university of groningen / faculty of spatial sciences

QUANTITATIVE APPROACH ON THE ASSOCIATION OF PUB DENSITY WITH LOCAL PROPERTY PRICES

Abstract. Pubs play an important within Britain's urban communities, contributing to cohesion and social life for ages. Since the pub used to have attractive power for local inhabitants, this study investigates whether premiums in property prices can be observed when being attached with higher pub densities in proximities. Given the fact that London suffered a severe pub loss from 2001 to 2017, residential property prices within this timeframe are used over 32 boroughs. Panel data is used to conduct fixed effects and semi-parametric estimation techniques allowing for linear as well as non-linear relationships between pub densities and local property prices. The results show that after controlling for unobserved heterogeneity, an increase in pub density is gifted with increasing property prices up until two pubs per square kilometer. Hereafter, increasing pub densities show a declining relationship with residential property prices. Those findings suggest an overrepresentation of negative externalities wedded to higher pub densities. Existing literature highlights the effect of pubs for surrounded areas in terms of noise, crime and disturbing behavior, but is not able to monetize such effects. This study intuitively gives some insides to what extent the presence of a pub can be considered as an amenity, reflected in local property prices.

Keywords: Pubs, Residential Property Prices, London Housing Market, Amenities, Local Preferences, Panel Data, Urban Landscape.

Title	A Quantitative Approach on the Association of Pub Density with Local Property Prices.
Version	Final
Date	2021-05-25
Author	Max Janssen
Student number	S3936236
E-mail	m.janssen.8@student.rug.nl
Supervisor	Mark van Duijn
Second reader	Xiaolong Liu

Table of Contents

1. Introduction 2
   1.1 Motivation 2
   1.2 Literature review 3
   1.3 Research problem statement 4

2. Theoretical Framework 6
   2.1 Role of the pub as an amenity 6
   2.2 Amenities on house prices 7
   2.3 Locational preferences for public amenities 8
   2.4 Determinants for house price dynamics 9

3. Data description 12
   3.1 Panel and time dimension 12
   3.2 Variables 12
   3.3 Preliminary analysis 13

4. Methodology 15
   4.1 Fixed effects model 16
   4.2 Random effects model 18
   4.3 Semiparametric analysis (non-linear relation test) 19

5. Results 22
   5.1 Fixed effects estimates 22
   5.2 Non-parametric estimates 24

6. Discussion 27

7. Conclusion 28

Reference list 30

Appendix 1: Diagnostics 35
   A: Testing for Fixed effects (F-test) 35
   B. Hausman model specification test 36
   C. Strict Exogeneity 37
   D. Multicollinearity Test 37
   E. Heteroscedasticity Test 38
   F. Serial(Auto) and Cross-sectional autocorrelation 39

Appendix 2: statistics 40
   Descriptive statistics Inner and Outer London 40

Appendix 3: Chow F-test on heterogeneity effect of GDP 41
   A. Poolabiliy-regression 41
   B. STATA Computation of Poolability 42
   C. Manual Computation of Poolability 42

Appendix 4: Robustness semi-parametric model 42

1. Introduction

1.1 Motivation

It is common sense that houses which are located in proximity to local amenities will attract homebuyers. An amenity that should not be undermined is the presence of a pub. Reason for this is a UK national survey in 2015, which investigated that a third of the homebuyers would check if their targeted property is near a good public house (Savills, 2019). In addition, Metro (2016: 1) quotes '*It's official: Having a local pub makes you happier'*. The newspaper refers here to a research of the University of Oxford, confirming that people who are regulars at local pubs tend to be happier, have more friends and have a better life satisfaction than those who don't. Professor Robin Dunbar, the researcher of the study states: '*Given the increasing tendency for our social life to be online rather than face-to-face, having relaxed accessible venues where people can meet old friends and make new ones becomes ever more necessary.*' (Digital Journal, 2016: 1)

Although pubs turned out to be important for urban life, the UK has lost 21,000 pubs since 1980. More than a half of those loses took place since 2006 (Institute of economic affairs, 2014). Also the BBC (2017: 1) quotes that '*London Loses 1,200 pubs in 15 years'* . London mayor Sadiq Khan was promoting an annual check of closing pubs to prevent even further closures, this in order to maintain the image of London as cultural capital and pubs as its product of identity. Several factors have contributed to this decline. One of those factors are the changes that took place in the pub sector. In 1989, the Beer Orders induced by the Parliament, forced large beer brewers to get rid of a proportional quantity of pubs in order to create a formation of large retailing companies (Pratten, 2007). Those companies purchased the majority of the pubs and selected a small range of suppliers, causing fewer possibilities for other breweries. Consequently, the rise of corporate pub companies like Wetherspoons induced the decline of smaller independently owned pubs (Preece, 2008). Other reasons that contributed to the increasing pub closure in the UK are due to the financial crisis in 2008, the decrease of alcohol prices in local supermarkets and growing "home entertainment facilities" such as television, video games and sound systems. Moreover, changes in pub visits are shifted to the weekend instead of distributed through the week (Smith and Foxcroft, 2009). The decline in pub density may result in the relapse of social cohesion and liveliness, resulting in a less attractive neighborhood to life in.

In contrary, too many local pubs can result in negative pricing effects. Evidently, bars and pubs are recurrently related with noise pollution, public disorder and violent behavior (Collins 1982; Lang & Rumbold 1997; Plant et al. 2002). When having higher pub densities in a particular neighborhood, it is likely that those negative pricing effects will have a greater influence on local properties. Despite that the role of the pub is frequently examined in previous literature studies as a (dis)amenity creator, there is still a lack of longitudinal studies, disentangling the economic value that a pub can give to local property prices. Since the presence of amenities as well as disamenities in proximity is the main consideration in a households' purchase preference in the UK (Daly et al. 2003), examining the association between pub densities and local property prices might give some insights to what extent pubs are welcome assets in the neighborhood. First thoughts are that this is not a linear process: low pub densities might result in lagging amenity levels, whereas too high pub densities might result in an overload of negative externalities such as noise pollution, vandalism and alcohol abuse. The aim of this research is to discover the association between different pub density levels and local property prices in 32 London boroughs between 2001 and 2017.

1.2 Literature review

Early literature already showed the importance of access and proximity to amenities for households' choice to settle in a certain area (Ding et al., 2010). Studies such as Cheshire & Sheppard (1995) shows that amenity and locational specific characteristics are reflected in land prices and therefore on property prices. 23 Percent of the homebuyers admits that a good local pub is one of the main considerations in choosing a location. This suggests that the presence of a pub in a certain location may have a positive effect on house prices (Tepilo, 2015). To clarify this percentage, pubs appear to offer opportunities to strengthen the social aggregation and communal activities. A range of studies confirmed the positive effects of pubs on local communities. Bowler & Everitt (1999) prove that the effect of pubs on community cohesion turned out to be greater than in other public spaces such as community halls, parks and libraries (Cabras & Lau, 2019). Changes in pub densities, subsequently can influence attractiveness and communal cohesion. Since the evidence for the positive contribution of pubs on social life is highlighted in academic literature, it is probable to state that house prices may exhibit higher prices to some extent of pub densities (Andrews & Turner 2011; Oldenburg & Brisett 1982; Cabras 2011).

On the other hand Bottoms and Wiles (1997) recognize a link between the location of crimes and the location of licensed premises. Crime and public disorder are negative pricing effects, that can overtake the positive amenity effects of having a pub in proximity. Even though the effect of criminal damage on house prices is negative, Gibbons (2004) uses local public house density (licensed premises per square kilometer) to capture the direct amenity effects from public houses on property prices and finds a positive effect. The results show that licensed premises like wine bars and pubs are beneficial for local property prices. Although this study finds positive effects of pubs per square kilometer on property prices, the study assumes that the effect is causal and linear. Hypothetically, the effect of pub densities on property prices is unlikely to be positively linear for an infinite value of pub densities. Moreover, the study uses data from 2001, which may be outdated, considering the fact that house prices as well as preferences for pub densities in proximity could have changed last decades. Moreover, the use of aggregate data makes it hard to say if the effect is really causal or not. Therefore, this study will serve as critical additional material with regards to the association between pub densities and local property prices in contemporaneous urban life.

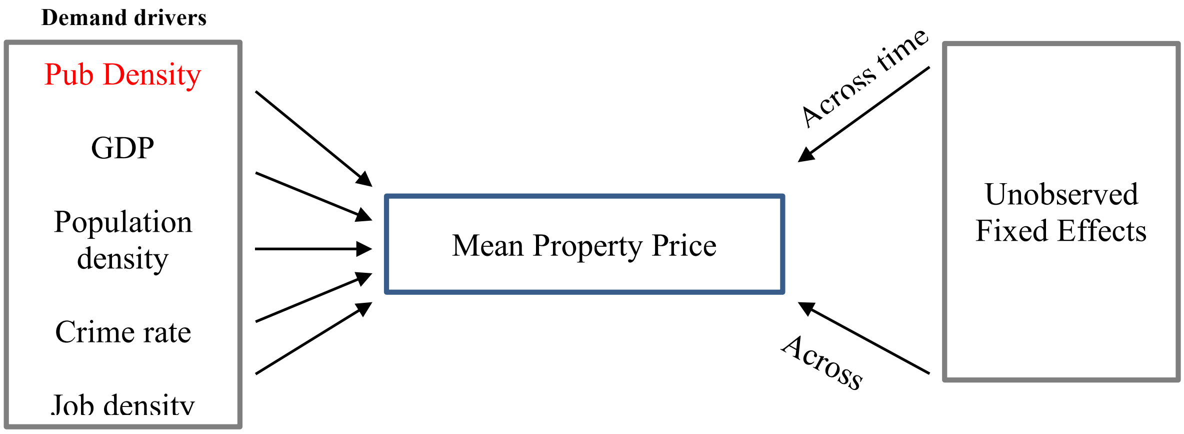
1.3 Research problem statement

In order to answer how pubs are effecting property prices, the following main question is established:

What is the effect of pub density on residential real estate prices?

The research aim of this study is measure the effect of pub density (pubs per square kilometer) on house prices. Regarding remarkable changes in the quantity of pubs last decades, data that is available over time and across sections is used to capture the importance of pub density on house prices. To make a better estimation of the association of pub density on property price, several macro-economic variables will be used to explain property prices. A panel data regression will be executed in order to analyze the association of pub densities with property prices over time and across different London Areas, categorized as “Boroughs”. The panel data is available from 2001 to 2017, so the association will be based on variable values in this period. Furthermore, different control variables will be added to the model in order to prevent omitted variable bias. In addition, it could be the case that pub density is not fully exogeneous, since pub densities could be influenced by house price developments in certain neighborhoods. It could be that pubs are replaced by housing or that pubs are replaced by substitutable institutes like coffee bars or lunch rooms. Changing house prices can alter the sorting of different income groups into the neighborhood. Those newcomers might have different preferences and thus can evoke the change of pubs in the neighborhood.

Figure 1: Conceptual model



Based on existing literature, the following questions are composed to operationalize the effect of pub density on residential real estate prices:

Question 1: What is the association of pubs with local property prices?

This research question will explore the relationship between pub densities and local property prices. As already mentioned, the effect association in a specific borough will by compared with other boroughs. The model will be executed by a panel data model given the fact that the dependent variable '*Property Price*' is a variable changing over time, explained by other variables, also changing over time. To control for location specific factors, variables such as job density, GDP, crime rates and population density are included in the model. Those variables are used as control variables to capture the unobserved house price dynamics as well as possible. Since the association of pub densities with residential property prices is unknown yet, the increased real estate prices of precedent years could follow a trend with the decline in pub densities in several boroughs in the city of London. Given that existing literature is not agreeing upon the contribution of pub densities on property price dynamics, it could be the case that positive and negative externalities are expressed in local property prices. More flexible, semi-parametric estimation model will be conducted to explore to what extent the estimated relationship is in line with current literary exhibitions.

