The Micro and Macro Effects of the Location of New Housing Supply

Interim Report for the Office of the Deputy Prime Minister

Gwilym Pryce
Department of Urban Studies
25 Bute Gardens
University of Glasgow
G12 8RS

email: g.pryce@socsci.gla.ac.uk

Contents:

1 Introduction

Plan:

1.1 Context

The recently published Interim Report from the Barker Review of Housing Supply (2003, p.58) estimates that the current number of new homes per annum would need to be doubled in order to "achieve the European trend rate", and "more than double to get real price stability". Macro estimates of this kind, however, implicitly contain a plethora of assumptions about the nature and operation of local housing markets. The coexistence of spiralling house prices in some areas and low demand in others demonstrates the complexity of the supply-demand mismatch at the micro level and raises the crucial question of where those new properties should be located.

Given the relative proximity of contrasting housing submarkets, it is possible that the actual impact on house price growth of a major expansion in housing supply (were that at all achievable) would depend on where those new houses are built. If new supply is to realize its potential impact on macro house price trends, careful thought needs to be given to the substance of local land planning strategies and whether the local economic and housing market effects of land release decisions are being fully considered. And it is not just the evidence-base of the decisions of local planners that are at question here. Are there sound economic grounds, for example, for the choice of the four expansion areas outlined in the sustainable communities plan? Has the estimated local impact been systematically measured and evaluated against alternative locations?

Whilst the questions may seem obvious, delivering on them is not. How the location of new houses relates to their final micro and macro impact is a relatively unexplored topic. Identifying the appropriate evaluation framework for new supply is a much more difficult task than simply stating that evaluation is needed. The situation is complicated further by questions of how and whether land release policy can be integrated with initiatives aimed at encouraging the regeneration of areas and the promotion of sustainable communities. Given the complexity of local housing markets and the social interactions that affect them (Meen and Meen, 2003), meaningful evaluation can really only be achieved from empirical analysis of housing markets based at the very local level.

Unfortunately, there have been relatively few published empirical studies in the UK of housing supply at all, and none of these extend to an analysis of the hypo-micro impacts of new construction on existing housing markets. To date, one of the few attempts to consider these issues in a UK context is an unpublished pilot study by researchers at the University of Glasgow for the Joseph Rowntree Foundation in 2003. Using housing supply and transactions data on Glasgow, Pryce and Gibb 2003 investigated whether there was a different impact of new construction on contiguous second hand housing markets depending on whether those areas were deprived areas or affluent. The results suggested a potentially profound asymmetry in the impact of new supply. For deprived areas, new construction was found to have a substantial positive impact on house price, but for affluent areas, the effect was barely traceable.

But in terms of establishing the optimal location of new supply, these results only scratch the surface. A great deal more needs to be learnt about the local effects of new construction and the goal of this report is to offer some suggestions as to how future research could provide a useful evidence base. This may seem a modest goal at first, but under-girding the public debate are some spectacularly complex theoretical and methodological issues. The aim is not to cavalierly unearth these issues – that would expose us to the kind of agnostic despair that so easily befalls practicioners of the 'dismal science'. Instead I hope to offer some detailed (though inevitably incomplete) practical suggestions that draw on some recent advances in modelling techniques and data collection.

1.2 Aims

The report attempts to address three core questions:

1. What is the macro house price effect of new supply?
2. How might new supply have a regenerative effect?
3. What is the role of submarkets in determining the effects of new supply?

It is immediately apparent that these are not separate but interrelated aspects of the effect of new supply. Note that there is an important hypothetical premis, that of a significant expansion in the release of land for residential use and in the rate of new construction. To some extent this may be viewed as a distraction from the core problem facing UK housing policy of how to increase the rate of new completions given the failure to do so in any significant way over the past thirty years (see the Barker Review Interim Report 2003). Whilst this is ultimately the more pressing question, in reality it is contingent upon the issues addressed in this report. For if the truth be told, the political pressures against land release will persist and indeed intensify at the least sign of planning liberalisation and as (if) more properties are built. Reform of supply-side housing policy is inevitably going to be a gradual process, where convincing justification for deregulation is likely to be a political necessity at every stage (and justifiably so). As such, understanding the local effects of land release will become increasingly important since it is the reports of local examples of success or failure that are often most influential in shaping public opinion. The more we can understand the local adjustment process of the market in response to new supply and how these feed through to macro house price changes the better informed will policy be and the more constructive (hopefully quite literally) the debate.

1.3 Plan

The plan of the report is roughly as follows. First I briefly review some of the literatures pertinent to local impact of housing supply (chapter 2). The literature of most obvious connection is the submarkets literature and I offer a critique with respect to the usefulness of current approaches to delivering identification of submarket boundaries that are of practical use. I then consider the crucial question of how the existence submarkets might impose spatial qualifications on the impact on adjacent second-hand dwellings, and in particular the potential for using the location of new supply to regenerate deprived areas (chapter 3). Chapters 4 and 5 return to the question of how submarkets might distort the macro price effect of new supply. Different rates of turnover of stock, and different demand elasticities across submarkets may have a crucial role to play in the long term price effect of the location of new supply. This is something of a leap in the role usually ascribed to submarkets which typically occupy a fairly minor place in the greater scheme of housing economics (though there have a limited number of papers that have discussed these issues at a fairly preliminary level, such as Jones and Watkins 1999). Bourassa et al (2003), for example, in answer to the question posed in the title of their paper, “Do housing submarkets really matter?”, conclude that they do because when properly accounted for, better specified hedonic price equations can be derived. What

I argue below, however, is that far more may be at stake than R-square of hedonic regressions. Perhaps submarkets hold the key to the optimal location of new supply, and indeed, to what the notion of "optimal location" might entail. This is the subject of the final chapter of the report which offers a methodological framework for expanding the knowledge base necessary to facilitate future decisions on the location of new supply.

2 Submarkets

Plan:

2.1 Introduction

Existing empirical literature on housing supply in the UK has tended to focus on its causes and effects at the city, regional and national levels, and in so doing, has overlooked the implications of new housing at the very local level. There is no need to review this supply literature since that task has competently and recently been executed by a number of authors (see White and Allmendinger 2003; Cullingworth and Nadin, 2002; Adams and Watkins, 2002; and not least, the Barker Review Interim Report, 2003). The need is now to consider more thoroughly the links with literatures usually viewed as tangential to the supply question, the most obvious of which (given the goals of the current report) is the extensive body of work that defines, describes and delineates submarkets. Whilst there is a brief discussion of the potential importance of submarkets to land planning decisions by Jones and Watkins (1999), the links between submarkets and housing supply remain conceptually undeveloped and there is almost a complete absence of empirical analysis.

The submarkets literature actually preceded the more dominant set of papers that came to be known as the "access-space model". As we explain below, access-space view of the world came to dominate because it offered a coherent theoretical framework for the structure of cities that linked household work/leisure trade-offs with location decisions and the spatial dimension of urban housing. Beginning with an initial monocentric simplification of the city, successive elaborations of models have been developed providing insights into how economic agents might trade off the consumption of space versus accessibility to a central business/amenity location (see Anas et al, 1998, for a comprehensive review). This new urban economics perspective painted a very different picture of the city to the fluid and somewhat unstructured patchwork described by the early submarkets proponents. But the concept of submarkets has subsequently enjoyed something of a renaissance. Access-space models overlooked the segmentation of local housing markets and Maclennan (1982), amongst others, argued that this led to a fundamental misrepresentation of how housing markets truly worked. There has followed a plethora of empirical studies attempting to identify the boundaries of submarkets for particular cities, many of which have found evidence for significant and persistent heterogeneity between urban housing segments.

I shall present below an overview of the submarkets literature and attempt to show how they may profoundly shape the impact of new supply. Of course, this connection is of little use if submarkets remain a theoretical abstraction. Unless their boundaries can be identified we are left none the wiser. Unfortunately, methodological difficulties encountered by empirical analyses of submarkets have meant that their delineation remains somewhat elusive. So a second goal of this chapter will be to suggest ways in which recent methodological advances can be employed to overcome these problems. I shall argue in later chapters, that once identified, submarket areas may form the useful basis for analysis of a range of factors that effect the impact of supply.

2.2 Filtering and Early Definitions of Submarkets

The conventional notion of an urban housing market is one of a single homogenous market:

> "A housing market area is the physical area within which all dwelling units are linked together in a chain of substitution… In a broad sense, every dwelling unit within a local housing market may be considered a substitute for every other unit. Hence, all dwellings may be said to form a single market, characterized by interactions of occupancy, prices and rents" (Rapkin et al, 1953, pp. 9-10 quoted in Grigsby, 1963, pp. 33-34).

Grigsby (1963, p.34), one of the early expositors of submarkets, argues that, "In reality, the housing market in a given area consists of groups of submarkets which are related to one another in varying degrees". Dwellings are to be considered in the same submarket if the degree of substitutability between them is sufficiently great to "produce palpable and observable cross-relationships in respect to occupancy, sales, prices and rents, or in other words, whether the units compete with one another as alternatives fo the demanders of housing space" (Rapkin et al, 1953, p. 10 quoted in Grigsby op cit).

Imediately we can see how submarkets are related to the subject of new supply, for the picture painted by Grigsby leads one to ask which set of existing dwellings will be considered substitutes for the new housing? The question is pertinent because it is those units with which the new supply will be competing that will experience the greatest downward pressure on price as a result of the new construction. The relevance new supply does not end there, however. The filtering model(s) on which the early submarkets literature was based, explained how new housing had a domino effect that ripples through the whole housing system. To illustruate, proponents of filtering would typically assume that high income households experience an increase in income, and then suppose further that they have a high income elasticity demand with respect to new construction (that is, as income rises, they will want to spend a larger proportion of their total income on that good). The result would be that high income households move into the new dwellings and "the houses that they vacated would fall in value, allowing a household of lower income to move in" (Fallis, p. 82). This process, known as "filtering" is replicated throughout the housing chain until the lowest income households have moved out of the lowest quality housing, which subsequently becomes vacant and is eventually demolished.

Note, however, that in the context in which this literature was written, the private rental sector plays an important role – vacant dwellings arise because landlords cannot find tenants for the lowest quality housing. In the UK context where many cities have a residualised rental sector, the process might succeed through low income mortgage borrowers in the lowest quality owner occupied housing failing to meet their mortgage payments (either because of a fall in household income or because they no longer wish to live there but are unable to sell). Mortgage borrowers might take possession of the dwellings but may then unable to resell them.

Whatever the precise process, the prediction of filtering is that if one puts quality new supply in at the top of the housing chain, once the cogs of the system have had opportunity to turn, out will pop a similar quantity of vacant dwellings at the bottom ready for demolition and everyone enjoys a boost in living standards in the process.

The notion of filtering was central to the policy debate in the US in the 1960s and 1970s over whether the state should subsidise new construction (filtering had a lower profile in the UK debate because the "postwar public housing programme … diverted attention from filtering concepts" Maclennan 1982, p. 25). Even if the poor are unable to afford the new dwellings being built, so the argument goes, they will benefit because "better quality housing will 'filter down' to them" (Fallis, p. 83). However, whilst it is possible to design a system of construction subsidies at various quality levels that reduces prices at all quality levels (see Sweeney 1974), an arbitary set of subsidies will not necessarily do this and "It is even possible that a construction subsidy will lower prices at high-quality levels and increase prices at lower quality levels – thus benefiting the rich and not the poor" (Fallis, p. 83).

Ironically, even a well oiled filtering process may have a potentially negative side effect if it operates in an economy with a diminutive private rented sector where large proportions of the poor in owner occupation (cf Burrows & Ford's "Half the Poor"). Were new supply to filter down in the form of falling real prices at the lowest end of the housing distribution, the brunt of the adjustment process would be bourne by those on lowest income (not asset-rich landlords). Note that there may already be downward pressures on prices at the lower end of the housing market even without any new supply. If, for example, the income elasticity of demand is greater for more luxurious and better placed dwellings (a theme picked up on later in the report – **see chapter 5** on income and demand elasticities) and if a good proportion of those in the lowest income groups have a fairly flat distribution of future wage offers. In other words, the greater the growth in the disparity in income and the greater the differences in the income elasticity of demand for housing of the lowest quality compared with that of the highest, then the greater the disparity in expected capital gains between lowest and greatest house price brackets. Not only will wealthier households enjoy more expensive houses, but they will enjoy greater proportional increases in values of those houses **(see chapter 5** below).

How can these phenomena be verified empirically? One consequence that could fairly easily be investigated is the extent of price divergence or convergence. Though there are multiple causes of this (not least due to movements in the income distribution – see Andrew and Meen 2003, 2004) of price divergence, its existence is relevant since it makes more potent the issue of supply-induced depression of the lower end of the housing market. A simple approach would be to consider whether the lowest 5% of house prices has diverged from, or converged towards, the prices of dwellings in the top 5%. It would be feasible, for example, using previous years of the CML 5% Survey of Mortgage Lenders and/or Land Registry data to examine this question. To illustrate, consider SASINES data on the City of Glasgow. Analysis reveals that, in 1991, the 5 centile house price was 9.63% of the 95th centile. By 2000, this ratio had fallen to 7.5%. The growing disparity is associated with a rise in the standard deviation of house prices from £37,627 to £49,402. Research is needed to ascertain whether such divergence is commonplace and whether our target should not just be to stabilise average prices but to reduce the standard deviation of prices within [illegible]es.

