Computers in Human Behavior Reports 6 (2022) 100194

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Quantile regression analysis of in-play betting in a large online gambling dataset

Seb Whiteford [a,*], Alice E. Hoon [b], Richard James [c], Richard Tunney [d], Simon Dymond [a,e]

[a] School of Psychology, Swansea University, Singleton Campus, Swansea, SA2 8PP, United Kingdom
[b] Swansea University Medical School, Singleton Campus, Swansea, SA2 8PP, United Kingdom
[c] School of Psychology, University of Nottingham, Nottingham, NG7 2RD, United Kingdom
[d] School of Psychology, Aston University, Birmingham, B4 7ET, United Kingdom
[e] Department of Psychology, Reykjavík University, Menntavegur 1, Nauthólsvík, 101, Reykjavík, Iceland

ARTICLE INFO

Keywords:
In-Play
Live-action
Gambling
Quantile regression
Internet betting

ABSTRACT

In-play betting involves making multiple bets during a sporting event and is an increasingly popular form of gambling. Behavioural analysis of large datasets of in-play betting may aid in the prediction of at-risk patterns of gambling. However, datasets may contain significant skew and outliers necessitating analytical approaches capable of examining behaviour across the spectrum of involvement with in-play betting. Here, we employ quantile regression analyses to investigate the relationships between in-play betting behaviours of frequency and duration of play, bets per day, net/percentage change, average stake, and average/percentage change across groups of users differing by betting involvement. The dataset consisted of 24,781 in-play sports bettors enrolled with an internet sports betting provider in February 2005. We examined trends in normally-involved and heavily-involved in-play bettor groups at the .1, .3, .5, .7 and .9 quantiles. The relationship between the total number of in-play bets and the remaining in-play betting measures was dependent on degree of involvement. The only variable to differ from this analytic path was the standard deviation in the daily average stake for most-involved bettors. The direction of some relationships, such as the frequency of play and bets per betting day, were reversed for most-involved bettors. Crucially, this highlights the importance of determining how these relationships vary across the spectrum of involvement with in-play betting. In conclusion, quantile regression provides a comprehensive account of the relationship between in-play betting behaviours capable of quantifying changes in magnitude and direction that vary by involvement.

1. Introduction

In-play (sometimes referred to as live-action) betting is a form of gambling which involves making bets during a (typically) live sporting event, such as betting that a specific player will score a goal before halftime in a football (soccer) match (Killick & Griffiths, 2019). Compared to fixed-odds betting, in which bets are made prior to the commencement of the event, in-play betting has increased in popularity in recent years. Most online bettors surveyed in the United Kingdom (UK) report experience with in-play betting and higher estimated prevalence rates of problem gambling are observed in younger bettors (Gambling Commission, 2021a). Several researchers have expressed concerns that the nature of the rapid cycling betting propositions associated with in-play gambling, relative to the typically slow cyclical and static nature of fixed-odds bets, may be problematic for individuals "at-risk" of developing significant gambling problems (Griffiths & Auer, 2013; Harris & Griffiths, 2018; Gainsbury et al., 2020). Despite its growing popularity, there have however been relatively few studies directly investigating in-play betting (Killick & Griffiths, 2019).

In tandem with the rise in popularity of in-play gambling, online betting via internet gambling providers has increased significantly, from 17.3% of UK adults surveyed in 2016 to 23.6% in 2020 (Gambling Commission, 2021b). The availability of gambling opportunities 24 hours a day via the internet and mobile devices enables users to play at any time, and from anywhere. Concerns about the combination of rapid-paced in-play bets and online gambling availability have led to restrictions for online in-play betting in some jurisdictions (e.g., Australia - Gainsbury et al., 2020). As this form of gambling becomes

* Corresponding author.
E-mail address: s.j.whiteford@swansea.ac.uk (S. Whiteford).

https://doi.org/10.1016/j.chbr.2022.100194
Received 18 November 2021; Received in revised form 28 January 2022; Accepted 30 March 2022
Available online 1 April 2022
2451-9588/

increasingly regulated, the widespread adoption of online gambling has the potential to yield large and extensive datasets for research purposes. Such datasets are critical to understanding how individuals use these services and how their gambling behaviours develop and change over time. Large online datasets of real-world gambling data, such as those provided by the internet betting service provider *bwin* have enormous potential for facilitating novel analyses of real-world gambling behaviours on a scale that would be extremely impractical for other approaches, such as behavioural gambling paradigms tested in-person (Deng et al., 2019). For example, Brosowski et al. (2012) examined behavioural data from over 27,000 gamblers from the *bwin* dataset that revealed an increased level of at-risk involvement for users who engaged with poker and in-play betting.

