

The hidden cost of smoking: Rent premia in the housing market

Cigdem Gedikli¹ | Robert Hill²  | Oleksandr Talavera³  |
Okan Yilmaz¹ 

¹School of Management, Swansea University, Swansea, United Kingdom

²Department of Economics, University of Graz, Graz, Austria

³Department of Economics, University of Birmingham, Birmingham, United Kingdom

Correspondence

Robert Hill, Department of Economics, University of Graz, 8010 Graz, Austria.
Email: robert.hill@uni-graz.at

Abstract

In this article, we provide novel evidence on the additional costs associated with smoking. While it may not be surprising that smokers pay a rent premium, we are the first to quantify the size of this premium. Our approach is innovative in that we use text mining methods that extract implicit information on landlords' attitudes to smoking directly from Zoopla UK rental listings. Applying hedonic, matching, and machine-learning methods to the text-mined data, we find a positive smoking rent premium of around 6%. This translates into £14.40 of indirect costs, in addition to £40 of weekly spending on cigarettes estimated for an average smoker in the United Kingdom.

KEYWORDS

contracting frictions, hedonic regression, matching, random forest, rental market, smoking, smoking rent premium, text mining

1 | INTRODUCTION

Tobacco smoking is one of the world's biggest public health problems (Goodchild et al., 2018). The prevalence of smoking is highest amongst people who are unemployed or working in routine and manual jobs with low income, and without formal educational qualifications (Action on Smoking & Health [ASH], 2019). The unfavorable health and labor market consequences lock the most

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Real Estate Economics* published by Wiley Periodicals LLC on behalf of American Real Estate and Urban Economics Association.

disadvantaged groups into an intergenerational cycle of poor physical/mental health and well-being, addiction and impoverishment (Action on Smoking & Health [ASH], 2018; Böckerman et al., 2015; Levine et al., 1997; Pichon-Riviere et al., 2020; van Ours, 2004; Weng et al., 2013).

In this article, we aim to move beyond the well-established health and labor market outcomes associated with smoking. Instead, we focus on the experience of smokers in the housing rental market. We provide novel evidence of additional “hidden” costs associated with smoking that have not been explored before.

There are a number of potential reasons why landlords may wish to charge smokers a rent premium. First, smoke can leave an odor that persists on surfaces. This could be especially a problem for furnished properties. Second, smoking can damage paintwork, which could be particularly undesirable in new or newly refurbished properties. This in turn, can hinder future lets and create a financial burden for landlords. Third, there is an increased fire hazard associated with smoking.¹ Finally, landlords may perceive nonsmoking tenants as potentially more reliable. Marking a property as “nonsmoking,” therefore, may help landlords identify tenants that will take better care of the property they are renting (over and above the concerns with smokers already listed). This relates to a point emphasized by Halket et al. (2021) that contracting frictions exist between landlords and tenants regarding maintenance.

As a result, landlords may charge a rental premium to compensate for the potential financial costs and the safety risk associated with smoking. Nevertheless, there is no prior evidence on whether such a premium exists and, if it exists, how large it is. Exploring these potential further costs associated with smoking is worthwhile as it helps uncover the true cost of smoking.

Using Zoopla Property data, which offer unique historical information on the UK real estate market, we present the first estimates of the rent premium associated with smoking in the long-term rental market. We do this using the implicit content in advertisements regarding landlords’ attitudes toward smoking.

We use textual information available in the description section of the Zoopla property data to capture whether houses listed for rent include keywords or clauses such as “no smoking,” “smoking not allowed,” “smoking is not permitted” and so on. Using this information, we investigate the extent to which these houses command different prices when compared with houses, which do not include clauses intended to exclude smokers. We measure the size of the smoking rental premium using both standard hedonic pricing models and matching methods (mainly coarsened exact matching) as well as machine learning techniques. While coarsened exact matching enables a comparison for the houses that are almost identical in observed characteristics, machine learning techniques further help us to exploit the data more fully and calculate a probability of a rental property excluding smokers even when a listing does not include a smoking-related clause. Across all these estimations, our results point to a rental premium for properties where smoking is not prohibited. The estimated weekly rental premium is around 6%, translating into £14.40 of additional indirect costs of smoking in the form of rental premium. This extra cost is substantial, especially when considering the direct weekly cost of £40 an average UK smoker spends on cigarettes. We also show that the premium is higher in more expensive regions of the United Kingdom. We attribute this finding to higher maintenance costs in more expensive regions, which gives landlords a stronger incentive to seek out more reliable tenants.

Section 2 reviews the existing literature. Section 3 explains the details of Zoopla property data and the methods we employ in exploiting the textual information to determine the houses where

¹A UK Home Office report documents that smokers’ materials were responsible for 8% of accidental dwelling fires and 34% of fire-related fatalities in 2018/19 See <https://tinyurl.com/39m2drew>, access date: November 5, 2021.

smoking is not allowed. The various estimation methods we use to measure the smoking rent premium are also explained here. In Section 4, we present and interpret our results. Finally, Section 5 concludes.

2 | LITERATURE REVIEW

Previous research has investigated the rent premia associated with certain property and/or tenant characteristics. There is a long-standing interest in analyzing the factors that determine rental prices. Building on the simple hedonic pricing models which include standard property-specific characteristics (such as number of bedrooms, bathrooms, size etc.), studies document that rental prices are also influenced by other factors such as characteristics of the neighborhood, market conditions, location of the property as well as characteristics of the potential tenants (e.g., their ethnicity, gender and sexual orientation). For example, positive rent premia are noted for green buildings in the commercial real estate market (e.g., Fuerst et al., 2012; Robinson & Sanserson, 2016), for cancellable rental contracts (Yoshida et al., 2016) and for having trees on a rental property's lot (Baranzini et al., 2010). Additionally, it is widely acknowledged that rent premia exist based on ethnicity, sexual orientation and gender (see Flage, 2018, for a review of the international evidence).

