Communities and Local Government

Identifying Submarkets at the Sub-Regional Level in England

planning

Identifying Submarkets at the Sub-Regional Level in England

University of Glasgow

May 2007
Department for Communities and Local Government: London

Research by:

Professor Gwilym Pryce
Department of Urban Studies and

Gareth Evans
Department of Geographical and Earth Sciences

both University of Glasgow

This research was originally commissioned under the New Horizons programme. The New Horizons programme assists Communities and Local Government's horizon-scanning capabilities by seeking proposals from outside researchers to explore issues beyond the scope of our main research programmes. It encourages new and innovative approaches to issues and research, and probes the margins between orthodox and unconventional research methods. In addition, the programme looks forward with a view to identifying potential new challenges facing the Department over the next decade.

Although this research was commissioned by Communities and Local Government, the findings and recommendations in this report are those of the author and DO NOT necessarily represent the views of Communities and Local Government.

Department for Communities and Local Government
Eland House
Bressenden Place
London
SW1E 5DU
Telephone: 020 7944 4400
Website: www.communities.gov.uk

If you require this publication in an alternative format please email alternativeformats@communities.gsi.gov.uk

Communities and Local Government Publications
PO Box 236
Wetherby
West Yorkshire
LS23 7NB
Tel: 08701 226 236
Fax: 08701 226 237
Textphone: 08701 207 405
Email: communities@twoten.com
or online via the Communities and Local Government website: www.communities.gov.uk

May 2007

Product Code: 07HC04605

Contents:

Executive Summary

One of the imperatives that has emerged following the Barker Review of Housing Supply (2004) is the need to improve our measurement and understanding of sub-regional housing markets. A great obstacle to evidence based housing policy and investment is the sheer complexity of local house prices in terms of their variation over both time and space. How can we derive structure and meaning from such a baffling degree of intricacy and volatility in the pattern of property transactions?

The theory of submarkets – the notion that sub-regional housing systems are not comprised of a single market but a whole patchwork of inter-connected micro-markets – lends meaning to the complex system we observe in the housing market data and offers a conceptual framework on which we can construct empirical techniques.

Existing techniques have either been atheoretical or of limited practical use in terms of assisting policy makers and market investors. This project sought to develop a method of submarket delineation that was theoretically sound yet readily applicable to most areas of the United Kingdom; that could be easily aggregated to a variety of geographical units, yet was not dependent on those units for its fundamental derivation; and that could incorporate data from multiple time periods and yet provide a visualisation of the data that would allow the researcher to see at a glance the structure of the housing system.

In response to these objectives a method was developed based on the estimation of house price inflation for every point in geographical space. On the basis that houses in the same submarket will appreciate at similar rates, we then clustered dwellings (according to their proximity to each other and their estimated rate of constant quality inflation) to derive a series of submarket areas.

The method could be described as belonging to the *spatial* submarkets stream of segmenting housing systems. Yet, it deviates from the typical approach used in that literature because it draws on market dynamics (rather than attributes of the market or of properties at a given point in time) as the primary organising principle by which properties are allocated to particular submarkets.

To demonstrate how the technique works, we applied the method to Kent and East Sussex, using Nationwide house price information for the period 1996 to 2004.

One of the advantages of our method is that the intermediate steps are potentially of use in themselves, helping us visualise and understand the structure of the sub-regional housing system. The constant quality price surfaces, for example, provide a detailed picture of the location value (land prices) of each point in geographical space.

The report also considers how the model could be developed to analyse a range of other issues. For example, it could provide a way of assessing the impact of regeneration initiatives, and gauging the spatial patterns in housing wealth inequality. It could also be used to simulate the house price effect of a range of flood risk and climate change scenarios.

The report suggests that the inflation surface approach be applied to other regions of the United Kingdom and that research continues into how the model could be further extended and improved.

1 Introduction

How can policy researchers and commercial organisations reliably determine the boundaries of housing submarkets? Following the Barker Review of Housing Supply (2004)[1] there has been renewed interest in improving the quality and usefulness of information provided to planners, developers and policy makers, and the question of defining submarkets has emerged as central to that goal. This is because more detailed examination of housing data reveals spectacular variations across space, both at the regional and sub-regional level. Even at neighbourhood and street level, we can observe immense variation. All headline figures of house price inflation and housing market change aggregate across space and therefore strip out much of that important information. Yet the sheer complexity and volume of information contained in geographical patterns of house prices leave us in need of a systematic way of structuring our analysis. Submarkets potentially provide us with the organising principle we seek.

To appreciate the value of such a principle, it is worth recalling the imperatives that have been highlighted with regard to our need to know more about housing submarkets:

- The impact of new supply on existing house prices will be driven by the nature and structure of the existing landscape of inter-locking housing submarkets. So, planners and developers need to be aware of submarkets if they are to anticipate the impacts of their decisions;

- House prices and supply do not respond uniformly to changes in demography, policy, and unanticipated shocks – there is a spatial pattern to the response which is conditioned by submarkets;

- Long-term differences in house price appreciation across housing submarkets have implications for the inequality of wealth accumulation;

- The distribution of ethnic groups, skill and income categories is spatially dependent so different sets of people will be affected by the asymmetric responses across submarkets.

So why has the analysis of housing submarkets not entered the standard toolbox of policy makers and information providers? While rich and varied academic literature has emerged on the topic of housing submarkets[2], a number of key challenges hinder the useful application of the concept:

- **Data** – by definition, in order to examine submarkets, one needs detailed geo-coded information on the prices and attributes of dwellings and hopefully also their neighbourhoods. Analysis of submarkets has tended to focus on regions where such data exists (such as the West of Scotland: Pryce 2004, Ismail 2005, Watkins 2001) rather than on where the analysis of submarkets is most needed. London and the South East, where much of the policy debate is currently focussed,

[1] Barker Review of Housing Supply, HM Treasury, March 2004 – http://www.hm-treasury.gov.uk/consultations_and_legislation/barker/consult_barker_index.cfm

[2] See the recent review of the literature by Pryce (2004). A notable milestone was Grigsby's 1963 monograph which argued that the urban housing system was unlikely to be made up of a single uniform market but a patchwork of interconnected submarkets. Since then, researchers have endeavoured to deepen our understanding of the meaning and nature of submarkets and to find ways of identifying and modelling them (Maclennan and Tu 1996; Watkins 2001; Goodman and Thibodeau 1998, 2003; Gelfand et al 2003).

has yet to be subject to rigorous submarket mapping (at least, not in ways that correspond to economic theory), and this is partly due to problems in sourcing the appropriate data.

- **Methods** – there is no one, agreed way to identify or describe submarkets. For the purposes of policy analysis many of the existing methods are not appropriate. Approaches that are not underpinned by simple intuition and that cannot be easily reproduced by other researchers are of limited value. Also, non-spatial definitions of submarkets, while theoretically interesting, are not amenable to policy application. If consideration of submarkets is to become a routine component of policy decisions, both by central government and by regional and sub-regional planning bodies, a method has to be developed that can be easily understood, applied to any area, and frequently updated. While tools already exist that offer detailed maps of statistical data such as labour market and Census information, such approaches do not offer a delineation of submarkets that is well grounded in economic theory and may be of limited use in helping planners and developers identify the optimal location of new supply or anticipate the impact of new supply.

- **Administrative Boundaries** – Pryce (2004) highlighted the problems of relying on a method that rests too heavily on data that is available only for relatively large administrative areas, or that uses the boundaries of such areas as the means of defining the submarkets. This is because submarkets do not respect administrative boundaries. However, while it might be desirable in principle to use a method to derive submarket boundaries that is totally independent of any existing boundaries, there is a draw-back when one comes to actually make use of the optimally-derived areas. Whether we like it or not, existing official economic and social data are published using standard administrative boundaries and so it may not be easy to mesh the derived submarkets with other published data for wider policy purposes. There is a need, therefore, to design a method that can produce a measure of submarket similarity that can be easily aggregated to any standard administrative area, while not being dependent on these boundaries in its fundamental derivation.

- **Spatial Scale** – Different policy and planning decisions require definitions of submarkets at different spatial scales. We seek a method that is relatively flexible in the spatial scale at which it can be applied. For example, we may wish to identify eight to ten key housing 'submarkets' for the whole of the South East. For a different set of policy considerations, however, we might want to consider the detailed housing market breakdown of a single county, such as Kent or East Sussex.

So our goal in this project has been to find a way of delineating submarkets that is amenable to analysis at a range of spatial scales, that is based on sound and reproducible economic analysis, that is not overly dependent on data units based on administrative boundaries yet is amenable to being represented using any standard boundary definition, and that is not limited to an ad hoc local dataset but could in principle be applied to any area of the United Kingdom.

It is perhaps at this point that the realisation dawns that the goals we have set ourselves are nothing short of impossible. Inevitably compromises will have to be made to greater or lesser degrees with regard to each of these objectives. Yet we hope that, in the pages which follow, the methods developed for the purposes of this project will present themselves to the reader as meeting most of these objectives sufficiently well to warrant their inclusion in the tool kits of policy makers in planning bodies and central government alike. If nothing else, we believe that these methods will help us understand and visualise more clearly the structure of the housing market. This in itself is a worthy goal.

1.1 Aims

The original aims of the project were as follows:

- To investigate the relative suitability of existing data sources (Council of Mortgage Lenders Survey of Mortgage Lenders data, HM Land Registry data, and Nationwide data) for the delineation of housing submarket boundaries in English regions;

- To investigate the development of a methodology that could produce useable outputs on a single English region (such as the South East) within a relatively short timeframe, and which could be reproduced for other regions in future without excessive cost or difficulty;

- To investigate whether measures of submarket variation could be achieved at varying spatial scales, and to investigate how easy it would be to map that information using standard software, and to summarise it for standard administrative areas;

- To present or offer suggestions on possible applications of the model used to derive these boundaries to research questions of interest to Communities and Local Government.

1.2 Plan

The remainder of the report is structured as follows:

- Chapter 2 examines the meaning of "submarkets" and the guiding principles for identifying them;

- Chapter 3 reports on the suitability of existing data and the combination of datasets selected for use in the final analysis;

- Chapter 4 sets out in detail the methodology developed in the project to define submarkets in the South East of England. The explanation is given by way of an illustration – ie via an application of the method to Kent and East Sussex;

- Chapter 5 discusses further applications of the method and directions for future research;

- Chapter 6 concludes and offers a number of recommendations.

2 Defining Submarkets

2.1 Introduction

In this chapter we offer a brief summary and critique of the literature and explain the implications for the choice of research methodology. In particular, we trace out the two broad streams in the submarkets literature and attempt to tease out the core aspects that we would like to be included in any new approach.

2.2 The non-spatial approach to submarkets

Two broad approaches have emerged towards the identification of submarkets. The first is quintessentially non-spatial. In this view of the world, two houses can be in the same market even if they are many kilometres apart. What binds them into the same submarket is the fact that they are close substitutes for one another and so potential buyers consider them both as legitimate contenders. Put more formally, both dwellings are elements of the same choice-set of potential buyers. This perspective on the nature of submarkets has been less concerned with mapping out the geographical boundaries of particular submarkets (because individual streets, even stand-alone dwellings, can be part of a separate and distant submarket) and more concerned with clustering dwellings by attribute types. There are two distinct drawbacks with this approach, however. First, it holds little prospect of being useful for policy analysis. This is because the political and administrative process in the United Kingdom and most other countries is fundamentally territorial. Members of Parliament, local authority councillors, and Members of the European Parliament are all elected at particular spatial scales. As a result, their obligation is to a particular constituency or ward and the allocation of resources is ultimately and inevitably decided along spatial lines. This is as true of housing and planning as it is of National Health Service expenditure, education, and police. True, some funds are allocated at individual household level (such as means-tested benefits) and some are almost entirely non-spatial in their allocation (such as national defence), but most have some spatial component or variation. Moreover, land use planning is generally preoccupied with intrinsically spatial phenomena. Zoning and land release are guided by the implications of contiguity – *where* one allows noisy, polluting industrial units to locate will have a profound effect on the well-being of residents and other firms. More subtly, *where* one allows new houses to be built will affect profoundly the eagerness of developers to build there and the demand of consumers to locate there. One of the imperatives for understanding the structure of housing submarkets is to identify the areas that are viewed as desirable – so that planners and developers can scratch where the market is itching. This is fundamentally a spatial problem.

The second drawback of the aspatial approach to defining submarkets is that dwellings have often been grouped without any prior knowledge or theory as to which attributes constitute a separate submarket. The researcher ultimately has to apply his own judgement, when really what we want to know is what the market considers the defining features of a submarket. We have said that properties are in the same submarket if they are close substitutes and if they are part of the same choice set of buyers. Ultimately prices are the most useful and easily measured signal of how consumers make housing choices and so the only convincing way to group properties into submarkets is to adopt a method that makes some reference to attribute prices or price dynamics. Currently, most analysis in this field has grouped properties by how similar the physical features of the house are and then looked at the impact on price, when really the logic should run the other way: first group properties by how consumers value them and then look at whether there happens to be any pattern in construction type across the groups. This is because location itself is an important determinant of whether a property is part of the same choice set – a three bedroom modern flat in London is unlikely to be part of the same submarket as an identical apartment in the Outer Hebrides. Grouping properties by type takes no account of that.

