UCLPRESS

Authors: Bin Chi[1], Adam Dennett[2], Thomas Oléron-Evans[3], Robin Morphet[4]
Affiliations: Centre for Advanced Spatial Analysis, University College London, 90 Tottenham Court Road, London W1T 4TJ, UK.[1]
Orcid ids: 0000-0002-5815-7428[1], 0000-0001-8036-0185[2], 0000-0003-1364-3606[3], 0000-0002-8203-8429[4]
Contact e-mail: bin.chi.16@ucl.ac.uk

Dear editors,

We hope to submit a manuscript titled "A new attribute-linked residential property price dataset for England and Wales, 2011 to 2019" for consideration by *"UCL Open: Environment"*.

In this article, we outlined one approach to overcame the house price data deficiencies in the UK by enriching transaction information from the official open Land Registry Price Paid Data (LR-PPD) and property size information from the official open Domestic Energy Performance Certificates (EPCs). This research contributes to the literature in four main ways: Firstly, this new linked data provides information of house price per square meter (HPM) in England and Wales, which could offer a clearer picture of the house price differences in these countries through normalising the transaction price by total floor area. Secondly, it covers a wider range of datasets that contains transaction price and total floor area with the biggest sample size. This offers greater flexibility for the exploration of house price ($£/m^2$) variation in England and Wales at different scales over postcode units between 2011 and 2019. Third, the linked data also contains the property energy information. This particular piece of information could support research on house price and property energy. Last but not least, as the LR-PPD is updated monthly and the EPCs are updated two or four times a year, the open access codes will allow future updates.

The linked, original EPCs and LR-PPD datasets and linkage codes are deposited and published in UK Data Service's ReShare (https://reshare.ukdataservice.ac.uk/854240/), which are open access data under a licence that precludes commercial use. The licence for each dataset is introduced along with the data in ReShare. UKDA can give reviewers the access to the data files if they e-mail to reshare@ukdataservice.ac.uk and mention the record number (854240) in their e-mail.

We confirm that this manuscript has not been published and is not under consideration for publication elsewhere.

Please let us know if you need any further information. Many Thank for your consideration of this manuscript.

Best wishes,

Bin Chi

bin.chi.16@ucl.ac.uk

UCL Centre for Advanced Spatial Analysis

90 Tottenham Court Road

London W1T 4TJ, UK

A new attribute-linked residential property price dataset for England and Wales, 2011 to 2019

Bin Chi[1], Adam Dennett[1], Thomas Oléron-Evans[1], Robin Morphet[1]

[1] Centre for Advanced Spatial Analysis (CASA), University College London, UK

Abstract

Current research on residential house price variation in the UK is limited by the lack of an open and comprehensive house price database that contains both transaction price alongside dwelling attributes such as size. This research outlines one approach which addresses this deficiency in England and Wales through combining transaction information from the official open Land Registry Price Paid Data (LR-PPD) and property size information from the official open Domestic Energy Performance Certificates (EPCs). A four-stage data linkage is created to generate a new linked dataset, representing 79% of the full market sales in the LR-PPD. This new linked dataset offers greater flexibility for the exploration of house price ($£/m^2$) variation in England and Wales at different scales over postcode units between 2011 and 2019. Open access linkage codes will allow for future updates beyond 2019.

Keywords: Land Registry Price Paid Data, Domestic Energy Performance Certificates, data linkage, England and Wales.

1.INTRODUCTION

Comparative international analyses of house prices are constrained by differences in definition, data structure, spatial/time scales and coverage. These limit both comparative analysis and within-country analysis of housing markets (Cannari and Faiella, 2008; Ciarlone, 2015). House price data deficiencies hinder research on residential house price variation in the UK, and limit understanding of the housing market. Modelling of UK-based house price changes dates back to the 1970s (Ball, 1973; McAvinchey and Maclennan, 1982) with much of the data used either aggregated to coarse geographies such as regions or districts or, conversely, associated with individual properties in a specific city. Aggregate sample mortgage data, mainly from building societies, such as the 5% sample survey of Building Society Mortgages and the Nationwide Building Society mortgage data, have been widely used(Ahlfeldt et al., 2012; Alexander and Barrow, 1994; Cook, 2003; Fleming and Nellis, 1992; Gray, 2015; Hudson et al., 2018; Law, 2018; Rosenthal, 1986; Wabe, 1971). These datasets lack local nuance but are also problematic due to the potential biases inherent in small samples(Hamnett, 1983; Jones and Bullen, 1993). Conversely, more detailed micro-level housing data such as the local estate agent survey data used by Orford(Orford, 2000) have offered opportunities for local housing analysis, but such datasets are not widely available.

Land Registry Price Paid Data (LR-PPD) have been published as open data since 2013. These data have been transformative for house price variation research in the UK(Cooper et al., 2013; Fingleton, 2006; Gray, 2012; Szumilo et al., 2017) as they are a comprehensive record of residential transactions at address level in England and Wales dating back to 1995(HM Land Registry, 2016). Although the Land Registry excludes some types of residential property sales (e.g. 'Right to buy' sales at a discount), these data still provide the most accurate picture of residential property sales at full market value in England and Wales(Marsden, 2015). The Office for National Statistics (ONS) has used the LR-PPD to calculate official house price statistics such as the House Price Statistics for Small Areas dataset (South and Henretty, 2017) and the official House Price Index(Office for National Statistics et al., 2016). Despite the utility of these data a lack of attribute information relating to the properties, such as total floor size information, is identified as one of the major shortcomings, as the impacts of stock mix on broader patterns cannot be fully accounted for (Law, 2018; Orford, 2010).

Since total floor area is identified as the most important determinant of house price variation (De Nadai and Lepri, 2018; Orford, 2010; Sirmans et al., 2006; Thwaites and Wood, 2005), two approaches have been developed in the UK to enhance the LR-PPD by adding total floor area. The first approach, created by Orford(Orford, 2010), adds an estimated total floor area to the LR-PPD. The estimated total floor area is derived from building footprints obtained from Ordnance Survey MasterMap and Environment Agency LiDAR data, but the methods are recognised as problematic for estimating the floor area of flats within a building(Orford, 2010), and for properties where the number of stories cannot be accurately inferred.