In the UK, there may be reason to believe that low demand areas tend to spiral downwards, not upwards, (Meen and Meen 2003) into areas of even less demand, and if the owners of such housing are not wealthy landlords, but homeowners at the bottom of the income gradient, then the impact of new supply may not be entirely desirable if it initiates a downward vortex. If poorly placed new supply results in the bottom falling out of the lower end of the housing market, we have to consider who will bear the cost. Clearly there are city and regional differences that may qualify these affects, not least due to the differences in the nature and scale of migration flows across the country (an influx of low income migrants, for example, may increase the price inflation of housing at the lower end of the housing chain relative to the top). These questions only heighten the need to augment our understanding of the location-dependent impacts of new supply and substantially increases the importance of taking into account the potential regenerative effect of locating new supply adjacent to deprived areas (**see chapter 3**).

2.3 New Urban Economics

Filtering theory took something of a back seat in the urban economics literature as a model of a rather different kind came to dominate economists' perspective of the city. This focussed on the trade-off that households make between locating near the city centre and face low commuting costs, and locating nearer the periphery and endure longer and more expensive travel to work. Travel costs cause land prices to fall further away from the city centre, and so there is an "access-space" choice to be made: live in a small plot and enjoy easy access to the central business district or benefit from a large plot on the edge of the city and endure the daily commute to work (Alonso 1964). While the filtering theory was hindered by ambiguities over definition due to the lack of formal analytical models (Fallis p.82)[1], the access-space theory was presented in elegant mathematical frameworks with explicit assumptions and formal, testable hypotheses. The rigour and transparency of the basic model made it amenable to further development by Muth, Romanos, Evans and others into what became the "dominant paradigm of urban economic research" (Maclennan, 1982).

This new framework, though not mutually exclusive to the filtering model, certainly offered a different perspective on the nature of cities. In the new urban economics world, cities were well ordered and symmetrical, formed on a flat, featureless plane, typically with a single business district

[1] Filtering came to mean different things to different authors. It was not clear whether filtering was primarily of properties or people and the submarket "matrix" suggested by Grigsby lacked prices and a true economic adjustment mechanism (Rothenburg et al.**??**)

at the centre, and where the trade-off between access and space usually led to more affluent workers locating in larger plots on the edge of the city. It was an economic system where all markets were in equilibrium, all individuals shared common sets of preferences, all markets in equilibrium and all households and firms have perfect knwoledge and foresight of market relevant information. And most importantly, all adjustment to economic shocks was instaneous.

2.4 Submarket Critique of the Access-Space model

There has subsequently been something of a rediscovery of the early foundations of and motivations for the submarkets literature as empirical studies revealed a more complex picture of the true structure of urban cities. The monocentric, instantly adjusting urban superstructure does not fit with reality. This, of course, was pointed out by the submarket and related literature long before the access-space framework had evolved. Ironically, Maclennan notes that Hoyt (usually seen as a precursor to the access-space theorists), in his 1939 analysis of housing market dynamics, actually presented a rationale for the form of a city to emerge not by distance to city centre, but by coagulation of residential types:

> "large cities were as much characterisd by residential sectors as they were by residential rings. The basis of his argument is as follows. In the early phase of urban development, the most affluent and influential social and economic group were not sufficiently numerous to occupy a complete residential ring of the city. Instead, they tended to gather within a well-defined area or sector on one side of the city centre. The location of this high income group then became a reference point for successively lower income groups. The rationale for this assertion was that lower income households have a preference to live near to their peer groups. Thus initially assuming a stationary urban economy, it could be expected that the neighbourhood of the dominant social, racial and economic group would produce rent gradients around the edge of the peer group sector." (Maclennan, 1982, p. 23).

This is essentially a precursor not only to the submarkets literature, but also to the "self-organising" and social interaction models of cities that have emerged in recent years based on complex mathematical models borrowed from the biological and physical sciences (see Meen and Meen 2003 for a review). Moreover, limits and imperfections to the information upon which households base their location decisions, and the inability of the market to fully adjust within short or even medium term time frames meant that markets were likely to be characterised by disequilibrium and segmentation. Filtering models, therefore offer a complementary dynamic short run adjustment process and lead to the potential for persistent segmentation, the corollary of which it is argued in this report, is the consideration of submarkets as a unit of analysis. As a general rule, land prices may still fall (at a decreasing rate) from the centre of a city, but there will be important anomalies arising from the location of amenities, the physical geography of the city, and the spatial nature of social interaction (Meen and Meen, 2003; see Hoang and Wakely, 2000, and Brueckner et al, 1999, for recent attempts at revising location theory). Most important of all, the adjustment to economic shocks, such as a major increase in new supply, will evolve as a complex process characterised by spatial asymmetries not typically predicted by simplistic macro models:

> "In a uniform national economy with no spatial frictions in the flow of raw materials, households, or capital, a well structured macroeconometric model of the housing system could be expected to accurately describe and forecast the price and output dynamics of the housing system ... the existence of space in the national economy could be expected to produce differences betwe average 'national' performance and regional level changes. This could arise in several ways. First, preferences for housing vis a vis other goods, or for housing tenures, or the regional efficiency of the housing industry could vary across regions. Second, the structure of demand, supply, finance and the planning system not only vary over regions, but the spatial fixity of these factors restricts, in the short and the long term, equilibrating flows of households or 1 construction inputs. Third, apparently national processes may not diffuse smoothly or impact simultaneously over space' (MacAvinchey and Maclennan 1982; p.44; quoted in O'Sullivan p.62).

2.5 Empirical Analysis of Submarkets

The definitions of submarkets offered by Grigsby and Rapkin et al do not lead easily to a single empirically measurable definition of housing markets. In this section I shall consider practical attempts to identify the segmentation of housing markets and the pros and cons of the approaches currently employed in the literature.

2.5.1 Markets defined by buyers and sellers: the fallacy of population flows

Perhaps the most obvious solution to the problem of delineating submarkets is to simply use the patterns of intra-urban housing flows to identify submarket boundaries. Certainly this is methodologically feasible: we could in principle follow home owners that move within the urban perimeter (SASINES, for example, which records origin and destination of buyers). Presumably households would seek to locate within the same submarket so we could trace the location of both contiguous and non-contiguous submarkets over space (if they exist). Indeed, there is even an apparent theoretical rationale, to the extent that one way of defining a markets is in terms of its particular set of buyers and sellers.

This is the approach taken by Scottish Homes to define Housing Market Areas:

> "… in terms of housing and labour markets, a self contained housing market would be an area in which the majority of those moving house (migration), without changing jobs, would

> stay, and an area in which the majority of the employed population both reside and work" (Scottish Homes, 1993: 20, quoted in Jones and Watkins, 1999, p. 100).

Accordingly, Jones and Mills (1996) use migration patterns to define HMAs for Strathclyde using "the criterion of spatial containment of buyers, and operationalised by reference to migration patterns" (Jones and Watkins, 1999, p. 100). Using a threshold of 50% of buyers moving within an area, they derive 23 HMAs. Raising the threshold to 60% reduced the number of HMAs to 15. Jones and Watkins (1999, p.100) note that,

> "The significance of their findings for planning policy is that any assessment of demand based on a local authority's administrative boundaries is likely to be inaccurate. The HMAs identified for Strathclyde do not in all cases conform to these areas. In many cases individual local authority areas are either too large or too small… Where even moderate size HMAs exist, a full assessment of local demand requires submarket analysis within the HMA or analysis of housing demand in particular settlement set the wider context"

The authors go onto argue that, "This conclusion mirrors recent debate about the usefulness of TTWAs as the appropriate level of analysis for assessing unemployment (Turok, 1997). In both labour and housing markets the definition of the extent of a spatial market represents normally only the first essential pre-requisite for more policy analysis". Whilst Jones and Watkins are quite right to question the usefulness of administrative boundaries for policy and economic analysis, the question is whether the HMA approach to boundary definition can be applied to the definition of submarkets (if the two are indeed at all distinct). Is the conundrum of submarket demarcation not fully solved by the most obvious of solutions?

We have at this point come full circle. For this approach is closely allied with the dilemma that caused confusion in the earliest filtering literature. Are submarkets defined by people or properties? The answer is both. It is of course people that move, not housing, so it is tempting to conclude that people flows will tell us what really want to know about submarkets. This is a fallacy, however, first identified as such by Grigsby in his classic text:

> … the conventional approach completely ignores one of the key features of housing markets, the fact that some of the closest linkages are between *completely different* housing types. The reason for this high crosselasticity, moreover, lies in the differences themselves... It is … necessary that the dwelling unit on the market be … a better alternative for the family in question. And to be a better alternative, <u>*it must be different*</u>. By contrast, it would take a significant drop in new home

> prices to motivate a family to discard its current dwelling and buy an identical structure in a new development two blocks down the street." (Grigsby 1963, p.40-41; emphasise mine)

The point is this: households that move (other than for job or family reasons) may actually be more likely to be switching to a different (hopefully preferable) submarket. Afterall, why else move? So unless we know in great detail the reasons for each and every observed move, population movements in themselves are unlikely to tell us a great deal about the location of submarkets.

Having said that, if we have successfully delineated submarkets by some other means (such as hedonic structural break scanning) then it would be interesting to trace moves between dwellings of similar total price and similar sets of attribute prices (presumably due to some non-housing change, such as job relocation) as it might possibly offer a means of identifying properties perceived as substitutes but spatially distant. As such, allied with other techniques, it might be a means of identifying non-contiguous submarkets. It might also be of interest to identify how often intra-city moves entail a submarket switch. By itself, however, the population-flows approach does little to disentangle the Gordian Knot of urban submarkets.

2.5.2 Clustering by Attribute Type

One of the most popular approaches has been to apply some form of principle component, cluster or factor analysis to bunch properties that fall into the same product group and can therefore be broadly viewed as substitutes. Typically authors then run hedonic price regressions on the separate product groups and demonstrate that the regression fit improves significantly as a result of splitting the sample. Maclennan and Tu (1996), for example, use principle components analysis to identify the key variables that explain variation in their data on Glasgow, and then apply cluster analysis to those variables. Bourassa et al (1999) follow a similar process using principle component analysis to extract a set of factors from the original set of variables from local government area and individual dwelling data on Sydney. They then apply cluster analysis to the scores of the most important factors to determine the segmentation of submarkets and finally run hedonic price regressions on the subsamples to show that the clustering procedure results in a model that is "significantly better than classifications derived from all other methods of constructing housing submarkets[2]" (p. 160). Further examples include Dale-Johnson (1982) and Goodman and Thibodeau (1998).

However the application of cluster analysis to identify submarkets contains an essential flaw. Like the population-flows approach, it is using non-economic criteria to delineate and economic entity. Unless the clustering uses market determined criteria (such as marginal prices) to determine the way in which attributes should be bunched then although hedonic regressions subsequently run on the outcomes appear to perform better, there is nothing to say that the optimal (i.e. market) clustering of attributes has been achieved. This criticism is akin to the second point of Greene's critique of factor analysis:

> "First, the results are quite sensitive to the scale of measurement in the variables. The obvious remedy is to standardize the variables, but, unfortunately, this has substantial effects on the computed results. Second, the principle components are not chosen on the basis of any relationship of the regressors to y , the variable we are attempting to explain. Lastly, the calculation makes ambiguous the interpretation of results. The principle components estimator is a mixture of all of the original coefficients. It is unlikely that we shall be able to interpret these combinations in any meaningful way." (Greene p. 273)

Furthermore, there is the question of how the clustering algorithm distinguishes between attributes that are substitutes from those that are complements. Most important to our current goal, the aspatial nature of this approach does not necessarily result in a set areas that can be used for subsequent analysis. The practical worth of this approach, therefore, other than to improve hedonic regression performance, is questionable.

2.5.3 Comparison of Spatial Segments

Another approach is to segment the housing market using some spatial criteria. This might be to use real estate agent definitions of submarkets (Palm, 1978; Michaels and Smith, 1990); or government areas (Goodman 1981); or socio-economic segmentation (Straszheim 1971 aggregates US census tracts on the basis of racial composition); or some other administrative boundary (Ball and Kirwan use census boundaries; Hancock 1981 uses post code sectors clustered contiguously using structural break testing). The total number of papers to use this approach either on its own or in combination with the product-group method is substantial as the following table taken from Watkins 2001, demonstrates.

What is common to the great majority of tests for spatial homogeneity is the employment of hedonic regression as the yard stick. These models relate selling price to dwelling attributes. The estimated coefficients on attributes can be interpreted (following Rosen 1974) as the marginal prices of those

[2] by "all other methods" they mean all other methods of clustering which they experimented with.

attributes. That is, the regression coefficients estimate the market value of an extra room, the existence of a garage, an extra square metre of garden space, and so on.

The “law of one price” dictates that, within any one market, each attribute will have the same market value – the converse gives rise to arbitrage and the price differential is rapidly competed away. Therefore, by testing for homogenous attribute prices the literature has found a way of formally testing for submarket coherence. Where there are significant “structural breaks” – shifts in attribute prices – between areas, those areas can be designated as separate submarkets. The process is analogous to testing for black holes -- though such holes cannot be observed directly, their existence imposes necessary conditions for the behaviour of surrounding observable phenomena. Submarket boundaries are similarly unobservable in any direct sence, but their existence, if of any import at all, will be reflected in price distortions to the city level house price surface.