There is also literature that explores the link between housing quality and health outcomes (see for example, Palacios et al., 2021). However, we are aware of only one previous paper that attempts to measure the link between smoking and rent. Using a hedonic model, Benjamin et al. (2001) find that vacationers to the Outer Banks of North Carolina are willing to pay a premium of 11.6% to rent properties that prohibit smoking during the peak rental season. Interestingly, this premium goes in the opposite direction to what we observe here. This finding may reflect differences between the long-term and holiday rental markets. In particular, owner-occupiers participate in the holiday rental market but not in the long-term rental market. Since owner-occupiers tend to be richer, their presence in the holiday rental market (where typically they are the majority) can create a demand for nonsmoking rentals that mostly does not exist in the long-term rental market.

Our use of textual information available in the description section of online Zoopla property data to detect the smoker preferences of landlords connects with a strand of the real estate literature, which stresses the importance of textual data in improving the property price predictions in hedonic models. Exclusively focusing on the US sale/rental market and benefiting from the textual information available in the Multiple Listing Service (MLS) data, many argue that listing agents' descriptions and remarks about houses contain hidden characteristics that may well influence their prices (see, for example, Haag et al., 2000; Luchtenberg et al., 2019; Nowak & Smith, 2017). Accordingly, these studies either identify a set of keywords or phrases based on the remarks section of the MLS data in order to signal sellers' or landlords' motivation, amenities, interior design, physical improvements, such as whether the property is recently refurbished, painted etc. (e.g., Soyeh et al., 2014) or they resort to machine/deep learning when mining textual information (e.g., Nowak et al., 2021; Zhou et al., 2019)—the methods which we also employ in our analysis.

These studies all prove the value of textual information and controlling for the additional property attributes in estimating property prices. They show that the words used in MLS listings affect sale price, time on market (Haag et al., 2000; Nowak & Smith, 2017; Pryce & Oates, 2008), the likelihood of visiting a property for sale (Luchtenberg et al., 2019), and rental prices (Zhou et al., 2019). Some focus on specific attributes hidden in the textual data; for example, Bond and Devine (2016) find that including phrases related to being green is associated with higher rental rates and

Leadership in Energy and Environmental Design (LEED) certification commands the highest rent premium.

Our paper also connects with the literature on price-rent ratios (Bracke, 2015; Halket et al., 2021; Heston & Nakamura, 2009; Hill & Syed, 2016). Price-rent ratios are observed to be larger at the high end of the market: a result that is attributed partly to contracting frictions between owners and renters over maintenance. We find a similar dynamic at work with the smoking premium.

3 | DATA AND METHODOLOGY

3.1 | The Zoopla data set

Our empirical investigation is based on rental listings from Zoopla property data.² Zoopla is one of the UK's leading providers of historical property listings data, with more than 27 million residential property records in their archives. It contains records of more than 8 million properties advertised for sale and/or rent. In this article, we employ the data for the properties, which were advertised for rent between the years 2012 and 2018.

Zoopla property data include a rich set of historical information on advertised properties' physical characteristics such as the number of bedrooms, floors, bathrooms and property type (e.g., flat; terrace house etc.). The data provide a detailed level of geographical information, enabling the location of rental properties to be determined at the local authority level and at even smaller units, such as post town.

Most important to our analysis is that Zoopla property data contain landlords' or listing agents' written descriptions of the rental property. Using these descriptions, we identify and generate a list of keywords and phrases, which we then use as indicators of landlords' attitudes toward smoking in the rental property. We determine clauses pointing to a negative attitude toward smoking; for example, "smoking is not allowed," "smoking is not permitted," "no smoking," and so on. If the listing for the property involves any of these phrases, it is classified as a nonsmoking property. We also capture positive phrases such as "smoking is allowed"; "smoking is permitted" to identify these properties as smoking friendly. The full list of these negative/positive indicator phrases is available in Table A1.

In addition to determining whether smoking is permitted in the rental property or not, the description section of Zoopla property data assists us in identifying additional nonstandard attributes of the properties (such as whether the property is recently refurbished or furnished) that are not readily available in the data. The variable definitions and key descriptive statistics are presented in Table 1.

3.2 | Hedonic estimation

In the first step, we employ a hedonic pricing model where rent³ is expressed as a linear combination of its key property attributes along with an indicator for smoking

²The data are provided by the Urban Big Data Centre (UDBC) licensed by Zoopla Limited (2018), <https://tinyurl.com/efw2p2oy>.

³We use listing rent (ask rent) as we do not have data on agreed rent.

TABLE 1 Descriptive statistics

	Whole sample	No smoking dummy = 1	No smoking dummy = 0	Mean difference
Log(<i>rental price</i>)	5.303 (0.549)	5.036 (0.369)	5.320 (0.554)	-0.285***
<i>No smoking</i>	0.059			
<i>Probability of not accepting smokers</i>	0.621 (0.122)			
Property characteristics				
Size				
<i>Number of bedrooms</i>	2.181 1.104	2.202 (0.963)	2.179 (1.113)	0.023***
<i>Number of floors</i>	(0.114) (0.455)	0.108 (0.443)	0.115 (0.456)	-0.007***
<i>Number of bathrooms</i>	0.592 (0.694)	0.602 (0.674)	0.591 (0.696)	0.011***
Status				
<i>Refurbished</i>	0.112	0.092	0.114	-0.022***
<i>Furnished</i>	0.190	0.353	0.180	0.174***
Type				
<i>Detached house</i>	0.123	0.168	0.120	0.049***
<i>Terraced house</i>	0.160	0.195	0.158	0.037***
<i>Flat</i>	0.498	0.381	0.505	-0.124***
<i>Other</i>	0.220	0.256	0.217	0.038***
Observations	2,726,268	162,507	2,563,761	

Note: Log(*rental price*) is the natural logarithm of weekly rent. *No smoking* takes the value of 1 if the property description explicitly disallows smoking and 0 otherwise. *Probability of not accepting smokers* is the implied probability of not accepting smokers. *Number of bedrooms*, *Number of floors*, and *Number of bathrooms* are continuous variables. *Refurbished* is a dummy variable that takes the value of 1 if the property is refurbished and 0 otherwise. *Furnished* is a dummy variable that takes the value of 1 if the property is furnished and 0 otherwise. Property type categories are *Detached house*, *Terraced house*, *Flat*, and *Other* (each is a dummy variable taking the value of 1 if the property belongs to the relevant category and 0 otherwise). *Other* includes property types such as a bungalow, cottage, lodge, etc. The second and third columns show the mean values of variables with respect to the values of no smoking dummy. The last column shows the differences in the mean values and *t*-test results on the equality of means. *t*-Test results reject the null hypothesis of equal means at 1 % significance level (***) for all variables.

