Analyzing the Effects of Subway Proximity on Real Estate Prices in London *

Aman Rana
1007034692
aman.rana@mail.utoronto.ca

Pierre Sarrailh
1006770111
pierre.sarrailh@mail.utoronto.ca

ECO375: Applied Econometrics
University of Toronto
Department of Economics
5th December, 2024

Abstract

Zhou et al. (2019) shows us that the completion of subway lines increases the prices of real estate in their vicinity. Our contribution will be to look at these effects in London. We find precise null effects by a multivariate linear regression of the log of house prices on distance to the nearest station in metres, with controls for time and district. We find a slope coefficient of -0.0001261 with a standard error of 0.00000145 on the subway distance variable. The effect is economically insignificant, even a 10km distance from a station (which is above the 95th On more relative terms, a standard deviation difference in distance is only associated with a percentile of our distance data) would lead to only a 0.2% decrease in house price. This statistically significant relationship of small effect leads us to conclude precise null effects. 14% of a standard deviation change in log house price. While the marginal effect of distance is precisely null, we find a threshold effect: the price of houses within 2km of a subway station is 0.18% higher than those further away, with economic and statistical significance after controlling for district and time fixed effects.

*Code and data supporting this analysis is available at: `https://github.com/Aman-Rana-02/Subway_Proximity_and_House_Prices`

1 Introduction

Accessibility to public transit has been long established as a key factor influencing the price of residential properties, particularly in dense urban environments where vehicle mobility is greatly constrained (Rietveld et al., 2007). In the context of the Greater London Area subway stations are the most efficient method of transportation for most commuters and thereby play a critical role in shaping the greater housing market. To this end in this paper we investigate how the euclidean distance from a residence to its nearest subway station affects market price.

By focusing on the Greater London Area our study contributes to already established relationships between subway proximity and house prices. Early studies such as those done by Muth (1969) are cornerstones in establishing a relationship between densification and land use in cities. More recently a greater number of studies have found a significant positive relationship between transit proximity and real estate prices in cities such as those in Shanghai (Zhou et al., 2019) where subway stations proximity was specifically correlated to increased house prices at statistically significant results. However London presents unique opportunities for research. As the historical pioneer in subway technology the city has had the opportunity to develop around an already established subway system and culture.

In our analysis we investigate two effects, the premium associated with being marginally closer to a subway station, as well as the threshold effect of being within a certain distance of a subway station. We find precise null effects for the marginal distance on house price. Rietveld et al. (2007) finds that the price effect of subway stations on commercial buildings is present at short distances, we find the same results for London house prices, setting a 2km threshold.

The rest of the paper is structured as follows: Section 2 describes the data used in this analysis. Section 3 outlines the methodology and model specifications used to measure the subway effect, and presents the results of our analysis. Section 4 discusses the limitations of our methodology and concludes. Appendix A holds figures and tables.

2 Data

We began our analysis by sourcing data from ONS, the national statistics arm of the UK government, which provides panel data containing each lease transfers (home sale) that occurred in the U.K. from 1995 to today (HM Land Registry, 2014). We prune this dataset to only contain observations within the Greater London Area. This dataset amounts to approximately

350 thousand observations. Each unit of observation represents a single lease transfer in London containing the date, coordinates, neighborhood, and house price in British Pounds during the transfer. We also sourced the coordinates of every subway station in the London Underground and the date at which it was opened (Wikipedia.org, 2014). We combine the subway and housing datasets, finding the closest subway station to each house at the time of the lease transfer using euclidean distances based on latitude and longitude. Delving into the summary statistics of our data, we can look at how the data is spread across our different variables. For our main variable of interest, the distance to subway stations, we see an average distance of 2821 meters with a variance of 2915 meters (Table A.2). On average houses are located quite far from a subway station but also contain vast amounts of variation. Houses are clustered around subway stations therefore we end up with long tails, a few houses that are very far away bringing kurtosis to our data. We truncate these outliers at 10km, and the distribution of the minimum distance to a subway station can be seen in Figure 2. Raw house price is also right skewed with a long tail, so we take the log of house prices to make our regression more interpretable and the distribution of log price can be seen in Figure 3. We see that our houses are very evenly spread between our neighborhood controls with the largest neighborhood only being 2 points off of the mean (Table A.2). Similarly, our lease transfers are also evenly spread between years with each year accounting for approximately 2% of our dataset (Table A.2). This is good for the robustness of our future models as every neighbourhood and year is well represented. On the other hand the spread between newly built and old homes is skewed with 97% of lease transfers coming from the sale of pre-owned homes(Table A.2). As a result it may be more sensitive to outliers.

3 Regression Analysis

For our regression analysis we build four different specifications to best capture the relationship between the distance to a subway station and the log of price. We choose a log-linear model so that our coefficients are interpretable as percentage changes in price. We inspect our variable of interest in two forms, continuous so we can see the marginal effect of each additional meter of distance on price, and as a dummy variable to see the effect of being subjectively close to a subway station. We look at Figure 1 to see the relationship between distance and price, and decide on 2000m as distinguishing being close to a station or far. For our analysis, we assume that our data is identical and independently distributed since we do not expect the lease transfer of one house to affect the lease transfer of another. In our data cleaning section we windsorize distance at 10km and can assume our data contains no large outliers. We use robust standard errors in our regressions, and therefore aren't concerned about the homoscedasticity of errors.

