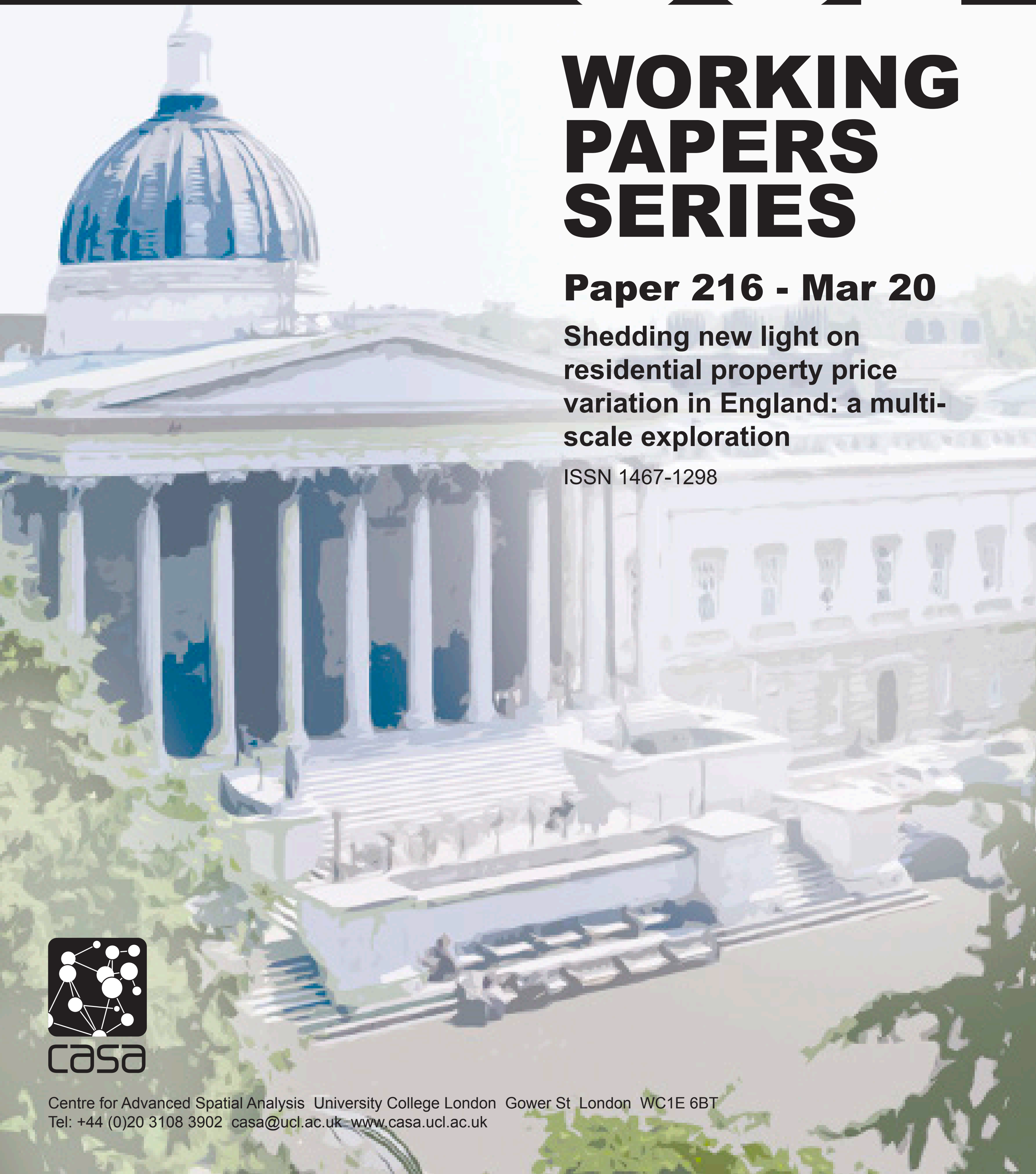
UCL CENTRE FOR ADVANCED SPATIAL ANALYSIS



WORKING PAPERS SERIES

Paper 216 - Mar 20

Shedding new light on residential property price variation in England: a multi-scale exploration

ISSN 1467-1298

Centre for Advanced Spatial Analysis University College London Gower St London WC1E 6BT
Tel: +44 (0)20 3108 3902 casa@ucl.ac.uk www.casa.ucl.ac.uk

Shedding new light on residential property price variation in England: a multi-scale exploration

Bin Chi, Adam Dennett, Thomas Oléron-Evans, Robin Morphet

Abstract

Exploring the nature of spatial and temporal variation in house prices is important because it can help better understand such issues as affordability and equity of access to housing. In the UK, research on house price variation has been hindered by a lack of extensive data linking the prices of properties at different places and times to their physical attributes. This paper addresses this gap through using a new dataset linking Land Registry Price Paid Data (PPD) to attribute data from Ordnance Survey (OS) and Energy Performance Certificate (EPC) datasets. The new data are used to investigate spatial disparities in England's house prices at four geographical scales (from local authority to individual address) between 2009 and 2016 – a period of sustained price rises after the global financial crisis of 2008. Multilevel variance components models are used to estimate house price variation at different levels of spatial aggregation, and we compare spatial disparities in transaction price against the house price per square metre at different levels. Our results suggest that accounting for the size of properties by using house price (£/m^2) offers a more accurate picture of house price variation than does the use of transaction prices at the same geographic scale. Spatial disparities in house price (£/m^2) are more apparent and are seen to be clustered at local authority level and highly clustered at MSOA level, with imbalances increasing during this eight-year period and highlighting the strong and growing influence of London on the national housing market.

Keywords: Housing, House price variation, multilevel modelling, England

1. Introduction

Shelter is a basic human need, but a majority of countries now face critical housing challenges (UN-HABITAT, 2011a, 2011b, 2012). In wealthier countries such as the UK, residential property prices, on the whole, have been rising since the mid-1990s (Knoll et al., 2017), but prices can exhibit huge sub-national geographic variations (Fotheringham et al., 1998; Huang et al., 2010). In the UK, prices in London can be prohibitively high, whereas in the de-industrialised North, properties have occasionally been purchased for a few pounds (Edwards, 2016; Nsubuga, 2018). These variations can be understood through examining published house price statistics (the term 'house price' is frequently used in relation to all types of residential property). These have been reported in Britain for a long time but local heterogeneity in stock composition, sales frequency and transaction volumes means that data analysis can be challenging, particularly when focusing on small area geographies. Transaction prices are frequently reported at regional or city level in order to smooth out some of this variation. Different organisations produce these statistics using different measurement methods and datasets (Chandler and Disney, 2014; Wood, 2015). The Nationwide Building Society produces a regional quarterly house price index (Nationwide Building Society, 2015) with its mortgage lending data while the Office for National Statistics (ONS) publishes a regional house price index (Office for National Statistics et al., 2016) based on Land Registry data. Both datasets are published at regional level, but they are not comparable due to differences in data definition and the housing characteristics considered when using the 'mix-adjusted' approach. Lack of comparability can present difficulties when researching house price variation (Gray, 2012; Hamnett and Reades, 2019).

Despite data challenges, quantitative analysis of house price variation at multiple geographic scales has been conducted within individual cities in the UK, such as London (Feng and Jones, 2016; Law, 2018) and Cardiff (Orford, 2002; Wang et al., 2015). In three of these four papers, transaction price was used to explore house price variation (Feng and Jones, 2016; Orford, 2002; Wang et al., 2015) using the Land Registry PPD. Furthermore, local variations in stock composition and other factors mean that simple aggregation for the purposes of studying price variations is problematic. One suggested solution has been to examine house price patterns using price (£/m^2), but to date, only Powell-Smith (2017) has achieved this with any degree of success.

Systematic analysis of house price variation at a variety of geographic scales and accounting for the local variations in stock characteristics could better aid both government and public in understanding housing inequalities in England. This research takes a first step by modelling the patterns of house price variation across England at multiple geographical scales whilst taking into account local variations in total floor area. House price variation in terms of both transaction price and house price normalised by total floor area, are explored utilising a new dataset which links entries in the Land Registry Price Paid Data (PPD) to building locations from Ordnance Survey (OS) datasets and total floor area from Domestic Energy Performance Certificates (EPCs) data (Chi et al. 2019). In this paper we first demonstrate the utility of this new linked dataset, and then explore how the factors that affect house prices can operate at different spatial scales. Accounting for variations in individual property characteristics (i.e. total floor area) permits these spatial effects to be revealed. The second section reviews relevant literature on the spatial dimension of house price variation. Our study area and data are introduced in the third. The fourth section presents the concept of multilevel variance components models and defines sixteen multilevel models used in this research. The fifth section presents and discusses the results. Finally, conclusions and recommendations for future research are provided.

2. The spatial dimension of house price variation

House prices are spatially auto-correlated in small areas but also spatially heterogeneous in different geographical locations (Basu and Thibodeau, 1998; Goodman and Thibodeau, 2003). Research shows the drivers behind house price variation are complex and operate at different geographical scales in England (Cook, 2005; MacDonald and Taylor, 1993; Szumilo et al., 2017; Yao and Fotheringham, 2016). At the broadest national or regional scales, house prices are influenced by macro-structural political, economic and demographic factors (Ferrari and Rae, 2013; Meen, 1999), such as regional economic development, infrastructure provision and policies affecting migration. At a scale that might include cities or travel-to-work areas, house prices are influenced by urban form, local economic conditions and amenities, and the availability of different transport modes (Downes, 2018; Smith, 2018). At local neighbourhood or electoral ward scales, house prices are influenced by local amenities, the character of neighbouring households, local public goods (i.e. schools and open space) and the availability of public transport (Orford, 2002). House prices in the same neighbourhood tend to be similar to each other, but may vary as a result of physical qualities such as dwelling size, age, structural design and historic value (Ahlfeldt et al., 2012; Goodman and Thibodeau, 1995; Kain and Quigley, 1970).

