A new insight into residential house price variation across England through linking Land Registry Price Paid Data and Domestic Energy Performance Certificates

Bin Chi*[1], Adam Dennett†[1] and Thomas Oléron-Evans‡[1]

[1]Centre for Advanced Spatial Analysis, University College London

January 23, 2019

Summary

In this paper we outline a method for integrating Land Registry Price Paid data and Domestic Energy Performance Certificates datasets in order to explore property level and area context level influences on house prices in England. Transaction price and its total floor area show a strong positive linear association in a majority of local authorities in England. This relationship varies between different geographic scales and by different property types across England, which has not been assessed on the temporal and spatial scale in England before.

KEYWORDS: House price, total floor area, correlation, house price variation, data linkage

1. Introduction

Housing is a major source of inequality in the UK, particularly in England (Dorling, 2014). In some areas of England the cost of renting or buying a house is becoming prohibitively expensive (Inman, 2017). Escalating housing costs reduce people's ability to buy or rent a dwelling. These housing affordability issues have been widely discussed in media and research communities (Collinson, 2014; John, 2015). Housing affordability is determined by residential house price and household income. A more nuanced understanding of residential house price in England will support an in-depth understanding of the housing affordability issue. However, data deficiencies are an obstruction to a comprehensive analysis. Presently, there is no comprehensive database which contains transaction price along with property characteristics in England and Wales (Wood, 2015). The current official house price dataset (Land Registry Price Paid Data) covers all residential transactions with a number of housing characteristics, but it does not contain any accurate housing size information. This is one of the most important house price variation determinants through house price modelling, especially for floor area (Orford, 2010). Building a comprehensive housing price database will produce an advanced understanding of the house price variation.

This research outlines a method for integrating Land Registry Price Paid Data (PPD) and Domestic Energy Performance Certificates (Domestic EPCs). The crucial linkage here adds floor space information in each property, enabling an assessment of the price paid per square metre – something which has not been assessed on a temporal and spatial scale in England before.

2. Data

2.1. Land Registry Price Paid Data

* bin.chi.16@ucl.ac.uk

† a.dennett @ucl.ac.uk

‡ thomas.evans.11@ucl.ac.uk

Land Registry PPD is an open administrative dataset from the UK's Her Majesty's Land Registry. It covers all transaction records in England in Wales since 1/1/1995.

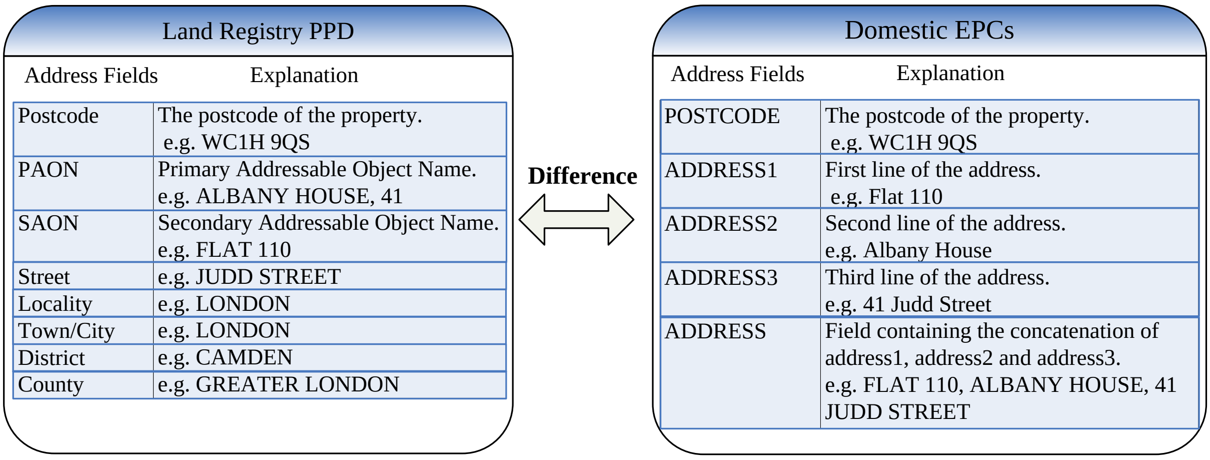
2.2. Domestic Energy Performance Certificates

Domestic EPCs released by the Department for Communities and Local Government (DCLG) contains property energy performance information and building stock information (i.e. total floor area, numbers of habitable rooms). The current public available EPC dataset starts from 1/1/2008 to 1/10/2016.

