0.012	-0.268
+ country-level controls†	1.488	-0.023	-0.291
+ turnout change × margin of victory	1.488	0.000	-0.291
+ turnout change × ENEP	1.491	0.003	-0.288
+ turnout change × ENEP change	1.479	-0.012	-0.300
+ ENEP × turnout	1.448	-0.031	-0.331
+ late deciders × ENEP	1.435	-0.013	-0.344
+ late deciders × turnout	1.427	-0.008	-0.352
+ late deciders × turnout change	1.427	0.000	-0.352

† Given the composition of the subset of data containing late decision-making measurements, it was not possible to include the electoral system or polling moratorium variables within country-level controls, as even reducing it to a binary variable failed to resolve issues of rank deficiency.
‡ There was an insufficient number of presidential elections within the subset of data for the election type and round two variables to be included within the model, as their scarcity resulted in rank deficient predictions.

The Robustness of ABI 2 Findings to Controls

Table B73: K-fold cross validated RMSE values for ABI 2 calculated from the subset of data for which all control variables were available (n = 5,430).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
ABI 2 ~ 1 (Null)	2.158	-	-
+ poll-level controls	2.139	-0.019	-0.019
+ pollster-level controls	2.136	-0.003	-0.022
+ country-level controls	1.731	-0.405	-0.427
+ additive election-level variables	1.706	-0.025	-0.452
+ turnout change × margin of victory	1.701	-0.005	-0.457
+ turnout change × ENEP	1.687	-0.014	-0.471
+ turnout change × ENEP change	1.671	-0.016	-0.487
+ ENEP × turnout	1.672	0.001	-0.486
+ turnout × ENEP change	1.673	0.001	-0.485

Table B74: K-fold cross validated RMSE values for ABI 2 calculated the subset of data for which all control variables were available alongside partisanship figures (n = 4,384).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
ABI 2 ~ 1 (Null)	1.855	-	-
+ poll-level controls	1.827	-0.028	-0.028
+ pollster-level controls	1.821	-0.006	-0.034
+ country-level controls†	1.429	-0.392	-0.426
+ additive election-level variables	1.388	-0.041	-0.467
+ turnout change × margin of victory	1.372	-0.016	-0.483
+ turnout change × ENEP	1.371	-0.001	-0.484
+ turnout change × ENEP change	1.371	0.000	-0.484
+ ENEP × turnout	1.370	-0.001	-0.485
+ turnout × ENEP change	1.359	-0.011	-0.496
+ ENEP × partisanship	1.359	0.000	-0.496
+ turnout change × partisanship	1.360	0.001	-0.495
+ turnout × partisanship	1.360	0.000	-0.495
+ turnout change × ENEP × partisanship	1.358	-0.002	-0.497
+ turnout change × ENEP change × partisanship	1.359	0.001	-0.496
+ ENEP × turnout × partisanship	1.358	-0.001	-0.497
+ turnout × ENEP change × partisanship	1.353	-0.005	-0.502

† Given the composition of the subset of data containing partisanship measurements, it was not possible to include the electoral system variable within country-level controls, as even reducing it to a binary variable failed to resolve issues of rank deficiency. Similarly, too few instances of compulsory voting were present to allow for its inclusion. The scarcity of these values resulted in rank deficient predictions due to the presence of constants in one or more of the train/test splits generated by the k-fold cross validation process.

Table B75: K-fold cross validated RMSE values for ABI 2 that result from the inclusion of election-level variables both additively and interactively across the subset of data for which all control variables were available alongside late decision-making figures (n = 1,557).

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
ABI 2 ~ 1 (Null)	1.078	-	-
+ poll-level controls	0.879	-0.199	-0.199
+ pollster-level controls	0.878	-0.001	-0.200
+ country-level controls†	0.835	-0.043	-0.243
+ additive election-level variables‡	0.805	-0.030	-0.273
+ turnout change × margin of victory	0.805	0.000	-0.273
+ turnout change × ENEP	0.805	0.000	-0.273
+ turnout change × ENEP change	0.805	0.000	-0.273
+ ENEP × turnout	0.805	0.000	-0.273
+ late deciders × ENEP	0.805	0.000	-0.273
+ late deciders × turnout	0.805	0.000	-0.273
+ late deciders × turnout change	0.805	0.000	-0.273

† Given the composition of the subset of data containing late decision-making measurements, it was not possible to include the electoral system or polling moratorium variables within country-level controls, as they brought about issues of rank deficiency.

‡ There was an insufficient number of presidential elections within the subset of data for the election type and round two variables to be included within the model, as their scarcity resulted in rank deficient predictions.

Table B76: K-fold cross validated RMSE values for ABI 2 that result from the inclusion of election-level interactions across the subset of data for which all control variables were available alongside late decision-making figures (n = 1,557). The model is run across a reduced feature set to explore prospective overfitting in the more complex specification.

Model	Average RMSE	Iterative Change in RMSE	Change in RMSE Compared to Null
ABI 2 ~ 1 (Null)	1.078	-	-
+ poll-level controls	0.879	-0.199	-0.199
+ pollster-level controls	0.878	-0.001	-0.200
+ country-level controls†	0.835	-0.043	-0.243
+ turnout change × margin of victory	0.835	0.000	-0.243
+ turnout change × ENEP	0.836	0.001	-0.242
+ turnout change × ENEP change	0.821	-0.015	-0.257
+ ENEP × turnout	0.817	-0.004	-0.261
+ late deciders × ENEP	0.803	-0.014	-0.275
+ late deciders × turnout	0.805	0.002	-0.273
+ late deciders × turnout change	0.805	0.000	-0.273

† Given the composition of the subset of data containing late decision-making measurements, it was not possible to include the electoral system or polling moratorium variables within country-level controls, as they brought about issues of rank deficiency.

‡ There was an insufficient number of presidential elections within the subset of data for the election type and round two variables to be included within the model, as their scarcity resulted in rank deficient predictions

The Robustness of SBP Findings to Controls

Table B77: Average proportion of correct classifications for SBP calculated from the subset of data for which all control variables were available. Accuracy is calculated as an average of averages from repeated k-fold cross validation across 6 balanced, down-sampled data frames of those elections for which all control variables were available (n = 6 × 4,956).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
SBP ~ 1 (Null)	49.97	-	-
+ poll-level controls	59.25	9.28	9.28
+ pollster-level controls	62.46	3.21	12.49
+ country-level controls	69.86	7.40	19.89
+ additive election-level variables	70.69	0.83	20.72
+ turnout change × margin of victory	70.86	0.17	20.89
+ turnout change × ENEP	70.65	-0.09	20.80
+ turnout change × ENEP change	70.61	-0.04	20.76
+ ENEP × turnout	70.57	-0.04	20.72
+ turnout × ENEP change	70.87	0.30	30.02

Table B78: Average proportion of correct classifications for SBP calculated from the subset of data for which all control variables were available. Accuracy is calculated as an average of averages from repeated k-fold cross validation across 5 balanced, down-sampled data frames of those elections for which all control variables were available alongside measurements of partisan sentiment within the electorate (n = 5 × 4,018).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
SBP ~ 1 (Null)	49.98	-	-
+ poll-level controls	59.14	9.16	9.16
+ pollster-level controls	63.37	4.23	13.39
+ country-level controls†	70.45	7.08	20.47
+ additive election-level variables	73.87	3.42	23.89
+ turnout change × margin of victory	73.92	0.05	23.94
+ turnout change × ENEP	73.93	0.01	23.95
+ turnout change × ENEP change	74.01	0.08	24.03
+ ENEP × turnout	74.11	0.10	24.13
+ turnout × ENEP change	74.31	0.20	24.33
+ ENEP × partisanship	74.31	0.00	24.33
+ turnout change × partisanship	74.32	0.01	24.34
+ turnout × partisanship	74.16	-0.16	24.18
+ turnout change × ENEP × partisanship	74.16	0.00	24.18
+ turnout change × ENEP change × partisanship	74.08	-0.08	24.10
+ ENEP × turnout × partisanship	74.18	0.10	24.20
+ turnout × ENEP change × partisanship	74.21	0.03	24.23

† Given the composition of the subset of data containing partisanship measurements, it was not possible to include the electoral system variable within country-level controls, as even reducing it to a binary variable failed to resolve issues of rank deficiency. Similarly, too few instances of compulsory voting were present to allow for its inclusion. The scarcity of these values resulted in rank deficient predictions due to the presence of constants in one or more of the train/test splits generated by the k-fold cross validation process.

Table B79: Average proportion of correct classifications for SBP that result from the inclusion of election-level variables both additively and interactively across the subset of data for which all control variables were available alongside late decision-making figures. Accuracy is calculated as an average of averages from repeated k-fold cross validation across 4 balanced, down-sampled subsets of elections (n = 4 × 1,474).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
SBP ~ 1 (Null)	49.86	-	-
+ poll-level controls	52.62	2.76	2.76
+ pollster-level controls	69.63	17.01	19.77
+ country-level controls	73.04	3.41	23.18
+ additive election-level variables	78.56	5.52	28.70
+ turnout change × margin of victory	78.56	0.00	28.70
+ turnout change × ENEP	78.56	0.00	28.70
+ turnout change × ENEP change	78.56	0.00	28.70
+ ENEP × turnout	78.56	0.00	28.70
+ turnout × ENEP change	78.56	0.00	28.70
+ late deciders × ENEP	78.56	0.00	28.70
+ late deciders × turnout	78.56	0.00	28.70
+ late deciders × turnout change	78.56	0.00	28.70

† Given the composition of the subset of data containing late decision-making measurements, it was not possible to include the electoral system or polling moratorium variables within country-level controls, as even reducing it to a binary variable failed to resolve issues of rank deficiency.

‡ There was an insufficient number of presidential elections within the subset of data for the election type and round two variables to be included within the model, as their scarcity resulted in rank deficient predictions.

Table B80: Average proportion of correct classifications for SBP that result from the inclusion of election-level interactions across the subset of data for which all control variables were available alongside late decision-making figures. Accuracy is calculated as an average of averages from repeated k-fold cross validation across 4 balanced, down-sampled subsets of elections (n = 4 × 1,474). The model is run across a reduced feature set to explore prospective overfitting in the more complex specification.

Model	**Average Accuracy (%)**	**Iterative Change in Accuracy (%)**	**Change in Accuracy Compared to Null (%)**
SBP ~ 1 (Null)	49.86	-	-
+ poll-level controls	52.62	2.76	2.76
+ pollster-level controls	69.63	17.01	19.77
+ country-level controls	73.04	3.41	23.18
+ turnout change × margin of victory	73.62	0.58	23.76
+ turnout change × ENEP	73.66	0.04	24.80
+ turnout change × ENEP change	75.63	1.97	26.77
+ ENEP × turnout	77.47	1.84	28.61
+ turnout × ENEP change	76.10	-1.37	27.24
+ late deciders × ENEP	78.58	2.48	29.72
+ late deciders × turnout	78.50	-0.08	29.64
+ late deciders × turnout change	78.50	0.00	29.64

† Given the composition of the subset of data containing late decision-making measurements, it was not possible to include the electoral system or polling moratorium variables within country-level controls, as even reducing it to a binary variable failed to resolve issues of rank deficiency.
‡ There was an insufficient number of presidential elections within the subset of data for the election type and round two variables to be included within the model, as their scarcity resulted in rank deficient predictions.

The Robustness of LVRC Findings to Controls

Table B81: Average proportion of correct classifications for LVRC calculated from the subset of data for which all control variables were available. Accuracy is calculated as an average of averages from repeated k-fold cross validation across 132 balanced, down-sampled data frames of those elections for which all control variables were available (n = 132 × 682).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
LVRC ~ 1 (Null)	49.88	-	-
+ poll-level controls	56.92	7.04	7.04
+ pollster-level controls	61.95	5.03	12.07
+ country-level controls	80.49	18.54	30.61
+ additive election-level variables	86.65	6.16	36.77
+ turnout change × margin of victory	87.18	0.53	37.30
+ turnout change × ENEP	87.91	0.73	38.03
+ turnout change × ENEP change	87.18	-0.73	37.30
+ ENEP × turnout	88.55	1.37	38.67
+ turnout × ENEP change	88.70	0.15	38.82

Table B82: Average proportion of correct classifications for LVRC calculated from the subset of data for which all control variables were available. Accuracy is calculated as an average of averages from repeated k-fold cross validation across 166 balanced, down-sampled data frames of those elections for which all control variables were available alongside measurements of partisan sentiment within the electorate (n = 166 × 438).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
LVRC ~ 1 (Null)	49.81	-	-
+ poll-level controls	60.56	10.75	10.75
+ pollster-level controls	63.93	3.37	14.12
+ country-level controls†	83.73	19.80	33.92
+ additive election-level variables	88.64	4.91	38.83
+ turnout change × margin of victory	89.38	0.74	39.57
+ turnout change × ENEP	90.16	0.78	40.35
+ turnout change × ENEP change	91.12	-0.04	40.31
+ ENEP × turnout	91.17	0.05	40.36
+ turnout × ENEP change	92.11	0.94	41.30
+ ENEP × partisanship	93.06	0.95	42.25
+ turnout change × partisanship	92.06	-1.00	41.25
+ turnout × partisanship	91.94	-0.12	41.13
+ turnout change × ENEP × partisanship	91.87	-0.07	41.06
+ turnout change × ENEP change × partisanship	90.84	-1.03	40.03
+ ENEP × turnout × partisanship	90.51	-0.33	39.70
+ turnout × ENEP change × partisanship	89.96	-0.55	39.15

† Given the composition of the subset of data containing partisanship measurements, it was not possible to include the electoral system variable within country-level controls, as even reducing it to a binary variable failed to resolve issues of rank deficiency. Similarly, too few instances of compulsory voting were present to allow for its inclusion. The scarcity of these values resulted in rank deficient predictions due to the presence of constants in one or more of the train/test splits generated by the k-fold cross validation process.

Table B83: Average proportion of correct classifications for LVRC that result from the inclusion of election-level variables both additively and interactively across the subset of data for which all control variables were available alongside late decision-making figures. Accuracy is calculated as an average of averages from repeated k-fold cross validation across 210 balanced, down-sampled subsets of elections (n = 210 × 222).

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
LVRC ~ 1 (Null)	49.40	-	-
+ poll-level controls	69.64	20.24	20.24
+ pollster-level controls	73.81	4.17	24.41
+ country-level controls	82.14	8.33	32.74
+ additive election-level variables	86.31	4.17	36.91
+ turnout change × margin of victory	86.31	0.00	36.91
+ turnout change × ENEP	86.31	0.00	36.91
+ turnout change × ENEP change	86.31	0.00	36.91
+ ENEP × turnout	86.31	0.00	36.91
+ turnout × ENEP change	86.31	0.00	36.91
+ late deciders × ENEP	86.31	0.00	36.91
+ late deciders × turnout	86.31	0.00	36.91
+ late deciders × turnout change	86.31	0.00	36.91

† Given the composition of the subset of data containing late decision-making measurements, it was not possible to include the electoral system or polling moratorium variables within country-level controls, as even reducing it to a binary variable failed to resolve issues of rank deficiency.

‡ There was an insufficient number of presidential elections within the subset of data for the election type and round two variables to be included within the model, as their scarcity resulted in rank deficient predictions.

Table B84: Average proportion of correct classifications for SBP that result from the inclusion of election-level interactions across the subset of data for which all control variables were available alongside late decision-making figures. Accuracy is calculated as an average of averages from repeated k-fold cross validation across 210 balanced, down-sampled subsets of elections (n = 210 × 222). The model is run across a reduced feature set to explore prospective overfitting in the more complex specification.

Model	Average Accuracy (%)	Iterative Change in Accuracy (%)	Change in Accuracy Compared to Null (%)
LVRC ~ 1 (Null)	49.40	-	-
+ poll-level controls	69.64	20.24	20.24
+ pollster-level controls	73.81	4.17	24.41
+ country-level controls	82.14	8.33	32.74
+ turnout change × margin of victory	85.12	2.98	35.72
+ turnout change × ENEP	85.72	0.60	36.32
+ turnout change × ENEP change	86.30	0.58	36.90
+ ENEP × turnout	87.50	1.20	38.10
+ turnout × ENEP change	88.10	0.60	38.70
+ late deciders × ENEP	88.10	0.00	38.70
+ late deciders × turnout	89.29	1.19	39.89
+ late deciders × turnout change	89.88	0.59	40.48

† Given the composition of the subset of data containing late decision-making measurements, it was not possible to include the electoral system or polling moratorium variables within country-level controls, as even reducing it to a binary variable failed to resolve issues of rank deficiency.
‡ There was an insufficient number of presidential elections within the subset of data for the election type and round two variables to be included within the model, as their scarcity resulted in rank deficient predictions.