Table 1: Classification of submarket studies

Authors	Study Area	Study Date	Sample Size	No. of Test Segments	Sub-markets Exist?	Definition Class
Straszheim (1975)	San Francisco Bay, USA	1965	28,000	81	Yes	Spatial
Schnare and Struyk (1976)	Boston, USA	1971	2,195	2/3/2	No	Demander Group, Spatial, and Structural
Ball and Kirwan (1977)	Bristol, UK	1970/ 1971	280	8	No	Spatial
Palm (1978)	San Francisco Bay, USA	1971 and 1978	344	2/7	Yes	Spatial
Sonstelie and Portney (1980)	San Mateo, USA	1969/ 1970	1,453	25	Yes	Spatial
Goodman (1981)	New Haven, USA	1967 - 1969	1,835	5/15	Yes	Joint
Dale-Johnson (1982)	Santa Clara, USA	1977	3,021	10	Yes	Structural
Gabriel (1984)	Beer Sheva, Israel	1982	89	3	Yes	Spatial
Bajic (1985)	Toronto, Canada	1978	385	3	Yes	Structural
Munro (1986)	Glasgow, UK	1983/ 1984	154	2	Yes/No	Spatial, and Demander Group
Maclennan *et al* (1987)	Glasgow, UK	1976 and 1985/ 1986	863 and 1257	5	Yes	Spatial
Michaels and Smith (1990)	Boston, USA	1977 - 1981 (pool	2,182	4	Yes	Spatial

		ed)				
Rothenberg *et al* (1991)	Des Moines, USA	1963 and 1971	1,360	6	Yes	Structural
Hancock (1991)	Tayside, UK	1977/ 1978 - 1986	28,053	6	Yes	Spatial
Allen *et al* (1995)	Clemson, USA	1991	215	3	Yes	Structural
Adair *et al* (1996)	Belfast, UK	1992	999	7	Yes	Joint (nested)
Maclennan and Tu (1996)	Glasgow, UK	1984 and 1990	1257 and 1342	25	Yes	Joint
Bourassa *et al* (1997)	Sydney and Melbourne, Australia	1991	2,307 and 2,354	5	Yes	Spatial and Joint

Source: Watkins 2001, p.2238.

The obvious question to ask at this point is if a valid approach to identifying submarkets has been achieved, why submarkets have not emerged as a widely used "analytical framework for applied research" (Watkins, 2001, p.2235). Watkins (2001) argues that limited dissemination of submarkets as an applied tool arises from the lack of agreement on how areas should be defined, and whether the definition should be along spatial or product-group ("structural") lines, and what set of statistical criteria should be used to determine whether a boundary has been identified.

In addition to the lack of cohesion and dissemination in the identification and application of submarkets, the following set of methodological problems have tended to be overlooked in the literature:

Administrative and A Priori Boundaries

First, the smallest spatial unit has tended to be some kind of administrative boundary as the basis for analysis. This is not only the case when a study tests whether there is a structural break in the marginal valuation of attributes between adjacent local authorities or post code districts, but also when some other criteria is being used (such as racial composition) since spatial unit for data on this criteria is usually an ad hoc administrative one. At worst, this tells us little about the actual sub-structure of the urban housing market because it is only testing for breaks along non-market (I.e. administrative) boundaries, whereas we know that in many places these boundaries can have little in common with social and economic spatial structures. In principle, one would like to at least test for whether the submarket boundaries do indeed fall along administrative lines.

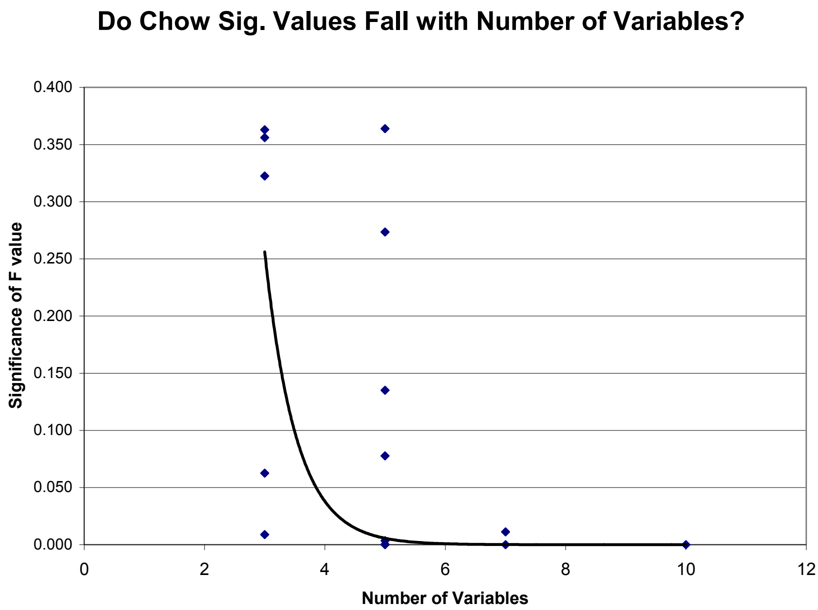
Over Specified

A second problem with existing studies is that the house price-attribute regressions – the "hedonic" models – tend to be over-specified. This is not usually a problem, particularly if one is going to use these models for prediction purposes, which is often the case (they are usually used to construct "mix adjusted" or "constant quality" house price indices). However, such models inevitably suffer from multicolinearity (linear correlations between the explanatory variables – Sheppard **1998**) and this results in unstable parameter estimates. Whilst this is not a problem when a regression is being used for prediction purposes (there is no deleterious effect on the efficiency of the predicted values) it does have major implications when one is testing for structural breaks, a point that has so far been overlooked in the literature. A highly colinear model will almost inevitably result in structural break tests coming out positive. It's rather like setting the sensitivity on one's car alarm to such a high setting that the alarm goes off at the slightest disturbance. The result? When the one hears the siren, one is more inclined to think that the alarm is detecting a gust of wind than a foiling burgular at work. Similarly, when over-specified regressions show evidence of structural breaks across space, one might be inclined to conclude that it is multicolinearity that is being detected rather than a submarket boundary.

Consider two very simple illustrations of this point (one empirically robust and one not). The first is a plot of significance values against number of variables from Chow results computed from a recent empirical submarkets paper (Watkins 2001):

	k	n	Chow F	r	dfu	sig F(r,dfu)
C&W	3	189	1.09	4	181	0.363
C&NW	3	82	3.67	4	74	0.009
C&E	3	101	2.32	4	93	0.063
C&S	3	176	0.58	4	168	0.322
C&SW	3	109	1.11	4	101	0.356
W&NW	5	201	1.1	6	189	0.364
W&E	10	220	5.61	11	198	0.000
W&S	7	295	5.57	8	279	0.000
W&SW	5	228	5.3	6	216	0.000
NW&E	5	113	4.79	6	101	0.000
NW&S	5	188	1.27	6	176	0.273
NW&SW	5	121	3.47	6	109	0.004
E&S	7	207	2.56	8	191	0.011
E&SW	5	140	1.95	6	128	0.078
S&SW	5	215	1.65	6	203	0.135

(Based on Watkins 2001)



Although there aren't enough observations here to draw any substantive conclusions (also note that Watkins' models benefit from being amongst the most parsimonious of those in the literature and so less susceptible to these problems) there is at least a hint of a negative relationship between significance values from Chow tests and the number of variables.

Another set of results, which are perhaps more conclusive with regard to this relationship between structural breaks and multicolinearity, emerges from the analysis described in the final section of this report where more than a million Chow tests are run on Glasgow GSPC data. It was found from repeated runs of the procedure that Chow significance values were substantially higher on average the more variables were introduced into the model and the greater the level of multicolinearity in the model. If this is true of regressions generally (not just hedonic house price equations), then it is possible that Chow's first test can, in certain circumstances, be considered as much a test for multicolinearity as test for structural breaks.

Sample Size

A third problem is that of sample size. There is a large variation in sample sizes across studies (and sometimes within studies), which makes it difficult to draw general conclusions from the literature regarding the prevalence of submarkets. In the studies reviewed by Watkins (2001) the sample sizes vary between 28,053 and 89. Potential ambiguity can arise from the fact that the larger the sample, the more certain you can be that a break of a given size is not due to sampling variation. So for a given size of observed break, significance values for tests will be much smaller in large samples.

One of the ways to simultaneously test this criticism in conjunction with the previous one would be to run a multiple regression model of Chow significance values against number of variables and number of observations:

$$\text{ChowSig} = a + b\,\text{k} + c\,\text{N}$$

It is anticipated that both b and c would turn out negative.

Testing Points Few and Far Between

Fourth, studies searching for structural breaks tend to select testing points that are few and arbitrary. It's impossible to define a continuous submarket boundary from such an approach as the testing points are too far apart and are insufficiently systematic. It is akin to attempting to draw a picture by joining the dots but where the dots are too sparse to be of assistance.

At what spatial scale should we test for submarket boundaries?

This is another dimension to the sample size question, and more specifically the spatial concentration of data used in empirical studies relative to the true spatial concentration of housing transactions. For example, if I only have data on 1% of transactions in a city, and the rate of transactions relative to the stock is not particularly high, then splitting a sample of say 400 down the middle to a test for structural break in the model will be testing for disjointness at a relatively large spatial scale. Since it is often difficult for the reader to extract from papers the true spatial scale at which submarkets are being tested for (particularly since, internationally, cities vary so dramatically in scale), it is difficult to draw general conclusions from the literature since every study may be testing for breaks at different spatial levels.

Are submarket boundaries discrete or gradual?

A question given very little attention in the empirical literature is whether boundaries are discrete or gradual. Theoretically, both are possible:

> The definition of submarkets is, of course, imperfect and somewhat arbitrary. "There may, indeed, be closer interrelationships between units of... two classes than between units within the same class." 6 This is because in the real world there are no clean cutoff points between two submarkets, the chain of substitutions being a continuum with sharp breaks or gaps being the exception rather than the rule. Nevertheless, where the distance between two units on the continuum is large, they become weak substitutes and the price or rent behavior of one does not affect the other. Or to put the matter differently, if the distance on the continuum is great, the units will be good substitutes for only a small number of families. (Grigsby, 1963, p.34)

The cumulative sum of empirical work to date really throws very little light on how precipitous or continuous boundaries tend to be across space, partly because analysis is ultimately based on discrete cells (such as post code sectors or census tracts) and so the breaks tested for are inevitably also discrete.

But can submarket boundaries be thought of in the same way one thinks of administrative boundaries? Can one, for example, straddle two submarkets in the way that one can have one foot in the East and one foot in the West by placing a foot either side of the Greenich Meridian? Do submarkets butt upto each other in the precise way that say the boundaries for local authorities do? Or the boundaries rather fuzzy – merging into one another with varying degrees of ambiguity? These are important questions since they determine how one should model the spatial effect of new supply. Should one divide the country up into discrete areas where the supply effect will be the same, or should one model the variation as a continuum?

The answer to this question may to some extent be answered by the grid search process described above.

2.6 The Way Forward

If the consequence of all this for micro housing market modelling is that existing methods do not lead us to well defined submarket areas that can be used as the spatial unit of analysis for the modelling the impact of new supply, the corollary is to ask, "what needs to be done to rectify the situation?"

It is proposed that what one would really like to be able to do is test for breaks at every point in the data using a consistent sample size and parsimonious hedonic model. This would allow us to observe where the probability of structural break is the highest – not just whether the test passes or fails, but the extent to which it passes or fails. Interim results from an initial attempt at such a method are described later on in the report.

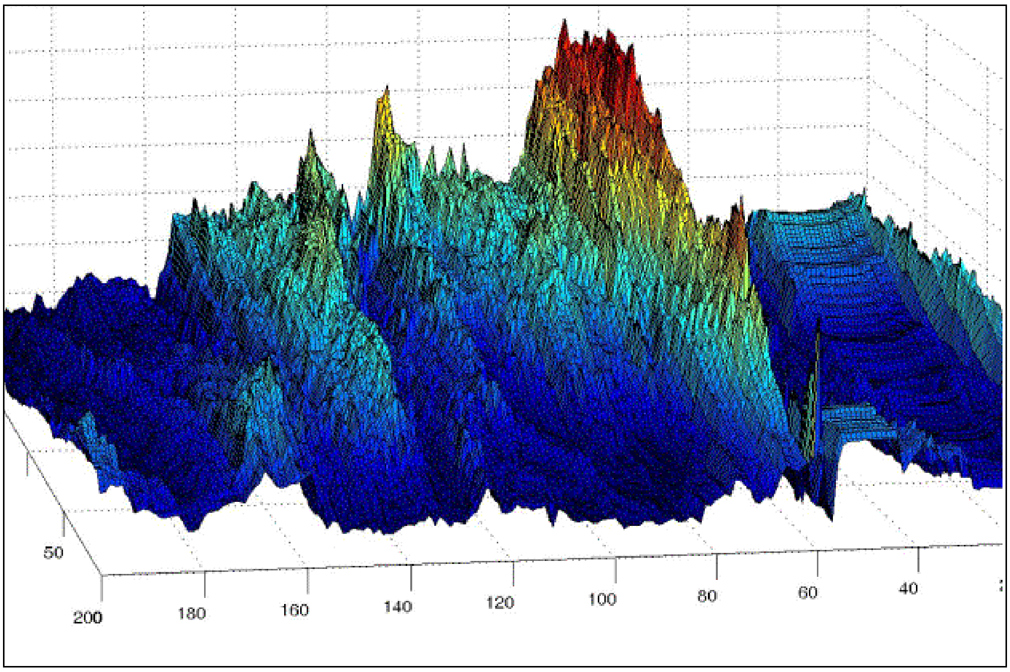
2.6.1 A Proposed Methodological Framework for Analysing the Impact of New Supply

The method, currently under development by the author (Pryce 2004), is to search for structural breaks at every point in the data. It starts off by splitting the data into horizontal bands – what I've called "Northing strata" (Appendix 1 describes the choice of the original data area). The program then searches along each stratum for East-West structural breaks using a moving "window" of 195 observations either side of the break point. This means that we lose 195 observations at each end of each stratum. The initial stratum contains 1,457 observations, so there are c1,060 structural break tests (1,457 – 2x195) in the first Northing stratum. I repeat this for each stratum until the total area is covered (if there are 1,300 northing strata, and c.1,060 structural break tests in each stratum, this means that we end up with nearly 1.5 million structural break tests).

