Property Heterogeneity and Convergence Club Formation among Local House Prices*

Mark J. Holmes
Department of Economics
Waikato University
New Zealand

Jesús Otero
Facultad de Economía
Universidad del Rosario
Colombia

Theodore Panagiotidis
Department of Economics
University of Macedonia
Greece

September 2018

Abstract

We investigate the extent of convergence club formation in local house prices. Our study is novel in a number of key ways. First, new insights are obtained through utilising a large disaggregated panel dataset comprising multiple types of housing (detached, semi-detached, terraced housing and flats) for a very large sample of 348 England and Wales local authorities. Second, we analyse through probit estimation those factors that drive convergence club formation. Third, we also consider within-city club formation as well as club formation during periods of house price bubbles. Using a study period of more than two decades, we find the presence of divergence or multiple house price convergence clubs rather than a single club. While location, distance, income, population density, congestion and education are significant in explaining convergence club formation, housing type is also an important factor with convergence less likely among detached, semi-detached and terraced housing than among flats.

JEL Classification: C2, C3, R1, R2, R3

Keywords: Convergence, clubs, house prices, England and Wales.

*We are grateful to Paul Carrillo (Editor) and two anonymous referees for their detailed comments, as well as participants at the 2018 Annual Conference of the Royal Economic Society (Brighton, UK) and the 2018 Annual Conference of the International Association for Applied Econometrics (Montreal, Canada). The usual disclaimer applies. E-mail addresses: mark.holmes@waikato.ac.nz (M.J. Holmes; corresponding author); jesus.otero@urosario.edu.co (J. Otero); tpanag@uom.gr (T. Panagiotidis).

1 Introduction

There is considerable value in understanding how relative regional house prices behave over time. Not only do they have the potential to influence relative regional economic activity, but also the affordability of housing, relocation costs and labour mobility between regions. While there is a relatively extensive literature that explores inter-regional house price convergence, there have been far fewer studies that explore intra-regional convergence. Indeed, little is known about the causes explaining the evolution of intra-regional house prices and the increased disparities within regions. This is because tradition has been to employ standard definitions of economic regions to study spatial differences in house prices by authors such as Meen (2002), Clark and Coggin (2009), and Fadiga and Wang (2009). One might consider the notion of convergence clubs that are based on the clustering of house price series according to some common factor such that there is movement over time leading towards a steady state. With this in mind, Kim and Rous (2012) note that the possibility of house price convergence club classifications that are not entirely compatible with any of the widely used definitions of economic regions has implications for the rationale behind including regional dummy variables. Moreover, estimated coefficients may not be capturing the full effect of important geographically-based omitted variables.[1]

In this paper, we examine the evidence for convergence clustering across England and Wales local authorities, on the basis of house price trends over the period 1995-2017. Given that we employ data from 348 authorities for four types of property for more than two decades, this is one of the largest datasets employed in the house price convergence literature; for a recent review see Meen (2016). It is of interest to investigate the characteristics of convergent and divergent subgroups as well as possible factors driving the

[1] Club convergence might also have significant policy implications in terms of identifying the regions where policy (government or from local authorities) can have more impact.

convergence clubs. With this in mind, we also explore whether the formation of convergence clubs is based on common factors among groups of local authorities leading them to converge to a similar house price level. A small number of studies have examined regional house price convergence using the club convergence and clustering procedure of Phillips and Sul (2007, 2009). For example, Montagnoli and Nagayasu (2015) find that house prices across UK regions can be grouped into four clusters, confirming the heterogeneity and complexity of the UK housing market. They document the dynamics of house price spillovers across regions. There are stronger spillover effects from the core regions such as London, but regional economic and financial developments are found to be important too. In the case of US states, Kim and Rous (2012) find strong evidence of multiple convergence clubs. They also examine the general characteristics of the various convergence and divergence subgroups as well as some important driving forces of convergence clubs. They find that housing supply regulation together with climate are important determinants of convergence club membership. Apergis et al. (2015) analyse the long run behaviour of provincial house prices in South Africa across nine provinces. In doing so, they consider three segments of the housing market and find that the nine provinces do not form a homogeneous convergence club. Data limitations prevent a formal analysis of the drivers behind convergence club formation. Blanco et al. (2016) identify convergence clubs in house prices among Spanish regions and find that regional house prices do not converge to a common trend and so confirms the existence of segmentation in the Spanish housing market. Spanish regions are found to be grouped across four separate convergence clubs. The estimation of an ordered logit model suggests that differences in population growth, size of the rental market, initial house supply and geographical situation have played significant roles in determining club membership.[2]

[2]In related literature, Yang et al. (2017) apply panel data unit root and stationarity tests to land sale prices in the German state of Lower Saxony to identify regions that

Another strand of the literature focuses more on the interaction of the housing market with macroeconomic variables. Kuttner and Shim (2016) gauge the effectiveness of non-interest rate policies on house prices whereas Plakandaras et al. (2017) discuss the effects of macroeconomic shocks on house prices for the US and the UK. Jordà et al. (2015) examine the interdependence between monetary policy and house prices. The earlier literature on the interaction between the macro economy and the housing market is reviewed in Goodhart and Hofmann (2007).

Whether or not local authority house prices drift apart, there are implications for relative affordability, labour mobility, labour mismatch, commuter times and potential localised house price bubbles. Essentially, these are issues relating to economics, financial stability and well-being. Our investigation differs from the above-mentioned studies in a number of key ways. First, we examine house price convergence clubs from an intra-regional perspective using a very large data set for 348 England and Wales local authorities. This is in sharp contrast to existing studies of inter-regional convergence that employ a far more limited number of time series based on aggregated regions. While existing studies of convergence club formation in house prices such as Kim and Rous (2012) and Tsai (2018) make a valuable contribution to our understanding of regional housing market interaction, they are based on more aggregated regional data. However, this extent of regional aggregation employed can mask insights in terms of what is going within the standard regional classifications in terms of convergence club formation and the drivers of club formation. Second, we take the extent of disaggregation further still through a breakdown of property types. In contrast to existing studies that provide insights into interregional convergence based on a single house price measure encompassing all property types, we distinguish between four prop-

exhibit similar land price behaviour over time. In turn, Kallberg et al. (2014), using a linear, dynamic, common-factor model, find that price co movement in the U.S. residential real estate market has increased between 1992 and 2008.

erty types. This enables us to investigate the extent to which convergence club formation is influenced by property type in addition to other drivers of convergence. Third, we offer an analysis of convergence club formation within cities. Studies by Jones and Leishman (2006), Liu et al. (2016), Oikarinen et al. (2018) and others point to the presence of heterogeneity of house price dynamics within cities. With our focus on local authorities across England and Wales, we also reflect on whether house price dynamics around convergence club formation within London paint a different picture compared to the rest of our sample.

We are therefore able to incorporate a significantly greater degree of heterogeneity in the UK housing market in our assessment of convergence.[3] It is well-known that districts within a given region can themselves be subject to great variation in terms of the house price behaviour and the local drivers that influence housing demand. Also, it may well be that housing type is at least as an important consideration as the distance between local authorities when it comes to assessing the extent of convergence club formation. Indeed, convergence club formation may well be more associated with certain types of housing than others. Indeed, heterogeneity in the housing market is important (see also Nemov et al. (2016)). These are key issues that our study explores by employing a disaggregated dataset. Given that housing affordability has been a key concern, borough price convergence or divergence across housing types is of potential interest to policy-makers with an eye on long-term affordability. The potential policy implications that arise from this are in terms of measures such as transport policy to address commuting issues, house building to address affordability, incentives for employment to be offered in cheaper areas and financial measures to address potential bubbles.