Bibliography

Aarts, Emmeke et al., ‘A Solution to Dependency: Using Multilevel Analysis to Accommodate Nested Data’, *Nature Neuroscience*, 17.4 (2014), 491 – 496

ABC News, *Australian election campaign begins* (2013), <https://www.abc.net.au/news/programs/the-world/2013-08-05/australian-election-campaign-begins/4866762> [accessed 04/08/2023]

ABC News, *Lube, legs and lies: Lacklustre start to federal election campaign sees politicians resort to personal attacks* (2019), <https://www.abc.net.au/news/2019-04-12/federal-election-2019-day-one-taxes-legs-lubricant/10996876> [accessed 04/08/2023]

ABC News, *Pakistan election campaign begins* (2013), <https://www.abc.net.au/news/programs/the-world/2013-03-27/pakistan-election-campaign-begins/4598540> [accessed 04/08/2023]

Abramowitz, Alan I., ‘An Improved Model for Predicting Presidential Election Outcomes’, *PS: Political Science and Politics*, 21.4 (1988), 843-847

Abramowitz, Alan I., ‘Visibility, Electability, and Candidate Choice in a Presidential Primary Election: A Test of Competing Models’, *The Journal of Politics*, 51.4 (1989), 977-992

Abrams, Mark, ‘The Opinion Polls and the 1970 British General Election’, *The Public Opinion Quarterly*, 34.3 (1970), 317-324

Abramson, Paul R., ‘Generational Replacement and Partisan Dealignment in Britain and the United States’, *British Journal of Political Science*, 8.4 (1978), 505-509

Abramson, Paul R. and others, ‘Comparing Strategic Voting Under FPTP and PR’, *Comparative Political Studies*, 43.1 (2010), 61 – 90

Adam, Barbara and Chris Groves, *Future Matters: Action, Knowledge, Ethics*, (Leiden: Brill, 2007)

Adams, James, Jay Dow, and Samuel Merill III, ‘The Political Consequences of Alienation-based and Indifference-based Voter Abstention: Applications to Presidential Elections’, *Political Behavior*, 28.1 (2006), 65 – 86

Adkins, Randall E. and Andrew J. Dowdle, ‘The Money Primary: What Influences the Outcome of Pre-primary Presidential Nomination Fundraising?’, *Presidential Studies Quarterly*, 32.2 (2002), 256-275

Afro Barometer, *Data* (2022), <https://afrobarometer.org/data> [accessed 05/03/2022]

Akoglu, Haldun, ‘User’s Guide to Correlation Coefficients’, *Turkish Journal of Emergency Medicine*, 18.3 (2018), 91 – 93

Al-Aswat, Asharq, *Campaigning begins for Tunisia’s parliamentary elections* (2014), <https://web.archive.org/web/20141008171651/http:/www.aawsat.net/2014/10/article55337236> [accessed 04/08/2023]

Albers, Casper Albers and Daniel Lakens, ‘When Power Analyses Based on Pilot Data are Biased: Inaccurate Effect Size Estimators and Follow-up Bias, *Journal of Experimental Social Psychology*, 74 (2018), 187 – 195

Alkharusi, Hussain, ‘Categorical Variables in Regression Analysis: A Comparison of Dummy and Effect Coding’, *International Journal of Education*, 4.2 (2012), 202 – 210

All Africa, *Kenya: Campaign Period to Officially Start on May 29 as IEBC Gazettes Election Data* (2022), <https://allafrica.com/stories/202201240205.html> [accessed 04/08/2023]

Allan, James P., Marc J. O’Reilly, and Richard Vengroff, ‘The Election Everybody Won? The Impact of Party System Change, Voter Turnout, and Strategic Voting in the 1998 Quebec Election’, *American Review of Canadian Studies*, 30.4 (2000), 497 – 519

American National Election Studies, *Data Center* (2022), <https://electionstudies.org/data-center/> [accessed 05/03/2022]

Anderson, Christopher J. and Pablo Beramendi, ‘Income, Inequality, and Electoral Participation’, in *Democracy, Inequality, and Representation in Comparative Perspective*, ed. by Pablo Beramendi and Christopher J. Anderson, (New York: Russell Sage Foundation, 2008)

Anderson, Leslie, ‘Neutrality and Bias in the 1990 Nicaraguan Pre-election Polls: A Comment on Bischoping and Schuman’, *American Journal of Political Science*, 38.2 (1994), 486 – 494

Anderson, Leslie, ‘Surprises and Secrets: Lessons from the 1990 Nicaraguan Election’, *Studies in Comparative International Development*, 27.3 (1992), 93 – 119

Anderson, Robert, James Tilly, and Anthony Heath, ‘Political Knowledge and Enlightened Preferences: Party Choice through the Electoral Cycle’, *British Journal of Political Science*, 35 (2005), 285-302

Andersson, Ulf, Alvaro Cuervo-Cazurra, and Bo Bernhard Nielsen, ‘From the Editors: Explaining Interaction Effects Within and Across Levels of Analysis’, *Journal of International Business Studies*, 45 (2014), 1063 – 1071

Andrews, Geoff, ‘The Italian General Election of 2006’, *Representation*, 42.3 (2006), 253 – 260

Annas, George J., ‘Precatory Prediction and Mindless Mimicry’, *The Hastings Center Report*, 18.6 (1988), 31

Ansolabehere, S. and E. Hersh, ‘Validation: What Big Data Reveal About Survey Misreporting and the Real Electorate’, *Political Analysis*, 20.4 (2012), 437 – 459

Ansu-Kyeremeh, Kwasi, ‘The Challenges of Surveying Public Opinion in an Emerging Democracy’, *International Journal of Public Opinion Research*, 11.1 (1999), 59-74

Arceneaux, Kevin, ‘Do Campaigns Help Voters Learn? A Cross-national Analysis’, *British Journal of Political Science*, 36 (2005), 159-173

Arcuri, Luciano and others, ‘Predicting the Vote: Implicit Attitudes as Predictors of the Future Behaviour of Decided and Undecided Voters’, *Political Psychology*, 29.3 (2008), 369 – 387

Ardevol-Abreu, Alberto and Homero Gil de Zuniga, ‘Obstinate partisanship: political discussion attributes effects on the development of unconditional party loyalty’, *International Journal of Communication*, 14 (2020), 324-345

Arzheimer, Kai and Jocelyn Evans, ‘A New Multinomial Accuracy Measure for Polling Bias’, *Political Analysis*, 22 (2014), 31 – 44

Arzheimer, Kai and Jocelyn Evans, ‘Estimating Polling Accuracy in Multiparty Elections Using Surveybias’, *The Stata Journal*, 16.1 (2016), 139 – 158

Ascher, William, ‘Political Forecasting: The Missing Link’, *Journal of Forecasting*, 1 (1982), 227 – 239

Asian Barometer, *Data Release* (2022), <http://www.asianbarometer.org/data/data-release> [accessed 05/03/2022]

Atkinson, Simon, ‘The Opinion Polls: In Praise of Measurement’, *International Journal of Market Research*, 59.4 (2017), 405 – 407

Auret, Lidia and Chris Aldrich, ‘Interpretation of Nonlinear Relationships Between Process Variables by Use of Random Forests’, *Minerals Engineering*, 35 (2012), 27 – 42

Austin, Peter C. and Jack. V. Tu, ‘Bootstrap Methods for Developing Predictive Models’, *The American Statistician*, 58.2 (2004), 131 – 137

Australian Election Study, *Interactive Charts* (2020), <https://australianelectionstudy.org/interactive-charts/> [accessed 05/03/2022]

Australian Electoral Commission, *2022 Federal Election Timetable* (2022), <https://www.aec.gov.au/Elections/federal_elections/2022/timetable.htm> [accessed 04/08/2023]

Australian Electoral Commission, *Election dates 1901 – present* (2023), <https://www.aec.gov.au/Elections/federal_elections/election-dates.htm> [accessed 04/08/2023]

Australian Electoral Commission, Federal Elections (2023), <https://www.aec.gov.au/Elections/federal_elections/> [accessed 04/08/2023]

AUTNES, *Austrian National Election Study* (2019), <https://www.autnes.at/en/page/3/?a=register&c=index&siteid=1> [accessed 05/03/022]

Baerwald, Hans H., 'Japan's 35th House of Representatives Election: The LDP Toys with a Return to 1954', *Asian Survey*, 20.3 (1980), 257 – 268

Bafumi, Joseph, Robert S. Erikson, and Christopher Wlezien, 'Forecasting House Sears from Generic Congressional Polls: The 2010 Midterm Election', *PS: Political Science and Politics*, 43.4 (2010), 633 – 636

Baker, S. A. Abu and others, 'Modelling Loss Data Using Composite Models', *Insurance: Mathematics and Economics*, 61 (2015), 146 – 154

Bale, Tim, 'Restricting the Broadcast and Publication of Pre-election and Exit Polls: Some Selected Examples', *Representation*, 39.1 (2002), 15 – 22

Bartels, Larry, 'Partisanship and Voting Behaviour, 1952-1996', *American Journal of Political Science*, 44 (2000), 35-50

Bartle, John, Agusti Bosch and Lluís Orriols, 'The Policy Mood in Spain: The Thermostat in a Warm Climate, 1978 – 2017', *European Political Science Review*, 12.2 (2020), 133 – 153

Bauer, Alexander and others, 'Mundus Vult Decipi, Ergo Decipiatur: Visual Communication of Uncertainty in Election Polls', *PS: Political Science & Politics*, 55.1 (2022), 102 – 108

Bautista, René and others, 'Studying Nonresponse in Mexican Exit Polls', *International Journal of Public Opinion Research*, 19.4 (2007), 492-503

Baylis, John, *British Defence Policy* (London: Palgrave Macmillan, 1989)

Begum, Imrana, General Election in Pakistan: A Critical Study', *FWU Journal of Social Sciences*, 16.3 (2022), 132 – 143

Belli, R. F. and others, ‘Reducing Over-reporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring’, *Public Opinion Quarterly*, 63.1 (1999), 90-108

Beltrán, Ulises and Marcos Valdivia, ‘Accuracy and Error in Electoral Forecasts: The Case of Mexico’, *International Journal of Public Opinion Research*, 11.2 (1999), 115 – 134

Benoit, Kenneth, ‘Like Déjà Vu All Over Again: The Hungarian Parliamentary Elections of 2002’, *Journal of Communist Studies and Transition Politics*, 18.4 (2002), 119 – 133

Benson, Edward G., Cyrus C. Young, Clyde A. Syze, ‘Polling Lessons from the 1944 Election’, *Public Opinion Quarterly*, 9.4 (1945), 467 – 484

Berelson, Bernard, Paul Lazarsfeld, and William McPhee, *Voting: A Study of Opinion Formation in a Presidential Election,* (Chicago: University of Chicago Press, 1954)

Bergman, Jakob and Björn Holmquist, ‘Poll of Polls: A Compositional Loss Model’, *Scandinavian Journal of Statistics*, 41.2 (2014), 301 – 310

Bernstein, R., A. Chadha, and R. Montjoy, ‘Overreporting Voting: Why it Happens and Why it Matters’, *Public Opinion Quarterly*, 65.1 (2001), 22 – 44

Bertenthal, Bennett I., ‘Dynamical systems: It’s about time’, in *Data Analytic Techniques for Dynamical Systems*, ed. by S. M. Boker and M. J. Wenger, (Hillside, NJ: Erlbaum, 2007), 1 – 24.

Best, Samuel J. and Benjamin Radcliff, *Polling America: An Encyclopedia of Public Opinion*, (Westport: Greenwood Publishing Group, 2005)

Bianconcini, Silvia, Silvia Cagnone, and Dimitris Rizopoulos, ‘Approximate Likelihood Inference in Generalized Linear Latent Variable Models Based on the Dimension-wise Quadrature’, *Electronic Journal of Statistics*, 11 (2017), 4404 – 4423

Biemer, Paul P., ‘Total Survey Error: Design, Implementation, and Evaluation’, *Public Opinion Quarterly*, 74.5 (2010), 817 – 848

Biemer, Paul, ‘Measuring Data Quality’, in *Telephone Survey Methodology*, ed. by Robert Groves and others, (New York: John Wiley and Sons, 1988)

Birch, Sarah, ‘Two-round Electoral Systems and Democracy’, *Comparative Political Studies*, 36.3 (2003), 319 – 344

Bischoping, Katherine and Howard Schuman, ‘Pens and Polls in Nicaragua: An Analysis of the 1990 Pre-election Surveys’, *American Journal of Political Science*, 36.2 (1992), 331 – 350

Blais, André and Christopher H. Achen, ‘Civic Duty and Voter Turnout’, *Political Behaviour*, 41 (2019), 473-497

Blais, André and others, ‘Assessing the psychological and mechanical impact of electoral rules: A quasi-experiment’, *Electoral Studies*, 31 (2012), 829-837

Bliese, Paul D., ‘Within-Group Agreement, Non-Independence, and Reliability’, in *Multilevel Theory, Research, and Methods in Organizations*, ed. by Katherine J. Klein and Steve W. J. Kozlowski, (San Francisco: Jossey-Bass, 2000), 349 – 381

Blum, Avrim, Adam Kalai, and John Langford, ‘Beating the Hold Out: Bounds for K-fold and Progressive Cross-validation’, *Proceedings of the 12th Annual Conference on Learning Theory*, 1 (1999), 1 – 6

Blumberg, Herbert H., Carolyn Fuller, and A. Paul Hare, ‘Response Rates in Postal Surveys’, *Public Opinion Quarterly*, 38.1 (1974), 113 -123

Blumenstiel, Jan Eric, and Thomas Plischke, ‘Changing Motivations, Time of the Voting Decision, and Short-term Volatility – The Dynamics of Voter Heterogeneity’, *Electoral Studies*, 37 (2015), 28 – 40

Blumenthal, Mark, ‘Polls, Forecasts, and Aggregators’, *Political Science and Politics*, 47.2 (2014), 297 – 300

Bobak, Carly A., Paul J. Barr, and A. James O’Malley, ‘Estimation of an Inter-Rater Intra-Class Correlation Coefficient That Overcomes Common Assumption Violations in the Assessment of Health Measurement’, *BMC Medical Research Methodology*, 18.1 (2018), 93 – 114

Bodor, Tamás, ‘Hungary’s “Black Sunday” of Public Opinion Research: The Anatomy of a Failed Election Forecast’, *International Journal of Public Opinion Research*, 24.4 (2012), 450 – 471

Bogart, Leo, ‘Politics, Polls, and Poltergeists’, Society, 35.4 (1998), 8 – 16

Bol, Damien and others, *MEDW 2014 Belgian National Election Study* (2017), <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/7GA3IT> [accessed 05/03/2022]

Bon, Joshua J., Timothy Ballard, and Bernard Baffour, ‘Polling Bias and Undecided Voter Allocations: US Presidential Elections, 2004 – 2016, *Journal of the Royal Statistical Society*, 182.2 (2019), 467 – 493

Borg, Sami and Kimmo Grönlund, *FSD2653 Finnish National Election Study 2011* (2011), <https://services.fsd.tuni.fi/catalogue/FSD2653?lang=en&study_language=en> [accessed 05/03/2022]

Bormann, Nils-Christian and Matt Golder, ‘Democratic Electoral Systems Around the World, 1946 – 2011’, *Electoral Studies*, 32 (2013), 360 – 369

Borrelli, Stephen, Brad Lockerbie, and Richard G. Niemi, ‘Why the Democrat-Republican Partisanship Gap Varies from Poll to Poll’, *The Public Opinion Quarterly*, 51.1 (1987), 115 – 119

Bowler, Shaun and David M. Farrell, ‘The Study of Election Campaigning’ in *Electoral Strategies and Political Marketing* ed. by Shaun Bowler and David M. Farrell, (London: Palgrave Macmillan, 1992), 1-23

Box, Brian and Joseph Giammo, ‘Late Deciders in U.S. Presidential Elections’, *The American Review of Politics*, 30 (2009), 333-355

Boyd Jr., Harper W.., and Ralph Westfall, ‘Interviewers as a Source of Error in Surveys’, *Journal of Marketing*, 19.4 (1955), 311 – 324

Brady, Henry, Sidney Verba, and Kay Lehman Schlozman, ‘Beyond SES: A Resources Model of Political Participation’, *American Political Science Review*, 89.2 (1995), 271 – 294

Braun, Daniela and others, ‘Issues that mobilize Europe. The role of key policy issues for voter turnout in the 2019 European Parliament election’, *European Union Politics*, 23.1 (2022), 120-140

Breiman, Leo, ‘Statistical Modeling: The Two Cultures’, *Statistical Science*, 16.3 (2001), 199 – 231

Bristol University LEMMA, *Comparing Groups Using Multilevel Modelling* (2022), <https://www.cmm.bris.ac.uk/lemma/mod/lesson/view.php?id=276&pageid=336&startlastseen=no> [accessed 12/08/2022].

Britannica, *United States Presidential Election Results* (2022), <https://www.britannica.com/topic/United-States-Presidential-Election-Results-1788863> [accessed 16/07/2022]

British Election Study, *Data* (2022), <https://www.britishelectionstudy.com/data/> [accessed 05/03/2022]

British Election Study, *1963 – 1970 Political Change in Britain* (2022), <https://www.britishelectionstudy.com/data-object/1963-1970-political-change-in-britain/> [accessed 16/08/2022]

British Polling Council, *Poll Methodology, Weighting, and Adjustment Systems* (2013), https://www.britishpollingcouncil.org/wordpress/wp-content/uploads/2013/11/andrew-cooper.pdf> [accessed 7 February 2022]

Brody, Richard and Lawrence Rothenberg, 'The Instability of Partisanship: An Analysis of the 1980 Presidential Election', *British Journal of Political Science*, 18 (1988), 445 – 465

Broh, C. Anthony, 'Horse-Race Journalism: Reporting the Polls in the 1976 Presidential Election', *Public Opinion Quarterly*, 44.4 (1980), 514 – 529

Brown, Lloyd B. and Henry W. Chappell Jr., 'Forecasting Presidential Elections Using History and Polls', *International Journal of* Forecasting, 15.2 (1999), 127 – 135

Brown, P. J., D. Firth, and C. D. Payne, 'Forecasting on British Election Night 1997', *Journal of the Royal Statistical Society*, 162 (1999), 211 – 226

Brown-Iannuzzi, Jazmin, Maxine Najle, and Will Gervais, 'The Illusion of Political Tolerance: Social Desirability and Self-reported Voting Preferences', *Social Psychological and Personality Science*, 10.3 (2019), 364 – 373

Brox, Brian and Joseph Giammo, 'Late Deciders in U.S. Presidential Elections', *American Review of Politics*, 30 (2009), 333-355

Buchanan, William, 'Electoral Predictions: An Empirical Assessment', *The Public Opinion Quarterly*, 50.2 (1986), 222 -227

Buhlmann, Marc and others, 'The Democracy Barometer: A New Instrument to Measure the Quality of Democracy and its Potential for Comparative Research', *European Political Science*, 11 (2012), 519 – 53

Buller, Jim and Toby S. James, 'Statecraft and the assessment of national political leaders: The case of New Labour and Tony Blair', *The British Journal of Politics and International Relations*, 14.4 (2012), 534-555

Burden, Barry, 'Voter Turnout and the National Election Studies', *Political Analysis*, 8.4 (2000), 389 – 398

Bürkner, Paul-Christian, 'brms: An R Package for Bayesian Multilevel Models Using Stan', *Journal of Statistical Software*, 80 (2017), 1 – 28

Bürkner, Paul-Christian, *Package 'brms'* (2021), <https://cran.r-project.org/web/packages/brms/brms.pdf> [accessed 10 February 2022]

Butler, David, Anthony King, and Fintan Hoey, *The British General Election of 1966* (London: Palgrave Macmillan, 1966)

Butler, David and Dennis Kavanagh, *The British General Election of 1992*, (London: Macmillan, 1992)

Butler, David and Dennis Kavanagh, *The British General Election of 2001* (London: Palgrave Macmillan, 2002)

Butler, David and M. Pinto-Duschinsky, *The British General Election of 1970*, (London: Macmillan, 1971)

Butler, David and Richard Rose, *The British General Election of 1959*, (London: Palgrave Macmillan, 1960)

Callegaro, Mario and Gaincarlo Gasperoni, 'Accuracy of Pre-election Polls for the 2006 Italian Parliamentary Election: Too Close to Call', *International Journal of Public Opinion Research*, 20.2 (2008), 148 – 170

Camarota, Steven and Karen Zeigler, *Projecting the 2012 Hispanic Vote Shares Nationally and in Battleground States* (2012), <https://cis.org/sites/cis.org/files/projecting-hispanic-vote-2012.pdf> [accessed 29/08/2020]

Campbell, Angus, 'Surge and Decline: A Study of Electoral Change', in *Elections and Political Order*, ed. by Angus Campbell and others, (New York: Wiley, 1966)

Campbell, Angus and others, *The American Voter* (New York: Wiley, 1960)

Campbell, Donald T., 'Systematic Error on the Part of Human Links in Communication Systems', *Information and Control*, 1 (1958), 334 – 369

Campbell, James E., 'Do Swing Voters Swing Elections?', in *The Swing Voter in American Politics*, ed. by William G. Mayer (Washington D.C.: Brookings Institution Press, 2007)

Campus, Donatella and Gianfranco Pasquino, 'Leadership in Italy: The Changing Role of Leaders in Elections and in Government', *Journal of Contemporary European Studies*, 14.1 (2006), 25 – 40

Campus, Donatella, 'The 2006 Election: More Than Ever, a Berlusconi-centred Campaign', *Journal of Modern Italian Studies*, 11.4 (2006), 516 – 531

Cantril, Hadley, 'Do Different Polls Get the Same Results?', *Public Opinion Quarterly*, 9.1 (145), 61 – 69