The next step is to plot the F values from these tests (they are comparable since each regression has the same size sample) on true a Cartesian spatial grid to yield what one might call an "MRI scan" for submarket boundaries. While I have done an East-West search for all Northing strata, so far I have only plotted results for one segment of Glasgow, and that at a lower resolution of breaks (i.e. at every 50th observation rather than every observation). Also, the current run was not done in one go for each strata so there are vertical "seams" in the graphical plots, and the process of transferring the results onto a true Cartesian grid has not been completed. Note therefore that the graph that follows is not a true spatial plot (areas with no observations are compressed).

Figure 2-1

F-values from East-West structural breaks across space
(areas with no observations are compressed)

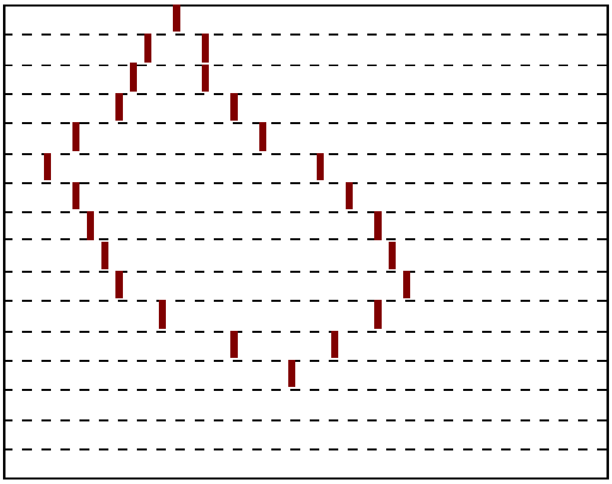


Nevertheless the graph reveals some important results. Most obviously, given the undulating nature of the plotted F values it confirms how important it is to test for structural breaks not just at a single point but at every point across wide areas. Where there are mountain peaks in the F-values, the structural breaks in the hedonic price regression are the strongest. This "continuum" approach provides perhaps the first glimpse of where the true submarket boundaries lie. If nothing else it takes the luck out of structural break modelling (in the same way that a "Geo-phyz" run gives an archaeological team a better chance of finding a site than randomly digging trenches).

Once the results for the whole of Glasgow are achieved and plotted onto a true Cartesian grid, the maximum points will be used to delineate East-West structural break boundaries, as depicted in the following stylised diagram:

Figure 2-2

"Mountatin range peaks" in F values
from East-West structural break searches

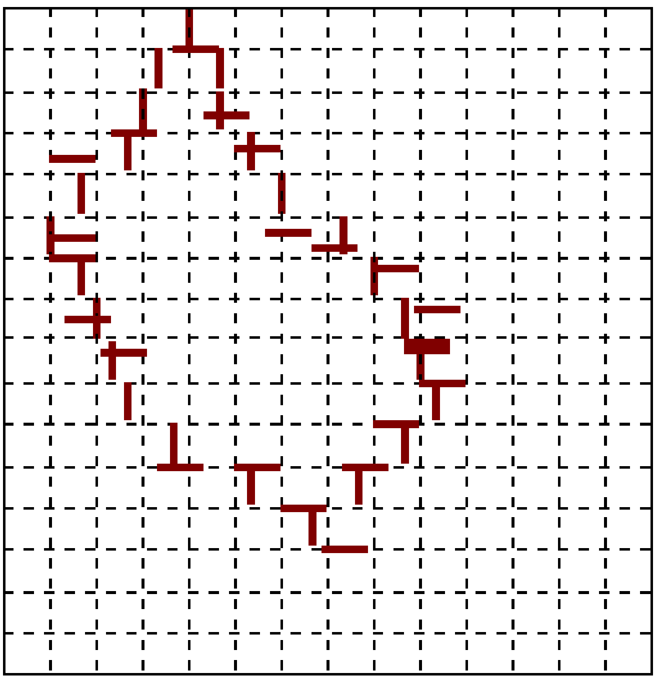


After this, the goal is to repeat the grid search for Easting strata (i.e. test for North-South Structural breaks), and locate North-South structural break peaks. Finally, two sets of peaks will be combined to form a pattern of horizontal and vertical mountain ranges that intersect to delineate submarket boundaries, as depicted below:

Figure 2-3

Vertical and Horizontal Peaks in F-Values

Combine to Define a Submarket



2.6.2 Are Submarket Boundaries Discrete or Fuzzy?

It is difficult to tell precisely the extent of gradation from the plot in Figure 2-1 given that it is not in Cartesian space, but it does offer a useful preview into whether the peaks and troughs in F-values (which reflect the extent of structural breaks) are likely to be precipitous or gradual slopes. The former would lead naturally to a discrete definition of submarkets, the latter to a more continuous one. The answer sems to be a rich mixture of both. At one end precipitous, at the other, gradual slopes. It is likely, though, that the answer to this question will depend on the aspect of inter-submarket variation we are interested in. For some purposes it will be useful to break up regions and cities into discrete submarket boundaries even though the true borders are of gradual gradation (in the same way that for some purposes it is useful to assign households into artificial income brackets, even though income is really a continuous variable). Defining submarket boundaries will be useful (if not essential) to phases of further analysis, particularly the consideration of price and turnover trends over time (see **chapter 4**).

2.6.3 At what spatial scale should we test for submarket boundaries?

The grid search method described above used a moving "window" of 400 observations. But what if we increased this to 1000 or even higher? Would the meaning of the results change? The answer is "yes", since for a given spatial density of observations, raising the sample size on which the structural break tests are based would raise the spatial level at which one is considering submarket boundaries. This relates to the criticism levelled earlier at the existing literature where the impact of spatial scale is not typically discussed. What this method offers, is a means for systematic analysis of the granularity of submarkets – that is, what spatial scale the most important segmentation of housing market occurs.

The "correct" spatial scale of submarket analysis depends crucially on the use to which such submarket identification is going to be put. If subsequent analysis is based on data available only at a relatively large spatial scale (e.g. post code regions or local authority level) then the submarkets should be defined accordingly.

Note that this method gives rise to the possibility of deriving a picture of how submarket boundaries pan out right across the whole of the UK. In this case, one might typically be interested in fairly large submarket areas, but the search process could equally employ a very micro definition even though a very large area is being scanned (the choice of 'window size' in the search algorithm can be applied to any overall search area). As such, one could picture a map of the UK with mountains and peaks plotted for the extent of structural breaks with a view to identifying just where the North-South divide in the housing market truly lie (one might, in fact, discover a whole series of divides, and ones that do not correspond at all to the usual Regional division of the country). Such an analysis is already possible in principle using the Nationwide data, or CML data (subject to spatial coding be released). The latter would be preferable since it would not be subject to the possible bias that might result from the particular Nationwide market share.

2.7 Do submarkets shift over time?

This is an important question since there are no theoretical results that freeze submarket boundaries (at whatever level) over time. Anecdotally, we may know of processes of gentrification some areas and relative decline in others, but one is generally aware of a series of regional intra-urban disparities that persist over time, some of which have existed for many decades (see the Hoyt, 1939, rationale for segmentation described above).

One of the advantages of the endogenous approach to submarket delineation described above is that it could in principle be applied to different years of data to ascertain, not only whether submarket boundaries shift, but the extent to which they shift and the way in which they shift (for example, the nature of the boundaries themselves might change, from being precipitous to gradual declines, or visa versa). Since the Nationwide data have existed for more than a decade, intermittent cross sectional scans of structural breaks in particular cities (namely those with enough observations in each year to make the analysis feasible) could be drawn up to monitor how submarket boundaries have shifted. Most encouragingly, the move to 100% sampling of mortgage transactions means that the CML data in future will provide an ideal basis for the scanning method and the new density of observations will allow it to be applied to a much wider range of cities and towns.

By analysing the movements of these boundaries over time, we might also discover that it is a practical way of investigating the kind of tipping point behaviour described by Meen and Meen (2003) as it would allow us to identify significant and rapid shifts in the relative fortunes of particular areas and would form the first step in a process of unpicking why and how the transition occurred (which in turn might help us understand the processes that are needed to induce rapid regeneration of an area). A complementary approach would be to apply the recently developed Geographically Weighted Regression technique developed by Stewart Fotheringham et al (see appendix to chapter 3) and related techniques that have been developed independently in the US (see, for example, Gelfand et al 2003 and Banerjee et al 2003).

2.7.1 An Illustration of Defining Submarkets by Attribute Price

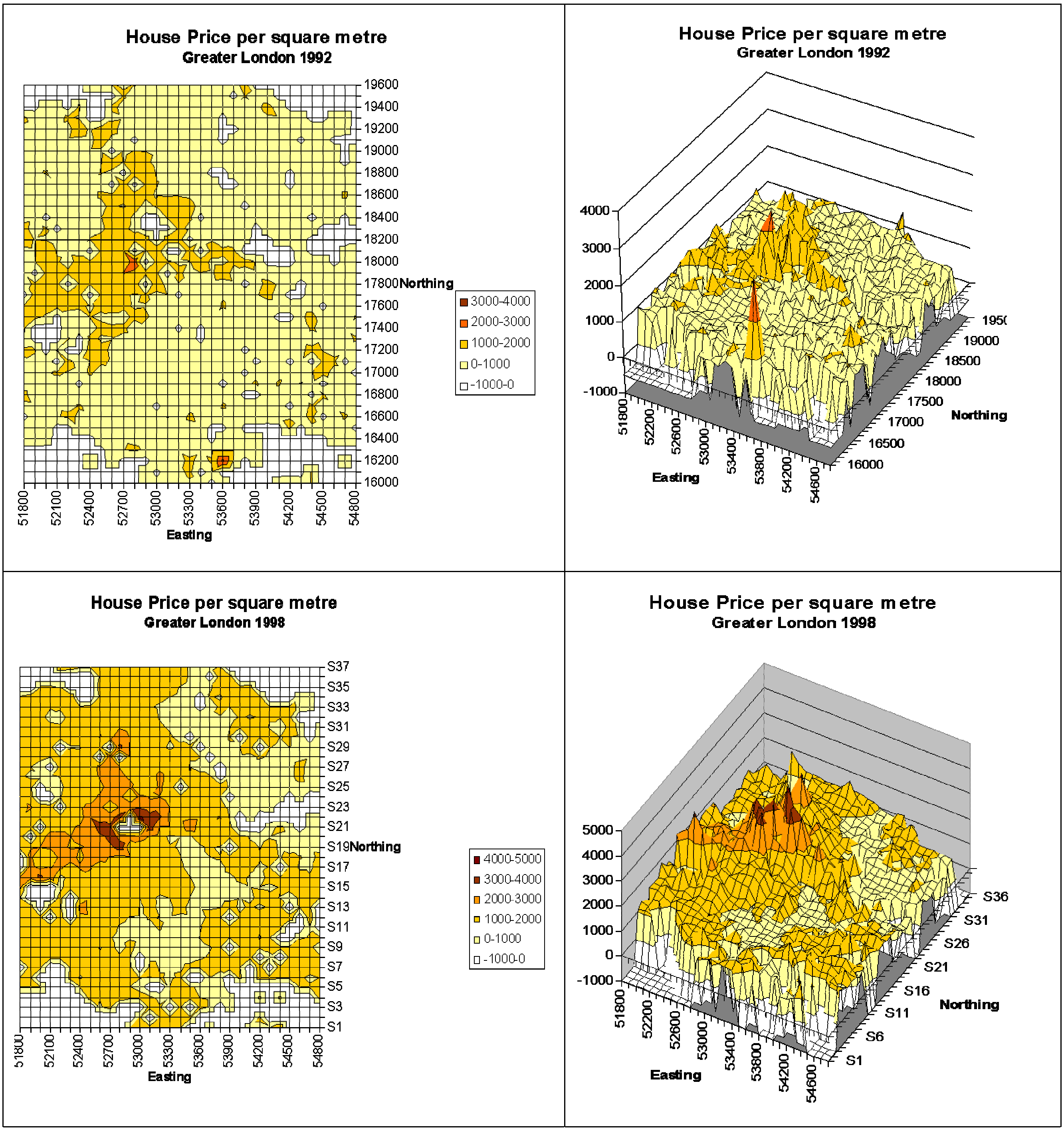
As a simple illustration of how variations in attribute prices across space might be used to define housing submarkets, consider the following plots of price per square foot derived from the Nationwide data. Note that the variable being plotted here is not the estimated attribute price of an extra room derived from hedonic regression (that holds constant all other measured attributes) but simply the ratio of house price to the total number of square metres of each property.

Comparing such snapshots and taking into account random sampling variation and differences in the scales used to draw the contours, it is apparent how one could observe how contours have shifted or stagnated over time. A more robust and precise result would emerge from the grid search of hedonic price regressions (or the Geographically Weighted Regression) but the presentation of results and the readily accessible nature of the output would be the same. Once derived, these submarkets will provide us with the spatial units necessary for further work:

- do house prices diverge or converge across submarkets?
- how do price changes in one submarket feed through to price changes in adjacent submarkets?
- how do turnover rates vary across submarkets?
- how do demand and supply elasticities vary across submarkets?
- how does time to sale vary between submarkets?
- how does LTV, MPPI take-up and other mortgage variables vary between submarkets?
- how does the regeneration impact of new supply vary across submarkets?

Alternatively (or additionally), we might be interested in defining submarkets by one or more of these variables. In the chapters that follow we consider these questions further, and examine in particular, how variations in demand elasticities, regeneration impacts, and turnover rates might be analysed.