The second approach is more direct and links LR-PPD with the total floor area information from Domestic Energy Performance Certificates(Fuerst et al., 2016, 2015, 2013; Powell-Smith, 2017; Simpson et al., 2018). Domestic Energy Performance Certificates (Domestic EPCs) is an open dataset released by the Department for Communities and Local Government (DCLG). It not only records a property's energy performance but also gives building attribute information (i.e. total floor area or number of habitable rooms). Despite this link being feasible, only two research studies have mentioned the linkage rate between LR-PPD and Domestic EPCs and no research has yet published the details of both the linkage method and linkage data(Powell-Smith, 2017; Simpson et al., 2018). Aiming to remedy this situation, we publish our own linkage codes alongside the open access and reusable house price per square metre dataset.

2. DATA DESCRIPTION AND DEVELOPMENT

2.1 LR-PPD and Domestic EPCs data
The LR-PPD dataset is open, available online and updated on a monthly basis (https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads). The LR-PPD used in this research was downloaded in 2019, which contains 16 items with 24,852,949 transactions in England and Wales between 1/1/1995 and 31/10/2019. For each transaction, there is a unique transaction identifier along with the property's transaction price, transaction date, address information (postode, PAON, SAON,street), property type (detached, semi-detached, terraced houses or flats/maisonettes), whether a property is newly built or whether it was sold at full market value (HM Land Registry, 2016) . For various reasons, not all the

properties within the dataset are sold at full market value, therefore these entries are excluded from the linkage exercise. These excluded entries comprise only 2.90% of the whole dataset.

EPCs have been required by law since 2008 for all properties sold, built or rented in England and Wales. Data from these certificates is open and available on-line from the Ministry for Housing, Communities and Local Government - MHCLG (https://epc.opendatacommunities.org/). The EPC dataset used in this research is the third version downloaded 20/10/2020 and contains certificates issued between 1/10/2008 and 31/5/2019 (MHCLG, 2019). It records 18,575,357 energy performance data records with 84 fields. It not only records a property's energy performance but also building stock information, such as its address, total floor area and number of habitable rooms.

2.2 Data linkage

Linkage between the PPD and Domestic EPC dataset is achieved though several phases dealing with successively more complex address matching challenges. Before matching, transactions in the LR-PPD without postcodes in the Domestic EPCs dataset are excluded – this accounts for 0.55% of the data - leaving a total of 23,999,656 transactions for matching. Figure 1 shows an example of the data linkage process, with the basic idea of linkage between these two datasets being to match by full postal delivery address (i.e. postcode plus detailed address strings). These two datasets both contain property information at address level but their address structures differ, thus basic data standardization is needed. First, all address strings in the Domestic EPCs are capitalised and stored in new variables. These newly created address variables are used to achieve an initial data linkage. To deal with more complex subsequent linkage passes, 180 new variables are created in the LR-PPD and 95 new variables are created in the Domestic EPCs.
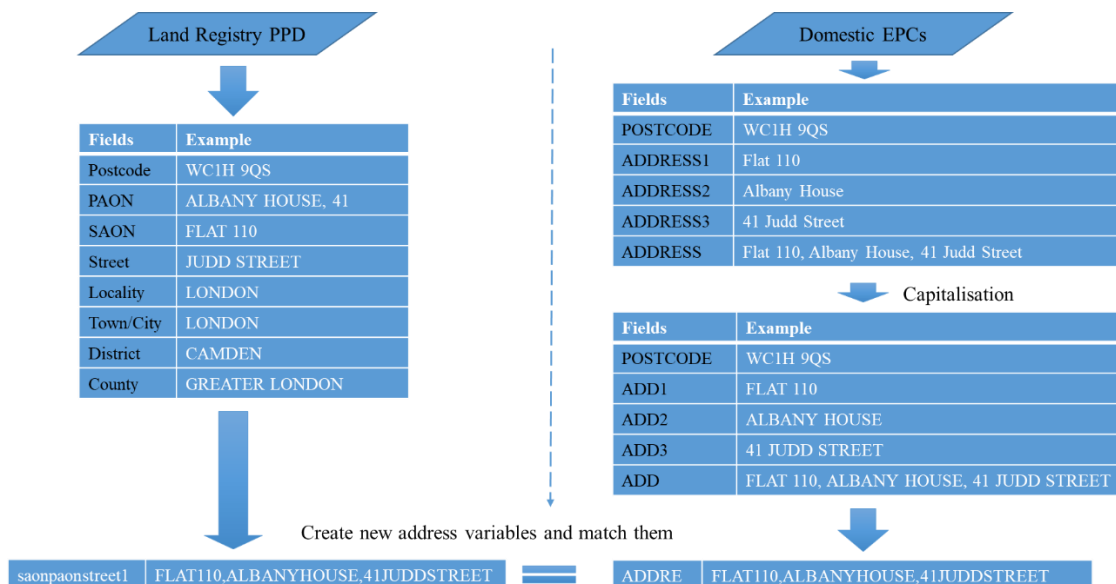

Figure 1 An example of data linkage process

A matching method containing a four-stage (251 matching rules) process was designed and is outlined in Figure 2. In the Domestic EPCs, each record is created using a unique identifier

named id. Each transaction in the LR-PPD has a unique identifier named transactionid. Taking Stage 1 as an example of the matching process; all the matches are based on a temporary address string (i.e. postcode+saonpaonstreet) with the algorithm testing whether postcode+saonpaonstreet in LR-PPD is equal to any postcode +ADDRE in Domestic EPCs. Where they match directly, records for both datasets are joined, removed from the original data and stored in a new temporary linked data table, DATA 1. For records where a match is not achieved on the first pass, the algorithm moves onto a further set of matching tests in Stage 2.