3.1 Univariate Linear Regressions

3.1.1 Univariate Specifications

$$\log(\text{Price}) = \beta_0 + \beta_1 \cdot \text{Min\_dist} + \epsilon \tag{1}$$

$$\log(\text{Price}) = \beta_0 + \beta_2 \cdot \text{close} + \epsilon \tag{2}$$

Where:

- $Log(Price)$ is the log of the price of the house
- β_0 is the intercept and not shared between the two specifications
- β_1 is the coefficient of the distance to the nearest subway station
- β_2 is the coefficient of the distance dummy, whether the house is close (within 2000m) to a subway station or not
- Min_dist is the distance to the nearest subway station
- ϵ is the error term

3.1.2 Univariate Results

We start with our simplest specifications. We regress the log of price on distance to get a basic understanding of the relationship between our variable of interest and our dependent variable. Using this specification we find statistically significant null effects for distance on price. Table A.1 shows us that the coefficient of our variable of interest is -.0000431 at above the 5% significance. The coefficient can be interpreted as a 0.0000431% decrease in house price for each additional meter of distance from a subway station, which is economically insignificant. Even houses positioned 10 km away from a subway station wouldn't even observe a difference above 1% in house price from the mean. Looking at the results from specification 2 we see that the coefficient is economically and statistically significant, with a coefficient representing a 0.31% increase in house price for houses within 2000m of a subway station (Table A.1). What this tells us is that the marginal effect of being a metre closer to a subway station is not significant. But, buyers' willingness to pay follows a piecewise function, and have some threshold of distance over which they are willing to pay more for a house. Essentially, above a certain threshold the distance from a subway station is not a consideration. In this case, whether you are within 2km of a subway station or not is a significant factor in determining house price, and adds a 0.33% premium.

However, these models are underspecified. They suffer from omitted variable bias, there are variables that comove with distance from a subway station that also affect house prices, and we delve into these controls in our multivariate analysis. We would expect the economic significance of distnace to increase with the inclusion of these controls. Since subway stations are typically central, we would expect neighbourhoods further away to have lower house prices. Omitting district controls would bias our coefficient to be more positive than it should be. We would also expect house prices to increase over time, whereas they decrease with min_dist. Omitting time controls would bias our coefficient to be more positive than it should be.

3.2 Multivariate Linear Regressions

3.2.1 Multivariate Specifications

$$\begin{aligned}\log(\text{Price}) = \beta_0 &+ \beta_1 \cdot \text{Min\_dist} + \beta_3 \cdot (\text{Min\_dist})^2 \\ &+ \beta_4 \cdot \text{oldnew} + \beta_5 \cdot \text{district} + \beta_6 \cdot \text{year} + \beta_7 \cdot (\text{district} \times \text{year}) + \epsilon\end{aligned} \tag{3}$$

$$\begin{aligned}\log(\text{Price}) = \beta_0 &+ \beta_2 \cdot \text{close} \\ &+ \beta_4 \cdot \text{oldnew} + \beta_5 \cdot \text{district} + \beta_6 \cdot \text{year} + \beta_7 \cdot (\text{district} \times \text{year}) + \epsilon\end{aligned} \tag{4}$$

Where:

- β_3 is the coefficient of the squared distance to the nearest subway station
- β_4 is the coefficient on the age of the house
- β_5 is a matrix of coefficients for district dummies
- β_6 is the coefficient of the year
- β_7 is a matrix of coefficients for the interaction of district and year dummies
- oldnew is a dummy variable for the age of the house
- district represents dummy variables for the district of the house
- year represents dummy variables for the year of the observation

We include a squared term for distance in our multivariate regression to account for the non-linear relationship between distance and price we find in Figure1.

We now include controls for the age of the house, the district of the house, and the year of the observation. The economic basis for this, is we expect there to be price variation between older and newer homes because of the quality of the home. We also expect price variation between districts as some districts hold premiums due to proximity to a city center or other amenities. We also expect price variation over time as the economy grows and inflation occurs. Consider that some districts receive more or less government funding and expansion. Since we might expect the time varying effect to be different across districts, we include an interaction term between district and year.

3.2.2 Multivariate Results

Table A.1 hides the coefficients of the controls, but a full table of the results can be found in the appendix under Table A.3 and Table A.3.

In our multivariate regression with the continuous polynomial for distance (Specification 3), we find that the coefficient is still economically insignificant (-0.0001261). As expected, controlling for district and time-varying effects increases the magnitude of our result, but we still conclude a precise null marginal effect of distance on house price.

In our final specification (Specification 4), we find that the coefficient of the dummy variable for being close to a subway station is still economically and statistically significant. However, much of the variation is now explained by the controls, and we find a more reasonable 0.18% increase attributable to being close (Within 2km) to a subway station. This aligns with the results of Rietveld et al. (2007), which finds that the price effect of subway stations is primarily in short distances.

Notably, our R^2 jumps from 0.03 in our univariate regression to 0.6 in our multivariate regression(Table A.3 and Table A.3). This is a good sign that our multivariate specifications which include controls are explaining much of the variation in logarithmic house prices.

Looking at the predicted versus actual plots in Figure 5 and Figure 6, we see that our models are well-specified, apart from at the lower tail. Most predicted values lie on the 45 degree line, however, we are systematically underpredicting the prices of the lowest priced homes. This is further reflected in the residual fits as seen in Figure 4 and Figure 7. We see that our residuals are randomly distributed around 0 apart from at the lower tail, where they are systematically positive.