Capanu, Marinela, Mithat Gonen, and Colin B. Begg, 'An Assessment of Estimation Methods for Generalized Linear Mixed Models with Binary Outcomes', *Stat Med*, 32.26 (2013), 1 – 24

Carlin, Bradley and Thomas Louis, ‘Identifying Prior Distributions that Produce Specific Decisions with Application to Monitoring Clinical Trials’, in *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*, ed. by John Donald, Kathryn Chaloner, and Arnold Zellner, (New York: John Wiley and Sons, 1996)

Carter, Woody, ‘Interviewer-related Error’, in *Encyclopedia of Survey Research Methods*, ed. by Paul J. Lavrakas, (Thousand Oaks, Sage Publications, 2008)

Castillo-Manzano, José I., Lourdes López-Valpuesta, and Rafael Pozo-Barajas, ‘At a Time of Insurgent Parties, Can Societies Believe in Election Polls? The Spanish Experience’, *Revista de Economía Aplicada*, 78.26 (2018), 81 – 97

Castle, Jennifer L., Michael P. Clements, and David F. Hendry, *Forecasting: An Essential Introduction*, (New Haven: Yale University Press, 2019)

Cavanagh, Thomas E., ‘Changes in American Voter Turnout, 1964-1976’, Political Science Quarterly, 96.1 (1981), 53-65

Ceccarini, Luigi and Ilvo Diamanti, ‘The Election Campaign and the Last-minute Deciders’, *Contemporary Italian Politics*, 5.2 (2013), 130-148

Centre for Systematic Peace, *Polity5: Regime Authority Characteristics and Transitions Datasets* (2018), <https://www.systemicpeace.org/inscrdata.html> [accessed 21/01/2022]

Cerina, Roberto and Raymond Duck, ‘Polling India via Regression and Post-stratification of Non-probability Online Sample’, *PLoS ONE*, 16.11 (2021), 1 – 34

Chadefaux, Thomas, ‘Conflict Forecasting and its Limits’, *Data Science*, 1.1 (2017), 7 – 17

Chang L., and J. A. Krosnick, ‘National Surveys via RDD Telephone Interviewing Versus the Internet: Comparing Sample Representativeness and Response Quality’, *Public Opinion Quarterly*, 73.4 (2009), 641 – 678

Chapel Hill Expert Survey, *1999-2019 Chapel Hill Expert Survey (CHES) Trend File* (2022), <https://www.chesdata.eu/1999-2019chestrend> [accessed 17/06/2022]

Chiaramonte, Alessandro, ‘The Unfinished Story of Electoral Reforms in Italy’, *Contemporary Italian Politics*, 7.1 (2015), 10-26

Chiaramonte, Alessandro and Roberto D’Alimonte, ‘The New Italian Electoral System and its Effects on Strategic Coordination and Disproportionality’, *Italian Political Science*, 13.1 (2018), 8-18

Chicago Daily Tribune, ‘Dewey Defeats Truman’, *Chicago Daily Tribune*, 3 November 1948

Chiru, Mihail and Sergiu Gherghina, ‘When Vote Loyalty Fails: Party Performance and Corruption in Bulgaria and Romania’, *European Political Science Review*, 4.1 (2012), 29 – 49

Clarke, P., ‘When Can Group Level Clustering Be Ignored? Multilevel Models Versus Single-level Models with Sparse Data’, *Journal of Epidemial Community Health*, 62 (2008), 752 – 758

Clift, Ben and Sean McDaniel, ‘Is this crisis of French socialism different? Hollande, the rise of Macron, and the reconfiguration of the left in the 2017 presidential and parliamentary elections’, *Modern & Contemporary France*, 25.4 (2017), 403-415

Clinton, Joshua D., John S. Lapinski, and Marc J. Trussler, ‘Reluctant Republicans, Eager Democrats? Partisan Nonresponse and the Accuracy of 2020 Presidential Pre-election Telephone Polls’, *Public Opinion Quarterly*, 86.2 (2022), 247 – 269

Cohodas, Nadine, *Strom Thurmond and the Politics of Southern Change*, (Mercer University Press, 1994)

Colander, David, and others, ‘The Financial Crisis and the Systemic Failure of the Economics Profession’, *Critical Review*, 21.2 (2009), 249 – 267

Coletto, David and Bryan Breguet, ‘The Accuracy of Public Polls in Provincial Elections’, *Canadian Political Science Review*, 9.1 (2015), 41 – 54

Collins, M. K., ‘The Effect of the Iranian Hostage Crisis on the 1980 Presidential Election’, *Tenor of Our Times*, 2.6 (2013), 28 – 35

Comparative Study of Electoral Systems, *Download Data and Documentation* (2021), <https://cses.org/data-download/download-data-documentation/> [accessed 05/03/2022]

Connelly, Gordon M., ‘The Questions the Polls Ask: Now Let’s Look at the Real Problem: Validity’, *Public Opinion Quarterly*, 9.1 (1945), 51 – 60

Converse, Jean M., ‘Predicting No Opinion in Polls’, *Public Opinion Quarterly*, 40 (1976), 515 – 530

Converse, Philip E., ‘Of Time and Partisan Stability’, *Comparative Political Studies*, 2.2 (1969), 139–71

Converse, Philip E. and Michael W. Traugott, ‘Assessing the Accuracy of Polls and Surveys’, *Science*, 234.4780 (1986), 1094 – 1098

Coppedge, Michael, ‘Democracy and Dimensions: Comments on Munck and Verkuilen’, *Comparative Political Studies*, 35.1 (2002), 35 – 39

Coppock, Alexander, ‘Did Shy Trump Supporters Bias the 2016 Polls? Evidence from a Nationally-representative List Experiment’, *Statistics, Politics and Policy*, 8.1 (2017), 29-40

Coram, Alex, ‘The Red Queen and the Dynamics of Resource Spending in Party Competition’, *British Journal of Political Science*, 40.2 (2010), 469-475

Coscuig, Anatolie, ‘The Influence of the Electoral System on the Quality of Electoral Predictions: Romanian Elections: 1992 – 2009’, *Studia UBB Politica*, 1 (2011), 79 – 118

Cowley, Philip and Dennis Kavanagh, *The British General Election of 2017*, (London: Palgrave Macmillan, 2018)

Cox, Gary W., *Making Votes Count: Strategic Coordination in the World’s Electoral Systems*, (New York: Cambridge University Press, 1997)

Craig, Stephen C., ‘The Decline of Partisanship in the United States: A Re-examination of the Neutrality Hypothesis’, *Political Behavior*, 7 (1985), 57-78

Crespi, Irving and Dwight Morris, ‘Question Order and the Measurement of Candidate Preference in the 1982 Connecticut Elections’, *Public Opinion Quarterly*, 48.3 (1984), 578-591

Crespi, Irving, *Pre-Election Polling: Sources of Accuracy and Error*, (New York: The Russell Sage Foundation, 1988)

Crespi, Leo P., ‘The Interview Effect in Polling’, *Public Opinion Quarterly*, 12.1 (1948), 99-111

Crewe, Ivor, ‘A Nation of Liars? Opinion Polls and the 1992 Election’, *International Journal of Market Research*, 35.4 (1993), 1 – 19

Crewe, Ivor, ‘The Opinion Polls: Confidence Restored?’, *Parliamentary Affairs*, 50.4 (1997), 569 – 586

Crewe, Ivor, ‘The Opinion Polls: Still Biased to Labour’, *Parliamentary Affairs*, 54.4 (2001), 650 – 665

Crewe, Ivor, ‘The Opinion Polls: The Election They Got (Almost) Right’, *Parliamentary Affairs*, 58 (2005), 684 – 698

Crewe, Ivor, Bo Särlvik, and James Alt, ‘Partisan Dealignment in Britain 1964 – 1974, *British Journal of Political Science*, 7.2 (1977), 129 – 190

Crilly, Rob, *Newsweek Recalls 125,000 Copies of its Souvenir Madam President Issue* (2016), <https://www.telegraph.co.uk/news/2016/11/10/newsweek-recalls-125000-copies-of-its-souvenir-madam-president-i/> [accessed 02/08/2020

Crossley, Archibald M. and Helen M. Crossley, 'Polling in 1968', *The Public Opinion Quarterly*, 33.1 (1969), 1 – 16

Crossley, Archibald M., 'Straw Polls in 1936', *Public Opinion Quarterly*, 1.1 (1937), 24 – 35

Crowell, Laura, 'Franklin D. Roosevelt's Audience Persuasion in the 1936 Campaign', *Speech Monographs*, 17.1 (1950), 48 – 64

Csaba Nikolenyi, 'Strategic Co-ordination in the 2002 Hungarian Election', *Europe-Asia Studies*, 567 (2004), 1041 – 1058

CSES, *Election Studies* (2022), <https://cses.org/data-download/download-data-documentation/election-studies/> [accessed 17/08/2022]

Curia, *Archives: Roy Morgan* (2022), <https://www.curia.co.nz/company/roy-morgan/> [accessed 12/09/2022]

Curtice, John and Nick Sparrow, 'The Past Matters: Eliminating the Pro-Labour Bias in British Opinion Polls', *International Journal of Market Research*, 52.2 (2010), 169 – 189

Curtice, John, 'So How Well Did They Do? The Polls in the 1997 Election', *Journal of the Market Research Society*, 39.3 (1997), 449 – 461

Daily Nation, *Presidential Candidates Present Papers to IEBC* (2013), <https://archive.ph/2013.02.20-235205/http:/elections.nation.co.ke/news/Presidential-candidates-present-papers-to-IEBC-/-/1631868/1674424/-/ty88lrz/-/index.html> [accessed 04/08/2023]

Dal Pozzolo, Andrea, Olivier Caelen, and Gianluca Bontempi, 'Comparison of Balancing Techniques for Unbalanced Dataset', *Mach. Learn. Gr. Univ. Libr, Bruxelles Belgium*, 16.1 (2010), 732 – 735

Daoust, Jean Francois, 'Blame it on Turnout? Citizens' Participation and Polls' Accuracy', *The British Journal of Politics and International Relations*, 23.4 (2021), 736 – 747

Dawn, *National Assembly stands dissolved as second successive democratic government completes five-year term* (2018), <https://www.dawn.com/news/amp/1411167> [accessed 04/08/2023]

Day, David, *John Curtin: A Life* (Sydney: Harper Collins, 1999)

Day, Richard and Kurt M. Becker, ‘Pre-election Polling in the 1982 Illinois Gubernatorial Contest’, *Public Opinion Quarterly*, 48.3 (1984), 606 – 614

de Bruin, Wandi Bruine and others, ‘Asking About Social Circles Improves Election Predictions Even with Many Political Parties’, *International Journal of Public Opinion Research*, 34.1 (2022), 1 – 11

De Cáceres, M., X. Font, and F. Oliva, ‘Assessing Species Diagnostic Value in Large Data Sets: A Comparison Between Phi-coefficient and Ochiai Index’, *Journal of Vegetation Science*, 19.6 (2008), 779 – 788

de Cáceres, Miquel and Pierre Legendre, ‘Associations Between Species and Groups of Sites: Indices and Statistical Inference’, *Ecology*, 90.12 (2009), 3566 – 3574

de Marchi, Scott, ‘Adaptive Models and Electoral Instability’, *Journal of Theoretical Politics*, 11.3 (1999), 393 – 419

De Neve, Jan-Emmanuel, ‘The Median Voter Data Set: Voter Preferences Across 50 Democracies’, *Electoral Studies*, 30.4 (2011), 865 – 871

De Stefano, Domenico, Francesco Pauli, and Nicola Torelli, ‘Pre-electoral Polls Variability: A Hierarchical Bayesian Model to Assess the Roles of House Effects with Application to Italian Elections’, *The Annals of Applied Statistics*, 16.1 (2022), 460 – 476

DeSart, Jay and Thomas Holbrook, ‘Campaigns, Polls, and the States: Assessing the Accuracy of State-wide Presidential Trial-heat Polls’, *Political Research Quarterly*, 56.4 (2003), 431 – 439

Dey, Eric L., ‘Working with Low Survey Response Rates: The Efficacy of Weighting Adjustments’, *Research in Higher Education*, 38 (1997), 215-227

Dierkes, Julian, ‘Mongolia in 2016’, *Asian Survey*, 57.1 (2017), 128 – 134

Diniz, Márcio and others, ‘The Sticker Collector’s Problem’, *The College Mathematics Journal*, 47.4 (2016), 255 – 263

Dowding, Keith and Charles Miller, ‘On Prediction in Political Science’, *European Journal of Political Research*, 58 (2019), 1001 – 1018

Dowding, Keith, *The Philosophy and Methods of Political Science*, (New York: Palgrave Macmillan, 2016)

Drummond, Chris and Robert C. Holte, ‘Class Imbalance and Cost Sensitivity: Why Under-sampling Beats Oversampling’, *Workshop on Learning from Imbalanced Datasets*, 11 (2003), 1 – 8

Duff, Brian and others, ‘Good Excuses: Understanding Who Votes with an Improved Turnout Question’, *Public Opinion Quarterly*, 71.1 (2007), 67 – 90

Duffy, Bobby and others, *Divided Britain? Polarisation and Fragmentation Trends in the UK*, (London: The Policy Institute, 2019)

Dulakshi Karunasingha, ‘Root Mean Square Error or Mean Absolute Error? Use Their Ratio as Well’, *Information Sciences*, 585 (2022), 609 – 629

Durand, Claire and André Blais, ‘Quebec 2018: A Failure of the Polls?’, *Canadian Journal of Political Science*, 53.1 (2020), 133 – 150

Durand, Claire and André Blais, ‘Why Did the Polls Go Wrong in the 1998 Quebec Election? The Answer from Post-election Polls’, *Bulletin de Méthodologie Sociologique*, 62 (1999), 43 – 47

Durand, Claire and others, ‘Report of the WAPOR Committee Reviewing the Pre-election Polls in the 2017 Presidential Election in Chile’, *WAPOR* (2018), 1 – 39

Durand, Claire and Timothy P. Johnson, ‘Review: What About Modes? Differences between Modes in the 21st Century’s Electoral Polls across Four Countries’, *Public Opinion Quarterly*, 85.1 (2021), 183 – 222

Durand, Claire, ‘Are Polls Biased Against Quebec Liberals?’, *Policy Options Montreal*, 23.3 (2002), 51 – 56

Durand, Claire, ‘The Polls of the 2007 French Presidential Campaign: Were Lessons Learned from the 2002 Catastrophe?’, *International Journal of Public Opinion Research*, 20.3 (2008), 275 – 298

Durand, Claire, André Blais, and Mylène Larochelle, ‘The Polls in the 2002 French Presidential Election: An Autopsy’, *Public Opinion Quarterly*, 68.4 (2004), 602 – 622

Durand, Claire, André Blais, and Sebastien Vachon, ‘Accounting for Biases in Election Surveys: The Case of the 1998 Quebec Election’, *Journal of Official Statistics*, 18.1 (2002), 25 – 44

Düring, Bertram and Oliver Wright, ‘On a Kinetic Opinion Formation Model for Pre-election Polling’, *Philosophical Transactions of the Royal Society A*, 380.2224 (2022), 1 – 20

Dyson, Simon, ‘Polls Apart? The 1990 Nicaraguan and 1992 British General Elections’, *The Political Quarterly*, 65.4 (1994), 425 – 431

Eberl, Jakob-Moritz, Lena Maria Huber, and Carolina Plescia, ‘A Tale of Firsts: the 2019 Austrian Snap Election’, *West European Politics*, 43.6 (2020), 1350 – 1363

Edwards, Owen Dudley, ‘Remembering the Kennedys’, Journal of American Studies, 18.3 (1984), 405 – 423

Elections Canada, *October 21st 2019 Federal Election: Election Results*, <https://enr.elections.ca/National.aspx?lang=e> [accessed 09/08/2020]

Élections Québec, *General Elections November 30, 1998 – December 14 (Masson)* (2020), <https://www.electionsquebec.qc.ca/english/provincial/election-results/general-elections.php?e=18&s=2#s> [accessed 10/07/2020]

Élections Québec, *General Elections September 12, 1994 – October 24 (Saint-Jean)* (2020),<https://www.electionsquebec.qc.ca/english/provincial/election-results/general-elections.php?e=27&s=2#s> [accessed 10/07/2020]

El-Horbaty, Yahia and Eman Hanafy, ‘Some Estimation Methods and Their Assessment in Multilevel Models: A Review’, *Biostatistics and Biometrics Open Access Journal*, 5 (2018), 1 – 8

Elliot, Graham and Robert P. Lieli, ‘Predicting Binary Outcomes’, *Journal of Econometrics*, 174.1 (2013), 15 – 26

Endersby, J. W., S. E. Galatas, and C. B. Rackaway, ‘Closeness Counts in Canada: Voter Participation in the 1993 and 1997 Federal Elections’, *The Journal of Politics*, 64.2 (2002), 610 – 631

Enns, Peter K. and others, *Why the Polls Missed in 2016: Was it Shy Trump Supporters After All?* (2016), <https://www.washingtonpost.com/news/monkey-cage/wp/2016/12/13/why-the-polls-missed-in-2016-was-it-shy-trump-supporters-after-all/> [accessed 11/06/2020]

Enten, Harry, *Macron Won, But the French Polls Were Way Off* (2017), <https://fivethirtyeight.com/features/macron-won-but-the-french-polls-were-way-off/> [accessed 17/07/2020].