Analysis of large, real-world online gambling datasets requires appropriate statistical techniques capable of detecting dynamic changes in behaviour and the impact of industry-operated responsible gambling tools (Auer & Griffiths, 2013, 2019; Catania & Griffiths, 2021; Philander, 2014; Ukhov et al., 2021). In jurisdictions where access to large datasets is either challenging or not possible, the availability of existing, older publicly available datasets may prove a useful analytical alternative. Despite this, the potential of new statistical approaches is only now being fully realised with both operator-provided and freely available datasets (Deng et al., 2019). For instance, previous research of a publicly available *bwin* dataset provided descriptive analyses of in-play behavioural measures and reported correlations between the total number of in-play bets and the duration or frequency of in-play gambling (LaBrie et al., 2007). Such an approach can be beneficial for preliminary analysis of the relationship between in-play behaviours, but conventional correlations (e.g., the Pearson correlation coefficient) make several assumptions likely to be violated in large online gambling datasets. Firstly, they require an absence of outliers, despite any real-world large gambling dataset almost certainly containing users who fall outside the normal range (LaBrie et al., 2007). In many fields, excluding outliers using relatively objective criteria such as 2.5/3 standard deviations from the group mean is a viable solution, but in the case of real-world gambling data, it is inefficient and potentially misleading to discard data from the 'most-involved' users. LaBrie et al. (2007) and others (e.g., LaPlante et al., 2008), have attempted to partially mitigate this issue (and consider how the behaviour of bettors who are more involved may differ from the "average" bettor) by separately analysing the top 1% of bettors in terms of money spent, or bets made. Secondly, even with such a group still included in the main analyses, correlation coefficients are likely to be skewed by outlying datapoints. Finally, this highlights another critical assumption of standard correlations that could be problematic for real-world gambling behaviours, in that it assumes a linear relationship between variables. In some cases, such as the duration of play and the total number of bets made, this might not be an entirely unreasonable assumption, but other in-play betting behaviours, such as average stake, almost certainly have a more nuanced relationship with the total number of in-play bets. While ordinary least squares (OLS) regression does not necessarily require a linear relationship between the dependent and independent variables, it does require normally distributed residuals which are of constant variance across the values of the independent variable (homoscedasticity).

Fortunately, alternative analysis techniques such as quantile regression, an extension of linear regression that estimates the conditional median rather than the conditional mean (e.g., OLS regressions; Koenker, 2017) can be employed to assess the association between two continuous gambling variables, when certain assumptions of standard regression or correlation analyses are not met. That is, quantile regression models do not require the relationship between in-play behaviours to be constant across levels of involvement with in-play betting or that all outliers be excluded (Koenker, 2017; Koenker & Hallock, 2001). In addition to these advantages, quantile regression enables a detailed examination of the magnitude of relationships between in-play gambling behaviours across the spectrum of involvement with in-play betting. As such it is possible to directly test whether specific in-play gambling behaviours such as average stake, and frequency of play vary across levels of involvement with in-play betting. If the relationships between other in-play betting behaviours and the total number of in-play bets do vary across levels of involvement (quantified as the total number of in-play bets within the study period), it provides preliminary support for approaches based on in-play betting as a function of levels of involvement with in-play betting and which may identify individuals at risk for problem gambling. To date, however, this analytic approach has yet to be applied to existing large, in-play betting datasets.

It is important to note that while here we have used the total of in-play bets across the study period as our primary measure of in-play gambling involvement, there are numerous other options such as net loss (Broda et al., 2008), frequency of play, or other composite measures (e.g., Russell et al., 2019). Given our focus on in-play betting, we concentrate on the *depth* of involvement with in-play betting, using the total number of bets within the study period. Within the broader context of research on problem gambling, however, other approaches such as the *breadth* of involvement or usage of multiple types of gambling activity may be beneficial (see LaPlante et al., 2014 for a comparison of depth and breadth effects in gambling involvement). It is not our intent to comment on which measure is the most appropriate way of quantifying individuals' degrees of involvement, simply to adopt a measure which, though raw, is an objective indication of involvement with in-play betting and examine how the relationship between this measure and other aspects of in-play gambling behaviour differ across various degrees of involvement.

The aim of the present study was therefore to apply quantile regression to in-play betting from a large online, publicly available gambling dataset. We sought to determine whether involvement with in-play betting and other features of in-play betting behaviour are consistent across the spectrum of involvement or if they vary in any systematic fashion. As such, the primary analysis involves quantile regressions of a variety of in-play betting behaviours (*duration*, *frequency*, *average stake*, *bets per day*, *bets per betting day*, *net change*, *percentage change* & the *standard deviation of daily average stakes*) on the total number of in-play bets during the study period. In this way, we aimed to characterise how the relationships between in-play betting behaviours can be quantified and compared in bettors with diverse degrees of involvement with in-play gambling.

2. Method

2.1. Sample

Internet betting service provider *bwin* Interactive Entertainment AG (*bwin*) provided data for 43,851 users who signed up between the 1st of February and the 30th of September 2005. The raw data is publicly available on the 'Transparency Project' webpage (http://www.thetransparencyproject.org/). After reducing the dataset to contain only the users who made in-play sports bets during the study period and for whom demographic data was available, 24,781 users remained. Of these, 22,736 of these users were male (91.75%) and 2045 were female (8.25%). Most users (23,980) made both fixed-odds and in-play bets within the study period, and 801 made in-play bets only.

The dataset includes in-play bettors from 64 countries, with most users based in Germany (14,386–58.05%), with the remainder from several countries (e.g., Poland: 1,626–6.56%, Turkey: 1,560–6.30%, Greece: 1,488–6.00%, Spain: 1,353–5.46%, France: 1,213–4.89%). Data about the ages of users was unavailable as it was removed during de-identification prior to public upload of the data.