4 Discussion

4.1 Limitations of Results

The analysis of our limitations can be broken into internal and external validity.

4.1.1 Internal Validity

Our regression is misspecified and has an omitted variable bias. One potential omitted variable could be house size. We would expect house size to explain some of the variation in house price since larger houses command higher prices. However without further analysis it would be difficult to estimate the effect of house size since larger houses are in less dense areas of the city where property is also generally cheaper. We also run into the problem of simultaneous causality, areas that become trendy and develop would command higher housing prices, which could justify a new subway station nearby. We attempt to control for this using time-invariant district fixed effects but that does not solve the time-variant relationship where growing neighborhoods get subway allocations. In a future work we would frame a subway being built as a treatment, and attempt a difference-in-difference analysis, so we can gauge economic significance and price effects of subways without having to consider some of the other omitted variables. In our analysis we also assumed that our data was i.i.d. However this may not be true as we saw in the 2008 financial crisis when many people sold houses at the same time, leading to a depreciation in house price, causing panic and further selling. This means that there are regimes where selling a house has a causal effect on another person selling a house.

4.1.2 External Validity

The external validity of our regression is limited. The population studied are homes in the Greater London Area, with observations from 1995 to 2024. We can use our model to make inferences on homes that fit within this sample space, however, expect pricing dynamics to be different between cities and regimes. For example, in smaller cities we might expect price variation to be independent of subway locations if there is sufficient coverage.

4.2 Conclusion

Houses closer to subway stations in the Greater London Area see higher prices than those further away. This analysis includes controls for time, district, and a polynomial fit for the distance and price relationship. The relationship is likely piece wise, beyond 2km away from a station the distance from a station has little effect on price. We would warn against assuming external validity, and exogenous factors like governance and housing density could affect inferences. Marginal distance from subway stations are statistically significant but economically insignificant in their relationship with house price. Being 'close' to a subway station however, is statistically and economically significant, even when controlling for district, building age, and time effects. We are wary about causality given that our residuals have a non-zero mean. Given the distribution of our residuals vs fitted the specification is missing a factor that explains the variation is lower priced home. This may be size, which we've discussed, or some other factor.

References

HM Land Registry (2014, jul). Price Paid Data. `https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads`. Accessed: 2024-09-15.

Muth, R. F. (1969). *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*. Chicago and London: The University of Chicago Press.

Rietveld, P., G. Debrezion, and E. Pels (2007, 02). The impact of railway stations on residential and commercial property value: A meta-analysis. *The Journal of Real Estate Finance and Economics 35*, 161–180.

Wikipedia.org (2014). List of London Underground stations. `https://en.wikipedia.org/wiki/List_of_London_Underground_stations`. Accessed: 2024-09-15.

Zhou, Z., H. Chen, L. Han, and A. Zhang (2019, Feb). The effect of a subway on house prices: Evidence from shanghai. *Real Estate Economics 49*(S1), 199–234.

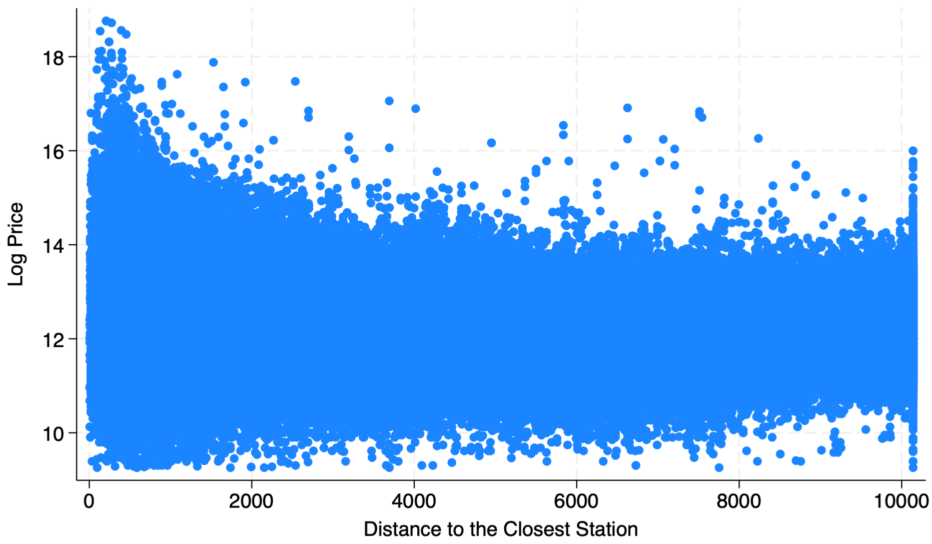

Figure 1: Scatter plot of log price against distance to the nearest subway station. Notice the non-linear relationship, and piece wise nature beyond 2km.

Appendix

A Appendix

A.1 Regression Summary and Figures

Coefficient	**1**	**2**	**3**	**4**
Intercept	12.52	12.22	14.57	14.57
	(0.00208)	(0.00202)	(1.826)	(1.826)
min_dist	-0.0000441		-0.000126	
	(0.000000470)		(0.00000145)	
close		0.314		0.189
		(0.00283)		(0.00290)
min_dist^2			1.09e-08	
			(1.35e-10)	
Controls	No	No	Yes	Yes

Table 1: Regression Coefficients and Standard Errors for Model Specifications 1-4. Controls hold district, year, house age, and year-district interactions.