Figure 2-4 House Prices per Square Metre in Greater London (1992 vs 1998)



3 Regeneration Effects of New Supply

Plan:

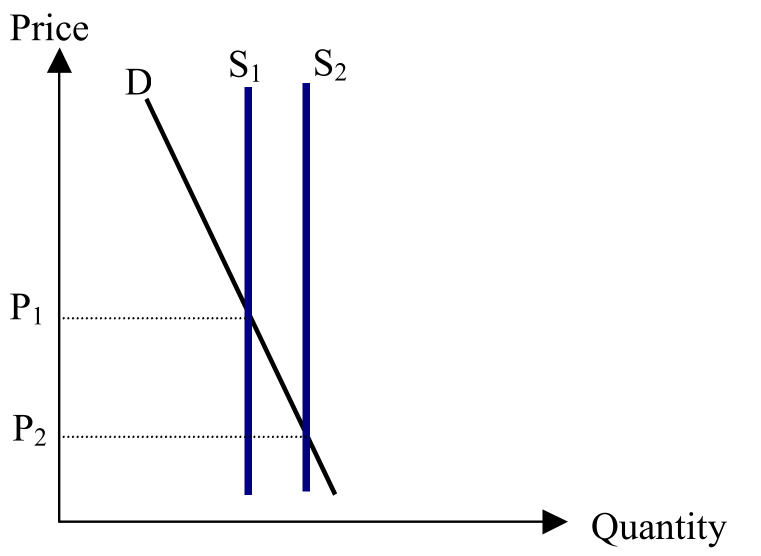
3.1 Introduction

The report has so far concentrated on how submarkets might shape the impact of new supply, a theme elaborated in more detail in chapters 4 and 5 below. Before we progress with this theme, however, it is worth considering how new supply has itself the potential to shape and influence submarkets. Most analyses of new supply tend to focus on the macro effects. The Barker Review, for example, has so far largely concentrated on the macro and regional real house price implications of new housing construction. But local price dynamics are also of importance and need to be at the forefront of the debate on housing supply. We know much less about the indirect price effects generated by new sites on existing housing markets at a micro scale. A pilot study by the Joseph Rowntree Foundation (Pryce and Gibb, 2003) offered some initial glimpses into the possible effects of new construction at the very micro scale. Their results, though preliminary, offer some insight into the possibility of using land planning as a means of area regeneration. This chapter summarises the rationale and findings of the Pryce and Gibb (2003) study and considers ways in which the investigation could be applied more broadly.

3.2 The Neighbourhood Effects of New Supply

Pryce and Gibb (2003) suggest two categories of influence of new construction on house prices. They first describe the direct affect, which corresponds to the conventional outcome following an outward shift of the supply curve. Given a downward sloping demand curve for housing, such a shift would, ceteris paribus, result in a fall in price. This scenario is depicted below in Figure 3-1, where S_1 and S_2 are the supply curves before and after land release, and P_1-P_2 is the fall in price that results.

Figure 3-1 Conventional Effect of Supply Shift



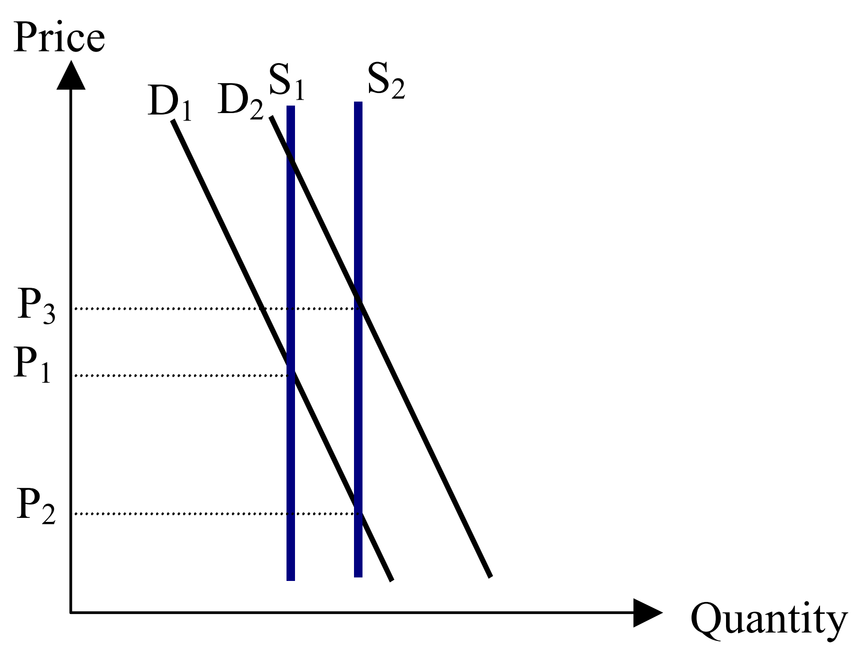
Pryce and Gibb argue that, although this would be the most obvious effect of an increase in new build, it is possible that it may in fact be dominated by indirect effects at the neighbourhood level, particularly if the new build is concentrated, of a substantial number of units, and of a superior quality to existing adjacent housing. Such a development has the potential to give rise to a positive externality effect, purely from the improvement in asthaetics that the new site brings, particularly if it is replacing derelict or former industrial land. In addition to the aesthetic externality effects, the prices of adjacent housing may benefit from an influx of higher income families:

> "… newly constructed dwellings on adjacent plots of land are likely to be purchased by a clientele quite distinct from than those most likely to purchase one of the established dwellings. The decision to purchase a new or old dwelling may for a significant proportion of purchasers precede the location decision – developers market their properties through different channels than owner occupiers, and may have offers that attract particular groups (such as the purchase of the customer's existing dwelling, particularly attractive to those needing to move in a hurry and/or who are having difficulty selling). The new dwellings may well be of a different size, construction type, price, and carry a very different ambience

and 'projected lifestyle' than the established second hand dwellings. The different demand, supply and market marginal valuation of attributes of the new dwellings may effectively distinguish them as a separate sub-market, competing with similar new build sites across the city rather than the adjacent second hand neighbourhood."

For adjacent areas that are already prosperous, therefore, the impact may be negligible or negative. But for deprived areas, there may be a regenerative effect, the new development acting as a signal of an influx of new investment and potentially upward spirals in demand and employment. Local retail and enterprise benefit from the inflow of more affluent households, unemployment rates fall, further retail and amenities emerge, and more workers seek to locate in the area causing house prices to rise further. This scenario is depicted in Figure 3-2 by the outward shift of supply resulting in a outward movement of the demand curve for neighbouring properties. The net result is a price rise from P_1 to P_3.

Figure 3-2 The Regenerative Effect of New Supply



3.3 Gaps in our Understanding of the Regeneration Effect

Whilst there are good intuitive reasons to believe that such regenerative effects of new supply can occur, there is very little analysis currently available that will help us gauge the extent or permanence of these effects. Is the regenerative effect sufficient to dominate the direct price effect or will it merely ameliorate a price fall? If the effect is dominant, how localised is its influence? Is it only properties in the nearest vicinity that benefit or is the impact widespread? How stark is the contrast between the impact of building new supply adjacent to already affluent localities compared with construction adjacent to deprived areas? How large does the new estate have to be for the regenerative impact to have any real effect and what are the long term price implications for both the adjacent markets and the new estate itself?

It might well be that the city-wide impact of the new supply is indeed to reduce prices overall, but there may be strategic reasons for releasing land away from the most affluent areas. If new supply can at the same time regenerate deprived areas and take the heat out of the housing boom in demand hot spots, city planners will have good reason to commend themselves. Such a strategy is unlikely to work on the basis of a serendipitous approach to the release of land. Without economic analysis of the likely urban contours of price response to the construction of a large estate, location decisions will be ill informed and the consequence of new construction unpredictable. Such an analysis requires and understanding of the submarket structure of the existing housing market (chapter 2), the dynamics of those submarkets (chapter 4) and the possible variation in demand elasticities across those submarkets (chapter 5).

There is also the difficult issue of where developers will be willing to build – releasing land does not in itself guarantee an increase in the stock of housing. Developers have to be confident that there will be adequate demand for proposed sites and this often is dependent on contingent adjacent development. The success of one site will be contingent on the development of adjacent sites or projects and so the fear of systemic failure can be a deterrent to development taking place at all. An important omission for many city developers is the availability of information on the liquidity and volatility performance of city submarkets – a point discussed further in chapter 4. Importantly, greater understanding by developers and policy makers about the possible positive externality and regenerative impacts of new supply could go some way to removing some of the uncertainty surrounding new supply. The purpose of the remainder of this chapter, therefore, will be to describe the approach taken by the Pryce and Gibb (2003) attempt at measuring the neighbourhood effects of new supply, and to suggest how estimation might be developed in future.

3.4 The Pryce and Gibb (2003) Approach to Estimating the Regeneration Effect

Pryce and Gibb (2003) develop a model of house purchase where each potential buyer (labelled "bidders") are assumed to have a unique set of preferences for housing attributes. Though personal valuations will vary from individual to individual, the valuations will form a regular distribution for each attribute. Not many potential bidders will place a very valuation on an attribute; not many will place a very high valuation, and most will place a valuation somewhere near the mean. Thus an approximately normal (i.e. 'bell' shaped) distribution of attribute valuations is assumed. Because of the symmetrical nature of the distribution this means that if more than 50% of bidders place a positive valuation on an attribute, then the average valuation will be positive, and visa versa. This approach allows the authors to make use of the standard output of regression analysis to estimate the proportion of successful bidders who place a positive valuation on a particular attribute.

The authors then go onto to conceive of the neighbourhood effect of new supply as a location "attribute" to the characteristics of existing properties. Some buyers will place a negative valuation on the location of a contiguous new estate, others will place a positive valuation on its proximity. The goal is to estimate the proportion of successful house purchases of second hand properties that place a positive valuation on the existence of adjacent new build. To achieve this aim, the notion of "proximity" has to be defined and the authors address this by including the distance to new build as part of the construction of the new supply location-attribute. More precisely, Pryce and Gibb compute, for every second-hand property transaction in their Glasgow database, the total number of recent newbuild on the nearest site divided by the distance to that site (a variable they denote as ACC1_D). They also compute the interaction of the total number of recent newbuild/distance with the area being in the lowest quartile of deprivation (denoted by ACC1_DL); and interaction of the total number of recent newbuild/distance with the area being in the highest quartile of deprivation (ACC1_DH). Having constructed the model, thay attempt to answer six questions:

1. What is the probability of a successful bidder placing a positive valuation on ACC1_D, cet par? That is, what proportion of bidders place a positive value on contiguous new construction?
2. What is the probability of a successful bidder placing a lower valuation on ACC1_DL than bidders generally cet par? That is, what proportion of bidders in areas of low deprivation place a value on contiguous new construction less than the average value placed by all bidders?
3. What is the probability of a successful bidder placing a higher valuation on ACC1_DH than bidders generally, cet par? That is, what proportion of bidders in areas of low deprivation place a value on contiguous new construction less than the average value placed by all bidders?
4. What is the expected valuation on ACC1_D, cet par? That is, for every unit of new construction built one metre away, by what % does the value of the second hand dwelling rise or fall?
5. What is the expected valuation on ACC1_DL, cet par? That is, for every unit of new construction built one metre away in an area of *low* deprivation, by what % does the value of the second hand dwelling rise or fall?
6. What is the expected valuation on ACC1_DH, cet par? That is, for every unit of new construction built one metre away in an area of *high* deprivation, by what % does the value of the second hand dwelling rise or fall?

The authors derive the desired probabilities by using the estimated standard errors of the estimated slope parameters for the attributes in the regression model as approximations to the standard deviations of the attribute valuations made by successful bidders:

standard error of slope parameter on attribute a $\approx$ standard deviation of valuation of a.

This approximation is then used to compute the desired probabilities, which are in fact equal to the one tail significance level on each regression slope parameter in a hedonic price model. The econometric model estimated is as follows:

$$P^H = f(X, \; CBD, \; DEP, \; ACC1\_D, \; ACC1\_DH, \; ACC1\_DL, \; TTS, \; Time)$$

where,

X	= dwelling characteristics
CBD	= distance to Glasgow city centre (taken as Queen Street Station)
DEP	= deprivation score (measure of existing social neighbourhood effects)
ACC1_D	= total recent new build on nearest site / Distance to nearest site
ACC1_DH	= interaction with high deprivation
ACC1_DL	= interaction with low deprivation
TTS	= Time to sale
Time	= quarter dummies

The second-hand transactions data was supplied by GSPC and the new construction data by Strathclyde Structure Plan Team. The latter covers the period 1998-2001 and includes information on location of site (grid reference, address) and nature of site including, site size and capacity (how many potential units), name of owner, tenure, name of builder, development type (brownfield/greenfield), number of completions, planned timetable of completions. The study was only a pilot analysis since too analyse the effect of *all* sites on each sale (there were a total of 18,000 property transaction observations and around 5,000 land supply sites) would require transposing a 90 million cell table, so the authors report only on analysis carried out on the first 4,000 sales of 1999.

3.4.1 Results

The final regression was run on a subsample of 900 cases which had a good overall fit (Adjusted R square = 0.75, indicating that around three quarters of the variation in the dependent variable was being explained by the model). The dwelling attribute variables are jointly highly statistically significant, as were the two main geographical variables, distance to city centre and deprivation score, and the time to sale variable, TOM. The authors estimate that around 72% of bidders place a positive value on contiguous new construction, and 63% of bidders in areas of low deprivation place a value on contiguous new construction less than the average value placed by all bidders. In contrast, 94% of bidders in areas of high deprivation place a value on contiguous new construction *greater* than the average value placed by all bidders.