Various modelling techniques have been proposed to capture spatial heterogeneity in property prices over a large area, such as traditional regression-based hedonic price models (Visser et al., 2008), geographically weighted hedonic regression models (Helbich et al., 2014; Lu et al., 2014; Yu, 2007), geographically and temporally weighted regression models (Fotheringham et al., 2015; Huang et al., 2010) and multilevel models (Jones and Bullen, 1993, 1994). Jones (1991) first applied multilevel models to analyse house price variation showing that they demonstrate a considerable improvement over the traditional linear hedonic price modelling. Subsequent research using multilevel models has explored house price variation across the world (Dong et al., 2015; Goodman and Thibodeau, 1998; Leishman, 2009). Orford (2002) applied multilevel modelling to estimate the effects of location upon Cardiff house prices suggesting that the overall house price variation is composed of variation within districts, within communities and across individual properties. Recently, Feng and Jones (2016) were the first to present London's house price variation at five geographical scales determined by postcode and census geography. These two geographical classifications showed that transaction price was hierarchical and highly clustered at smaller geographical scales.

Current research suggests house price variations can occur across geographies but may also vary with house characteristics. Aggregate statistics for house prices at a large geographical scale will mask variation at smaller scales. Moreover, most previous hedonic analyses are applied to regional level or single city examples. A framework that integrates housing market analyses in England at different scales does not currently exist. This suggests an incomplete understanding of the spatial determinants of the variations in Britain's housing market.

3. Study area and Data

3.1 Study area and geographical scales

The study area is England. It contains nine regions: the North East, the North West, Yorkshire and the Humber, East Midlands, West Midlands, East of England, the South East, the South West and London. Administratively, England is divided into 326 local authority districts within which are 6791 smaller statistical geographies known as Middle Layer Super Output Areas (MSOAs)[1]. Nested within these are 32,842 Lower Layer Super Output Areas (LSOAs)[2]. England's house prices are frequently aggregated at regional, local authority, MSOA and LSOA levels. Regional level house prices have been well explored and shown as exhibiting a ripple effect pattern (Cook and Watson, 2016;

MacDonald and Taylor, 1993). Few studies have explored house price variance at small geographical levels. Thus this research considers three small geographical levels (from Local Authority to LSOA) plus the individual address level.

3.2 House price data

The Land Registry PPD[3] provides the most accurate picture of full market sale prices in the UK residential housing market (Marsden, 2015), but the dataset contains only a limited set of other attributes about each property. This is a barrier to analysis of house price variance as attributes such as property size are important determinants of price (De Nadai and Lepri, 2018; Orford, 2010). Thus this paper uses a new linked dataset (Chi et al., 2019) combining the Land Registry PPD with Domestic EPCs data containing information on total floor area. This enables house price variance analysis in terms of transaction both normalised and non-normalised by floor area. The new linked dataset contains 4,682,468 transactions between 2009 and 2016, representing 80% of the full market housing sales over this period. Figure A in supplementary material A plots total transaction price against house price per square metre and reveals only weak covariance, suggesting we are justified in exploring both in this paper.

4. Methodology

4.1 Multilevel variance components model

The multilevel variance components model is a statistical model without explanatory variables (Ren et al., 2013; Raudenbush, 2002). In exploring house price variations, it simultaneously decomposes the total house price variance at and across different geographical scales, thus quantifying the extent of spatial effects on house prices. Properties can be viewed as being nested within different geographical jurisdictions. Given a house price dataset, which records transactions occurring within different local authorities, a two-level variance components model could be formulated, in which level 1 is the property level and level 2 is local authority level. This multilevel variance components model can be written as:

$$h_{ij} = \beta_0 + l_j + e_{ij} \qquad (1)$$

$l_j \sim N(0, \sigma_l^2)$
$e_{ij} \sim N(0, \sigma_e^2)$
here h_{ij} is the individual house price for the *i*th transaction in local authority j (e.g. Camden) in a given year, β_0 is the fixed effect, representing the overall mean house price, and r_j and e_{ij} are the random effects of the variance components model, representing respectively, the residuals at the local authority level and the individual property level. The random effect arises from unobserved heterogeneity in characteristics that affect house prices (Feng and Jones, 2015; Snijders and Bosker, 2011). The deviation l_j measures the extent to which the mean house price in jurisdiction j varies from the overall mean house price (β_0), whilst e_{ij} represents the deviation of transaction price i from the mean price in its local authority j. Residuals at local authority level and individual level are assumed to be independent and to follow normal distributions with zero means and constant variance

of σ_r^2 and σ_e^2, respectively. Moreover, residuals at the same or different levels are assumed to be uncorrelated with one another.

Figure 1 illustrates house prices for nineteen individual properties in two local authorities (Camden and Sheffield) in England. Individual house prices are shown as black circles, the overall mean house price (β_0) is represented by the thick black horizontal line, and the mean house prices for Camden and Sheffield are shown as blue horizontal lines. Camden has an above average mean (positive l_j), Sheffield has a below-average mean (negative l_j). Each individual house price (e.g. h_{11} in Camden), is equal to the overall mean house price (β_0) plus its local authority level residual (l) and its individual-level residual (e_{11}).

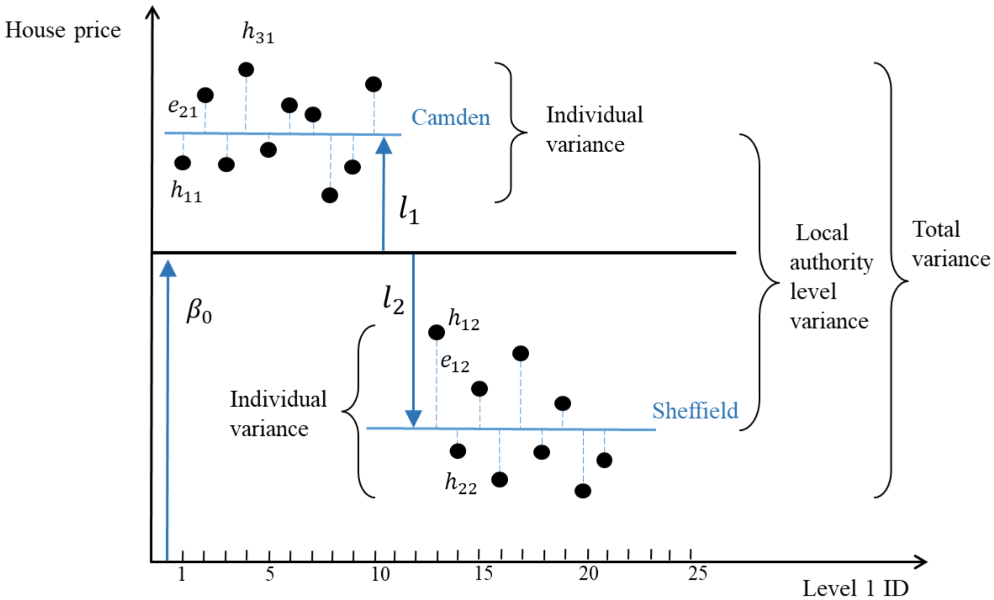


Figure 1. A graphical illustration of the two-level variance components model

The multilevel variance components model is decomposed by the overall mean house price (fixed effect) and the house price variation at each level (random effects). It treats the units at each level as a random sample from a larger population with an assumed distribution, decomposing the overall variance into two variance parts (σ_l^2 and σ_e^2). σ_l^2 is the variance at local authority level, presenting variability among group house prices; σ_e^2 is the residual variation at individual level. These terms are often called the "between-group variance" and the "within-group variance". We use 'group' as a generic term to describe clusters of individuals at one specific geographical level.

The two-level variance components model can be extended to three or more levels to examine the location effects at multiple scales simultaneously. Such an extension is straightforward, simply requiring the introduction of additional random effect terms. For example, a three-level model might have properties nested in MSOAs, which are nested within local authorities. This is written as:

$$h_{ikj} = \beta_0 + l_j + m_{kj} + e_{ikj} \qquad (2)$$

$l_j \sim N(0, \sigma_l^2)$
$m_{kl} \sim N(0, \sigma_m^2)$

$e_{ikj} \sim N(0, \sigma_e^2)$

here h_{ikj} is the individual house price for ith transaction in MSOA k of local authority j, while β_0, l_j, e_{ikj}, σ_l^2 and σ_e^2 have the same meaning as before in equation 1. The new random term m_{kj} measures the extent to which the mean house price of MSOA k deviates from the mean house price in local authority j. m_{kj} is assumed to follow a normal distribution with zero mean and constant variance σ_m^2. Residuals at the same level are uncorrelated with each other, and residuals at different levels are also uncorrelated with each other.