not being allowed in the rental property. Accordingly, the following OLS regression is estimated:

$$\ln(R_i) = \alpha + \beta \text{No Smoking}_i + X_i\gamma + D_i\delta + \varepsilon_i$$

The dependent variable in the hedonic model is the logarithm of the rent of the *i*th rental property. *No Smoking* is our main variable of interest, which is a dummy variable that takes the value 1 if smoking is not allowed in the rental property and value 0 in the absence of any such restriction. The vector X_i contains standard attributes of the rental houses, including the type of property (detached house, terraced house, flat, and other), number of bedrooms, number of bathrooms,

number of floors.⁴ It also includes additional characteristics which we identified using the textual information available in the data. These are dummy variables indicating whether the house is furnished (1 if furnished, 0 otherwise) and whether the house is refurbished (1 if refurbished, 0 otherwise). D_i is a vector of dummy variables indicating the county-level location (the data set includes listings from 85 counties) and the year in which the listing enters the market. The vectors γ and δ denote the shadow prices of the physical and location/year-of-listing characteristics of the properties.

In this model, the coefficient β captures the effect of the presence of a nonsmoking clause in a listing on the rental price. A negative coefficient indicates that landlords ask for a premium for allowing smoking in the property. Accordingly, the size of this coefficient reflects the rent premium (or penalty) of smoking in the housing market.

3.3 | Coarsened exact matching

Estimates of the smoking premium obtained from hedonic modeling may be misleading if landlords select nonsmokers into particular types of properties.⁵ In order to overcome such problems inherent in a standard hedonic pricing model that relies on OLS estimation, we employ matching methods to reduce the imbalance in the covariates between the treated and control groups and improve causal inferences. Our preferred matching technique is coarsened exact matching (CEM), which is a relatively new matching technique that employs monotonic imbalance bounding (Iacus et al., 2012). CEM is proven to be more efficient in terms of achieving lower levels of imbalance, model dependence and bias than other commonly used matching techniques such as propensity score matching (Blackwell et al., 2010; Iacus, et al., 2012). More pragmatically, CEM is faster than other matching methods and works well with larger data sets like the one we use in our analysis (King & Nielsen, 2019).⁶

Following the notation in Blackwell et al. (2010), assume that $Y_i(1)$ is the potential outcome (here the rent) if unit i receives the treatment and $Y_i(0)$ is the potential outcome if the same unit is in the control group. For each observed unit i only one of these potential outcomes is observed, and we set $T_i = 1$ if i is treated and $T_i = 0$ otherwise. The observed outcome can be shown as: $Y_i = T_i Y_i(1) + (1 - T_i)Y_i(0)$. The treatment effect (in our case, the effect of the “no smoking restriction”) on unit i can then be expressed as $TE_i = Y_i(1) - Y_i(0)$ and the average treatment effect (ATT) on the treated is $ATT = \frac{1}{n_T} \sum_{i \in T} TE_i$, where $n_T = \sum_{i \in T} T$ and $T = \{1 \leq i \leq n : T_i = 1\}$. CEM addresses the abovementioned observational data problem related to nonrandom assignment to treated and control groups and these groups not being identical before treatment. It does so by temporarily coarsening the data, applying exact matching to

⁴ The information about size of the properties (e.g., square meters) is not typically included in property listings in the United Kingdom. However, there are legal restrictions related to the size of a bedroom: for example, the size of a single bedroom has to be at least 6.51 square meters. See <https://www.legislation.gov.uk/ukxi/2018/616/regulation/2/made> (accessed on November 5, 2021).

⁵ This means that rental properties where smoking is restricted could have observably different characteristics than properties where smoking is not restricted. For instance, nonsmoking properties might be more likely to be refurbished. By using CEM, we ensure that the data are trimmed so that a more similar empirical distribution is achieved regarding the covariates between the treated and control groups.

⁶ As a robustness check, we still repeated our analysis using propensity score matching, nearest neighbour matching, and inverse probability weighting (IPW). Results from these attempts were qualitatively similar to what we found using CEM.

the coarsened data, identifying the matched and unmatched units and then keeping only the original—uncoarsened—values of the matched data.

To be more specific, CEM first coarsens the covariates (X) by recoding⁷ and generates a set of strata, which contain the same coarsened values of X . Following this, it assigns these strata to the original data. Finally, it retains any observation whose stratum contains at least one treated and one control unit and drops the others. Once this procedure is accomplished, these strata constitute the basis for calculating the treatment effect.

We match the rental houses which we identify as “nonsmoking” with those without a ban on smoking on the basis of the type of the rental property, number of bedrooms, number of bathrooms, whether they are furnished and whether they are refurbished. To control for any confounding effects, we also use the year of entering the market when matching the houses. Finally, to account for locational differences, we include county dummies in our CEM estimations. However, given that there are 85 counties, and this may reduce the quality of the matches, we also aggregate counties and match houses over larger regions. Here we use the 11 NUTS1 regions for Great Britain (we exclude Northern Ireland due to a lack of data).⁸

As with any other matching method, CEM is a data preprocessing technique (Blackwell et al., 2010). It ensures that the observations in the original data are pruned and the remaining data are better balanced in terms of treated and control groups. Hence, a more similar empirical distribution is achieved regarding the covariates. When the data are exactly balanced or when one-to-one exact matching is performed, a difference in mean outcome values between the treated and control groups produces an estimate of the causal effect. When the match is approximately balanced or the match is not exact, then it might be necessary to control for the covariates in the model and estimate parametric models. Even when this is the case, the derived estimates are less model-dependent and statistical bias is lower when compared to the estimations based on the data, which are not preprocessed by matching methods. We achieved good levels of balance in our matched data between the treated and control units when we used county when matching the properties (see Table A2, A3)⁹. Therefore, we calculated the average treatment effect of smoking by simply including CEM weights in our linear regression, as suggested by Blackwell et al. (2010).¹⁰

Additionally, we perform a CEM analysis separately for each county across the years using the same set of characteristics described above. We then compare the size of the smoking rent premium observed in each county with the average rental prices observed at county-level. Doing

⁷ This coarsening and recoding of the covariates can be performed by using user-defined cut points; for example, continuous variables can be grouped into meaningful categories or Likert-scale variables (e.g., strongly agree, agree, disagree, strongly disagree) may be aggregated if necessary. Alternatively, CEM’s automatic binning algorithm can be used. In this article, our covariates were mostly categorical with finely defined categories or dummy variables; therefore, we mainly relied on CEM’s algorithms.