The rest of the paper is structured as follows. The following section provides a review of the Phillips and Sul convergence club methodology. The

[3]In a recent paper Pavlidis et al. (2017) show that cross-sectional aggregation can significantly affect the persistence properties of house prices.

third section describes the data set that we employ. Our data set is far more detailed than hitherto applied in convergence club studies. The fourth section reports and discusses our results. Based on 1386 time series, we find evidence in support of four convergence clubs for England and Wales house prices. In the fifth section, we analyse the drivers in terms of what determines convergence club membership. The final section concludes.

2 Econometric methodology: A brief review

Recent panel studies of long-run convergence among UK house prices such as Holmes and Grimes (2008), Abbott and De Vita (2012) and Holmes et al. (2018) have employed variations of panel unit root testing and cointegration-based methodologies. By way of contrast, the Phillips and Sul clustering algorithm that we employ here provides an empirical modelling of long-run equilibria within a heterogeneous panel, outside of the co-integration setup. An algorithm based on a $\log(t)$ regression approach clusters regions with a common unobserved factor in their variance. In this respect, sigma-convergence as opposed to beta-convergence deals with the reduction in the variance of the intra-regional house price distribution over time. In addition to detecting panel convergence, if present, a key advantage of the Phillips and Sul clustering algorithm test is that it can reveal whether club formation is also present. Also, this test does not necessitate any specific assumptions regarding the non-stationarity of the variables and allows for cases where individual series may be transitionally divergent. Indeed, the method by Phillips and Sul (2007) enables the detection of convergence where other methods such as stationarity tests fail insofar as stationary time series methods are unable to detect the asymptotic co-movement of two time series and therefore erroneously reject the convergence.[4] Drawing from the detailed ex-

[4]LeSage (1999) observes that spatial econometrics differs from standard estimation techniques in two dimensions that are present in the former but not in the latter, namely spatial dependence and spatial heterogeneity. The approach that we adopt in this pa-

positions in Phillips and Sul (2007, 2009), let us consider the following time varying factor representation of a panel data X_{it}:

$$X_{it} = g_{it} + a_{it}, \tag{1}$$

where g_{it} denotes permanent common components that give rise to cross section dependence, a_{it} denotes transitory components, $i = 1, ..., N$ is the number of individuals in the panel, and $t = 1, ..., T$ is the number of time observations. In this representation, X_{it} contains both common and idiosyncratic terms, so that in order to separate one from the other equation (1) is transformed using:

$$X_{it} = \left(\frac{g_{it} + a_{it}}{\mu_t} \right) \mu_t = \delta_{it} \mu_t, \tag{2}$$

for all i and t, where δ_{it} is a time varying idiosyncratic element, while μ_t is a time varying component that is common to all individuals in the panel. In equation (2), Phillips and Sul model the term δ_{it} in a semiparametric form as:

$$\delta_{it} = \delta_i + \sigma_i \xi_{it} L(t)^{-1} t^{-\alpha}, \tag{3}$$

where δ_i is fixed, $\xi_{it} \sim iid(0, 1)$ across i but weakly dependent over t, and $L(t)$ is a slowly varying function such that $L(t) \to \infty$ as $t \to \infty$; an example of the function $L(t)$ is $\log(t)$. Phillips and Sul observe that the formulation in (3) ensures that for all $\alpha \geq 0$, δ_{it} converges to δ_i, and so these conditions become the null hypothesis of interest. Within this framework, the authors develop a test of the null hypothesis of convergence based on:

per can be viewed as hybrid. Indeed, the notion of spatial lag is related to the idea of neighbouring observations affecting one another so that, for example, areas of high house values might be adjacent to other high value areas. Thus, in terms of the relationship between convergence clubs and spatial models, it should be noted that the formation of convergence clubs is based on common degrees of spatial correlation across local authorities. In analysing the drivers of convergence club formation, we employ systematic spatial sampling in order to assess the robustness of our findings to potential spatial interactions among neighbouring local authorities.

$$H_0 : \delta_i = \delta \quad \text{and} \quad \alpha \geq 0, \tag{4}$$

against the alternative of divergence:

$$H_A : \delta_i \neq \delta \quad \forall i \quad \text{or} \quad \alpha < 0. \tag{5}$$

The test is implemented in three steps. The first step involves constructing the cross sectional variance ratio H_1/H_t given by:

$$H_t = \frac{1}{N} \sum_{i=1}^{N} (h_{it} - 1)^2 , \tag{6}$$

where h_{it} is referred to as the relative transition parameter, which traces out the transition path of individual i in relation to the average of the panel, that is:

$$h_{it} = \frac{X_{it}}{N^{-1} \sum_{i=1}^{N} X_{it}}. \tag{7}$$

In equations (6) and (7), the notion of convergence implies, for a given number of time series in the panel N, that $h_{it} \to 1$ and therefore $H_t \to 0$ as $t \to \infty$. In the second step one runs the so-called log t regression:

$$\log \left(\frac{H_1}{H_t} \right) - 2 \log L(t) = \hat{a} + \hat{b} \log t + \hat{u}_t, \tag{8}$$

for $t = [rT], [rT] + 1, ..., T$, with $r > 0$. In this regression, Phillips and Sul use $L(t) = \log(t + 1)$ and further observe that the estimate of the slope coefficient $\hat{b} = 2\hat{\alpha}$, where $\hat{\alpha}$ is the estimate of α in the null hypothesis. The regression is estimated using a fraction rT of the observations, where the trimming parameter r is recommended to be set equal to 0.3 (in our empirical application we examine the robustness of the results when this parameter is varied taking on the values of 0.15, 0.20, 0.25 and 0.30).

Lastly, in the third step, the notion of convergence is tested using a one-sided t test of the inequality part of the null hypothesis $\alpha \geq 0$, where the

test statistic is constructed using a heteroskedasticity and autocorrelation consistent standard error. Using a 5% significance level, the null hypothesis is rejected when $t_{\hat{b}} < -1.65$.

Phillips and Sul (2007, 2009) indicate that rejection of the null hypothesis of convergence does not rule out the possibility of convergence in subgroups of panel individuals. Thus, to allow for this possibility, the authors develop an algorithm with the purpose of determining the number of potential convergence clubs as well as their conforming members (also permitting the categorisation of some time series as non-convergent). The original version of the Phillips and Sul algorithm was implemented in GAUSS, although versions of it in R and Stata have been produced by Schnurbus et al. (2017) and Du (2017), respectively. For the purposes of our empirical analysis, we shall be using the latter version of the code. The algorithm involves four steps, which can be sketched as follows (the interested reader is referred to these references for more details):

- Step 1: Isolate and remove the cyclical component from X_{it} using a suitable filter to extract trend and cycle information from a time series, and order the resulting smoothed version series according to the value of the last observation or, in case of highly volatile series, the average value of the last half of observations.

- Step 2: Form a core primary group of k^* individuals using the maximum t_k, subject to $t_k > -1.65$, from the sequential $\log(t)$ regression based on the k highest individuals in the panel, where $2 \leq k < N$.

- Step 3: Add one individual at a time to the core primary group with k^* members, estimate the $\log(t)$ regression again, and add the new individual to the core primary group if the resulting t-statistic is greater than the criterion c^*. Phillips and Sul refer to this step as the sieve condition, and recommend setting $c^* = 0$ in small T samples, while

for large T the asymptotic 5% critical value -1.65 is the recommended value.