2.3 The spatial approach to submarkets

The second stream of literature on submarkets has tended to group properties by area and test for differences in the estimated prices of property attributes across these areas. The main problem with this literature is that it has relied heavily on standard administrative boundaries to group properties. This has two main drawbacks. First, it means that the test-points for breaks between submarkets are arbitrary. The problem is exacerbated by the fact that techniques used to test for breaks are not sufficiently precise to allow one to apply a fixed threshold for what constitutes a break. Depending on how one sets up the regression equation, one can make all break points exceed the standard statistical threshold of 5% significance, or none at all. So if one adopts this approach, one really needs to test for break-points at every point in geographical space and consider the relative degree of structural breaks across space (Pryce 2004). Second, it rules out the possibility that there may indeed be a relatively small pocket of dwellings that are either a unique submarket or are part of a different submarket than that of the surrounding neighbourhoods. In some situations, being able to identify 'islands' of properties such as these can be useful. They might, for example, give us an insight into what makes consumers view particular locations as being more or less desirable, and more importantly, actually allow us to quantify those effects. This is important because these location factors will qualify the impact of new supply and other policy decisions, but without knowing whether it is school catchment, proximity to power lines, flood risk, spectacular views, access to public transport or particular combinations of these that most influence consumer decisions, we are left to guess where and which type of houses should be built.

2.4 Developing a hybrid approach

So ideally we need a hybrid of the spatial and aspatial definitions of submarkets. And indeed, this has been attempted in the literature. But the result has typically been a doubling in the complexity of the technique and an overdependence on the clustering of property types by arbitrary criteria. And there are further problems that have dogged both the spatial, aspatial and hybrid approaches which we need to overcome.

2.4.1 Isolating quality effects from quantity effects

A full tank of petrol will cost more than half a tank. The price differential does not indicate the existence of two separate submarkets! Crucially, the unit price is the same. Price differences only reliably indicate submarket differences if: (a) they relate to identical quantities; and (b) they are persistent. With regard to (a), the great problem in housing economics is that no two houses are the same. Attribute price differences (differences in the additional value of having a garage or a garden etc) may simply reflect differences in quantity. A large house will cost more than a small house for the same reason that a full tank of petrol costs more than half a tank.

This point is so important that we believe it should form the starting point of submarket analysis rather than an afterthought to the methodology. One of the goals of the analysis below is to do just this – to make the isolation of quantity from quality the defining feature of the methodology being proposed.

2.4.2 Persistence and dynamics

Point (b) above is also important. Temporary differences in unit prices do not necessarily indicate the existence of submarkets. Information imperfections, the messiness of the buying and selling process and other market frictions can all throw up sizeable but fleeting aberrations. This is an important point because the great majority of submarket studies in both the United States and the United Kingdom have been cross sectional. That is, they have looked only at a snapshot in time, albeit in great detail, but a snapshot nonetheless. So it is impossible to ascertain whether the criteria being used to identify submarket boundaries – such as attribute price shifts – are merely temporary blips.

We are aware of only two studies that attempt a truly dynamic consideration of submarkets: Jones et al 2003 and Pryce and Gibb 2006. Jones et al 2003 argue that for localities to be considered as separate submarkets, not only must their attribute prices be different at a particular point in time, but also the dynamics of house prices must be independent: 'we consider whether price differences between submarkets have been eroded by a process of arbitrage operating through supply-side responses and/or migration flows' (p.1315). They employ cointegration tests as a practical method of investigation and find that, 'a stable system of housing submarkets persists throughout the study period' (p.1315).

Pryce and Gibb (2006) consider liquidity – the time it takes to sell a house – as a possible criterion for the examination and delineation of submarkets. They contend that the shape of the "hazard curve" – the graph that plots the likelihood of a property selling given the length of time the property has been on the market – should be the same for properties in the same submarket. This is because the shape of the curve reflects the buyer's perception of the state of the market, and so if two properties are in the same market, they should have similar shaped curves.

The main weakness of these two studies is that they both use pre-defined initial demarcations of submarkets. Even if these delineations were correct in the initial period considered, in the case where long periods of time are being studied, it would be inappropriate to ignore the possibility that submarket boundaries potentially shift (and even disappear!) over time.

Perhaps more profoundly, the implications of both Jones et al 2003 and Pryce and Gibb 2006 is that dynamics should, in fact, be another of the defining principles by which submarkets are initially identified. Similarities, or indeed differences, in attribute prices at particular snapshots in time can be coincidental and can reflect omitted variables – such as an adequate measure of size. Persistence in differences or similarities is much more revealing. Significant differences in the trajectories of the price of a homogenous unit of housing (that is, controlling for quality and quantity variations) is highly likely to indicate the existence of separate submarkets. This is the corollary of the "law of one price" in a dynamic context. Products in the same market will have the same unit price. As the market price changes, the unit price of all products in that market will also change (albeit subject to a degree of noise). So monitoring changes in the unit price of housing services across space – if indeed it can be achieved – will produce a more robust method of identifying submarket boundaries. Where there are shifts – up or down – in the rate of price appreciation of a unit of house service, there is a boundary between two different markets. The converse, of course, does not necessarily hold. It is possible for two areas to have similar rates of house price appreciation, yet belong to separate submarkets. This is true, however, of all indicators of submarket breaks. Should this be a concern, the researcher can always consider multiple criteria – while it is possible that two areas will pass one criteria for unified market and still in actual fact belong to separate submarkets, it is far less likely that they will pass multiple tests.

A further advantage of making the changes to the unit price of housing measurement of submarket variation, is that house price inflation is of interest in, and of, itself. Rising prices indicate that demand exceeds supply. Falling prices indicate that supply exceeds demand. Only where demand equals supply are prices stable. So variations in price appreciation reveals where excess demand (supply) is most prominent.

2.5 The key criteria

The approach we seek to develop in an attempt to meet the criterion set out in the introduction, and which avoids the methodological pitfalls listed above, is one which plots surfaces in geographical space of the rate of change over a prolonged period in the unit price of housing services, and then clusters dwellings according to their contiguity and their similarity in rate of price change.

2.5.1 Why an inflation surface?

By estimating the rate of appreciation for every point in geographical space we give ourselves the option to group or average these values according to any criterion we choose. This flexibility would allow us, for example, to express our description of submarkets in terms of post code sectors, wards, Census output areas, parliamentary constituencies, parish boundaries, districts, counties, or regions. We simply take the average or median for whichever administrative unit we prefer.

2.5.2 Why a prolonged period?

The longer the period considered, the smaller the likelihood that differences in the rate of change will be due to fleeting market frictions, and the greater the chances that observed differences reflect genuine submarket boundaries. It would also allow researchers to consider the possibility that market boundaries have shifted.

2.5.3 Why cluster by contiguity?

Plotting a surface of the rate of change in the price of a unit of housing will allow one to see instantly the pattern of submarkets across space. This may reveal areas that are not contiguous to have similar rates of change and hence belong to the same submarket. But the further two areas are from each other, the less likely a particular buyer is to actually include them both in their choice-set, even though price trajectories are similar. Also, for the reasons discussed above, for policy and planning purposes it is often useful to identify contiguous areas. Once defined, these areas can always be combined non-spatially if desired (ie area A and area C are defined separately because they are not contiguous, but may be considered part of the same submarket).

2.6 A time-location value signature model

The basic model developed here is an extension of the work by Fik et al (Real Estate Economics, 2003) and Pryce (2004b). Fik et al (2003) included latitude and longitude interactions in a regression of dwelling attributes on price and then used this equation to derive a "location value signature" for every point in the area covered by the data. Pryce (2004b) extends the Fik et al (2003) approach to include continuous time interactives (interacted with both attributes and latitude and longitude to account for movements and twists in the price surface over time) complemented by year and season dummies to capture step shifts in attribute values. The end result is a "***time-location value signature***" (TLVS) model.

Our plan is to utilise the Fik et al (2003)/Pryce (2004b) method to ascribe to each property in the data a unique location value for each month. In other words, we estimate how the value of each property changes over time. From these estimated price changes we can then interpolate across areas where there are few transactions using nearest-neighbour surface fitting techniques.

3 Data Audit

3.1 Introduction

Ultimately the deciding factor in the analysis of submarkets is the availability of data. However ingenious or precise our method, if data restrictions do not allow us to apply it to areas of greatest interest, or indeed to roll it out for the whole of the United Kingdom, then it's unlikely to be worth further investment. In this chapter we briefly summarise the range of data on offer and our rationale for selecting the Nationwide data as the final basis for the model.

3.2 HM Land Registry

3.2.1 Pros

The great advantage of the HM Land Registry data is that it includes all property transactions and is fully spatially coded (though there tend to be a fair proportion of properties with missing address details). It therefore has the greatest geographical coverage and provides observations at the greatest available spatial density. It is also available for a reasonably large number of years (from 1996 onwards) and is regularly updated.

3.2.2 Cons

The greatest disadvantage is the absence of attribute variables. Property type (detached, semi-detached, terrace or flat) is recorded, but there are no measures of size such as number of bedrooms or price per square metre.

3.3 Council of Mortgage Lenders

3.3.1 Pros

House prices data is available from the Survey of Mortgage Lenders for twenty five years or more and has a fair selection of property attribute variables (including type, bedrooms, and garage). In recent years, the survey has moved to almost one hundred percent of Council of Mortgage Lenders member mortgage transactions and therefore provides potentially large numbers of observations on almost all areas of the United Kingdom.

3.3.2 Cons

In the data available to us, only the last couple of years had full spatial coding (ie the post code unit of the dwelling). As a result, the cumulative inflation surface could only be estimated for a very short time span. In future years, however, the data could form the ideal platform for time-location value signature analysis.

3.4 Nationwide

3.4.1 Pros

The Nationwide data has precise spatial coding. It is also available in spatially coded format for a reasonable time period (1996 onwards), and includes a rich array of dwelling attributes. Most useful of all is the fact that it includes a record of total floor-area (measured in square metres).

3.4.2 Cons

The main disadvantage with the Nationwide data is that it only records mortgage transactions with a particular lender. To some extent this is compensated for by the fact that our proposed analysis is fundamentally spatial and so we are able to bolster the data with a good range of location indicators that pick-up peculiarities of the surrounding neighbourhood, and hence minimise the effect of sample selection bias. The other issue is sample size. In some areas, the number of observations is relatively small, but this can be compensated for by building a model that includes the full range of years available and that spans a sufficiently large geographical area.

3.5 Conclusion

On balance, it was decided that the Nationwide dataset was appropriate. After spending considerable time trying to develop methods that could be used with the various data sources, the Nationwide data emerged as the preferred option. All the results that follow are based on Nationwide data, but we should emphasise that the techniques we develop could in principle be applied to other data sources, particularly as these other measures are continually improving over time.

4 An Application to Kent

4.1 Introduction

Perhaps the best way to explain the approach we have developed is to apply it to a particular area. In this chapter we therefore attempt to explain the econometric and Geographical Information System (GIS) methodology by way of illustration. It will become clear as we proceed that the approach is in fact made up of a series of steps and that each of those steps is useful in its own right because each unveils some aspect of the nature of the housing market. The chapter will begin by discussing briefly the criteria for selecting the geographic area. A price per square metre map is then presented and discussed, followed by a constant quality price per square metre map. An inflation surface is then presented. Finally, an example of how submarkets could be defined from this surface is given based on grouping postcode sectors.