2. Theoretical Framework

In this section of the research, theoretical arguments are used to justify the prepared hypothesis that the pub density may have a positive effect on residential property prices. The theoretical framework is based on three segments to identify the role of pubs and the link with local property prices. First will be explained what the role of the pub is in daily life and why the pub can be seen as local amenity. Second, theory will elaborate the effect of local amenities on house prices. In the end, some theory about the key determinants of house prices will be provided in order to find the variables that could fit into the model. Those variables will be used in order to make the estimates of pub density on house prices as trustworthy as possible.

2.1 Role of the pub as an amenity

To consider the pub as an amenity, there will be suggested that the pub as a meeting place contributes to the satisfaction of social interaction and leisure activities. Apardian & Reid (2020) argue that the pub can be seen as a “Third Place”. Oldenburg (1997) describes the third place as informal public gathering places. Oldenburg considers our homes as the “First Place”, the office as “Second Place” and the sees the “Third Place” as just as essential. The extent to which an individual enjoys living is mainly based on the company of those who live around them. Therefore, the Third Place must be provided and be available at a place where people live. Oldenburg & Brisett (1982: 269) even quote: ‘*tavern, or bar, is without doubt the dominant third place in our society*’. Given the fact that the pub has an inclusive character and conversation is the main activity, the role of the pub in creating social networks should not be undermined. Cohen (2004) strengthens this argument by arguing that being active in a social network can deliver some benefits in reducing stress, loneliness and some clinical illnesses. Cabras (2011) states in his article that the presence of pubs plays an important role in stimulation the social fabric of the neighborhood. To summarize the role of pubs in local communities, amenities such as pubs are vital in creating a higher level of community cohesion and therefore deliver some valuable assets for local residents. Andrews & Turner (2011: 542) elaborate this by stating: '*The public house has a pivotal function within both rural and urban communities and contributes to the economy, cohesion and social life in Britain and has done so for hundreds of years*'. Previous research papers have focused on the effect of community cohesion at higher urban levels (Putnam 2000; Tolbert et al. 1998). These studies explored the positive role of community cohesion in creating a community with a higher overall well-being. Furthermore, Moorman et al. (1992) connect increasing levels of neighborhood involvements and community cohesion to extended economic activities.

On the other hand, Green et al. (2007) highlight the presence of violence, vandalism, public disorder and noise pollution in proximity of bars. Zhu et al. (2004) use a data analysis of census tracts to investigate the violent crime rates when the alcohol outlet density is increasing. Their results suggest that the presence of alcohol serving premises are fundamental in explaining violent crime rates in proximity. In line with Zhu et al.(2004), Scribner et al. (1995) recognize that higher levels of alcohol outlets are related to higher levels of assaults. Norström (2000) and Van Oers & Garretsen (1993) strengthen those findings with arguing that there is a strong association between the density of bars and greater levels of criminal violence and traffic injuries per neighborhood. More than physical disamenities, noise pollution is also on key detractor when having a higher pub density in the neighborhood. Summing up all the pros and cons that are attached to the pub as a public good, no concise answer can be given on whether the pub can be seen as an amenity or disamenity per se.

2.2 Amenities on house prices

Many studies addressed the importance of amenities on the attractiveness of an area. The main point that should be addressed is whether the increase of attractiveness leads to higher property prices. The attractiveness of areas is not fully dependent of agglomeration benefits such as wages and rents but also on urban amenities (Glaeser et al., 2001). Areas are meant to be successful consumer havens in order to maintain attractiveness. Cities or areas that offer better natural amenities or a wide range of consumer goods and local public services are seen as attractive (Glaeser & Gottlieb, 2009). Individual households are tending to base their choice on where to settle down on utility. Migration across locations will continue as long as differences in utility offered by a certain location will exist. According to Florida (2008) cities are place where people meet and interact with each other in order to produce new ideas and knowledges. This attraction of the creative class may help in developing the cities popularity. De decision where to live is for the creative class commonly based on the lifestyle quality such as access to urban amenities. In addition, Shimizu et al. (2014) argue that the role of cities has changed from a "place of production" to a "place of consumption". Moreover, Shimizu et al. (2014) describe that when the supply of land is fixed, de demand for areas with high level of amenities will increase and will be reflected in housing prices. In contrary, high levels of amenities with negative externalities will decrease the housing prices. Moreover, "Concentration of diverse amenities enhances the appeal of an area and, as a result, households will seek to reside there, even if it means paying higher rent" (Shimizu et al., 2014)

Gillard, Q. (1981) argues that differences in property values are positively affected by a variety of urban amenities such as greenbelts, open space, water access. Disamenities such as air pollution, groundwater contamination, airport noise and traffic noises and other undesirable land use characteristics are determined to have a negative impact on house prices. The most important point is that the effect of adjacent land uses on house values varies by location (Nelson, 1985). The beneficial effect of having a certain amenity in proximity will depend on the dominance of positive externalities relative to negative ones, created by the amenity itself. For example, living near a factory can cause negative externalities such as air pollution, but in contrary, can decrease the travel time when working in the factory. In conclusion, the importance of amenities on the attractiveness of an area is shaped by the composition and variety of amenities and the externalities that are produced by them.

2.3 Locational preferences for public amenities

It is commonly known that locations with a strong economy are often attached with a lively and vibrant character. Individual inhabitants of those areas are often more skilled and earn a higher nominal wage compared to areas with lower levels of amenities. Having a higher nominal wage in general will increase the ability for an individual to enjoy a broader range of consumption and leisure (Florida 2002, Gleaser et al. 2001). Higher skilled individuals show a tendency of to live in more expensive parts of the city. Moreover those individuals are willing to spend a higher proportion of their income to local amenities and leisure events. Gagliardi & Schlüter (2015: 23) quote: *"Large metropolitan areas where investments in beautification and local amenities are reflected in higher housing price and cost of living might experience* *significant changes in their demographic composition".* This indicates that the clustering of well-educated and rich people will indirectly increase the demand for amenities, whereas lower income groups will do the opposite.

In addition, the value that residents place in having a local amenity is an important factor in de demand for a certain good. In this case, the term "option value" is used in a situation where an individual is uncertain whether he will demand a certain good regarding the supply of that same good in the future (Freeman, 1985). The option value is directly related to the maximum willingness to pay for a certain good in the future concerning the supply at a certain point in time. The option value is attached with the willingness to preserve a good even though there is no likelihood the individual is ever going to use that particular good. Due to uncertainty about future demand and supply of a certain good, the option value can either be positive or negative. Freeman (2014:165) quotes '*if an individual was uncertain about future income and the demand for the good in question was a positive function of income, then that demand uncertainty option value is unambiguously negative for risk-averse individuals*'. To elaborate this, when an individual is uncertain about their demand (driven by income) for a certain good, the willingness to maintain access to the good in question will be lower than a risk-neutral/taking individual. Therefore the option value includes how individual risk influences the decision to pay for preservation of local amenities or goods.

2.4 Determinants for house price dynamics

Regarding the fact that a house is probably the most valuable asset for most of the UK households, several studies have done research to the composition of house prices. Most of them have tried to cover the key determinants of house price determinants. Rosen (1974) uses a hedonic price model to estimate the values of different housing characteristics to house prices. Rosen (1974) uses housing characteristics like size, location and access to several amenities to explain the value of a house as bundled package of characteristics. The model assumes that the marginal benefits of moving to a better neighborhood, larger house or living closer to desirable amenities will lead to a premium regarding house prices. The model of Rosen emphasizes the willingness to pay for a range of attributes.

In addition to Rosen, DiPasquale & Wheaton (1992) illustrate how real estate is impacted by a nation's macroeconomics and its financial markets. DiPasquale & Wheaton (1992) use the four quadrant model to explain rents, asset prices, construction and the housing stock by exhibit different exogenous shocks. The framework, divided in four quadrants is meant to explain how exogenous changes lead to new equilibria of (1) the asset market, (2) the property market rents, (3) the asset market construction and (4) the property market stock change. The model assumes mutual interdependence of the different quadrants to come to an equilibrium. This framework identified that effective house prices tend to increase when construction prices do increase. Increase construction prices can lead to a decrease in new built units, which influences the housing stock negatively. Scarcity in housing stock, therefore leads to higher rents and property prices. It is a common finding, that areas where housing supply is more limited due to legal constraints or geographical constraints, house price dynamics tend behave in a different way. Glaeser et al. (2008) for example, explains that regions with greater supply constraints experienced a significantly higher housing boom in the period 1982 until 2007 than other regions.

Subsequently, Capozza et al. (2002) strengthen the importance of macroeconomic factors regarding house price dynamics. This study elaborates the effect of different variables and economic situations on house prices. In this model, population growth, construction costs and income were included. According to Day (2018) the density of population has a positive effect on house prices, because of rising demand for houses. Capozza et al. (2002) further mention the effect of construction costs on house prices. Capozza et al. (2002) argue that the effect of construction costs depends on the submarket and do not have a certain linearity. Local authorities can for example stimulate the building of new houses, given the fact that construction costs could be regulated in a certain way. At last, income growth or GDP is one of the variables Capozza (2002) mentions in his paper. Due to income growth, households do have more money to spend, which logically reflects in a higher supply of money. Therefore, people are more capable of buying a house or renovating a house they prefer. Putting more money into local houses could result in higher house prices in a particular area.

Many studies did research to locational effects on house prices. Alonso (1964) and Brueckner (1996) for example recognize the relationship between land value and distance to the CBD. Due to the fact that commonly is assumed that the demand for a house is derived from the demand for land, house prices tend to increase when being located closer to the CBD in a monocentric city. Moreover, Richardson et al. (1974) imply that neighborhood characteristics are significant in explaining house prices. This neighborhood characteristics that measure local public services are variables like crime rates, employment and facilities. Schwartz et al. (2003) find that almost one-third of the house prices increases were attached to declining crime rates. Moreover, property crime reduces house prices by almost three percent according to Thaler (1978) and Buck et al. (1991). In addition, Li and Brown (1980) emphasize the importance of locational attributes by including "micro-neighborhood" characteristics such as aesthetic quality, noise pollution and specific distances to several amenities. Regarding the significant differences among locations (neighborhoods, boroughs, areas) the impact of certain amenities on house prices can differ a substantial amount when comparing those locations.

Because pubs do offer social benefits, putting a monetary value on their presence in a neighborhood is rather difficult. A solution to express the marginal effect of losing/adding a pub in monetary terms is to use property prices as economic value. Benefits of having a pub in proximity may vary across location and time. Thus, in this research will be investigated wat the effect and relationship is between changes in pub density on property prices between 2001 and 2017 in the City of London. A panel dataset with macro-economic variable of the London boroughs will be used to analyze the time and location specific effects. Although extensive literature puts some general benefits to the presence of local pubs (i.e. social well-being), counter arguments that recognizes the negative external effects of high pub densities (i.e. crime, noise pollution) makes it hard to resolutely draw a conclusion on the question whether higher pub densities will lead to higher real estate prices. Whereas we hypothesize that the benefits of having a pub in proximity will offset the negative externalities in general, we also believe that the premium that higher pub densities may deliver to local real estate prices is different per density level. Ones can argue that having a pub in proximity is always an asset to have, but when they are over-represented, the negative external effects may overturn the positive benefits of consuming a pub. Therefore, we hypothesize that the pub density effect is positive until a certain tipping point. After that point is reached, pubs will bring more complications than being an added value.