Variance partition coefficients (VPC) can be used to interpret the variance components in a multilevel model. Taking the local authority level as an example, the VPC of the local authority level is calculated as the ratio of the local authority variance to the total variance. It represents the proportion of the house price variance that can be attributed to differences between local authorities. The VPC ranges from 0 to 1; 0 signifying no between group differences and 1 signifying no within group differences. A higher VPC at a particular level indicates larger differences between groups at that level. In the three level variance components model, the total house price is decomposed into three variance components: individual variance (σ_e^2), MSOA level variance (σ_m^2) and local authority level variance (σ_l^2). The equation for VPC at individual level are presented in equation 3, following the equations at MSOA (equation 4) and local authority level (equation 5).

$$VPC_e = \frac{\sigma_e^2}{\sigma_l^2+\sigma_m^2+\sigma_e^2} \qquad (3)$$

$$VPC_m = \frac{\sigma_m^2}{\sigma_l^2+\sigma_m^2+\sigma_e^2} \qquad (4)$$

$$VPC_l = \frac{\sigma_l^2}{\sigma_l^2+\sigma_m^2+\sigma_e^2} \qquad (5)$$

The intraclass correlation coefficient (ICC) offers another approach to interpreting the variance component. It measures the expected correlation (similarity) of observations within the same group at a particular level (Bartholomew et al., 2008). It is expressed as a ratio of variances, comparing the house price variance that is between groups at a particular level to the total variation (Finch et al., 2014). In terms of ICC's and VPC's algebraic form, the ICC at any given level is the sum of the VPC at this level and the higher levels. Thus the ICC at MSOA level is the sum of VPC at MSOA level and local authority level. The ICC of the highest level (local authority) coincides with its VPC. The ICC ranges from 0 to 1, with higher values indicating a greater degree of clustering (meaning data is more similar within groups, with larger differences between groups). Equations for the ICC from individual level to local authority level are shown in equations 6 to 8:

$$ICC_e = \frac{\sigma_l^2+\sigma_m^2+\sigma_e^2}{\sigma_l^2+\sigma_m^2+\sigma_e^2} = 1 \qquad (6)$$

$$ICC_m = \frac{\sigma_l^2+\sigma_m^2}{\sigma_l^2+\sigma_m^2+\sigma_e^2} \qquad (7)$$

$$ICC_l = VPC_l = \frac{\sigma_l^2}{\sigma_r^2+\sigma_l^2+\sigma_e^2} \qquad (8)$$

4.2 Exploring spatial influences on price variation

A four-level variance components model was used to estimate house price variance at the different levels of spatial aggregation. Level 1 is the individual residential property, level 2 is the LSOA level, level 3 is the MSOA level and level 4 is the local authority level.

Models for each year from 2009 to 2016 are created for both total transaction price and house price (£/m^2) (Table 1).

Table 1. The candidate five-level variance component models[4]

Model	Model name	Equations
TP	TP2009 - TP2016	$tp_{ijkh} = \beta_0 + l_h + m_{kh} + o_{jkh} + e_{ijkh}$
HP	HP2009 - HP2016	$hp_{ijkh} = \beta_0 + l_h + m_{kh} + o_{jkh} + e_{ijkh}$
Note: tp is the transaction price and hp is house price per square metre in a certain year. For example, the tp in Model TP 2009 is the transaction price in 2009. The parameter β_0 is the overall mean house price in England, l_h, m_{kh}, o_{jkh} and e_{ijkh} are the residuals of level 4 to 1.		

5. Results and discussion

Models presented in Table 1 were run using MLwiN 3.03 (Charlton et al., 2019). Likelihood ratio tests are used to establish whether the four-level variance components model fits the data significantly better than the null single-level model. Each four-level model in Table 1 is preferred to its null single-level model based on the near zero p-value of the likelihood ratio test. In addition, each four-level multilevel model was compared to a set of three-level models formed by dropping one geographic level for each comparison (e.g. dropping the LSOA level in the four-level model). All comparisons showed a significant increase in explanatory power with increasing numbers of levels according to the near zero p-values obtained from likelihood ratio tests. Hence the following discussions are based on the estimated coefficient values for the four-level variance component models.

5.1 Overall house price change and house price variance

Following the financial crisis, both the estimated mean transaction price and mean house price (£/m^2) show the same increasing trend from 2009 to 2016 (Figure C1.A in supplementary material C). Overall house price variation increased between 2009 and 2016 (Figure C1.B in supplementary material C). The trend of house price variance differs depending on whether transaction price is normalised by the floor area of the property – henceforth we shall term the normalised transaction price as simply the house price (£/m^2). The variance of house price (£/m^2) increased between 2014 and 2015, while the variance of the transaction price decreased. Comparing the data in 2015 to 2014, a smaller number of full market value residential sales together with fewer sales at extremely high prices are the main reasons for the decrease in transaction price variance. This may be due to the increasing Stamp Duty Land Tax (SDLT) rates on higher bands at the end of 2014 limiting purchases of more expensive dwellings (Scanlon et al., 2017). One explanation for the trend discrepancy between transaction price and house price (£/m^2) is the different mix of stock sold in different years. For example, a higher proportion of large dwellings (total floor area bigger than 250 m^2) with high transaction prices (over £5 million) were sold in London's housing market in 2014, but a lower proportion of these dwellings were sold in London in 2015 (Figure C2 in supplementary material C). Using the normalised price (£/m^2) approach, these large dwellings may have a low house price (£/m^2), however, the small dwellings with high transaction prices may have a higher house price (£/m^2). Therefore, the variance of house price (£/m^2) could increase. The overall transaction price variance is smaller than house price (£/m^2) variance. This not only means that normalised transaction prices by the floor area are more concentrated, but also that differences in total floor area contribute greatly to transaction price variance.

5.2 House price variance at four geographic scales

For both transaction price and house price(£/m^2), VPC results reveal (Figure 2) that house price variance does exist between local authorities, within-local-authority-between-MSOAs and also within-MSOA-between-LSOAs. House price (£/m^2) variance across different levels follows a similar pattern as transaction price variance, but with a higher variability at the same level. Comparing VPC at the same level and same year for transaction prices against house price (£/m^2) reveals that once variations in property size are controlled for, spatial effects become much stronger. This reveals that controlling for floor area offsets much of the house price variation among individual properties and correspondingly increases VPC at higher geographic levels (i.e. level 4 to level 2). Local authority effects (compared to other spatial effects) had more of an influence on transaction prices in 2016 than in 2009, but when floor area is accounted for, this change in the local authority's influence is even more noticeable. Meanwhile, MSOA or LSOA effects are stable from 2009 to 2016. The conclusion drawn here is that the house price (£/m^2) aggregated at geographic level (i.e. local authority level or MSOA level or LSOA level) offers a more accurate picture of the England's housing market than transaction price.

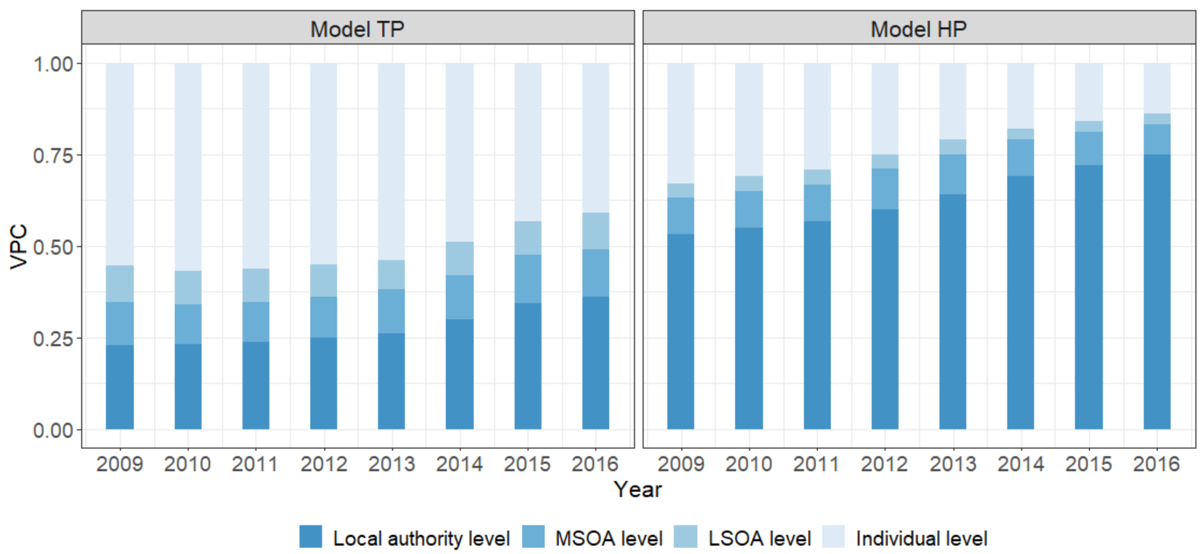


Figure 2. VPC results for *model TP and HP*

5.3 House price (£/m^2) *clustering at four geographic levels between 2009 to 2016*

The ICC at a given level shows the degree of house price clustering. The ICC results of model HP are presented in Table 2. ICC at local authority level is 0.53 in 2009 and continues increasing to 0.75 by 2016, this illustrates that house price (£/m^2) are clustered at local authority level. ICC at MSOA level is 0.63 in 2009 and continues increasing to 0.83 by 2016. Meanwhile, the ICC at MSOA level shows negligible improvement at LSOA level. This suggests that house prices (£/m^2) at MSOA level are highly clustered and variations within the same MSOA unit are quite small between 2009 and 2016. This also suggests that using the mean house price(£/m^2) at MSOA level gives a relatively clear house price picture (2009-2016) and very little additional explanatory power is gained from observing house price variations at a more granular geographical scale. This spatial association is helpful as this highly auto-correlated relationship between the house price (£/m^2) at the MSOA level, makes predicting house prices at this level more reliable.

Table 2. ICC results for the HP models

Level	Model HP							
	2009	2010	2011	2012	2013	2014	2015	2016
Local authority level	0.53	0.55	0.56	0.60	0.64	0.69	0.72	0.75
MSOA level	0.63	0.65	0.67	0.72	0.75	0.79	0.81	0.83
LSOA level	0.67	0.69	0.71	0.75	0.79	0.82	0.84	0.86
Individual level	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

5.4 House price (£/m^2) variation at local authority level

The ICC at MSOA level is equal to the VPC at MSOA level plus local authority level. Owing to a noticeable VPC increase at local authority level between 2009 and 2016, ICC at MSOA level shows a strong increase. House prices (£/m^2) became more highly clustered at MSOA level between 2009 and 2016 as shown by the increase in ICC, which is largely due to the noticeable VPC increase at local authority level. Owing to the total house price (£/m^2) variance increases between 2009 and 2016 (Figure C1), increasing house price variance at local authority level is the main reason behind the VPC increase between 2009 and 2016 (Figure 2). Figure 3 is a graphical illustration of this increasing variation from 2009 to 2016 and shows estimated residuals at local authority level (l_j) as scatter plots. Each point represents the residual of one local authority and similarly coloured points belong to the same region. Residuals are ranked across England. The red horizontal line is the zero residual line, which presents the overall mean house price in England (β_0). It is obvious that the house price variance at local authority level is largely due to local authorities in London with extremely high house prices.