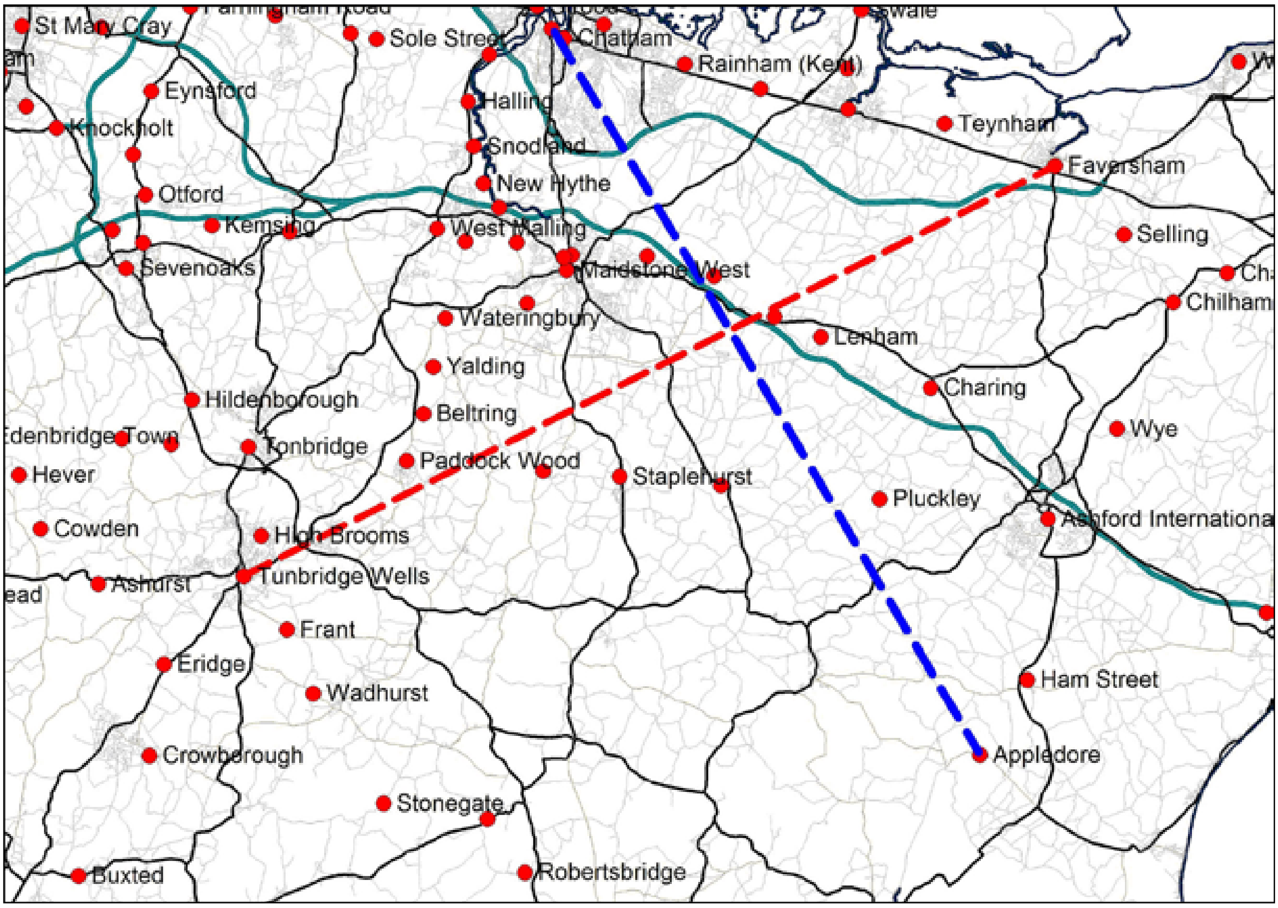
4.2 Selected area

The Nationwide data is available for most of the United Kingdom over the period 1996 to 2004. So the range of possible areas to choose from for illustrative purposes is immense. Given the policy focus on the South East, this seems the most appropriate region. However, this still leaves us with a relatively large land mass and one that is of unusual shape (there's a hole in the middle called Greater London!). Because of the doughnut-like shape of the South East, there are difficulties in fitting a single surface to the whole area without including London. Also, there are advantages to focusing on a much smaller area, such as a particular county, to illustrate the spatial precision of the estimates.

So ideally, we wanted an area roughly the size of a typical county with a range of natural and man-made features which included coastline, an estuary or major river, motorways, rail, and a mixture of proximity to a major city and rural land. In the end, we opted to fit the model to an area of land that incorporates Kent and East Sussex. However, the area was too large to fit on A4 paper at a scale that would allow the reader to identify place names and key landmarks. Consequently, in the maps below we only present a selection of the area ranging from St Mary Cray in the top left, to Whitstable in the top right, to St Mary's Bay in the bottom right, to Buxted in the bottom left.

The area is illustrated in Figure 4-1, along with the routes of the two cross sections highlighted by the two dashed lines, one red (running from Tunbridge Wells railway station to Faversham station) and one blue (running from Rochester railway station to Appledore railway station). It is along these lines that we shall take cross-sections to examine the shape of the various estimated surfaces. Place names on the maps refer to railway stations as these allow us to identify precise points in space (rather than the vague area often alluded to by the name of a town or city).

Figure 4-1 Map area selected for illustration of the model with cross-section routes highlighted



Source: Derived from Ordnance Survey Meridian 2 map information (© Crown Copyright/database right 2005. An Ordnance Survey/(Datacentre) supplied service). Place names refer to railway stations.

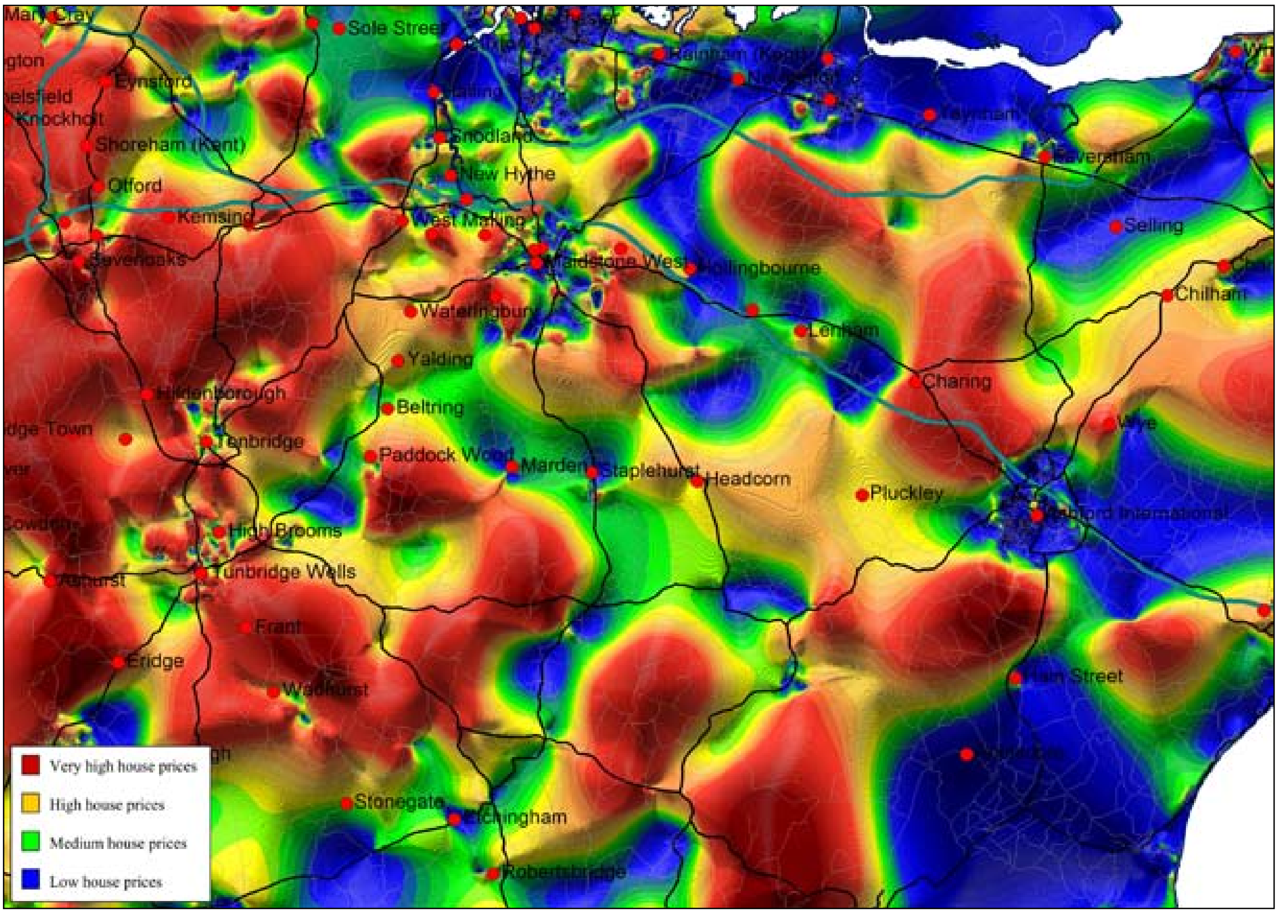
4.3 Price per square metre

Analysing variations in price per unit of space is the simplest way of looking for submarkets. This is based on "The Law of One Price" – the micro-economic theoretical dictum that each market can have but a single prevailing price. It follows that persistent price differences across areas reflect the existence of different markets. The first step in our analysis, therefore, is to produce a detailed price per square metre map to give us an initial glimpse into where submarket boundaries may lie.

The result is presented in Figure 4-2. The map depicts huge variation in price per square metre even across a relatively small geographical area. In one sense, the peaks may be considered as different submarkets from the troughs, but the difference may be due to clusters of property types. For example, price per square metre may be higher for houses than flats, but a buyer with a given budget may legitimately be weighing up whether to buy a large flat or a small house. The two property types are therefore fairly close substitutes for that particular buyer.

A more demanding definition of submarkets would require us to control for attribute variations. In other words, how does the unit price of housing vary across space if we hold all other dwelling attributes constant (ie what would be the price per square metre for the same house placed at different locations on our map)? A stricter definition would require us to look at differences in the rate of change of the unit price of housing, holding all other dwelling attributes constant. But let's consider the intermediate definition first.

Figure 4-2 Observed price per square metre



Source: Derived from the authors' econometric model, estimated using Nationwide data (1996-2004), and Ordnance Survey Meridian 2 map information (© Crown Copyright/database right 2005. An Ordnance Survey/(Datacentre) supplied service).

4.4 Constant quality house price surface

The simple price per square metre approach assumes we are comparing "like-for-like". As such, price per unit of space may vary simply because the height of rooms or quality of build or type of dwelling varies. A more robust approach will attempt to control for variations in dwelling attributes. The great advantage of the Nationwide data over HM Land Registry records is that it records a rich array of information on dwelling attributes including floor area, number of rooms, number of bathrooms, type of dwelling (detached house, semi detached house, terraced house, country cottage, detached bungalow, semi detached bungalow, purpose built flat, purpose built maisonette, flat conversion, and maisonette conversion), age of dwelling, whether the dwelling has a garage (single or double), parking, central heating type (none, full gas, full electric, full oil, full solid fuel, part gas, part electric, part oil, part solid fuel) and the full postcode.

The Nationwide data was augmented with Mosaic, Hometrack and Ordnance Survey data to create a fairly detailed profile of the postcode unit of each and every property in the data, including the average distance between dwellings, the average footprint area of each property, height above sea level of the postcode centroid, and the proportion of dwellings in the postcode in each property type and age category (detached, flats, bungalows, semi-detached, pre-1920, 1920-1945, 1946-1979, post 1979). We also calculated the value of these variables for the nearest adjacent postcode to capture spillover effects across neighbourhoods.

The latitude and longitude interactions were calculated along similar lines to Fik et al (2003) up to the power of three. We also interacted these variables with a time variable (months since January 1996), year dummies (to account for step shifts over time), and district dummies (to account for step shifts in the whole surface across districts). The combination of these interactions and the aforementioned neighbourhood variables, gave the model sufficient flexibility to account for the variation in prices across space. Indeed, this is reflected in the adjusted R2 of the final model (reported in table 4-1) of over 85%. As in Fik et al (2003) a stepwise process was used to eliminate variables with significance levels below 0.001 (t ratio > 2.5).

Table 4-1 Time-location value signature model for Kent and East Sussex

Variable	Coefficient	t-ratio
Number of bathrooms	0.072314	22.08
Number of bedrooms	0.029749	15.36
Detached house	0.181338	40.83
Semi-detached house	0.046306	14.57
Detached bungalow	0.210457	34.72
Semi-detached bungalow	0.172146	25.09
Maisononette – converted	–0.05975	–5.55
Maisononette – purpose built	–0.15884	–6.76
Age of dwelling squared	1.41E-06	20.56
Floor area (m2)	–0.00453	–91.86
Garage – double	0.198636	32.64
Garage – single	0.101838	34.38
Parking	0.097155	31.37
Full gas central heating	0.079309	25.25
Full oil central heating	0.130107	14.25
Part electric central heating	0.025681	4.37
Part gas central heating	0.028321	4.33
Flat – converted	–0.18095	–33.74
Flat – purpose built	–0.08324	–14.81

Table 4-1 Time-location value signature model for Kent and East Sussex (continued)		
Variable	**Coefficient**	**t-ratio**
Area effects:		
Average distance between dwellings	0.004028	19.17
Dominant number of bedrooms	–0.01292	–7.93
Proportion of properties that are bungalows	0.229652	18.5
Proportion of properties that are "other flats"*	0.079409	8.39
Proportion of properties built 1920-45	–0.03987	–4.25
Proportion of properties built 1946-79	–0.12886	–20.15
Proportion of properties that are purpose built flats	0.088702	10.74
Proportion of properties that are detached	0.104938	10.53
Proportion of properties that are semi detached	0.060762	7.68
Proportion of properties that built pre1920	–0.09673	–14.98
Average property size (m2)	0.002604	22.57
Nearest adjacent postcode: dominant no. bedrooms	–0.00703	–4.55
Nearest adjacent postcode: % other flats	0.039046	4.66
Nearest adjacent postcode: % 1920-45	–0.04549	–5.31
Nearest adjacent postcode: % purpose built flats	0.043475	6.54
Nearest adjacent postcode: average size	0.001148	11.64
Nearest adjacent postcode: % detached	0.031999	3.88
Nearest adjacent postcode: % semi detached	0.024423	3.57
+ x,y,t, year and district interactions		
Number of observations	35,904	
Adj R-squared	0.8514	

Dependent variable = price per metre squared (measured in logs).