To discover how the pub as an amenity will influence local house prices by location, one hypothesis is established:

Hypothesis: 'The association of pub density with local property prices is different for several pub density intervals'

3. Data description

3.1 Panel and time dimension

This research focusses on the association of pub density on property prices in the City of London between 2001 and 2017. The panel that is used do consist of 32 boroughs divided over Inner and outer London. Boroughs where data was incomplete have been excluded from the dataset, since this data could lead to invalid estimation results.

3.2 Variables

Dependent variable: Property Price

DCLG Land Registry (2018) provided the data for mean property price on local authority level per year. The model will be built based on the effect of independent variables on the dependent variable “Property Price”. The mean of the property prices are based on an Arithmetic mean. This implicates that the summed values of sold properties are divided by the amount of sales of the concerning units. Therefore, there should be noticed that boroughs with frequently more sales will have more reliable property price averages than others. Subsequently, the value of the properties is expressed in GBP.

Independent variable: Pub density

In order to estimate the effects of pub density on property prices, the Office for National Statistics (2018) made some datasets available about the quantity of pubs per borough. To account for the size of the area where the pubs are located in, the amount of pubs are divided by size of the borough in km^2. To work with pub density, there will be assumed that all pubs are evenly distributed over the area. The limitation here is that not the direct effect can be measured from a near located pub on a property price, due to the absence of locational information. Another problem could occur in the absence of info about the size and character of a pub . The size and character of a pub can influence the attraction or detraction of a pub since their target audience and visitors may be different.

Control variables

Based on existing literature, four variables are used to control for the effect of pub density on property prices. These variables are GDP per capita, population density (as number of inhabitants per km^2), crime rate (given value based on committed crimes per borough) and job density (jobs per km^2). Based on the literature, table 1 shows if the expected coefficient of a variable is positive or negative. Due to the fact that house prices are determined by demand and supply, the different variables are marked as being a demand or a supply driver in house price determining. At the bottom of the table, the dummy variables "Boroughs" and "Year" are mentioned separately.

Table 1. Variable description

Variable	Aggregation	Source	Unit	Type of Determinant	Expected Sign
Property price	London borough	DCLG Land Registry (2018)	Level		
Pub density	London borough	Office for National Statistics (2018)	Ratio	Demand	Positive
GDP	London borough	Office for National Statistics (2019)	Level	Demand	Positive
Population density	London borough	Office for National Statistics (2018)	Ratio	Demand	Positive
Crime rate	London borough	MPS and Home Office (2017)	Index	Demand	Negative
Job density	London borough	Offic National Statistics (2019)	Ratio	Demand	Positive
Boroughs	London borough	DCLG Land Registry (2018)	Dummy	Demand	None
Year	Year	DCLG Land Registry (2018)	Dummy	Demand	None

3.3 Preliminary analysis

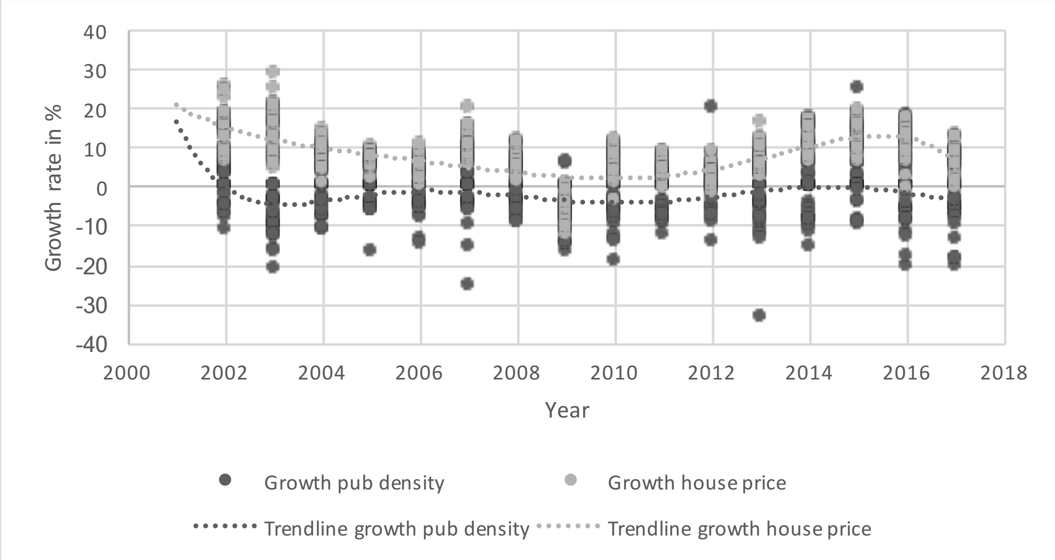
Tables 2 show the descriptive statistics of the dependent variable and the independent variables. There can be noticed that all variables do have the same amount of observations, which can be explained by 17 timestamps (years) divided over 32 boroughs. In addition, there can be observed that there are more observations in Outer London than in Inner London, since Outer London do consist of more authorities. Because every panel has the same amount of observations, we can call the this a strongly balanced dataset. This is important in order to make a valid and reliable regression. In this descriptive statistic table is not log transformed yet in order to get an original interpretation of the available data. When specifying the model, *property price* will be transformed into logarithm to make the variable more normally distributed and . Key thing to mention is that on average, Inner London has more than five times higher pub density than Outer London (appendix 2). Moreover, property prices in Inner London are also higher than in Outer London. The inequality in property prices is larger in Inner London, where the maximum price is more than twelve times higher than the minimum, whereas in Outer London the maximum price is less than eight times higher than the minimum. Also the range between maximum and minimum pub density is higher in Inner London compared with Outer Londen. The maximum pub density is almost twice as high in Inner London than in Outer London. This phenomena can be explained by the higher amount of nightlife facilities clustered in Camden, Chelsea, Islington and Westminster.

Table 2. Descriptive Statistics London

Variable	Obs	Mean	Std.Dev.	Min	Max
Property price	544	309000	159000	84625	1280000
Pub density	544	4.085	4.616	.524	23.377
Crime rate	544	113.095	52.666	45.01	424.368
GDP	544	37690.38	38923.96	10383	283000
Population/km2	544	6720.588	3382.525	1963.327	15842.73
Job/km2	544	4772.909	. 5986.367	698.9508	34725.28

In Figure 2, the change in pub density between 2001 and 2017 is plotted against the change in mean property prices in the City of London. The figure only involves data from 2001 to 2017. Growth rates of pub densities and house prices is computed with regards to their preceding year. The figure indicates that the pub density in the City of London are steadily decline between zero and five percent each year, whereas house prices tend to grow between three and fitheen percent each year. The first guess should indicate a negative relationship between pub density growth and house price growth since house prices are steadily growing whereas pub densities tend to decline over the years. Nevertheless, those are just preliminary intuitions, uncontrolled for time and neighbourhood specific effects.

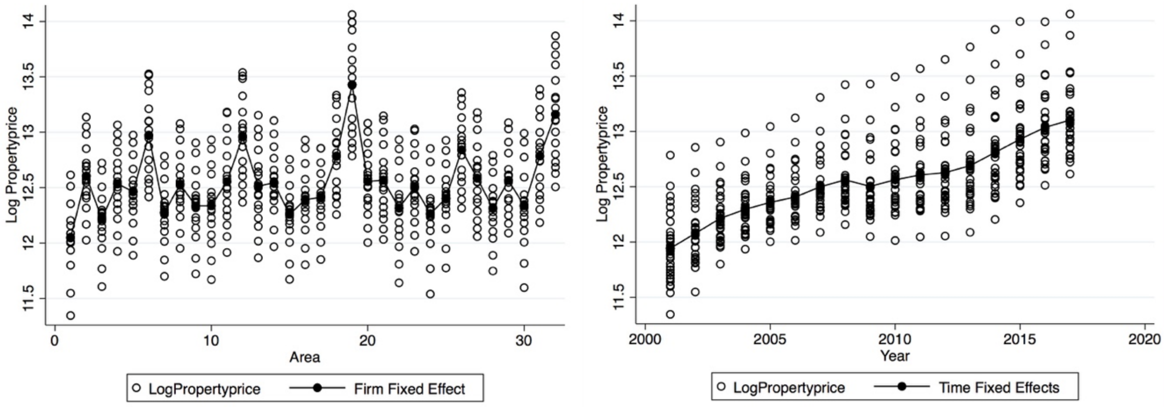
Figure 2. Descriptive statistics pub density vs Change in mean property price between 2001 and 2017



4. Methodology

To test the hypothesis whether the effect of pub density on property prices is positive, there is need to test which kind of model should be used to give the best estimation results. The data that will be used is quantified as Panel data. This implicates that the data is longitudinal and is distributed over several cross-sectional dimensions. This means that entities are observed across time. The entities in this research are London boroughs and are distributed over 17 year. There are several advantages for the use of panel data. Baltagi (2009) argues that panel data allows to control for individual heterogeneity. The data suggests that the panels are heterogeneous. Moreover, panel data gives a broader variety of data and provides in general less collinearity among the variables. Some drawbacks of panel data that should be taken into account are cross-sectional dependency which means that panels could be correlated to each other. Furthermore, due to the longitudinal characteristic of panel data, errors can be correlated over time (autocorrelation), this will threaten the efficiency of the research results. To give a first impression whether some time or locational heterogeneity (fixed effects) are present in dataset of log property price, we sort the data by borough and by year. After that, the mean property price per year (over all boroughs) and per borough (over all years) are executed. Due to the observation that there is quite a lot of variation between boroughs (figure 3), firm fixed effects are considered to be important in when formulating a model (Torres-Reyna, 2007: 13). Moreover, figure 3 showcases an upwards trend in mean property prices over the years. This urges the suggestion that the should be controlled for time specific effects. The heterogeneity that is shown in figure 3 are unobserved variables that do not change over firm and time (Torres-Reyna, 2007: 14)

Figure 3. Area and Year Fixed effects



Based on the Hausman test, executed in appendix 1B, this study will make use of Fixed effects models. First, we will linearly estimate the effect of pub density on log property price, accounted for year and borough fixed effects. In appendix 1a, we reject the null hypothesis that the coefficients for all years are jointly equal to zero. Thus, the effect of unobserved heterogeneity over time and borough is strong enough we should account for it (Torres-Reyna, 2007: 31). Second, we will relax on the assumption of a complete parametric (linear) model by allowing for a non-linear relationship between pub density and log property price.

4.1 Fixed effects model

As a result of cross-sectional differences, de issue of heterogeneity bias occurs when considering Panel data analysis. Since every panel/entity has unique and specific characteristics, the model should include additional parameters to control for this. With panel data, there is a possibility to account for cross-sectional as well as for temporal heterogeneity. To account for individual specific characteristics that can influence the predictor, the Fixed effects model focusses on the relation between predictors and variables within a specific entity over time (Torres-Reyna, 2007). Therefore, the Fixed effects model uses the "Within-estimator", which measures the effect of parameters within individual panels. The Fixed effects model absorbs all time-invariant effects, which makes that the model cannot contain any other variable that only depends on entity. When also correcting for time fixed effects, the model cannot include variables that only depend on time. When this assumption does not hold, multicollinearity will be present between the 'static' variable and the fixed effects. Another characteristic of the fixed effects model is that the entity or year fixed effects should not be correlated with each other as well as their error terms. In this case, there is cross-sectional or temporal dependency in the error term of the fixed effects (appendix 1f). Therefore, one of the assumptions of the fixed effects model is that $E(v_{it}|x_{it}) = 0$ for all times and cross sections. This means that the model is strictly exogenous and assumes that the idiosyncratic error term is not correlated with the parameters and completely unaffected by Y. Another assumption of the fixed effects model are that the regressors are independent and thus free of multicollinearity (appendix 1d). Furthermore, the idiosyncratic error term v_{it} is assumed to be homoscedasticity (constant variance over time) and free of cross-sectional/serial autocorrelation over time and space (appendix 1e).