Enten, Harry, *Shy Voters Probably Aren’t Why the Polls Missed Trump* (2016), <https://fivethirtyeight.com/features/shy-voters-probably-arent-why-the-polls-missed-trump/> [accessed 11/06/2020]

Enyedi, Zsolt and Gábor Tóka, ‘The Only Game in Town: Party Politics in Hungary’, in *Party Politics in New Democracies*, ed. by Paul Webb and Stephen White, (Oxford: Oxford University Press, 2007)

Erceg-Hurn David M. and Vikki M. Mirosevich, ‘Modern Robust Statistical Methods: an Easy Way to Maximize the Accuracy and Power of Your Research’, *American Psychologist*, 63.7 (2008), 591 – 601

Erikson, Robert S., ‘Why Do People Vote? Because they are Registered’, *American Politics Research*, 9.3 (1981), 259 – 276

Erikson, Robert S. and Christopher Wlezien, ‘Presidential Polls as a Time Series: The Case of 1996’, *The Public Opinion Quarterly*, 63.2 (1999), 163-177

Erikson, Robert, Costas Panagopoulos, and Christopher Wlezien, ‘Likely (and Unlikely) Voters and the Assessment of Campaign Dynamics’, *Public Opinion Quarterly*, 68.4 (2004), 588 – 601

Erikson, Robert S., Costas Panagopoulos, and Christopher Wlezien, ‘The Crystallization of Voter Preferences During the 2008 Presidential Campaign’, *Presidential Studies Quarterly*, 40.3 (2010), 482-496

Esser, Frank, ‘Dimensions of Political News Cultures: Sound Bite and Image Bite News in France, Germany, Great Britain and the United States’, *The International Journal of Press/Politics*, 13.4 (2008), 401 – 428

Evrensel, Astrid, *Voter Registration in Africa: A Comparative Analysis*, (Johannesburg: EISA Publishing, 2010)

Farrell, David M., Jane Suiter, and Clodagh Harris, ‘The Challenge of Reforming a “Voter-friendly” Electoral System: The Debates Over Ireland’s Single Transferable Vote’, *Irish Political Studies*, 32.2 (2017), 293 – 310

Fauvelle-Aymar, Christine, Michael S. Lewis-Beck, and Richard Nadeau, ‘French Electoral Reform and the Abstention Rate’, *Parliamentary Affairs*, 64.1 (2011), 45-60

Fawagreh, Khaled, Mohamed Medhat Gaber, and Eyad Elyan, ‘Random Forests: From Early Developments to Recent Advancements’, *Systems Science and Control Engineering*, 2 (2014), 602 – 609

Felson, Marcus and Seymour Sudman, ‘The Accuracy of Presidential Preference Primary Polls’, *Public Opinion Quarterly*, 39.2 (1975), 232 – 236

Fenwick, Ian and others, ‘Classifying Undecided Voters in Pre-election Polls’, *Public Opinion Quarterly*, 46 (1982), 383 – 391

Feric, Ivana and Vesna Lamza Posavec, ‘Opinion Polls, Voters’ Intentions and Expectations in the 2011 Croatian Parliamentary Elections’, *European Quarterly of Political Attitudes and Mentalities*, 2.4 (2013), 4-15

Fernández, Alberto and others, *Learning from Imbalanced Data Sets* (Cham: Springer, 2018)

Ferrante, Marco and Monica Saltalamacchia, ‘The Coupon Collector’s Problem’, *Material Matemátics*, 2.35 (2014), 1 – 35

Field, H., 'The Accuracy of Pre-election Polls', *Bulletin of the Oxford University Institute of Economics and Statistics*, 13.7 (1951), 201 – 222

Fienberg, Stephen E., 'The Sun-Times Straw Poll, 1968 and 1970: A Statistical Appraisal', *Journal of the American Statistical Association*, 67.338 (1972), 292 – 297

Finkel, Steven E., 'Re-examining the Minimal Effects Model in Recent Presidential Elections', *Journal of Politics*, 55 (1993), 1 – 21

Fisher, Stephen and Rosalind Shorrocks, 'Collective Failure? Lessons from Combining Forecasts for the UK's Referendum on EU Membership', *Journal of Elections, Public Opinion and Parties*, 28.1 (2018), 59 – 77

Fisher, Stephen D. and others, 'From Polls to Votes to Seats: Forecasting the 2010 British Election', *Electoral Studies*, 30 (2011), 250 – 257

FiveThirtyEight, *How FiveThirtyEight's House, Senate and Governor Models Work* (2020), <https://fivethirtyeight.com/methodology/how-fivethirtyeights-house-and-senate-models-work/> [accessed 8 February 2022]

FiveThirtyEight, *Latest Polls* (2020), <https://projects.fivethirtyeight.com/polls/president-general/> [accessed 23 December 2020].

FiveThirtyEight, *Latest Polls* (2022), <https://projects.fivethirtyeight.com/polls/president-general/> [accessed 15/03/2022]

Flickinger, Richard S. and Donley T. Studlar, 'One Europe, Many Electorates? Models of Turnout in European Parliament Elections After 2004', *Comparative Political Studies*, 40.4 (2007), 383-404

Fligner, Michael A. and Timothy J. Killeen, 'Distribution-Free Two-sample Tests for Scale', *Journal of the American Statistical Association*, 71.353 (1976), 210 – 213

Ford, Robert and Matthew Goodwin, *Revolt on the Right: Explaining Support for the Radical Right in Britain* (London: Routledge, 2014)

Ford, Robert and others, 'From Polls to Votes to Seats: Forecasting the 2015 British General Election', *Electoral Studies*, 41 (2016), 244 – 249

Forsberg, Ole J., 'US Election Polls: A Quick Postmortem', *Significance*, 18.1 (2021), 4 – 5

Frederick Solt, 'Does Economic Inequality Depress Electoral Participation? Testing the Schattschneider Hypothesis', *Political Behaviour*, 32 (2010), 285 – 301

Freedman, P. and K. Goldstein, 'Building a Probable Electorate from Pre-election Polls: A Two-stage Approach', *Public Opinion Quarterly*, 60.4 (1996), 574 – 587

Freiberg, Henry C., ‘The Problem of Validity vs. Reliability in Public Opinion Polls’, *Public Opinion Quarterly*, 6.1 (1942), 87 – 98

Friedenson, Bernard, ‘Dewey Defeats Truman and Cancer Statistics’, *Journal of the National Cancer Institute*, 101.16 (2009), 1157 – 1157

Fuchs, Marek, ‘Total Survey Error (TSE)’, in *Encyclopedia of Survey Research Methods*, ed. by Paul J. Lavrakas, (Thousand Oaks, Sage Publications, 2008)

Fumagalli, Laura and Emanuela Sala, ‘The Total Survey Error Paradigm and Pre-election Polls: The Case of the 2006 Italian General Elections’, *ISER Working Paper Series*, 29 (2011), 1 – 41

Gallagher, Michael, ‘Electoral Systems and Voting Behaviour’, in *Development in West European Politics*, ed. by Martin Rhodes, Paul Heywood, and Vincent Wright (New York: St. Martin’s Press, 1997)

Gallagher, Michael, *Irish Elections 1948-77: Results and Analysis* (Oxford: Routledge, 2009)

Gallego, Aina, ‘Understanding Unequal Turnout: Education and Voting in Comparative Perspective’, *Electoral Studies*, 29.2 (2010), 239 – 247

Gallup, *Gallup Presidential Election Trial-Heat Trends, 1936 – 2008* (2008), <https://news.gallup.com/poll/110548/gallup-presidential-election-trial-heat-trends.aspx> [accessed 21/05/2020]

Gallup, George and Saul Forbes Rae, *The Pulse of Democracy: The Public Opinion Poll and How it Works*, (New York: Simon and Schuster, 1940)

Gallup, George and Saul Rae, *The Pulse of Democracy: The Public-Opinion Poll and How It Works*, (New York: Simon and Schuster, 1940)

Gallup, George H., *The Gallup Poll: Public Opinion 1935 – 1971. Volume One: 1935 – 1948*, (Random House: New York, 1972)

Gallup, George, ‘Polls and the Political Process – Past, Present, and Future’, *Public Opinion Quarterly*, 29.4 (1965), 544 – 549

Gallup, George, ‘Pollsters, Not Prophets’, *Society* (13 (1976), 19 – 23

Gallup, George, ‘The Gallup Poll and the 1950 Election’, *The Public Opinion Quarterly*, 15.1 (1951), 16 – 22

Gallup, George, *The Sophisticated Poll Watcher’s Guide*, (Princeton: Princeton Opinion Press, 1972)

Gattermann, Katjana and Claes H. De Vreese, ‘The role of candidate evaluations in the 2014 European Parliament elections: Towards the personalization of voting behaviour?’, *European Union Politics*, 18.3 (2017), 447-468

Geer, John Gray, *Public Opinion and Polling Around the World: A Historical Encyclopaedia*, (Santa Barbara: ABC-CLIO, 2004)

Gelman, Andrew and Gary King, ‘Why Are American Presidential Election Campaign Polls So Variable When Votes are so Predictable?’, *British Journal of Political Science*, 23.4 (1993), 409 – 551

Gelman, Andrew and Jennifer Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*, (Cambridge: Cambridge University Press, 2006)

Gelman, Andrew, ‘Failure and Success in Political Polling and Election Forecasting’, *Statistics and Public Policy*, 8.1 (2021), 67 – 72

Gelman, Andrew, ‘Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper)’, *Bayesian Analysis*, 1 (2006), 515 – 534

German Longitudinal Election Study, *Download Data* (2022), <https://gles-en.eu/download-data/> [accessed 05/03/2022]

Gherghina, Sergiu and Mihail Chiru, ‘Determinants of Legislative Voting Loyalty Under Different Electoral Systems: Evidence from Romania’, *International Political Science Review*, 35.5 (2014), 523 – 541

Gibbins, Roger, ‘Another New West: Environmentalism and the New Policy Agenda’, in *How Ottawa Spends 1991-92: The Politics of Fragmentation*, ed. by Frances Abele, (Ottawa: Carleton University Press, 1991)

Gidengil, Elisabeth and others, ‘Priming and campaign context: Evidence from recent Canadian elections’ in Do Political Campaigns Matter? Campaign Effects in Elections and Referendums ed. by David M. Farrell and Rudiger Schmitt-Beck, (London: Routledge, 2002), 76-92

Giugal, Aurelian and others, ‘Reforming an Electoral System – An Experiment That Failed: Romania 2008-2012’, *Representation*, 56.1 (2020), 111-126

Giuliani, Marco, ‘Making Sense of Pollsters’ Errors. An Analysis of the 2014 Second-order European Election Predictions’, *Journal of Elections, Public Opinions and Parties* 29.2 (2019), 162 – 178

Global Barometer Surveys, *Surveys* (2022), <https://www.globalbarometer.net/survey_do> [accessed 05/03/2022]

Gold, Howard J., ‘Third Party Voting in Presidential Elections: A Study of Perot, Anderson, and Wallace’, *Political Research Quarterly*, 48.4 (1995), 751 – 773

Goldman, Eric F., ‘Poll on the Polls’, *The Public Opinion Quarterly*, 8.4 (1945), 461 – 467

Goldstein, Harvey, William Browne, and Jon Rasbash, ‘Partitioning Variation in Multilevel Models’, *Understanding Statistics*, 1.4 (2002), 223 – 231

Goldstein, Henry, *Multilevel Statistical Models: Fourth Edition*, (New Jersey: John Wiley and Sons)

Golosov, G. V., ‘The Effective Number of Parties: A New Approach’, *Party Politics*, 16.2 (2010), 171 – 19

Gong, Joonho and Hyunjoong Kim, ‘RHSBoost: Improving Classification Performance in Imbalance Data’, *Computational Statistics and Data Analysis*, 111 (2017), 1 – 13

Goodin, Robert and John Dryzek, ‘Rational Participation: The Politics of Relative Power’, *British Journal of Political Science*, 10.3 (1980), 273 – 292

Goot, Murray, *Did Late Deciders Confound the Polls?* (2019), <https://insidestory.org.au/did-late-deciders-confound-the-polls/> [accessed 20/08/2022]

Goot, Murray, ‘How Good are the Polls? Australian Election Predictions, 1993 – 2019’, *Australian Journal of Political Science*, 56.1 (2021), 35 – 55

Goot, Murray, ‘Labor’s 1943 Landslide: Political Market Research, Evatt, and the Public Opinion Polls’, *Labour History*, 106 (2014), 149 – 166

Gopoian, J. David and Sissie Hadjiharalambous, ‘Late-deciding voters in presidential elections’, *Political Behavior*, 16 (1994), 55-78

Gorard, Stephen, ‘What is Multi-level Modelling For?’, *British Journal of Educational Studies*, 51.1 (2003), 46 – 63

Górecki, Maciej A. and Michal Pierzgalski, ‘Legislated Candidate Quotas and Women’s Descriptive Representation in Preferential Voting Systems’, *European Journal of Political Research*, 61 (2022), 154 – 174

Gosnell, Harold F. and Sebastian de Grazia, ‘A Critique of Polling Methods’, *The Public Opinion Quarterly*, 6.3 (1942), 378-390

Greaves, Lara M. and others, ‘How low can we go? Declining survey response rates to New Zealand electoral roll mail surveys over three decades’, *Political Science*, 72.3 (2020), 228-244

Green, Donald P., Alan S. Gerber, and Suzanna L. De Boef, ‘Tracking Opinion Over Time: A Method for Reducing Sampling Error’, *Public Opinion Quarterly*, 63.2 (1999), 178 – 192

Greene, Kenneth F., ‘Campaign Effects and the Elusive Swing Voter in Modern Machine Politics’, *Comparative Political Studies*, 54.1 (2021), 77 – 109

Groves, Robert and Emilia Peytcheva, ‘The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-analysis’, *Public Opinion Quarterly*, 72.2 (2008), 167 – 189

Groves, Robert M. and Lars Lyberg, ‘Total Survey Error: Past, Present, and Future’, *Public Opinion Quarterly*, 74.5 (2010), 849 – 879

Groves, Robert M., *Survey Errors and Survey Costs*, (Hoboken: John Wiley and Sons, 2004)

Hainsworth, Paul, ‘The Return of the Left: the 1997 French Parliamentary Election’, *Parliamentary Affairs*, 51.1 (1998), 71 – 83

Halcoussis, Dennis, Anton D. Lowenberg, and G. Michael Phillips, ‘An Empirical Test of the Comey Effect on the 2016 Presidential Election’, *Social Science Quarterly*, 101.1 (2019), 161 – 171

Hancock, John T., and Taghi M. Khoshogftaar, ‘Survey on Categorical Data for Neural Networks’, *Journal of Big Data*, 7.1 (2020), 1 – 41

Haner, Charles F. and Norman C. Meir, ‘The Adaptability of Area-Probability Sampling to Public Opinion Measurement’, *Public Opinion Quarterly*, 15.2 (1951), 335 – 352

Harel, Ofer and others, ‘Multiple Imputation for Incomplete Epidemiologic Studies’, *American Journal of Epidemiology*, 187.3 (2018), 576 – 584

Hargittai, Eszter and Gokce Karaoglu, ‘Biases of Online Political Polls: Who Participates?’, *Socius: Sociological Research for a Dynamic World*, 4 (2018), 1 – 7

Harris, Louis, ‘Election Polling and Research’, *Public Opinion Quarterly*, 21.1 (1957), 108 – 116

Harris, Louis, *Reagan Now Far Behind Carter as the President’s Lead Continues to Grow*, *The Chicago Tribune*, 3 January 1980, ABC News-Harris Survey, pp. 1 – 3

Hawkins, Douglas M., ‘The Problem of Overfitting’, *Journal of Chemical Information and Computer Sciences*, 44 (2004), 1 – 12

Haynes, Fred E., ‘The Significance of the Latest Third-party Movement’, *The Mississippi Valley Historical Review*, 12.2 (1925), 177 – 186

Heerwegh, D., 'Mode Differences Between Face-to-Face and Web Surveys: An Experimental Investigation of Data Quality and Social Desirability Effects', *International Journal of Public Opinion Research*, 21.1 (2009), 111 – 121

Hellwig, Timothy and Eva Coffey, 'Public Opinion, Party Messages, and Responsibility for the Financial Crisis in Britain', *Electoral Studies*, 30.3 (2011), 417-426

Henderson, Michael, 'Finding the Way Home: The Dynamics of Partisan Support in Presidential Campaigns', *Political Behaviour*, 37 (2015), 889 – 910

Hennessy Bernard C., and Erna R. Hennessy, 'The Prediction of Close Elections: Comments on Some 1960 Polls', *The Public Opinion Quarterly*, 25.3 (1961), 405-411

Herbst, Susan 'Surveys in the Public Sphere: Applying Bourdieu's Critique of Opinion Polls', *International Journal of Public Opinion Research*, 4.3 (1992), 220 – 229

Herbst, Susan, *Numbered Voices: How Opinion Polling Has Shaped American Politics*, (Chicago: University of Chicago Press, 1995)

Herrnson, Paul S., and James M. Curry, 'Issue Voting and Partisan Defections in Congressional Elections', *Legislative Studies Quarterly*, 36.2 (2011), 281 – 307

Herr, Richard, 'Partisan Chickens Coming Home to Roost in the 2010 Tasmanian Election: Consequences of the 1998 Reduction in Size of the Parliament, *Public Administration Today*, 22 (2010), 33 – 35

Highfill, Robert D., 'Effects of News of Crime and Scandal upon Public Opinion', *Journal of Criminal Law and Criminology*, 17.1 (1926), 40-103

Hillygus, D. Sunshine and Simon Jackman, 'Voter Decision Making in Election 2000: Campaign Effects, Partisan Activation, and the Clinton Legacy', *American Journal of Political Science*, 47.4 (2003), 583 – 596

Hillygus, D. Sunshine, 'The Evolution of Election Polling in the United States', *Public Opinion Quarterly*, 75 (2011), 962 – 981

Himmelweit, Hilde T. and others, 'Memory of Past Vote: Implications of a Study of Bias in Recall', *British Journal of Political Science*, 8.3 (1978), 365 – 375

Ho, Tin Kam, 'Random Decision Forests', *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, 1 (1995), 278 – 282

Hodder-Williams, Richard, *Public Opinion Polls and British Politics*, (London: Routledge, 1970)

Hoek. Janet A. and Philip J. Gendall, ‘A New Method of Predicting Voting Behaviour’, *International Journal of Market Research*, 35.4 (1993), 1 – 14

Holbrook, Thomas M., *Do Campaigns Matter?,* (Thousand Oaks: Sage, 1996)

Holtz-Bacha, Christina, ‘Polls, Media and the Political System’, in *Opinion Polls and the Media: Reflecting and Shaping Public Opinion* ed. by Christina Holtz-Bacha and Jesper Stromback (Basingstoke: Palgrave Macmillan, 2012), 267-281

Hopkins, David A., *Red Fighting Blue: How Geography and Electoral Rules Polarize American Politics*, (Cambridge: Cambridge University Press, 2017)

Hox, Joop J. Hox, Mirjam Moerbeek, and Rens van de Schoot, *Multilevel Analysis: Techniques and Applications*, (New York: Routledge, 2017)

Høyland, Bjørn and Martin G. Søyland, ‘Electoral Reform and Parliamentary Debates’, *Legislative Studies Quarterly*, 44.4 (2019), 593 – 615

Hug, Simon, *Altering Party Systems: Strategic Behaviour and the Emergence of New Political Parties in Western Democracies* (Ann Arbor: The University of Michigan Press, 2001)

Hughes, Rachael A. and others, ‘Accounting for Missing Data in Statistical Analyses: Multiple Imputation is Not Always the Answer’, *International Journal of Epidemiology*, 48.4 (2019), 1294 – 1304

Hutcheson, Derek S., and Elena A. Korosteleva, ‘Patterns of Participation in Post-Soviet Politics’, *Comparative European Politics*, 4 (2006), 23 – 46

Hyndman, Rob J. and Anne B. Koehler, ‘Another Look at Measures of Forecast Accuracy’, *International Journal of Forecasting*, 22.4 (2006), 679 – 688

Hyndman, Rob J. and George Anthanasopoulos, *Forecasting: Principles and Practice,* (Melbourne: Otexts, 2018)

Ibenskas Raimondas and Allan Sikk, ‘Patterns of party change in Central and Eastern Europe, 1990-2015’, *Party Politics*, 23.1 (2017), 43-54

Ignazi, P., ‘The Crisis of Parties and the Rise of New Political Parties’, *Party Politics*, 2.4 (1996), 549-566

Igo, Sarah E., ‘”A Gold Mine and a Tool for Democracy”: George Gallup, Elmo Roper, and the Business of Scientific Polling, 1935 – 1955’, *Journal of the History of the Behavioral Sciences*, 42.2 (2006), 109 – 134

Immerzeel, Tim and Mark Pickup, ‘Populist radical right parties mobilizing ‘the people’? The role of populist radical right success in voter turnout’, *Electoral Studies*, 40 (2015), 347-360

Interfax, *Electoral Code Becomes Effective in Ukraine* (2020), <https://en.interfax.com.ua/news/general/633561.html> [accessed 24/08/2023].