* "Other flats" – ie other than purpose built flats.

Source: Author's regression model estimated using Nationwide data (1996-2004)

We then used this model to predict the value of price per square metre for each property in the data for December 2004, holding dwelling attributes (but not spatial/neighbourhood factors) constant. These predictions were plotted as a surface and mapped onto the standard geographical details used in the previous two maps. The first of the maps (Figure 4-4) shows the full area covered by the model. The second of the maps zooms in to show only the area selected for comparison (Figure 4-5). Perhaps surprisingly the results show a high degree of variation in price per square metre which suggests that much of the variation in price is down to the effect of location (rather than dwelling size or type).

The extent of the variation is highlighted if we take a cross section of the inflation landscape. The course of the cross sections is illustrated in Figure 4-1 by the two dashed lines, and the results are graphed in Figure 4-3, first for the section running from Tunbridge Wells railway station to Faversham station and second for the section running from

Rochester railway station to Appledore railway station. Both graphs illustrate the huge variation in constant-quality price per square metre. The graphs also illustrate the flexibility offered by combining econometric techniques with Geographic Information System analysis tools. Cross sectional extracts of this kind could be taken for any two points on the map.

Figure 4-3 Cross sections of the constant quality price per square metre (December 2004)

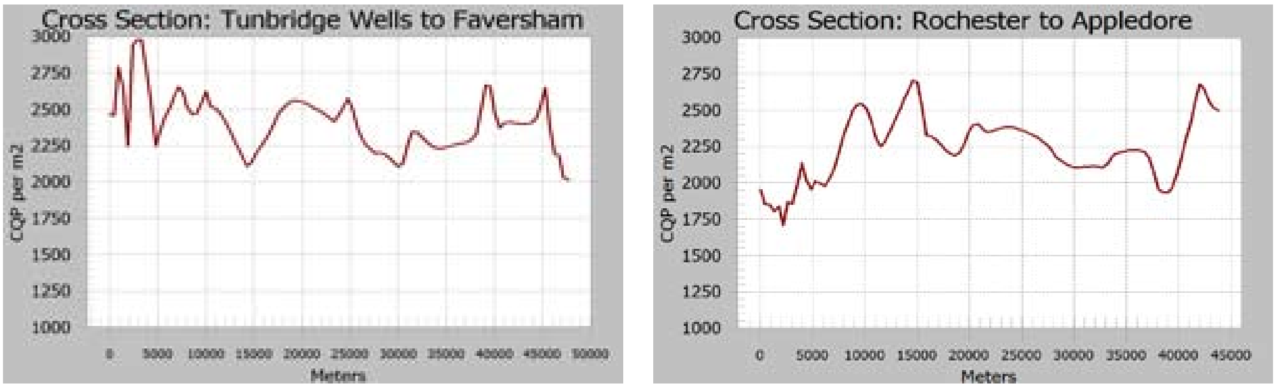
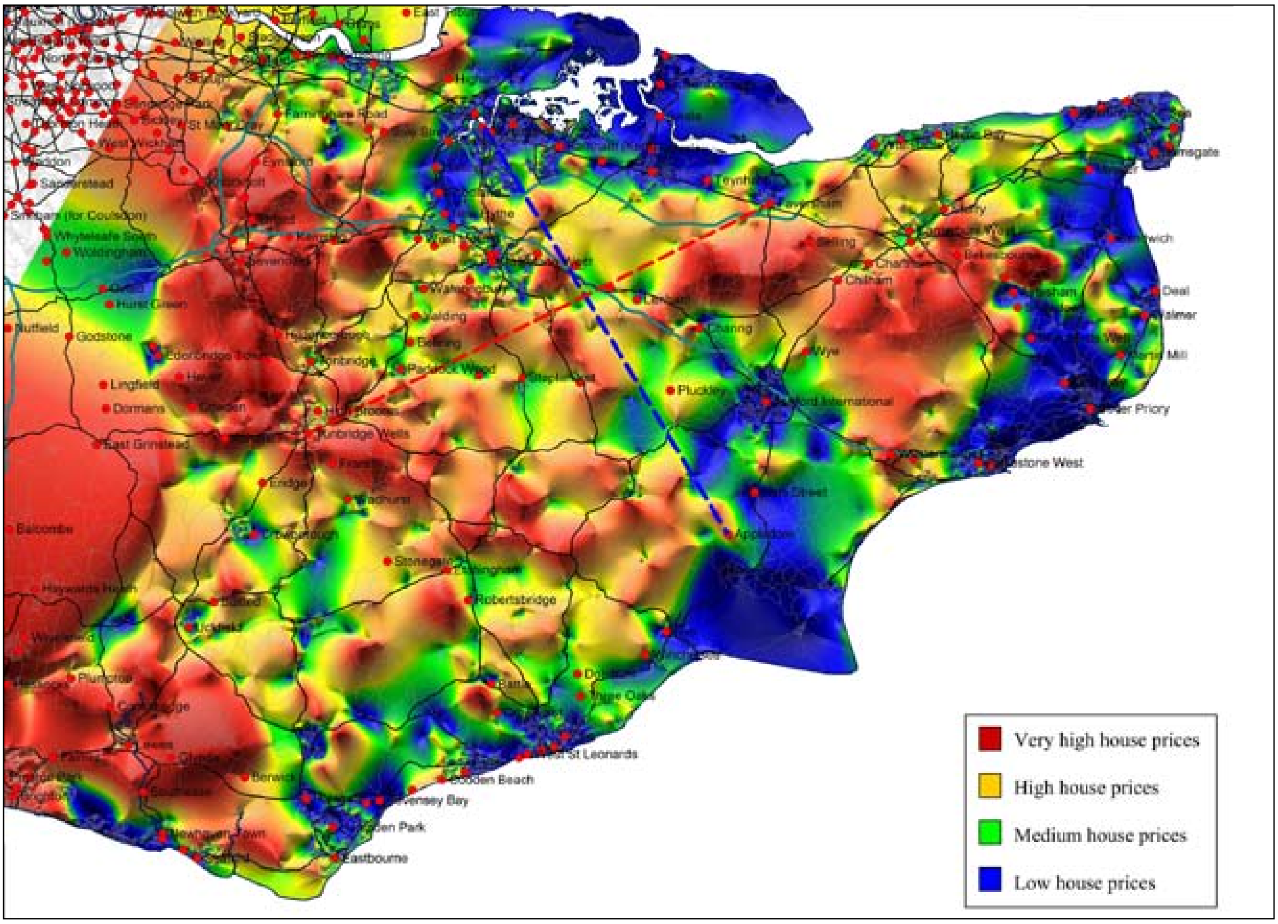
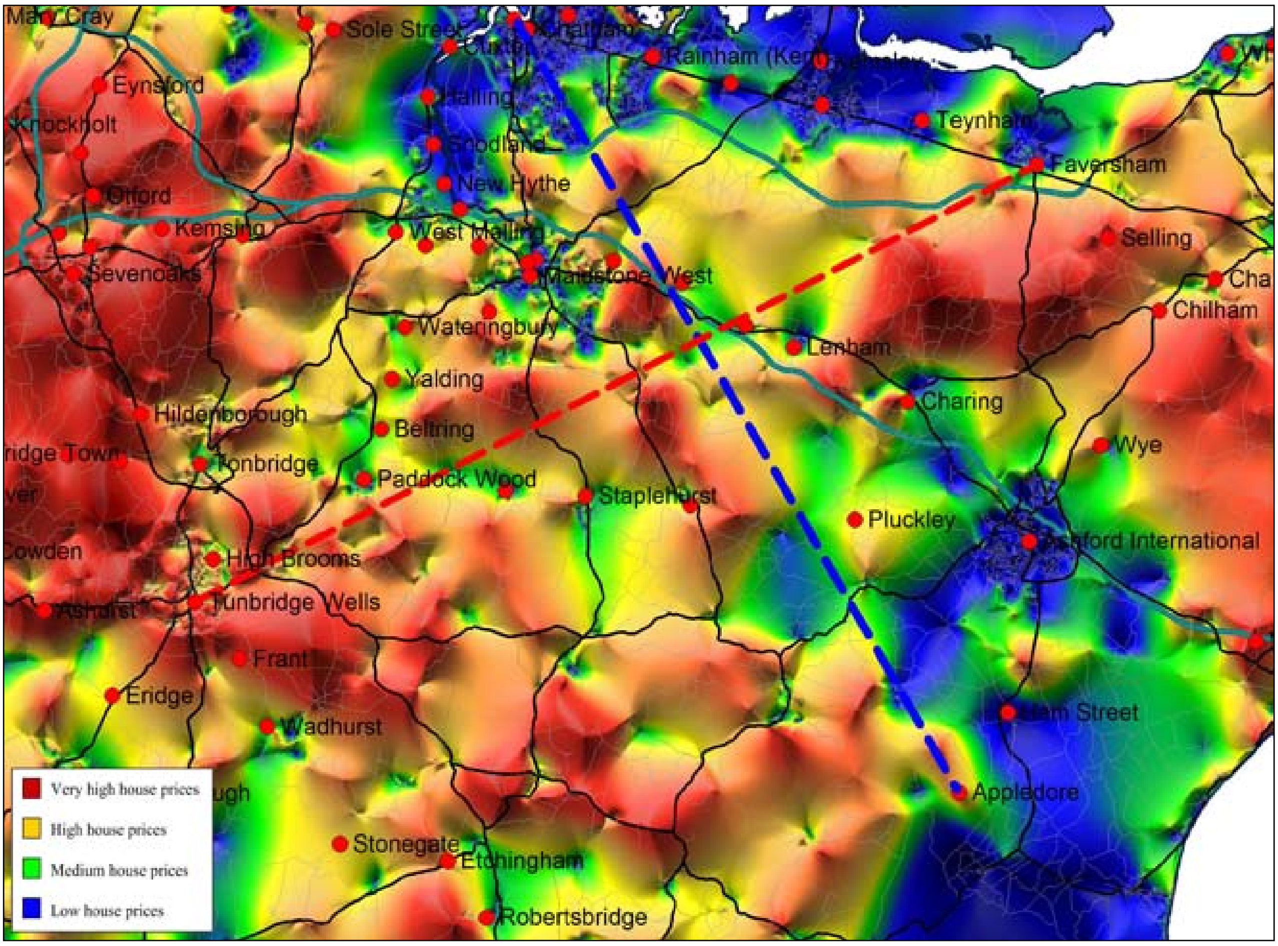


Figure 4-4 Constant quality price per square metre – full area covered by the model (December 2004)



Source: Derived from the authors' econometric model, estimated using Nationwide data (1996-2004), and Ordnance Survey Meridian 2 map information (© Crown Copyright/database right 2005. An Ordnance Survey/(Datacentre) supplied service).

Figure 4-5 Constant quality price per square metre (December 2004)



Source: Derived from the authors' econometric model, estimated using Nationwide data (1996-2004), and Ordnance Survey Meridian 2 map information (© Crown Copyright/database right 2005. An Ordnance Survey/(Datacentre) supplied service).

4.5 House price inflation surface

The ultimate purpose of the model was to estimate house price inflation for every point in space. This is an ambitious aim given that, at the moment, official house price inflation figures are only available at relatively high levels of aggregation, such as standard regions or local authorities. So the question is how can a model such as this be used to derive an *inflation* surface?

We have already explained how we derived the *price* surface for December 2004. The next step was to estimate a comparable constant quality price surface for an earlier time period. The relatively long time period for which spatially coded Nationwide data were available allowed us to estimate a second start of our data series (January 1996) – nine years and eleven months prior to our December 2004 surface. Given the argument put forward above that the longer the time period considered the better (to iron-out any temporary glitches in market adjustment) this seemed to be the optimal period for the estimation of the second surface because it gave us the largest time-span possible.

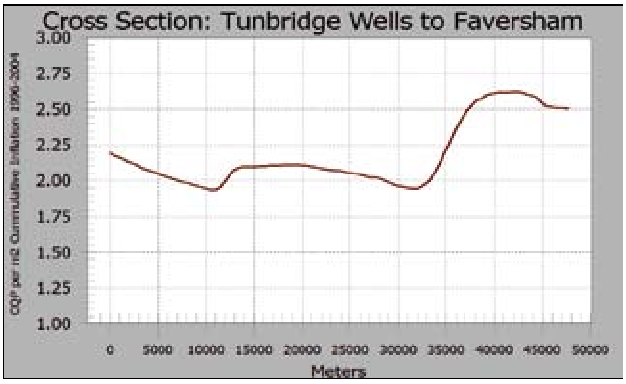
The difference between the two surfaces was then calculated by subtracting the inflation estimates at each point in space in the January 1996 surface from the inflation estimates at the same set of points in December 2004. This yielded a set of spatially coded price changes for which we could plot an inflation surface representing cumulative changes in constant quality house price over the nine year eleven month period between January 1996 and December 2004.