To enable a detailed analysis of gambling behaviour across degrees of involvement with in-play betting, the dataset was subdivided into two groups. The first, containing all users who had a total number of in-play bets within 5 median absolute deviations (MAD – see Leys et al., 2013) around the group median (0–111 in-play bets across the study period), and the second group who made more than the median plus 5 MAD in-play bets (>111 in-play bets during the study period). This subdivision resulted in 20,891 users in the "*normally-involved in-play bettor*" (NIB) group, and 3,890 users in the "*most-involved in-play bettor*" (MIB) group. This approach is comparable to previous analyses of "most-involved bettors" defined as the top 1% of bets made within the study period (Broda et al., 2008; LaPlante et al., 2008; LaBrie et al., 2007). Here, our subdivision allows for a larger number of bettors in the most-involved group, which enables us to examine this group in more detail. Additionally, this approach has the benefit of encapsulating users within a given range of the median degree of involvement, only separating users who fall outside of this degree of involvement with in-play betting into the MIB group. This differs from the percentile approach used in other studies in that a MIB group defined by percentiles would include a certain percentage of users, even if their involvement with in-play betting falls within a given range of the median involvement.

2.2. Measures

The raw dataset contains daily aggregations of betting activity, summing the number of bets, total stakes, and total winnings (which can be return on bets made on previous days). From these variables, several other measures which quantify various aspects of individuals in-play betting behaviours were calculated. These measures included: *duration,* defined as the difference in days between the first and last in-play bet; *frequency* of play which was calculated as the percentage of betting days between the first and last in-play bet; *total in-play bets* which is simply the sum of bets made across their duration of play, as well as the average number of *bets per day* (across their entire duration of play) and *bets per betting day* (average number of bets including only days when bets were made). Measures quantifying the amount of money spent (in Euros) include the *total stake* (the sum of all stakes across the study period), the *average stak*e (average amount of money staked on each bet), and the standard deviation (*SD) for the daily average stake* (calculated by determining the average stake on each betting day and estimating the SD for each individual across betting days). To quantify the overall financial impact of in-play betting, the *net change* (the summed total of money lost/won over the whole study period), and the *percentage change* (the returns/winnings as a percentage of the total amount staked across the study period) were calculated from the daily betting aggregations.

2.3. Procedures

We conducted a secondary data analysis of daily in-play betting activity provided by *bwin* dataset. Prior to analysis, all data were thoroughly checked for inconsistencies, such as incomplete daily aggregation, days with no activity, or missing demographic data. To simplify this process, noninformative observations were removed from the dataset (e.g., observations for a user on a day where no bet, stake or win occurred – 186,489 observations in total), then all gambling products other than sports betting (e.g., online casino games) were removed from the data (74,220 observations) Any remaining entries where one user had multiple activities for a single gambling product on a single day (477 users) were combined to create an accurate daily aggregation, and any users without complete demographic data (1,609 users) were removed from the dataset.

We received ethical approval from the Swansea University School of Psychology Ethics Committee (5247) to conduct secondary data analyses with this dataset.

2.4. Statistical analysis

The statistical analyses follow three broad objectives. Firstly, to examine each of the behavioural characteristics across the MIB and NIB groups, and their respective degrees of involvement with in-play betting using quantile descriptive statistics (quantified using *total number of in-play bets*). Secondly, to employ quantile regressions to quantify the relationships between these other behavioural characteristics of in-play betting and the total number of in-play bets. Finally, to compare the estimated coefficients from the quantile regressions across quantiles of involvement with in-play betting to ascertain whether the strength of the relationships between the total number of in-play bets and these other in-play betting behavioural characteristics are influenced by an individual's degree of involvement with in-play betting.

Quantile regression offers several distinct advantages over conventional OLS regression analyses for the analysis of large gambling datasets. Specifically, in datasets with many individuals, who each display complex and differing patterns of play (such as the present *bwin* dataset) there are frequently issues with outliers and skewed data. These are problematic as they violate the assumptions for OLS regressions but are still potentially informative datapoints for in individuals whose gambling behaviours fall at the extreme tails of the sample distribution. As quantile regression estimates the relationship between variables for a specific quantile or percentile, it is robust to even extreme outliers, and more importantly can quantify the strength of the relationship between variables for locations other than the mean, allowing a more thorough investigation of the relationship between gambling behaviours at all levels of gambling involvement.

Separate multiple quantile regressions were fit to the .1, .3, .5, .7 and .9 quantiles in the NIB and MIB groups using the package "*quantreg*" (Koenker, 2021) in *R* (R Core Team, 2021). In each case, the models estimated incorporated the *total number of in-play bets* as the dependent variable, with each other measure of in-play betting activity (*duration*, *frequency*, *average stake*, *bets per day*, *bets per betting day*, *net change*, *percentage change* & the *standard deviation of daily average stakes*) as predictor variables.

3. Results

Quantifying and comparing the relationships between the total number of in-play bets and other in-play betting behaviours requires a multi-step approach, as noted in Section 2.4 above. Firstly, to visually examine the differences in behaviour across the spectrum of involvement with in-play betting, quantile descriptive statistics are reported. Secondly, the estimated coefficients from the quantile regressions are plotted, alongside OLS estimates to illustrate how the relationships between the in-play betting behaviours vary across degrees of involvement with in-play betting. Finally, Tests of Equality of Distinct Slopes (Koenker, 2021) are reported along with direct coefficient comparisons to quantify when and how the relationships between the *total number of in-play bets* and other in-play betting behaviours differ.