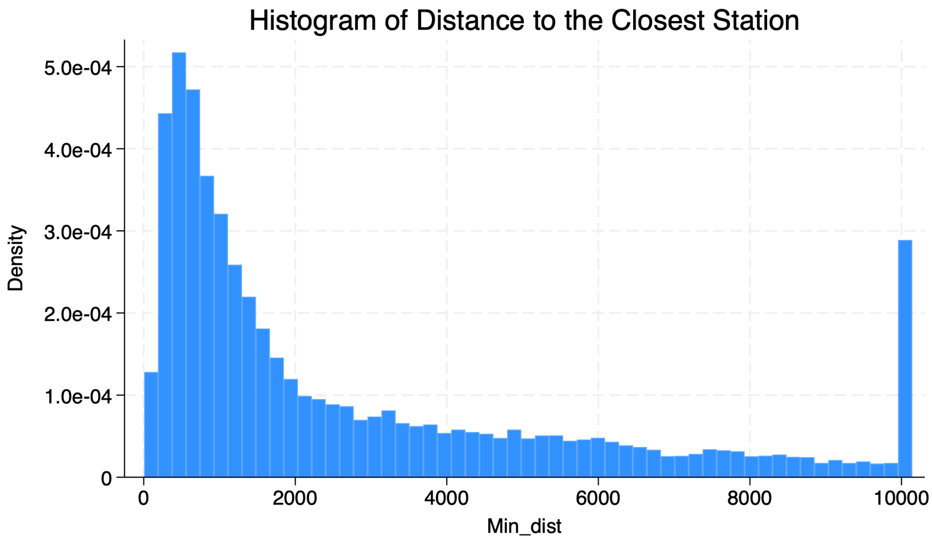

Figure 2: Distribution of the minimum distance to a subway station. The raw distribution is right skewed with a long tail so we windsorize at 10km.

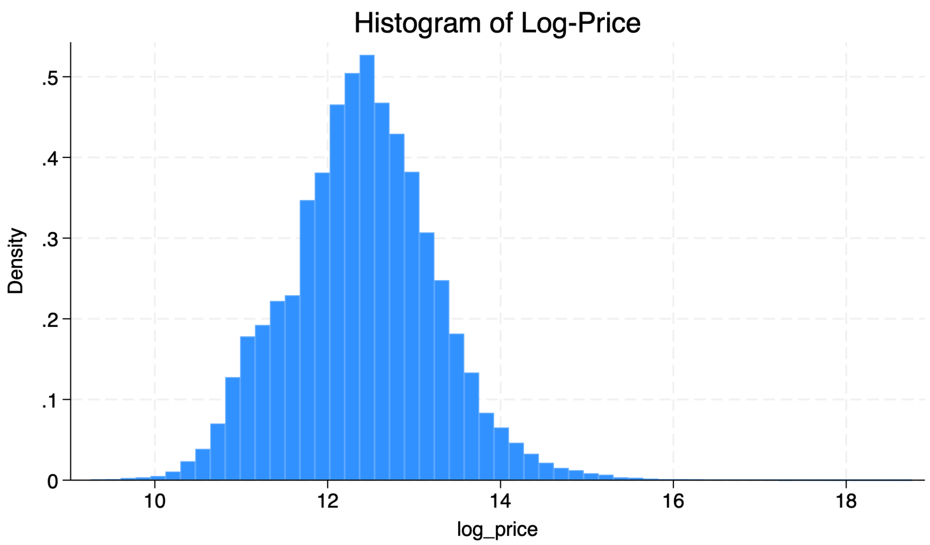

Figure 3: Distribution of log price. The log transformation accounts for the long tails in house price's distribution, and makes our regression nicely interpretable.

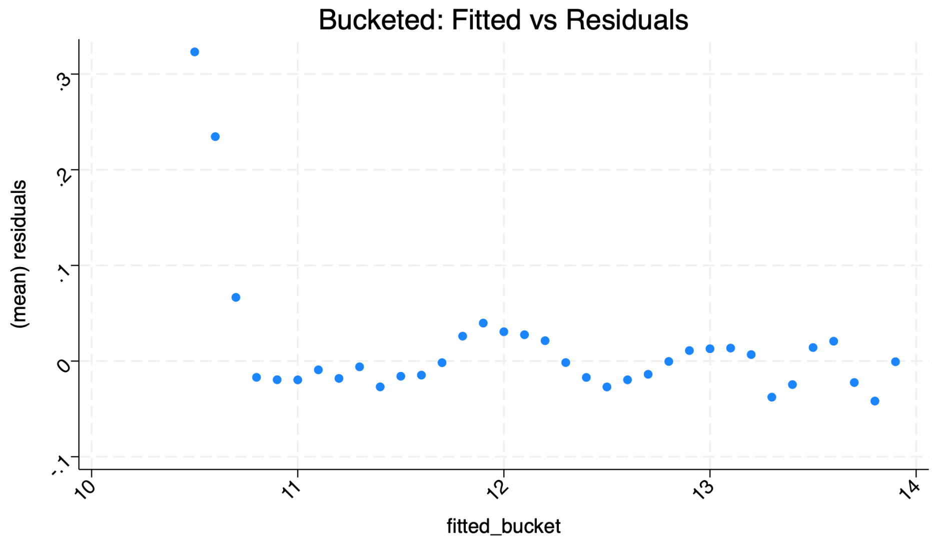

Figure 4: Fitted vs Residuals plot for the multivariate regression. The residuals are distributed randomly around 0 apart from at the lower tail where they break down. Results are after bucketing predicteds into bins for easier interpretation. Results are from the regression in Specification 3.