Our approach should not only prove useful for the current project (because variations in price inflation may be considered as another way of identifying submarket boundaries on the basis that the prices of properties that are considered close substitutes – and hence in the same submarket – should move in unison) but may also have many important alternative uses. For example, the inflation model could be used to simply identify "hotspots" in the region or indeed those areas that are struggling. This would help inform government and house-builders as to the most appropriate location for new housing.

The cross sections and full inflation surface are plotted in Figure 4-6 and Figure 4-7 respectively. It is immediately clear from comparison of the inflation surface map with the price levels maps that the most expensive areas have not necessarily experienced the largest increase in value. We can also observe considerable variation in price appreciation even after dwelling attributes have been controlled for. This suggests that different submarkets do indeed exist in the Kent and East Sussex areas.

We now have the basis for defining submarkets. Areas with similar rates of price change can be classified as belonging to the same submarket. The only question that remains is how – in practical terms – do we decide whether a particular point in space belongs in submarket A or submarket B? It is to this question we shall now turn.

Figure 4-6 Inflation surface plot (with relief shading)



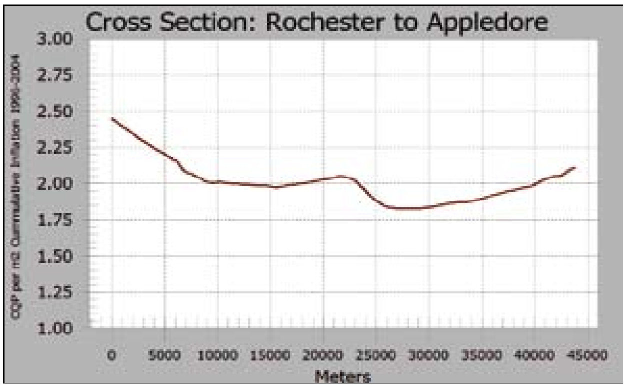
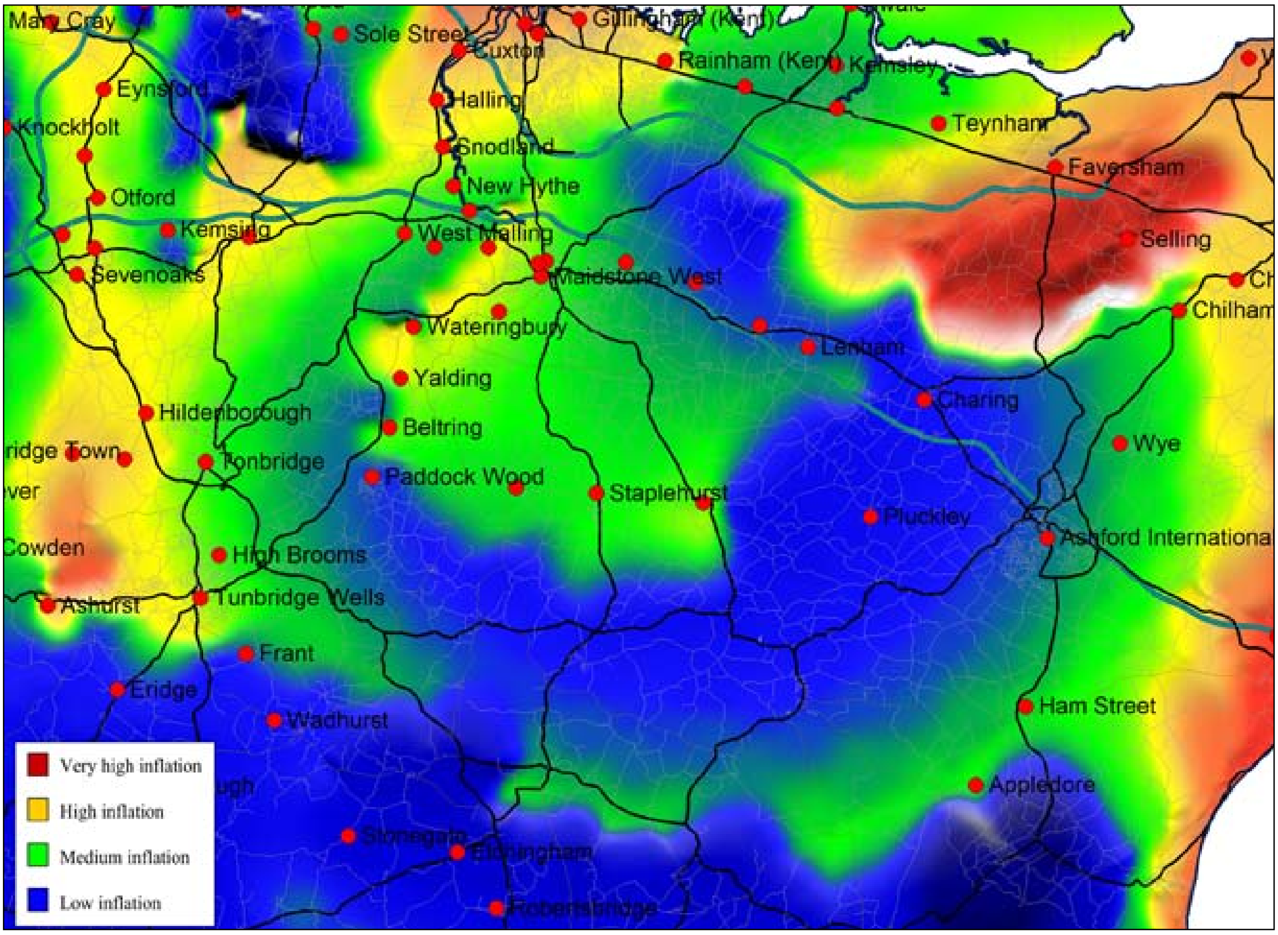


Figure 4-7 Constant quality house price per square metre cumulative inflation surface (1996-2004)



Source: Derived from the authors' econometric model, estimated using Nationwide data (1996-2004), and Ordnance Survey Meridian 2 map information (© Crown Copyright/database right 2005. An Ordnance Survey/(Datacentre) supplied service).

4.6 Clustering by postcode sector

Cluster analysis is the name given to a broad set of mathematical algorithms that facilitate the grouping of observations in a dataset into meaningful categories. It has been used in a variety of different situations across the full range of applied academic disciplines. For example, cluster analysis has been used to group together genes of similar types, helping scientists to derive order out of the bewildering complexity of gene sequences.

The groups of observations you end up with depend crucially on three factors. First, the clustering algorithm employed. It has been said that there are as many cluster analysis methods as there are researchers using cluster analysis! We have opted to use Ward's method, one of the simplest and most often used approaches.

The second important factor in determining the final groupings is the number of clusters one wants to derive. This is usually one of the inputs required by the cluster algorithm. If one decides on eight clusters (in our case, this would equate to eight submarkets) then the clustering procedure will keep dividing the dataset until each observation belongs to one of eight clusters. In each case, the algorithm will allocate the observation to the group with the most similar observations. How many groups should one opt for? Fortunately, there are procedures for helping us decide but even so, there is usually a degree of judgement required because such procedures rarely yield an unambiguous answer. For example, careful examination of the Dendrogram (discussed below) might suggest that the data either group naturally into six clusters, or into ten clusters, or into forty-four clusters. The researcher will then make a decision based on which is of most practical use. Forty-four may be too many to be of much use, and six too few, so one might end up opting for ten clusters.

The third key input into the clustering process is the criteria by which observations are to be clustered. For example, if one were to cluster species of flower into a handful of broad categories, one could choose from a range of different clustering criteria, such as plant height, number of petals, colour and root depth. Different criteria would lead to different groupings. Normally, one's choice of criteria would be governed by the theoretical and/or practical motivations that underpin one's desire to cluster the observations in the first place. Note that several criteria can be used.

In the case of submarkets, while cluster analysis has been applied by a number of researchers, the choice of clustering criteria has not been thought through with sufficient depth to make the derived submarkets have a clear interpretation. For example, simply clustering properties by attribute types is theoretically unsound since we have no way of knowing how the market rates different attributes in terms of their substitutability. Having reached the point of deriving constant quality price inflation estimates, we believe that we are now in a position to apply cluster analysis in a way that more closely adheres to the theoretical foundation established in the submarkets literature.

The three variables we need to use in our cluster analysis if we want to derive spatial submarkets are latitude, longitude and constant quality house price inflation. Latitude and longitude are included since we want properties that are far apart to have less of a chance of being included in the same submarket. If we are not interested in the spatial aspect of submarkets, we can omit the latitude and longitude variables and therefore do away with cluster analysis altogether – we simply group properties by band of house price inflation (ie whether they are in the blue, green, yellow or red areas of Figure 4-7) and this amounts to drawing contour lines to group observations. As discussed earlier, aspatial approaches are likely to lead to delineations of submarkets that are of little policy use. Therefore, by including latitude and longitude, we know that, other things being equal, observations adjacent to each other are more likely to be placed in the same group, and we end up with a more useful set of boundaries.

Having decided on our criteria for clustering observations (constant quality house price inflation, latitude and longitude) we then needed to decide what to use as our observations. In principle, we could take each dwelling (or an even finer set of points on our inflation surface) as the initial units of our cluster analysis. This would lead to a very precise set of boundaries. Given that the spatial coding of our original observations was only down to postcode unit level (rather than individual address level) it would perhaps suggest a spurious degree of precision to group below postcode unit level. In practice, even postcode units would entail an excessively large set of calculations (beyond the power of our software), so to illustrate we use postcode sectors as the initial cluster unit. In other words, we first average-up the inflation estimates to postcode sector level, leaving us with 342 units, and we then group these sectors using Ward's clustering method.

Of course, by moving to postcode sectors as our unit of aggregation we have committed ourselves to an imperfect process (there is no reason why submarket boundaries should follow those of post code sectors) but by comparing these clusters with the aforementioned maps, it is likely that the submarkets derived from clustered sectors will be more than accurate enough for policy analysis at the regional level. By using post code sectors, the derived submarket areas can now be integrated with published socio-economic data (such as unemployment rates and population statistics). This is therefore a useful exercise since it also illustrates how the results from inflation surface modelling could be exported to a given set of standard administrative units. One should note, however, that we could have just as easily aggregated our inflation estimates up to ward level, or output-area level, or indeed any spatial level. This is the advantage of using a continuous surface as the basis of our clustering methodology.

To summarise, then, the sequence of steps used to derive submarkets from house price data on the South East of England are as follows:

1. Estimate the time-location value signature for two points in time;

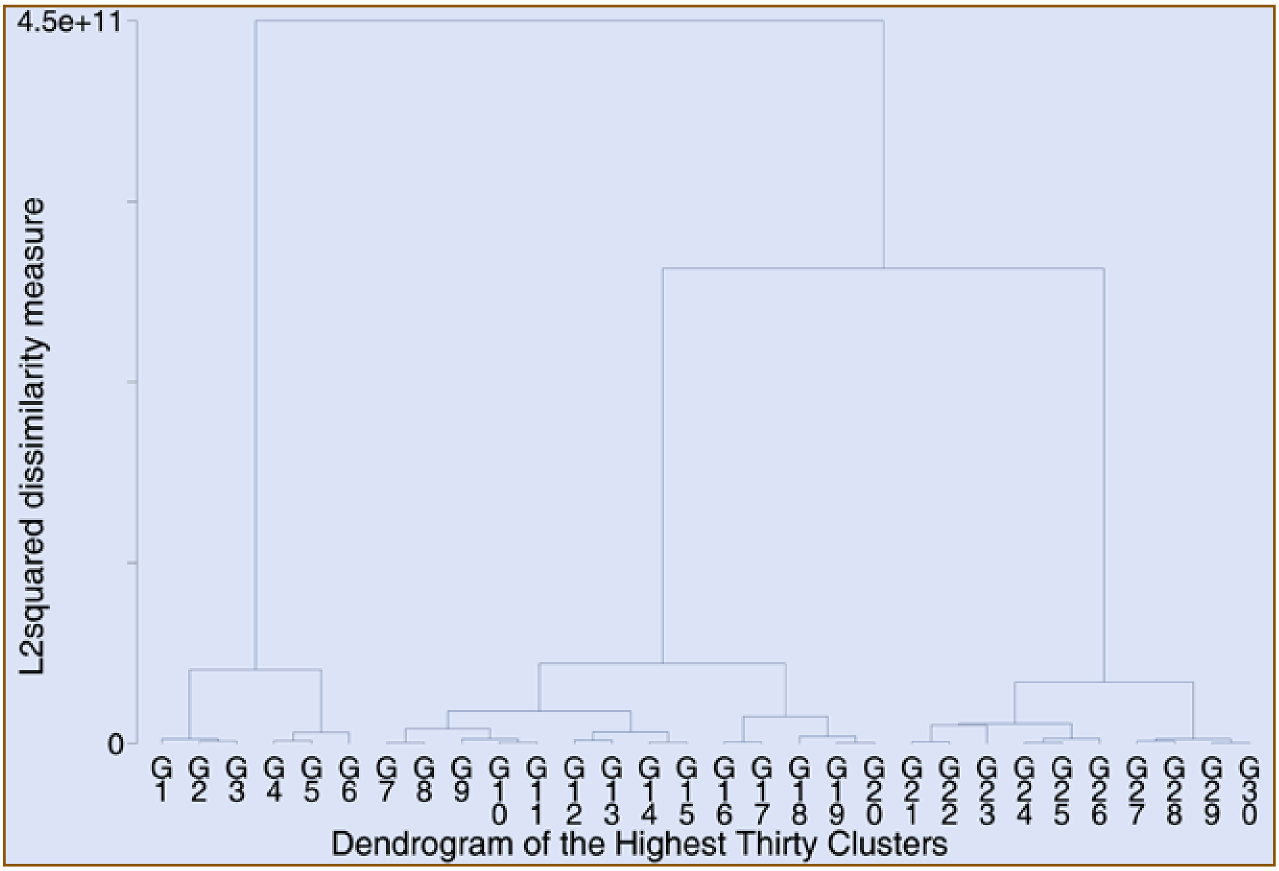
2. Compute the cumulative rate of inflation between the two periods and plot the *inflation surface;*

3. Calculate the average rate of inflation at the chosen level of spatial aggregation (we used postcode sectors);

4. Apply Ward's clustering method (to the 342 postcode sectors) using the preferred clustering criteria (we used three variables: latitude, longitude, and the estimated rate of constant quality house price inflation).