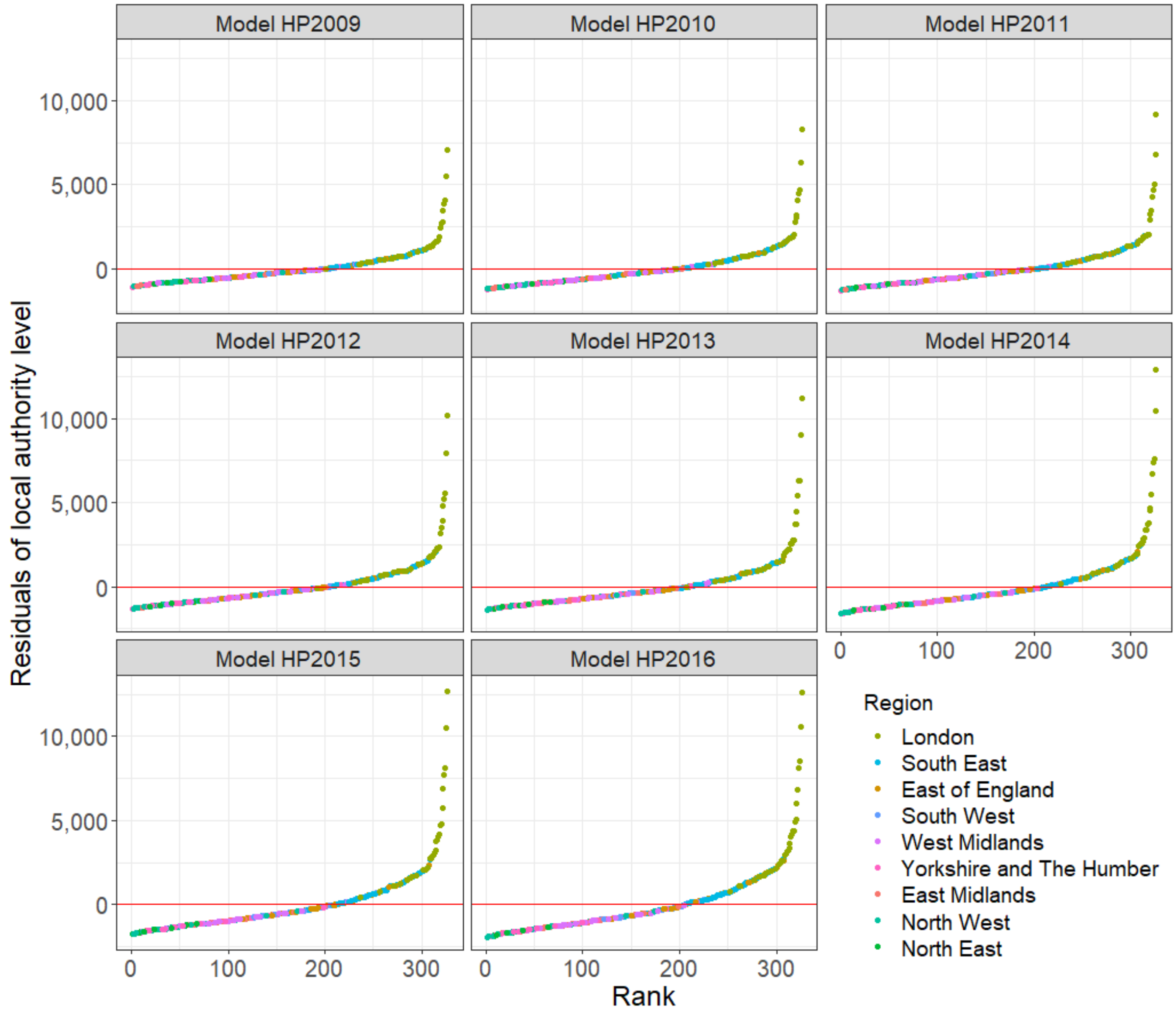


Figure 3. Residuals at local authority level in England in the HP models

Examining local authority residuals further by plotting them for different regions (supplementary material D), shows that, with the exception of Barking and Dagenham, local authorities' house prices in London are consistently above the overall mean house price in England, with a continuously widening house price difference. London can be classed as an 'outlier' region in England and maintains its position as the most expensive region. London's local authorities display a more rapidly increasing house price than the local authorities in other regions. This London effect dominates the increasing house price variation at local authority level from 2009 to 2016. Meanwhile, relatively small house price increases in the North East and the East of England also make a small contribution to the widening differential in regional house prices.

5.4.1 London's local authority house price variation

Figure 4 displays local authority residuals in London (third column plots in supplementary material D) and offers a graphical illustration of London's increasing house price variation from 2009 to 2016. Most local authorities in London have a mean price a lower than 2000 £/m^2. Kensington and Chelsea (KC) consistently comes top in this ranking, followed by

Westminster (W) and then Camden (C). Some other local authorities also show a substantial increase in house prices, contributing to the increasing local authority house price difference. For example, Southwark (S), Lambeth (L), Hackney (H) and Tower Hamlets (HM) were lower than 2,000 level (grey dashed line) from 2009 to 2011, but after that, their increasing prices started to exceed the 2,000 level. Consequently, house prices (£/m^2) at local authority level in London have become more polarised, with central and south-western boroughs such as Westminster, Kensington and Chelsea, Hammersmith and Fulham, Camden and Richmond pulling further away from the 2,000 level.

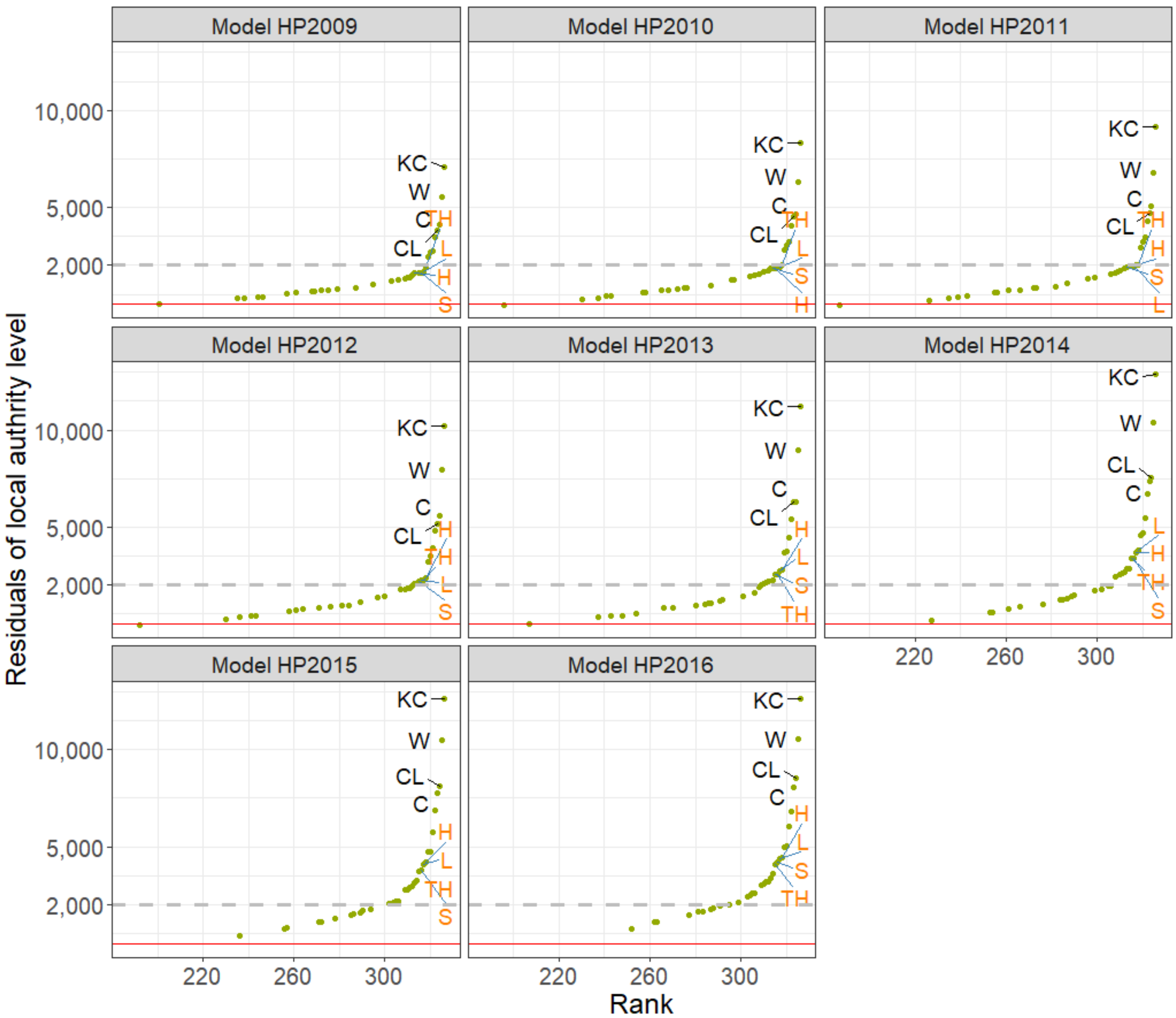


Figure 4. Residuals at local authority level in London in the HP models

Looking at the rank of local authority residuals within London (E), Kensington and Chelsea consistently tops the price league with five other central London boroughs (Westminster, Camden, City of London, Hammersmith and Fulham, and Islington) consistently high. At the other end of the scale, Barking and Dagenham consistently exhibits the lowest prices. The other fourteen local authorities generally keep their rank orders or drop slightly within London, however Local authorities that increased their rank order from 2009 to 2016 are mainly located in outer London to the east and southwest of the centre (Green text label in supplementary material E and F).