Problems emerge where one property may have more than one Domestic EPC. Where this is the case, only property transactions with just one successfully linked EPC will be moved from the temporary DATA 1 and directly stored in the final linked-EPC PPD dataset. Property transactions with successful links to more than one EPC are stored in a separate dataset, DATA 3. These data are filtered to select all Domestic EPCs for which total floor area is neither NULL nor 0 and then linked where the EPC inspection date or lodgement date is closest to the transaction date in the LR-PPD. This result will then be stored in the final linked-EPC PPD dataset. Stages 2 to 4 follow a similar process to Stage 1. The linked-EPC PPD dataset is the data linkage result. These data linkage results link back to the original Domestic EPCs and to the LR-PPD by their unique identifiers.
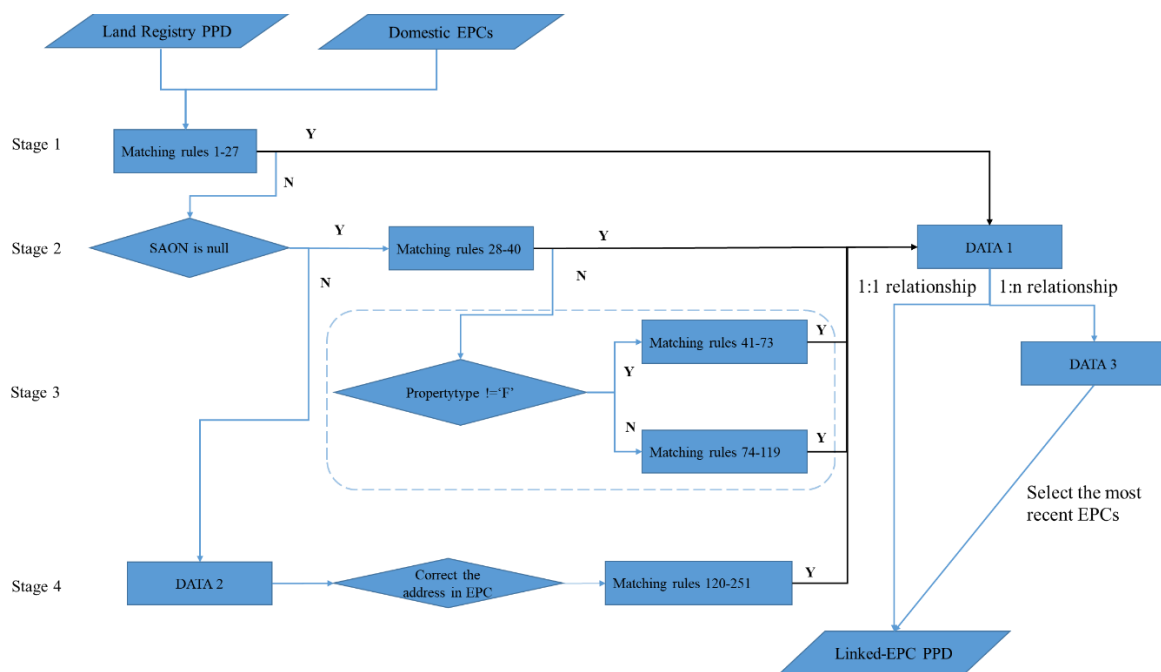

Figure 2 Workflow of the four-stage data linkage between LR-PPD and Domestic EPCs

Following the four-stage data linkage, 16,846,834 transaction records in England and Wales between 1995 and 2019 were successfully linked with Domestic EPCs. These comprise the linked dataset. The match rate of transactions in England is shown in Figure 3. The match rate between 2011 and 2019 is higher than 90%, while the match rate of the rest of the period is considerably lower, this is mainly due to the EPCs dataset only covering the period between 1/10/2008 and 31/8/2019. The match rate of 56.20% in 2008 is particularly low but rapidly increases to over 88% after 2010. Since the match rate before 2008 is significantly lower than for the period after 2008, only the linked data between 2009 and 2019 are used to conduct the
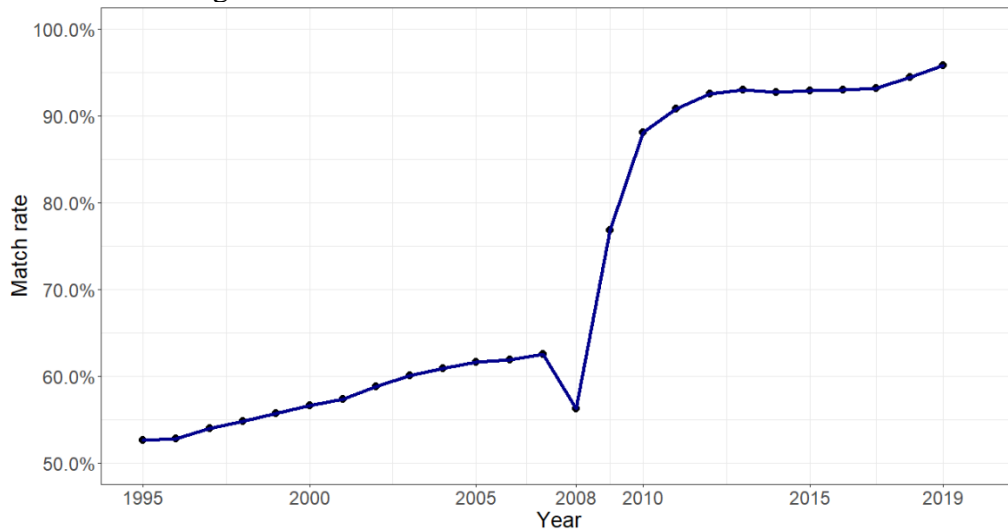
evaluation of data linkage.