International IDEA, *Voter Turnout Database* (2022), <https://www.idea.int/data-tools/world-view/40> [accessed 17/08/2022].

Inter-parliamentary Union, *France Parliamentary Chamber: Assemblée Nationale Elections Held in 1997* (1997), <http://archive.ipu.org/parline-e/reports/arc/2113_97.htm> [accessed 13/07/2020]

Ipsos, *Political Monitor Archive* (2022), <https://www.ipsos.com/en-uk/political-monitor-archive> [accessed 12/09/2022]

Irish Statute Book, Electoral Act, 1992 (2023), <https://www.irishstatutebook.ie/eli/1992/act/23/section/39/enacted/en/html> [accessed 04/08/2023]

Irwin, G. A., and J. J. van Holsteyn, ‘What are they waiting for? Strategic Information for Late Deciding Voters’, *International Journal of Public Opinion Research*, 20.4 (2008), 483-493

Italian National Election Studies, *Data Request* (2022), <http://www.itanes.org/dati/#> [accessed 05/03/2022]

Iyengar, Shanto and Adam Simon, ‘News Coverage of the Gulf Crisis and Public Opinion: A Study of Agenda-setting, Priming, and Framing’, *Communication Research*, 20.3 (1993), 365-383

Jackman, Simon, ‘Pooling the Polls Over an Election Campaign’, *Australian Journal of Political Science*, 40.4 (2005), 499 – 517

Jackman, Simon and Gary N. Marks, ‘Forecasting Australian elections: 1993, and all that’, *Australian Journal of Political Science*, 29.2 (1994), 277-291

Jackson, Natalie and Michael S. Lewis-Beck, ‘Causes of 2020 Polling Error’, in *Polarization and Political Party Factions in the 2020 Election*, ed. by Jennifer C. Lucas, Tauna S, Sisco, and Christopher J. Galdieri London: Rowman and Littlefield, 2022)

Jackson, Keith and Alan McRobie, *New Zealand Adopts Proportional Representation: Accident? Design? Evolution?*, (New York: Routledge, 1998)

Jackson, Robert A., ‘Differential Influences on Participation in Midterm Versus Presidential Elections’, *The Social Science Journal*, 37.3 (2000), 385 – 402

Jager, Kai, ‘Not a New Gold Standard: Even Big Data Cannot Predict the Future’, *Critical Review*, 28.3 (2016), 335 – 355

Jennings, Will and Christopher Wlezien, ‘Election Polling Errors Across Time and Space’, *Nature Human Behaviour*, 2.4 (2018), 276 – 283

Jennings, Will and Christopher Wlezien, ‘The Timeline of Elections: A Comparative Perspective’, *American Journal of Political Science*, 60 (2016), 219 – 233

Jennings, Will and Christopher Wlezien, *Replication Data for: Election Polling Errors Across Time and Space* (2018), <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/8421DX> [accessed 31 January, 2022]

Jennings, Will, ‘The Polls in 2017’ in *Political Communication in Britain*, ed. by D. Wring, R. Mortimore, and S. Atkinson, (London: Palgrave MacMillan, 2019)

Jennings, Will, Michael Lewis-Beck, and Christopher Wlezien, ‘Election Forecasting: Too Far Out?’, *International Journal of Forecasting*, 36.3 (2020), 949 – 962

Jérôme, Bruno, Véronique Jérôme, and Michael S. Lewis-Beck, ‘Polls Fail in France: Forecasts of the 1997 Legislative Election’, *International Journal of Forecasting*, 15 (1999), 163 – 174

Jervis, Robert, *Why Intelligence Fails: Lessons from the Iranian Revolution and the Iraq War*, (New York: Cornell University Press, 2011)

Jiang, Gaoxia and Wenjian Wang, ‘Error Estimation Based on Variance Analysis of K-fold Cross-validation’, *Pattern Recognition* 69 (2017), 94 – 106

Johnson, Wendell, *People in Quandaries: The Semantics of Personal Adjustment*, (New York: Harper and Row, 1946)

Johnston, Ron and others, ‘Was the 2016 United States’ presidential contest a deviating election? Continuity and change in the electoral map – or “Plus ça change, plus ç’est la mème géographie”’, *Journal of Elections, Public Opinion and Parties*, 27.4 (2017), 369 – 388

Jolly, Seth and others, ‘Chapel Hill Survey Trend File, 1999 – 2019’, *Electoral Studies*, 75 (2022), 102420

Jones, Bill and Dennis Kavanagh, *British Politics Today*, (Manchester: Manchester University Press, 2003)

Jones, Mark P., ‘Electoral Laws and the Effective Number of Candidates in Presidential Elections’, *The Journal of Politics*, 61.1 (1999), 171 – 184

Jones, Mark P., *Electoral Laws and the Survival of Presidential Democracies* (Notre Dame: University of Notre Dame Press, 1995)

Jordan, David M., *FDR, Dewey, and the Election of 1944*, (Bloomington: Indiana University Press, 2011)

José da Cruz Belo, António, ‘Accuracy of Polls in Portugal’, paper delivered at the WAPOR 58th Annual Conference (2005)

Jowell, Roger and others, ‘The 1992 British Election: The Failure of the Polls’, *Public Opinion Quarterly*, 57.2 (1993), 238 – 263

Jusko, Karen Long, ‘Systematic Error’, in *Encyclopedia of Survey Research Methods*, ed. by Paul J. Lavrakas, (Thousand Oaks, Sage Publications, 2008)

Kac, Mark, ‘Some Reflections of a Mathematician on the Nature and Role of Statistics’, *Advances in Applied Probability*, 7 (1975), 5 – 11

Kaid, Lynda Lee and Jesper Strömbäck, “Election News Coverage Around the World: A Comparative Perspective”, *The Handbook of Election News Coverage Around the World* (New York: Routledge, 2009)

Kantar, *Kantar Public UK Polling Archive* (2022), <https://www.kantar.com/expertise/policy-society/kantar-public-uk-polling-archive> [accessed 12/09/2022]

Kaplan, Edward and Arnold Barnett, ‘A New Approach to Estimating the Probability of Winning the Presidency’, *Operations Research*, 51.1 (2003), 32-40

Karlson, Rune, Georg Lutz, and Patrik Ohberg, ‘Candidate Campaigns in Comparative Perspective’, in *Parliamentary Candidates Between Voters and Parties: A Comparative Perspective*, ed. by Lieven de Winter, Rune Karlsen, and Hermann Schmitt (Oxford: Routledge, 2021)

Kasprzyk. Daniel, and Lee Giesbrecht, ‘Reporting Sources of Error in U.S. Federal Government Surveys’, *Journal of Official Statistics*, 19.4 (2003), 343 – 363

Katz, Daniel and Hadley Cantril, ‘Public Opinion Polls’, *Sociometry*, 1 (1937), 155-179

Katz, Daniel, ‘The Polls and the 1944 Election’, *Public Opinion Quarterly*, 8.4 (1944), 468 – 482

Katz, Daniel, 'The Public Opinion Polls and the 1940 Election', *Public Opinion Quarterly*, 5.1 (1941), 52 – 78

Katz, Richard S., 'Reforming the Italian Electoral Law, 1993', in *Mixed-Member Electoral Systems: The Best of Both Worlds*, ed. by Matthew Soberg Shugart and Martin P. Wattenberg, (Oxford: Oxford University Press, 2003), 96-112

Kaur, Amandeep and Robin Kumar, 'Comparative Analysis of Parametric and Non-parametric Tests', *Journal of Computer and Mathematical Sciences*, 6.6 (2015), 336 – 342

Keeter, Scott and others, 'Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey', *Public Opinion Quarterly*, 70.5 (2006), 759 – 779

Kellner, Peter, *General Election Polls 2017: How the Pollsters Got it Wrong* (2017), <https://www.standard.co.uk/news/politics/general-election-polls-how-the-pollsters-got-it-wrong-a3560936.html> [accessed 11/06/2020]

Kendall-Taylor, Andrea, Natasha Lindstaedt, and Erica Frantz, *Democracies and Authoritarian Regimes* (Oxford: Oxford University Press, 2019)

Kenett, Ron S. and Thomas C. Redman, *The Real Work of Data Science: Turning Data into Information, Better Decisions, and Stronger Organisations*, (Chichester: Wiley, 2019)

Kenett, Ron S., Danny Pfeffermann, and David M. Steinberg, 'Election Polls – A Survey, A Critique, and Proposals', *Annual Review of Statistics and its Application*, 5.1 (2018), 1 – 24

Kenett, Ron, Danny Pfeffermann, and David Steingberg, 'Election Polls – A Survey, A Critique, and Proposals', *Annual Review of Statistics and Its Application*, 5 (2018), 1 – 24

Kennedy, Courtney and others, 'An Evaluation of the 2016 Election Polls in the United States', *Public Opinion Quarterly*, 82.1 (2018), 1 – 33

Kennedy, Fiachra, 'The 2002 General Election in Ireland', *Irish Political Studies*, 17.2 (2002), 95 – 106

Kennedy, Ryan, Stefan Wojcik, and David Lazer, 'Improving Election Prediction Internationally', Science, 355.6324 (2017), 515 – 520

Kennedy, Ryan, Stefan Wojcik, and David Lazer, Supplementary Materials for Improving Election Prediction Internationally (2017), <https://www.science.org/doi/suppl/10.1126/science.aal2887/suppl_file/kennedy.sm.pdf> [accessed 14/08/2022]

Kim, Hae-Young Kim, ‘Analysis of Variance (ANOVA): Comparing Means of More Than Two Groups’, *Restorative Dentistry & Endodontics*, 39.1 (2014), 74 – 77

Kim, Heemin, and Richard C. Fording, ‘Voter Ideology in Western Democracies: An Update’, *European Journal of Political Research*, 42 (2003), 95 – 105

Kim, Ji-Hyun., ‘Estimating Classification Error Rate: Repeated Cross-Validation, Repeated Hold-out and Bootstrap’, *Computation Statistics and Data Analysis*, 53.11 (2009), 3735 – 3745

Kimball, Spencer, ‘2016 Presidential State-wide Polling – A Substandard Performance: A Proposal and Application for Evaluating Pre-election Poll Accuracy’, *American Behavioral Scientist*, 63.7 (2019), 768 – 788

King, Gary and Langche Zeng, ‘Logistic Regression in Rare Events Data’, *Political Analysis*. 9.2 (2001), 137 – 163

King, Gary, Robert O. Keohane, Sidney Verba, *Designing Social Inquiry*, (Princeton: Princeton University Press, 1994)

Kirkpatrick S., ‘Political Attitudes and Behaviour: Some Consequences of Attitudinal Ordering’, in *Political Attitudes & Public Opinion* ed. by D. Nimmo and C. Bonjean, (New York: David McKay Company, 1972), 386-404

Knapp, Andrew, ‘Proportional but Bipolar: France’s Electoral System in 1986’, *West European Politics*, 10.1 (1987), 89-114

Koff, Stephen P., ‘Public Opinion Polls and the 1970 British Election’, *Il Politico*, 37.3 (1972), 466 – 482

Kohut, Andrew, ‘A Review of the Gallup Pre-election Methodology in 1980’, *The Proceedings of the Survey Research Methods Section of the American Statistical Association*, 4 (1981), 41 – 46

Kohut, Andrew, ‘Illinois Politics Confounds the Polls’, *Public Opinion*, 5 (1983), 42 – 43

Kondylis, Athanassios, Ali S. Hadi, and Mark Werner, ‘The BACON Approach for Rank-deficient Data’, *Statistics in the Twenty-First Century*, 8.3 (2012), 359 – 379

Kou, S. G. and Michael E. Sobel, ‘Forecasting the Vote: A Theoretical Comparison of Election Markets and Public Opinion Polls’, *Political Analysis*, 12 (2004), 277-295

Kovács, Zoltán and Gyorgy Vida, ‘Geography of the New Electoral System and Changing Voting Patterns in Hungary’, *Acta Geobalcanica*, 1.2 (2015), 55-64

Kraska-Miller, M., *Nonparametric Statistics for Social and Behavioral Sciences*, (Boca Raton: CRC Press, 2014)

Kraut, Robert E. and John B. McConahay, ‘How Being Interviewed Affects Voting: An Experiment’, *Public Opinion Quarterly*, 37.3 (1973), 398 – 406

Krishna, C. S., *C. S. Krishna: The Problem with Pre-poll Surveys* (2014), <https://www.business-standard.com/article/opinion/c-s-krishna-the-problem-with-pre-poll-surveys-114051000919_1.html> [accessed 27/05/2022]

Krotki, Karol, ‘Sampling Error’, in *Encyclopedia of Survey Research Methods*, ed. by Paul J. Lavrakas, (Thousand Oaks, Sage Publications, 2008)

Krzywinski, Martin, Naomi Altman, and Paul Blainey, ‘Nested Designs’, *Nature Methods*, 11 (2014), 977 – 978

Kuhn, Max and Kjell Johnson, *Applied Predictive Modelling*, (New York: Springer, 2013)

Kuhn, Max, ‘Building Predictive Models in R Using the Caret Package’, *Journal of Statistical Software*, 28.5 (2008), 1 – 26

Kuhn, Raymond, ‘The French Presidential and Parliamentary Elections, 2002’, *Representation* 39.1 (2002), 44 – 56

Kuhn, Thomas, *The Structure of Scientific Revolutions*’, (Chicago: The University of Chicago Press, 1962), p. 121.

Kumar, Pradeep and others, ‘Classification of Imbalanced Data: Review of Methods and Applications’, *IOP Conference Series: Materials Science and Engineering*, 1099 (2021), 1 – 20

Kyiv Post, *Ukrainian parliament reduces presidential campaign to 90 days* (2009), <https://archive.kyivpost.com/article/content/ukraine-politics/ukrainian-parliament-reduces-presidential-campaign-45841.html> [accessed 04/08/2023]

Laakso, Markku and Rein Taagepera, ‘Effective Number of Parties: A Measure with Application to West Europe’, *Comparative Political Studies*, 1979 (12.1), 3 – 27

Ladd, Everett Carll, ‘The Brittle Mandate: Electoral Dealignment and the 1980 Presidential Election’, *Political Science Quarterly*, 96.1 (1981), 1 – 25

Ladd, Everett Carll., ‘The Election Polls: An American Waterloo’, Chronicle of Higher Education, November 22, 1996, A52

Ladd, Everett Carll, ‘The Pollsters’ Waterloo’, Wall Street Journal, November 19, 1996, A22

Ladner, Andreas and Henry Milner, 'Do voters turn out more under proportional than majoritarian systems? The evidence from Swiss communal elections', *Electoral Studies*, 18.2 (1999), 235-250

Lago, Ignacio and others, *MEDW 2016 Spanish National Election Study* (2017), <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/XHBLOT> [accessed 05/03/2022]

Lai, Nicholas, *Clinton Defeats Trump: Polling Failure or Media Failure?* (2018), <https://bpr.berkeley.edu/2018/05/22/clinton-defeats-trump-polling-failure-or-media-failure/> [accessed 25/07/2020]

Laing, Lionel H., 'The Pattern of Canadian Politics: The Elections of 1945', *American Political Science Review*, 4 (1946), 760 – 765

Lakens, Daniel, 'Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for T-tests and ANOVAs', *Frontiers in Psychology*, 4 (2013), 863

Lanf, Kurt and Gladys Engel Lang, 'The Impact of Polls on Public Opinion', *The Annals of the American Academy of Political and Social Science*, 472.1 (1984), 129 – 142

Lang, Kurt and Gladys Engel Lang, 'The Impact of Polls on Public Opinion', *The Annals of the American Academy of Political and Social Science*, 472.1 (1984), 129-142

Langer, Gary and Daniel M. Merkle, 'Likely Voter', in *Encyclopedia of Survey Research Methods*, ed. by Paul J. Lavrakas, (Thousand Oaks, Sage Publications, 2008)

Latino Barometer, *Data* (2022), <https://www.latinobarometro.org/latContents.jsp> [accessed 05/03/2022]

Lau, Richard R. and David P. Redlawsk, 'Advantages and Disadvantages of Cognitive Heuristics in Political Decision Making', *American Journal of Political Science*, 45.4 (2001), 951 – 971

Lau, Richard R., 'An Analysis of the Accuracy of Trial Heat Polls During the 1992 Presidential Election', *The Public Opinion Quarterly*, 58.1 (1994), 2 – 20

Laver, Michael and Kenneth Benoit, 'The Evolution of Party Systems between Elections', *American Journal of Political Science*, 47.2 (2003), 215-233

Laver, Michael, Kenneth Benoit, and Nicolas Suager, 'Policy Competition in the 2002 French Legislative and Presidential Elections', *European Journal of Political Research*, 45.4 (2006), 667 – 697

Lavrakas, Paul J., 'Random Error', in *Encyclopedia of Survey Research Methods*, ed. by Paul J, Lavrakas, (Thousand Oaks: Sage Publications, 2008)

Le, Huyen and others, Bumps and Bruises: Mining Presidential Campaign Announcements on Twitter', *Proceedings of the 28th ACM Conferences on Hypertext and Social Media*, 1 (2017), 215-224

Lee, Wonjae and Kangwon Seo, 'Downsampling for Binary Classification with a Highly Imbalanced Dataset Using Active Learning', *Big Data Research*, 28 (2022), 1 – 19

Lemiuex, Peter H., 'Political Issues and Liberal Support in the February 1974, British General Election', *Political Studies*, 25.3 (1977), 323 – 342

Lemoine, Nathan P., 'Moving Beyond Noninformative Priors: Why and How to Choose Weakly Informative Priors in Bayesian Analysis', *Oikos*, 128 (2019), 912 – 928

Leslie Kish, *Survey Sampling*, (New York: John Wiley and Sons, 1965)

Lewis-Beck, Michael S., 'Election Forecasting: Principles and Practice', *British Journal of Politics and International Relations*, 7 (2005), 145-164

Lewis-Beck, Michael S. and Tom W. Rice, 'Forecasting presidential elections: a comparison of naïve models', *Political Behavior*, 6.1 (1984), 9-21

Lewis-Beck, Michael S. and Tom W. Rice, Forecasting U.S. House Elections', *Legislative Studies Quarterly*, 9.3 (1984), 475-486

Li, Fan, Yiming Yang, and Eric P. Xing, 'From Lasso Regression to Feature Vector Machine', *Advances in Neural Information Processing Systems*, 18 (2005), 1 – 8

Li, Xiang, Nandan Sudarsanam, and Daniel D. Frey, 'Regularities in Data from Factorial Experiments', *Complexity*, 11.5 (2006), 32 – 45

Liao, Pei-shan, 'Social Desirability Bias and Mode Effects in the Case of Voting Behaviour', *Bulletin de Methodologie Sociologique*, 132.1 (2016), 73 – 83

Liljequist, David, 'Intraclass Correlation – A Discussion and Demonstration of Basic Features', *PLoS One*, 14.7 (2019), 1 – 35

Lin, Jih-wen, 'How Are the Power of the President Decided? Vote Trading in the Making of Taiwan's Semi-presidential Constitution', *International Political Science Review*, 38.5 (2016), 659 – 672