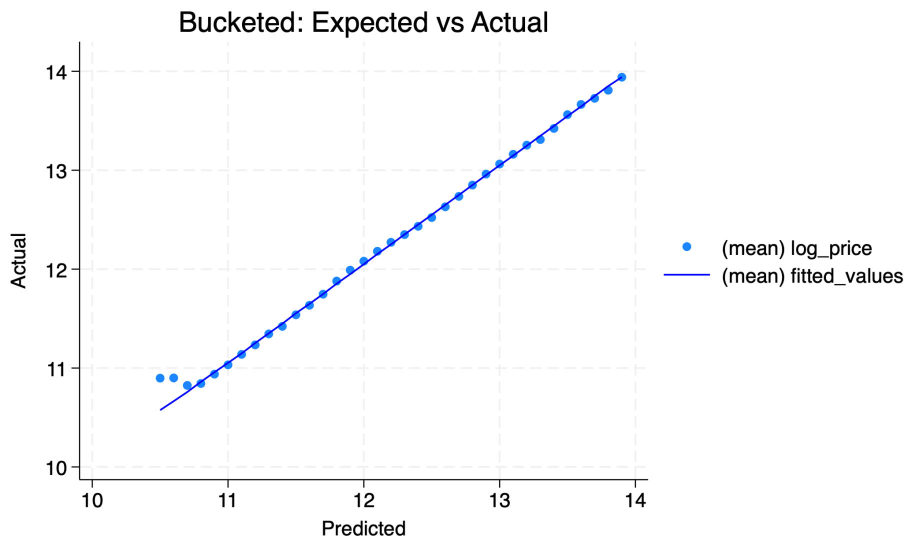

Figure 5: Bucketed Expected vs Actual plot for the multivariate regression. The expected values are bucketed into 10 bins and the actual values are averaged within each bin we do this becuase with over 340000 observations the plots otherwise are difficult to interpret. Results are from the regression in Specification 3.

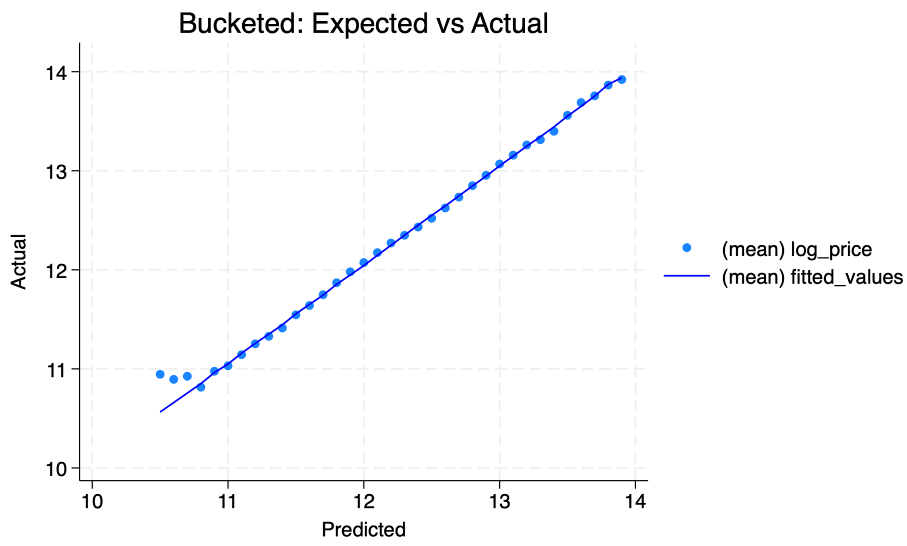

Figure 6: Bucketed Expected vs Actual plot for the multivariate regression with dummy. The expected values are bucketed into 10 bins and the actual values are averaged within each bin we do this because with over 340000 observations the plots otherwise are difficult to interpret. Results are from the regression in Specification 4.