Figure 3 Match rate of linked house price in England and Wales,1995-2019

2.3 Technical Validation
(1) Evaluation of the data linkage between 2009 and 2019
Match rates offer a crude way to quantify the matching performance, but visual comparison of the house price frequency distributions for the new linked data and original LR-PPD data reveals a clearer picture of matching performance. Histograms of the logarithm of transaction price from both datasets are shown in Figure 4. In each graph, the distribution of the linked data (blue) is overlaid onto the distribution of the original LR-PPD dataset(white). The area of visible white bars represents the proportion of un-matched cases. Importantly, there was no significant loss of information as a result of un-matched cases in the data linkage between 2010 to 2019.
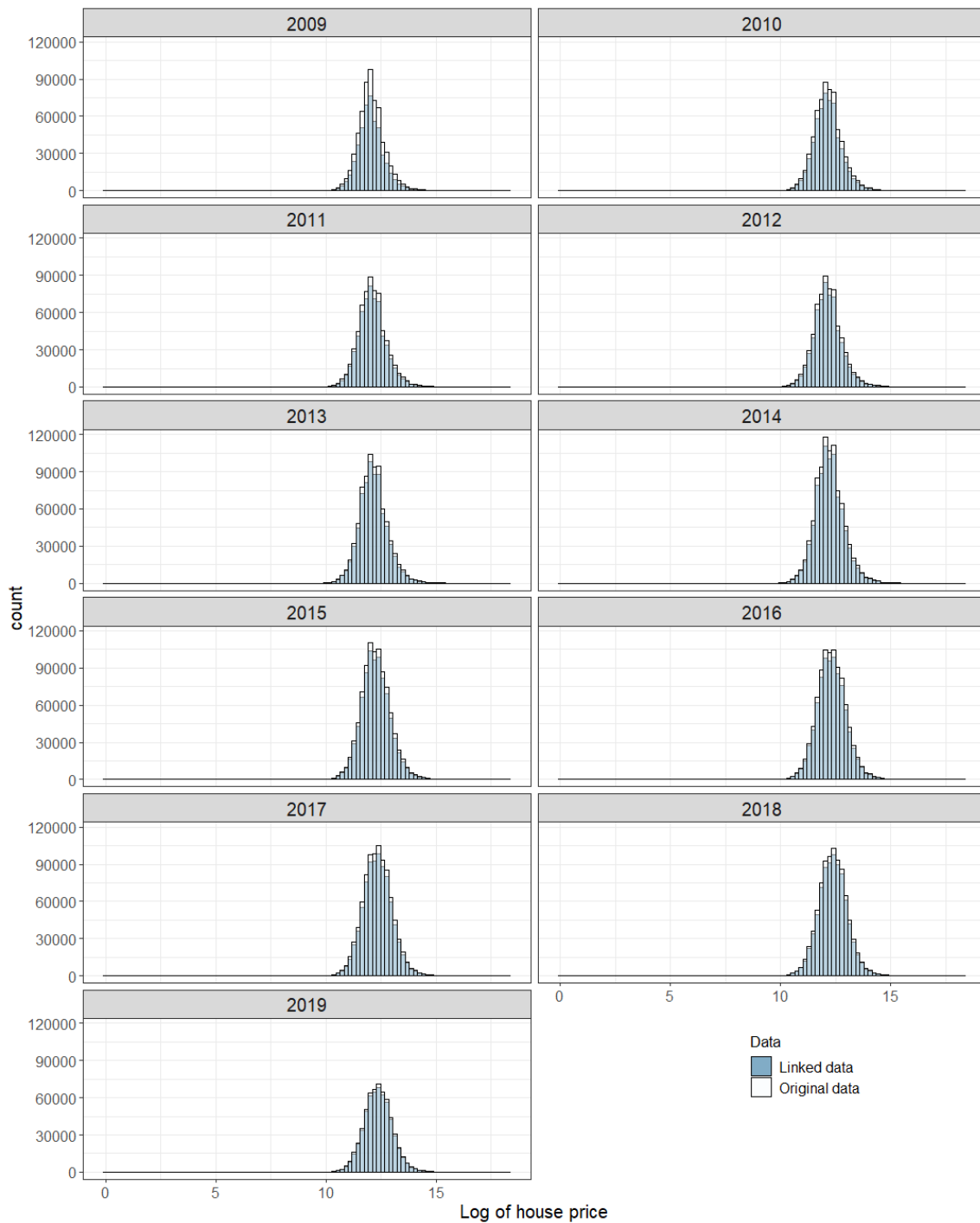
Figure 4 House price distribution of original data and linked data, 2009-2019

The Kolmogorov–Smirnov test (K-S test) and the Jeffreys divergence (J-divergence) can be used to quantify the extent of house price information lost. The Kolmogorov–Smirnov (K-S) test is a nonparametric test that examines the differences in the shape of a distribution. The K-S test statistic, D, is based on the maximum absolute difference between two cumulative distribution functions. Here, the test will be used to quantify the difference of two house price distributions (original data versus linked data). The Jeffreys divergence (J-divergence), derived from information theory, is a function used to establish the distance of one probability distribution to another (Jeffreys, 1946; Nielsen, 2010; Rohde, 2016). To calculate the J-

divergence, the data from two different samples must first be assigned to k different categories. In the case of this research, these categories are a simple subdivision of the log house price into bins. The J-divergence is then defined as

$$J = \sum_{j=1}^{k} p^j \ln\left(\frac{p^j}{q^i}\right) + \sum_{j=1}^{k} q^j \ln\left(\frac{q^j}{p^i}\right) \quad (1)$$

where k is the number of categories, p^j is the proportion of data points in category j in the original house price data, and q^j is the proportion of data points in category j in the linked house price data. The final divergence measure, J, ranges from 0 to 1. If the distribution of both data samples across all the categories is the same, J will be 0. Larger values of J indicate greater differences between the two distributions.