3. Method

3.1. Combination with Domestic Energy Performance Certificates

Land Registry PPD and Domestic EPCs offer the property information at address level, but their address information structures are different (Figure 1). Thus, a matching method containing 20 matching rules is created to combine Land Registry PPD and Domestic EPCs data. Figure 2 shows the match rate of Land Registry PPD by year.



Land Registry PPD

Address Fields	Explanation
Postcode	The postcode of the property. e.g. WC1H 9QS
PAON	Primary Addressable Object Name. e.g. ALBANY HOUSE, 41
SAON	Secondary Addressable Object Name. e.g. FLAT 110
Street	e.g. JUDD STREET
Locality	e.g. LONDON
Town/City	e.g. LONDON
District	e.g. CAMDEN
County	e.g. GREATER LONDON

Difference

Domestic EPCs

Address Fields	Explanation
POSTCODE	The postcode of the property. e.g. WC1H 9QS
ADDRESS1	First line of the address. e.g. Flat 110
ADDRESS2	Second line of the address. e.g. Albany House
ADDRESS3	Third line of the address. e.g. 41 Judd Street
ADDRESS	Field containing the concatenation of address1, address2 and address3. e.g. FLAT 110, ALBANY HOUSE, 41 JUDD STREET

Figure 1 Address information in Land Registry PPD and Domestic EPCs

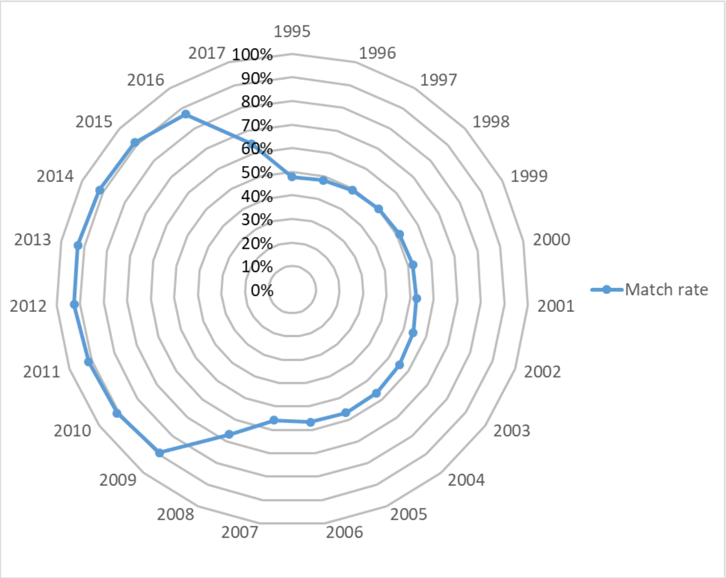


Figure 2 Match rate of Land Registry Price Paid Data,1995-2017

Following the combination of these two datasets, 14,570,679 transaction records were successfully linked for the whole of Land Registry PPD in England and Wales (1/1/1995-31/7/2017). The available EPC data time coverage is shorter, which results in a relatively low match rate for the periods before 2008 or after 2016.

3.2. Evaluation of house price information lost after data linkage

The Kolmogorov–Smirnov test (K-S test) and calculation of Jeffreys divergence (J-divergence) are used to understand the extent of the transaction information lost after linkage with Domestic EPCs. The K-S test is a nonparametric test that examines the differences in the shape of a distribution, in which statistic D is based on maximum absolute difference between two cumulative distribution functions. A K-S test result (statistic D) can be used to quantify the difference of two house price distributions (Original data VS linked data). In addition, J-divergence, derived from information theory, is a function used to establish the distance of one probability distribution to another (Jeffreys, 1946; Rohde, 2016). J-divergence is defined as:

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

where k is the number of categories, p^j is the probability of category j in the original data, and q^j is the probability of category j in the finally successful linked data. J ranges from 0 to 1. When the probability of two datasets across all the categories is the same, J will be 0. A larger value of J indicates a greater difference between two datasets.