3.1. Quantile descriptives for in-play gambling behaviours

To illustrate how aspects of in-play gambling vary across the spectrum of involvement with in-play betting (operationalised as the *total number of in-play bets* across the study period), the descriptive statistics (mean & standard deviation) for each behavioural measure, across groups (NIB & MIB) and quantile within groups (.1, .3, .5, .7 & .9) are shown in Table 1 (non-parametric alternatives can be found in the Supplementary Materials). Notably, while several measures display relatively linear relationships (e.g., *duration*, *bets per day*, or *total stake* in euro), other measures display distinctly different relationships with the total number of in-play bets placed across the study period. For example, while net change (the total amount in euros won/loss across the study period) increases in magnitude as the number of in-play bets increase, the percentage change (the amount won/lost as a percentage of the total amount of money staked) decreases in magnitude across all quantiles in both the NIB and the MIB groups (as the *number of total in-play bets* increase). Other behaviours, such as frequency of play show a similarly nuanced relationship, with frequency decreasing with increasing

Table 1
Descriptive statistics (mean & SD) for in-play betting activity across degrees of involvement with in-play betting.

In-Play Bet group:	Normally-involved bettor group (NIB)					Most-involved bettor group (MIB)				
	[1,2]	(2,6]	(6,15]	(15,37]	(37,111]	(112,149]	(149,206]	(206,315]	(315,577]	(577, 21230]
Duration (days)	8.55 (30.47)	38.71 (59.16)	63.28 (71.76)	88.44 (78.25)	112.12 (78.81)	134.24 (76.14)	145.26 (75.29)	153.79 (69.07)	171.05 (60.65)	190.38 (49.24)
Frequency (%)	86.87 (30.61)	45.92 (41.02)	32.3 (34.03)	27.88 (29.4)	27.09 (25.26)	27.5 (22.04)	29.24 (20.52)	32.17 (20.01)	36.16 (18.53)	51.77 (19.66)
Total bets	1.4 (0.49)	4.27 (1.1)	10.41 (2.55)	24.89 (6.33)	65.51 (20.96)	128.92 (11.08)	175.53 (16.4)	253.02 (31.08)	423.2 (73.46)	1498.64 (1690.12)
Bets per betting day	1.21 (0.41)	2.16 (1.14)	2.98 (2.04)	4.07 (3.19)	5.71 (4.3)	7.27 (5.06)	7.93 (6.55)	8.62 (6.67)	9.95 (7.05)	16.16 (12.07)
Bets per day	1.08 (0.56)	1.29 (1.58)	1.37 (2.38)	1.62 (3.31)	2.03 (4.04)	2.25 (3.52)	2.78 (5.67)	3.03 (5.75)	3.66 (5.56)	8.44 (9.01)
Average Stake (€)	14.13 (31.01)	9.7 (23.17)	8.51 (21.2)	8.01 (18.28)	9.84 (24.39)	10.4 (25.42)	12.96 (28.34)	13.56 (27.4)	14.74 (31.46)	15.22 (26.33)
SD Daily average stake	7.14 (21.17)	6.15 (18.99)	6.19 (17.31)	6.65 (27.13)	8.72 (20.73)	9.84 (22.24)	11.75 (23.55)	13.18 (23.73)	13.74 (26.38)	13.78 (21.12)
Total Stake (€)	19.18 (43.41)	40.06 (89.86)	89.86 (246.89)	198.53 (464.84)	659.28 (1713.67)	1325.67 (3197.48)	2288.47 (5063.66)	3448.64 (7169.03)	6418.18 (14469.01)	20065.11 (36339.63)
Net Change (€)	−7.27 (36.84)	−7.3 (80.05)	−14.91 (92.84)	−27.96 (116.71)	−59.49 (256.78)	−108.41 (380.05)	−128.04 (419.5)	−223.27 (770.82)	−305.67 (1009.27)	−1138.84 (2343.31)
Percentage Change (%)	−41.7 (113.46)	−30.69 (69.76)	−24.21 (45.11)	−18.56 (30.18)	−14.12 (19.61)	−12.51 (14.88)	−10.36 (13.53)	−9.85 (11.07)	−9.35 (9.92)	−8.54 (7.96)
N	4428	4047	4137	4118	4161	789	775	772	777	777

Note. SD for daily average stakes are calculated within-subjects (across days) then averaged across participants.

numbers of in-play bets in the NIB group and increasing frequency with greater numbers of in-play bets in the MIB group. Additionally, average stake (in Euro), shows a U-shaped pattern in the NIB group, but increases relatively linearly in the MIB group.

3.2. *Quantile regression analyses*

The estimated quantile regression coefficients for each predictor in both the NIB and MIB groups are shown in Fig. 1. It is clear from Fig. 1 that the quantile coefficient estimates (blue circles) are frequently dissimilar to the OLS estimates (solid red line) in both groups, and that they predominantly fall outside of the 95% CI for the OLS estimates, which