The standard functional form of a fixed effects model is:

(1) $Y_{it} = (a+u_i) + X'\beta_{it} + v_{it}, \quad i=1,\ldots,N$ and $t = 1,\ldots,T$ (Park, 2011)

i = cross-sectional dimension
t = time dimension,
a = constant,
u_i =entity specific effect (fixed effect),
β = coefficient to be estimated
X' = matrix of the observed explanatory variables
v_{it} = is the time varying error term (idiosyncratic error term).

The fixed effects model demeans entity specific time-constant unobserved heterogeneity. To do so, only the variation within the entity is left: $(Y_{it} - \bar{Y}_i) = \beta(X_{it} - \bar{X}_i) + (v_{it} - \bar{v}_i)$. Noted is that u_i is not in the demeaned within-equation. This implies that for the fixed effects model, time invariant heterogeneity (fixed effect) is allowed Brüderl & Ludwig (2019). The fixed effects estimator is consistent even if $E(x_{it},u_i)$ is not zero. Since the fixed effects is part of the intercept and time invariant, u_i is allowed to be correlated with the other regressors. Hence, OLS assumption 2 is not violated (exogeneity assumption).

Table 3. Assumptions for Fixed effects models

Assumption	**Test Method**
Strict exogeneity	Correlation Matrix
No multicollinearity	Variance Inflation Factor
Homoscedasticity	Modified Wald test for groupwise heteroscedasticity in fixed effect regression model
No autocorrelation	Wooldridge test for autocorrelation in Panel data
No cross-sectional dependency	Pesaran's test for cross-sectional dependency (not an issue for small panels)

4.2 Random effects model

Another estimation method in panel data is the *Random effects model*. Barili et al.(2018: 317) state: *'Random effects models assume that there may be different underlying true effects estimated in each trial which are distribute about an overall mean'*. To elaborate this, the variation across different panels is assumed to be random and not correlated with the independent variables and predictors. Therefore, the random effects model allows the model to have time invariant variables. If there will be assumed that difference across panels will influence the outcome of the dependent variable, the random effects model will be considered. When $E(x_{it},u_i) = 0$, the entity specific errors are not correlated with the parameters, so there is no need to use fixed effects. The random effects model implicates that the errors in u_i are not caused by fixed entity specific effects, but random (Torres-Reyna, 2007). In fixed effects, all unobserved heterogeneity was kept by the intercept, whereas in the random effects model, entity specific effects needs to be specified by additional variables. Due to the fact that u_i is not allowed to correlate with the independent variables, omitted variable bias is more of an issue in the random effects model. For the purpose that u_i is assumed to be random effects, usually u_i is normally distributed. Furthermore no time-constant unobserved heterogeneity and no time varying unobserved heterogeneity is allowed in random effects. The random effects model is composed as follows:

(2) $Y_{it} = a + \beta X_{it} + (u_i + v_{it}) \quad i = 1 \ldots N$ (Park, 2011)

Where the composition of the error term, consists of:

(3) $w_{it} = u_{it} + v_{it}$ (Park, 2011)

Here a is called a "*grand mean*" which mean that all entities do have a common intercept. u_i is assumed to be the error component for the random effects. within-entity specific errors, whereas v_{it} refers to the between-entity errors for the independent variables.

Table 4. Assumptions for Random effects models

Assumption	**Test method**
Normality of the within-entity specific errors ($\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$	Histogram residuals
Normality of the between-entity errors ($u_{it} \sim N(0, \sigma_u^2)$	Histogram residuals
No time-constant unobserved heterogeneity $E(x_{it}, u_i)$, random-effects	Correlation matrix
No time-varying unobserved heterogeneity $E(x_{it}, v_{it})$, strict exogeneity	Correlation matrix
No autocorrelation	Wooldridge test for autocorrelation in Panel data
No cross-sectional dependency	Pesaran's test for cross-sectional dependency

4.3 Semiparametric analysis (non-linear relation test)

In order to obtain a better knowledge about the effect of pub density on mean property prices, semiparametric model estimations are used to capture a plausible non-linear relationship between pub density and mean property prices. Semiparametric regression models are less restrictive since they allow for non-linear functional forms of one of the regressors, whereas other regressors are estimated linearly. In general, semiparametric models are used to fit a parametric model where the functional form of one of the explanatory variables is unknown. Therefore, semiparametric models are known as partially linear (Rouwendal et al, 2014). Formulation of a standard partial linear model can be defined as follows:

(4) $y_{it} = X'_{it}\beta + f(z_{it}) + e_{it}$

Parameter z is defined as a random variable, whereas X' is a vector p-dimensional random variables. This means that the expected value of y_{it} given values of $X'$and z equals equation 4. Put differently:

(5) $\mathbb{E}[y_{it}|X', z] = y_{it} = X'_{it}\beta + f(z_{it})$

Equation 5 assumes the error term having zero mean and being homoscedastic (Lokshin, 2006). Moreover, Lokshin et al.(2006:2) describe the semiparametric part as folllows: *'The function f is a smooth, single valued function with a bounded first derivative'*. In the model $X'_{it}\beta$ will be estimated parametric (linearly), whereas $f(z_{it})$ will be estimated non-parametrically (non-linearly).

In this study, Yatchew's difference estimator (1997) is used to explore the relationship between property prices and pub densities non-parametrically, controlled for linearly estimated control variables. To correctly execute the differenced estimation, the data should be resorted based on the variable of interest, which is z_{it} $(pubdensity)$. This means that pub density is ordered from the lowest density to the highest density. Ordering the subsequent values of pub density close to each other will remove the non-parametric effect in first difference. Differencing a standard semiparametric model like model 4 in first order yields:

(6) $\left[y_{i(t)} - y_{i(t-1)}\right] = [X'_{i(t)} - X'_{i(t-1)}]\beta + \left[f\left(z_{i(t)}\right) - f\left(z_{i(t-1)}\right)\right] + e_{i(t)} - e_{i(t-1)}, with\ n = 2, \dots T$

Because f is marked as a single valued function with a bounded first derivative, the first difference of $f(z_{it})$ will go to zero when the sample size increases (Verardi, 2013). Therefore, $f(z_{it})$ will be cancelled out when the number of observations increase. Parameter vector $X'_{it}\beta$ can be estimated without estimating $f(z_{it})$ specifically. With the non-parametric term $f(z_{it})$ being removed, the remaining parametric (linear) part of the regression will be estimated with simple Ordinary Least Squares (OLS) estimation. . The estimated differenced parameter $\hat{\beta}$ of equation 6 has the following distribution:

(7) $[X'_{i(t)} X'_{i(t-1)}]\hat{\beta}diff \rightarrow N(\beta, \frac{1}{T} * \frac{1.5\sigma_e^2}{\sigma_u^2})$ (Lokshin, 2006) , T = sample size

Where σ_e^2 is the variance of the error term and σ_u^2 the conditional variance of X' given the value of z. This means that the variance of $X'$is dependent of the variance of z. When the differenced estimations of $\hat{\beta}$ are gained in equation 6, non-parametric estimation can be used to estimate the values of f. Finally, $f(z_{it})$ can be estimated by regressing the difference between the dependent variable and the estimated difference vector β on z_{it} semiparametrically:

(8) $(y_{it} - X'\hat{\beta}diff) = X'(\beta - \hat{\beta}diff) + f(z_{it}) + e_{it} = f(z_{it}) + e_{it}$

Due to the fact that $\hat{\beta}diff$ will rally to the true values of β, the parametric part X' will cancelled out and $f(z_{it})$ is left. $X'\hat{\beta}diff$ will be diminished from y_{it} by using a smoothing procedure. This procedure will start once the parametric effect is removed. In essence, a smoothed locally weighted least squares regression will be executed on the ordered data of pub density. In essence, the smoothing procedure fits a line symmetrically between the point estimates, lying on certain levels of pub density. The smoothing procedure can be done according to different bandwidths. Larger bandwidth values are leading to higher degrees of smoothing (Found, 2013). The standard semi-parametric model in this study uses the default value of 0.8 as bandtwith, but will also robustly check for less smoothed values functions of pub density. Yatchew (2000) justifies the use of a semiparametric model by using a V-statistic that is constructed as follows:

(9) $V = \sqrt{mn}\left(\frac{S_p^2 - S_{sp}^2}{S_{sp}^2}\right) -> N(0,1)$

This test consists of m-order differenced data, n - numbers of observations, S_p^2 being the estimated mean squared residuals of the parametric regression and S_{sp}^2 being the mean squared residuals of the semi-parametric model. The null hypothesis is that the regression function has a parametric form against the alternative hypothesis that the regression function has a semiparametric form (Yatchew, 1997). The Semiparametric model that is built and will be estimated by Yatchew's difference estimator (1997) is constructed as follows:

(9) $\text{LogPropertyprice}_{it} = f(\text{Pubdensity}_{it}) + \beta_1 GDP_{it} + \beta_2 Crimerate_{it} + \beta_3 Populationdensity_{it} + \gamma_i + \delta_t + e_{it}$

f(z) = Non-parametric function of Pub density
β = Coefficient of linearly estimated parameter
γ_i = Borough Fixed effects
δ_t = Year Fixed effects

In order to execute the semi-parametric model stated in equation 9, the *plreg* command in stata is used to explore the non-linear relationship between pub density and the logarithm of property price (Yatchew, 1988).

5. Results

5.1 Fixed effects estimates

The results, identifying the effect of pub density on the logarithm of property prices are presented in table 5. The Fixed effects model is built up by doing a stepwise regression. In models 1-4, Pooled OLS results are showing a positive coefficient of pub density on log Property price. After adding all control variables into the equation, the effect of pub density on log property price is positive and significant at 0.1% level (model 4). The coefficient implies that if one pub is added per square kilometer, property prices are increasing with 10.5%, ceteris paribus. Control variables GDP, crime rate and population density are all significant at 0.1% level and show consistent signs from methodological point of view. Even though those results seems quite acceptable, the preliminary analysis on the data (chapter 4) confirms the use of borough and year fixed effects to control for unobserved heterogeneity across time and space. After controlling for year and borough fixed effects, there can be observed that hardly any statistical significant variable is left in the estimation (model 6). Since the R^2 jumps from 0.59 to 0.972 after controlling for borough and year fixed effects, we can conclude that a lot of variation in log property price is sourced in year and borough specific fixed effects. Finally, model 7 represents the model including all control variables, borough fixed effects and year fixed effects. In model 7, Discroll and Kraay standard errors are used to correct for cross-sectionally correlated, heteroscedastic and auto correlated errors. The model shows a negative sign in the coefficient of pub density, although we cannot interpret this due to insignificance. The same analysis holds for GDP, which cannot reject the null hypothesis of no association with log property price. Crime rate and population density are showing both a negative sign and are significant at 5% and 1% level. The analysis shows that when crime rate increases with one unit, property prices decrease with 0.137% ceteris paribus. For population density, one unit increase will result in property prices to decrease with 0.000209% ceteris paribus, which is inconsistent with the methodology and previous regression models.

After analyzing the association of pub density with mean log property prices with Fixed effects models, regression results are showcasing non-significant results. Moreover, the signs in the coefficient of pub density is such different across all models that this study might suggest a non-linear relationship between pub density and property prices. The next section will further tackle this phenomena by allowing for semiparametric series.