Statistic D results (Figure 3) were bigger than 0.03 before 2008, while they were lower than 0.03 afterwards. This means more house price information was lost before 2008. Moreover, the extent of house price information lost after linkage in 2008 and 2017 is higher than the period from 2009 to 2016. The J-divergence result also shows the similar result. Therefore, the linked house price dataset from 2009 to 2016 is more reliable.

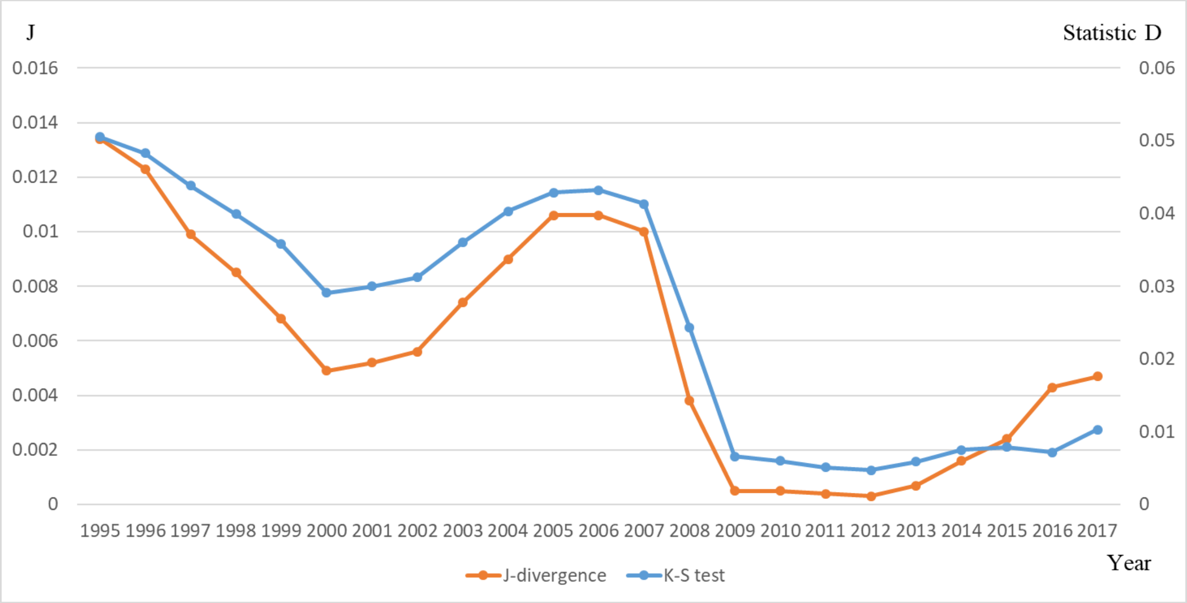
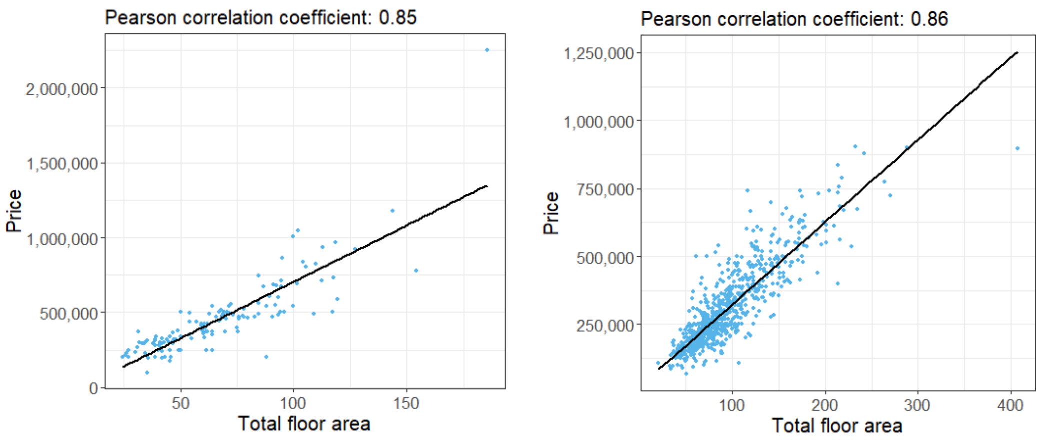