- Step 4: Form a second group of individuals for which the sieve condition stated in step 3 fails. Estimate the $\log(t)$ regression on this group and verify if the resulting t statistic is greater than -1.65. If this condition is satisfied then conclude that there are two convergence clubs, namely the core primary group and the second group. Otherwise, repeat steps 1 through 3 to see if the second group can be subdivided into smaller convergence clubs. If in step 2 there is no k for which $t_k > -1.65$, conclude that the remaining individuals in the panel do not contain a convergence subgroup, and so the remaining individuals exhibit divergent behaviour.

3 Data

We employ quarterly data on the median price (measured in pounds sterling per square metre) paid for detached houses (d), flats (f), semi-detached houses (s) and terrace houses (t) in the 348 local authorities (and boroughs) that comprise England and Wales, which yields a total of $348 \times 4 = 1392$ price series.[5] The source of the sale prices is the HM Land Registry database of residential property sales to individuals in England and Wales from 1 January 1995 to 28 February 2017, which contains more than 22 million observations on prices based on housing transactions. To obtain prices per square metre, we use the floor area per property as taken from Energy Performance of Buildings Data in England and Wales provided by the Department for Communities and Local Government.[6] With this database, we estimate the

[5] Although in practice real estate agents may be tempted to push the advantage of an end-terrace house, these properties are not usually viewed as fully equivalent to a semi-detached house.

[6] The house price data were downloaded from *https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads* on 28 April 2017. The floor area data were

median floor area (in m^2) of each property at the level of postcode district, and the resulting values are subsequently merged with the file that contains the sale prices using the corresponding property postcode. The resulting prices per square metre are then used to compute time series of median prices for each local authority and property type over the sample period 1995q1 to 2017q1, for a total of $T = 89$ time observations.[7] Price series with more than 25% missing observations are dropped from the empirical analysis, while for series with 25% or fewer missing observations the missing values are estimated using linear interpolation. All in all, we end up with 1386 quarterly time-series of house prices.[8]

To gain some initial insight on the evolution of property prices over time, Figure 1 presents plots of the time-series of (the log of) median property prices grouped by region, as classified by the Office for National Statistics (ONS). The price series are displayed at this (larger) level of aggregation because of the very large number of local authorities involved in our empirical analysis. In each figure, we also include the median property price in England and Wales as a measure of the overall development of property prices. Visual inspection of the time series plots in these figures suggests that within each property type, the price series are non-stationary and tend to move together as time passes. There could be support for the hypothesis of price cointegration not only across local authorities, but also across the types of property, which are topics that have received a great deal of attention in the literature; see, inter alia, Holmes and Grimes (2008), Holly et al. (2011), Holmes et al. (2011), Abbott and De Vita (2012) and, more recently, Holmes et al. (2018). Another feature that is apparent from the graphs is that prices

downloaded from *https://epc.opendatacommunities.org/* on 1 July 2017.

[7]Given that the price database does not include sale transactions for March, the price information in 2017q1 may be viewed as preliminary.

[8]The six time series not included in the analysis correspond to the prices in the City of London for detached, semi-detached and terrace houses, as well as prices in the Isles of Scilly for detached, semi-detached and terrace houses.

in the London, East, South East and South West regions are above the median level in the whole of England and Wales. In addition, an examination of prices within each region reveals that the value of the square metre tends to be more expensive for detached houses.

4 Club convergence analysis

We begin the presentation of our empirical analysis by reporting the results when the log(t) test is applied across 1386 local authorities and four different property types in England and Wales. For the purposes of the econometric analysis, these house price series are considered in logarithms, and after removing their underlying cyclical component using the Hamilton (2018) filter.[9] In addition, we discard the first 25% of the sample period, that is the first 22 time observations for each house price series; qualitatively similar results are obtained when this trimming parameter is varied by taking on the values of 30% (as recommended by Phillips and Sul), 20% and 15%.

Ordinary least squares estimation of the log(t) regression yields an estimate of the slope coefficient equal to -0.351, with an associated HAC standard error of 0.027, and a t-statistic of -13.123. This result indicates that the slope coefficient is statistically different from zero, supporting the view of house price divergence across all 1386 series. Developing further this initial result, we examine the possibility of conforming smaller clubs of convergence through the application of the clustering algorithm outlined before. Table 1 presents the initial classification of convergence clubs, the statistical tests of potential club merging, and the final classification. As can be seen in the

[9]Illustrating their approach, Phillips and Sul use the Hodrick and Prescott (1997) (HP) filter which is widely popular in empirical work in time series macroeconomics. However, in a recent paper, Hamilton (2018) criticises the HP filter mainly because of the introduction of spurious dynamic relations. To overcome this and other limitations, Hamilton proposes the removal of the cyclical component of a series by running an OLS regression of the variable of interest against a constant and its eighth, ninth, tenth and eleventh lagged values, and using the resulting fitted values as an estimate of the underlying trend component.

table, the evidence points towards the existence of four convergence clubs, which can be visualised in Figures 2 to 5. Here, we present a map of England and Wales where the shading corresponds to local authority convergence club membership. The area corresponding to the London boroughs is magnified in order to give them reasonable exposure. Across property types, these maps provide initial evidence for the potential role of contiguity in the context of club formation. Indeed, the clubs with the highest median house prices can be observed in London and the South of England, with clubs of lower median prices occurring in Wales and in the intermediate and northern parts of England. While a casual inspection of these figures indicates that contiguity and geographic identity may be a key factor in explaining membership of a particular convergence club across England and Wales or within London, this is not always the case. For example, there are many instances across England and Wales where localised or isolated pockets of particular club membership exist. With this in mind, we later explore other drivers of convergence club membership.

The aggregated UK study by Montagnoli and Nagayasu (2015) also finds evidence of four house price clusters for the 12 regional house price series that they consider. With the increased degree of regional disaggregation in our study which allows for greater heterogeneity with respect to housing types, we find that house prices fall into the same number of convergence clubs. Our evidence, therefore, points to the England and Wales housing market being removed from an overall converged state (which would occur under the presence of a single convergence club).

Our results thus far show patterns of spatial correlation in home price appreciation across local authorities in England and Wales. At this point, two further issues that arise concerning the extent to which convergence clubs can be also identified in areas within a city, as well as across countries. To address the first issue, we perform the convergence clubs analysis focusing only on the 32 London boroughs and the City of London. The $\log(t)$ regression

results yield a slope coefficient equal to -0.665, with an associated HAC standard error of 0.116, and a t-statistic of -5.736, which unequivocally rejects the null hypothesis of convergence. Subsequent application of the Phillips and Sul merging algorithm also provides support for the existence of four convergence clubs. Interestingly, when one compares the composition of the resulting convergence clubs in London, there is a great deal of similarity in terms of whether house prices in London are analysed as part of England and Wales, or separately. As for the second issue, Tsai (2018) recently uncovered evidence that housing prices in several European countries converge towards house prices in Germany. Thus, it is of interest to examine whether or not the disaggregated house price series in England and Wales also converge to those in Germany or vice versa.[10] The incorporation of the German house price series along with those for the 1386 local authorities in England and Wales leaves the convergence club results largely unaffected; that is, the $\log(t)$ regression results yield a slope coefficient equal to -0.350, with an associated HAC standard error of 0.027, and a t-statistic of -13.117 which, once again, rejects the null of convergence. As before, the merging algorithm provides support for the presence of four convergence clubs. Rather than being classified as part of a non-convergent club, the German price series being classified in the fourth club along with the local authorities where, on average, the lowest house prices are observed.