To compute the J-divergence, the original data and linked data are divided into 100 bins, the 100 bins are created based on the 100 equal intervals of log house price in the original data in a given year. The results of the J-divergence and K-S tests are shown in Figure 5. p-values of all the K-S tests are less than 0.05 (the conventional default threshold for statistical significance), indicating a statistically significant difference between the original house price data and the linked house price data. The D statistic is relatively low (less than 0.007) after 2010. This demonstrates that the house price datasets before and after linkage are highly similar after 2009. The J-divergence results also show that the linked data exhibits relatively low information loss after 2009. Given the information lost in terms of J-divergence is slightly higher in 2010 compared to the loss after 2010, the newly created house price data from 2011 to 2019 is more representative than that for other years. Therefore we keep the 2011 to 2019 time period.
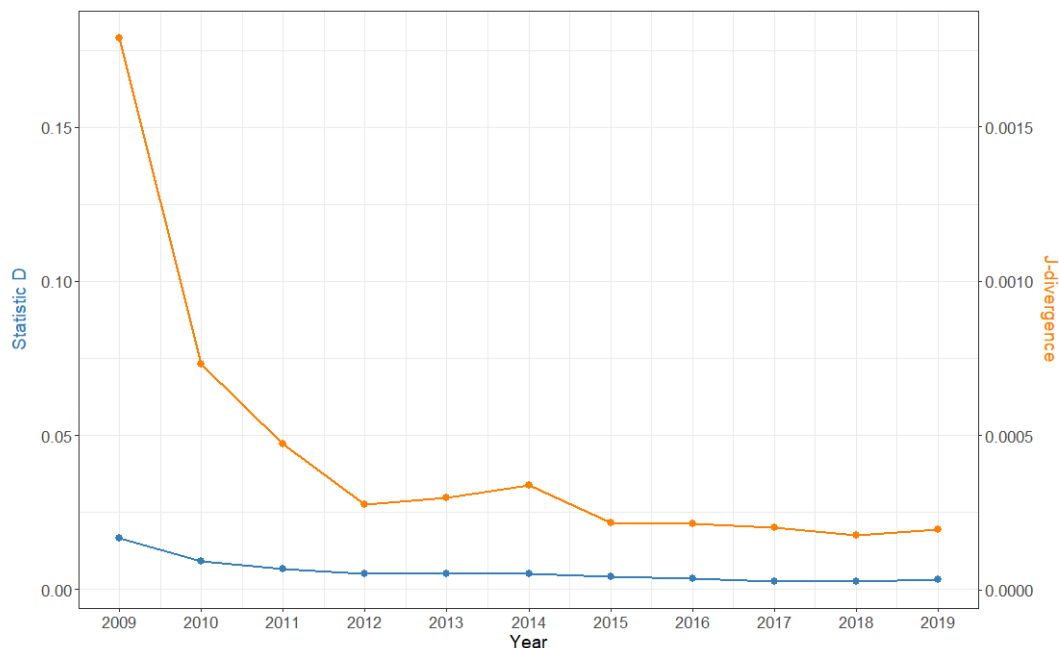

Figure 5 Results of K-S test and J-divergence method

(2) Linked dataset between 2011 and 2019

There were 7,249,259 full market value transactions in England and Wales between 1/1/2011 and 31/10/2019. 6,753,335 of these have been successfully linked to EPC records. The overall match rate for this period is 93.15%. To support more advanced understanding of match rate

spatially, the National Statistics Postcode Lookup -NSPL (November 2019 version) is used to geo-reference both the linked data and orgianl pre-linked LR-PPD by postcode to 2011 Census Output Area (OA) code, Lower Layer Super Output Area (LSOA) code and Middle Layer Super Output Area (MSOA) code (South and Henretty, 2017). Then the ONS hiericerchiacal lookup table (Office for National Statistics, 2019) is used to relate the OAs with Local authorities (LAs) and Regions information. 28 linked transactions and 3001 transactions in LR-PPD were lost during this process.

With the geo-referenced data, the overall match rates between 2011 and 2019 by LA (Figure 6) are not equally distributed. 92% of LAs in England and Wales have a match rate over 90%. Only two LAs(City of London and Isles of Scilly) have a match rate under 80%, these are 71.65% and 76.65% respectively. The remaining 8% of LAs (26 LAs) show a match rate between 80% and 89.81%.
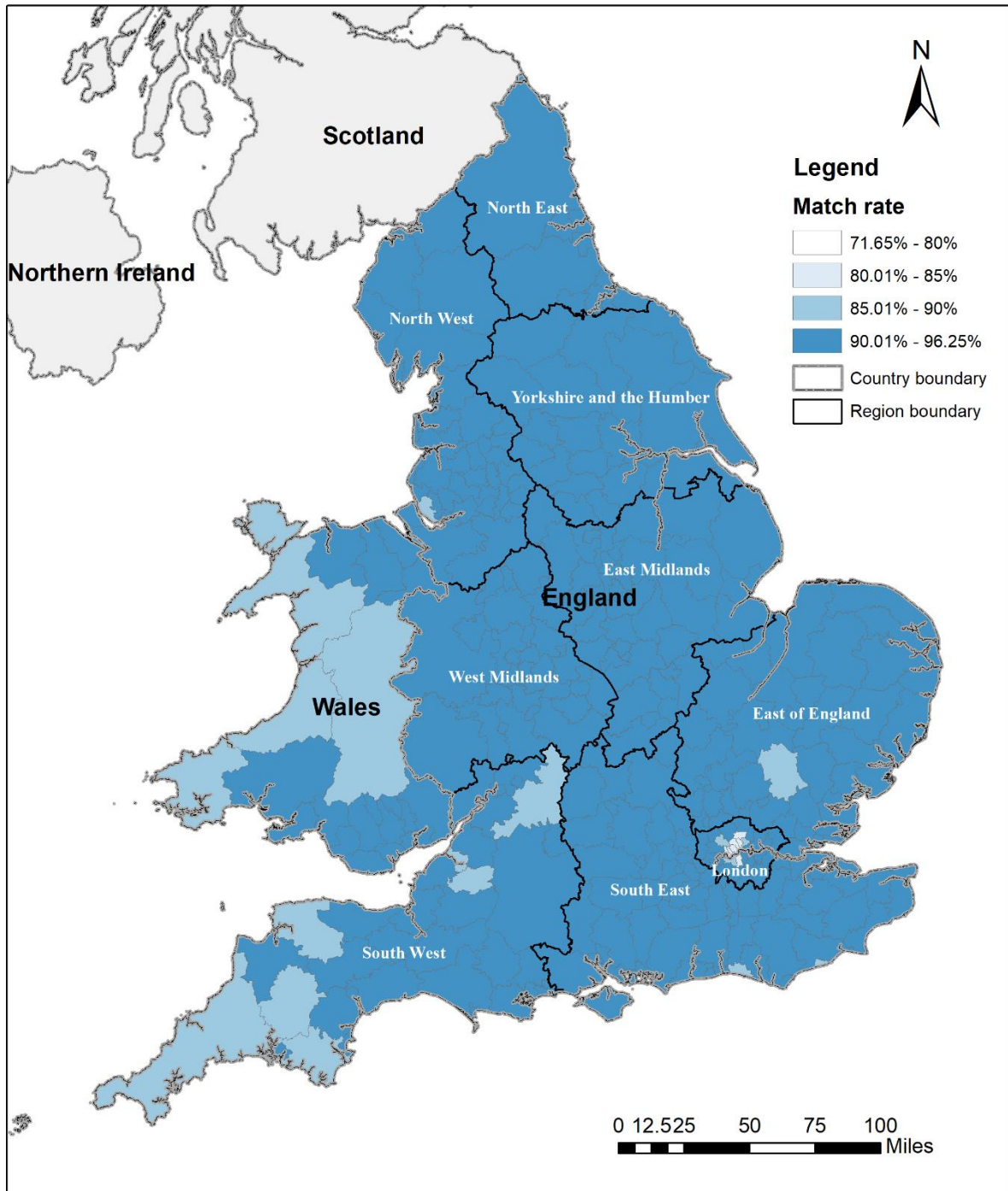
Figure 6 Overall match rate at local authority level between 2011 to 2019