Lin, Tse-min and Brian Roberts, 'Markets and Politics: The 2000 Taiwanese Presidential Election', in *Topics in Analytical Political Economy*, ed. by Melvin Hinich and William A. Barnett, (Amsterdam: Elsevier, 2007)

Lindstaedt, Natasha, *Democratic Decay and Authoritarian Resurgence* (Bristol: Bristol University Press, 2001)

Lipset, Seymour Martin, ‘The Wavering Polls’, *The Public Interest*, 43 (1976), 70 – 89

Liu, Raymond and Duncan F. Gillies, ‘Overfitting in Linear Feature Extraction for Classification of High-dimensional Image Data’, *Pattern Recognition*, 53 (2016), 73 – 86

Liu, X. Y., J. Wu, and Z. H. Zhou, ‘Exploratory Undersampling for Class-Imbalance Learning’, *IEEE Transactions on Systems, Man, and Cybernetics*, 39.2 (2009), 539 – 550

Lloyd, Ryan and Mathieu Turgeon, ‘Polling in New Democracies and Electoral Malpractice: The Case of Brazil’, *International Journal of Public Opinion Research*, 33.4 (2021), 1039 – 1049

Loewen, Peter John, Henry Milner, and Bruce M. Hicks, ‘Does Compulsory Voting Lead to More Informed and Engaged Citizens? An Experimental Test’, *Canadian Journal of Political Science*, 41.3 (2008), 655 – 672

Lohr, Sharon L., *Sampling: Design and Analysis*, (Boca Raton: CRC Press, 2019)

Lopez-Arevalo, Ivan and others, ‘A Memory-efficient Encoding Method for Processing Mixed-type Data on Machine Learning’, *Entropy*, 22.12 (2020), 1 – 21

Louth, Jonathan and Lisa Hill, ‘Compulsory Voting in Australia: Turnout with and Without it’, *Australian Review of Public Affairs*, 6.1 (2005), 25 – 37

Louwerse, Tom, ‘Improving Opinion Poll Reporting: The Irish Polling Indicator’, *Irish Political Studies*, 31.4 (2016), 541 – 566

Luengo, Óscar G. and Jaime Peláez-Berbell, ‘Exploring the Accuracy of Electoral Polls During Campaigns in 2016: Only Bad Press?’, *Contemporary Social Science*, 14.1 (2017), 43 – 53

Lydgate, William, *What America Thinks*, (New York: Thomas Y. Crowell, 1944)

Lynn, Peter and Roger Jowell, ‘How Might Opinion Polls be Improved? The Case for Probability Sampling’, *Journal of the Royal Statistical Society*, 159.1 (1996), 21 – 28

Lyons, Pat, ‘Public Opinion in the Republic of Ireland – 2002’, *Irish Political Studies*, 18 (2003), 6 – 23

Madsen, Jens Koed, *The Psychology of Micro-targeted Election Campaigns*, (Oxford: Palgrave Macmillan, 2019)

Magalhães, Pedro C., ‘Pre-election Polls in Portugal: Accuracy, Bias, and Sources of Error’, *International Journal of Public Opinion Research*, 17.4 (2005), 399 – 421

Magalhães, Pedro C., 'What Are (Semi)Presidential Elections About? A Case Study of the Portuguese 2006 Elections', *Journal of Elections, Public Opinion and Parties*, 17.3 (2007), 263-291

Maggini, Nicola, 'The evolution of turnout in European elections from 1979 to 2009', in *The European Parliament Elections of 2014*, ed. by Lorenzo De Sio, Vincenzo Emanuele, and Nicola Maggini, (Rome: CISE, 2014), 31 – 36

Mainwaring, Scott and Matthew S. Shugart, 'Juan Linz, Presidentialism, and Democracy: A Critical Appraisal', *Comparative Politics*, 29.4 (1997), 449 – 471

Mainwaring, Scott and Matthew Soberg Shugart, 'Conclusion: Presidentialism and the Party System', in *Presidentialism and Democracy in Latin America*, ed. by Scott Mainwaring and Matthew Soberg Shugart (New York: Cambridge University Press, 1997)

Makridakis, Spyros, 'Forecasting Accuracy and the Assumption of Constancy', *Omega*, 9.3 (1981), 307 – 311

Makridakis, Spyros, Rob J. Hyndman, and Fotios Petropoulos, 'Forecasting in Social Settings: The State of the Art', *International Journal of Forecasting*, 36 (2020), 15 – 28

Makridakis, Spyros, Robin M. Hogarth, and Anil Gaba, 'Why Forecasts Fail. What to Do Instead', *MIT Sloan Management Review*, 51.2 (2010), 83 – 90

Mann, Henry B., 'On a Problem of Estimation Occurring in Public Opinion Polls', *The Annals of Mathematical Statistics*, 16.1 (1945), 85 – 90

Mansillo, Luke and Simon Jackman, 'National Polling and Other Disasters', in *Morrison's Miracle: The 2019 Australian Federal Election*, ed.by Anika Gauja, Marian Sawer, and Marian Simms (Canberra: Australian National University Press, 2020),

Margalit, Dan and Joseph Rabinoff, *Interactive Linear Algebra*, (Atlanta: Georgia Institute of Technology, 2017)

Market Research Society, *Public Opinion Polling on the 1970 Election*, (London: MRS, 1972)

Market Research Society, *Report of the MRS Inquiry Into the 1992 General Election Opinion Polls*, (London: MRS, 1992)

Markus, Gregory, 'Political Attitudes During an Election Year: A Report on the 1980 NES Panel Study', *American Political Science Review*, 76 (1982), 538 – 560

Marthaler, Sally, 'La course au centre: Policy Convergence and Partisanship in France, 1981 – 2002', *French Politics, Culture and Society*, 28.2 (2010), 75 – 95

Marsh, Michael, 'Electoral Context', *Electoral Studies*, 21.2 (2022), 207 – 217

Mattes, Robert, ‘Public Opinion Research in Emerging Democracies’ in *The SAGE Handbook of Public Opinion Research*, ed. by Michael W. Traugott and Wolfgang Donsbach, (London: SAGE Publications, 2007), 113-122

McCutcheon, Allan L., ‘Sampling Bias’, in *Encyclopedia of Survey Research Methods*, ed. by Paul J. Lavrakas, (Thousand Oaks, Sage Publications, 2008)

McDonald, Danny L. and others, *Federal Elections 2000: Elections Results for the U.S. President, the U.S. Senate, and the U.S. House of Representatives*, (Washington D.C., Federal Election Commission, 2001)

McElroy, Gail and Michael Marsh, ‘Why Opinion Polls Got It Wrong in 2002’, in *How Ireland Voted 2002*, ed. by Michael Gallagher, Paul Mitchell, and Michael Marsh, (Basingstoke: Palgrave Macmillan, 2003)

McHugh, Mary L., ‘The Chi-square Test of Independence’, *Biochemia Medica*, 23.2 (2013), 143 – 149

McNeish, Daniel, ‘Estimation Methods for Mixed Logistics Models with Few Clusters’, *Multivariate Behavioral Research*, 51 (2016), 790 – 80

Meisel, John, ‘Unresolved Ambiguity: Quebec After the Election of 1998’, *Government and Opposition*, 34.3 (1999), 333 – 351

Melamed, David and Mike Vuolo, ‘Assessing Differences Between Nested and Cross-classified Hierarchical Models’, *Sociological Methodology*, 49.1 (2019), 220 – 257

Mellon, Jonathan *All You Have to Do is Ask: Measuring Uncertainty in Vote Intention* (2017), <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2958302> [accessed 11/07/2022]

Mellon, Jonathan and Christopher Prosser, ‘Missing Non-voters and Misweighted Samples: Explaining the 2015 Great British Polling Miss’, *Public Opinion Quarterly*, 81.3 (2017), 661 – 687

Mellon, Jonathan and others, ‘Aggregate Turnout is Mismeasured’ (2018), available at SSRN: https://ssrn.com/abstract=3098436 [accessed 23/04/2022]

Meltzer, Allan and Scott Richard, ‘A Rational Theory of the Size of Government’, *Journal of Political Economy*, 89.5 (1981), 914 – 927

Mercaldo, Sarah Fletcher and Jeffrey D. Blume, ‘Missing Data and Prediction: The Pattern Submodel’, *Biostatistics*, 236 – 252

Merkle, Daniel M., ‘Nonresponse Bias’, in in *Encyclopedia of Survey Research Methods*, ed. by Paul J. Lavrakas, (Thousand Oaks, Sage Publications, 2008)

Miller, Arther H., ‘Sex, Politics, and Public Opinion: What Political Scientists Really Learned from the Clinton-Lewinsky Scandal’, *PS: Political Science & Politics*, 32.4 (1999), 721-729

Miller, Mungo, ‘The Waukegan Study of Voter Turnout Prediction’, *Public Opinion Quarterly*, 16.3 (1952), 381 – 398

Miller, Peter V., ‘The Authority and Limitations of Polls’, in *Navigating Public Opinion: Polls, Policy and the Future of American Democracy*, ed. by Jeff Manza, Fay Lomax Cook, and Benjamin J. Page, (New York: Oxford University Press, 2002)

Miller, Peter V., ‘Which Side Are You On? The 1990 Nicaraguan Poll Debacle’, *Public Opinion Quarterly*, 55 (1991), 281 – 302

Minji, Lee, *National Assembly Passes Electoral Reform Amid Opposition Lawmakers' Protest* (2019), <https://en.yna.co.kr/view/AEN20191227003854315> [accessed 24/08/2023]

Miranda, Mario J., and Paul R. Fackler, *Lecture Notes in Computational Economic Dynamics* (1997), <http://fmwww.bc.edu/ec-p/software/Miranda/chapt6.pdf> [accessed 11 February 2022]

MIT Election Data and Science Lab, *Data* (2022), <https://electionlab.mit.edu/data> [accessed 12/09/2022]

Mitofsky, Warren J., ‘The 1980 Pre-election Polls: A Review of Disparate Methods and Results’, *The Proceedings of the Survey Research Methods Section of the American Statistical Association*, 4 (1981), 47 – 52

Mitofsky, Warren J., ‘Was 1996 a Worse Year for Polls than 1948?’, *Public Opinion Quarterly*, 62 (1998), 230 – 249

Moe, Richard, *Roosevelt's Second Act: The Election of 1940 and the Politics of War*, (New York: Oxford University Press)

Moon, Nick, *Opinion Polls: History, Theory and Practice*, (Manchester: Manchester University Press, 1999)

Moore, D. W., *The Superpollsters: How they Measure and Manipulate Public Opinion in America*, (New York: Four Walls Eight Windows, 1992)

Morgan, John and Philip C. Stocken, ‘Information Aggregation in Polls’, *American Economic Review*, 98.3 (2008), 864 – 896

Mortimore, Roger and others, ‘BPC/MRS Enquiry into Election Polling 2015: Ipsos Mori Response and Perspective’, *International Journal of Market Research*, 59.3 (2017), 285 – 300

Moser, Robert G., ‘Electoral Systems and the Number of Parties in Post-Communist States’, *World Politics*, 51.3 (1999), 359 – 384

Mosteller, Frederick and others, *The Pre-election Polls of 1948: Report to the Committee on Analysis of Pre-election Polls and Forecasts*, (New York: Social Science Research Council, 1949)

Mummolo, Jonathan, Erik Peterson, and Sean Westwood, ‘The Limits of Partisan Loyalty’, *Political Behaviour*, 43 (2021), 949-972

Mundfrom, Daniel J., Michelle DePoy Smith, and Lisa W. Kay, ‘The Effect of Multicollinearity on Prediction in Regression Models’, *General Linear Model Journal*, 44.1 (2018), 24 – 28

Murphy, Gary and Theresa Reidy, ‘Presidential Elections in Ireland: From Partisan Predictability to the End of Loyalty’, *Irish Political Studies*, 27.4 (2012), 615 – 634

Murray, G. R., C. Riley, and A. Scime, ‘Pre-election Polling: Identifying Likely Voters Using Iterative Expert Data Mining’, *Public Opinion Quarterly*, 73.1 (2009), 159 – 171

Nagal, Jack, ‘Voter Turnout in New Zealand General Elections, 1928-1988’, *Political Science*, 40.2 (1988), 16-38

Nakagawa, Shinichi, Paul C. D. Johnson, and Holger Schielzeth, ‘The Coefficient of Determination R^2 and Intra-class Correlation Coefficient from Generalized Linear Mixed-Effects Models Revisited and Expanded’, *Journal of the Royal Society Interface*, 14 (2017), 134

Narangoa, Li, ‘Mongolia in 2011: Resources Bring Friends and Wealth’, *Asian Survey*, 52.1 (2012), 81 – 87

Nardulli, Peter F., Jon K. Dalager, and Donald E. Greco, ‘Voter Turnout in U.S. Presidential Elections: An Historical View and Some Speculation’, *PS: Political Science and Politics*. 29.3 (1996), 480 – 490

New Zealand Election Study, *About the New Zealand Election Study* (2020), <http://www.nzes.org/> [accessed 05/03/2022]

Newell, James L.,‘The Italian Election of May 2006: Myths and Realities’, *West European Politics*, 29.4 (2006), 802 – 813

News 24, *Kibaki: I deserve another term* (2007), <https://web.archive.org/web/20090112031044/http:/www.news24.com/News24/Africa/News/0,,2-11-1447_2193347,00.html> [accessed 04/08/2023]

News Room 1015, *Federal Election 2016: Start of Week 2 of the 8 Week Campaign* (2016), <https://1015fm.com.au/2016/05/federal-election-2016-start-of-week-2-of-the-8-week-campaign/> [accessed 04/08/2023]

Nez, Heloise, 'Podemos: The Emergence of a New Political Party in Spain', in *Contemporary Voting in Europe: Patterns and Trends*, ed. By Alexis Chommeloux and Elizabeth Gibson-Morgan (London: Springer Nature, 2017)

Ngufor, Che and Janusz Wojtusiak, 'Learning from Large Distributed Data: A Scaling Down Sampling Scheme for Efficient Data Processing', *International Journal of Machine Learning and Computing*, 4.3 (2013), 216

Nickerson, David W., 'Do Voter Registration Drives Increase Participation? For Whom and When?', *The Journal of Politics*, 77.1 (2014), 88 – 101

Nicolas Sauger, 'Assessing the Accuracy of Polls for the French Presidential Election: The 2007 Experience', *French Politics*, 6 (2008), 116 – 136

Nijman S. W. J., and others, 'Missing Data is Poorly Handled and Reported in Prediction Model Studies Using Machine Learning: A Literature Review', *Journal of Clinical Epidemiology*, 142 (2022), 218 – 229

Noelle-Neumann, Elisabeth, 'About Methodological Progress in Survey Research', *Dialectica*, 16.4 (1962), 307 – 328

Northcott, Robert, 'Opinion Polling and Election Predictions', Philosophy of Science, 82.5 (2015), 1260 – 1271

Norpoth, Helmut, 'To Change or Not to Change Horses: The World War II Elections', *Presidential Studies Quarterly*, 42.2 (2012), 324 – 342

Norris, Pippa, 'Choosing Electoral Systems: Proportional, Majoritarian, and Mixed Systems', *International Political Science Review*, 18.3 (1997), 297 – 312

Norwegian Center for Research Data, *National Election Surveys* (2022), <https://www.nsd.no/nsddata/serier/norske_valgundersokelser_eng.html> [accessed 05/03/2022]

O'Grady, Kevin E., and Deborah R. Medoff, 'Categorical Variables in Multiple Regression: Some Cautions', *Multivariate Behavioural Research*, 23.2 (1988), 243 – 260

Oelbermann, Kai-Friederike, Antonio Palomares, and Friedrich Pukelsheim, 'The 2009 European Parliament Elections: From Votes to Seats in 27 Ways', *European Electoral Studies*, 5.1 (2010), 148 – 182

Oganesyan, Rafael, *Back to the Future? The 2018 Parliamentary Elections and the Armenian Voter* (2018), <https://evnreport.com/elections/back-to-the-future-the-2018-parliamentary-elections-and-the-armenian-voter/?fbclid=IwAR0xmG4C82glJpoDUe6ItAvC9eP54_8T4bLoSq6CluaxpJW7WQOx6QYHSY8> [accessed 05/03/2022]

Ogden, Helen, *Fitting GLMMs with glmmsr* (2018), <https://cran.r-project.org/web/packages/glmmsr/vignettes/glmmsr-vignette.pdf> [accessed 17 February 2022]

Opinium, *Political Polling* (2022), <https://www.opinium.com/resource-center/> [accessed 12/09/2022]

Orr, Graeme and George Willians, 'Electoral Challenges: Judicial Review of Parliamentary Elections in Australia', *Syndey Law Review*, 23 (2001), 54 – 94

Pack, Mark, *Pollbase: Opinion Polls Database from 1943-Today* (2019), <https://www.markpack.org.uk/opinion-polls/> [accessed 23 December 2020]

Paletz, David L. and Others, 'Polls in the Media: Content, Credibility, and Consequences', *Public Opinion Quarterly*, 44 (1980), 495 – 513

Panagopoulos, Costas, 'Polls and Elections Accuracy and Bias in the 2020 U.S. General Election Polls', *Presidential Studies* Quarterly, 51.1 (2021), 214 – 227

Panagopoulos, Costas, 'Pre-election Poll Accuracy in the 2008 General Elections', *Presidential Studies Quarterly*, 39.4 (2009), 896 – 907

Panagopoulos, Costas, Kyle Endres, and Aaron C. Weinschenk, 'Pre-election Poll Accuracy and Bias in the 2016 U.S. General Elections', *Journal of Elections, Public Opinion and Parties*, 28.2 (2018), 157 – 172

Panagopoulos, Costas, Kyle Endres, and Aaron C. Weinschenk, 'Pre-election Poll Accuracy and Bias in the 2016 U.S. General Elections', *Journal of Elections, Public Opinion and Parties*, 28.2 (2018), 157 – 172

Parliament of Canada, *Length of Federal Elections* (2023), <https://web.archive.org/web/20150924131725/http:/www.parl.gc.ca/Parlinfo/Compilations/ElectionsAndRidings/LengthCampaigns.aspx> [accessed 04/08/2022]

Parry, Hugh J. and Helen M. Crossley, 'Validity of Responses to Survey Question', *Public Opinion Quarterly*, 14 (1950), 61 – 80

Parten, Mildred, *Surveys, Polls, and Samples: Practical Procedures*, (New York: Harper and Brothers, 1950)

Patterson, Thomas E., ‘Of Polls, Mountains: U.S. Journalists and Their Use of Election Surveys’, *Public Opinion Quarterly*, 69.5 (2005), 716 – 724

Pederson, Mogens N., ‘The Dynamics of European Party Systems: Changing Patterns of Electoral Volatility’, *European Journal of Political Research,* 7 (1979), 1-26

Pepinsky, Thomas B., ‘A Note on Listwise Deletion Versus Multiple Imputation’, *Political Analysis*, 26.4 (2018), 480 – 488

Perry, Paul, ‘A Comparison of the Voting Preferences of Likely Voters and Likely Non-voters’, *Public Opinion Quarterly*, 37.1 (1973), 99 – 109

Perry, Paul, ‘Certain Problems in Election Survey Methodology’, *Public Opinion Quarterly,* 43.3 (1979), 312 – 325

Perry, Paul, ‘Election Survey Procedures of Gallup Poll’, *Public Opinion Quarterly,* 24 (1960), 531 – 542

Persson, Mikael, Maria Solevid and Richard Öhrvall, ‘Voter Turnout and Political Equality: Testing the ‘Law of Dispersion’ in a Swedish Natural Experiment’, *Politics*, 33.3 (2013), 172 – 184.