Finally, although the Phillips and Sul convergence approach allows for episodes where the time series are transitionally divergent, it is useful to examine whether convergence patterns differ when house prices exhibit periods of aberrant variation (if at all present). This leads us to the area

[10] We have put together a German series by taking the quarterly Bank of International Settlements (BIS) series for the index of average house price in Germany. Then, using data that indicated that the average detached property price in Germany was 260,000 Euros in July 2014, we re-scaled the index throughout into Euros, and converted it into a pound sterling measure of the German house price using the Euro/Pound exchange rate. Admittedly, this is a rough and ready measure because of comparability issues in the housing markets of the countries.

of literature that deals with the identification of price bubbles, when price movements are measured in relation to some 'fundamentals', or exuberance; see, e.g. Phillips and Yu (2011) and more recently Hu and Oxley (2018). In terms of econometric testing, Phillips, Wu and Yu (PWY) (2011) and Phillips, Shi and Yu (PSY) (2015) advocate the use of recursively calculated right-tailed augmented Dickey and Fuller (ADF) (1979) statistics to "date-stamp" episodes of bubbles/exuberance. More specifically, PWY test the unit-root null against the alternative of a single periodically collapsing explosive period through repeated estimation of the ADF regression on a forward expanding sample sequence, and obtain the test as the sup value of the resulting sequence of ADF statistics; the test is therefore referred to as SADF. PSY, on the other hand, test the unit-root null against the alternative of multiple periodically explosive periods by allowing more flexible windows of estimation, and so the test is referred to as the generalised SADF test, or GSADF for short.

Applying the more general GSADF test on the (logarithm of the) individual house prices, including a constant and a lag order of one, indicates that the unit root null hypothesis is rejected at the 5% significance level in 800 out of the 1386 time series. For these cases, the episodes of exuberance, more often than not, tend to occur during the first half of the 2000s and also at the end of our study period.[11] Some insight into the potential effect of these episodes of localised exuberance on the number of convergence clubs can be gained through an expanding window estimation of the Phillips and Sul convergence procedure. Thus, we start off by performing the estimation for the sub-sample 1995q1-2006q1; then 1995q1-2006q2; and so on until reaching the last observation available, that is 2017q1. Figure 6 summarises the results of the recursive estimation (where some caution must be exercised when in-

[11] To perform inference, the required critical values were estimated through Monte Carlo simulations, based on 2000 replications. We are grateful to Itamar Caspi for kindly providing us with an Eviews code to implement the GSADF test; see Caspi (2017) for more details.

terpreting the results because the sample period for the recursive estimation starts much later than that of the GSADF test). As can be seen, the number of convergence clubs exhibits a U-shape pattern, which appears to suggest that the number of convergence clubs increases during periods of localised price exuberance.

5 Drivers of club convergence

In contrast to Montagnoli and Nagayasu (2015), we now turn to consider the potential drivers of convergence club formation. For this, we first estimate a binary (probit) model and find the variables that are expected to have an effect on the likelihood of any two local authorities being part of the same convergence club. To do this, it is important to bear in mind that ideally, one would need to assemble a consistent database of the relevant explanatory variables across all 348 local authorities over the sample period. However, in several instances, data availability prevents us from calculating the average value of a variable during the totality of the study period, but only in some specific years. Thus, caution must be exercised when interpreting the effects because potential variations in levels over the whole time span may have not being accounted for.[12]

In order to measure geographic separation, it is necessary to obtain the latitude and longitude of the population-weighted centroids of the local authorities in England and Wales. Once the coordinates are available, we use the Stata programme `osrmtime` developed by Huber and Rust (2016), which employs the Open Source Routing Machine software based on Open Street

[12] An alternative estimation that is worth considering is that of a hazard duration model, which is more flexible regarding the use of censored data, and can also be viewed as a generalisation of the probit models that we use in this paper. Of special relevance, for instance, through hazard duration models one would be able to check whether (some of) the determinants of club membership vary over time. Unfortunately, this alternative route imposes major implications in terms of obtaining sufficient time series data on all our covariates, and for this reason is not considered in the paper, but postponed for future research.

Maps, to determine the shortest route by car between any two pairs of geographic coordinates, and calculate the corresponding optimal distance (in metres) and optimal travel time (in seconds). For brevity, in what follows we only present the results based on optimal distance, since those based on optimal travel time are qualitatively similar.[13] Positive income movements across local authorities, which are related to productivity and wages, should attract flows of workers thus pushing up local housing demand and in turn house prices. For a measure of income, we draw on the data available from income and tax by borough and district or unitary authority (taxpayers only) for the tax year 2014-15, which is published by the ONS. For demographic influences we consider the logarithm of population density (per hectare) using data taken from the 2011 population census. Another variable of potential importance is crime, which enters the analysis with the intention of capturing security conditions in the boroughs. Being a negative attribute, crime is expected to adversely impact on house prices. For crime rates, we employ police data for on the number of burglaries from dwellings. Again, the data are for 2011 which matches the population census year. In terms of a measure of school quality across our large sample, a problem that emerges is that Wales is no longer consistently included with the rest of the UK in national education league tables. As a result of this, we employ local authority data on the education attainment of the (16-64 years old) population and re-express this in terms of the percentage of the local population with an educational qualification. Lastly, there is evidence that traffic congestion affects urban housing prices importantly (see, e.g., Hou (2017)), and so there is the question of whether it also matters for convergence patterns. We proceed on the basis that the measurement of congestion concerns time cost. On the one

[13]The results produced by the software `osrmtime` take into account the fact that the distance (duration time) through the road transportation network between local authorities i and j is not necessarily the same as that between j and i; thus, for the purposes of our the econometric analysis we use the simple average between the two distances (duration times).

hand, we already have data on population density which one might expect to be positively correlated with congestion on roads. On the other hand, there is also workplace commuting distance data in the sense that longer commuting distances are likely to be associated with more congestion for the local population going to work. Thus, we draw data from the 2011 census on the distance travelled to work for the different local authorities.

The results reported in Table 2 indicate that the probability that any two price series are members of the same convergence club is influenced by a number of key drivers. Estimation of a probit model suggests that larger distances between the local authorities and greater differences in population density, income, crime, education quality and total congestion lead to a smaller likelihood that any two house price series belong to the same convergence club. If we consider the different types of properties, the evidence suggests that pairs of flats are the most likely to belong to the same convergence club. In varying degrees, terraced, semi- and detached housing pairs are less likely to be in the same house price convergence club. With semi- and detached housing as arguably the more heterogeneous property types in our sample, this is perhaps to be expected. It is perhaps more surprising that terraced housing which one might regard as a relatively homogeneous property type compared with semi- and detached housing, also has a negative coefficient. Compared to flats and semi-detached housing, the price variation within terraced housing is considerable when looking across local authority areas. This factor will influence the standing of terraced housing when it comes to club formation. In terms of the regional dummies, we find that price series drawn from the East Midlands, North East, North West, South West, Wales, West Midlands, and Yorkshire and the Humber are more likely to belong to the same convergence club compared to the East and South East. Interestingly, there is no significant effect if both series are from either inner- or outer-London. Holmes et al. (2018) in comparing the housing stocks of Inner and Outer London local authorities, find that flats constitute

a much larger share of the local housing stock than terraced housing in the former. Also, terraced housing is relatively more important as a share of the housing stock in the Outer London local authorities. This suggests that different forms of housing may not necessarily be driving a differing positioning on the part of Inner and Outer London towards convergence club formation. Instead, a further consideration here is the the possibility that the presence of the congestion charge that dichotomises inner- and outer-London has not affected club formation.