These four steps lead not to a single set of submarkets, but to a range of submarkets depending on how many we would like. This is demonstrated in the Dendrogram (Figure 4-8). Each postcode sector is clustered into an initial set of groups (see the first row of branches at the bottom of the Dendrograph – the more branches there are at a particular level, the more groups). These groups are clustered again into a smaller number of groups. The process continues until the desired number of groups is obtained. It is difficult to say without at least a degree of ambiguity what the optimum number of groups (ie submarkets) is, and in a sense that depends on the use to which they are going to be put. In some situations, one wants to derive a potentially large number of submarkets that reflect the boundaries of the individual neighbourhoods in which each property is located. On this basis we might select a hundred or more groups, and we would end up with one hundred very small submarkets. However, for other purposes, we might only be interested in defining five to ten larger submarket areas. And this is the approach adopted below. The Dendrogram offers us some support for this choice – the length of each vertical line measures the dissimilarity between the groups at any given level. The greatest degree of dissimilarity occurs at three groups (so there is a case for splitting the whole of Kent and East Sussex into just three submarkets). However, the scale of the dissimilarity is quite large, so there is also good reason to make the cut at six groups or even ten.

Figure 4-8 Clustering postcode sectors using Ward's method



To illustrate what would happen if we opted for the latter, the graph in Figure 4-9 plots the centroids (the centre point) of each postcode sector and colours them according to which of the ten groups the centroid belongs to. The next step (Figure 4-10) is to add-in the postcode sector boundaries and shade each of the postcode sectors according to the group it belongs to (using a separate shade for each of the ten groups identified from the cluster analysis). Figure 4-10 is plotted for the whole Kent and East Sussex area, with darker areas representing higher constant quality price inflation. It represents the final step in our estimation process – it is a map of submarket boundaries presented in terms of commonly used administrative areas (postcode sectors), yet based on an estimation procedure that is not contingent on the boundaries of those areas.

To allow the reader to compare the submarkets map of Figure 4-10 with the original inflation surface, we reproduce the latter in Figure 4-11. Comparing Figure 4-10 and Figure 4-11, it would appear that the clustering procedure has done a good job of disentangling the complexity of the inflation surface. It has managed to identify coherent areas (avoiding, for example, the production of submarket "islands" within larger submarket areas), while grouping sectors with broadly similar rates of price change.

Figure 4-9 Postcode sector centroids: colour coded by submarket (10 Groups)

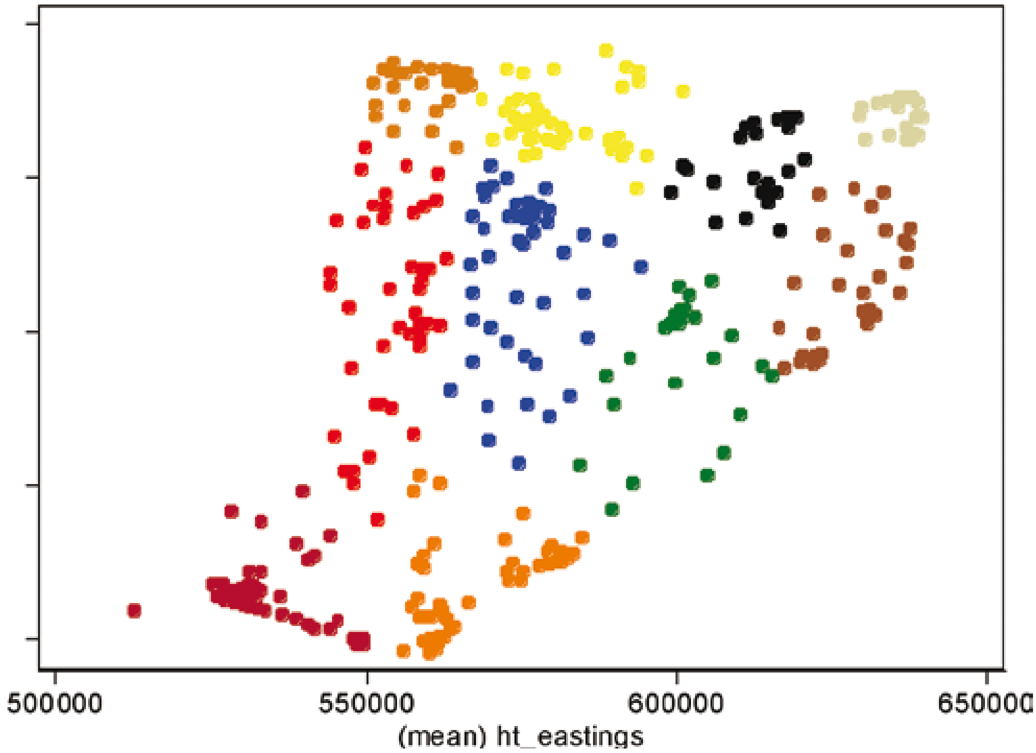
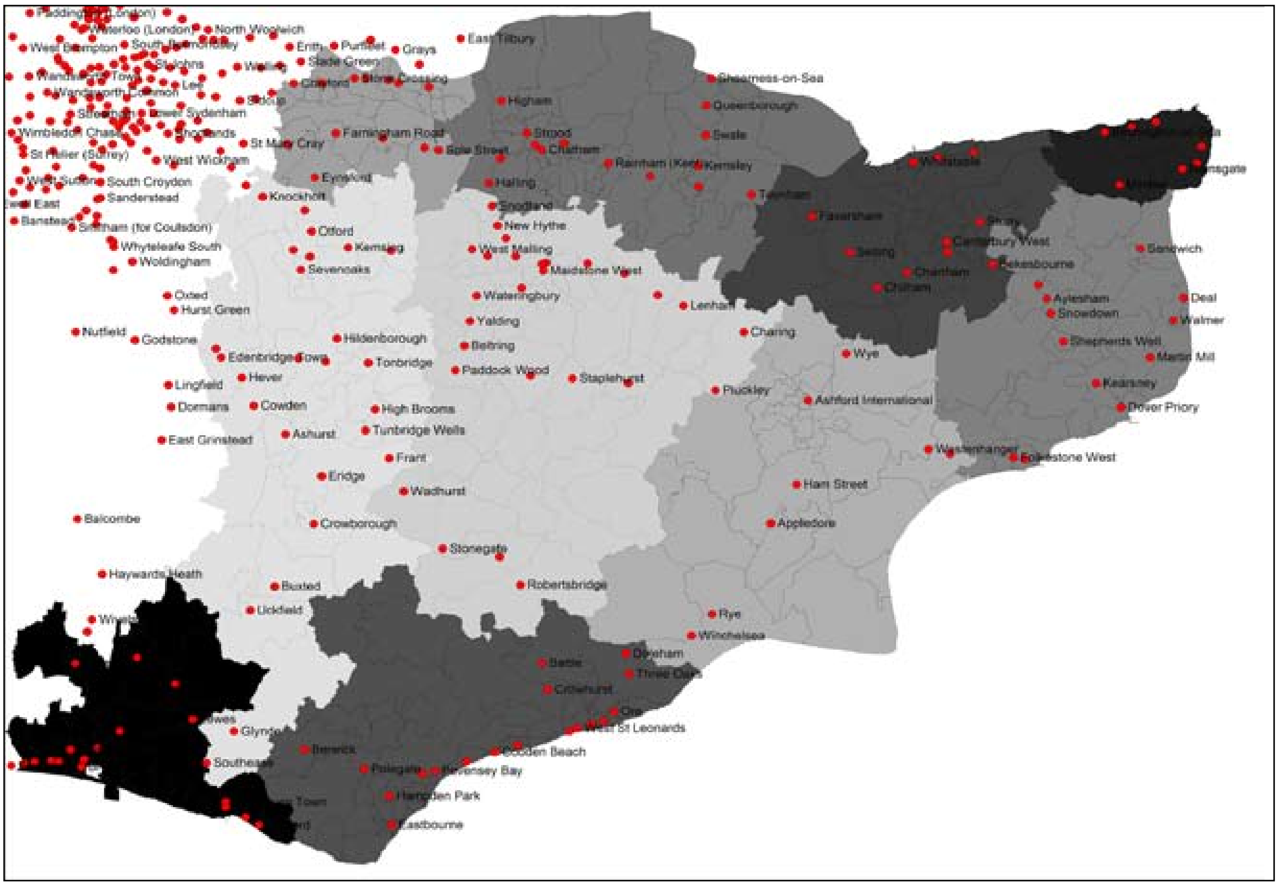


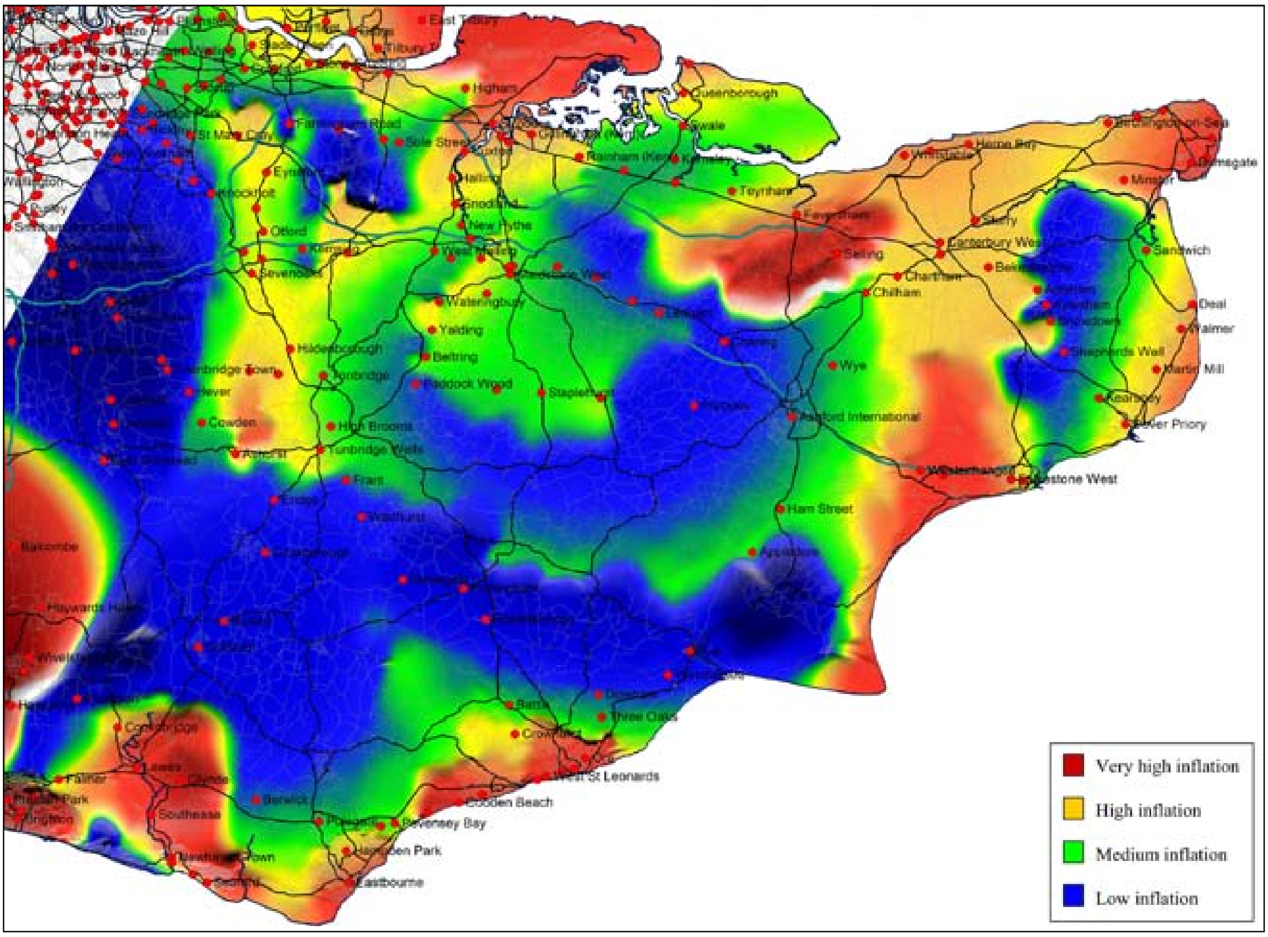
Figure 4-10 Postcode sectors: shaded by submarket



(10 Groups – Darker areas have higher constant quality price inflation)

Source: Derived from the authors' econometric model, estimated using Nationwide data (1996-2004), and Ordnance Survey Meridian 2 map information (© Crown Copyright/database right 2005. An Ordnance Survey/(Datacentre) supplied service).