Table 5. Regression results of Pooled OLS and Fixed effects models

	Pooled OLS (1)	Pooled OLS (2)	Pooled OLS (3)	Pooled OLS (4)	Fixed effects (5)	Fixed effects (6)	Fixed effects (7)
Pubdensity	0.0406^{***}	0.0136^{*}	0.1000^{***}	0.105^{***}	-0.0265	0.00144	0.00144
	(0.00414)	(0.00583)	(0.00833)	(0.00830)	(0.0374)	(0.0192)	(0.00895)
GDP		0.00000403^{***}	0.00000283^{***}	0.00000252^{***}	0.00000569	0.00000118	0.00000118
		(0.000000564)	(0.000000553)	(0.000000559)	(0.00000284)	(0.000000957)	(0.00000115)
Crimerate			-0.00837^{***}	-0.00838^{***}	-0.00295	-0.00137^{*}	-0.00137^{*}
			(0.000481)	(0.000483)	(0.00151)	(0.000646)	(0.000547)
Population density				0.000000803^{***}	0.00000859^{***}	-0.00000209^{*}	-0.00000209^{**}
				(0.000000232)	(0.00000216)	(0.000000821)	(0.000000542)
Constant	12.38^{***}	12.34^{***}	12.97^{***}	12.77^{***}	10.62^{***}	12.57^{***}	12.57^{***}
	(0.0205)	(0.0206)	(0.0410)	(0.0816)	(0.687)	(0.252)	(0.183)
Borough Fixed Effects	No	No	No	No	Yes	Yes	Yes
Year Fixed Effects	No	No	No	No	No	Yes	Yes
N	544	544	544	544	544	544	544
R^2	0.195	0.245	0.592	0.602	0.766	0.973	0.9728
adj. R^2	0.193	0.242	0.590	0.599	0.764	0.972	0.9726

Notes: Dependent variable is the natural log of mean property price. The Fixed Effects models are expressing the Within-R^2 for the fixed effects models. Heteroscedastic robust standard errors are in parentheses in model 1-6. Discroll and Kraay standard errors are in parentheses in model 7.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In addition to table 5, a heterogeneity check is executed on different levels of GDP, since the association of pub density and local property prices might depend on GDP levels of people residing in the neighbourhood. First, we divide the sample group by the 50 percent lowest income observations and 50 percent highest income observations. Second, we run separate regressions with log property price as dependent variable with pub density, crime rate, population density and year fixed effects as independent variables. Since the data is sorted on

GDP-levels, GDP as well as borough fixed effects[1] are excluded from the model since this would result in multicollinearity issues. Finally, we conduct a Chow test (1960) to check equality between the sets of coefficients in two linear regressions. The null hypothesis of the test is that there is no breaking point in the data. This means that Pooled regression is appropriate. The alternative hypothesis states that the sample should be divided into two groups. Differences in coefficients are displayed in appendix 3a. The regression results show that in the low GDP group, property prices increase with 9.14% when pub density increases by one unit, ceteris paribus. This effect is significant at 0.1% level. When comparing this with the high GDP group, we see that this effect is 5.11% lower. The lower effect of pub density on property prices compared to low income groups is significant at 0.1% level. Appendix 3a as well as Appendix 3b show that the effect of pub density on log property price is significantly different between the two groups at 0.1% level. When comparing the effect of all the independent variables between the two GDP groups we find an F-value of 9.94 computed by STATA (appendix 3b) as well as manually (appendix 3c). This value exceeds the critical F-value (at 5% level). Therefore, we can argue that association of the independent variables with log property prices is different when being sorted in contrasting GDP-levelled areas. The regression results of table 13 in appendix 3a showcasing a lower coefficient of pub density when observing high GDP's than when observing low GDP's. Even though all parameter estimates in table 13 are significant at 0.1% level, we cannot fully interpret the coefficients as being causal since the model is not accounted for borough specific effects. Moreover, the specification of the model is diverging from the earlier composed fixed effects model and serves just as a heterogeneity check between GDP groups.

5.2 Non-parametric estimates

In order to get a better understanding on the pattern of the effect of pub density on mean property prices, semi-parametric model estimation is used. Since, we explored an unknown relationship of pub density with mean property prices, semi-parametric estimates can help to investigate the functional form of the relationship. The parametric control variables X' are estimated by first order-differencing, whereas f is modeled with a smoothed function. By using such more flexible estimation method, we allow for non-linear relationships between the main

[1] Borough Fixed effects cannot be included since the high GDP group consists of more boroughs than the low income group. This would result in different numbers of parameters in both groups and thus unfeasible for a conducting a chow test.

explanatory variable and the dependent variable. The model is an extended version of the fixed effects model, controlled for all control variables and fixed effects. The model is executed in STATA by using the *plreg* command and uses a default bandwidth of 0.8, which represents the smoothness of *f.* The V-test statistic that is computed by STATA after regressing the model yields a significance test on Pub density with V =12.035 with a significance of 0.01% level. Therefore, we reject the null hypothesis that the regression function has a parametric form.

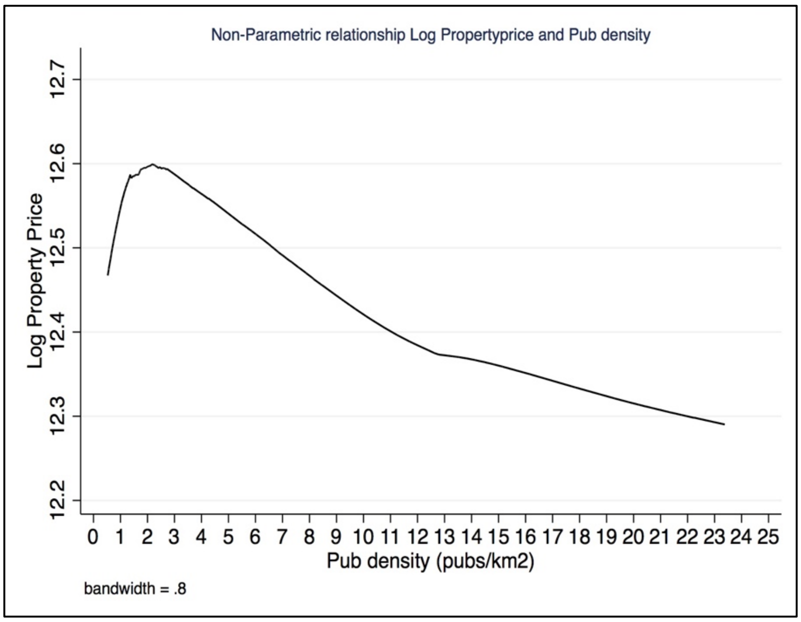
The results in table 6 show that the effect of pub density on mean property prices is the strongest between zero and two pubs per km^2. The estimation results yield a positive sign where mean property prices increase with 7.88% when one pub is added per km^2, ceteris paribus. This effect is significant at 0.1% level. Surpassing the density of two pubs per km^2, there can be observed that the effect of pub density is decreasing property prices with 2.25% per unit increase in km^2, ceteris paribus. This effect is negative and significant at 0.1% level. In the last category of pub densities from 12.5 km^2 to 23.5 km^2, the effect of an extra pub per km^2 is further decreasing, but displays a smaller fall of property prices, compared to the previous density category. Within this category, when the amount of pubs per km^2 increase with one unit, mean property prices decrease with 0.838%, ceteris paribus. This effect shows significance at 0.01% level.

When observing the effect signs of the control variables, population density represents a negative sign in the coefficient results. Even though, the significance is covered at 0.1% level, the sign is contradictive when comparing it with the preliminary expectation. The coefficient implies that one unit increase in population density will decrease house prices with 0.000196%, ceteris paribus. In contrary, GDP shows a positive sign, fully as expected, but is insignificant. Therefore, the result cannot be interpreted. The last control variable Crime rate is consistent with literature, since this coefficient is either negative and significant at 1% level. Therefore, we can interpret the coefficient as follows: when crime rate increases with one unit, property price decrease with 0.0828%, ceteris paribus. After regressing pub density and all control variables on the logarithm of property prices, the model gives an R-squared of 0.979. This means that the variation in the logarithm of property prices is for 97.9% explained by the independent variables.

Table 6. Regression results of Semi-parametric models by Yatchew's difference estimator (1998). Next to the table is a graphical representation of the effect of pub density on log property price, allowing for a non-linear relationship

	Semi-Parametric
	(1)
Pub density km²	
0 – 2 km²	0.0787503***
	(0.0021082)
2 – 12.5	-0.0225158***
	(0.0000765)
12.5 – 23.5	-0.0083769***
	(0.0000718)
Population density	-0.00000196***
	(0.000000416)
GDP	0.000000437
	(0.000000622)
Crime rate	-0.000828**
	(0.000300)
Borough Fixed Effects	Yes
Year Fixed Effects	Yes
N	543
R^2	0.981
adj. R^2	0.979

Notes: Dependent variable is the natural log of mean property price. The coefficients of pub density are manually computed by regressing the estimated values of log property price (plreg) on pub density Standard errors are given within parentheses:* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$



Since the *plreg* function in STATA can also allow for different smoothing bandwidths, the semiparametric function of *f(pubdensity)* is robustly checked in appendix 4. The appendix displays four bandwidths ranging from 0.05 to 0.80 to check whether the association between pub density and log property prices is robust over different sentitivities. The lower the bandwidth, the more accurate the estimation is done to the closes pub density. The main results of table 6 is quite robust ranging from 0.4 and 0.8 as bandwidth. When using very sensitive bandwidths of lower than 0.1, we can see that the association between log property price and pub density represents a small increasing slope between 6-8 pubs per square kilometer, but is steadily decreasing afterwards in line with table 6. Generally speaking, after checking for different sensitivities of the non-linear relations between pub density and log property prices, we can say that above two pubs per square meters, house prices are declining almost asymptotically.

6. Discussion

We observe in the final fixed effects model (model 5) that the effect of pub density showcases a positive effect on the Logarithm of property price. This effect should be in line with Gibbons (2004) and Cabras (2011), who argue that pubs play a role in the creation of community cohesion and social fabric in the neighbourhood. Moreover, the comment is made that some pubs are preferred in the neighbourhood, unconcerned of the travelling time. Respected the insignificance of pub density as explanatory variable for property prices, we cannot agree upon those theories. Based on linearly regressed data, there cannot be concluded that pub density has a significant positive effect on property prices. Taking the conflicting literary arguments of the effect of pub density on house price dynamics into consideration, the semiparametric estimation (table 6) was better able to capture a consensus. Although, mentioned literature touched upon the beneficial importance of pubs for urban life (Cabras, 2011), others were highlighting the risk factors of having high pub densities in local communities. Results of Green et al. (2007) showed that areas with higher concentrations pubs are attached with increasing levels of vandalism, public disorder and noise pollution. Marking those phenomena as negative externalities, ones can argue that there could be a tipping point where the negative externalities overtake the positive ones. As can be observed in table 6, when the pub density is above 2 pubs per km^2 we observe a negative association between the density of pubs and log property prices after controlling for year and neighbourhood specific attributes. When we assume those results to be true, the possibility could be indeed that individuals are willing to pay a premium for having "some" pubs in proximity, but are willing to get a discount in their house price when having an over-representation of pubs within the neighbourhood. After allowing for a less restrictive non-linear model, we see that preferences of individuals are not constantly increasing or decreasing, but might differ for unique values of the concerned explanatory variable. Therefore, the results answer the main research: '*What is the association of pubs with local property prices?'*, to certain extent. We indeed find some non-linear patterns between pub densities and residential property prices, but given the fact that the model is highly explained by unobserved heterogeneity (fixed effects) and extensive literature is not agreeing upon marginal benefits for house prices, we should be careful interpreting the results.