Figure 3 The result of K-S test and J-divergence method

4. Results

After linking house price with total floor area information from the Domestic EPCs, a strong positive linear association between transaction price and total floor area (as measured by the Pearson correlation coefficient) can be observed within individual local authorities. Figure 4 displays examples of this relationship for two sample local authorities.



a. City of London

b. Epsom and Ewell

Figure 4 Transaction price against total floor area in local authorities, 2009

We are also able to observe the geography of this relationship. Figure 5, based on linked data in 2009, shows the extent of linear association between transaction price and total floor area in each local authority across England. For 99% of local authorities, the correlation coefficient between price and total floor area (ρ) is larger than 0.5. 80% of local authorities show ρ is larger than 0.7; using the total floor area distribution in one of these local authorities, 70% the residential house price variation can be estimated. Lower correlations reveal areas where other contextual factors are having an increased influence on house prices and these can be observed in parts of London, Manchester, Liverpool and South Yorkshire.

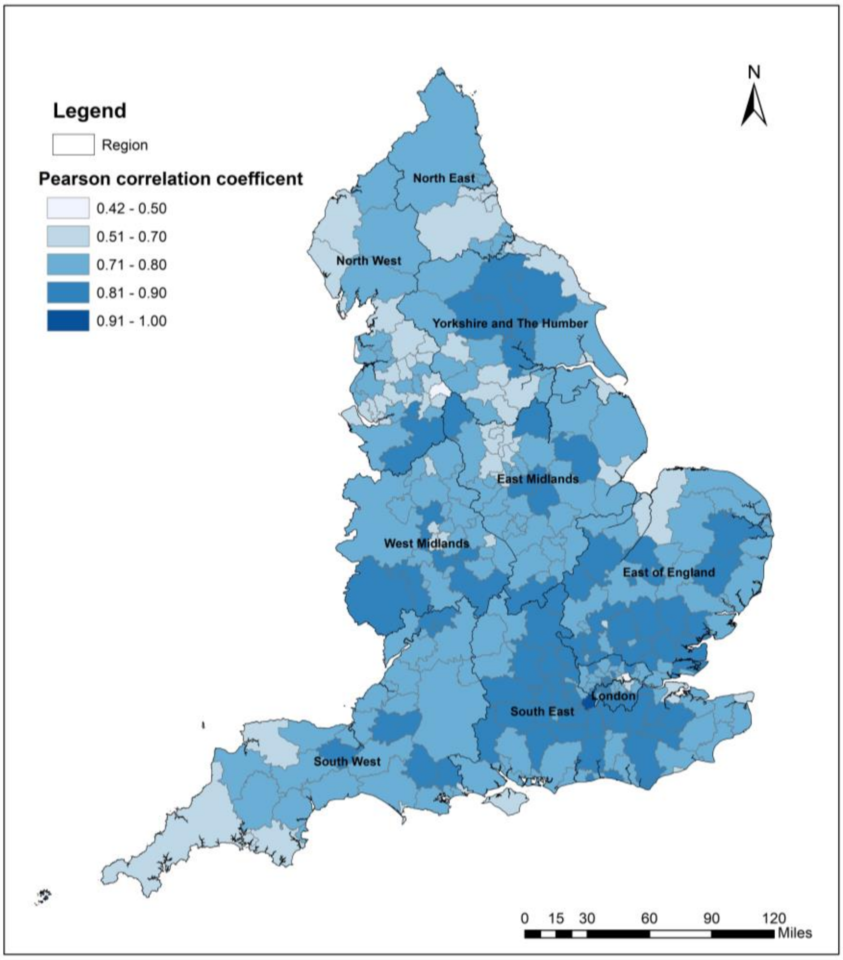


Figure 5 Pearson correlation coefficient at local authority level in England, 2009