It is likely that the fundamental drivers of house prices are highly correlated, and so it is worth examining whether the coefficient estimates are robust to potential spatial interactions among neighbouring local authorities. To account for potential spatial effects, we employ systematic spatial sampling which involves using for estimation purposes only those observation pairs that are x kilometres away from their nearest neighbours in all four directions; see, e.g. Müller (2005) for an illustration. Within this context, we compute the first and second quartiles of the empirical distribution of the distance variable (that is, 140 and 226 kilometres, respectively), and use only the observation pairs above these values to estimate the corresponding probit models; see models 2 and 3 in Table 2, respectively. As can be seen, the resulting coefficient estimates are qualitatively similar, maintaining their signs and statistical significance.

Finally, another question whose answer lies within the reach of the econometric modelling strategy employed in the paper is that of the drivers or factors that determine the probability that a local authority belongs to a specific convergence club. For this, an ordered probit model provides a suitable tool of analysis. Table 3 summarises the estimated coefficients of the ordered probit model along with their associated marginal effects which are easier to interpret. According to the results, increases in income, density, education and congestion make membership of clubs 1-3 (4) more (less) likely in varying degrees. This is mirrored with the estimated coefficient on crime

not being statistically different from zero for any of the four clubs.

6 Concluding remarks

Compared with earlier studies of regional house price convergence, the employment of a large data set comprising four types of housing for each of 348 local authorities over a twenty year period provides an ideal opportunity for new insights. The presence of multiple convergence clubs rather than just a single club impacts directly on the assessment of relative affordability across regional housing markets with further implications for labour mobility, labour mismatch and commuter times. We also find that convergence club formation may also arise during times of localised house price bubbles. Additionally, our study points to UK house price clustering being not just about the location of the local authorities, but also about the type of housing. The cheaper properties within each local authority are terraced housing and flats and are therefore most relevant to the affordability issue for lower income households. Across local authorities, we find that convergence is more likely for flats but less likely for terraced housing. Indeed, our findings imply relative divergence in the case of terraced housing and therefore potential for relative affordability to worsen. This issue is exacerbated with a strongly significant role for income differentials in driving club membership. As a caveat, it should of course be pointed out that the strong correlations we observe between these drivers and convergence club formation do not necessarily mean that the drivers cause club formation. With this in mind, we cautiously offer some policy recommendations. Policy measures aimed at addressing the relative affordability of cheaper housing might, therefore, be strongly borough-focused dependant upon the significance of these two types of housing within the local housing stocks. We also find that crime and congestion are significant factors in convergence club formation. Measures that can bring crime rates and congestion levels down towards a common nation-

wide standard are likely to facilitate movements towards a single convergence club for an integrated England and Wales housing market.

References

Abbott, A. and G. De Vita (2012). Pairwise convergence of district-level house prices in London. *Urban Studies 49*(4), 721–740.

Apergis, N., B. D. Simo-Kengne, and R. Gupta (2015). Convergence in provincial-level South African house prices: Evidence from the club convergence and clustering procedure. *Review of Urban & Regional Development Studies 27*(1), 2–17.

Blanco, F., V. Martín, and G. Vazquez (2016). Regional house price convergence in Spain during the housing boom. *Urban Studies 53*(4), 775–798.

Caspi, I. (2017). Rtadf: testing for bubbles with Eviews. *Journal of Statistical Software 81*(C), Code Snippet 1.

Clark, S. and T. Coggin (2009). Trends, cycles and convergence in U.S. regional house prices. *The Journal of Real Estate Finance and Economics 39*(3), 264–283.

Dickey, D. A. and W. A. Fuller (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association 74*(366), 427–431.

Du, K. (2017). Econometric convergence test and club clustering using Stata. *The Stata Journal 17*(4), 882–900.

Fadiga, M. and Y. Wang (2009). A multivariate unobserved component analysis of US housing market. *Journal of Economics and Finance 33*(1), 13–26.

Goodhart, C. and B. Hofmann (2007). *House Prices and the Macroeconomy: Implications for Banking and Price Stability.* Oxford: Oxford University Press.

Hamilton, J. D. (2018). Why you should never use the Hodrick-Prescott filter. *The Review of Economics and Statistics forthcoming.*

Hodrick, R. J. and E. C. Prescott (1997). Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking 29*(1), 1–16.

Holly, S., M. H. Pesaran, and T. Yamagata (2011). The spatial and temporal diffusion of house prices in the UK. *Journal of Urban Economics 69*(1), 2–23.

Holmes, M. J. and A. Grimes (2008). Is there long-run convergence among regional house prices in the UK? *Urban Studies 45*(8), 1531–1544.

Holmes, M. J., J. Otero, and T. Panagiotidis (2011). Investigating regional house price convergence in the United States: Evidence from a pair-wise approach. *Economic Modelling 28*(6), 2369–2376.

Holmes, M. J., J. Otero, and T. Panagiotidis (2018). Climbing the property ladder: An analysis of market integration in London property markets. *Urban Studies forthcoming.*

Hou, Y. (2017). Traffic congestion, accessibility to employment, and housing prices: A study of single-family housing market in Los Angeles county. *Urban Studies 54*(15), 3423–3445.

Hu, Y. and L. Oxley (2018). Do 18th century 'bubbles' survive the scrutiny of 21st century time series econometrics? *Economics Letters 162*(C), 131–134.

Huber, S. and C. Rust (2016). osrmtime: Calculate travel time and distance with openstreetmap data using the open source routing machine (OSRM). *The Stata Journal 16*(2), 416–423.

Jones, C. and C. Leishman (2006). Spatial dynamics of the housing market: An interurban perspective. *Urban Studies 43*(7), 1041–1059.

Jordà, O., M. Schularick, and A. M. Taylor (2015). Betting the house. *Journal of International Economics 96*(S1), 2–18.

Kallberg, J. G., C. H. Liu, and P. Pasquariello (2014). On the price comovement of U.S. residential real estate markets. *Real Estate Economics 42*(1), 71–108.

Kim, Y. S. and J. J. Rous (2012). House price convergence: Evidence from US state and metropolitan area panels. *Journal of Housing Economics 21*(2), 169–186.

Kuttner, K. N. and I. Shim (2016). Can non-interest rate policies stabilize housing markets? evidence from a panel of 57 economies. *Journal of Financial Stability 26*(C), 31–44.

LeSage, J. P. (1999). *The Theory and Practice of Spatial Econometrics*. Toledo, OH: University of Toledo.

Liu, C. H., A. Nowak, and S. S. Rosenthal (2016). Housing price bubbles, new supply, and within-city dynamics. *Journal of Urban Economics 96*(C), 55–72.

Meen, G. (2002). The time-series behavior of house prices: A transatlantic divide? *Journal of Housing Economics 11*(1), 1–23.

Meen, G. (2016). Spatial housing economics: A survey. *Urban Studies 53*(10), 1987–2003.

Montagnoli, A. and J. Nagayasu (2015). UK house price convergence clubs and spillovers. *Journal of Housing Economics 30*(C), 50–58.

Müller, D. (2005). Stata in space: Econometric analysis of spatially explicit raster data. *The Stata Journal 5*(2), 224–238.

Nemov, P. T., E. R. Larsen, and D. E. Sommervoll (2016). Thick-market effects, housing heterogeneity, and the determinants of transaction seasonality. *Economic Journal 126*(598), 2402–2423.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedastic and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Oikarinen, E., S. C. Bourassa, M. Hoesli, and J. Engblom (2018). U.S. metropolitan house price dynamics. *Journal of Urban Economics 105*(C), 54–69.