The average valuation of ACC1_D, ACC1_DL, ACC1_DH are given in the model in terms of the % impact on selling price of a one unit increase. For example, the coefficient on ACC1_D is 0.0018, which suggests that for every unit of new construction built one metre away, the value of the second hand dwelling rises by 0.18%, ceteris paribus. If the second hand property is in an area of low deprivation, this positive impact is reduced to 0.03%, whereas if the dwelling is an area of high deprivation, this positive impact is increased to 33%. Clearly, then, the effects of new construction on the price of second hand properties are minimal in most areas, and particularly small in areas of low deprivation. However there was evidence to suggest that the effects of new construction may have a substantial positive effect on second hand properties in the most deprived areas.

3.5 Suggestions for Future Research

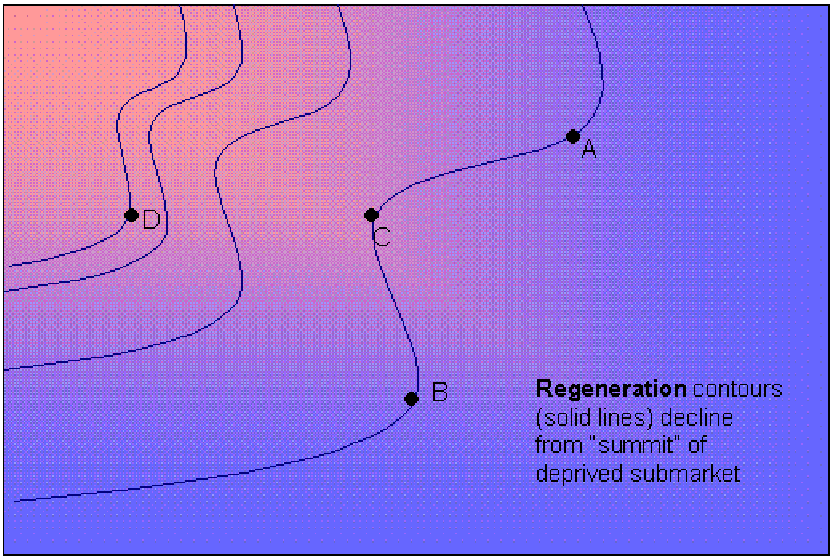
The most involved aspect of the Pryce and Gibb study was the acquisition, cleaning and preparation of the local authority land supply data with the GSPC transactions data. Clearly, this would be a very time consuming process to repeat for each major city in the UK (though, arguably a very valuable exercise as the merged data could be used for a range of other analyses). The advent of the newly expanded CML data means that estate agent data is no longer so critical to the analysis (though there is again much to be gained from seeking to negotiate with other large estate agencies across the UK for similar kinds of data). Most importantly, both the CML and Land Registry data have information on the sales of new dwellings. Since most developers only construct a house if they already have a guaranteed purchaser, such figures are likely to reflect total new supply. Whilst the Land Registry data do not identify the boundaries of construction sites, they do provide location information of the new dwellings which could be used to compute distance to sales of second hand dwellings. (Distance from a CML sale, for example, to the nearest new house, second nearest new house, third nearest new house etc. could be calculated by applying Pythagoras' Theorem to easting and northing information). These distances could be entered into hedonic price regressions as means of estimating the local impact of new supply. It will also enable us to identify whether there is a difference between the single plot new build and large new estates (are there, for example, "tipping points" in the size of new estate). As such, the replication of the Pryce and Gibb analysis at a much larger scale is now feasible.

There are in addition a range of other issues not incorporated into the Pryce and Gibb study which are worthy of further exploration:

3.5.1 Gradations in the Regeneration Effect

A recurring theme in this report is whether the variations in regeneration effects should be measured using methods that segment cities and regions into subdivisions of approximately equal regenerative impact, or whether the gradations in regenerative impact should be calibrated along a continuum, arriving at a surface plot or a series of contour lines. Geographically Weighted Regression (see appendix at the end of this chapter) would be an ideal medium to investigate such questions. The goal would be to establish how steeply the contours of the regenerative effect decline away from the most deprived areas, as depicted in the stylised diagram of Figure **3-3**. Locating new construction at point D would result in a much greater regenerative impact that at sites A, B and C. How steep these contours decline is a question of considerable import given that developers are least likely to want to build adjacent to depressed areas. However, if the contours of decline of the regeneration effect can be established, it may be possible to identify the location(s) of optimal trade-off between developer's most desired location and the most beneficial for the city as a whole. It may be, for example, that developers greatly prefer site A to sites B and C. There is, however, no difference in the regeneration effects between these three sites and so such a map would allow planners to choose A over points B and C.

Figure 3-3 Contours of the Regeneration Effect



3.5.2 Consideration of more than the nearest site:

- Only so far looked at nearest new build site but perhaps we should extend this to say the 5 nearest sites.Some properties very close to several sites and so we need to take account of this.
- Perhaps use average of 5 nearest sites?

3.5.3 Derelict Land

- Need to look at effect of derelict land
- this also can have a neighbourhood effect
- conversion of unsightly brownfield to a shiny new housing estate likely to have a positive neighbourhood effect

3.5.4 Cross-elasticity of demand:

- effect of new construction on 2nd hand house prices will be partly effected by the price of those dwellings.
- If competitively priced, will drive down local 2nd hand house prices.
- If relatively upmarket, expensive new build, may have an anticipated gentrification effect and $\Rightarrow$ $\uparrow$ local demand $\Rightarrow \uparrow P^H$
- could use SASINES data on new build prices
- Future work:
  Time to Sale and Bid-Offer Spreads
- Instead of price as the dependent variable, we could model time on the market or Bid-Offer Spreads instead
- We would then be able to see how new construction and land use affect time on the market or the difference between asking and selling prices in different areas

3.5.5 Commercial Construction:

- Does new commercial construction have a more beneficial effect on adjacent areas than residential construction?
- Is there a difference in the relative benefits of commercial vs residential construction on depressed rather than booming areas?
- Affluent neighbourhoods may for example benefit less from the employment effect of new local business accommodation

3.5.6 Time to Sale and Offers Over

- What is the impact of new construction on the time on the market of second hand properties?
- Does the impact vary depending on the affluence/socio-economic make-up of the area?
- What factors influence the gap between asking and selling prices?

3.5.7 Other issues:

- For interaction term with high deprived areas, use highest decile (rather than highest quartile) once you have enough observations.
- Spatial autocorrelation
- Systematic determination of the location of new build/land release: i.e. in areas that predicted no effect would take place.
- Sample selection effect – do missing values have a systematic determination?
- Polycentricity – should other CBDs be included? (Paisley, East Kilbride…)

- Interactions between attributes:
- (Type) with (Attributes) where:
- (Type) = House, flat etc.
- (Attributes) = bathrooms, bedrooms, number of public rooms, Garden, garage, GCH, spacious, alarm, mature established area, bay etc.
- (Temperature) with GCH
- (Temperature) = (DG, spacious, winter sale, ensuite, Victorian, stone, traditional)
- Views with garden, deprivation, CBD, conservatory, **new build**, summer sale.
- Garden with summer sale.
- **Mature est area** with **new build**
- Alarm with deprivation
- Parking with distance to CBD, deprivation
- Demolitions
- Sample size problems for number of dwellings in deprived areas
- Colin Jones' 3 papers.
- Remove/analyse new dwellings from GSPC records
- Effect of current construction: appears to be strongly negative – due to uncertainty? Noise factor? Myopia of buyers? But future expected new build has a positive effect…
- Land banking
- Deprivation score of next nearest post code sector divided by the distance to that post code sector
- What is the impact of **remaining capacity** on house prices:
- Impact of land banking
- Interaction with brownfield and greenfield.

3.6 Appendix to Chapter 3: Geographically Weighted Regression

Consider the following model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

where y is the dependent variable, x_1 and x_2 are the independent variables, β_0, β_1, and β_2, are the parameters to be estimated, and ε is a random error term, assumed to be normally distributed. A GWR version of the model would permit the parameter estimates to vary locally:

$$y(\boldsymbol{g}) = \beta_0(\boldsymbol{g}) + \beta_1(\boldsymbol{g})x_1 + \beta_2(\boldsymbol{g})x_2 + \varepsilon$$

where ***(g)*** indicates that the parameters are to be estimated at a location whose coordinates are given by the vector **g**. The OLS solution is traditionally derived from,

$$\beta = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

compared with GWR weighting scheme:

$$\beta(\mathbf{g}) = (\mathbf{X}^T\mathbf{W}(\mathbf{g})\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}(\mathbf{g})\mathbf{Y}$$

"The weights are chosen such that those observations near the point in space where the parameter estimates are desired have more influence on the result than observations further away. Two functions we have used for the weight calculation have been (a) bi-square and (b) Gaussian. In the case of the Gaussian scheme, the weight for the *i*th observation is:

$$w_i(\mathbf{g}) = e^{-(d/h)^2}$$

where $\underline{d}$ is the Euclidean distance between the location of observation *i* and location **g**, and $\underline{h}$ is a quantity known as the bandwidth. (There are similarities between GWR and kernel regression). One characteristic that is not immediately obvious, is that the locations at which parameters are estimated need not be the ones at which the data have been collected.
The resulting parameter estimates may be mapped in order to examine local variations in the parameter estimates." (Fotheringham web site).

3.6.1 GWR References:

Fotheringham, A.S., Brunsdon, C., and Charlton, M.E., 2002, Geographically Weighted Regression: The Analysis of Spatially Varying Relationships, Chichester: Wiley (to be published later this year)

Fotheringham, A.S., Brunsdon, C., and Charlton, M.E., 2000, Quantitative Geography, London: Sage

Brunsdon, C., Aitkin, M., Fotheringham, A.S., and Charlton, M.E., A comparison of random coefficient modelling and geographically weighted regression for spatially non-stationary regression problems, Geographical and Environmental Modelling, 3(1), 47-62

Brunsdon, C., Fotheringham, A.S., and Charlton, M.E., 1999, Some notes on parametric signficance tests for geographically weighted regression, Journal of Regional Science, 39(3), 497-524

Fotheringham, A.S., Brunsdon, C., and Charlton, M.E., 1998, Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis, Environment and Planning A, 30(11), 1905-1927

Brunsdon, C., Fotheringham, A.S., and Charlton, M.E., 1998, Geographically weighted regression - modelling spatial non- stationarity, Journal of the Royal Statistical Society, Series D-The Statistician, 47(3), 431-443

Brunsdon, C., Fotheringham, A.S., and Charlton, M.E., 1998, Spatial nonstationarity and autoregressive models, Environment and Planning A, 30(6), 957-993

A. Stewart Fotheringham, M.E. Charlton and C. Brunsdon "Measuring Spatial Variations in Relationships with Geographically Weighted Regression", Chapter 4 in Recent

Developments in Spatial Analysis, Spatial Statistics, Behavioral Modeling and Neurocomputing M.M. Fischer and A. Getis (eds.), Springer-Verlag: London,

Chris Brunsdon, A. Stewart Fotheringham and M.E. Charlton, 1997, "Geographical Instability in Linear Regression Modelling - A Preliminary Investigation" pp 149-158 in New Techniques and Technologies for Statistics II IOS Press: Amsterdam/Oxford/Washington

M.E. Charlton, A. Stewart Fotheringham and C. Brunsdon, 1997, "The Geography of Relationships: An Investigation of Spatial Non-Stationarity" Chapter 2, pp 23-47 in Spatial Analysis of Biodemographic Data J-P. Bocquet-Appel, D. Courgeau and D. Pumain (eds.) John Libbey Eurotext: Montrouge.

A. Stewart Fotheringham, M.E. Charlton and C. Brunsdon, 1997, "Two Techniques for Exploring Non-stationarity in Geographical Data" Geographical Systems, 4: 59-82.

A. Stewart Fotheringham, 1997, "Trends in Quantitative Methods I: Stressing the Local" Progress in Human Geography, 21: 88-96

C. Brunsdon, A.Stewart Fotheringham and M.E. Charlton, 1996, "Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity", Geographical Analysis, 28(4), 281-298

A. Stewart Fotheringham, M.E. Charlton and C. Brunsdon, 1996, "The Geography of Parameter Space: An Investigation into Spatial Non-Stationarity", International Journal of Geographic Information Systems, 10: 605-627

4 Spatial Variation in Market Dynamics

Plan:

4.1 Introduction

So far the report has considered the literature on the existence and nature of submarkets and whether there may be a neighbourhood effect of new supply. While there has been some discussion of dynamic processes, such as the evolution of submarket boundaries over time, the concepts have largely been couched in static terms or conceived of as snapshots in time. In this chapter, we bring our attention to bear on some of the distinctively dynamic aspects of submarkets and it is here that connections with macro housing policy are most clearly made. In particular, the issue of how macro price indices are calculated is discussed and it is shown that whether or not the new supply falls into the category of dwellings that trade relatively frequently will have a profound impact on its final effect on price trends because such indices are dominated by properties that sell frequently. However, even if new supply does indeed fall into this category, the benefits may be purely illusionary reflecting computation bias in price indices rather than any true economic benefit. For example, if length of stay at least in part reflects household satisfaction with the dwelling and its surroundings, then one of the goals of supply policy should be to increase the number of properties which people will be happy to inhabit long tern. The extent and importance of submarket variation in price volatility is then discussed and is presented as an important determinant of local supply investment. Finally, the chapter considers how dynamics may vary across submarkets by looking at possible models of duration properties off and on the market.