We are able to unpick these relationships further by altering the scale of analysis. In some local authorities, house price and total floor area show a stronger linear relationship when moved to a smaller area of analysis, such as Middle Layer Super Output Area (MSOA) level and property type is controlled for. One sample is shown in Figure 6, where in Richmond, local variations in floor area are particularly important for the price of semi-detached houses.

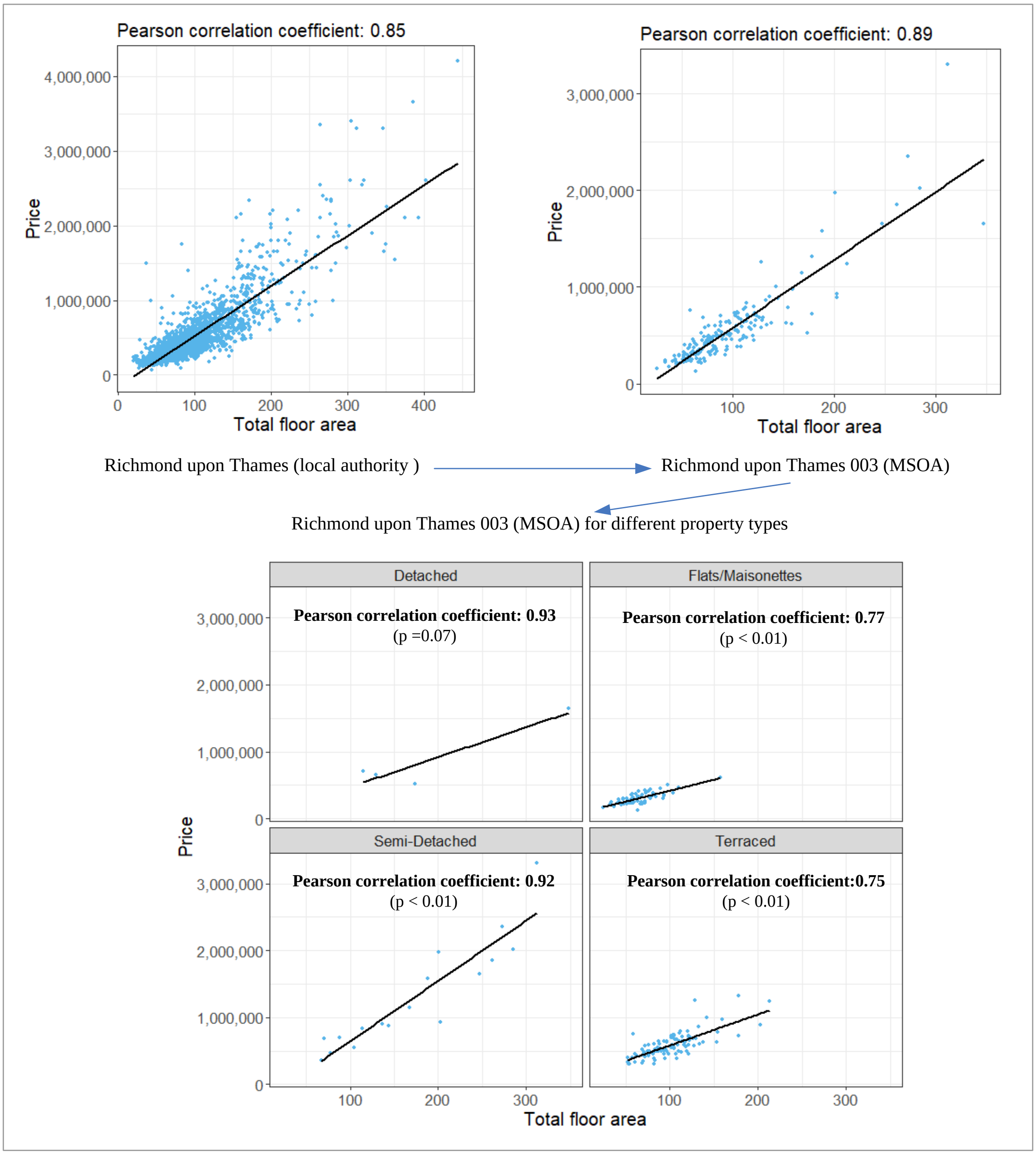


Figure 6 Transaction price against total floor area in Richmond upon Thames, 2009

5. Discussion

This research develops a method to overcome one limitation of the Land Registry, linking it with Domestic EPCs. The match rate is higher than previous research (Powell-Smith, 2017; Simpson et al., 2018). As a 100% match rate is not achieved, one statistical test and one differences measure are used to understand the degree of information lost after linkage. A comprehensive attribute housing price database is created which contains house price, property type, duration, old or new, dwelling total floor area and number of habitable rooms. This valuable new dataset advances explorations in house price variation, offering new insights into the housing market across England.

Based on the comprehensive housing price database, house price and total floor area show a moderate or strong linear relationship in local authorities across England. A stronger linear relationship was observed in some MSOAs and for individual property types within individual MSOAs, but these relationships are not consistent across all MSOAs. Moreover, as is shown in Figure 6, for some areas, low counts of transaction after Domestic EPCs linkage for certain property types (such as detached houses in Richmond upon Thames) mean that correlations are not statistically significant. Further study will explore how this relationship changes at MSOA level by different property types. Total floor area is one measure of property size, but others, such as building volume and plot size, are also worthy of investigation since they also impact house price variation. More descriptive and statistical analysis between house price and different property sizes need to be conducted.

6. Acknowledgements