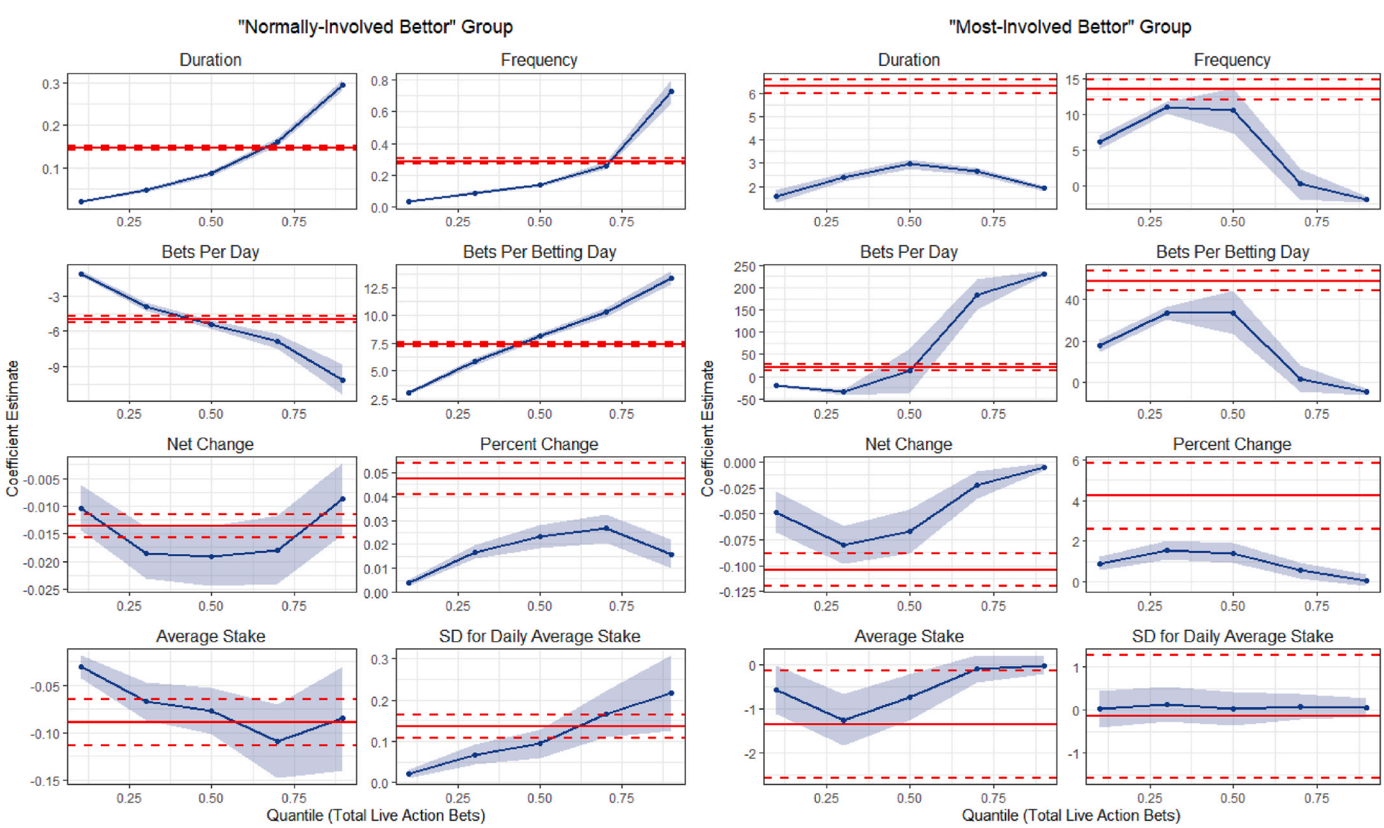

Figure 1. Quantile regression coefficient estimates for the .1, .3, .5 .7 and .9 quantiles in the NIB and MIB groups. *Note*. Blue circles indicate the point estimates of the quantile regression coefficients at each quantile with their associated 95% confidence intervals shaded in blue. Solid red lines show the location of the OLS-regression coefficient estimate, and dashed red lines show the 95% CI around the OLS estimate.

implies that the OLS estimates are generally biased estimators of the relationships between the number of in-play bets and the other in-play betting characteristics. The SD for daily average stakes in the MIB group is a clear exception to this general observation, and it is evident that there, and to a lesser degree with the average stake for the MIB group, the quantile coefficient estimates fall mostly inside the 95% CI for the OLS regression (except for the .7 & .9 coefficients for average stake with the MIB group).

Crucially, it is also evident that for most variables (except for the *average stake* & *daily average stake* for the MIB – noted above), the strength of the relationship between the total number of in-play bets and the other predictor variables is strongly influenced by the number of in-play bets placed. Taking *duration* or *frequency* in the NIB group as an example, the relationship between these variables and the total number of in-play bets is strongly influenced by the overall number of in-play bets, with a relatively weak relationship at low numbers of total in-play bets and a much stronger relationship at higher degrees of involvement. Notably, both variables show dissimilar patterns in the NIB and MIB groups, with a consistent increase in coefficient magnitudes across the NIB and a decline at higher levels of total bets in the MIB group.

In the NIB group, bets per day is negatively related to the total number of in-play bets in the study period and this relationship is stronger among bettors with a higher total number of in-play bets, however this pattern reverses in the most-involved bettor group, wherein bets per day becomes a strong positive predictor of total number of in-play bets for those who made more than the median number of bets in the MIB group (quantiles .5, .7 &.9). Conversely, bets per betting day is a positive predictor of total in-play bets in the NIB group, with increasingly large coefficients with increasing total number of in-play bets. This contrasts with the MIB group where the strength of the association between bets per betting day and total number of in-play bets falls in the .7 and .9 quantiles.