5.4.2 London's house price effect in England's housing market

London's high house price (£/m^2) also affects neighbouring housing markets. supplementary material G displays the residuals at local authority level in England at 2009 with a separate map of housing markets in nearby London. Residuals at local authority level in England are grouped into 8 sub-groups using the natural breaks method. The local authorities with darker shades show relatively high house prices when compared to the price in England. These relatively high price areas in the 'Home Counties' in the South East and the East of England all are all within the commuter belt with many living in these areas still working in London and properties consequently being linked to the London housing market. The 11 local authorities labelled in supplementary material G are all in the top 30 price league in England in 2009. London's effect on the South East appears to be stronger than on the East of England, which could well correspond to the higher density of commuter rail routes to the South West of the capital when compared to the North East.

Setting aside London's effect within its housing market and on the housing market in nearby local authorities, the property price (£/m^2) still shows considerable variability at local authority level in the remaining eight regions. There are three groups of local authorities exhibiting relatively high price areas distant from London. These are on the South coast - South Ham, Christchurch and East Dorset; the Cotswolds, straddling the north east part of the South West the West and East Midlands; and moving to the North of England, a third group comprising contiguous local authorities near to large national parks (e.g. South Lakeland). Local authorities within these three groups also contribute to house price differentials at local authority level. Interestingly, the London effect may still be operating here, as these areas are characterised by relatively high volumes of second-home ownership, with owners registering other homes in London (Dennett, 2013).

6. Conclusions and future research

Understanding the nature and extent of differentials in property prices at different geographical levels leads to a better understanding of housing markets. This study compares house price variation in transaction house price and house price (£/m^2) across England by using a new linked dataset. Examining standardised prices in this way exposes spatial differences (obscured by differences in local stock mix and transaction patterns) more prominently. This confirms that house price (£/m^2) offers a more meaningful picture of house price variation than transaction price alone. Examining house price variation at four geographic scales across England using a four-level variance components model suggests that house price drivers operate differently at different geographical scales and that these effects changed between 2009 and 2016. House price differentials between local authorities are quite large and this spatial effect increases at MSOA level. House prices (£/m^2) are generally very similar within MSOAs, with little to be gained from exploring variations at the lower LSOA level. This is a useful practical finding and demonstrates that aggregating house price (£/m^2) at MSOA level might be sufficient to give a clear picture of local housing markets or sub-markets in England.

Overall house price variability in England shows an increase from 2009 to 2016. In 2009, 53% of house price variation existed between local authorities. The magnitude of disparities increased 1.42 times in the following eight years. This increasing imbalance follows from London's house prices increasing more rapidly than that of any other region. While looking at house price (£/m^2) variation between local authorities by plotting the residuals of local authority level change, we found that some local authorities in the central part of London are the main source of this increasing local authority effect. Moreover, London affects not only house price differences between regions but also its nearby local authorities. Local authorities in the South East and East of England which are near to London show the highest house price within their regions. Of the top 30 local authorities with the most expensive house prices, 19 of them are located in London and the remaining 11 local authorities are located outside London but within its travel to work area. The current housing policy (e.g. Right to Buy) which differentiates between London and outside, would be more consistent if based on house prices (£/m^2), thus including some of the more expensive areas bordering London. Excepting the local authority housing markets nearby London, some local authorities near South Ham, Cotswold and South Lakeland also show a high house price spatial clustering. These areas contribute to house price imbalances after excluding the near-London effect.

This research has demonstrated that multilevel variance components modelling can offer a model-based descriptive analysis for the exploration of house price variation at multiple geographic scales, providing a new insight into spatial house price disparities across England. Having created a new time-series dataset linking transaction prices to a variety of housing attributes and established a methodology, we intend to extend this work through a more thorough exploration of house price variance at local authority and MSOA level by considering the interacting effects of time, location and key local factors such as plot size, land use structure, housing density, local physical and socio-economic environments. Understanding the underlying mechanisms of house price variation offers the potential for deeper insights into pressing housing inequality issues in England.

Notes

1. MSOA is a Census geographic level in the UK. It is the next lower hierarchic level after local authority. One MSOA contains 5,000 to 15,000 residents. Resource: https://www.towerhamlets.gov.uk/Documents/Borough_statistics/Research-tools-and-guidance/RB-Census2011-Census-Geography-Guide-2013-05.pdf and https://census.ukdataservice.ac.uk/use-data/guides/boundary-data.aspx

2. LSOA is a Census geographic level in the UK and is smaller than MSOA. One LSOA contains 1,000 to 3,000 residents. It is the next lower hierarchic level after MSOA. Resource: https://www.towerhamlets.gov.uk/Documents/Borough_statistics/Research-tools-and-guidance/RB-Census2011-Census-Geography-Guide-2013-05.pdf and https://census.ukdataservice.ac.uk/use-data/guides/boundary-data.aspx

3. The Land Registry PPD is an open administrative dataset from Her Majesty's Land Registry data. It is based on a complete register of all residential property sales since 2013, at full market value in England and Wales but excludes some other types of residential property sales; details are available at:

https://www.gov.uk/guidance/about-the-price-paid-data#data-excluded-from-price-paid-data.

4. The model used in this research is based on the transaction price and house price per square metre; we do not take logarithms of prices, as is common practice across some of the literature for the following reasons. First, the VPC pattern across the four-levels is similar whether using house price or log house price as the output variable (Table B in supplementary material B). The conclusions of this research would therefore be the same under both methods. Second, when comparing the estimated mean price and observed mean price from the house price model (i.e. HP2009) with the log house price model, the two models show similar results and the estimated price in the house price model shows a higher association with the observed mean (Figure B). Given these two reasons, this research uses the simpler and more easily interpreted model based on the unlogged price.

Acknowledgements

The authors would like to thank Mrs Jane Galbraith and Professor Kelvyn Jones for providing valuable guidance and basic knowledge of variance analysis. The author also would like to thank Rob Liddiard for sharing his expertise regarding EPC data during the earlier stages of this research.

Funding

This research was funded by the China Scholarship Council (CSC No. 201708060184).

References

Ahlfeldt GM, Holman N and Wendland N (2012) An assessment of the effects of conservation areas on value. Report, The London School of Economics and Political Science, UK, May.

Bartholomew DJ, Steele F, Moustaki I, et al. (2008) *Analysis of Multivariate Social Science Data, Second Edition*. Boca Raton: CRC Press.

Basu S and Thibodeau TG (1998) Analysis of Spatial Autocorrelation in House Prices. *The Journal of Real Estate Finance and Economics* 17(1): 61–85.

Chandler D and Disney R (2014) Measuring house prices: a comparison of different indices. Report, Institute for Fiscal Studies, UK, May.

Charlton C, Rasbash J, Jones K, et al. (2019) *MLwiN Version 3.03*. Centre for Multilevel Modelling, University of Bristol.

Chi B, Dennett A, Oléron-Evans T, et al. (2019) *Creating a new dataset to analyse house prices in England*. CASA working paper 213.

Cook S (2005) Regional house price behaviour in the UK: application of a joint testing procedure. *Physica A: Statistical Mechanics and its Applications* 345(3): 611–621.

Cook S and Watson D (2016) A new perspective on the ripple effect in the UK housing market: Comovement, cyclical subsamples and alternative indices. *Urban Studies* 53(14): 3048–3062.

De Nadai M and Lepri B (2018) The economic value of neighborhoods: Predicting real estate prices from the urban environment. In: *IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, Italy, 7 August 2018, pp. 323–330

Dennett (2013) "Cartograms!!!" Blog-post. http://adamdennett.co.uk/2013/06/04/cartograms/ (accessed 1/12/2019)

Dong G, Harris R, Jones K, et al. (2015) Multilevel Modelling with Spatial Interaction Effects with Application to an Emerging Land Market in Beijing, China. *PLOS ONE* 10(6): e0130761.

Downes S (2018) 'Location, location, location' still determines house prices. Available at: https://www.whatinvestment.co.uk/location-location-location-is-still-the-most-important-factor-in-house-prices-2614766/ (accessed 2 November 2018).

Edwards M (2016) The housing crisis and London. *City* 20(2): 222–237.

Feng Y and Jones K (2015) Comparing multilevel modelling and artificial neural networks in house price prediction. In: *2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM)*, July 2015, pp. 108–114.

Feng Y and Jones K (2016) Postcode or Census geography? An examination of neighbourhood classifications for house priced predictions. In: *22nd Annual Pacific Rim Real Estate Society Conderence*, Queensland, Australia 17-20 January 2016, pp.1-12.

Ferrari E and Rae A (2013) The spatiality of housing market volatility and selective migration. *Town planning review* 84(1): 107–125.

Finch WH, Bolin JE and Kelley K (2014) *Multilevel Modeling Using R*. 1 edition. Boca Raton, FL: CRC Press.

Fotheringham AS, Charlton ME and Brunsdon C (1998) Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis. *Environment and Planning A: Economy and Space* 30(11): 1905–1927.

Fotheringham AS, Crespo R and Yao J (2015) Geographical and Temporal Weighted Regression (GTWR). *Geographical Analysis* 47(4): 431–452.

Goodman AC and Thibodeau TG (1995) Age-Related Heteroskedasticity in Hedonic House Price Equations. *Journal of Housing Research* 6(1): 25–42.

Goodman AC and Thibodeau TG (1998) Housing Market Segmentation. *Journal of Housing Economics* 7(2): 121–143.

Goodman AC and Thibodeau TG (2003) Housing market segmentation and hedonic prediction accuracy. *Journal of Housing Economics* 12(3): 181–201.