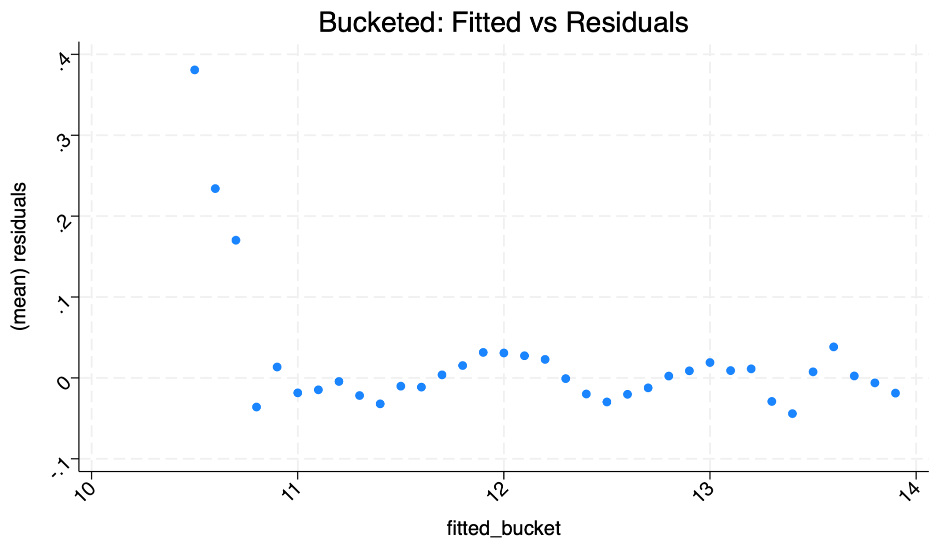

Figure 7: Fitted vs Residuals plot for the multivariate regression with dummy. The residuals are distributed randomly around 0 apart from at the lower tail where they break down. Results are after bucketing predicteds into bins for easier interpretation. Results are from the regression in Specification 4.

A.2 Data Summary Statistics

Table 2: Numeric Variables Summary Statistics

	(1)				
	mean	sd	Var	min	max
Min_dist	2821.537	2915.578	8500594	5.747817	10141.07
min_dist_squared	1.65e+07	2.83e+07	8.00e+14	33.0374	1.03e+08
log_price	12.3983	.8569691	.734396	9.25913	18.76283
Observations	347373				

Table 3: Old/New Summary Statistics

	(1)	
	Old/New	
	b	pct
N	334362	96.25446
Y	13011	3.745542
Total	347373	100
Observations	347373	

Table 4: District Summary Statistics

	(1)	
	District	
	b	pct
BARKING AND DAGENHAM	6773	1.949777
BARNET	13020	3.748132
BEXLEY	11626	3.346835

BRENT	8631	2.484649
BROMLEY	17304	4.981389
CAMDEN	8276	2.382453
CITY OF LONDON	434	.1249377
CITY OF WESTMINSTER	9881	2.844493
CROYDON	17283	4.975344
EALING	11824	3.403834
ENFIELD	13130	3.779799
GREENWICH	10877	3.131216
HACKNEY	7398	2.129699
HAMMERSMITH AND FULHAM	7949	2.288318
HARINGEY	9809	2.823766
HARROW	7131	2.052837
HAVERING	11603	3.340214
HILLINGDON	12621	3.63327
HOUNSLOW	9267	2.667738
ISLINGTON	7174	2.065215
KENSINGTON AND CHELSEA	8030	2.311636
KINGSTON UPON THAMES	8255	2.376408
LAMBETH	14930	4.297974
LEWISHAM	13391	3.854934
MERTON	10079	2.901492
NEWHAM	9985	2.874432
REDBRIDGE	10658	3.068172
RICHMOND UPON THAMES	11298	3.252412
SOUTHWARK	10059	2.895735
SUTTON	10342	2.977203
TOWER HAMLETS	6635	1.910051
WALTHAM FOREST	12275	3.533666
WANDSWORTH	19425	5.591972

Total	347373	100
Observations	347373	

Table 5: Year Summary Statistics

	(1)	
	Year	
	b	pct
1995	11273	3.245215
1996	13798	3.972099
1997	15791	4.545834
1998	14966	4.308337
1999	17317	4.985131
2000	15518	4.467244
2001	17110	4.925541
2002	18197	5.238461
2003	15726	4.527122
2004	16029	4.614348
2005	13882	3.996281
2006	17206	4.953177
2007	16474	4.742453
2008	7402	2.130851
2009	6962	2.004186
2010	8703	2.505376
2011	8340	2.400877
2012	8537	2.457589
2013	10055	2.894583
2014	11275	3.245791
2015	10621	3.05752
2016	9442	2.718116

2017	8588	2.47227
2018	8278	2.383029
2019	7836	2.255788
2020	7420	2.136032
2021	11412	3.285229
2022	9337	2.687889
2023	6825	1.964747
2024	3053	.8788824
Total	347373	100
Observations	347373	

A.3 Extensive Regression Results

Table 6: Regression Results

	(1)
	log_price
Min_dist	-0.0000441***
	(0.000000470)
Constant	12.52***
	(0.00208)
R-squared	0.0225
Observations	347373

Standard errors in parentheses

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

Table 7: Regression results for the close/far dummy variable. Shows us the relationship is economically piece wise.

	(1)
	log_price
Min_dist	0.314***
	(0.00283)
Constant	12.22***
	(0.00202)
R-squared	0.0326
Observations	347373

Standard errors in parentheses

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

Table 8: Multivariate regression results, min_dist is a polynomial and in-context with district, year, house age, and district-year interaction fixed effects.