This research is being co-funded by the China Scholarship Council (CSC No. 201708060184) and Overseas Research Scholarships from University College London.

References

Collinson P (2014) Data reveals full extent of house affordability crisis in England. *The Guardian*, 23 May. Available at: https://www.theguardian.com/money/2014/may/23/data-reveals-full-extent-house-affordability-crisis-england.

Dorling D (2014) *All That Is Solid: How the Great Housing Disaster Defines Our Times, and What We Can Do About It*. Penguin UK.

Inman P (2017) Housing crisis: more than 200,000 homes in England lie empty. Available at: http://www.theguardian.com/society/2017/apr/20/over-200000-homes-in-england-still-lying-empty-despite-housing-shortages.

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(1007): 453–461.

John B (2015) What is ‘affordable housing’? Available at: http://blog.shelter.org.uk/2015/08/what-is-affordable-housing/.

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. DOI: 10.1068/b35082.

Powell-Smith A (2017) House prices by square metre in England & Wales. Available at: 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(3): 847–870. DOI: 10.1111/rssa.12153.

Simpson P, Nesheim L, Halket J, et al. (2018) Estimating the benefits of transport investment. Available at: https://www.ifs.org.uk/publications/13241.

Wood R (2015) A comparison of UK residential house price indices. (21): 212–227.

Biographies

Bin Chi is a PhD student in the Bartlett Centre for Advanced Spatial Analysis (CASA), University College London. Her PhD research on understanding the spatial temporal patterns of housing price and housing affordability using statistical, mathematical and GIS techniques. Her research interests are in spatial temporal data analytics and modelling.

Adam Dennett is an Associate Professor and Director of the Bartlett Centre for Advanced Spatial Analysis (CASA), University College London. Adam is a Geographer, with interests broadly in the areas of population, quantitative methods, GIS and spatial analysis.

Thomas P. Oléron-Evans is a Lecturer in the Bartlett Centre for Advanced Spatial Analysis (CASA) at University College London. Thomas is a mathematician, whose research interests include optimisation, game theory and agent-based modelling and the application of these tools to interdisciplinary problems.