To confirm that the quantile regression coefficient estimates vary across the quantiles of total in-play bets, tests of equality of distinct slopes were conducted (Table 2). As indicated by Fig. 1, and confirmed in Table 2, all predictors except the SD for daily average stakes in the MIB group differ across the quantiles of total in-play bets. As such, the relationship between the total number of in-play bets and other aspects of in-play betting behaviour are heavily influenced by a user's degree of involvement.

4. Discussion

This study presents the first quantile regression analysis of real world in-play betting from a large online gambling dataset. The magnitude and direction of the relationships between the degree of involvement with in-play betting (quantified in terms of *total in-play bets* across the study period), and other in-play betting behaviours were heavily influenced by the user's total involvement with in-play betting. As such, this study demonstrates that quantile regression can yield detailed analyses of the relationships between gambling behaviour in large online datasets (Philander, 2014). It further emphasises the importance of the quantile approach over conventional analyses examining the hypothetical *mean* gambler, as OLS regression estimates were generally poor and/or biased estimates of the magnitude of the relationships between in-play behaviours across the spectrum of involvement with in-play gambling. In addition to these general contributions, the analyses reported here have revealed several specific patterns in the strength of the relationships between the involvement with in-play gambling and other in-play betting behaviours which warrant further discussion.

As outlined in the Introduction, the present study is the first application of quantile regression to in-play betting behaviour and, as such, there is a dearth of directly relevant previous research with which to compare findings. Previous research reporting correlations between in-play betting behaviours is however directly comparable (LaBrie et al., 2007). Here, we report that *duration* and *frequency* of play are positive predictors of the total number of in-play bets (at increasing magnitude across the assessed quantiles of involvement for the NIB group), but that the strength of these associations declines above the 0.5 quantile in the MIB group. LaBrie et al. (2007) reported a positive correlation ($r = 0.70$) between the *total number of in-play bets* and *duration*, and a negative correlation with *total in-play bets* and *frequency* of play ($r = -0.29$). That finding is consistent with the results reported here for *duration*, but not for *frequency* in the NIB group or for the MIB group below the .5 quantile. Given the rapid decline in the strength of the relationship between *frequency* of play and the *total number of in-play bets* above the .5 quantile in the MIB group, this suggests that the correlation reported in LaBrie et al. (2007) may have been driven by the most heavily involved bettors. The quantile approach reported here thus allows for a more detailed examination of these relationships across the spectrum of levels of involvement.

Similar observations can be made for the relationship between *bets per day* and *the total number of in-play bets*; we reported an increasingly negative relationship in the NIB group, which becomes positive at the .5 quantile in the MIB group. In contrast, LaBrie et al. (2007) reported a positive correlation ($r = 0.81$) between the same measures. Additional differences in analysis can be found in the *net loss/change* and *percentage loss/change* measures; LaBrie et al. (2007) observed a positive correlation ($r = 0.41$) for *net loss* and *total in-play bets* and a negative correlation for *percentage lost* ($r = -0.32$). For the quantile regression estimates reported here, *net change* was negatively associated with the *total number of in-play bets* at all quantiles and *percentage change* was a positive predictor, although the strength of this relationship varied across degrees of involvement with in-play betting. Taken together, these discrepancies with previous research (with the caveat that no direct comparison of these effects is possible due to differing analytical approaches, though one should expect at least consistent directionality in the relationships between in-play betting behaviours), highlight the potential utility of the quantile regression approach for examining patterns of play in in-play betting behaviour across the levels of involvement with in-play gambling.

Quantile regression is a demonstrably useful addition to the analytical toolbox for understanding the relationships between in-play betting behaviours as potential predictors of at-risk gambling. There is an established literature on analysis of behavioural tracking methods (Auer et al., 2020; Balem et al., 2021; Catania & Griffiths, 2021; Challet-Bouju et al., 2020; Ukhov et al., 2021) and growing interest in applying data science techniques to large online gambling datasets (Deng et al., 2019). As the availability of these types of datasets improves, so must the toolbox to quantify and characterise the underlying behaviours. Any real-world gambling dataset will have outliers and extremes of behaviour which can bias the estimates of many traditional analysis

Table 2
Tests of equality of distinct slopes.

	Normally-Involved Bettors	Most-Involved Bettors
Duration	$F(4, 80601) = 835.45$, $p < .001$	$F(4, 19446) = 223.31$, $p < .001$
Frequency	$F(4, 80601) = 329.67$, $p < .001$	$F(4, 19446) = 274.19$, $p < .001$
Average Stake	$F(4, 80601) = 18.8$, $p < .001$	$F(4, 19446) = 27.33$, $p < .001$
SD Daily Average Stake	$F(4, 80601) = 30.87$, $p < .001$	$F(4, 19446) = 0.82$, $p = .51$
Bets Per Day	$F(4, 80601) = 275.33$, $p < .001$	$F(4, 19446) = 1588.6$, $p < .001$
Bets Per Betting Day	$F(4, 80601) = 561.6$, $p < .001$	$F(4, 19446) = 200.94$, $p < .001$
Net Change	$F(4, 80601) = 25.18$, $p < .001$	$F(4, 19446) = 67.69$, $p < .001$
Percentage Change	$F(4, 80601) = 466.15$, $p < .001$	$F(4, 19446) = 94.31$, $p < .001$