	(1)
	log_price
Min_dist	-0.000126***
	(0.00000145)
min_dist_squared	1.09e-08***
	(1.35e-10)
N	0
	(.)
Y	0.115***
	(0.00475)
Year=1995	0
	(.)
Year=1996	0.0600***
	(0.00653)
Year=1997	0.192***
	(0.00648)
Year=1998	0.320***
	(0.00663)
Year=1999	0.483***
	(0.00652)
Year=2000	0.662***
	(0.00685)
Year=2001	0.783***
	(0.00684)

Year=2002	0.940***
	(0.00698)
Year=2003	1.061***
	(0.00729)
Year=2004	1.136***
	(0.00756)
Year=2005	1.183***
	(0.00804)
Year=2006	1.255***
	(0.00826)
Year=2007	1.381***
	(0.00870)
Year=2008	1.383***
	(0.00997)
Year=2009	1.358***
	(0.0106)
Year=2010	1.435***
	(0.0108)
Year=2011	1.454***
	(0.0113)
Year=2012	1.501***
	(0.0116)
Year=2013	1.552***
	(0.0118)
Year=2014	1.672***
	(0.0123)
Year=2015	1.770***

	(0.0127)
Year=2016	1.844***
	(0.0134)
Year=2017	1.874***
	(0.0143)
Year=2018	1.895***
	(0.0147)
Year=2019	1.899***
	(0.0154)
Year=2020	1.954***
	(0.0158)
Year=2021	1.975***
	(0.0154)
Year=2022	2.032***
	(0.0164)
Year=2023	2.006***
	(0.0171)
Year=2024	2.001***
	(0.0189)
BARKING AND DAGENHAM	0
	(.)
BARNET	2.765
	(1.732)
BEXLEY	5.589***
	(1.409)
BRENT	-18.80***
	(1.985)

BROMLEY	5.656***
	(1.398)
CAMDEN	-6.014**
	(2.238)
CITY OF LONDON	-23.00
	(15.16)
CITY OF WESTMINSTER	-17.20***
	(2.563)
CROYDON	5.677***
	(1.403)
EALING	-1.319
	(1.711)
ENFIELD	2.589
	(1.548)
GREENWICH	-3.600*
	(1.594)
HACKNEY	-29.40***
	(1.884)
HAMMERSMITH AND FULHAM	-4.303*
	(2.087)
HARINGEY	-11.89***
	(1.858)
HARROW	5.828**
	(1.884)
HAVERING	11.91***
	(1.403)
HILLINGDON	7.313***

	(1.485)
HOUNSLOW	5.766**
	(1.857)
ISLINGTON	-5.135*
	(2.103)
KENSINGTON AND CHELSEA	-12.13***
	(2.716)
KINGSTON UPON THAMES	4.581**
	(1.757)
LAMBETH	-9.541***
	(1.563)
LEWISHAM	-17.67***
	(1.465)
MERTON	-0.689
	(1.881)
NEWHAM	-13.05***
	(1.550)
REDBRIDGE	1.208
	(1.579)
RICHMOND UPON THAMES	9.189***
	(1.756)
SOUTHWARK	-17.08***
	(1.809)
SUTTON	8.695***
	(1.509)
TOWER HAMLETS	2.964
	(1.962)

WALTHAM FOREST	-22.81***
	(1.452)
WANDSWORTH	-2.433
	(1.529)
BARKING AND DAGENHAM × Year	-0.00161*
	(0.000762)
BARNET × Year	-0.00268**
	(0.000863)
BEXLEY × Year	-0.00424***
	(0.000702)
BRENT × Year	0.00799***
	(0.000990)
BROMLEY × Year	-0.00416***
	(0.000696)
CAMDEN × Year	0.00186
	(0.00111)
CITY OF LONDON × Year	0.0102
	(0.00756)
CITY OF WESTMINSTER × Year	0.00753***
	(0.00128)
CROYDON × Year	-0.00424***
	(0.000699)
EALING × Year	-0.000691
	(0.000852)
ENFIELD × Year	-0.00268***
	(0.000771)
GREENWICH × Year	0.000413

	(0.000795)
HACKNEY × Year	0.0133***
	(0.000939)
HAMMERSMITH AND FULHAM × Year	0.000992
	(0.00104)
HARINGEY × Year	0.00456***
	(0.000926)
HARROW × Year	-0.00429***
	(0.000938)
HAVERING × Year	-0.00737***
	(0.000699)
HILLINGDON × Year	-0.00506***
	(0.000740)
HOUNSLOW × Year	-0.00425***
	(0.000926)
ISLINGTON × Year	0.00134
	(0.00105)
KENSINGTON AND CHELSEA × Year	0.00515***
	(0.00135)
KINGSTON UPON THAMES × Year	-0.00350***
	(0.000875)
LAMBETH × Year	0.00345***
	(0.000778)
LEWISHAM × Year	0.00744***
	(0.000730)
MERTON × Year	-0.000965
	(0.000937)

NEWHAM × Year	0.00496***
	(0.000772)
REDBRIDGE × Year	-0.00205**
	(0.000787)
RICHMOND UPON THAMES × Year	-0.00569***
	(0.000874)
SOUTHWARK × Year	0.00720***
	(0.000902)
SUTTON × Year	-0.00568***
	(0.000752)
TOWER HAMLETS × Year	-0.00282**
	(0.000979)
WALTHAM FOREST × Year	0.00988***
	(0.000723)
WANDSWORTH × Year	0
	(.)
Constant	14.07***
	(1.525)
R-squared	0.633
Observations	347373

Standard errors in parentheses

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

Table 9: Regression results using the close/far dummy Min_dist is either 0 or 1. Shows us that the dummy effect is still economically and statistically significant in-context with our controls.