Gray D (2012) District House Price Movements in England and Wales 1997–2007: An Exploratory Spatial Data Analysis Approach. *Urban Studies* 49(7): 1411–1434.

Hamnett C and Reades J (2019) Mind the gap: implications of overseas investment for regional house price divergence in Britain. *Housing Studies* 0(0): 388–406.

Helbich M, Brunauer W, Vaz E, et al. (2014) Spatial Heterogeneity in Hedonic House Price Models: The Case of Austria. *Urban Studies* 51(2): 390–411.

Huang B, Wu B and Barry M (2010) Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science* 24(3): 383–401.

Jones K (1991) Specifying and Estimating Multi-Level Models for Geographical Research. *Transactions of the Institute of British Geographers* 16(2): 148–159.

Jones K and Bullen N (1993) A Multi-level Analysis of the Variations in Domestic Property Prices: Southern England, 1980-87. *Urban Studies* 30(8): 1409–1426.

Jones K and Bullen N (1994) Contextual models of urban house prices: A comparison of fixed- and random-coefficient models developed by expansion. *Economic Geography* 70(3): 252–272.

Kain JF and Quigley JM (1970) Measuring the Value of Housing Quality. *Journal of the American Statistical Association* 65(330): 532–548.

Knoll K, Schularick M and Steger T (2017) No Price Like Home: Global House Prices, 1870-2012. *American Economic Review* 107(2): 331–353.

Law S (2018) *A multi-scale exploration of the relationship between spatial network configuration and housing prices using the hedonic price approach. A Greater London case study*. PhD thesis. University College London.

Leishman C (2009) Spatial Change and the Structure of Urban Housing Sub-markets. *Housing Studies* 24(5): 563–585.

Lu B, Charlton M, Harris P, et al. (2014) Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data. *International Journal of Geographical Information Science* 28(4): 660–681.

MacDonald R and Taylor MP (1993) Regional House Prices in Britain: Long-Run Relationships and Short-Run Dynamics. *Scottish Journal of Political Economy* 40(1): 43–55.

Marsden J (2015) *House prices in London – an economic analysis of London's housing market*. November. Greater London Authority. Available at: https://www.london.gov.uk/sites/default/files/house-prices-in-london.pdf (accessed 21 November 2016).

Meen G (1999) Regional House Prices and the Ripple Effect: A New Interpretation. *Housing Studies* 14(6): 733–753.

Nationwide Building Society (2015) *Nationwide House Price Index Methodoldy*. Available at: https://www.nationwide.co.uk/-/media/MainSite/documents/about/house-price-index/nationwide-hpi-methodology.pdf (accessed 4 November 2018).

Nsubuga J (2018) Amazing transformation of Liverpool streets where houses cost just £1. *Metro*. Available at: http://metro.co.uk/2018/02/03/amazing-transformation-liverpool-streets-houses-cost-just-1-7283928/ (accessed 26 March 2018).

Office for National Statistics, Land Registry, Registers of Scotland and Land & Property, et al. (2016) *Development of a single Official House Price Index*. Available at: https://www.ons.gov.uk/economy/inflationandpriceindices/methodologies/developmentofasingleofficialhousepriceindex (accessed 10 October 2018).

Orford S (2002) Valuing Locational Externalities: A GIS and Multilevel Modelling Approach. *Environment and Planning B: Planning and Design* 29(1): 105–127.

Orford S (2010) Towards a Data-Rich Infrastructure for Housing-Market Research: Deriving Floor-Area Estimates for Individual Properties from Secondary Data Sources. *Environment and Planning B: Planning and Design* 37(2): 248–264.

Powell-Smith A (2017) House prices by square metre in England & Wales. Available at: https://houseprices.anna.ps (accessed 19 November 2018).

Raudenbush SW and Bryk AS (2002) *Hierarchical Linear Models: Applications and Data Analysis Methods* 2nd ed. *Thousand Oaks: Sage Publications*.

Ren Z, Wang J, Liao Y, et al. (2013) Using spatial multilevel regression analysis to assess soil type contextual effects on neural tube defects. *Stochastic Environmental Research and Risk Assessment* 27(7): 1695–1708.

Smith NB (2018) We used big data to predict house prices – and this is what we found. Available at: https://www.architectsjournal.co.uk/opinion/we-used-big-data-to-predict-house-prices-and-this-is-what-we-found/10024129.article (accessed 1 November 2018).

Snijders TAB and Bosker RJ (2011) *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling, second edition. London etc.: Sage Publishers*

Szumilo N, Laszkiewicz E and Fuerst F (2017) The spatial impact of employment centres on housing markets. *Spatial Economic Analysis* 12(4): 472–491.

UN-HABITAT (2011a) Affordable Land and Housing in Asia. Available at: https://unhabitat.org/books/affordable-land-and-housing-in-asia-2/ (accessed 12 April 2018).

UN-HABITAT (2011b) Affordable Land and Housing in Europe and North America. Available at: https://unhabitat.org/books/affordable-land-and-housing-in-europe-and-north-america-2/ (accessed 14 February 2018).

UN-HABITAT (2012) Affordable Land and Housing in Africa. Available at: https://unhabitat.org/books/affordable-land-and-housing-in-africa/ (accessed 12 April 2018).

Usborne S (2018) The story of No 20: how does a London family home end up empty in 2018? *the Guardian*, 20 February. Available at: http://www.theguardian.com/society/2018/feb/20/how-does-london-family-home-end-up-empty (accessed 26 March 2018).

Visser P, Dam FV and Hooimeijer P (2008) Residential Environment and Spatial Variation in House Prices in the Netherlands. *Tijdschrift voor economische en sociale geografie* 99(3): 348–360.

Wang Y, Potoglou D, Orford S, et al. (2015) Bus stop, property price and land value tax: A multilevel hedonic analysis with quantile calibration. *Land Use Policy* 42: 381–391.

Wood R (2015) A comparison of UK residential house price indices. *BIS Papers*, 21(6): 212–227.

Yao J and Fotheringham AS (2016) Local Spatiotemporal Modeling of House Prices: A Mixed Model Approach. *The Professional Geographer* 68(2): 189–201.

Yu D (2007) Modeling Owner-Occupied Single-Family House Values in the City of Milwaukee: A Geographically Weighted Regression Approach. *GIScience & Remote Sensing* 44(3): 267–282.

Supplementary Material A

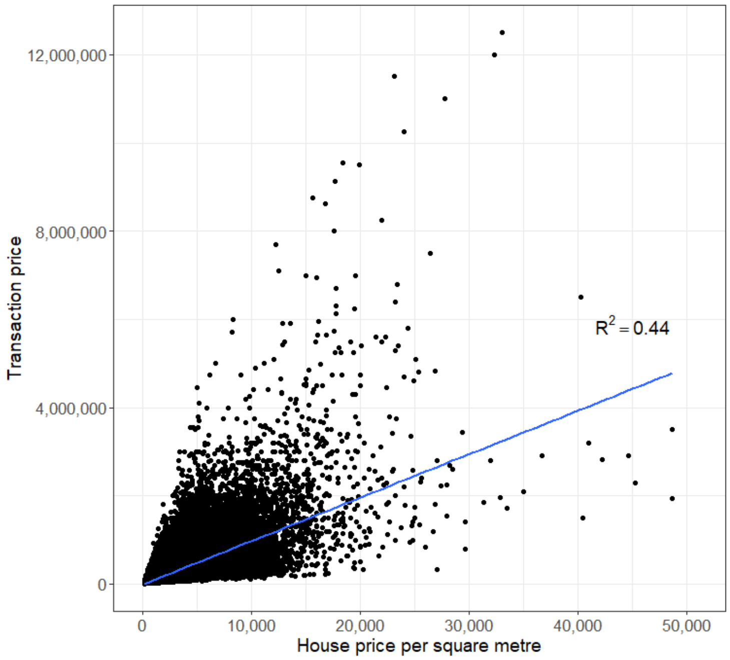


Figure A. Scatter plot of transaction price and house price per square metre in England, 2009

Supplementary Material B

Table B VPC results for TP2009 and HP20009 compared with results from corresponding models using log prices

Level	TP2009	Log version of TP2009	HP2009	Log version of HP2009
Local authority level	0.23	0.35	0.53	0.52
MSOA level	0.12	0.16	0.10	0.12
LSOA level	0.10	0.10	0.04	0.05
Individual level	0.56	0.39	0.33	0.31

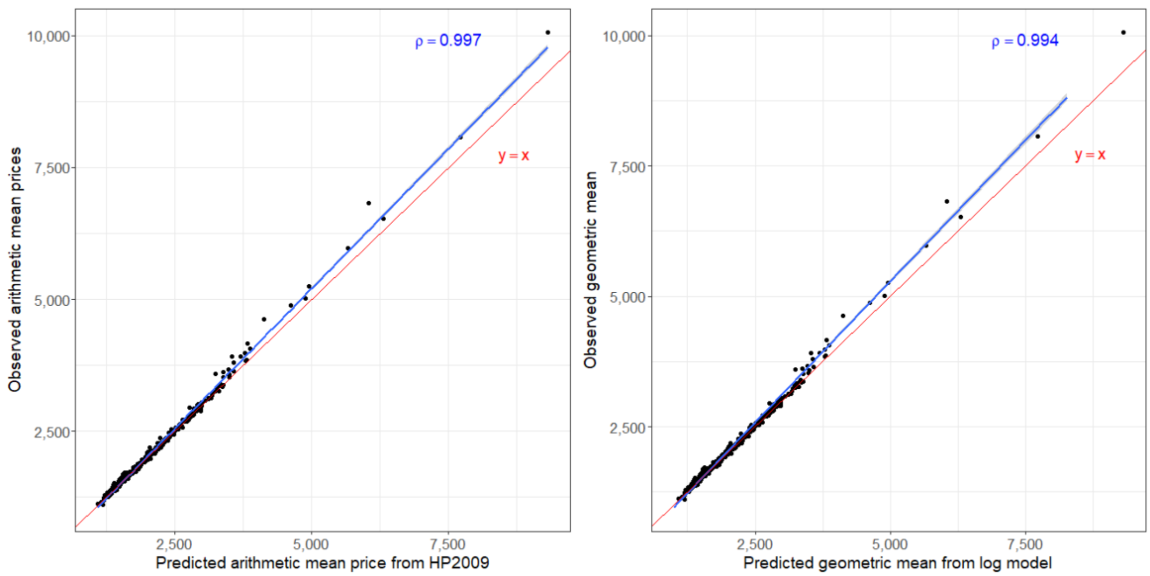


Figure B. the relationship between estimated mean house price and observed mean house price from HP2009 and a corresponding model using log prices.