⁸ NUTS denotes the Nomenclature of Territorial Units for Statistics.

⁹ The matching statistics when we use more aggregated regional data are presented in Table A3. Although using regions improved the imbalance between the treated and control groups even more (see the L_1 statistic in Table A3), matching using county can control for more unobservables. Moreover, using regions translates into more coarsening, resulting in fewer strata and a more diverse observations within them. Therefore, using regions may result in higher imbalance and model dependence. Nevertheless, for comparison purposes and robustness checks, the estimation results based on the data matched at regional level are also presented in the Results section (see column 2, Table 3).

¹⁰ Including property characteristics in the model has virtually no impact on the estimated smoking coefficient. Weights are included because, under matching with replacement, multiple properties in the treated group can be matched with the same property in the untreated group. Hence there will typically be more observations in the untreated group than in the treated group.

so enables us to explore any variation in the potential smoking rent premium across different parts of the rental market.

3.4 | Probabilistic categorizing of listings that are silent on smoking using random forests

One problem with our data set is that the number of listings that explicitly allow smoking is relatively small. It is likely, therefore, that our text mining methods have not identified all non-smoking properties. As a consequence, our empirical estimates of the smoking rent premium may be too low. To partially address this problem, we employ Random Forests—a machine learning approach—to categorize listings with unknown information into listings in which smoking is either allowed or not.¹¹

Our approach follows several steps:

- (i) We extract data that explicitly indicate preferences for nonsmoking (Smoking not allowed = 1) and smoking (Smoking allowed = 0) tenants. The sample has 162,507 and 3,578 listings in each category, respectively. Given that the former group is much larger in size, we balance the data so that the numbers of observations in each group are similar. To do so, we created subcategories based on the intersection of the following attributes: location, number of bedrooms, number of bathrooms, and number of floors.
- (ii) Within each subcategory, we match observations indicating tolerance to smokers with the same, randomly selected sample of listings that do not allow smoking. If a particular intersection does not have any listings, which allow for smoking, it is dropped.
- (iii) We divide our balanced data set of 6970 observations into training and testing data sets with the 80:20 split. A binary variable indicating whether smoking is not allowed is the main output in this machine learning exercise. The vector of inputs (or features) is created from two sources, quantitative and textual data. The former (e.g., region or bathrooms) includes all available categorical information transformed into binary variables. The latter (listing descriptions) were processed by using *word2vec* to extract 100 features per word, which were later averaged over all words in each listing description. The minimum occurrence of a word is 20.
- (iv) The random forest classifier from *sklearn* Python library is chosen as our main approach. Parameters of the classifier were optimized using grid search by looking through depth of trees [30, 40, 50, 100], number of trees [50, 100, 200, 250, 300, 400, 500], and number of features (sqrt(features), log2(features), auto). The best accuracy rate (using testing data) is 84% when we have 500 trees with depth of 30 and log(2) for max features. The most primary drivers are text-based embeddings with feature importance values varying from 0.007 to 0.014. Having developed the model, we categorize listings that do not explicitly mention smoking preference and estimate the probability of not allowing smoking.

¹¹ It is possible that some rental property owners may not have a particular preference toward smoker or nonsmoker tenants, or they may choose not to disclose a preference at least at the listing stage but a lower rent or security deposit may be offered to nonsmokers at the negotiation stage. Alternatively, landlords may also consider the possibility of accepting a smoker at the negotiation stage (assuming that the potential tenant informs the landlord that (s)he is a smoker). While a dichotomous no smoking variable used in the previous models may not be able to capture these elements, the machine-learning approach addresses these possibilities by estimating a probability of being a nonsmoker rental unit even when the listing does not disclose a preference towards smoking.

TABLE 2 Hedonic pricing model. OLS estimations

	(1)	(2)	(3)
<i>No smoking</i>	−0.046*** (0.001)	−0.030*** (0.001)	−0.026*** (0.001)
Property characteristics			
Size			
<i>Number of bedrooms</i>		0.216*** (0.000)	0.233*** (0.000)
<i>Number of floors</i>		−0.043*** (0.000)	−0.032*** (0.000)
<i>Number of bathrooms</i>		0.041*** (0.000)	0.038*** (0.000)
Status			
<i>Refurbished</i>			0.019*** (0.001)
<i>Furnished</i>			0.014*** (0.000)
Type			
<i>Terraced house</i>			−0.083*** (0.001)
<i>Flat</i>			0.053*** (0.001)
<i>Other</i>			−0.010*** (0.001)
Observations	2,726,268	2,726,268	2,726,268
R^2	0.542	0.731	0.737

The dependent variable $\log(\text{rental price})$ is the natural logarithm of weekly rent. *No smoking* takes the value of 1 if the property description explicitly disallows smoking and 0 otherwise. *Number of bedrooms*, *Number of floors*, and *Number of bathrooms* are continuous variables. *Refurbished* is a dummy variable that takes the value of 1 if the property is refurbished and 0 otherwise. *Furnished* is a dummy variable that takes the value of 1 if the property is furnished and 0 otherwise. Property type categories are *Detached house*, *Terraced house*, *Flat*, and *Other* (each is a dummy variable taking the value of 1 if property belongs to the relevant category and 0 otherwise). *Other* includes property types such as a bungalow, cottage, lodge, etc. All specifications include constant, county, and year dummies. Robust standard errors in parentheses. ***, **, and * correspond to 1%, 5%, and 10% statistical significance levels, respectively.

- (v) Once these probabilities have been estimated, we rerun the hedonic model from Section 3.2, this time inserting these probabilities as the dependent variable.