Arbitrating the regression results of the control variables, we can argue that the negative sign of population density on log property price is not following expectations. Day (2018) proposes the rise of property prices in areas with higher population densities, which is not in line with the regression results of both the fixed effects and the semi-parametric model. Subsequently, the coefficient of GDP is in line with theoretical suggestions of Capozza et al. (2002). He hypothesizes and find that GDP has a positive correlation with property prices. Higher income levels should allow people to buy more expensive properties. Even though the coefficient of this model is consistent with Capozza's theory, insignificance of the coefficient makes it unfavourable to draw the conclusion of an agreement. To finish off, Crime rate shows a negative coefficient which is in accordance with Schwartz et al. (2003) who highlight the positive effect of reducing crime rates for local real estate prices. The fixed effects as well as the semi-parametric model show significant results for the effect of crime rate on log property prices. Therefore, we can confirm with our findings the results in the mentioned literature.

7. Conclusion

The analysis that is carried out in this study uses pub density as determinant for property price dynamics in de City of London from 2001 to 2017. Due to extensive literary suggestions about the importance of the pub in Britain's urban life, we hypothesized that the effect of a higher pub density should have in general a positive influence on residential mean property prices in the London's neighbourhoods. This study explored that the use of aggregated data is not able to estimate significant results for the effect of pub density of log property prices, using linear regression estimation like the fixed effects model. After allowing for more flexible relationship between pub densities and log property prices, controlled for year and neighbourhood specific characteristics, the results show significance as well as theoretically plausible results. The results fulfil the hypothesis that *'The effect of pub density is different for several pub density intervals'*. The results show a 7.88% increase in property prices when pub density increases between zero and two pubs/km^2. After those levels of pub densities are passed, the increase in pub density will show negative correlations with property prices.

Critical assessments on this study are the use of aggregated data to explore micro level neighbourhood characteristics. In our model, we assumed that all time and neighbourhood static characteristics are caught in the fixed effects. The high explanatory power of the fixed effects after adding them into the model reveals the importance of locational specific attributes that could not be explained by macro data. The fixed effects model shows that 37.1% of the variation in log property price is explained by year and neighbourhood specific effect, which makes it hard to say whether data is really capable of capturing a causal effect. Even though, the effect of increasing pub densities on mean log property prices might not be causal, this study was able to explore the correlations that were found using semi-parametric regression model. When implicating further research on the effect of pubs in proximity on local real estate prices, micro data including locational characteristic of specific neighbourhood amenities should be in favour to use. Those implementations could solve this study's limitation on the assumption that pubs are equally distributed across the neighbourhood and characteristics of the pub's establishment that are absent. Moreover, it could be interested for further research to explore whether those correlation are similar for cities with different drinking and nightlife cultures.

Nevertheless, this study contributed to the lack of longitudinal research on the effect of pub presence on local property prices. In addition, a general negative association of pub density on property price developments could be an explanation for the pub loss that has developed during this research period. If too overwhelming pub densities are indeed attached to decreasing housing prices, the lowering demand for pubs could have induced the closure of pubs since 2001.

Reference list

Alonso, W. and Rent, T., 1964. *Location and land use*. Cambridge, Mass: Harvard Univ. Press.

Andrews, D. and Turner, S., 2012. Is the pub still the hub?. *International Journal of Contemporary Hospitality Management*, 24(4), pp.542-552.

Apardian, R. and Reid, N., 2020. Going out for a Pint: Exploring the Relationship between Craft Brewery Locations and Neighborhood Walkability. *Papers in Applied Geography*, 6(3), pp.240-255.

Baltagi, B., 2009. *Econometric analysis of panel data*. Chichester: John Wiley & Sons Inc.

Barili, F., Parolari, A., Kappetein, P. and Freemantle, N., 2018. Statistical Primer: heterogeneity, random- or fixed-effects model analyses?†. *Interactive CardioVascular and Thoracic Surgery*, 27(3), pp.317-321.

BBC News. 2021. *London loses 1,200 pubs in 15 years*. [online] Available at: <https://www.bbc.com/news/uk-england-london-39641056> [Accessed 6 May 2021].

Brooks, C., 2019. *Introductory econometrics for finance*. Cambridge, UK: Cambridge University Press.

Brüderl, J. and Ludwig, V., 2019. [online] Ls3.soziologie.uni-muenchen.de. Available at: <https://www.ls3.soziologie.uni-muenchen.de/studium-lehre/archiv/teaching-marterials/panel-analysis_april-2019.pdf> [Accessed 18 May 2020].

Brueckner, J., 1996. *Urban growth models with durable housing: an overview*. [S.l.]: [s.n.].

BUCK, A., HAKIM, S. and SPIEGEL, U., 1991. Casinos, Crime, and Real Estate Values: Do They Relate?. *Journal of Research in Crime and Delinquency*, 28(3), pp.288-303.

Cabras, I. and Lau, C., 2019. The availability of local services and its impact on community cohesion in rural areas: Evidence from the English countryside. *Local Economy: The Journal of the Local Economy Policy Unit*, 34(3), pp.248-270.

Cabras, I., 2011. Industrial and Provident Societies and Village Pubs: Exploring Community Cohesion in Rural Britain. *Environment and Planning A: Economy and Space*, 43(10), pp.2419-2434.

Capozza, D., Hendershott, P., Mack, C. and Mayer, C., 2002. *Determinants of Real House Price Dynamics*. Cambridge, Mass: National Bureau of Economic Research.

Cheshire, P. and Sheppard, S., 1995. On the Price of Land and the Value of Amenities. *Economica*, 62(246), p.247.

Cohen, S., 2004. Social relationships and health. Am Psychol 59: pp.676–684.

Collins, J., 1982. *Drinking and crime*. London: Tavistock.

Daly, J., Gronow, S., Jenkins, D. and Plimmer, F., 2003. Consumer behaviour in the valuation of residential property. *Property Management*, 21(5), pp.295-314.

De Hoyos, R. and Sarafidis, V., 2006. Testing for Cross-Sectional Dependence in Panel-Data Models. *The Stata Journal: Promoting communications on statistics and Stata*, 6(4), pp.482-496.

Ding, W., Zheng, S. and Guo, X., 2010. Value of access to jobs and amenities: Evidence from new residential properties in Beijing. *Tsinghua Science and Technology*, 15(5), pp.595-603.

DiPasquale, D. and Wheaton, W., 1992. The Markets for Real Estate Assets and Space: A Conceptual Framework. *Real Estate Economics*, 20(2), pp.181-198.

Drukker, D., 2003. Testing for Serial Correlation in Linear Panel-data Models. *The Stata Journal: Promoting communications on statistics and Stata*, 3(2), pp.168-177.

Florida, R., 2008. *Cities and the creative class*. New York, NY: Routledge.

Found, A., 2014. *Essays in Municipal Finance: Doctoral dissertation*. [online] Central.bac-lac.gc.ca. Available at: <https://central.bac-lac.gc.ca/.item?id=TC-OTU-65660&op=pdf&app=Library&oclc_number=1032919383> [Accessed 15 June 2021].

Foxcroft, D. and Smith, L., 2009. [online] Drugslibrary.wordpress.stir.ac.uk. Available at: <http://drugslibrary.wordpress.stir.ac.uk/files/2017/04/UK-alcohol-trends-FULL.pdf> [Accessed 8 July 2020].

Freeman, A., Herriges, J. and Kling, C., 2014. *The Measurement of environmental and resource values*. Washington, D.C.: Resources for the Future.

Freeman, A., III, A., 1985. Supply Uncertainty, Option Price, and Option Value. *Land Economics*, 61(2), p.176.

Gagliardi, L. and Schlüter, T., 1981. *The role of education for amenity based sorting in British cities*. [online] Ideas.repec.org. Available at: <https://ideas.repec.org/p/ehl/lserod/65017.html> [Accessed 6 March 2020].

Gibbons, S., 2004. The Costs of Urban Property Crime. *The Economic Journal*, 114(499), pp.F441-F463.

Gillard, Q., 1981. THE EFFECT OF ENVIRONMENTAL AMENITIES ON HOUSE VALUES: THE EXAMPLE OF A VIEW LOT. *The Professional Geographer*, 33(2), pp.216-220.

Glaeser, E. and Gottlieb, J., 2009. The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States. *Journal of Economic Literature*, 47(4), pp.983-1028.

Glaeser, E., Gyourko, J. and Saiz, A., 2008. Housing Supply and Housing Bubbles. *SSRN Electronic Journal*,.

Glaeser, E., Kolko, J. and Saiz, A., 2001. Consumer City. *SSRN Electronic Journal*,.

Goodhart, C. and Hofmann, B., 2008. [online] Ecb.europa.eu. Available at: <https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp888.pdf?ec420f6e04a07efadb15fbbd5d42> [Accessed 6 May 2021].

Green, J. and Plant, M., 2007. Bad bars: A review of risk factors. *Journal of Substance Use*, 12(3), pp.157-189.

Green, W., 2008. [online] People.stern.nyu.edu. Available at: <http://people.stern.nyu.edu/wgreene/StochasticFrontierModels.pdf> [Accessed 6 March 2021].

Hausman, J., 1978. Specification Tests in Econometrics. *Econometrica*, 46(6), p.1251.

Hoechle, D., 2007. Robust Standard Errors for Panel Regressions with Cross-Sectional Dependence. *The Stata Journal: Promoting communications on statistics and Stata*, 7(3), pp.281-312.

John, L., 2019. *Is it a good idea to buy a home near a pub?*. [online] Savills.fr. Available at: <https://www.savills.fr/blog/article/277752/residential-property/is-it-a-good-idea-to-buy-a-home-near-a-pub.aspx> [Accessed 6 May 2021].

Kentish, F., 2016. *Having a local makes you happier*. [online] Metro. Available at: <https://metro.co.uk/2016/01/25/its-official-having-a-local-pub-makes-you-happier-5642775> [Accessed 8 July 2020].

Koster, H., van Ommeren, J. and Rietveld, P., 2013. Is the sky the limit? High-rise buildings and office rents. *Journal of Economic Geography*, 14(1), pp.125-153.

Lang, E. and Rumbold, G., 1997. The effectiveness of community-based interventions to reduce violence in and around licensed premises: a comparison of three Australian models. *Contemporary Drug Problems*, 24(4), pp.805-826.

Li, M. and Brown, H., 1980. Micro-Neighborhood Externalities and Hedonic Housing Prices. *Land Economics*, 56(2), p.125.

Moorman, C., Zaltman, G. and Deshpande, R., 1992. Relationships between Providers and Users of Market Research: The Dynamics of Trust within and between Organizations. *Journal of Marketing Research*, 29(3), p.314.

NELSON, A., 1985. A Unifying View of Greenbelt Influences on Regional Land Values and Implications for Regional Planning Policy. *Growth and Change*, 16(2), pp.43-48.

Norström, T., 2000. Outlet density and criminal violence in Norway, 1960-1995. *Journal of Studies on Alcohol*, 61(6), pp.907-911.

Oldenburg and Brisett, 1982. *269*. [online] Illinois-online.org. Available at: <http://illinois-online.org/krassa/ps410/Readings/Third%20Places/The%20Third%20Place.pdf> [Accessed 6 March 2020].

Oldenburg, R., 2015. [online] Illinois-online.org. Available at: <http://illinois-online.org/krassa/ps410/Readings/Third%20Places/The%20Third%20Place.pdf> [Accessed 23 March 2020].

Ons.gov.uk. 2015. London Boroughs. [online] Available at: https://www.ons.gov.uk/file?uri=/methodology/geography/ukgeographies/administrativegeography/england/londonboroughseng2013map.pdf [Accessed 16 May 2020].