Pavlidis, E., E. Martinez-Garcia, and V. Grossman (2017). Detecting periods of exuberance: A look at the role of aggregation with an application to house prices. Technical Report 325, Federal Reserve Bank of Dallas.

Phillips, P. C. B., S. Shi, and J. Yu (2015). Testing for multiple bubles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review 56*(4), 1043–1077.

Phillips, P. C. B. and D. Sul (2007). Transition modeling and econometric convergence tests. *Econometrica 75*(6), 1771–1855.

Phillips, P. C. B. and D. Sul (2009). Economic transition and growth. *Journal of Applied Econometrics 24*(7), 1153–1185.

Phillips, P. C. B., Y. Wu, and J. Yu (2011). Explosive behavior in the 1990s NASDAQ: When did exhuberance escalate asset values? *International Economic Review 52*(1), 201–226.

Phillips, P. C. B. and J. Yu (2011). Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics 2*(3), 455–491.

Plakandaras, V., R. Gupta, C. Katrakilidis, and M. E. Wohar (2017). Time-varying role of macroeconomic shocks on house prices in the US and UK: Evidence from over 150 years of data. Technical report, University of Pretoria.

Schnurbus, J., H. Haupt, and V. Meier (2017). Economic transition and growth: A replication. *Journal of Applied Econometrics 32*(5), 1039–1042.

Tsai, I.-C. (2018). House price convergence in euro zone and non-euro zone countries. *Economic Systems 42*(2), 269–281.

Yang, X., M. Ritter, and M. Odening (2017). Testing for regional convergence of agricultural land prices. *Land Use Policy 64*(C), 64–75.

Table 1: Classification of convergence clubs

Initial classification				Tests of club merging				Final classification			
Club	β coeff.	t-stat.	p-value	Club	β coeff.	t-stat.	p-value	Club	β coeff.	t-stat.	p-value
[1]	-0.165	-1.605	(0.054)	[1+2]	-0.225	-6.255	(0.000)	[1]	-0.165	-1.605	(0.054)
[2]	0.071	0.491	(0.688)	[2+3]	0.074	0.563	(0.713)	[2]	-0.099	-0.905	(0.183)
[3]	0.014	0.074	(0.529)	[3+4]	0.019	0.118	(0.547)	[3]	0.194	6.450	(1.000)
[4]	0.125	0.706	(0.760)	[4+5]	0.061	0.288	(0.613)	[4]	-0.027	-0.423	(0.336)
[5]	0.091	0.439	(0.670)	[5+6]	0.185	1.035	(0.850)				
[6]	0.357	1.237	(0.892)	[6+7]	0.180	0.617	(0.731)				
[7]	-0.373	-0.628	(0.265)	[7+8]	0.154	1.120	(0.869)				
[8]	0.177	1.352	(0.912)	[8+9]	0.093	3.598	(1.000)				
[9]	0.194	6.450	(1.000)	[9+10]	-0.095	-2.130	(0.017)				
[10]	-0.027	-0.423	(0.336)								

Note: The numbers in brackets denote the number of convergence clubs. t-statistics are based on Newey and West (1987) heteroskedasticity and autocorrelation (HAC) standard errors. Lower tail probability values in parentheses.

Table 2: Estimation results from probit model

	Marginal effects					
	Model 1		Model 2		Model 3	
ln(Distance)	-0.009‡	(0.000)	-0.014‡	(0.000)	-0.013‡	(0.000)
Abs. Dif. ln(Density)	-0.035‡	(0.001)	-0.032‡	(0.001)	-0.030‡	(0.001)
Abs. Dif. ln(Income)	-1.450‡	(0.006)	-1.584‡	(0.006)	-1.562‡	(0.008)
Abs. Dif. ln(Crime)	-0.025‡	(0.001)	-0.025‡	(0.001)	-0.027‡	(0.001)
Education ratio	-0.157‡	(0.008)	-0.186‡	(0.009)	-0.154‡	(0.011)
ln(Total congestion)	-0.287‡	(0.003)	-0.178‡	(0.004)	-0.065‡	(0.005)
Property dummy variables						
Detached	-0.045‡	(0.002)	-0.059‡	(0.002)	-0.054‡	(0.003)
Flat	0.095‡	(0.002)	0.099‡	(0.003)	0.084‡	(0.003)
Semi-detached	-0.009‡	(0.002)	-0.017‡	(0.003)	-0.015‡	(0.003)
Terrace	-0.027‡	(0.002)	-0.026‡	(0.003)	-0.024‡	(0.003)
Region dummy variables						
East	-0.090‡	(0.004)	-0.126‡	(0.008)	-0.133‡	(0.020)
East Midlands	0.374‡	(0.004)	0.387‡	(0.014)		
Inner London	-0.094	(0.061)				
Outer London	-0.019	(0.056)				
North East	0.063‡	(0.017)				
North West	0.172‡	(0.005)	0.053‡	(0.010)	0.202‡	(0.023)
South East	-0.139‡	(0.003)	-0.084‡	(0.006)	0.009	(0.023)
South West	0.009†	(0.005)	0.014†	(0.008)	0.090‡	(0.020)
Wales	0.130‡	(0.008)	0.062‡	(0.013)	0.006	(0.017)
West Midlands	0.457‡	(0.003)				
Yorkshire and The Humber	0.251‡	(0.010)	0.335‡	(0.032)		
Observations	958420		721232		484076	
Pseudo R^2	0.1184		0.1098		0.1005	

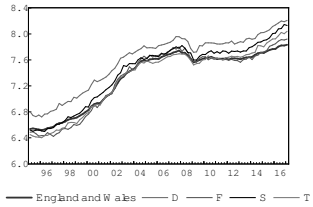
Note: The dependent variable takes value 1 when any two price series are members of the same convergence club (0 otherwise). Model 1 includes all observation pairs. Models 2 and 3 include observation pairs that are 140 and 226 kilometres away from their nearest neighbours in all four directions, respectively. Property (region) dummy variables take the value of one if a price pair involves the same property type (region), and zero otherwise. Heteroskedasticity consistent standard errors in parentheses. † $p < 0.10$, ‡ $p < 0.05$.

Table 3: Estimation results from ordered probit model

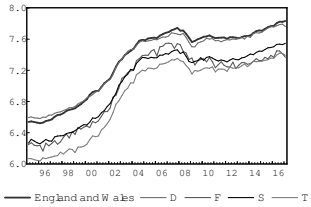
	Coeff.	Marginal effects			
		Club=1	Club=2	Club=3	Club=4
ln(Density)	-0.354‡	0.001‡	0.045‡	0.076‡	-0.121‡
	(0.043)	(0.000)	(0.006)	(0.011)	(0.015)
ln(Income)	-4.433‡	0.007‡	0.559‡	0.951‡	-1.517‡
	(0.487)	(0.003)	(0.071)	(0.122)	(0.167)
ln(crime)	-0.019	0.000	0.002	0.004	-0.006
	(0.056)	(0.000)	(0.007)	(0.012)	(0.019)
Education	-8.786‡	0.013‡	1.109‡	1.885‡	-3.007‡
	(1.618)	(0.006)	(0.201)	(0.375)	(0.545)
ln(Congestion)	-0.408‡	0.001	0.051‡	0.087‡	-0.140‡
	(0.188)	(0.000)	(0.025)	(0.040)	(0.065)
Observations	1386				
Pseudo R^2	0.287				

Note: The dependent variable takes value 1 for price series in convergence club 1, and so on until value 4 for price series in convergence club 4. The threshold parameters are not reported to save space. Heteroskedasticity consistent standard errors in parentheses. † $p < 0.10$, ‡ $p < 0.05$.