Looking at annual match rates across LAs in England and Wales (Figure 7), 70% of LAs represent an annual match rate over 90% from 2011 to 2019, while 98% of the LAs represent annual match rate over 80%. Figure 7 colours the six LAs with annual match rates lower than 80%. They are Isles of Scilly, City of London, Camden, Hammersmith and Fulham, Kensington and Chelsea and Westminster. Only two LAs (City of London and Isles of Scilly), both of which are small in terms of their numbers of transactions, show an obvious fluctuation during this 9-year period. The rates between 2011 and 2019 are, for the remaining 346 LAs, quite stable over time with a slight fall after 2015. Overall, the majority of LAs with a high match rate in 2011 maintained a high rate subsequently.
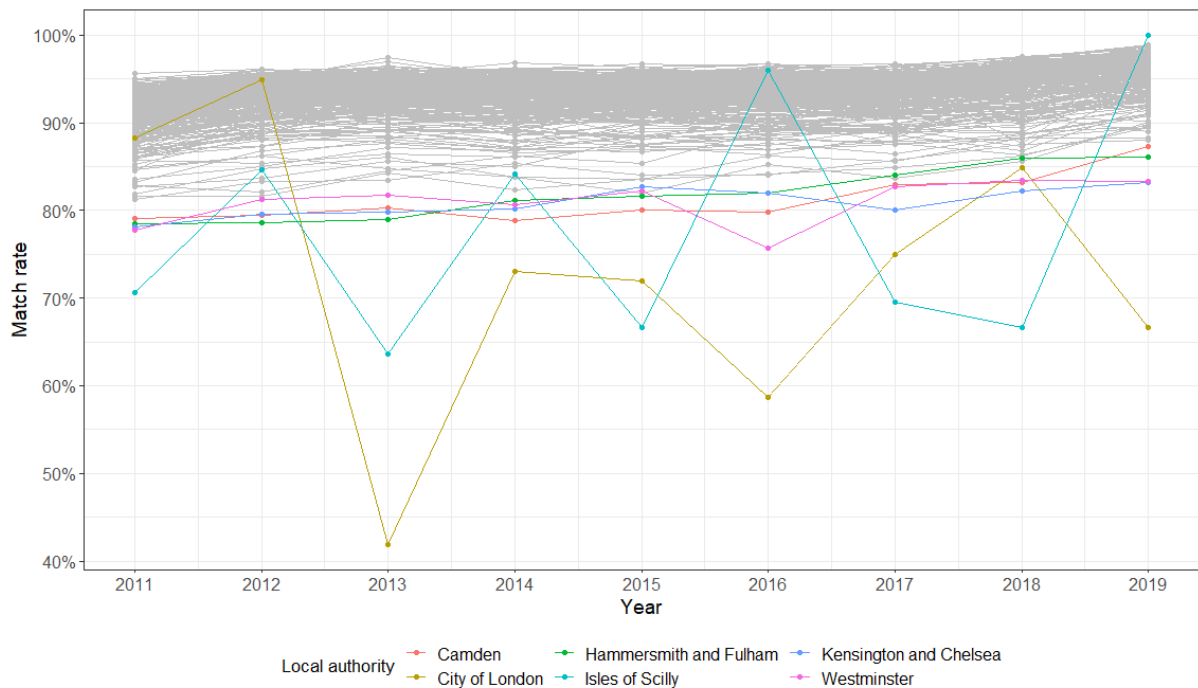
Figure 7 Match rate across local authority in England and Wales, 2011-2019

Properties that feature in the LR-PPD (1/1/2011-31/10/2019) are not fully available in Domestic EPCs (1/10/2008-31/8/2019), this is the main reason for unequal match rates across local authorities. For 18,980 transactions (2011-2019) relating to 6,375 postcode units Domestic EPCs cannot be found. For example, Domestic EPCs in the City of London at postcode "EC2Y 9BB" are not available hence transactions in "EC2Y 9BB" cannot be successfully matched, 0.26% of house price transactions in the LR-PPD (1/1/2011-31/10/2019) fail to link for this reason. Some transactions in the LR-PPD can relate to a postcode unit which is also identified in the EPC data but contain no matching property identifiers. For example, one flat sold in 2011 in Camden failed to match because Domestic EPCs are not available for this property. This potential reasons for non-availability of property records in Domestic EPCs could be that records have been incorrectly loaded by the surveyor or that the property owner has opted out.