Peytcheva, Emilia and Robert M. Groves, ‘Using Variation in Response Rates of Demographic Subgroups as Evidence of Nonresponse Bias in Survey Estimates’, *Journal of Official Statistics*, 25.2 (2009), 193-201

Philpot, Tasha S., Daron R. Shaw, and Ernest B. McGowen, ‘Winning the Race: Black Voter Turnout in the 2008 Presidential Election’, *Public Opinion Quarterly*, 73.5 (2009). 995-1022

Pickup, Mark and others, ‘Why Did the Polls Overestimate Liberal Democrat Support? Sources of Polling Error in the 2010 British General Election’, *Journal of Elections, Public Opinion and Parties*, 21.2 (2011), 179 – 209

Pickup, Mark and Richard Johnston, ‘Campaign Trial Heats as Election Forecasts: Measurement Error and Bias in 2004 Presidential Campaign Polls’, *International Journal of Forecasting*, 24.2 (2008), 272 – 284

Plowman, D. E. G., ‘Public Opinion and the Polls’, *The British Journal of Sociology*. 13.4 (1962), 331 – 349

Polson, Nicholas G., and James G. Scott, ‘On the Half-Cauchy Prior for a Global Scale Parameter’, *Bayesian Analysis* 7.4 (2012), 887 – 902

Popper, Karl R., *Objective Knowledge: An Evolutionary Approach*, (Oxford: Clarendon Press, 1972)

Potdar, Kedar, Taher S. Pardawala, and Chinmay D. Pai, ‘A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers’, *International Journal of Computer Applications*, 175.4 (2017), 7 -9

Press, William H. and others, *Numerical Recipes: The Art of Scientific Computing*, (Cambridge: Cambridge University Press, 2007)

Prosser, Christopher and John Mellon, ‘The Twilight of the Polls? A Review of Trends in Polling Accuracy and the Causes of Polling Misses’, *Government and Opposition*, 53.4 (2018), 757 – 790

Prosser, Christopher and others, ‘Tremors but no Youthquake: Measuring Changes in the Age and Turnout Gradients at the 2015 and 2017 British General Elections’, *Electoral Studies*, 64 (2020), 102129

Puleston, Jon, *Are We Getting Worse at Political Polling?* (2017), <https://shop.esomar.org/uploads/public/events-and-awards/events/2017/congress/documents/ESOMAR-Congress-2017_185446_61_Puleston.pdf> [accessed 16/09/2022].

Queen’s University, *Canadian Opinion Research Archive* (2022), <https://www.queensu.ca/cora/data-holdings> [accessed 05/03/2022]

Rabe-Hesketh, Sophia, Anders Skrondal, and Andrew Pickles, ‘Reliable Estimation of Generalized Linear Mixed Models Using Adaptive Quadrature’, *The Stata Journal*, 2 (2002), 1 – 21

Radchenko, Sergey and Mendee Jargalsaikhan, ‘Mongolia in the 2016-17 Electoral Cycle: The Blessings of Patronage’, *Asian Survey*, 57.6 (2017), 1032 – 1057

Rahat, Gideon, ‘Candidate Selection: The Choice Before the Choice’, *Journal of Democracy*, 18.1 (2007), 157-170

Rallings, Colin and Michael Thrasher, ‘Opinion Polling and the Aftermath of the 1992 General Election’, *Contemporary British History*, 7.1 (1993), 187 – 197

Rami Zeedan, ‘The 2016 US Presidential Elections: What Went Wrong in Pre-election Polls? Demographics Help to Explain’, *J – Multidisciplinary Scientific Journal*, 2.1 (2019), 84 – 101

Ranstam J. and J. A. Cook, ‘LASSO Regression’, *Journal of British Surgery*, 105.10 (2018), 1348 – 1349

Ratcliff, Shaun, ‘Voter Behaviour’, in *Australian Politics and Policy: Senior Edition*, ed. by Peter J. Chen and others, (Sydney: Sydney University Press, 2019)

Rau, Eli G., 'Partisanship as Cause, Not Consequence, of Participation', *Comparative Political Studies*, 55.6 (2022), 1021 – 1058

Raudenbush, Stephen and Anthony Bryk, *Hierarchical Linear Models: Applications and Data Analysis Methods*, (London: Sage, 2002)

Raykov, Tenko, 'Intraclass Correlation Coefficients in Hierarchical Designs: Evaluation Using Latent Variable Modelling', *Structural Equation Modelling*, 18 (2011), 73 – 90

Raymer, James and Philip Howell Rees, 'Framework for Guiding the Development and Improvement of Population Statistics in the United Kingdom', *Journal of Official Statistics*, 31.4 (2015), 699 – 722

Real Clear Politics, General Election: Bush vs. Kerry (2004), <https://www.realclearpolitics.com/epolls/2004/president/us/general_election_bush_vs_kerry-939.html> [accessed 17/07/2020]

Reed, Steven R. and Michael F. Thies, 'The Causes of Electoral Reform in Japan', in *Mixed-Member Electoral Systems: The Best of Both Worlds*, ed. by Matthew Soberg Shugart and Martin P. Wattenberg, (Oxford: Oxford University Press, 2003), 152-172

Rentsch, Anthony, Brian F. Schaffner, and Justin H. Gross, 'The Elusive Likely Voter: Improving Electoral Predictions with More Informed Vote-propensity Models', *Public Opinion Quarterly*, 83.4 (2019), 782 – 804

Renwick, Alan, 'Electoral Reform in Europe since 1945', *West European Politics*, 34.3 (2011), 456-477

Renwick, Alan, Chris Hanretty, and David Hine, 'Partisan Self-interest and Electoral Reform: The New Italian Electoral Law of 2005', *Electoral Studies*, 28.3 (2009), 437-447

Rhee, June Woong, 'How Polls Drive Campaign Coverage: The Gallup/CNN/USA Today's Coverage of the 1992 Presidential Campaign', *Political Communication*, 13.2 (1996), 213 – 229

Riera, Pedro and Francisco Cantú, 'Electoral Systems and Ideological Voting', *European Political Science Review*, (2022), 1 – 19

Rigsarkivet, *The Danish Election Survey 1971 – 2019* (2022), <https://www.sa.dk/en/the-danish-election-survey-1971-2019/> [accessed 05/03/2022]

Rindfuss, Ronald R. and others, 'Do Low Survey Response Rates Bias Results?', *Demographic Research*, 32 (2015), 797 – 828

Rivard, Alex B., 'It is not me, it is you: The emergence of secessionist parties in Western democracies', *Nations and Nationalism*, (2023), 1-20

Robinson, Claude E., ‘Pre-election Polls in the 1942 Elections’, *Public Opinion Quarterly*, 7.1 (1943), 139 – 144

Robinson, Claude E., *Straw Votes: A Study of Political Prediction*, (New York: Columbia University Press, 1932)

Robson, W., ‘The Survey on Nationalisation’, *Political Quarterly*, 30.2 (1959), 111

Rogers, Lindsay, *The Pollsters: Public Opinion, Politics, and Democratic Leadership*, (New York: Alfred A. Knopf, 1949)

Roper, Burns W., ‘Are Polls Accurate?’, *The Annals of the American Academy of Political and Social Science*, 472.1 (1984), 24 – 34

Roper, Burns W., ‘Election Poll Errors: A Problem for Us All’, *AAPOR News*, 10.2 (1983), 1

Rose, Richard, ‘The Polls and Election Forecasting in February 1974’ in H. Penniman (ed.), *Britain at the Polls* (Washington DC: AEI, 1974)

Rose, Richard, *The Polls and the 1970 General Election*, Occasional Paper No. 7, (Strathclyde: University of Strathclyde Survey Research Centre, 1970)

Rosenberg, Alexander, *Philosophy of Social Science*, (Boulder: Westview Press, 1988)

Rosenstone, Steven J., Roy L. Behr, and Edward H. Lazarus, *Third Parties in America: Citizen Response to Major Party Failure*, (Princeton: Princeton University Press, 1984)

Ross, Amanda and Victor L. Willson, *Basic and Advanced Statistical Tests*, (Boston: Sense Publishers, 2017)

Rossi, Richard J., *Mathematical Statistics: An Introduction to Likelihood Based Inference*, (Hoboken: John Wiley and Sons, 2018),

Roy, Jason and Christopher Alcantara, ‘The Election Timing Advantage: Empirical Fact or Fiction?’, *Electoral Studies*, 31.4 (2012), 774 – 781

Royston, Patrick, ‘Multiple Imputation of Missing Values’, *The Stata Journal*, 4.3 (2004), 227 – 241

Russell, Micah and others, ‘Toward a Novel Laser-based Approach for Estimating Snow Interception’, *Remote Sensing*, 12 (2020), 1146 – 1157

Rubin, Donald B., ‘Inference and Missing Data’, *Biometrika*, 63.3 (1976), 581 – 590 (p. 581); Andrew Gelman and Jennifer Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*, (Cambridge: Cambridge University Press, 2006)

Saad, Lydia, *Late Upsets Are Rare, but Have Happened* (2008), <https://news.gallup.com/poll/111451/late-upsets-rare-happened.aspx> [accessed 21/05/2020]

Saaty, Thomas L. and Luis G. Vargas, *Prediction, Projection and Forecasting*, (Boston: Kluwer Academic Publishers, 1991)

Sainani, Kristin L., 'Dealing with Missing Data', *PM&R*, 7.9 (2015), 990 – 994

Sanders, David, 'Pre-election Polling in Britain, 1950-1997', *Electoral Studies*, 22 (2003), 1-20

Sanders, Nada R., 'Measuring Forecast Accuracy: Some Practical Suggestions', *Production and Inventory Management Journal*, 38.1 (1997), 43 – 46

Sawasdee, Siripan Nogsuan, 'Electoral Integrity and the Repercussions of Institutional Manipulations: The 2019 General Election in Thailand', *Asian Journal of Comparative Politics*, 5.1 (2020), 52-68

Schafer, Joseph L., 'Multiple Imputation: A Primer', *Statistical Methods in Medical Research*, 8 (1999), 3 – 15

Schaffer, Cullen 'Overfitting Avoidance as Bias', *Machine Learning*, 10 (1993), 153 – 178

Schaffner, Brian F., and Matthew J. Streb, 'The Partisan Heuristic in Low-information Elections', *Public Opinion Quarterly*, 66 (2002), 559 – 581

Scheiner, Ethan, 'The electoral system and Japan's partial transformation: party system consolidation without policy realignment', *Journal of East Asian Studies*, 12.3 (2012), 351-380

Schielzeth, Holger and Shinichi Nakagawa, 'Nested by Design: Model Fitting and Interpretation in a Mixed Model Era', *Methods in Ecology and Evolution*, 4.1 (2012), 14 – 24

Schmitt-Beck, Rudiger and Julia Partheymuller, 'Why Voters Decide Late: A Simultaneous Test of Old and New Hypotheses at the 2005 and 2009 German Federal Elections', *German Politics*, 21.3 (2012), 299 – 316

Selb, Peter and others, *Bias and Variance in Multiparty Election Polls*, <https://www.polver.uni-konstanz.de/typo3temp/secure_downloads/105568/0/f8a48094815e56ff0112346ba44c2136cc6e487a/Disentangling.pdf> [accessed 12/12/2021]

Seligson, Mitchell A., 'Improving the Quality of Survey Research in Democratizing Countries', *PS: Political Science*, 38.1 (2005), 51-56

Seymour Martin Lipset, ‘Interpreting the Polls’, in *Advances in Consumer Research*, ed. by Beverlee B. Anderson, (Cincinnati: Association for Consumer Research, 1976)

Shamir, Jacob, ‘Pre-election Polls in Israel: Structural Constraints on Accuracy’, *Public Opinion Quarterly*, 50.1 (1986)

Shapiro, Robert Y., ‘What We Relearned and Learned from the 2016 Elections: Comment on Gelman and Azari’, *Statistics and Public Policy*, 4.1 (2017), 1-3

Shaw, Daron R., ‘A Study of Presidential Campaign Event Effects from 1952 to 1992’, *The Journal of Politics*, 61.2 (1999), 387 – 422

Shcherbakov, Maxim Vladimirovich and others, ‘A Survey of Forecast Error Measures’, *World Applied Sciences Journal*, 24 (2013), 171 – 176

Shieh, Gwowen, ‘A Comparison of Two Indices for the Intraclass Correlation Coefficient’, *Behavioural Research Methods*, 44 (2012), 1212 – 1223

Shinn, Glenn, Matt Baker, and Gary Briers, ‘Response Patterns: Effect of Day of Receipt of an E-mailed Survey Instrument on Response Rate, Response Time, and Response Quality’, *Journal of Extension*, 45.2 (2007), 1 – 7

Shirani-Mehr Houshmad and others, ‘Disentangling Bias and Variance in Election Polls’, *Journal of the American Statistical Association 113.552* (2018), 607 – 614

Shrestha, Noora, ‘Detecting Multicollinearity in Regression Analysis’, *American Journal of Applied Mathematics and Statistics*, 8.2 (2020), 39 – 42

Sigelman, Lee and Pamela Johnston Conover, ‘The Dynamics of Presidential Support during International Conflict Situations: The Iranian Hostage Crisis’, *Political Behaviour*, 3.4 (1981), 303 – 318

Silver, Nate, *How FiveThirtyEight’s House Model Works* (2018), <https://fivethirtyeight.com/features/2018-house-forecast-methodology/> [accessed 8 February 2022]

Silver, Nate, *How Our Primary Model Works* (2020), <https://fivethirtyeight.com/features/how-fivethirtyeight-2020-primary-model-works/> [accessed 8 February 2022]

Silver, Nate, *How The FiveThirtyEight Senate Forecast Model Works* (2014), <https://fivethirtyeight.com/features/how-the-fivethirtyeight-senate-forecast-model-works/> [accessed 8 February 2022]

Skalaban, Andrew, ‘Do the Polls Affect Elections? Some 1980 Evidence’, *Political Behaviour*, 10 (1988), 136 – 150

Skibba, Ramin, ‘Why the Polls Missed Trump’, *Nature*, 539.17 (2016), 339

Smith, Craig Allen, *Presidential Campaign Communication: The Quest for the White House*, (Malden: Polity Press, 2010

Smith, T. M. F., ‘Public Opinion Polls: The UK General Election, 1992’, *Journal of the Royal Statistical Society*, 159.3 (1996), 535 – 545

Smith, Tom W., ‘Freedom to Conduct Public Opinion Polls Around the World’, *International Journal of Public Opinion Research*, 16 (2004), 215 – 233

Smith, Tom W., ‘The First Straw? A Study of the Origins of Election Polls’, *Public Opinion Quarterly*, 54.1 (1990), 21-36

Söderlund, Peter, ‘Candidate-centred Electoral Systems and Voter Turnout’, *West European Politics*, 40.3 (2017), 516 – 533

Sohlberg, Jacob and J. Alexander Branham, ‘Just a Difficult Election to Poll? How Context Affects Polling Accuracy’, *Survey Methods: Insights from the Field (SMIF)*, 1 (2020), 1 – 13

Solt, Frederick, ‘Economic Inequality and Democratic Political Engagement’, *American Journal of Political Science*, 52.1 (2008), 48 – 60

Sommet, Nicolas and Davide Morselli, ‘Keep Calm and Learn Multilevel Logistic Modeling: A Simplified Three-step Procedure Using Stata, R, Mplus, and SPSS’, *International Review of Social Psychology*, 30 (2017), 203 – 218

Sontakke. Sumedh and others, ‘Classification of Cardiotocography Signals Using Machine Learning’, in *Intelligent Systems and Applications*, ed. By Kohei Arai, Supriya Kapoor, and Rahul Bhatia, (Cham: Springer, 2018)

Soumbatiants, Souren, Henry W. Chappel Jr., and Eric Johnson, ‘Using State Polls to Forecast U.S. Presidential Election Outcomes’, *Public Choice*, 127 (2006), 207 – 223

South African National Election Study, *Comparative National Elections Project, South Africa 2015* (2020), <https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/603> [accessed 05/03/2022]

Squire, Peverill, ‘Why the Literary Digest Failed’, *Public Opinion Quarterly*, 52.1 (1988), 125 – 133

SSJDA Direct, *Social Science Japan Data Archive* (2022), <https://ssjda.iss.u-tokyo.ac.jp/Direct/> [accessed 05/03/2022]

Stacks, John F., ‘Nation: Where the Polls Went Wrong’, *Time Magazine*, 1 December 1980, pp. 21 – 24

Stadje, Wolfgang, ‘The Collector’s Problem with Group Drawings’, *Advances in Applied Probability*, 22.4 (1990), 866 – 882

Statista, Voter turnout in the presidential elections in France between 1965 and 2022, by round (2023), <https://www.statista.com/statistics/1068866/participation-rate-voter-turnout-presidential-elections-france/> [accessed 16/08/2023]

Stec, Jeffrey A., ‘Bias’, in *Encyclopedia of Survey Research Methods*, ed. by Paul J. Lavrakas, (Thousand Oaks, Sage Publications, 2008)

Steenbergen, Marco R., and Bradford S. Jones, ‘Modelling Multilevel Data Structures’, *American Journal of Political Science*, 46.1 (2002), 218 – 237

Sterne, Jonathan and others, ‘Multiple Imputation for Missing Data in Epidemiological and Clinical Research: Potential and Pitfalls’, *The British Medical Journal*, 338 (2009), 1

Stockton, Hans, ‘How Rules Matter: Electoral Reform in Taiwan’, *Social Science Quarterly*, 91.1 (2010), 21-41

Stoker, Laura, ‘Judging Presidential Character: The Demise of Gary Hart’, *Political Behaviour*, 15.2 (1993), 193 – 223

Stokes, Donald E., ‘A Variance Components Model of Political Effects’, *Mathematical Applications in Political Science*, 1.1 (1965), 61 – 85

Stout, Christopher T. and Reuben Kline, ‘I’m Not Voting for Her: Polling Discrepancies and Female Candidates’, *Political Behaviour*, 33 (2011), 479 – 503

Stoutenborough, James W., ‘Margin of Error (MOE)’, in *Encyclopedia of Survey Research Methods*, ed. by Paul J. Lavrakas, (Thousand Oaks, Sage Publications. 2008)

Stovall, James Glen, ‘Incumbency and News Coverage of the 1980 Presidential Election Campaign’, *The Western Political Quarterly*, 37.4 (1984), 621 – 631

Sturgis, Patrick and others, ‘An Assessment of the Causes of the Errors in the 2015 UK General Election Opinion Polls’, *Journal of the Royal Statistical Society: Statistics in Society Series A*, 181.3 (2018), 757 – 781.