Figure 4-11 Inflation surface plot for Kent & East Sussex



Source: Derived from the authors' econometric model, estimated using Nationwide data (1996-2004), and Ordnance Survey Meridian 2 map information (© Crown Copyright/database right 2005. An Ordnance Survey/(Datacentre) supplied service).

5 Future Development and Application

5.1 Introduction

In this chapter we consider how the techniques presented here could be applied in different settings. A key question might be how the model could be used to improve our understanding of the way in which submarkets affect the impact of new construction. There are also specific environmental and policy issues that could be explored using the time-location value signature approach. The broader research theme of neighbourhood well-being and regeneration could also benefit from the time-location value signature methodology as house prices ultimately reflect variations in the quality of life across space.

5.2 Impact of new housing supply and other spatial shocks

Pryce (2004) concluded that 'we know very little about the indirect price effects generated by new sites on existing housing markets at a micro scale'. Unfortunately, that is still the case. While the method suggested by Pryce and Gibb (2006) remains a potentially useful way of addressing this issue, it would perhaps be better to develop that approach within the broader framework of a time-location value signature estimation. One advantage is the facility that the time-location value signature offers for visualising the impact of new supply on the shape and position of the price per square metre surface and also on the inflation surface.

The same is true with regard to the analysis of the impact of other potential shocks to the housing system. For example, one of the variables we could add to our model is the estimated flood risk associated with each postcode unit in our data. As flood risk rises (due to rising sea levels, increased storminess and other consequences of climate change) there will be a house price impact. This effect is important because house price changes will change the levels of housing equity (and hence the wealth many households are relying on to fund their retirement), the distribution of wealth, and the propensity for mortgage default, particularly in areas like the South East where debt gearing is relatively high.

Note also that one of the variables in our model is a measure of housing density (ie the average distance between dwellings in each postcode unit). We have not held this constant when predicting our constant quality price index since we have classified it as a location variable rather than a dwelling attribute. However, there is the potential to model the effect on our house price surfaces of increasing density in particular areas, though a more complete analysis would require the model to be extended to incorporate displacement effects. If density rises in area A, price appreciation may fall in that area, but prices in area B, which is a low density area, may actually rise because the new construction has caused low density housing to become relatively scarcer.

Similarly, the model could be used to quantify the impact of school performance on house values.

5.3 Council tax revaluation

In the preceding empirical analysis we have used price per square metre as the dependent variable. However, if instead we used the total selling price as the dependent variable, the model could be adapted to become an Automatic Valuation Model (AVM) and used to estimate the value of properties for the purposes of Council Tax revaluation or property tax reform.

5.4 Trajectories in neighbourhood well-being

Inequalities in quality of life between neighbourhoods and their diverging trajectories have become a core concern of social and urban policy in the United Kingdom. The National Strategy Action Plan (2001) of the Social Exclusion Unit[3], for example, begins with the following statement, 'Over the past twenty years, hundreds of poor neighbourhoods have seen their basic quality of life become increasingly detached from the rest of society. People living just streets apart became separated by a gulf in prosperity and opportunity'. The document goes on to list the core causes of neighbourhood decline (including economic change and the decline of old industries and the associated changes in the labour market and patterns of unemployment, family breakdown, and changes in housing demand), and how government policy in the past has failed to address these problems, sometimes even exacerbating them. The Social Exclusion Unit then set out two long term goals for government policy: first to reduce worklessness, crime and improve health, skills, housing and the physical environment of the poorest neighbourhoods; second, to narrow the gap between the most deprived neighbourhoods and the rest of the country. These goals are to be met through coordinated "joined-up" initiatives and partnerships rather than ad hoc, independent interventions.

Both the diagnosis and proposed solution raise important questions for neighbourhood research. The first is how one should measure the quality of life associated with living in a particular neighbourhood. The second is the definition and delineation of "neighbourhood". Third, is the question of how one gauges the relative importance of different drivers of neighbourhood change and the related question of how regeneration initiatives should be evaluated in a way that leads to true evidence based policy. The time-location value signature techniques developed here have been presented in the context of defining housing submarkets. However, they may also have the potential to be applied to the three broader questions and it is to these we shall now briefly turn.

5.4.1 How should we measure spatial variations in the quality of life?

This question is important because it affects which areas are identified as having low quality of life, and how we understand the pattern of neighbourhood wellbeing to have changed over time. For example, an area may have low levels of crime and unemployment, but people may not want to live there because of the lack of amenities or poor schooling. Any single measure of wellbeing will be inadequate. Even deprivation indices are intrinsically problematic because they are faced with the unavoidable problem of how different factors should be weighted (see, for example, Jonathan Bradshaw's assessment of the English Indices of Deprivation, Neighbourhood Renewal Unit 2003). Also, training the dynamic movement of particular neghbourhoods is difficult because the indices are not designed to measure change over time, rather they give a snapshot of deprivation at one point in time.

The time-location value signature method could offer an alternative (and to some extent, complementary) measure of wellbeing based on the *value people are willing to place on locating in an area*, one that utilises the weights that residents *themselves* place on the different factors that affect quality of life, rather than the theoretically derived weights used by the designers of deprivation indices. This method utilises data that is frequently updated, available for an extensive historical time period, available for most areas of the

[3] A New Commitment to Neighbourhood Renewal: A National Strategy Action Plan (2001) – http://www.neighbourhood.gov.uk/publications.asp?did=85

United Kingdom, and includes a very large number of observations recorded with unique and precise spatial identifiers.

The maps presented in chapter 3 are related to well-being because the current value of a house reflects not just its physical construction and size but also the perceived value of locating in a particular neighbourhood. Indeed, even after we had controlled for variations in dwelling attributes, there remained considerable variation in prices due to the value of particular locations. This location value is derived from the perceived quality of life associated with living in that location (in terms of school quality, safety from crime, access to amenities, access to employment, environmental quality, area aesthetics etc) and the expected capital gain from anticipated house price inflation. Note that capital gain is itself contingent on expectations about the future desirability and quality of life in the neighbourhood, and has an important effect on the affluence of home owners. Falling house prices can severely restrict the mobility of residents, whereas rising house prices create valuable equity which owners can draw upon to fund refurbishments, health care, education and other expenditures that have a significant effect on life outcomes and wellbeing. In principle, if one can control for key differences in dwelling type, then any remaining differences in house prices between localities will reflect differences in (perceived/expected) quality of life (plus a random error term). This allows us to compare how quality of life within a neighbourhood has faired over time relative to the average quality of life of the region as a whole.

Perceptions of a neighbourhood are themselves important since the stigma or prestige associated with a particular location may effect the chances of getting a job and other life-outcomes (Atkinson and Kintrea 2001), but they have also an important empirical property in that they allow us to incorporate into our indicator of neighbourhood wellbeing factors that would otherwise be difficult to measure (such as aesthetic factors, noise levels etc) but which affect the perceived quality of an area. These factors, while observationally elusive, will affect the sale price of dwellings in the area and so will be captured in our derived measure of location value.

An important subsidiary question to how we measure the quality of life is how we define and delineate neighbourhoods. Galster (2003, p.154) defines a neighbourhood as 'the bundle of spatially based attributes associated with clusters of residences, sometimes in conjunction with other land uses'. Households consume neighbourhoods 'through the act of occupying a residential unit and using the surrounding private and public spaces, thereby gaining some degree of satisfaction or quality of residential life' (Galster 2003, p.155). Galster explains that,

> 'In order for potential consumers to make bid offers for a commodity they must have some modicum of information about the quantity and quality of that commodity and what likely benefit they would receive from its consumption. Real estate markets have been shown to meet this criterion for a large number of spatially based attributes. Indeed, this is the foundation of over three decades of empirical work estimating 'hedonic indices'. These studies have shown that [spatially based attributes] … can be converted into a measure of household consumers' willingness to pay.'
> (Galster 2003, p.155).

Galster goes on to argue that pricing of individual attributes can be problematic because of market imperfections (such as insider-dealing and asymmetric information). However, over a prolonged period, and when a sufficiently large number of observations are considered, temporary frictions in the market should even themselves out, and relative differences in the location value of different neighbourhoods will closely reflect real differences in the quality of life associated with those localities.

5.4.2 How can we define neighbourhoods?

Galster's useful conceptualisation of neighbourhoods does not lead to a set of unambiguous guidelines for how neighbourhood boundaries should be defined. The problem is aptly illustrated by the Social Exclusion Unit statement that 'People living just streets apart' can have very different levels of affluence and opportunity (Social Exclusion Unit, 2001, p.7). This means that an area defined by administrative boundaries, such as a local authority or ward, can be characterised by a significant degree of heterogeneity. This is similar to the problem that has been discussed at length in the submarkets literature, alluded to in Chapter 2 and reviewed in detail in Pryce (2004). In other words, where one reads "submarket boundary" can one read "neighbourhood boundary", particularly when one is defining submarkets at very small spatial scales? The correlation of the two has yet to be demonstrated but it would be a worthwhile and interesting avenue of further research to improve our understanding of how neighbourhood boundaries correlate to, and interact with, submarket boundaries.

5.4.3 What are the true drivers of regeneration?

The need for evidence based policy is reinforced by the apparent failure of previous initiatives to regenerate localities. Bramley et al (2000) suggest that existing policy responses to problems of low demand (such as renewal areas, empty property strategies etc) have had limited success or have simply been inappropriate. Earlier reviews of policy initiatives (such as Robson et al 1994) have also drawn fairly negative conclusions, as have more recent attempts at evaluation. Gripaios (2002, p.568) concludes that, 'despite years of policy intervention, there is little evidence that the situation is improving'. More project-specific evaluations are often more positive about the achievements of intervention, but Dabinett et al's (Department for Environment, Transport and the Regions, 2002, p.54) review of the evidence base for partnership based regeneration policy warns that, 'a substantial proportion of relevant literature is designed to promote the concept of partnership working, and is not based on empirical research'.

Similar questions were raised by Dabinett et al over the quality of research evaluating other aspects of regeneration policy. Regarding area-based regeneration initiatives, Dabinett et al note that 'a considerable body of evidence is available examining the practice of area-based regeneration and the specific impacts that individual initiatives have achieved', but 'much of the evidence base relates to improvements in the quality of physical environments and changes in employment and unemployment for local residents. There is far less on regeneration outcomes such as improving health and education or reducing crime' and, 'there has been little clear evidence from the evaluations of past area-based initiatives about who benefited from the interventions undertaken'. There are fundamental methodological problems because 'much of the evidence focuses on individual projects with insufficient attention paid to context; this tends to inhibit the transfer of reliable findings across projects'. A particular problem is how to disentangle the effects of intervention from changes in socio-economic variables that would have taken place anyway: ‘evidence has been unable adequately to address the central problem of how to bring together disparate outcomes in order to assess overall cost effectiveness, or of how to assess the extent to which multiple interventions are more cost-effective than single focus initiatives’ (Dabinett et al, Department for Environment, Transport and the Regions 2002).

True evidence based policy can only emerge from a sound understanding of the underlying forces operating to transform places. This is a prerequisite for meaningful evaluation of intervention and can only come from a systematic examination of a wide range of neighbourhoods over a prolonged period, not just deprived areas over a few years. Because the time-location value signature measure could be rolled-out for the whole of the United Kingdom, it has the potential to pave the way for a more empirical approach to the evaluation of regeneration policy.

There remain, of course, significant barriers to overcome in order to construct a truly systematic and geographically widespread evaluation of public intervention at the local level. The most obvious difficulty lies in locating and quantifying all the sources of public investment due to the absence of a central database of the many types of funding from various tiers of government (local, national and European). Nevertheless, improvements in data recording procedures in future, combined with the techniques presented here, could lead to a more detailed and rigorous way of quantifying the impact on well-being of public expenditure at the local level.

6 Conclusion and Future Developments

Housing submarkets are important. They determine the impact of new supply on the surrounding housing system; they condition the response of house prices to policy interventions, new supply, and economic shocks; and they have implications for the pattern of housing wealth and related social inequalities. Without a sound understanding of the nature and operation of housing submarkets, we shall have a very shallow understanding of the structure of the urban system and we shall be ill-equipped to make appropriate policy decisions at the sub-regional level.