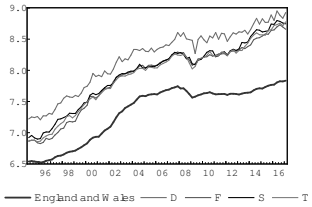
Figure 1: Median quarterly house prices per m^2 by region and property type



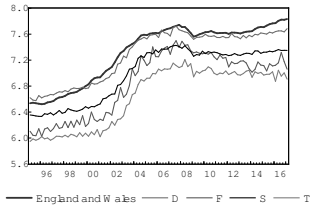
(a) East



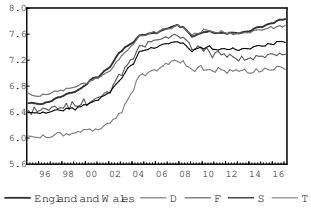
(b) East Midlands



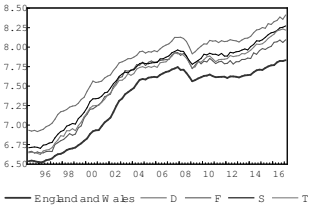
(c) London



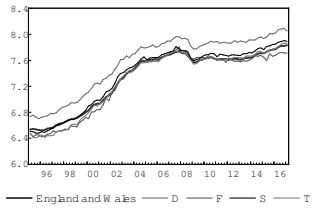
(d) North East



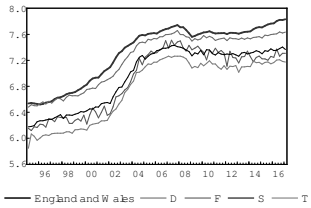
(e) North West



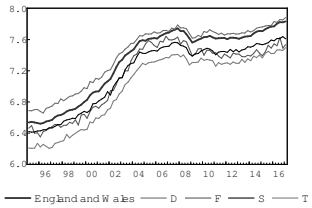
(f) South East



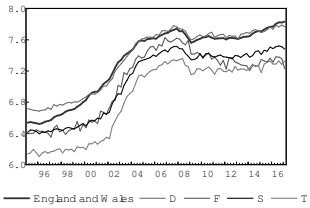
(g) South West



(h) West



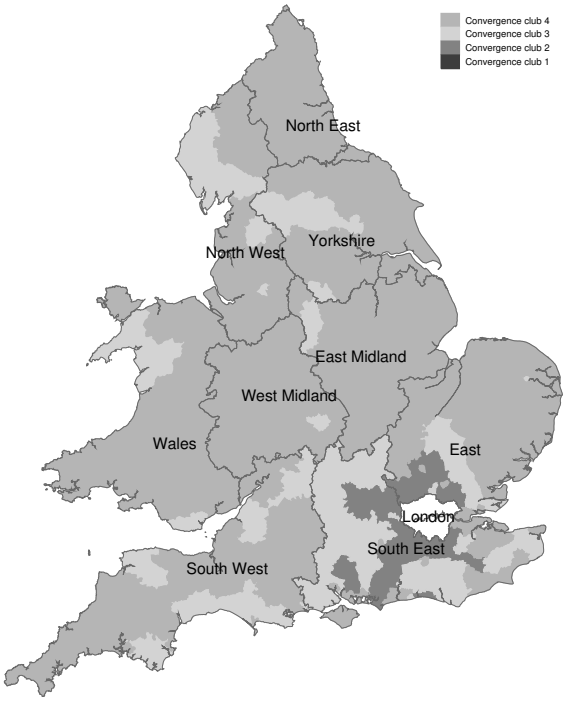
(i) West Midlands



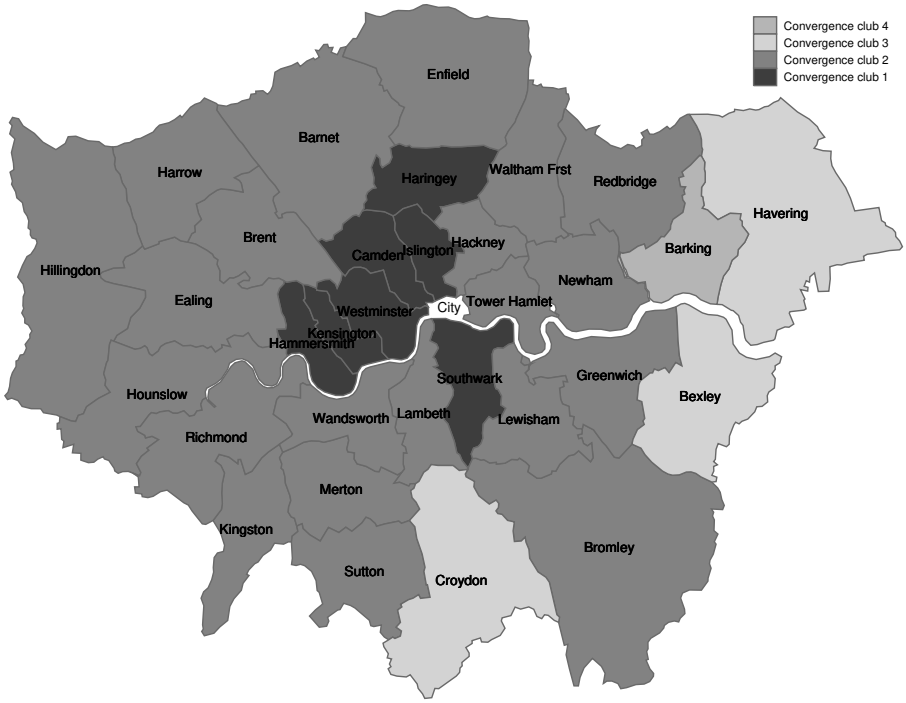
(j) Yorkshire & The Humber

Note: Property types are detached houses (D), flats (F), semi-detached houses (S) and terrace houses (T).