4.2 Variations in the Frequency of Sale Across Space

One of the complicating factors in understanding the impact of new supply on house price trends is the variation in the frequency of sale of properties across different parts of the country, and even within a particular city. Turnover rates also vary procyclically with the housing market cycle (Gatzalf and Haurin, 1998). This variation in the rate at which dwellings enter and leave the market will obviously affect the impact on price indices of new supply. This is because price indices are usually based on transactions data without controlling for the frequency of sale. As a result, some properties will enter the market many times over a ten year period, whereas others will only enter once or not at all. Clearly, it is those properties that repeatedly enter the market that will drive a house price index if that index is computed on the average of all transactions. The question is whether properties that trade frequently can be considered as a separate submarket (either because of location or attributes) from those that rarely come on to the market. If so, then an important question to ask is: to which of these submarkets will newly constructed dwellings belong? The answer will depend again on the location and attributes of those new properties.

Suppose first that a new dwelling enters the infrequently sold submarket. As such, it is purchased upon completion by a family that do not move for the next ten years. If the construction of this dwelling and those of a similar ilk successfully relieve price inflation in the infrequently traded sector, then to what extent will this be reflected in price indices based on the aggregate volume of trade? The answer is not very much. For one thing, alleviating price pressures in a low turnover area like Argyll and Bute may do very little to affect prices in a high turnover such as the West End sector of Glasgow. The two are separate submarkets and viewed as such by prospective purchasers. Furthermore, movements in Argyll and Bute prices will have less of an effect on Scotland-wide house price indices than those in the West End simply because each property in Argyll and Bute carries a lower weight in the computation of the headline house price index.

4.2.1 What Target?

This raises the important issue of what measure macro supply policy should use as its target. If the target is to reduce house price inflation, then policy makers have to be aware that current measures carry with them an intrinsic bias towards frequently traded properties. Does this computational bias reflect the desired policy bias? It may be, for example, that there are differential house price inflation rates between different sectors of the housing market according to frequency of sale. If less frequently traded property is experiencing higher inflation than high turnover dwellings, then there may be a case for giving greater policy priority to alleviating demand pressures in the infrequently traded sector. The case might be supported if there is reason to believe that length of stay is a measure of consumer satisfaction. Given the emotional and pecuniary upheaval associated with moving house, a family will only consider moving if they anticipate a significant improvement in living standards from doing so. A major reason for moving will therefore be dissatisfaction with current living conditions due to lack of space, unsuitable layout or due to neighbourhood problems. A policy that achieves a major increase in the stock of frequently sold dwellings, and in so doing, reduces the average price for such dwellings, may have only achieved its superficial policy goal, but in reality effected an increase in the proportion of the total housing stock with which homeowners are generally dissatisfied. The policy will only exacerbate the price differential between desirable and undesirable properties.

Understanding the role of frequency of sale is not just of relevance to policy makers. The bias it implies for house price indices has the potential to distort private sector investment decisions. In the least, the lack of an appropriately adjusted house price index will be a source of uncertainty for potential investors in either new construction or potential landlords. To make appropriate financial decisions, such investors need to be able to readily compare the performance of the housing sector with that of other tradable assets such as stocks and bonds and so lack of information in the housing asset market relative to other asset markets will further reduce the attractiveness of housing construction as a destination for investment funds.

4.2.2 An Example of house price bias from the West of Scotland

To illustrate the kind of biases endemic in existing price indices, consider the following table which lists the number of properties in each West of Scotland local authority that sold either once, twice, three times, four times or five or more times in the 1991 to 2000 period. The table also presents the proportion of sales in each area that fall into each of these repeat sales categories. The data are drawn from SASINES records on the West of Scotland and demonstrates the kind of analysis that could be done for the rest of the UK using Land Registry data. There is clearly considerable variation in repeat sales even within the West of Scotland. In the City of Glasgow, for example, nearly 30% of properties transacted sold twice, and 10% sold three times. This contrasts with Argyll and Bute where less than 18% sold twice and only 3.6% sold three times. Overall, 63.3% of properties that sold came on the market only once, 25.9% sold twice, 8.4% sold three times, 1.9% sold four times and 0.5% sold five or more times. It is likely that there are similar intra and inter settlement disparities in the proportion of properties sold at all.

4.2.3 Bias in rates of change?

For house price indices to be distorted by frequency of sale, however, there would have to be different patterns of house price values for different rates of turnover. Is there evidence for this kind of bias in a UK context (most studies of these issues are from a US context – for example: Gatzlaff and Haurin 1994, 1997, 1998; Fisher et al 2003; Hwang and Quigley 2004)? In **tables 4-2 and 4-3** it can be seen that the mean house price tends to be lower for properties which frequently sell (categories with small samples – less than 200 observations – should excluded because of the high variation in prices and the absence of any mix-adjustment). A notable exception is the City of Glasgow which is a very heterogeneous area and likely to be biased by the West End which is a generally considered a separate submarket (it is a high value area with high turnover). Further analysis would be needed to ascertain whether the effect was caused by the heterogeneity of the City of Glasgow (we would like to know, for example, whether *within* the West End, more frequently traded properties tend to be of lower value). Nevertheless, it is clear that there is good reason to believe that house prices vary systematically by frequency of sale and that grouping all properties together without accounting for this non-randomness is likely to result in house price indices giving a biased picture of the level of prices at a given point in time. There is also evidence here to support the argument that in many areas, properties that remain off the market for long periods yield higher yields of "satisfaction" (whether due to location, size or quality) as they tend to sell for a higher price than frequently sold dwellings. This might reflect simple lifecycle patterns or it might also be the result of information asymmetries in the housing market (buyers know less than sellers about the true quality of the dwelling and the desirability of its location) and this can result in the stock of dwellings for sale at a given point in time being characterised by a disproportionate number of poor quality properties (a process called "adverse selection – see Akerlof's **1963** seminal theoretical paper on the "Market for Lemons").

Table 4-1 Variation in the Frequency of Sale of Properties in the West of Scotland

	Number of times a dwelling has sold in the 1991-2000 period					
	1	**2**	**3**	**4**	**5+**	**All**
Argyll & Bute	14,815	3,422	687	96	73	19,093
	77.6%	17.9%	3.6%	0.5%	0.4%	100.0%
City of Glasgow	83,971	40,255	14,199	3,089	838	142,352
	59.0%	28.3%	10.0%	2.2%	0.6%	100.0%
East Ayrshire	16,136	5,396	1,509	332	35	23,408
	68.9%	23.1%	6.5%	1.4%	0.2%	100.0%
East Dunbartonshire	13,796	6,774	1,967	432	65	23,034
	59.9%	29.4%	8.5%	1.9%	0.3%	100.0%
East Renfrewshire	13,696	5,944	1,989	486	165	22,280
	61.5%	26.7%	8.9%	2.2%	0.7%	100.0%
Inverclyde	13,521	4,560	1,232	305	124	19,742
	68.5%	23.1%	6.2%	1.5%	0.6%	100.0%
North Ayrshire	21,235	6,484	1,839	352	52	29,962
	70.9%	21.6%	6.1%	1.2%	0.2%	100.0%
North Lanarkshire	41,634	16,570	5,685	1,388	349	65,626
	63.4%	25.3%	8.7%	2.1%	0.5%	100.0%
Renfrewshire	27,292	10,742	3,211	677	140	42,062
	64.9%	25.5%	7.6%	1.6%	0.3%	100.0%
South Ayrshire	18,310	6,364	1,731	328	65	26,798
	68.3%	23.8%	6.5%	1.2%	0.2%	100.0%
South Lanarkshire	41,467	18,552	6,643	1,747	744	69,153
	60.0%	26.8%	9.6%	2.5%	1.1%	100.0%
West Dumbartonshire	874	403	126	30	5	1,438
	60.8%	28.0%	8.8%	2.1%	0.4%	100.0%
West Dunbartonshire	11,015	4,398	1,278	302	53	17,046
	64.6%	25.8%	7.5%	1.8%	0.3%	100.0%
Total	317,762	129,864	42,096	9,564	2,708	501,994
	63.3%	25.9%	8.4%	1.9%	0.5%	100.0%

Table 4-2 Average House Prices in 1991 by no. times sold in previous 10 years

	City of Glasgow			East Renfrewshire			North Ayrshire		
	Mean	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**
1	£ 37,288	£ 43,320	9192	**£ 65,173**	£ 42,210	1235	£ 34,165	£ 26,116	2399
2	£ 37,676	£ 27,346	4138	£ 61,638	£ 35,940	607	**£ 35,754**	£ 21,518	755
3	**£ 38,028**	£ 27,370	1492	£ 57,829	£ 28,508	221	£ 32,625	£ 18,211	236
4	£ 37,325	£ 18,648	343	£ 53,135	£ 24,692	54	£ 35,084	£ 29,721	39
5+	£ 38,934	£ 18,816	103	£ 38,345	£ 18,689	15	£ 22,550	£ 8,603	4
All	£ 37,477	£ 37,627	15268	£ 62,912	£ 38,900	2132	£ 34,405	£ 24,732	3433

	Renfrewshire			East Ayrshire			East Dunbartonshire		
	Mean	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**
1	**£ 41,103**	£ 72,331	2895	**£ 33,254**	£ 27,616	1688	**£ 64,404**	£ 41,465	1406
2	£ 38,038	£ 22,466	1159	£ 32,814	£ 18,042	634	£ 59,634	£ 36,380	742
3	£ 37,969	£ 23,037	363	£ 33,258	£ 21,741	200	£ 55,912	£ 31,089	244
4	£ 37,008	£ 18,206	75	£ 39,025	£ 20,607	43	£ 48,358	£ 28,829	60
5+	£ 35,874	£ 22,070	17	£ 22,848	£ 6,030	3	£ 46,475	£ 58,754	6
All	£ 39,975	£ 59,502	4509	£ 33,230	£ 25,017	2568	£ 61,686	£ 38,970	2458

	South Lanarkshire			Argyll & Bute			South Ayrshire		
	Mean	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**
1	**£ 38,627**	£ 34,187	4018	**£ 43,248**	£ 39,470	1618	**£ 47,515**	£ 33,222	1783
2	£ 37,164	£ 26,054	1999	£ 44,813	£ 31,904	343	£ 47,282	£ 29,145	703
3	£ 36,781	£ 22,691	782	£ 48,289	£ 33,436	81	£ 40,610	£ 20,368	210
4	£ 35,590	£ 18,257	200	£ 49,023	£ 33,865	13	£ 38,817	£ 16,339	44
5+	£ 34,908	£ 15,532	86	£ 33,197	£ 11,162	24	£ 41,729	£ 12,090	16
All	£ 37,880	£ 30,398	7085	£ 43,622	£ 37,861	2079	£ 46,757	£ 31,165	2756

	North Lanarkshire			West Dumbartonshire			Inverclyde		
	Mean	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**
1	£ 30,933	£ 25,665	3807	**£ 32,372**	£ 23,967	1309	**£ 36,581**	£ 34,611	1296
2	£ 31,494	£ 18,954	1583	£ 32,760	£ 19,008	558	£ 34,371	£ 34,460	416
3	**£ 31,565**	£ 16,507	587	£ 30,861	£ 15,828	181	£ 33,627	£ 22,444	123
4	£ 30,832	£ 15,438	142	£ 36,015	£ 18,041	46	£ 31,976	£ 14,070	28
5+	£ 34,503	£ 21,817	31	£ 33,500	£ 18,053	7	£ 59,811	£ 46,267	7
All	£ 31,153	£ 23,109	6150	£ 32,428	£ 21,989	2101	£ 35,913	£ 33,762	1870

Table 4-3 Average House Prices in 2000 by no. times sold in previous 10years

	City of Glasgow			East Renfrewshire			North Ayrshire		
	Mean	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**
1	£54,716	£ 50,744	9382	**£98,288**	£106,418	1548	**£ 48,074**	£ 43,868	2165
2	**£57,202**	£ 51,253	4991	£82,881	£ 53,830	579	£ 44,403	£ 26,711	840
3	£55,827	£ 40,150	1929	£81,285	£ 45,267	173	£ 42,366	£ 24,266	258
4	£52,912	£ 35,444	409	£64,299	£ 43,925	47	£ 39,101	£ 34,378	45
5+	£50,802	£ 35,421	107	£53,746	£ 23,727	23	£ 42,873	£ 24,966	8
Total	£55,512	£ 49,402	16818	£92,177	£ 91,529	2370	£ 46,565	£ 38,785	3316

	Renfrewshire			East Ayrshire			East Dunbartonshire		
	Mean	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**
1	**£59,377**	£ 49,949	2777	**£49,320**	£167,645	2039	**£ 89,050**	£ 68,914	1403
2	£46,410	£ 29,625	1251	£42,165	£ 21,329	672	£ 81,549	£ 56,465	718
3	£46,528	£ 27,457	402	£41,306	£ 15,805	189	£ 79,464	£ 51,879	201
4	£40,986	£ 22,838	79	£39,881	£ 16,689	56	£ 62,744	£ 27,121	47
5+	£35,475	£ 21,167	16	£30,456	£ 15,055	7	£ 65,669	£ 58,961	9
Total	£54,245	£ 43,514	4525	£46,963	£139,552	2963	£ 85,366	£ 63,578	2378

	South Lanarkshire			Argyll & Bute			South Ayrshire		
	Mean	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**
1	**£57,961**	£ 46,401	4867	**£70,301**	£193,571	1330	**£ 63,964**	£ 55,403	1876
2	£53,297	£ 63,628	2168	£62,922	£ 52,241	320	£ 59,780	£ 38,398	750
3	£44,992	£ 27,455	802	£67,252	£ 65,024	64	£ 50,969	£ 26,532	232
4	£45,693	£ 24,788	222	£99,245	£ 66,275	11	£ 47,442	£ 19,694	47
5+	£40,167	£ 18,165	95	£62,800	£ 49,241	7	£ 44,235	£ 9,214	7
Total	£54,904	£ 49,763	8154	£68,979	£171,676	1732	£ 61,537	£ 49,356	2912

	North Lanarkshire			West Dumbartonshire			Inverclyde		
	Mean	**SD**	**n**	**Mean**	**SD**	**n**	**Mean**	**SD**	**n**
1	**£46,240**	£ 37,045	4939	**£46,491**	£ 30,615	1294	**£ 49,914**	£ 50,833	1521
2	£42,189	£ 24,198	2008	£42,313	£ 21,620	558	£ 48,035	£ 33,953	553
3	£39,942	£ 20,878	725	£39,408	£ 18,942	175	£ 47,349	£ 26,385	166
4	£36,475	£ 17,385	191	£40,053	£ 11,624	42	£ 37,815	£ 17,469	39
5+	£35,187	£ 17,821	41	£39,961	£ 19,237	5	£ 37,372	£ 14,939	17
Total	£44,340	£ 32,590	7904	£44,623	£ 27,388	2074	£ 48,977	£ 45,271	2296

The really crucial question, however, is whether there are different rates of house price change across the different rates of property turnover. If so, frequently sold properties could be characterised as a different submarket. In **Table 4.2** it can be seen that the increase in prices tends to be greater for properties that sell only once (where the sample is less than 200 in either 1991 or 2000 the figures should be treated with caution as the standard deviation of house prices is so large that very large samples are needed to give reliable estimates). As a result, for most of the local authorities listed, using the change in average of all properties under estimates the rate of growth of houseprices. Note that if a repeat sales index were used, in most Las this would result in even greater bias. In some areas the difference is enormous. In Renfrewshire, for example, the per centage increase in average prices from 1991 to 2000 was double that of either properties that sold twice or three times (44.5% compared with 22% and 22.5%), and four times that of properties sold four times (44.5% compared with 10%). An exception to the rule is again the City of Glasgow, the figures for which are probably distorted by the West End.