4 | RESULTS AND DISCUSSION

4.1 | Hedonic results

Table 2 reports the results from the first step of our analysis, where a hedonic pricing model is estimated by OLS. Specification 1 includes our key variable of interest indicating whether smoking is

TABLE 3 Average treatment effects (ATT) based on CEM

	(1)	(2)
No smoking	−0.063*** (0.001)	−0.044*** (0.001)
No of matched observations	2,515,743	2,567,114

In Column (1), county is used to match the rental properties in terms of location. In Column (2), region is used to match rental properties. In total, we have 85 counties and 11 regions. In both specifications, number of bedrooms, bathrooms, floors, property type, property being furnished, refurbished are used when matching the properties. Standard errors are in parenthesis. ***Statistically significant at 1%.

permitted or not along with the control dummies for the year and counties. The coefficient for “No Smoking” in this specification is -0.046 , indicating that houses, where smoking is not allowed, are associated with 4.6% lower rental prices compared to those without any ban on smoking. When we control for size-related variables such as number of bedrooms, bathrooms and floors, the rental price difference goes down to 3% (Specification 2). Finally, in the last specification (Specification 3), we include additional variables to control for the type of the property and whether the house is refurbished and/or furnished. Controlling for these characteristics does not change the size of the rental premium of smoking dramatically. Even after controlling for all the property characteristics, the coefficient of “No Smoking” is still negative and statistically significant showing a 2.6% lower weekly rent for the houses where smoking is not allowed. The control variables, which we include in our model, have the expected sign and magnitude. The rent is positively correlated with the number of bedrooms, bathrooms, and the property being refurbished or furnished.

4.2 | Coarsened exact matching results

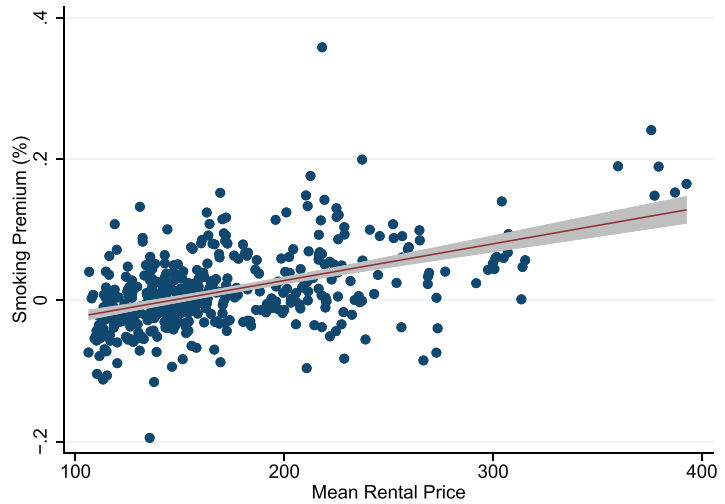
The standard hedonic pricing model points to a rent premium for houses where smoking is allowed. Nevertheless, this evidence may be distorted by the way landlords select houses to allow smoking or not. Therefore, in the second step of our analysis, we use CEM and attempt to compare the rental price of pairs of smoking and nonsmoking rental properties that are almost identical across all the covariates. Given our large sample size, we did not experience any sample size issues while performing matching; our control and treated groups are highly balanced (Table A2 in the Appendix presents the summary statistics and measures describing the quality of our matched data). The estimation results based on data preprocessed by CEM are less model-dependent and biased, enabling more credible causal inferences with regard to the effect of smoking on rental prices.

4.3 | Cross-section variation in the smoking rent premium

Table 3 shows the average treatment effect for houses where smoking is not allowed as -0.063 . This is notably larger than the figure produced by the OLS estimation that relied on the same covariates (last column in Table 2). It indicates that for houses where smoking is not allowed, the rent is 6% lower than those where smoking is not banned.

A county-level CEM analysis demonstrates that the amount of the premium differs along the rental price distribution. When we plot the amount of the smoker rent premium against the mean

FIGURE 1 Smoking premium and mean rental price
Note: Each point corresponds to a county at a given year. The red line represents the fitted values and the shaded area represents 95% confidence interval. The smoking premium is calculated using CEM. Subsamples with fewer than 50 nonsmoking houses are dropped to ensure the quality of the matching process.
 [Color figure can be viewed at wileyonlinelibrary.com]



rental prices at county-level (see Figure 1), we observe an upward sloping curve indicating that the smoking rent premium is larger at the higher-end of the rental market.

A parallel can be drawn here with the literature on housing market price-rent ratios (or rental yields—which are the reciprocals of price-rent ratios). A number of papers have observed an upward sloping price-rent ratio curve (for example, Bracke, 2015; Halket et al., 2021; Heston & Nakamura, 2009; Hill & Syed, 2016). Multiple factors may be contributing to this finding. However, Halket et al. (2021) stress the role played by contracting frictions regarding maintenance between landlords and tenants. These frictions become more important for more expensive properties.

As a result of contracting frictions, better quality properties with higher maintenance costs tend to be selected into the owner-occupied rather than rental market. Furthermore, some of the characteristics associated with higher maintenance costs are typically omitted from the data set (e.g., higher-quality kitchens, bathrooms, and finishes). It follows from this that owner-occupied properties at the higher end of the market are generally of better quality than rentals that are matched with them on the observable characteristics. This by itself can explain why an upward-sloping relationship is observed between price and the price-rent ratio.

Returning to our smoking context, the incentive to find a reliable tenant is therefore greater at the higher-end of the rental market, where contracting frictions are larger. To the extent that a tenant identifying as a nonsmoker is interpreted by the landlord as a signal of reliability, it follows that the smoker rent premium should be larger at the higher-end of the market. Our findings in Figure 1 are consistent with this view.