Figure 2: Convergence clubs for detached houses



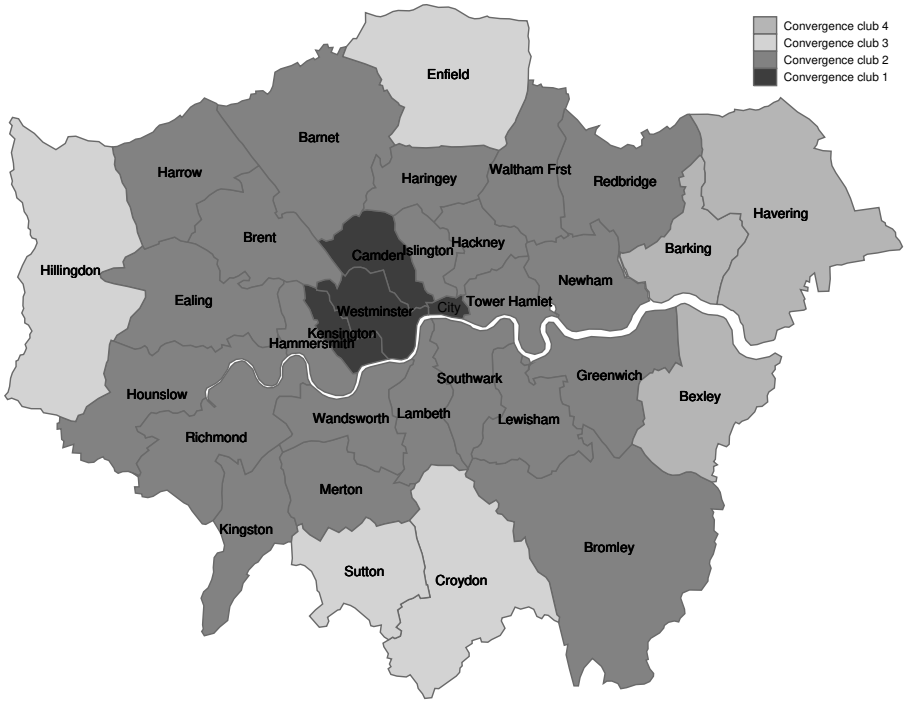
(a) England and Wales



(b) London

Figure 3: Convergence clubs for flats

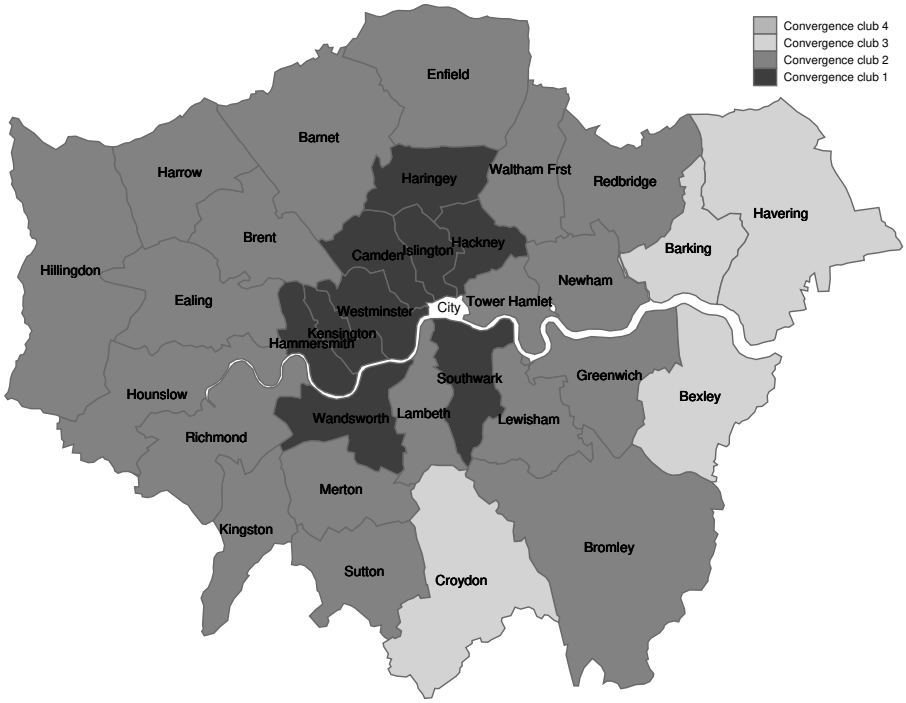
(a) England and Wales



(b) London

Figure 4: Convergence clubs for semi-detached houses

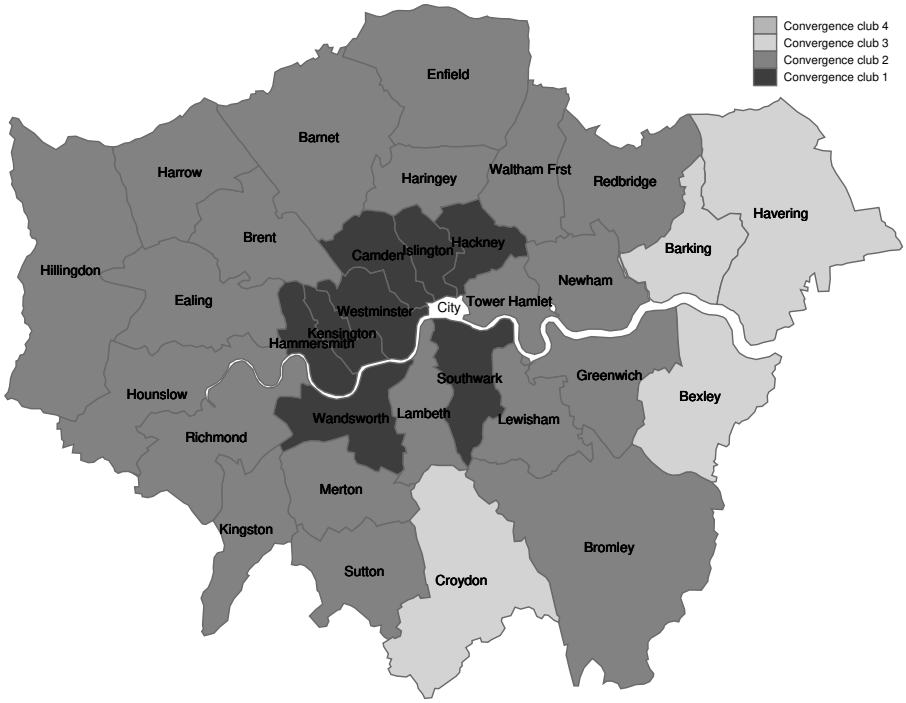
(a) England and Wales



(b) London

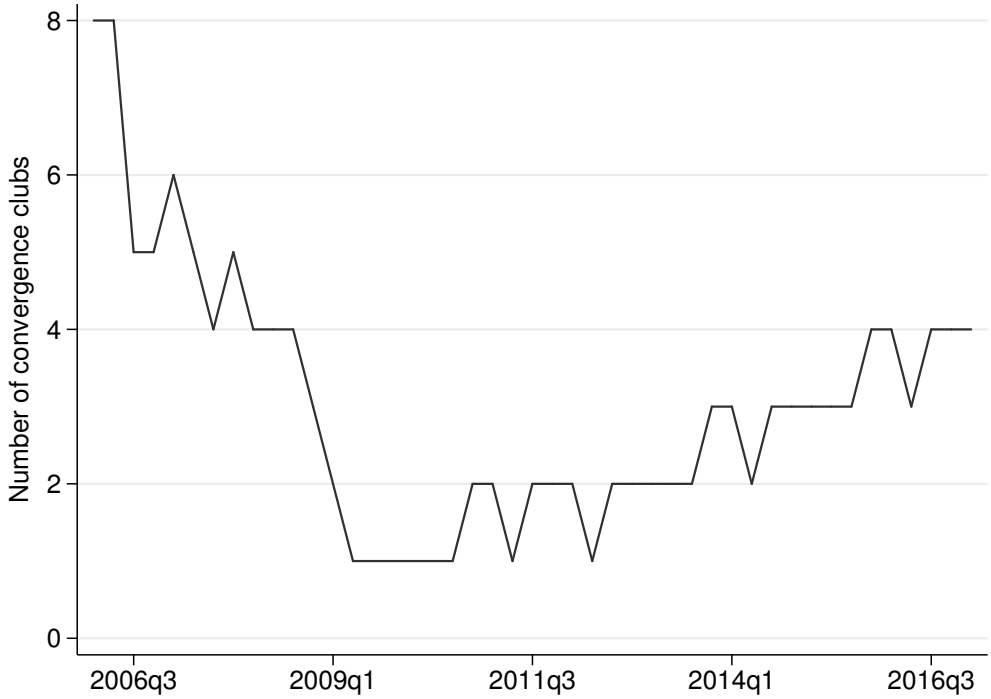
Figure 5: Convergence clubs for terrace houses

(a) England and Wales



(b) London

Figure 6: Recursive convergence club analysis



Note: The picture depicts the number of convergence clubs based on an expanding window with a 10-year initial estimation.