2.4 Data cleaning

6,753,307 records of linked data can be geo-referenced by linking the NSPL between 1/1/2011 and 31/10/2019 in England and Wales. This data comprises the transaction information in the LR-PPD together with property size (total floor area and number of habitable rooms) in EPCs. Some properties' total floor area and number of habitable rooms are recorded in EPCs with missing or untenable values (e.g. total floor area records as 0.01). This data is excluded prior to analysis. All the excluded transactions along with cleaning methods are listed in Table 1, which accounts for 15.11% of the linked geo-referenced data.

Table 1 List of transactions exclude from the linked geo-referenced data

No.	Method	Transaction numbers	Proportion
1	Transaction's total floor area or number of habitable rooms are NA value or 0.	1,016,247	99.59%

2	Transaction's total floor area is small than 9 m^2 or bigger than 974 m^2.	555	0.05%
3	Transaction's price per total floor area is bigger than 50000 £/m^2 or transaction price per total floor area is small than 200 £/m^2 .	841	0.08%
4	Transaction's floor size per habitable room is bigger than 100 m^2.	887	0.09%
5	Transaction's number of habitable rooms are bigger than 20.	476	0.05%
6	Transaction's floor size per habitable room is smaller than 6.51 m^2.	1,463	0.14%
Overall		1,020,469	100%

After removing the transactions listed in the table above, 5,732,838 transactions are left. This represents 79.11% of full market property sales in the LR-PPD in England and Wales between 1/1/2011 and 31/10/2019. This linked dataset, like the LR-PPD, fully covers all the regional areas, local authorities and MSOAs in England and Wales. The LR-PPD covers 99.99% of LSOAs and this is also the same for the final linked data. Although the newly linked data is not as comprehensive as the LR-PPD, it is the largest open access house price dataset in England and Wales (1/1/2011-31/10/2019) containing both the transaction price and total floor area.

3. DATASET ACCESS

The final linked dataset details 5,732,838 transactions in England and Wales (1/1/2011 - 31/10/2019). It not only adds in a property's total floor area and the number of habitable rooms, but also includes a new unique identifier (i.e. id) and other non-address fields (except LMK_KEY field) in the Domestic EPC dataset. Codes for other commonly used spatial units from Output Area (OA) to region are also included in this dataset. It contains 105 fields written in upper or lower case. All the fields written in upper case come from Domestic EPCs, the 33 remaining fields written in lower case are introduced in Github (https://github.com/BINCHI1990/Link-LR-PPD-and-Domestic-EPCs).

The linked, original EPCs and LR-PPD datasets are stored in CSV format and deposited in UKDA ReShare (Chi et al., 2020a). Postcode and address elements in the linked data stem from address information in LR-PPD, which is subject to Royal Mail copyright. The Royal Mail confirmed on 25/8/2020 that this linked data can be shared both by the first author and by the UK Data Service on the same terms as the original datasets. Therefore the linked data is under a licence that precludes commercial use. Meanwhile, the data linkage is conducted in R and stored in PostGIS. They are also open available in the UKDA ReShare under the same lincence (Chi et al., 2020a).

4. POTENTIAL DATASET USE AND REUSE

The newly linked dataset offers directly useable information on house price per square meter along with transaction price, total floor area, number of habitable rooms, transaction date and commonly used geographical area identifiers at and over postcode geographic level in England and Wales. Since the LR-PPD data for the most recent two months may be incomplete due to the delay between the property transaction and its registration in Land Registry(HM Land Registry, 2016), we suggest researchers use transactions before 31/8/2019. This could support quantitative house price research in terms of house price variation within England and Wales after 2011 at multi geographical scales over postcode level (Chi et al., 2020b). It also can be used to explore the relationship between house price and a property's energy performance (Department of Energy & Climate Change, 2013; Fuerst et al., 2016, 2015). In addition, since the LR-PPD is updated monthly and the EPCs are updated two or four times a year, the open access codes will allow for future updates and thereby maintain a continuously updated dataset of residential property prices in England and Wales.

5. CONCLUSIONS

The linkage method was orignally created to enrich the geo-referenced house price dataset in England before 31/7/2017 (Chi et al., 2019), it still shows a smiliar performance when updated with new published house prices and covers Wales as shown in this research. Within the linkage, properties in the LR-PPD and Domestic EPC dataset have slightly different names (e.g. 'CLEATOR STREET' VS 'CLEATER STREET'). We manually correct this type of mismatched address string for the properties located in England and record this correction within the linkage codes. This contributes to a less than 1% increase in the total matching rate. Our futher linkage research is to focus on fixing this issue in Wales and for the newly updated transactions in England.

We expect that this new dataset will enable new research directions in UK housing analysis. To date, most hedonic house price models have had to contend with the confounding influence of variations in dwelling size in different housing market areas. This new dataset will enable more parsimonious models of price variation to be explored where proxies for size can be dispensed with.

REFERENCES

Ahlfeldt, G.M., Holman, N., Wendland, N., 2012. An assessment of the effects of conservation areas on value. English Heritage.

Alexander, C., Barrow, M., 1994. Seasonality and Cointegration of Regional House Prices in the UK. Urban Studies 31, 1667–1689.

Ball, M.J., 1973. Recent Empirical Work on the Determinants of Relative House Prices. Urban Studies 10, 213–233.

Cannari, L., Faiella, I., 2008. House Prices and Housing Wealth in Italy (SSRN Scholarly Paper No. ID 1162109). Social Science Research Network, Rochester, NY.

Chi, B., Dennett, A., Oléron-Evans, T., Morphet, R., 2020a. A new attribute-linked residential property price dataset for England and Wales, 2011-2019. Colchester, Essex: UK Data Service. doi:10.5255/UKDA-SN-854240.