	(1)
	log_price
Min_dist=0	0
	(.)
Min_dist=1	0.189***
	(0.00290)
N	0
	(.)
Y	0.119***
	(0.00476)
Year=1995	0
	(.)
Year=1996	0.0605***
	(0.00663)
Year=1997	0.192***
	(0.00649)
Year=1998	0.321***
	(0.00663)
Year=1999	0.482***
	(0.00655)
Year=2000	0.661***
	(0.00681)
Year=2001	0.782***

	(0.00684)
Year=2002	0.939***
	(0.00696)
Year=2003	1.060***
	(0.00732)
Year=2004	1.135***
	(0.00752)
Year=2005	1.182***
	(0.00793)
Year=2006	1.255***
	(0.00795)
Year=2007	1.380***
	(0.00827)
Year=2008	1.384***
	(0.00966)
Year=2009	1.360***
	(0.0100)
Year=2010	1.438***
	(0.00992)
Year=2011	1.457***
	(0.0103)
Year=2012	1.502***
	(0.0105)
Year=2013	1.554***
	(0.0106)
Year=2014	1.673***
	(0.0109)

Year=2015	1.771***
	(0.0113)
Year=2016	1.846***
	(0.0117)
Year=2017	1.877***
	(0.0122)
Year=2018	1.899***
	(0.0126)
Year=2019	1.903***
	(0.0130)
Year=2020	1.957***
	(0.0134)
Year=2021	1.979***
	(0.0133)
Year=2022	2.035***
	(0.0139)
Year=2023	2.010***
	(0.0146)
Year=2024	2.005***
	(0.0166)
BARKING AND DAGENHAM	0
	(.)
BARNET	2.225
	(1.945)
BEXLEY	5.123**
	(1.974)
BRENT	-19.32***

	(2.103)
BROMLEY	5.577**
	(1.858)
CAMDEN	-6.402**
	(2.130)
CITY OF LONDON	-23.19**
	(7.644)
CITY OF WESTMINSTER	-17.68***
	(2.043)
CROYDON	4.450*
	(1.851)
EALING	-2.391
	(1.985)
ENFIELD	2.017
	(1.944)
GREENWICH	-4.107*
	(2.007)
HACKNEY	-29.34***
	(2.163)
HAMMERSMITH AND FULHAM	-4.642*
	(2.135)
HARINGEY	-12.52***
	(2.030)
HARROW	5.490*
	(2.198)
HAVERING	10.56***
	(1.962)

HILLINGDON	6.397**
	(1.954)
HOUNSLOW	5.913**
	(2.080)
ISLINGTON	-5.364*
	(2.195)
KENSINGTON AND CHELSEA	-12.54***
	(2.152)
KINGSTON UPON THAMES	3.984
	(2.123)
LAMBETH	-9.866***
	(1.897)
LEWISHAM	-17.46***
	(1.927)
MERTON	-1.193
	(2.029)
NEWHAM	-14.91***
	(2.059)
REDBRIDGE	-0.242
	(2.023)
RICHMOND UPON THAMES	8.947***
	(1.985)
SOUTHWARK	-18.27***
	(2.031)
SUTTON	8.082***
	(2.017)
TOWER HAMLETS	1.581

	(2.261)
WALTHAM FOREST	-23.41***
	(1.948)
WANDSWORTH	-3.230
	(1.829)
BARKING AND DAGENHAM × Year	-0.00201*
	(0.000911)
BARNET × Year	-0.00281***
	(0.000719)
BEXLEY × Year	-0.00439***
	(0.000738)
BRENT × Year	0.00788***
	(0.000822)
BROMLEY × Year	-0.00447***
	(0.000659)
CAMDEN × Year	0.00168*
	(0.000839)
CITY OF LONDON × Year	0.00998**
	(0.00375)
CITY OF WESTMINSTER × Year	0.00740***
	(0.000783)
CROYDON × Year	-0.00402***
	(0.000654)
EALING × Year	-0.000548
	(0.000746)
ENFIELD × Year	-0.00280***
	(0.000718)

GREENWICH × Year	0.000254
	(0.000760)
HACKNEY × Year	0.0129***
	(0.000860)
HAMMERSMITH AND FULHAM × Year	0.000783
	(0.000842)
HARINGEY × Year	0.00448***
	(0.000775)
HARROW × Year	-0.00451***
	(0.000882)
HAVERING × Year	-0.00709***
	(0.000731)
HILLINGDON × Year	-0.00500***
	(0.000726)
HOUNSLOW × Year	-0.00472***
	(0.000808)
ISLINGTON × Year	0.00107
	(0.000880)
KENSINGTON AND CHELSEA × Year	0.00498***
	(0.000853)
KINGSTON UPON THAMES × Year	-0.00362***
	(0.000834)
LAMBETH × Year	0.00322***
	(0.000686)
LEWISHAM × Year	0.00692***
	(0.000707)
MERTON × Year	-0.00110

	(0.000774)
NEWHAM × Year	0.00549***
	(0.000795)
REDBRIDGE × Year	-0.00172*
	(0.000771)
RICHMOND UPON THAMES × Year	-0.00597***
	(0.000745)
SOUTHWARK × Year	0.00740***
	(0.000776)
SUTTON × Year	-0.00579***
	(0.000767)
TOWER HAMLETS × Year	-0.00252**
	(0.000921)
WALTHAM FOREST × Year	0.00979***
	(0.000721)
WANDSWORTH × Year	0
	(.)
Constant	14.57***
	(1.826)
R-squared	0.631
Observations	347373

Standard errors in parentheses

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