Park, H., 2011. [online] Iuj.ac.jp. Available at: <https://www.iuj.ac.jp/faculty/kucc625/method/panel/panel_iuj.pdf> [Accessed 6 February 2021].

Plant, M., Plant, M. and Thornton, C., 2002. People and Places: some factors in the alcohol-violence link. *Journal of Substance Use*, 7(4), pp.207-213.

Pratten, J., 2007. Reporting on responsible drinking: a study of the major UK pub-owning companies. *Business Ethics: A European Review*, 16(1), pp.62-70.

Preece, D., 2008. Change and continuity in UK public house retailing. *The Service Industries Journal*, 28(8), pp.1107-1124.

Putman, R., 2000. Bowling alone: the collapse and revival of American community. *Choice Reviews Online*, 38(04), pp.38-2454-38-2454.

Richardson, H., Vipond, J. and Furbey, R., 1974. Determinants of Urban House Prices. *Urban Studies*, 11(2), pp.189-199.

Robinson, P., 1988. Root-N-Consistent Semiparametric Regression. *Econometrica*, 56(4), p.931.

Rosen, S., 1974. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1), pp.34-55.

Sandle, D., Graham, K., Graham, K. and Graham, K., 2016. *The secret of happiness? A local pub - Digital Journal*. [online] Digital Journal. Available at: <https://www.digitaljournal.com/life/a-good-pub-close-by-makes-you-happier/article/456287> [Accessed 14 February 2020].

Schwartz, A., Susin, S. and Voicu, I., 2003. Has falling crime driven New York City's real estate boom?. *Journal of Housing Research*, pp.101-135.

Scribner, R., MacKinnon, D. and Dwyer, J., 1995. The risk of assaultive violence and alcohol availability in Los Angeles County. *American Journal of Public Health*, 85(3), pp.335-340.

Shimizu, C., Yasumoto, S., Asami, Y. and Clark, T., 2014. [online] Hermes-ir.lib.hit-u.ac.jp. Available at: <http://hermes-ir.lib.hit-u.ac.jp/hermes/ir/re/26840/wp009.pdf> [Accessed 6 May 2021].

Snowden, C., 2014. *Closing Time: Who's killing the British pub? — Institute of Economic Affairs*. [online] Institute of Economic Affairs. Available at: <https://iea.org.uk/publications/research/closing-time-whos-killing-the-british-pub> [Accessed 3 October 2020].

Rouwendal, J., van Marwijk, R. and Levkovich, O., 2014. The Value of Proximity to Water in Residential Areas. *SSRN Electronic Journal*,.

Tepilo. 2021. *Our 2015 Buyer Barometer - the Complete Report*. [online] Available at: <https://www.tepilo.com/blog/2015/09/our-2015-buyer-barometer> [Accessed 6 April 2020].

Thaler, R., 1978. A note on the value of crime control: Evidence from the property market. *Journal of Urban Economics*, 5(1), pp.137-145.

Tolbert, C., Lyson, T. and Irwin, M., 1998. Local Capitalism, Civic Engagement, and Socioeconomic Well-Being. *Social Forces*, 77(2), p.401.

Torres-Reyna, O., 2007. [online] Princeton.edu. Available at: <https://www.princeton.edu/~otorres/Panel101.pdf> [Accessed 12 March 2020].

Van Oers, J. and Garretsen, H., 1993. The geographic relationship between alcohol use, bars, liquor shops and traffic injuries in Rotterdam. *Journal of Studies on Alcohol*, 54(6), pp.739-744.

Verardi, V., 2021. *Semiparametric regression in STATA*. [online] Stata.com. Available at:<https://www.stata.com/meeting/uk13/abstracts/materials/uk13_verardi.pdf> [Accessed 24 June 2021].

Walford, N., Everitt, J. and Napton, D., 1999. *Reshaping the countryside*. New York: CABI Pub., pp.147-156.

Wiles and Bottoms, 1994. *Environmental criminology*. Oxford: The Oxford Handbook of Criminology, pp.585-658.

Wooldridge, J., 2011. *Student's solutions manual and supplementary materials for Econometric analysis of cross section and panel data, 2nd ed.*. Cambridge (Mass.): The MIT Press.

Yatchew, A., 1997. An elementary estimator of the partial linear model. *Economics Letters*, 57(2), pp.135-143.

Yatchew, A., 1988. *Some tests of nonparametric regression models*. [S.l.]: [s.n.].

Yatchew, A., 2003. *Semiparametric Regression for the Applied Econometrician (Themes in modern econometrics)*. Cambridge University Press.

ZHU, L., 2004. ALCOHOL OUTLET DENSITY AND VIOLENCE: A GEOSPATIAL ANALYSIS. *Alcohol and Alcoholism*, 39(4), pp.369-375.

Appendix 1: Diagnostics

A: Testing for Fixed effects (F-test)

To confirm the need for plausible borough or year fixed effects, an F-test is executed in order to test whether there is a significant effect based on time (year) and space (borough) characteristics. The null hypothesis of the F-test is that all parameters are zero. Therefore, the model tests jointly for all year and borough parameters if they are equal to zero. This test explores whether there are some year or/and borough specific effects present in the data that should be accounted for. Therefore, the hypothesis for the borough fixed effects is $H_0{:}\gamma_1 = \cdots = \gamma_{32} = 0$, whereas for the year fixed effects the hypothesis is $H_0{:}\delta_{2001} = \cdots = \delta_{2017} = 0$. The test results of appendix 2 shows that the test statistics of the year fixed effects as well as the borough fixed effects are jointly different from zero at the 1% level. Therefore, we can confirm the use of firm and time fixed effects in the model.

Table 7. Testing for Fixed effects (F-test), source: STATA (2021)

(1) 2.area = 0	(1) 2002.Year = 0
(2) 3.area = 0	(2) 2003.Year = 0
(3) 4.area = 0	(3) 2004.Year = 0
(4) 5.area = 0	(4) 2005.Year = 0
(5) 6.area = 0	(5) 2006.Year = 0
(6) 7.area = 0	(6) 2007.Year = 0
(7) 8.area = 0	(7) 2008.Year = 0
(8) 9.area = 0	(8) 2009.Year = 0
(9) 10.area = 0	(9) 2010.Year = 0
(10) 11.area = 0	(10) 2011.Year = 0
(11) 12.area = 0	(11) 2012.Year = 0
(12) 13.area = 0	(12) 2013.Year = 0
(13) 14.area = 0	(13) 2014.Year = 0
(14) 15.area = 0	(14) 2015.Year = 0
(15) 16.area = 0	(15) 2016.Year = 0
(16) 17.area = 0	(16) 2017.Year = 0
(17) 18.area = 0	F(16, 492) = 242.31
(18) 19.area = 0	Prob > F = 0.0000
(19) 20.area = 0	
(20) 21.area = 0	
(21) 22.area = 0	
(22) 23.area = 0	
(23) 24.area = 0	
(24) 25.area = 0	
(25) 26.area = 0	
(26) 27.area = 0	
(27) 28.area = 0	
(28) 29.area = 0	
(29) 30.area = 0	
(30) 31.area = 0	
(31) 32.area = 0	
F(31, 492) = 193.10	
Prob > F = 0.0000	

B. Hausman model specification test

The choice whether to use a Fixed effects or Random effects model is tested by the Hausman specification test (1978). The model tests whether the parameters are correlated with the entity specific errors. Thus, the H_0 of the Hausman specification test is that $E(x_{it},u_i)$ is zero. When the H_0 is accepted the estimated beta is consistent as well as efficient for the random effects estimator. Therefore the random effects estimator will be appropriate. When $E(x_{it},u_i)$ is not zero, the beta is still consistent but inefficient under H_0. This means that H_0 (read random effects estimator) will be inefficient and thus the fixed effects estimator should be used. If $E(x_{it},u_i)$ is not zero, the betas of the random effects model are inconsistent while the betas of the fixed effects model are. The computation of the hausman test is executed as follows:

$H_0 : \hat{\beta}_{RE} = \hat{\beta}_{FE}$

$H = (\hat{\beta}_{RE} - \hat{\beta}_{FE})'[\hat{V}(\hat{\beta}_{RE}) - \hat{V}(\hat{\beta}_{FE})]^{-1}(\hat{\beta}_{RE} - \hat{\beta}_{FE})$

$\hat{\beta}$ = estimated beta

$\hat{V}$ = estimated variance

There can be observed that when the difference between the estimated betas of both estimators is very large, then there is a large probability of rejecting H_0. A large difference in estimated beta means there is a bias in the estimated beta and therefore, the test-statistic will be very large. When determining which model is appropriate for the sake of this study, the hausman test is executed for the following model:

(10) $\text{LogPropertyprice}_{it} = \beta_0 + \sum_{i=2}^{32} \beta_{0,i} + \sum_{t=2}^{17} \beta_{0,t} + \beta_1 \text{Pubdensity}_{it} + \beta_2 GDP_{it} + \beta_3 Crimerate_{it} + \beta_4 Populationdensity_{it} + e_{it}$

i = 1,…32 (32 different boroughs)

t = 1,…17 (2001-2017)

Table 8. Hausman specification test (1978) for Fixed vs. Random effects model, source: STATA (2021)

	Coef.
Chi-square test value	20.381
P-value	0.000

Regarding a p-value of 0.000 (table 7), the test rejects H0 at 1% level. This indicates that the fixed effects model stated in equation 4 is consistent and efficient with the used data.

Since the fact that we strive for an estimator that is efficient, consistent and unbiased, some of the fixed effects assumptions need to be tested, before results can be presented. Table 3 summarizes all the assumption for the Fixed effects model, whereas in this chapter will be discussed what the assumptions contain and how can be dealt with this. The execution of the assumption tests can be found in Appendix 1.

C. Strict Exogeneity

The first assumption of the Fixed effects model is *strict exogeneity assumption*. This means that the parameters are uncorrelated with the idiosyncratic error term in each time and space dimension. To technical notate: $E(e_{it}|x_{it}) = 0$ for all i and t. This means that the error term should be spatial and serial uncorrelated. However, because the fixed effects are not demeaned in the equation $E(a_i|x_{it}) = 0$ is no longer required to be a unbiased estimator. The fixed effects estimator gives consistent estimates in all situations when we suspect that individual specific unobserved heterogeneity is correlated with the observed variables.