Chi, B., Dennett, A., Oléron-Evans, T., Morphet, R., 2020b. Shedding new light on house price variation in England: a multi-scale exploration. Environment and Planning B: Urban Analytics and City Science. Available from: https://journals.sagepub.com/doi/10.1177/2399808320951212.

Chi, B., Dennett, A., Oléron-Evans, T., Morphet, R., 2019. Creating a new dataset to analyse house prices in England.CASA working paper 213. Available from: https://www.ucl.ac.uk/bartlett/casa/publications/2019/sep/casa-working-paper-213.

Ciarlone, A., 2015. House price cycles in emerging economies. Studies in Economics and Finance 32, 17–52.

Cook, S., 2003. The Convergence of Regional House Prices in the UK. Urban Studies 40, 2285–2294.

Cooper, C., Orford, S., Webster, C., Jones, C.B., 2013. Exploring the Ripple Effect and Spatial Volatility in House Prices in England and Wales: Regressing Interaction Domain Cross-Correlations against Reactive Statistics. Environment and Planning B: Urban Analytics and City Science 40, 763–782.

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

Department of Energy & Climate Change, 2013. An investigation of the effect of EPC ratings on house prices. Available from: https://www.gov.uk/government/publications/an-investigation-of-the-effect-of-epc-ratings-on-house-prices.

Fingleton, B., 2006. A cross-sectional analysis of residential property prices: the effects of income, commuting, schooling, the housing stock and spatial interaction in the English regions. Papers in Regional Science 85, 339–361.

Fleming, M.C., Nellis, J.G., 1992. Development of standardized indices for measuring house price inflation incorporating physical and locational characteristics. Applied Economics 24, 1067–1085.

Fuerst, F., McAllister, P., Nanda, A., Wyatt, P., 2016. Energy performance ratings and house prices in Wales: An empirical study. Energy Policy 92, 20–33.

Fuerst, F., McAllister, P., Nanda, A., Wyatt, P., 2015. Does energy efficiency matter to home-buyers? An investigation of EPC ratings and transaction prices in England. Energy Economics 48, 145–156.

Fuerst, F., McAllister, P.M., Nanda, A., Wyatt, P., 2013. Is Energy Efficiency Priced in the Housing Market? Some Evidence from the United Kingdom (SSRN Scholarly Paper No. ID 2225270). Social Science Research Network, Rochester, NY.

Gray, D., 2015. Are prices of New dwellings different? A spectral analysis of UK property vintages. Cogent Economics & Finance 3, 993860.

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

Hamnett, C., 1983. Regional Variations in House Prices and House Price Inflation 1969-81. Area 15, 97–109.

HM Land Registry, 2016. How to access HM Land Registry Price Paid Data [WWW Document]. How to access HM Land Registry Price Paid Data. Available from: https://www.gov.uk/guidance/about-the-price-paid-data#data-excluded-from-price-paid-data.

Hudson, C., Hudson, J., Morley, B., 2018. Differing house price linkages across UK regions: A multi-dimensional recursive ripple model. Urban Studies 55, 1636–1654.

Jeffreys, H., 1946. An Invariant Form for the Prior Probability in Estimation Problems. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 186, 453–461.

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

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 (Doctoral). UCL (University College London).

Marsden, J., 2015. House prices in London – an economic analysis of London's housing market. Greater London Authority.

McAvinchey, I.D., Maclennan, D., 1982. A Regional Comparison of House Price Inflation Rates in Britain, 1967-76. Urban Studies, 19 43–57.

MHCLG, 2019. Updates to Energy Performance Certificate open data. Available from: https://news.opendatacommunities.org/updates-to-energy-performance-certificate-open-data/.

Nielsen, F., 2010. A family of statistical symmetric divergences based on Jensen's inequality.

Office for National Statistics, 2019. Output Area (2011) to Built-up Area Sub-division to Built-up Area to Local Authority District to Region (December 2011) Lookup in England and Wales. Available from: http://geoportal.statistics.gov.uk/datasets/output-area-2011-to-built-up-area-sub-division-to-built-up-area-to-local-authority-district-to-region-december-2011-lookup-in-england-and-wales.

Office for National Statistics, Land Registry, Registers of Scotland and Land & Property, Services Northern Ireland, 2016. Development of a single Official House Price Index. Available from: https://www.ons.gov.uk/economy/inflationandpriceindices/methodologies/developmentofasingleofficialhousepriceindex

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, 248–264.

Orford, S., 2000. Modelling Spatial Structures in Local Housing Market Dynamics: A Multilevel Perspective. Urban Studies 37, 1643–1671.

Powell-Smith, A., 2017. House prices by square metre in England & Wales. . Available from: https://houseprices.anna.ps.

Rohde, N., 2016. J-divergence measurements of economic inequality. Journal of the Royal Statistical Society: Series A (Statistics in Society) 179, 847–870.

Rosenthal, L., 1986. Regional house price interactions in the UK, 1975–81: a cross-spectral analysis. Applied Economics 18, 1011–1023.

Simpson, P., Nesheim, L., Halket, J., Mysliwski, M., 2018. Estimating the benefits of transport investment. . Available from: https://www.ifs.org.uk/publications/13241.

Sirmans, G.S., MacDonald, L., Macpherson, D.A., Zietz, E.N., 2006. The Value of Housing Characteristics: A Meta Analysis. J Real Estate Finan Econ 33, 215–240.

South, B., Henretty, N., 2017. House price statistics for small areas: Using administrative data to give new insights. Statistical Journal of the IAOS 33, 609–614.

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

Thwaites, G., Wood, R., 2005. The Measurement of House Prices (SSRN Scholarly Paper No. ID 707043). Social Science Research Network, Rochester, NY.

Wabe, J.S., 1971. A Study of House Prices as a means of Establishing the Value of Journey Time, the Rate of Time Preference and the Valuation of some Aspects of Environment in the London Metropolitan Region. Applied Economics 3, 247–255.