Why has submarket analysis not entered the standard tool-box of policy analysis? The primary reason is that no method currently exists that derives submarkets in a way that is both consistent with economic theory and amenable to application and interpretation by policy practitioners. Our goal in this project was to develop a method of delineating and quantifying submarkets that would address this omission. More specifically, we sought to find a way of delineating submarkets that was: (1) amenable to analysis at a range of spatial scales, (2) based on sound and reproducible economic analysis, (3) not overly dependent on administrative boundaries in its fundamental calculations, (4) amenable to being represented using any standard boundary definition, and (5) not limited to an ad hoc local dataset but could in principle be applied to any area of the United Kingdom.

This report has outlined the method we developed in our attempt to achieve these goals. Our approach is comprised of the following four steps:

1. Estimate constant quality house price surfaces for two separate points in time;

2. Compute the cumulative rate of inflation between the two periods and plot the *inflation surface;*

3. Calculate the average rate of inflation at the chosen level of spatial aggregation;

4. Apply clustering analysis using an appropriate set of clustering criteria (we used three variables: latitude, longitude, and the estimated rate of constant quality house price inflation).

The advantage of our approach is that each of these four steps yields useful outputs for the decision making process. Step 1 gives us a visual snapshot of the constant quality price structure of the housing market. This is useful because it allows one to see at a glance the variation in location value (land prices) at a given point in time, and to identify where demand is high or low relative to supply. Step 2 yields an unbroken picture of how house price inflation varies across geographical space. This is a significant improvement on current measures of inflation which tend to be published for discrete spatial units – such as regions or local authority boundaries. Step 3 allows us to translate the data from our inflation surface into a format that allows it be meshed with other data sources. Step 4 provides a meaningful way of dividing up a region into its constituent housing submarkets.

The fact that each of these intermediate steps is useful in its own right means that, even if the local authority or planning body or private sector analyst did not have the resources to pursue all four phases at first blush, they would gain useful outputs from partial completion while setting in place the analytical prerequisites for the next stage.

Our method may also be amenable to a variety of fruitful “short-cuts”. For example, one might use HM Land Registry data to derive simple average inflation rates for each postcode sector and then use Step 4 to cluster these postcode sectors into contiguous segments. While this would not match the analytical rigour of our inflation surface method, it is likely to lead to a fair approximation, though further research is needed on this and other potential short-cuts.

A further benefit of our approach is that it provides a modelling framework that could potentially be applied to a much wider range of policy problems than just the derivation of submarkets. For example, the constant quality surface could be interpreted as a model of how the quality of life varies across space and over time. This in turn means that it could be used as a way of gauging area deprivation and for evaluating or simulating the impact on wellbeing of policy interventions. The framework may also provide a way of modelling the impact on housing of a range of environmental factors, such as flood-risk/climate change and pollution.

6.1 Future developments

Our suggestions for future development of this research are as follows:

- To roll-out the method used in Kent for the whole of the United Kingdom;
- To explore ways in which the model could be improved (by including additional spatial variables and by merging other price data, for example);
- To compare the submarket boundaries calculated using the inflation-surface approach with those of alternative methods and to explore the relative performance of the various “short-cuts” that could be used to approximate our method;
- To explore how the inflation-surface approach could be used to consider other policy relevant factors, such as the impact of (i) new supply; (ii) increased densities; (iii) rising sea-levels and flood risk; (iv) school performance; (v) council tax revaluation; and (vi) urban regeneration.

Glossary

Attribute variations – differences in property characteristics – particularly those that affect value (number of rooms, whether a property is detached, whether there is a garage etc).

Automatic valuation model (AVM) – an automated system for estimating the value of a property based on evidence from recent sales of similar properties.

Cluster analysis – a mathematical procedure that groups items according to some pre-specified criteria (such as proximity and similar rates of price change).

Cointegration tests – a statistical method for identifying whether or not there is a long-term relationship between variables.

Constant quality inflation – a way of calculating price changes that controls for variations in the attributes of the 'good' in question. Because houses are so heterogeneous, apparent changes in average price may in fact reflect changes in the mix of property types coming onto the market. By estimating changes in the value of a "constant quality dwelling" – a hypothetical uniform property – one can, in principle, produce a more reliable measure of house price inflation.

Continuous time interactives – multiplication of a variable by the time the observation was recorded relative to the first period in the data. By including such terms in a model it is possible to estimate how time has an effect on the relationship between particular variables.

Cumulative inflation surface – the total percentage change in price that has occurred between two periods (such as the change in house prices between 1996 quarter 1 and 2004 quarter 4) calculated at every point in geographical space. One can conceive of this variable as a "surface" or "geographical landscape" where the hills and valleys represent areas of high and low inflation.

Dendrogram – a tree-like diagram that helps one to visualise the optimal grouping of items. The length of branches indicates where the optimal number of groups is likely to lie. For example, if one has a hundred items, they might naturally fall into 6 groups, in which case the length of the "branches" of the dendrogram between this level of grouping and an attempt to cluster the items into a higher number of groups would be relatively long.

Geographical Information System (GIS) – a way of visualising and analysing data using digital mapping techniques.

Housing density – in this report, housing density refers to the average proximity of dwellings in a particular area. In neighbourhoods where dwellings tend to be far apart (due to large gardens, for example), housing density will be low. In high density areas, on the other hand, dwellings tend to be close together. Housing density can be measured, for example, using the average distance between properties in each postcode unit.

Inflation landscape – this is another term for the "*cumulative inflation surface*" defined above. It plots a "landscape" or "surface" in geographical space that represents the highs and lows of estimated inflation. It is a means of depicting the way in which house price changes vary across neighbourhoods. If all houses in all areas increased in value at exactly the same rate, the inflation landscape would be perfectly flat.

Inflation surface – this is another name for the "*cumulative inflation surface*" and "*inflation landscape*" terms defined above.

Inflation surface modelling – inflation surfaces can be drawn by simply plotting a map of average house price change in each neighbourhood and then "joining up the dots" in three dimensional space. This entails fitting a surface to join up each of those inflation estimates. Modelling is only required if one wants to control for variations in dwelling attributes (see "constant quality inflation" defined above).

Latitude and longitude interactions – multiplication of a variable by the latitude and/or longitude of each of its observations. By including such terms in a model it is possible to estimate how location has an effect on the relationship between particular variables.

Location value signature – the term coined by Fik et al (2003) to describe the value that arises specifically from a property's location. Because the location of each house is different, each will have a slightly different outlook, different set of neighbours, different position on the street, and different set of distances to amenities. The particular bundle of costs and benefits associated with each location will give each property a unique spatial "signature". Thus, even if two properties have identical structures, they are unlikely to have identical values because their location signatures will be different.

Mathematical algorithm – a set of rules for calculation. For example, a very simple algorithm is the set of rules used to calculate the average: "divide the sum of values by the number of values". More complex algorithms are needed to derive logical methods for clustering observations, or estimating the relationship between variables.

Nearest-neighbour surface fitting techniques – a way of estimating the surface that best approximates a series of points plotted in three dimensional space. It is a sophisticated way of "joining the dots".

R^2 – also called the *coefficient of determination*. It is a measure of goodness of fit of a statistical model. The closer the value is to one, the better the model fits the data.

Regression equation – a way of representing the relationship between two or more variables in a mathematical formula.

Season dummy – "dummy" is statistical jargon for a binary variable – one that has values that are either zero or one. A season dummy is simply a binary variable where the values are all zero except for the season in question, which would be coded as one. For example, a "*first quarter dummy*" would have all values equal to zero except for observations that occurred in January, February or March, which would have the value of one. By including seasonal dummies in a model, it is possible to estimate the effects of seasonal variation.

Submarket delineation – estimation of the boundaries of market segments. Like neighbourhoods, housing submarkets do not have pre-specified boundaries – their perimeters emerge organically and shift over time as the composition of the population varies and changes occur in local amenities and property characteristics. *Submarket delineation* is an attempt to work out where those boundaries lie at a particular point in time.

Stepwise process – an automated statistical procedure that removes insignificant variables from a regression equation.

Time-location value signature – this incorporates a time dimension to the location value signature explained above. In other words, a property may have a unique set of location attributes (position on the street, good/bad neighbours, access to quality schools and amenities etc.), but the value of those attributes will change overtime.

Ward's method – this is a particular type of clustering technique. It is regarded as one of the best methods for grouping items into clusters. Ward's method "considers all pairs of clusters and asks how much 'information' would be lost if that pair were to be amalgamated. The pair chosen is then the one which involves the least loss of information" (p.22 of Bartholomew et al 2002).

Zoning – the process of allocating and regulating land use. For example, a planner might stipulate that all development in a one zone is to be residential, while all development in another zone is to be industrial. Zoning might be used, for example, as a way of minimising the negative effect on residents of the noise, congestion and pollution associated with industrial and commercial land use.

Bibliography

Atkinson, R. and K. Kintrea (2001)

– Disentangling Area Effects: Evidence from Deprived and Non-deprived Neighbourhoods, Urban Studies

Barker, K. (2003)

– Review of Housing Supply: Interim Report, HMSO, London

Barker, K. (2004)

– Review of Housing Supply: Final Report, HMSO, London

Bartholomew, D. J., Steele, F., Moustaki, I. and J. Galbraith (2002)

– The Analysis and Interpretation of Multivariate Data for Social Scientists, Chapman and Hall, Florida.

Bramley, G., Pawson, H. and H. Third (2000)

– Low Demand Housing and Unpopular Neighbourhoods, Department for Environment, Transport and the Regions, London

Dabinett, G., Lawless, P., Tyler, P., and J. Rhodes (2002)

– A Review of the Evidence Base for Regeneration Policy and Practice, Department for Environment, Transport and the Regions, London

Fik, T.J., Ling, D.C., and G.F. Mulligan (2003)

– Modelling spatial variation in housing prices. Real Estate Economics

Galster, G. (2003)

– Neighbourhood Dynamics and Public Policy, in O'Sullivan, A. and K. Gibb. Housing Economics and Public Policy, Blackwell, Oxford

Gelfand, A.E., Kim, H.J., Sirmans, C.F., and S, Banerjee (2003)

– Spatial Modelling with Spatially Varying Coefficient Processes. Journal of the American Statistical Association

Goodman, A.C. and T.G. Thibodeau (1998)

– Housing Market Segmentation. Journal of Housing Economics

Goodman, A.C. and T.G. Thibodeau (2003)

– Housing Market Segmentation and Hedonic Prediction Accuracy. Journal of Housing Economics

Grigsby, W. (1963)

– Housing Markets and Public Policy. Philadelphia: University of Pennsylvania Press.

Gripaios, P. (2002)

– The Failure of Regeneration Policy in Britain. Regional Studies

Ismail, S. (2005)

– Hedonic Modelling of Housing Markets using Geographical Information System (GIS) and Spatial Statistics: A Case Study of Glasgow, Scotland. PhD thesis, Department of Property, University of Aberdeen.

Jones, C., Leishman, C., and C. Watkins (2003)

– Structural Change in a Local Urban Housing Market. Environment and Planning

Maclennan, D. and Y. Tu (1996)

– Economic perspectives on the structure of local housing systems. Housing Studies

Neighbourhood Renewal Unit (2003)

– Review for The Neighbourhood Renewal Unit of Blueprint for the Index of Multiple Deprivation – at Small Area Level, Office of the Deputy Prime Minister, London (academic assessment by Jonathan Bradshaw).

Ordnance Survey (2005)

– Meridian 2 Map data, accessed, under licence, from Edina Digimap (© Crown Copyright/ database right 2005. An Ordnance Survey/(Datacentre) supplied service).

Pryce, G. (2004)

– The Micro and Macro Effects of the Location of New Housing Supply, Office of the Deputy Prime Minister (now Communities and Local Government), London.

Pryce, G. and K. Gibb (2006)

– Submarket Dynamics of Time to Sale, Real Estate Economics

Robson, B., Bradford, M., Deas, I., Hall, E., Harrison, E., Parkinson, M.,Evans, R., Garside, P., Harding, A. and F. Robinson (1994)

– Assessing the Impact of Urban Policy, HMSO, London.

Social Exclusion Unit (2001)

– A New Commitment to Neighbourhood Renewal: National Strategy Action Plan, SEU, Cabinet Office, London.

Thomas, B. and D. Dorling (2004)

– Know your Place, Shelter.

Watkins, C. (2001)

– The Definition and Identification of Housing Submarkets. Environment and Planning

Acknowledgements

We acknowledge the support of Communities and Local Government (Housing Markets and Planning Analysis Division) and the Royal Institution of Chartered Surveyors (RICS). Any views expressed in this research report are not necessarily shared by Communities and Local Government or the RICS. We would also like to thank David Forrest from the Department of Geographical and Earth Sciences at the University of Glasgow for his advice regarding the creation of geo-coded topological variables. Particular thanks also to Nationwide and Experian for their generous provision of data. All remaining errors are our own.