approaches. Given that these outliers and heavily involved users contribute real and informative data points, we must use methods that can accommodate heterogeneity and extract the maximum possible information not only from the extremes but from bettors at all levels of involvement. As an example of how this works in practice, Deng et al. (2021) investigated Pareto estimates for online casino gambling, examining their association with voluntary self-exclusion as an index of gambling harm. In addition to demonstrating that the top 20% of gamblers (in terms of total bets) accounted for up to 92% of the total gambling activity in their sample, they reported significantly higher rates of voluntary self-exclusion in the top 20% relative to the remaining 80%. This, in conjunction with the strong variation in relationships between in-play betting behaviours along the spectrum of involvement with in-play betting further emphasises the importance of using techniques such as quantile regression which can examine the relationships between betting behaviours across the spectrum of gambling involvement, rather than attempting to examine gambling behaviour as though it can be adequately accounted for by an analysis of "average" behaviour alone.

While we have focused on examining the relationships between in-play betting behaviours across the spectrum of involvement with in-play betting, other approaches are possible to investigate how such relationships might be used to early identify and potentially reduce gambling related harm. Most notably, even though the approach here considers users across quantiles of involvement, examining how these relationships change over time is likely to be critical to understanding how in-play betting develops from normal use to those at-risk of gambling problems/harm. In a longitudinal analysis of sports betting behaviour, LaPlante et al., (2008) noted that newly subscribed internet bettors rapidly adapt, reducing their participation number of bets and stake size over time, but that heavily involved bettors did not follow this general pattern, especially for in-play bettors. Given this clear difference in the adaptation of in-play betting behaviour over time in new subscribers, a combined approach examining the nature of this adaptation over time across a range of quantiles of involvement would be beneficial for determining how adaptation occurs in users at other degrees of involvement and may be critical to differentiating the at-risk bettors from those who tend to reduce their gambling over time.

The present study has several limitations with implications for future research. Firstly, the current dataset consists of in-play betting data gathered in 2005, where all in-play betting was done using computers prior to the advent of the smartphone. Second, the number of in-play bettors has increased in the interim (Gambling Commission, 2021a). As a result of this increase in popularity, it is likely that the range of bets offered will have evolved over time and may impact aggregate gambling behaviour. Third, users from the *bwin* dataset may also have engaged with land-based gambling venues or other online platforms. Indeed, online gamblers have an average of three online gambling accounts (Gambling Commission, 2021a). Fourth, we quantified levels of involvement using the total number of in-play bets but alternatives such as monetary loss might produce differing relationships with other in-play betting behaviours (Xuan & Shaffer, 2009). Fifth, the present data is aggregated by day, which limits the analysis opportunities available in terms of temporal or sequential modelling of gambling behaviour. It would be of benefit to future researchers if gambling industry operators provided time-stamped data at the level of individual bets, such that a bet-by-bet analysis of in-play gambling during and across events was possible. Finally, we could further our understanding of the heterogeneity in gambling behaviour through the combination of quantile regression with restricted cubic splines (Marrie et al., 2009). While quantile regression enables the evaluation of the relationship between the independent variables and a continuous dependent variable across the complete range of the dependent variable, restricted cubic splines can be used to further assess nonlinearity, as well as to represent and better fit these complex nonlinear relationships (Gauthier et al., 2020).

In conclusion, we report a new approach to quantifying how relationships between in-play betting behaviours vary across the spectrum of involvement. It is however increasingly essential to consider gambling behaviour and in-play betting at the level of the individual bettor and to consider how these relationships between betting behaviours change across time. A fuller understanding of how such in-play behaviour evolves will aid in the identification of users at-risk of problem gambling and related harm.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was funded by a grant from the International Center for Responsible Gaming.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chbr.2022.100194.

References

Auer, M., & Griffiths, M. D. (2013). Voluntary limit setting and player choice in most intense online gamblers: An empirical study of gambling behaviour. *Journal of Gambling Studies, 29*, 647–660.

Auer, M., & Griffiths, M. D. (2019). Predicting limit-setting behavior of gamblers using machine learning algorithms: A real-world study of Norwegian gamblers using account data. *International Journal of Mental Health and Addiction*. https://doi.org/10.1007/s11469-019-00166-2

Auer, M., Malischnig, D., & Griffiths, M. D. (2020). Gambling before and during the COVID-19 pandemic among European regular sports bettors: An empirical study using behavioral tracking data. *International Journal of Mental Health and Addiction*, 1–8. https://doi.org/10.1007/s11469-020-00327-8

Balem, M., Perrot, B., Hardouin, J. B., Thiabaud, E., Saillard, A., Grall-Bronnec, M., & Challet-Bouju, G. (2021). Impact of wagering inducements on the gambling behaviors of on-line gamblers: A longitudinal study based on gambling tracking data. *Addiction*. https://doi.org/10.1111/add.15665

Broda, A., LaPlante, D. A., Nelson, S. E., LaBrie, R. A., Bosworth, L. B., & Shaffer, H. J. (2008). Virtual harm reduction efforts for internet gambling: Effects of deposit limits on actual internet sports gambling behavior. *Harm Reduction Journal, 5*(1), 1–9. https://doi.org/10.1186/1477-7517-5-27

Brosowski, T., Meyer, G., & Hayer, T. (2012). Analyses of multiple types of online gambling within one provider: An extended evaluation framework of actual online gambling behaviour. *International Gambling Studies, 12*(3), 405–419. https://doi.org/10.1080/14459795.2012.698295