4.4 | Results obtained by probabilistically classifying listings using random forests

Finally, we replicate the hedonic pricing model estimations, with a new variable to capture landlords' preferences toward smoking. As mentioned in Section 3, we use a machine learning method to compute the probability of not allowing smoking if a listing does not include any keywords related to smoking. Accordingly, we create a new variable, which takes the value of 1 if the description includes any of the negative phrases listed in Table A1 and 0 if the description includes any of the positive phrases. For the listings, which do not include any smoking-related keywords, this

TABLE 4 Hedonic pricing model. OLS estimations using probability of accepting smokers

	(1)	(2)	(3)
<i>Probability of not accepting smokers</i>	−0.244*** (0.002)	−0.054*** (0.001)	−0.070*** (0.001)
Property characteristics			
Size			
<i>Number of bedrooms</i>		0.216*** (0.000)	0.232*** (0.000)
<i>Number of floors</i>		−0.043*** (0.000)	−0.033*** (0.000)
<i>Number of bathrooms</i>		0.041*** (0.000)	0.039*** (0.000)
<i>Refurbished</i>			0.018*** (0.001)
<i>Furnished</i>			0.012*** (0.000)
Type			
<i>Terraced house</i>			−0.085*** (0.001)
<i>Flat</i>			0.053*** (0.001)
<i>Other</i>			−0.010*** (0.001)
Observations	2,726,268	2,726,268	2,726,268
R ²	0.544	0.731	0.737

The dependent variable $\log(\text{rental price})$ is the natural logarithm of weekly rent. *Probability of not accepting smokers* is implied probability of not accepting smokers. *Number of bedrooms*, *Number of floors*, and *Number of bathrooms* are continuous variables. *Refurbished* is a dummy variable that takes the value of 1 if the property is refurbished and 0 otherwise. *Furnished* is a dummy variable that takes the value of 1 if the property is furnished and 0 otherwise. Property type categories are *Detached house*, *Terraced house*, *Flat*, and *Other* (each is a dummy variable taking the value of 1 if property belongs to the relevant category and 0 otherwise). *Other* includes property types such as a bungalow, cottage, lodge etc. All specifications include constant, county and year dummies. Robust standard errors in parentheses. ***, **, and * correspond to 1%, 5%, and 10% statistical significance levels, respectively.

variable takes the probability values computed by the machine-learning process. This new variable represents the probability of not accepting smoking instead of a dummy variable capturing a restriction on smoking. Table 4 shows the results from this estimation. While the size of the coefficient for the key variable is very large before controlling for property characteristics, it drops to −0.070 after the inclusion of the full set of covariates. Nevertheless, a smoking premium of 7%, is larger than the results suggested by the previous model, which includes a dummy for the smoking restriction.¹² This is as expected since in the previous model it is probable that some nonsmoking properties were incorrectly identified as allowing smoking.

Our results show that regardless of the method we choose and the type of variable we use to capture landlords' preferences, we find a rent premium for the houses where smoking is not

¹² Considering the mean value of the probability of not accepting smokers (0.621), this result translates to an average effect of 4.3%.

prohibited. Moreover, this premium is stronger when we employ CEM and, when we assign restriction probabilities to listings, which do not explicitly express a preference using a machine learning approach. We believe both matching and machine learning methods contribute to more precise estimation of the smoking rent premium. These two approaches generate similar estimates of the smoking premium that range between 6% and 7%.

5 | CONCLUSION

In this article, we investigate the hidden costs associated with smoking by focusing on the rental market. Using a unique historical property data set and benefiting from the textual information therein available, we distinguish between rental properties where smoking is restricted or not. We then investigate the extent to which houses where smoking is not restricted command higher rental prices when compared to nonsmoking houses.

We use various estimation methods, ranging from a standard hedonic pricing model to matching methods and more sophisticated machine-learning techniques. Results from all these approaches consistently show that houses where smoking is not restricted charge a rent premium. The extent of this premium is noteworthy. If we consider the results from the data matched by the CEM method, there is a 6% rent premium related to smoking. As the mean weekly rent is around £240 in our sample, this premium corresponds to an average of £14.40 pounds per week. To put this number into context, the average number of cigarettes consumed by an adult smoker in England is 10 per day¹³ and the average retail price of 20 filtered cigarettes is £11.45.¹⁴ This translates into an average of £40 weekly spending as a direct cost of smoking. Accordingly, the £14.40 rental premium, as a hidden indirect cost, paid by an average smoker increase the total cost of smoking by a third.

This is a significant additional financial cost, especially considering a recent report by Action on Smoking and Health, which points to the increased poverty rates in smoking households in the United Kingdom after accounting for the money spent on tobacco (ASH, 2019). It is documented that while the poverty rate across the general population is 19.2%, this figure is 28.4% for private tenants who smoke. Private tenants are shown to spend 5.8% of their disposable income on tobacco products. In this article, we show that the total cost would be noticeably higher if the smoking rent premium is taken into account.

Further research is required into the causes of the smoking rent premium. Also, although it is worthwhile to explore further dynamics such as potential regional heterogeneities in future work, we find robust evidence of the presence of a rent premium associated with smoking. The social and economic costs tobacco smoking imposes on individuals and their families are substantial, and this article shows that, in fact, they might be even larger than we think.

ACKNOWLEDGMENTS

We thank Tho Pham, participants of Royal Economic Society 2022 Conference and seminar participants at the University of Sheffield for valuable comments and suggestions. Finally, we thank Urban Big Data Centre for giving access to the data.

¹³ See <https://tinyurl.com/lrei773h> (accessed on November 5, 2021).

¹⁴ See <https://tinyurl.com/lb3dkrml> (accessed on November 5, 2021).