Supplementary Material C

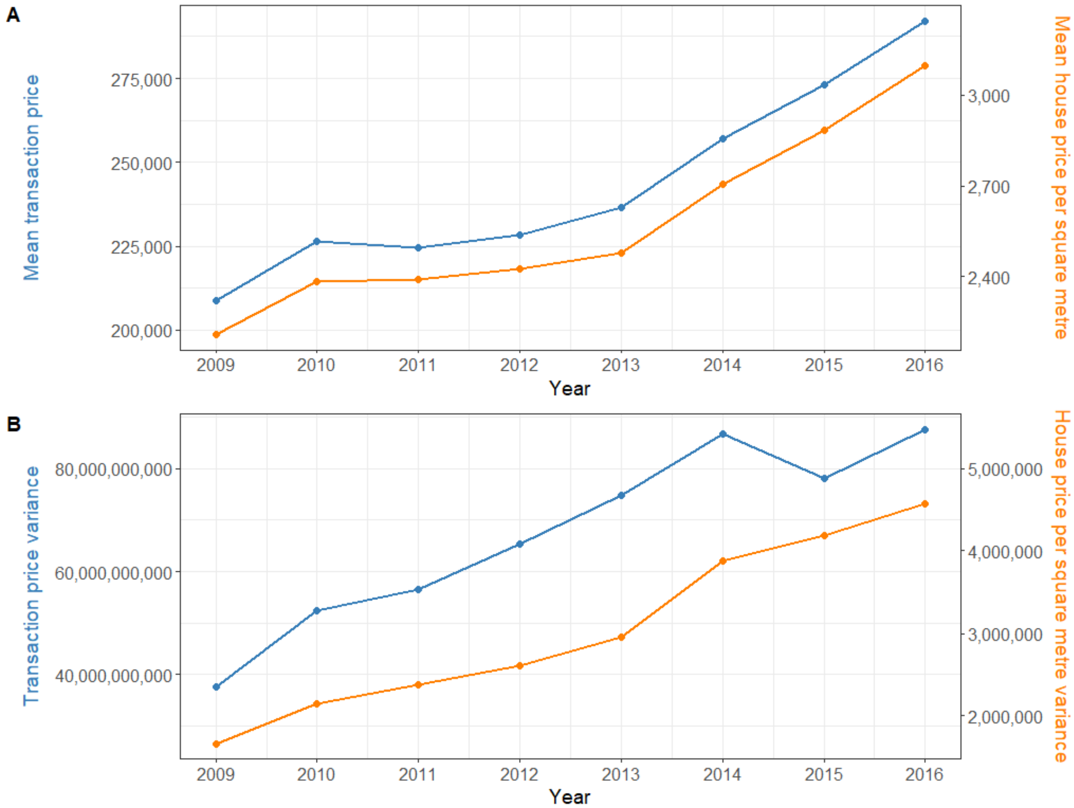


Figure C1. Change of overall mean house price change and house price variance

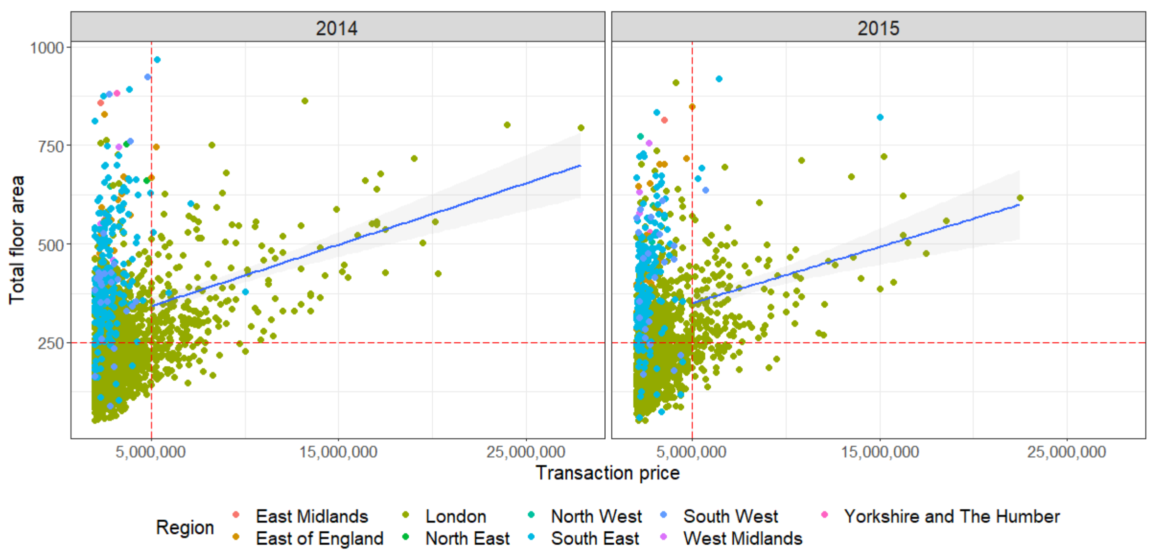


Figure C2. Scatter plots of transaction price against total floor area in 2014 and 2015