Catania, M., & Griffiths, M. D. (2021). Understanding online voluntary self-exclusion in gambling: An empirical study using account-based behavioral tracking data. *International Journal of Environmental Research and Public Health, 18*(4), 2000. https://doi.org/10.3390/ijerph18042000

Challet-Bouju, G., Hardouin, J. B., Thiabaud, E., Saillard, A., Donnio, Y., Grall-Bronnec, M., & Perrot, B. (2020). Modeling early gambling behavior using indicators from online lottery gambling tracking data: Longitudinal analysis. *Journal of Medical Internet Research, 22*(8), Article e17675. https://doi.org/10.2196/17675

Deng, X., Lesch, T., & Clark, L. (2019). Applying data science to behavioral analysis of online gambling. *Current Addiction Reports, 6*(3), 159–164. https://doi.org/10.1007/s40429-019-00269-9

Deng, X., Lesch, T., & Clark, L. (2021). Pareto distributions in online casino gambling: Sensitivity to timeframe and associations with self-exclusion. *Addictive Behaviors, 120*, 106968. https://doi.org/10.1016/j.addbeh.2021.106968

Gainsbury, S. M., Abarbanel, B., & Blaszczynski, A. (2020). The relationship between in-play betting and gambling problems in an Australian context of prohibited online in-play betting. *Frontiers in Psychiatry, 11*. https://doi.org/10.3389/fpsyt.2020.574884

Gambling Commission. (2021a). Taking a more in-depth look at online gambling: Publication of online survey looking at online gambling behaviour in 2020. Available online at: https://www.gamblingcommission.gov.uk/statistics-and-research/publication/taking-a-more-in-depth-look-at-online-gambling.

Gambling Commission. (2021b). *Gambling behaviour in 2020: Findings from the quarterly telephone survey – statistics on participation and problem gambling year to December 2020*. Available online at: https://www.gamblingcommission.gov.uk/statistics-and-research/publication/year-to-december-2020.

Griffiths, M. D., & Auer, M. (2013). The irrelevancy of game-type in the acquisition, development, and maintenance of problem gambling. *Frontiers in Psychology, 3*, 621. https://doi.org/10.3389/fpsyg.2012.00621

Harris, A., & Griffiths, M. D. (2018). The impact of speed of play in gambling on psychological and behavioural factors: A critical review. *Journal of Gambling Studies, 34*(2), 393–412. https://doi.org/10.1007/s10899-017-9701-7

Killick, E. A., & Griffiths, M. D. (2019). In-play sports betting: A scoping study. *International Journal of Mental Health and Addiction, 17*(6), 1456–1495. https://doi.org/10.1007/s11469-018-9896-6

Koenker, R. (2017). Quantile regression: 40 years on. *Annual Review of Economics, 9*, 155–176. https://doi.org/10.1146/annurev-economics-063016-103651

Koenker, R. (2021). *quantreg: Quantile regression*. R package version 5.86 https://CRAN.R-project.org/package=quantreg.

Koenker, R., & Hallock, K. F. (2001). Quantile regression. *The Journal of Economic Perspectives, 15*(4), 143–156. https://doi.org/10.1257/jep.15.4.143

LaBrie, R. A., LaPlante, D. A., Nelson, S. E., Schumann, A., & Shaffer, H. J. (2007). Assessing the playing field: A prospective longitudinal study of internet sports gambling behavior. *Journal of Gambling Studies, 23*(3), 347–362. https://doi.org/10.1007/s10899-007-9067-3

LaPlante, D. A., Nelson, S. E., & Gray, H. M. (2014). Breadth and depth involvement: Understanding Internet gambling involvement and its relationship to gambling problems. *Psychology of Addictive Behaviors, 28*(2), 396. https://doi.org/10.1037/a0033810

LaPlante, D. A., Schumann, A., LaBrie, R. A., & Shaffer, H. J. (2008). Population trends in Internet sports gambling. *Computers in Human Behavior, 24*(5), 2399–2414. https://doi.org/10.1016/j.chb.2008.02.015

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Marrie, R. A., Dawson, N. V., & Garland, A. (2009). Quantile regression and restricted cubic splines are useful for exploring relationships between continuous variables. *Journal of Clinical Epidemiology, 62*(5), 511–517.

Philander, K. S. (2014). Identifying high-risk online gamblers: A comparison of data mining procedures. *International Gambling Studies, 14*, (1), 53–63. https://doi.org/10.1080/14459795.2013.841721

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. URL https://www.R-project.org/.

Russell, G. E., Williams, R. J., & Sanders, J. L. (2019). The relationship between memory associations, gambling involvement, and problem gambling. *Addictive Behaviors, 92*, 47–52. https://doi.org/10.1016/j.addbeh.2018.12.015

Ukhov, I., Bjurgert, J., Auer, M., & Griffiths, M. D. (2021). Online problem gambling: A comparison of casino players and sports bettors via predictive modeling using behavioral tracking data. *Journal of Gambling Studies, 37*, 877–897. https://doi.org/10.1007/s10899-020-09964-z

Xuan, Z., & Shaffer, H. (2009). How do gamblers end gambling: Longitudinal analysis of Internet gambling behaviors prior to account closure due to gambling related problems. *Journal of Gambling Studies, 25*(2), 239–252. https://doi.org/10.1007/s10899-009-9118-z