ORCID

Robert Hill  <https://orcid.org/0000-0002-7483-9701>

Oleksandr Talavera  <https://orcid.org/0000-0002-4799-778X>

Okan Yilmaz  <https://orcid.org/0000-0002-0553-8518>

REFERENCES

- Action on Smoking and Health [ASH]. (2018). Smoking in the home: New solutions for a Smokefree Generation. Available at <https://tinyurl.com/1mn7oady>
- Action on Smoking and Health ASH (2019). Smoking and poverty. Available at <https://tinyurl.com/n0j3cn7w>
- Baranzini, A., Schaerer, C., & Thalmann, P. (2010). Using measured instead of perceived noise in hedonic models. *Transportation Research Part D: Transport and Environment*, 15(8), 473–482.
- Benjamin, J., Jud, G., & Winkler, D. (2001). The value of smoking prohibitions in vacation rental properties. *Journal of Real Estate Finance and Economics*, 22(1), 117–128.
- Blackwell, M., Iacus, S., King, G., & Porro, G. (2010). cem: Coarsened exact matching in Stata. *Stata Journal*, 9(4), 524–546.
- Böckerman, P., Hyytinen, A., & Kaprio, J. (2015). Smoking and long-term labour market outcomes. *Tobacco Control*, 24(4), 348.
- Bond, S. A., & Devine, A. (2016). Certification matters: Is green talk cheap talk? *The Journal of Real Estate Finance and Economics*, 52(2), 117–140.
- Bracke, P. (2015). House prices and rents: Micro evidence from a matched dataset in Central London. *Real Estate Economics*, 43(2), 403–431.
- Flage, A. (2018). Ethnic and gender discrimination in the rental housing market: Evidence from a meta-analysis of correspondence tests, 2006–2017. *Journal of Housing Economics*, 41, 251–273.
- Fuerst, F., Rottke, N. B., Zietz, J., & Reichardt, A. (2012). Sustainable building certification and the rent premium: A panel data approach. *Journal of Real Estate Research*, 34(1), 99–126.
- Goodchild, M., Nargis, N., & Tursan d'Espaignet, E. (2018). Global economic cost of smoking-attributable diseases. *Tobacco Control*, 27(1), 58.
- Haag, J. T., Rutherford, R. C., & Thomson, T. A. (2000). Real estate agent remarks: Help or hype? *Journal of Real Estate Research*, 20(1), 205–215.
- Halket, J., Nesheim, L., & Oswald, F. (2021). The housing stock, housing prices, and user costs: The roles of location, structure, and unobserved quality. *International Economic Review*, 61, 1777–1814.
- Heston, A., & Nakamura, A. O. (2009). Questions about the equivalence of market rents and user costs for owner occupied housing. *Journal of Housing Economics*, 18, 273–279.
- Hill, R. J., & Syed, I. (2016). Hedonic price-rent ratios, user cost, and departures from equilibrium in the housing market. *Regional Science and Urban Economics*, 56, 60–72.
- Iacus, S., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20(1), 1–24.
- King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis*, 27(4), 435–454.
- Levine, P. B., Gustafson, T. A., & Velenchik, A. D. (1997). More bad news for smokers? the effects of cigarette smoking on wages. *Industrial and Labor Relations Review*, 50(3), 493–509.
- Luchtenberg, K. F., Seiler, M. J., & Sun, H. (2019). Listing agent signals: Does a picture paint a thousand words? *The Journal of Real Estate Finance and Economics*, 59(4), 617–648.
- Nowak, A. D., Price, B. S., & Smith, P. S. (2021). Real estate dictionaries across space and time. *The Journal of Real Estate Finance and Economics*, 62(1), 139–163.
- Nowak, A., & Smith, P. (2017). Textual analysis in real estate. *Journal of Applied Econometrics*, 32(4), 896–918.
- Nowak, A. D., & Smith, P. S. (2020). Quality-adjusted house price indexes. *American Economic Review: Insights*, 2(3), 339–56.
- Palacios, J., Eichholtz, P., Kok, N., & Aydin, E. (2021). The impact of housing conditions on health outcomes. *Real Estate Economics*, 49, 1172–1200.
- Pichon-Riviere, A., Alcaraz, A., Palacios, A., Rodriguez, B., Reynales-Shigematsu, L. M., Pinto, M., Castillo-Riquelme, M., Torres, E. P., Osorio, D. I., Huayanay, L., Munarriz, C. L., de Miera-Juárez, B. S., Gallegos-Rivero,

- V., Gallegos-Rivero, V., De La Puente, C., Del Pilar Navia-Bueno, M., Caporale, J., Roberti, J., Virgilio, S. A., & Bardach, A. (2020). The health and economic burden of smoking in 12 Latin American countries and the potential effect of increasing tobacco taxes: An economic modelling study. *The Lancet Global Health*, 8(10), e1282–e1294.
- Pryce, G., & Oates, S. (2008). Rhetoric in the language of real estate marketing. *Housing Studies*, 23(2), 319–348.
- Robinson, S. J., & Sanderford, A. R. (2016). Green buildings: Similar to other premium buildings? *The Journal of Real Estate Finance and Economics*, 52(2), 99–116.
- Soyeh, K. W., Wiley, J. A., & Johnson, K. H. (2014). Do buyer incentives work for houses during a real estate downturn? *The Journal of Real Estate Finance and Economics*, 48(2), 380–396.
- Urban Big Data Centre (UDBC). (2018). Private sector rents in UK cities: Analysis of Zoopla rental listings data. <https://tinyurl.com/17w3dqwr>
- van Ours, J. C. (2004). A pint a day raises a man's pay; but smoking blows that gain away. *Journal of Health Economics*, 23(5), 863–886.
- Weng, S. F., Ali, S., & Leonardi-Bee, J. (2013). Smoking and absence from work: Systematic review and meta-analysis of occupational studies. *Addiction*, 108(2), 307–319.
- Yoshida, J., Seko, M., & Sumita, K. (2016). The rent term premium for cancellable leases. *The Journal of Real Estate Finance and Economics*, 52(4), 480–511.
- Zhou, X., Tong, W., & Li, D. (2019). Modeling housing rent in the atlanta metropolitan area using textual information and deep learning. *ISPRS International Journal of Geo-Information*, 8(8), 349.
- Zoopla Limited. Economic and Social Research Council. Zoopla Property Data [data collection]. Urban Big Data Centre, University of Glasgow.