Su, Chien-chou and Mandy Sha, ‘Poll Accuracy Measures in a Quasi-two-way Election: An Empirical Assessment of Taiwan’s 2000 Presidential Election’, *NPUST Humanities and Social Sciences Jikan*, 19.4 (2007), 505 – 527

Sultanov, Murodjon and others, ‘Modelling End-of-season Soil Salinity in Irrigated Agriculture Through Multi-Temporal Optical Remote Sensing, Environmental Parameters, and In Situ Information’, *Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 86.5 (2018), 221 – 233

Susan, Seba and Amitesh Kumar, ‘The Balancing Trick: Optimized Sampling of Imbalanced Datasets – A Brief Survey of the Recent State of the Art’, *Engineering Reports*, 3.4 (2021), 1 – 24

Swedish National Data Service, *Swedish National Election Studies (SNES)* (2022), <https://snd.gu.se/en/catalogue/collection/snes> [accessed 05/03/2022]

SWISSUbase, *Swiss Election Study (Selects) 2019* (2022), <https://www.swissubase.ch/en/catalogue/studies/13846/16586/overview> [accessed 05/03/2022]

Swyngedouw, Mark, *Belgian General Election Study 2007* (2008), <https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:64978> [accessed 05/03/2022]

Székely Gábor J., and Maria L. Rizzo, ‘Brownian Distance Covariance’, *The Annals of Applied Statistics*, 3.4 (2009), 1236 – 1265

Taagepera, Rein, ‘The Number of Parties as a Function of Heterogeneity and Electoral System’, *Comparative Political Studies*, 32.5 (1999), 531 – 548

Taagepera, Rein and Bernard Grofman, ‘Rethinking Duverger’s Law: Predicting the Effective Number of Parties in Plurality and PR Systems – Parties Minus Issues Equals One’, *European Journal of Political Research*, 13 (1985), 341 – 352

Taagepera, Rein and Matthew Soberg Shugart, *Seats and Votes: The Effects and Determinants of Electoral Systems*, (New Haven: Yale University Press, 1989)

Takao, Yasuo, ‘Aging and Political Participation in Japan: The Dankai Generation in a Political Swing’, *Asian Survey*, 49.5 (2009), 852 – 872

Taleb, Nassim Nicholas, *The Black Swan: The Impact of the Highly Improbable*, (New York: Random House, 2007),

Tamman, Maurice, *Clinton Has 90 Percent Chance of Winning: Reuters/Ipsos States of the Nation* (2016), <https://www.reuters.com/article/us-usa-election-poll-idUSKBN1322J1> [accessed 14/07/2022]

Tankard Jr., James W., ‘Public Opinion Polling by Newspapers in the Presidential Election Campaign of 1824’, *Journalism Quarterly*, 49 (1972), 361-365

Tavits, Margit, 'Direct Presidential Elections and Turnout in Parliamentary Contests', *Political Research Quarterly*, 62.1 (2009), 42 – 54

Taylor, Humphrey and others, 'The Record of Internet-based Opinion Polls in Predicting the Results of 72 Races in the November 2000 US Elections', *International Journal of Market Research*, 43.2 (2001), 127 – 135

Teer, F. and J. Spence, *Political Opinion Polls*, (London: Hutchinson, 1973)

Terris, Fay, 'Are Poll Questions Too Difficult?', *Public Opinion Quarterly*, 13.2 (1949), 314-319

Tetlock, Philip E., *Expert Political Judgement: How Good Is It? How Can We Know?*, (Princeton: Princeton University Press, 2005)

The American Presidency Project, *Voter Turnout in Presidential Elections* (2020), <https://www.presidency.ucsb.edu/statistics/data/voter-turnout-in-presidential-elections> [accessed 04/07/2020]

The American Presidency Project, *Voter Turnout in Presidential Elections* (2020), <https://www.presidency.ucsb.edu/statistics/data/voter-turnout-in-presidential-elections> [accessed 30/06/2020]

The Financial Times, *UK General Election Poll Tracker* (2019), < https://www.ft.com/content/263615ca-d873-11e9-8f9b-77216ebe1f17> [accessed 23 December 2020]

The Guardian and ICM, *All Guardian/ICM Poll Results* (2019), https://docs.google.com/spreadsheets/d/1oHcxlAbkTJmqfOxYQM22cvjjjRf5pETIF30x7L-qybc/edit#gid=0 [accessed 19 January 2019]

The Harris Poll, *Poll Archive* (2022), <https://harvardharrispoll.com/> [accessed 12/09/2022]

The Literary Digest, '2,386,052 Straws Forecast Tuesday's Tempest', *The Literary Digest*, 1 November 1924, Topics of the Day, pp. 5-8

The Literary Digest, 'Final Returns in the Digest's Presidential Poll', *The Literary Digest*, 3 November 1928, Topics of the Day, pp. 5-7

The Literary Digest, 'How the Straws Say the Election Will Go', *The Literary Digest*, 23 October 1920, Topics of the Day, p. 14

The Literary Digest, 'Landon, 1,293,669; Roosevelt, 972,897: Final Returns in the Digest's Poll of Ten Million Voters', *The Literary Digest*, 31 October 1936, Topics of the Day, pp. 5-6

The Literary Digest, 'Landon, 1,293,669; Roosevelt, 972,897: Final Returns in the Digest's Poll of Ten Million Voters', *The Literary Digest*, 31 October 1936, Topics of the Day, pp. 5-6.

The Literary Digest, 'Political Reports from 3,000 Communities', *The Literary Digest*, 28 October 1916, Topics of the Day, p. 1087

The Literary Digest, 'Roosevelt Bags 41 States Out of 48', *The Literary Digest*, 5 November 1932, Topics of the Day, pp. 8-9

The Literary Digest, 'What Went Wrong with the Polls?', *The Literary Digest*, 14 November 1936, pp. 7-8

The New York Times, *Iceland to Adopt Electoral Reform* (1959), <https://www.nytimes.com/1959/04/12/archives/iceland-to-adopt-electoral-reform.html> [accessed 24/08/2023]

The New York Times, *Presidential Forecast: Hillary Clinton Has a 91% Chance to Win* (2016), <https://www.nytimes.com/newsgraphics/2016/10/18/presidential-forecast-updates/newsletter.html> [accessed 14/07/2022]

Thies, Clifford F., 'Polls and elections: The Chicago record poll and the election of 1896', *Presidential Studies Quarterly*, 48.1 (2018), 127-138

Thomas J. D., and R. A. Hulquist, 'Interval Estimation for the Unbalanced Case of the One-way Random Effects Model', *Annals of Statistics*, 6 (1978), 582 – 587

Tingsten, Herbert, *Political Behaviour: Studies in Electoral Statistics* (London: P.S. King and Son, 1937)

Tomarken. Andrew J., and Ronald C. Serlin, 'Comparison of ANOVA Alternatives Under Variance Heterogeneity and Specific Noncentrality Structures', *Psychological Bulletin*, 99.1 (1986), 90 – 99

Traugott, Michael W., 'Assessing Poll Performance in the 2000 Campaign', *Public Opinion Quarterly*, 65.3 (2001), 389 – 419

Traugott, Michael W., 'Can We Trust the Polls? It All Depends', *The Brookings Review*, 21.3 (2003), 8-11

Traugott, Michael W., 'The Accuracy of the National Pre-election Polls in the 2004 Presidential Election', *Public Opinion Quarterly*, 69.5 (2005), 642 – 654

Traugott, Michael W. and Clyde Tucker, 'Strategies for Predicting Whether a Citizen Will Vote and Estimation of Electoral Outcomes', *Public Opinion Quarterly*, 48 (1984), 330 – 343

Traugott, Michael W. and Paul J. Lavrakas, *The Voter's Guide to Election Polls*, (Morrisville: Lulu Press, 2016)

Tsakalidis, Adam and others, 'Predicting Elections for Multiple Countries Using Twitter and Polls', *IEEE Intelligent Systems*, 30.2 (2015), 10 – 17

Tudor, Jack and Matthew Wall, 'A Moving Target? An Analysis of Electoral Context and Variation in Polling Accuracy Across Post-war British General Elections', *Journal of Elections, Public Opinion and Parties*, (2021), 1 – 24

Tudor, Jack, *Are Certain Elections More Predictable than Others? A Series of Bivariate Analyses of the Impact of Electoral Characteristics Upon the Predictability of British National Elections* (2017), <https://ssrn.com/abstract=3601958> [accessed 07/04/2021]

Tufte, Edward R., 'The Relationship Between Seats and Votes in Two-party Systems', *The American Political Science Review*, 67.2 (1973), 540 – 554

Tuson, Matthew and others, 'Predicting Future Geographic Hotspots of Potentially Preventable Hospitalisations Using All Subset Model Selection and Repeated k-fold cross validation', *International Journal of Environmental Research and Public Health*, 18 (2021), 1 – 21

Twyman, Joe, 'Getting it Right: YouGov and Online Survey Research in Britain', Journal of Elections, Public Opinion and Parties, 18.4 (2008). 343 – 354

Tyagi, Shivani and Sangeeta Mittal, 'Sampling Approaches for Imbalanced Classification Problem in Machine Learning', in *Proceedings of ICRIC 2019*, ed. by Pradeep Kumar Singh and others (Cham: Springer, 2020)

UK Parliament, *Dissolution of Parliament* (2023), <https://www.parliament.uk/about/how/elections-and-voting/general/dissolution/> [accessed 04/08/2023]

UK Political Info, *1970 General Election Results Summary* (2020), <http://www.ukpolitical.info/1970.htm> [accessed 03/07/2020]

UK Political Info, *1974 February General Election Results Summary*, <http://www.ukpolitical.info/1974Feb.htm> [accessed 09/08/2020]

UK Political Info, *1974 February General Election Results Summary* (2022), <https://www.ukpolitical.info/1974Feb.htm> [accessed 16/07/2022]

UK Political Info, *1974 February General Election Results Summary* (2020), <http://www.ukpolitical.info/1974Feb.htm> [accessed 03/07/2020]

UK Political Info, *1987 General Election Results Summary* (2020), <http://www.ukpolitical.info/1987.htm> [accessed 08/07/2020]

UK Political Info, *1992 General Election Results Summary* (2020), <http://www.ukpolitical.info/1992.htm> [accessed 08/07/2020]

United States Government, *Presidential Election Process* (2023), <https://www.usa.gov/election> [accessed 04/08/2023]

US Department of Justice, *CRS Report for Congress: Croatia 2003 Elections and New Government* (2004), <https://www.justice.gov/file/199811/download> [accessed 04/08/2023]

Valentino, Nicholas A., John Leslie King, and Walter W. Hill, 'Polling and Prediction in the 2016 Presidential Election', *Computer*, 50.5 (2017), 110 – 115

Valentino, Nicholas A. and others, 'Election Night's Alright for Fighting: The Role of Emotions in Political Participation', *The Journal of Politics*, 73.1 (2011), 156-170

van der Meer, T. W. G., H. van der Kolk and R. Rekker, *Dutch Parliamentary Election Study 2017 (DPES/NKO 2017)* (2017), <https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:101156> [accessed 05/03/2022]

van Eck, Liezl, Stephanus E. Visagie, and Hendrik C. de Kock, 'Fairness of Seat Allocation Methods in Proportional Representation', *ORiON*, 21.2 (2005), 93 – 110

van Ginkel, Joost R., and others, 'Rebutting Existing Misconceptions About Multiple Imputation as a Method for Handling Missing Data', *Journal of Personality Assessment*, 120.3 (2020), 297 – 308

Vavreck, Lynn, *The Message Matters: The Economy and Presidential Campaigns*, (New Jersey: Princeton University Press, 2009)

Vignati, Rinaldo and Giancarlo Gasperoni, 'The Predictive Ability of Pre-election Polls in Italy: A Regional Focus', *Statistica Applicata – Italian Journal of Applied Statistics*, 25.3 (2013), 287 – 302

Visa, Sofia and Anca Ralescu, 'Issues in Mining Imbalanced Data Sets – A Review Paper', *Proceedings of the Sixteenth Midwest Artificial Intelligence and Cognitive Science Conference*, 1 (2005), 1 – 7

Visser, Penny S. and others, 'Improving Election Forecasting: Allocation of Undecided Respondents, Identification of Likely Voters, and Response Order Effects', in *Election Polls, the News Media, and Democracy*, ed. by P. Lavrakas and M. Traugott, (New York: Chatham House, 2000)

Voss, D. Stephen, Andrew Gelman, and Gary King, ‘A Review: Pre-election Survey Methodology: Details from Eight Polling Organisations, 1988 and 1992’, *The Public Opinion Quarterly*, 59.1 (1995), 98 – 132

Vowles, Jack, ‘Electoral System Change, Generations, Competitiveness and Turnout in New Zealand, 1963 – 2005’, *British Journal of Political Science*, 40.4 (2010), 875 – 895

Vrigazova, Borislava, ‘The Proportion for Splitting Data into Training and Test Set for the Bootstrap in Classification Problems’, *Business Systems Research*, 12.1 (2021), 228 – 242

Walden, Graham R., *Public Opinion Polls and Survey Research: A Selective Annotated Bibliography of U.S. Guides and Studies from the 1980s*, (New York: Routledge, 2014)

Walther Steven T. and others, *Federal Elections 2016: Election Results for the U.S. President, the U.S. Senate, and the U.S. House of Representatives*, (Washington D.C., Federal Election Commission, 2017)

Wandall, H. H., ‘Medical Education in Denmark’, *Academic Medicine*, 36.9 (1961), 1059 – 1078

Wang, Sam, *Looking Ahead* (2016), <https://election.princeton.edu/articles/looking-ahead/> [accessed 14/07/2022]

Wells, Anthony, *UK Polling Report: Survey and Polling News from YouGov's Anthony Wells* (2019), <https://ukpollingreport.co.uk/> [accessed 23 December 2020].

Wells, Anthony *Understanding Margin of Error* (2011), <https://yougov.co.uk/topics/politics/articles-reports/2011/11/21/understanding-margin-error> [accessed 01/09/2022]

Werker, Claudia and Thomas Brenner, ‘Empirical Calibration of Simulation Models’, *ECIS Working Paper Series*, 200413 (2004), 1 – 30

West, Brady T. and Annelies G. Blom, ‘Explaining Interviewer Effects: A Research Synthesis’, *Journal of Survey Statistics and Methodology*, 5.2 (2017), 175 – 211

Westwood, Sean Jeremy, Solomon Messing, and Yphtach Lelkes, ‘Projecting Confidence: How the Probabilistic Horse Race Confuses and Demobilises the Public’, *The Journal of Politics*, 82.4 (2020), 1530-1544

Wheeler, M., *Lies, Damn Lies, and Statistics: The Manipulation of Public Opinion Polls in America*, (New York: Liveright, 1976)

Whewell, William, The *Philosophy of the Inductive Sciences*, (London: John W. Parker Publishing, 1840)

Whiteley, Paul, ‘Why Did the Polls Get It Wrong in the 2015 General Election? Evaluating the Inquiry into Pre-election Polls’, *The Political Quarterly*, 87.3 (2016), 437-442

Whiteley, Paul F., ‘Forecasting Seats from Votes in British General Elections’, *British Journal of Politics and International Relations*, 7 (2005), 165 – 173

Wijeysundera, Duminda N., and Sindhu R. Johnson, ‘Surveys, Samples, and Botched US Presidential Predictions’, *European Journal of Anaesthesiology*, 29.10 (2012), 462-464

Willocq, Simon, ‘Explaining Time of Vote Decision: The Socio-Structural, Attitudinal, and Contextual Determinants of Late Deciding’, *Political Studies Review*, 17.1 (2018), 53 – 64

Wlezien, Christopher, ‘Presidential Election Polls in 2000: A Study in Dynamics’, *Presidential Studies Quarterly*, 33 (2003), 172 – 187

Wlezien, Christopher and Robert S. Erikson, ‘Campaign Effects in Theory and Practice’, *American Politics Research*, 29.5 (2001), 419 – 436

Wlezien, Christopher and Robert S. Erikson, ‘The Horse Race: What Polls Reveal as the Election Campaign Unfolds’, *International Journal of Public Opinion Research*, 19.1 (2007), 74 – 88

Wlezien, Christopher and Stuart N. Soroka, ‘Electoral Systems and Opinion Representation’, *Representation*, 51.3 (2015), 273-285

Wolak, Mathew E., Daphne J. Fairbairn, and Yale R. Paulsen, ‘Guidelines for Estimating Repeatability’, *Methods in Ecology and Ecology and Evolution*, 3.1 (2012), 129 – 137

Worcester, Robert M., ‘Winning in the Rain’, *New Society*, 31.643 (1975), 394

Worcester, Robert, ‘Political Polling: 95% Expertise and 5% Luck’, Journal of the Royal Statistical Society: Series A (Statistics in Society), 159.1 (1996), 5 – 20

Worcester, Robert, ‘The Polls Have a Lot to Answer For’, *British Public Opinion*, 15 (1992), 1 – 3

Wozniak, Kevin H., Brian R. Calfano, and Kevin M. Drakulich, ‘A “Ferguson Effect” on 2016 Presidential Vote Preference? Findings from a Framing Experiment Examining “Shy Voters” and Cues Related to Policing and Social Unrest’, *Social Science Quarterly*, 100.4 (2019), 1023-1038

Wright, Gerald C., ‘Errors in Measuring Vote Choice in the National Election Studies’, *American Journal of Political Science*, 37.1 (1993), 291 – 316

Wright, Malcolm J., David P. Farrar, and Deborah F. Russell, 'Polling Accuracy in a Multiparty Election', *International Journal of Public Opinion Research*, 26 (2014), 113 – 124

Wright, Stephen G., 'Voter Turnout in Runoff Elections', *The Journal of Politics*, 51.2 (1989), 385-396

Wright, Vincent, 'The French General Election of March 1978: La Divine Surprise', *West European Politics*, 1.3 (1978), 24 – 52

Wu, Bingzhe and others, 'Reducing Overfitting in Deep Convolutional Neural Networks Using Redundancy Regularizer', in *Artificial Neural Networks and Machine Learning – ICANN 2017*, ed. by Alessandra Lintas and others, (Cham: Springer, 2017), 49 – 55.

Ying, Xue, 'An Overview of Overfitting and its Solutions', *Journal of Physics: Conf. Series*, 1168 (2019), 1 – 7

YouGov, *Political Tracker Archive* (2022), <https://yougov.co.uk/topics/politics/explore/topic/Political_tracker_archive> [accessed 12/09/2022]

YouGov, *Politics & Current Affairs* (2022), <https://yougov.co.uk/topics/politics/survey-results> [accessed 15/03/2022]

Yu, Zhaoxia and others, 'Beyond T Test and ANOVA: Applications of Mixed-Effects Models for More Rigorous Statistical Analysis in Neuroscience Research', *Neuron*, 110.1 (2022), 21 – 35

Zaller J. and S. Feldman, 'A Simple Theory of Survey Response: Answering Questions versus Revealing Preferences', *American Journal of Political Science*, 36.3 (1992), 579 – 616

Zeedan, Rami, 'The 2016 US Presidential Elections: What Went Wrong in Pre-election Polls? Demographics Help to Explain', *J – Multidisciplinary Scientific Journal*, 2.1 (2019), 84 – 101

Zhou, Zhenkun, 'Why Polls Fail to Predict Elections', *Journal of Big Data*, 8 (2021), 1-28