Supplementary Material D

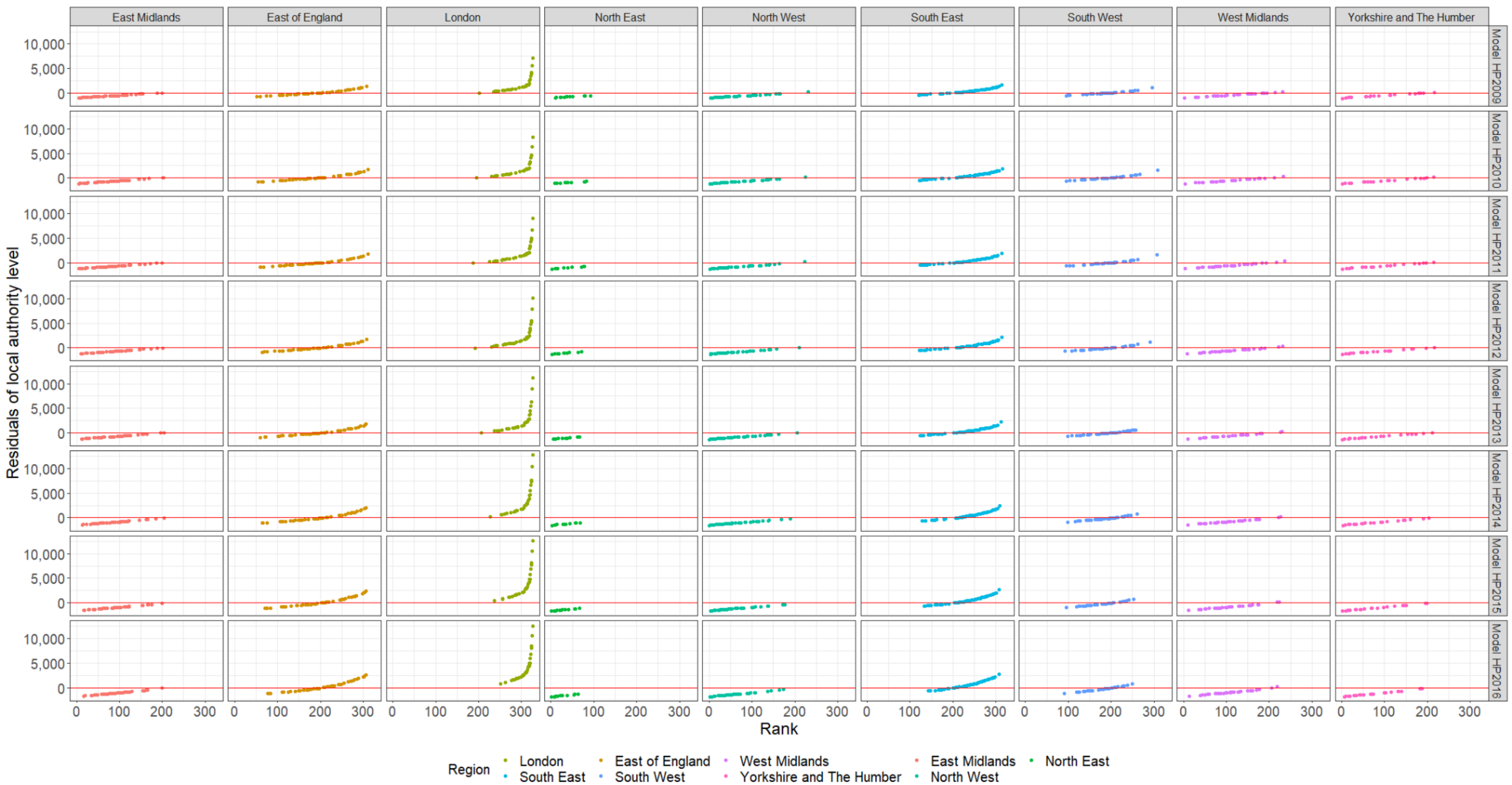


Figure D. Local authority level residuals by region in England

Supplementary Material E

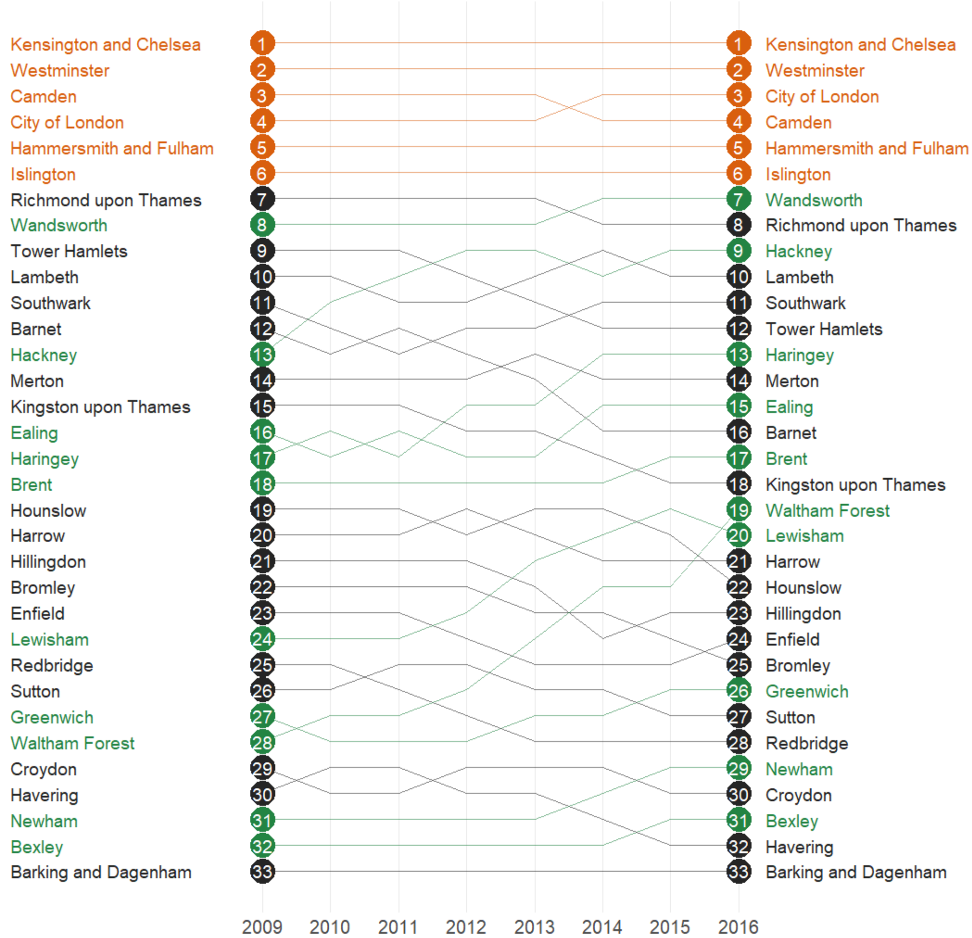


Figure E. Ranks of local authority's residual in London, from 2009 to 2016

Supplementary Material F

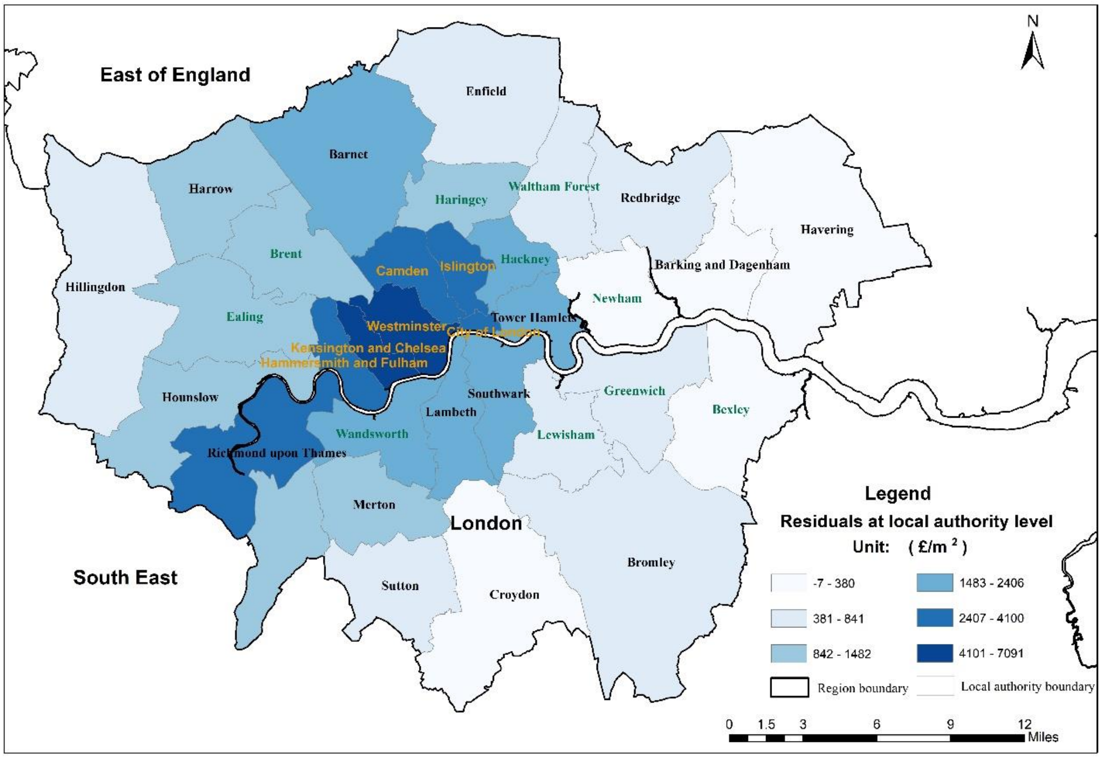


Figure F. Residuals at local authority level in London, 2009

Supplementary Material G

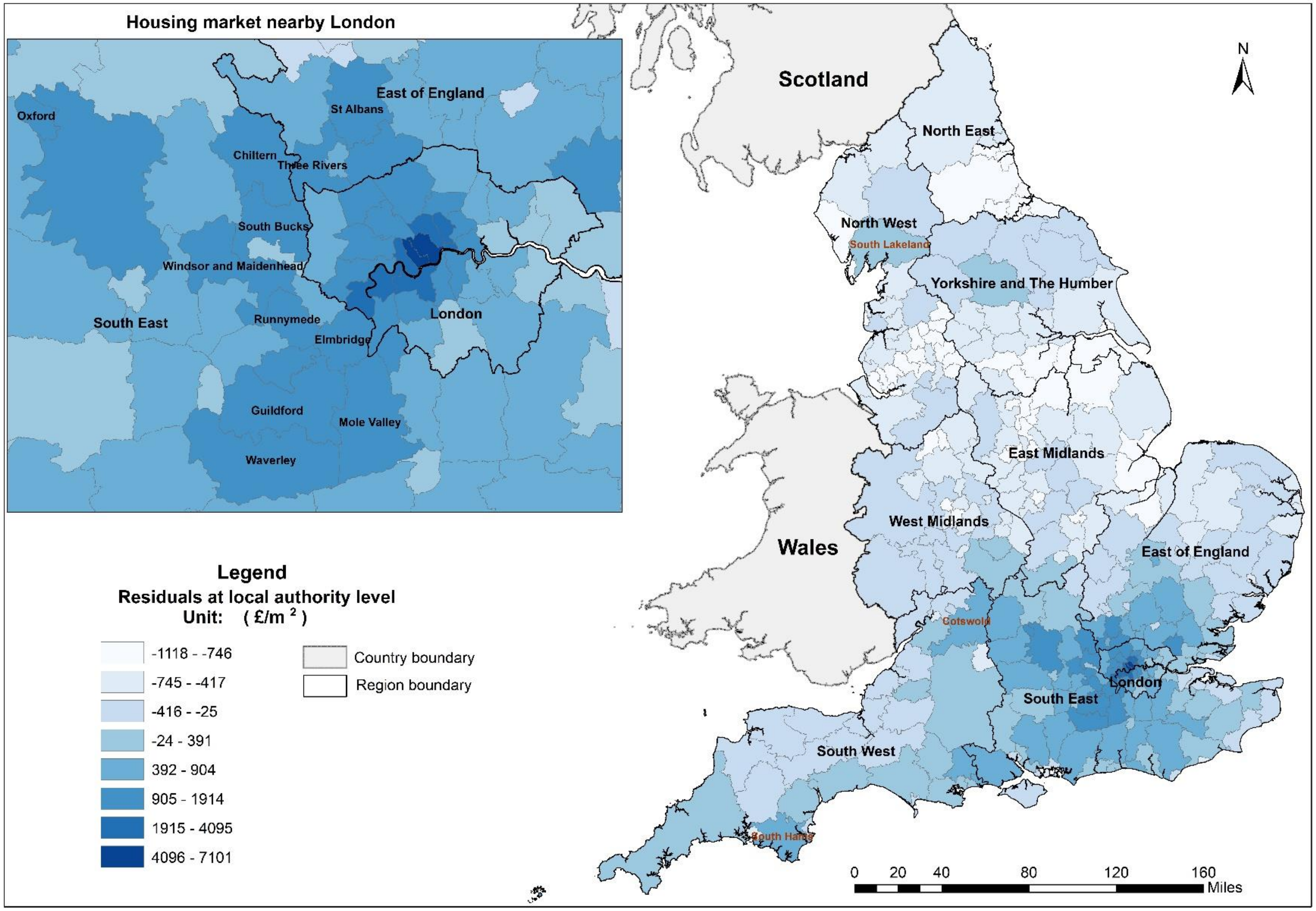


Figure G. House price difference at local authority level in England, 2009