How to cite this article: Gedikli, C., Hill, R., Talavera, O., & Yilmaz, O. (2022). The hidden cost of smoking: Rent premia in the housing market. *Real Estate Economics*, 1–19. <https://doi.org/10.1111/1540-6229.12399>



APPENDIX

TABLE A1 Phrases used to restrictions on smoking

Negative phrases		
cannot accept dss applicants or smok	no housing benefit, pets, smok	not accept pets or smok
cannot accept pets, dss or smok	no housing benefit, smok	not accept pets, housing benefit, or smok
cannot accept sharers, smok	no housing benefits or smok	not accept pets, housing benefit, smok
exclusions 1. dss, 2. students, 3. pets, 4. smok	no housing benefits, pets, or smok	not accept pets, smok
exclusions 1. dss, 2. students, 3. smok	no housing benefits, pets, children, or smok	not accept pets, students, housing benefit, or smok
exclusions 1. pets, 2. dss 3. students, 4. smok	no housing benefits, pets, smok	not accept pets, students, housing benefits, smok
exclusions 1. pets, 2. smok	no housing benefits, pets, students, or smok	not accept pets, students, smok
exclusions 1. smok	no housing benefits, smok	not accept smok
is not suitable for smok	no pet, smok	not accept students, housing benefit, pets, or smok
no benefit applicants, children, pets or smok	no pets/dss/smok	not accept students, housing benefit, smok
no benefit applicants, pets or smok	no pets /smok	not accept students, smok
no benefit applicants, smok	no pets dss or smok	not accepted: pets, smok
no benefit or smok	no pets dss students or smok	not accepted: smok
no benefits or smok	no pets housing benefit or smok	not accepted: students, pets, smok
no benefits, pets or smok	no pets or smok	not accepted: students, smok
no benefits, smok	no pets, benefits, or smok	not accepted smok
no children/pets/smok	no pets, children, or smok	not allow smoking inside
no children or smok	no pets, children, dss, or smok	not available for dss applicants or smok
no children, pets, or smok	no pets, children, smok	not available to students, housing benefit recipients, those with pets, or smok
no children, pets, smok	no pets, dss or smok	not be suitable for pets, smok
no children/pets/smok	no pets, dss, children or smok	not consider pets, smok
no dhss, pets or smok	no pets, dss, or smok	not consider smok

(Continues)



TABLE A1 (Continued)

Negative phrases	
no dhss, smok	no pets, dss, students, or smok
no dhss, smoking or pets	no pets, housing benefit or smok
no dss/pets/smok	no pets, housing benefits, or smok
no dss applicants or smok	no pets, smok
no dss applicants, pets, or smok	no pets, students, or smok
no dss applicants, smok	no pets, dss, or smok
no dss or smok	no pets/children or smok
no dss pets or smok	no pets/dss/smok
no dss, children, smok	no pets/smok
no dss, fasttrak, pets, or smok	no sharers, smok
no dss, hb, pets, or smok	no smok
no dss, pets, or smok	no students, children, pets, smok
no dss, pets, children, or smok	no students, dss or smok
no dss, pets, children, smok	no students, dss, pets, or smok
no dss, pets, or smok	no students, pets, or smok
no dss, pets, smok	no students, smok
no dss, pets, students, or smok	no: smok
no dss, smok	no: dss, pets, or smok
no dss, students, or smok	non smok
no dss, students, pets, or smok	none smok
no dss, students, pets, children, or smok	non-smok
no dss/ housing benefit, pets, or smok	nosmok
no dss/ housing benefit, pets, children, or smok	not able to accept hb, pets, or smok
	smok
	not permitted to smok
	not suitable for children, pets, students, or smok
	not suitable for pets or smok
	not suitable for pets, smok
	not suitable for pets, students, or smok
	not suitable for smok
	not suitable for the following: pets, smok
	not suitable for the following: smok
	not willing to accept pets or smok
	smokers: outside only
	smokers & housing benefits are not considered
	smokers and housing benefits are not accepted
	smokers are ineligible
	smokers are not considered
	smokers not considered
	smokers or housing benefit not permitted
	smokers & housing benefits are not considered
	smokers: no
	smokers: not allowed
	smoking (inside the property) will not be accepted
	smoking (inside) is not accepted
	smoking (inside) not accepted
	smoking and pets are not

(Continues)



TABLE A1 (Continued)

Negative phrases		
no dss/pets or smok	not accept dss, pets, and smok	smoking and pets are not permitted
no dss/pets/smok	not accept dss, pets, or smok	smoking is not permitted
no dss/smok	not accept dss, smok	smoking is strictly forbidden
no hb, pets, or smok	not accept housing benefit or allow pets or smok	smoking not accepted
no housing benefit/pets/smok	not accept housing benefit, pets or smok	smoking not permitted
no housing benefit/smok	not accept housing benefit, pets, sharers or smok	unable to accept housing benefit or smok
no housing benefit or smok	not accept housing benefit, pets, smok	unable to accept housing benefit, smok
no housing benefit, children, pets, or smok	not accept housing benefit, smok	unable to accept pets, smok
no housing benefit, pets, or smok	not accept housing benefits, pets or smok	
no housing benefit, pets, or smok	not accept housing benefits/smok	
Positive phrases		
accepted: students, pets, smok	allow pets, students, and smok	smoking considered
smokers accepted	dss, pets, and smokers allowed	smokers possible
accepted: pets, smok	yes smok	possible smokers
accepted: students, smok	smokers allowed	negotiable smok
smokers: allowed	smoking and pets are not permitted unless prior permission is agreed	
smokers welcome	smokers considered	

TABLE A2 Coarsened exact matching: match statistics (based on county-level geographical data)

Matching summary		
Number of strata	58,934	
Number of matched strata	20,422	
	Smoking is not restricted	Smoking is restricted
All	2,563,761	162,507
Matched	2,355,424	160,319
Unmatched	208,337	2188
Multivariate L_1 distance	0.309	

Note: The L_1 statistic is a measure of imbalance across the treated and control groups and, $L_1 = 0$ in case of perfect global balance and larger values point to a larger imbalance with $L_1 = 1$ indicating the complete separation (Blackwell et al., 2010). The second column reports the difference in means between the treated and control groups.

TABLE A3 Coarsened exact matching: match statistics (based on 11 NUTS1 regions-level geographical data)

Matching summary when region is used instead of county		
Number of strata	37,846	
Number of matched strata	13,809	
	Smoking is not restricted	Smoking is restricted
All	2,563,761	162,507
Matched	2,405,719	161,395
Unmatched	158,042	1112
Multivariate L_1 distance	0.000	

Note: The L_1 statistic is a measure of imbalance across the treated and control groups and, $L_1 = 0$ in case of perfect global balance and larger values point to a larger imbalance with $L_1 = 1$ indicating the complete separation (Blackwell et al., 2010). The second column reports the difference in means between the treated and control groups.