D. Multicollinearity Test

If explanatory variables are independent of each other, adding or removing a variable from the model would not lead to a change in the value of the coefficient of another variable. (Brooks, 2019). In practice, there will always be a correlation between different explanatory variables, but a small degree of correlation will not cause inefficiency of the model. When explanatory variables proves to be highly correlated, a problem occurred and is called *Multicollinearity.* When two or more variables are highly correlated with one and another, *perfect multicollinearity* occurs. In this situation, the model will be inefficient and not consistent. Highly correlated variables will intend that the "true" value of the coefficients are not warranted. To check whether multicollinearity is a problem, a correlation matrix or the variance inflation factor (VIF) is used. The rule of thumbs regarding the VIF is that a variable is not correlated with one and another variable when VIF<10. Table 9 gives the result of the test of multicollinearity on the independent variables. After omitting *Job density* from the equation, we see that there is no reason to assume that multicollinearity is a problem. In the first correlation matrix. Job Density is heavily correlated with the other regressor. Therefore, the results can be misleading due to less reliable statistical inferences. After removing job density from the equation, we see all variables having considerably lower VIF<10 (rule of thumb)

Table 9. Correlation matrix, source: STATA (2020)

	LogProperty price	Pub density	Crime rate	GDP	Population density	Job density	VIF
Log Property price	1						
Pub density	0.441***	1					5.719
Crime rate	0.0127	0.820***	1				3.578
GDP	0.487***	0.795***	0.614***	1			16.58
Population density	-0.0107	-0.306***	-0.257***	-0.142***	1		1.151
Job density	0.455***	0.850***	0.715***	0.962***	-0.203***	1	21.361

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	LogPropertyprice	Pubdensity	Crimerate	GDP	Population density	VIF
Log Propertyprice	1					
Pub density	0.441***	1				5.578
Crimerate	0.0127	0.820***	1			3.096
GDP	0.487***	0.795***	0.614***	1		2.842
Population density	-0.0107	-0.306***	-0.257***	-0.142***	1	1.138

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

E. Heteroscedasticity Test

If the errors of a model do not have a constant variance, the distribution of the residuals is called to be "Heteroscedastic". When this occurs the model violates the assumption of being Homoscedastic in the error term. Heteroscedasticity causes incorrectness of the standard errors. Cross-sectional studies often have significantly different value among the cross-sections. Therefore, they are more likely to have the presence of heteroscedasticity. In cross-sectional time series data (panel data), most inconstant variances in the error term are due to differences specific cross sectional units. The consequences of having heteroscedastic errors will induce the beta's to be inefficient and leads to biased estimation of the model (Brooks, 2019). Solutions for dealing with heteroscedasticity are: (1) Transforming variables by rescaling them by using logarithms or some measure of size. This minimalizes the dispersion of extreme values. (2) Using *heteroscedasticity-consistent standard errors* accounts for situations where the variance of the errors is varying over time. Econometric data software like STATA have the option 'robust' to implement this. The modified Wald statistic test for groupwise heteroscedasticity calculates the residuals of a fixed effects model in a regression model (Green, 2009, p. 598). Here, the null hypothesis is that the model is homoscedastic. Appendix 1 executes the modified Wald statistic test for groupwise heteroscedasticity. The test rejects the null hypothesis that the residuals of the fixed effects model are homoscedastic at 1% level. Therefore, robust standard errors should be used to correct for this.

Table 10. Modified Wald test for groupwise heteroscedasticity in fixed effect regression model, source: STATA (2020)

H0: sigma(i)^2	= sigma^2 for all i
Chi2 (32)	= 2748.96
Prob>chi2	= 0.000

H0: The null hypothesis is homoscedasticity. Above we reject the null and conclude heteroscedasticity. To fix this, a robust standard error is used in the regressions to solve this

F. Serial(Auto) and Cross-sectional autocorrelation

Serial correlation in panel data is one of the concerns when estimating a panel data model. Serial correlation implicates the intertemporal dependency of the error term from one period to another. Autocorrelation biases the standard errors and the model will turn out to less efficient. Drukker (2003) implicates the use of Wooldridge's test for autocorrelation for autocorrelation in panel data. Wooldridge (2002) uses residuals from a first-differenced regression and tests their lags. The null hypothesis of the test assumes no first order autocorrelation. Due to the fact that the panel of this study consists of 17 years, there is no reason to assume that there is serial correlation present within the panels. Baltagi (2009) argues that serial correlation is only a problem in macro panels with 20 to 30 years of data.

Furthermore Cross-sectional correlation is another issue that may occur in panel data analysis. Hoyos & Sarafidis (2006: 482) states that: *"A growing body of the panel-data literature concludes that panel-data models are likely to exhibit substantial cross-sectional dependence in the errors, which may arise because of the presence of common shocks and unobserved components that ultimately become part of the error term, spatial dependence, and idiosyncratic pairwise dependence in the disturbances with no particular pattern of common components or spatial dependence".* Reasons for cross-sectional dependency are rooted in the interwoven economic and financial systems between different cross-sections. To test for spatial dependence in a regression, Pesaran's test (2004) for cross-sectional dependency is used. The null hypothesis of the test is that the residuals are uncorrelated across different cross-sections. Appendix 1 shows the test results of Pesaran's test for cross-sectional dependency. The null hypothesis of the test is that there is cross-sectional dependency in the error term of the fixed effects model. The test statistic cannot reject the null hypothesis. Thus, cross-sectional dependency is present in the data. To correct for cross-sectional correlation, Driscoll and Kraay standard errors are used to account for cross-sectional dependent errors. Hoegle (2007) pleads for the use of Driscoll and Kraay standard errors when cross-sectional dependency is present, since those errors are better calibrated in such situation. When there is no cross-sectional dependency, robust standard errors will perform better. Hoegle (2007), furthermore states that the implementation of Driscoll and Kraay standard errors are robust to disturbances being heteroscedastic, contemporaneously cross-sectionally correlated errors and first order auto-correlated errors.

1. **Cross sectional dependence: Pasaran test for cross sectional dependence**

- Pesaran's test of cross sectional independence = -2.350, Pr = 0.0188
  Average absolute value of the off-diagonal elements = 0.501
- H0: The null hypothesis is that there is cross sectional dependence. Above we reject the null and conclude cross sectional dependence. To fix this, Driscoll and Kraay standard errors are used in the model

2. **Serial autocorrelation**

- Wooldridge test for autocorrelation in panel data
  F(1, 31) = 424.869
  Prob > F = 0.0000
- H0: The null hypothesis is that there is no first-order autocorrelation. Above we reject the null and conclude first-order autocorrelation. Driscoll and Kraay standard errors are used in the model to control for this.

Appendix 2: statistics

Descriptive statistics Inner and Outer London

Table 11. Descriptive Statistics Inner London

Variable	Obs	Mean	Std.Dev.	Min	Max
Property price	221	379000	206000	103000	1280000
Pub density	221	7.887	5.262	1.167	23.377
Crime rate	221	148.743	62.516	65.831	424.368
GDP	221	56169.64	55044.51	10383	283000
Population/km2	221	10271.57	2080.988	6465.828	15842.73
Job density	221	9169.823	7433.135	1918.405	34725.28

Table 12. Descriptive Statistics Outer London

Variable	Obs	Mean	Std.Dev.	Min	Max
Property price	323	261000	90014.82	84625	632000
Pub density	323	1.484	.552	.524	2.776
Crime rate	323	88.705	23.2	45.01	151.424
GDP	323	25046.67	9490.827	12077	71294
Population/km2	323	4290.965	1326.532	1963.327	7696.732
Job density	323	1764.494	579.2787	698.9508	3631.516

Appendix 3: Chow F-test on heterogeneity effect of GDP

A. Poolabiliy-regression

Table 13. Regression results comparing the coefficients between low and high GDP groups

	Chow F-test Pooled OLS
GDPhighlow	0.723***
	(0.1113)
Pubdensity	0.0914***
	(0.0091)
Crimerate	-0.00413***
	(0.0005)
Populationdensity	0.00000147***
	(2.17e-07)
GDPhighlow*Pubdensity	-0.0511***
	(0.1032)
GDPhighlow*Crimerate	0.00341***
	(0.0007)
GDPhighlow*Populationdensity	-0.00000358***
	(3.05e-07)
Constant	11.70***
	(0.0817)
Year Fixed Effects	Yes
N	544
R^2	0.860
adj. R^2	0.849

Notes: Dependent variable is the logarithm of Property price. GDPhighlow is a dummy variable, where 1=high income and 0 =low income. Standard Errors given within parentheses: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

B. STATA Computation of Poolability

Table 14. Contrasts of marginal linear predictions as a result of table 13

	df	F	P>F
GDPhighlow	1	78.49	0.000
GDPhighlow*Pubdensity	1	24.47	0.000
GDPhighlow*Crimerate	1	21.38	0.000
GDPhighlow*Population	1	137.54	0.000
GDPhighlow*Year	16	3.66	0.000
Overall	20	9.94	0.000
Denominator	504		

H0: There is no breaking point in the data. Pooled regression is appropriate.

C. Manual Computation of Poolability

$$F(20,504) = \frac{RSS_{total} - (RSS_{lowGDP} + RSS_{highGDP})}{k} / \frac{RSS_{lowGDP} + RSS_{highGDP}}{N_{lowGDP} + N_{highGDP} - 2*k} = \frac{19.2001251 - (3.27866803 + 10.4910762\ )}{20} / \frac{(3.27866803 + 10.4910762\ )}{272 + 272 - 2*20} = 9.94$$

$$F(20,504)_{crit\ at\ 5\%} = 1.59143456$$

$$F_{value} > F_{crit} = 9.94 > 1.59143456 = seperate\ regression\ is\ appropriate$$

Appendix 4: Robustness semi-parametric model

Figure 4. Semi-parametric regression of log Property price on Pub density.

Bandwidth = 0.05

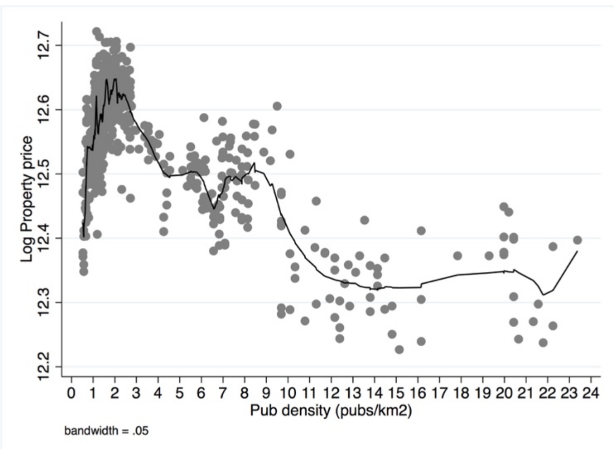


Figure 5. Semi-parametric regression of log Property price on Pub density.

Bandwidth = 0.10

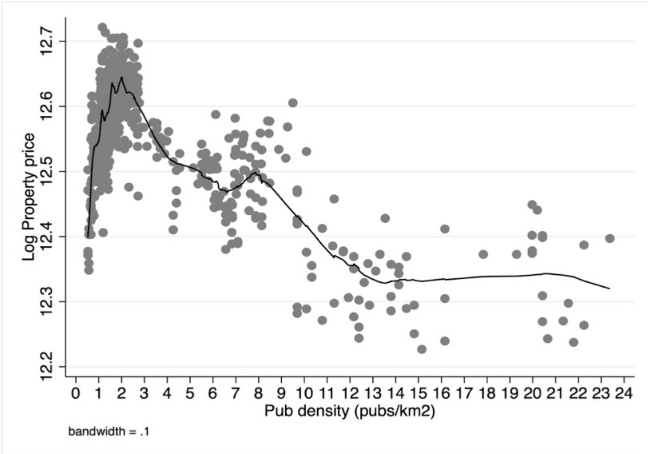


Figure 6. Semi-parametric regression of log Property price on Pub density.

Bandwidth = 0.40

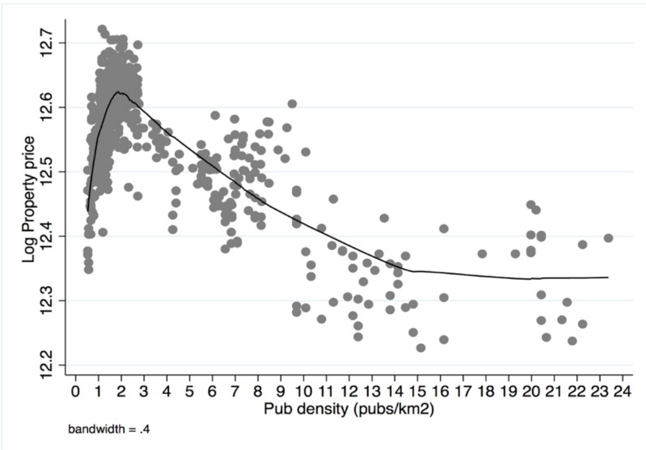


Figure 7. Semi-parametric regression of log Property price on Pub density.

Bandwidth = 0.80