It could be argued that, to some extent, repeat sale bias may be mitigated by standard methods for controlling for the mix of dwellings coming onto the market (hedonic techniques, for example). For this mitigation to be effective, repeat sale patterns would have to fall along dwelling attribute lines, and the relevant attributes would have to be adequately controlled for in the mix-adjustment procedure. It is highly unlikely that existing mix adjusted indices adequately capture this effect, however, since there is likely to be a sub-city level spatial dimension to the repeat sales process and existing indices do not have this level of spatial refinement (see chapter 2 of Meen and Andrew, 1998, for a summary of how existing house price indices are computed).

Table 4-4 % Change in Average House Price by No. Times sold in 1999-2000 Period

	City of Glasgow		East Renfrewshire		North Ayrshire	
	Mean	**min n**	**Mean**	**min n**	**Mean**	**min n**
1	46.7%	9192	**50.8%**	1235	**40.7%**	2165
2	**51.8%**	4138	34.5%	579	24.2%	755
3	46.8%	1492	40.6%	173	29.9%	236
4	41.8%	343	21.0%	47	11.4%	39
5+	30.5%	103	40.2%	15	90.1%	4
All	48.1%	15268	46.5%	2132	35.3%	3316

	Renfrewshire		East Ayrshire		East Dunbartonshire	
	Mean	**min n**	**Mean**	**min n**	**Mean**	**min n**
1	**44.5%**	2777	**48.3%**	1688	38.3%	1403
2	22.0%	1159	28.5%	634	36.7%	718
3	22.5%	363	24.2%	189	**42.1%**	201
4	10.7%	75	2.2%	43	29.7%	47
5+	-1.1%	16	33.3%	3	41.3%	6
All	35.7%	4509	41.3%	2568	38.4%	2378

	South Lanarkshire		Argyll & Bute		South Ayrshire	
	Mean	**min n**	**Mean**	**min n**	**Mean**	**min n**
1	**50.1%**	4018	**62.6%**	1330	**34.6%**	1783
2	43.4%	1999	40.4%	320	26.4%	703
3	22.3%	782	39.3%	64	25.5%	210
4	28.4%	200	102.4%	11	22.2%	44
5+	15.1%	86	89.2%	7	6.0%	7
All	44.9%	7085	58.1%	1732	31.6%	2756

	North Lanarkshire		West Dumbartonshire		Inverclyde	
	Mean	**min n**	**Mean**	**min n**	**Mean**	**min n**
1	**49.5%**	3807	**43.6%**	1294	**36.4%**	1296
2	34.0%	1583	29.2%	558	39.8%	416
3	26.5%	587	27.7%	175	40.8%	123
4	18.3%	142	11.2%	42	18.3%	28
5+	2.0%	31	19.3%	5	-37.5%	7
All	42.3%	6150	37.6%	2074	36.4%	1870

4.3 Correcting for Sample Selection Bias

One way of viewing the sample selection problem is as one of omitted variable bias (Heckman 1979) where the omitted variable in the house price equation is the probability of the property coming onto the market. Gatzlaff and Haurin (1998) take this view and use logistic regression to estimate the probability of a property coming onto the market. This estimated probability[1] can then be entered into the sale price equation to correct for sample selection bias. Unfortunately, the applicability of this approach to the UK context is limited since comprehensive data on unsold properties are rarely available. (Note that there are also problems with the data set used by Gatzalff and Haurin in that their data is “limited to single-family detached homes with between 600 and 6000 square feet of living area and less than five acres of land” op cit, p.209).

An alternative approach would be to make use of the repeat sales information that could potentially be gleaned from SASINES and Land Registry data. If fifteen to twenty years of Land Registry data can be compiled for a particular area, while information on properties that do not sell at all would not be available, we would be able to examine the nature of any property that sells at least once during that period. Duration modelling techniques could then be applied to explain the length of time the property remains off the market using techniques that control for “censored” observations -- properties that sold once but currently remain off the market. Since duration off the market varies with market buoyancy, this bias could be controlled for by predicting the hazard rate for each property for a set of “controlled” market conditions. This hazard rate could then be entered into the house price equation to control for sample selection bias.

Alternatively, one might conceive of long-stay vs short-stay properties as different submarkets. In other words, the property types and locations are sufficiently different in the eyes of purchasers that they cannot be conceived of as close substitutes. If so, it might be more appropriate to estimate separate price regressions for long-stay vs short-stay properties. As such, the duration analysis described above would be used to categorize properties into those that frequently sell, and those that

---

[1] More precisely, the inverse Mills ratio is calculated.

infrequently sell. Grid-search procedures could be used to test for "structural breaks" (i.e. shifts) in the house price parameters. Perhaps more usefully, different rates of house price change could be investigated amongst properties of different rates of turnover. It might be, for example, that properties that sell infrequently experience a higher rate of growth than those that frequently come onto the market.

Applying duration analysis would also overcome an important potential weakness in the Gatzlaff and Haurin (1998) analysis. By applying logit rather than a duration model approach, Gatzlaff and Haurin implicitly assume that there is no "duration dependence" in the process by which properties come onto the market. That is, the longer a property remains off the market does not in any way affect the probability of it entering the market in the next period. Without testing whether this assumption is generally valid, a source of bias in the adjustment developed by these authors cannot be ruled out. Moreover, the distinction between properties that sell and those that do is to some extent a false dichotomy, or at least an incomplete one. The real issue is frequency of sale. A window of ten years of all property transactions, for example, will not include all properties since some will not sell at all. However, it will most probably include all types of properties. Even though a there may exist a type of property that sells once in twenty years, provided this class of properties is of reasonable size, it is likely that a number of these properties will trade within the ten year window. So the ten year window should give a random sample of all levels of frequency of sale. Application of censored duration techniques should adequately control for those properties that sell only once in this period.

If a single, overall house price index was still desirable, it could be calculated on the basis of median house prices with repeat sales removed. This does not remove the bias completely since a large number of properties will not sell at all, even in the fifteen or twenty years of Land Registry data analysed. The longer the period considered, the less of a problem this will be.

4.4 Variations in Price Trends across submarkets

Theil (1954) specifies three conditions for aggregation – that is, the conditions under which units might be legitimately combined across space into a single unit of analysis. Parameters have to be first homogenous; second, dependent; and third, convergent within the area of aggregation. These requirements are closely related to the dynamic aspects of the question discussed in earlier in the report regarding the means by which submarkets should be defined and demarcated. Meen (1996) makes use of these conditions to examine the unity of housing markets at the regional level but they might legitimately be applied to the sub-city segmentation of the housing market, though there has been very little UK work done at this level.

One aspect of Theil's conditions that has been explored at the submarket level is that of dependence. Jones et al (2003) construct repeat sales indices for six Glasgow submarkets (as defined by Watkins 2001) and apply cointegration techniques to determine whether these submarkets remain distinct overtime. The criterion for independence is the absence of a cointegrating relationship between the repeat sales indices for the different housing market segments. Two pairs of comparisons fail this test and as a result the six pre-defined submarkets collapse to four when this dynamic definition of dependence is used.

There are, however, a number of severe problems with the method used by Jones et al. First there are problems associated with the initial delineation of submarkets (see the critique in chapter 2 above). Second, the authors employ repeat sales indices as the basis for the analysis and the corollary of the discussion of frequency of sale earlier in this chapter is that there are likely to be several problems in applying the repeat sales approach to computing price indices. First, the proportion of repeat sales is likely to vary between submarkets. Two, this variation is not random but correlated with nature and quality of the properties. Third, repeatedly sold properties will, as a result, have different average price levels than infrequently sold properties. Fourth, the rate of change of prices in the repeat sales group may well be quite different to the price inflation of properties that sell only once or not at all. In effect, Jones et al have considered the behaviour of a submarket of repeatedly sold properties within a submarket defined by cross sectional means on properties not necessarily sold more than once. This "submarket within a submarket" is then compared with other "submarkets within submarkets" and found to be independent in some cases. The interpretation is ambiguous, however, since spatial submarket boundary may have changed over time, as might the structure of repeat sales relative to non-repeat sales, and the relationship of the repeat sales sub-submarket to the spatial submarket may vary across the different housing segments.

The application of cointegration analysis (Meen 1996, Jones et al 2003), however, is legitimate and offers an appealing way of testing whether persist over time. A more robust way of applying this approach in future would be to control for variation in repeat sales (using duration analysis, for example) and then use this adjustment to develop consistent indices for each submarket. The cointegration tests would then give more meaningful results. Movements in the submarket boundaries should also be investigated since tests for Theil's dependence requirement become meaningless if there are shifts in the homogeneity conditions. A practical solution in future analysis would be to run the grid search procedure suggested in **chapter two** on selected years of the data to verify that submarket boundaries have not shifted. Where they have, it may be possible to identify irreducible cores of each submarket that remain distinct over time and it would be for these submarket cores that adjusted price indices could be designed and compared.

4.5 Submarket Price Volatility

A major deterrent to investment in new construction is the level of risk associated with development and a key component of this risk arises from the volatility of prices. The Barker Review Interim Report (2003, p.64) estimates that, "For a 1 per cent increase in house prices, gross development profit on some sites can increase by almost 8 per cent". In practice, the perceived risk is greatly exacerbated by the uncertainty surrounding the true volatility of prices at the local level. There is almost complete absence of robust, appropriately adjusted price indices at the submarket level for most cities in the UK. Compared with the great volume of information available on the performance of other assets, it is unsurprising that both the level of investment in new construction and the magnitude of supply elasticities are so low. There is growing recognition on both sides of the Atlantic of the importance of local volatility in property prices and rents. Current research by Deng et al (2004, p.1) argues that, "The space market tends to be local in nature as supply and demand for space can vary considerably across locations. The capital market tends to be more national and little variation in discount rates from location to location". They attempt to compare growth and volatility across commercial property submarkets defined by economic characteristics of metropolitan statistical areas (MSAs). By relating variations in individual property income growth to economic events in specific economic sectors, their goal is to develop "a risk measure for investing in commercial property that can identify the extent to which a particular MSA is vulnerable to market downturns (such as the loss of one major industry)".

Given the paucity of volatility research in the UK at the sub-city level, there is no shortage of avenues of future research. The first is the need to develop submarket based price indices that correct for sample selection bias. A natural corollary of this will be an analysis of the variations in volatility between submarkets. The next step is to consider the determinants of volatility and most importantly in our current context, to ascertain the relationship between submarket price volatility and new construction at the sub-city level. There is much to do, but there is also much to be gained from research in this area. Land Registry data potentially provide the means to construct indices that are adjusted for repeat sales sample selection, though the absence of property characteristics means that mix adjustment of the attribute kind cannot be made. In future, the availability of basic attribute information in the now near 100% survey of mortgage lending transactions opens the way for quality mix adjusted indices to be developed at submarket level. These indices will still need to be adjusted for sample selection bias (particularly if they are to be constructed at the submarket level, and it is crucial to investment decisions that they are) and there are good prospects for repeat sales adjustments from the Land Registry data to be applied in future to the Council for Mortgage Lenders data.

4.5.1 The Role of Credit in Determining Volatility

A potentially important factor in shaping the dynamics of local housing markets is the nature and availability of credit, and in particular, spatial variations in the structure of mortgage finance. Since the great majority of house purchases require a mortgage of some type (variations in the proportion of non-mortgage transactions is itself of interest), differences in LTVs (loan to value ratios) and other mortgage characteristics can potentially cause significant asymmetries in the impact of interest rate changes and can themselves influence local housing market dynamics. Stein (1995), for example, develops a model of trade in the housing market where purchasers require a down-payment to purchase a new home. Stein shows how variations in loan to value ratios can influence the volatility of house prices and time-to-sale of properties. Lamont and Stein (1999) investigate the Stein (1995) hypotheses using US city-level data on the relationship between homeowner borrowing patterns and house-price appreciation rates. They find that in cities where a greater fraction of homeowners have high loan-to-value ratios do indeed appear to have house prices that react more sensitively to changes in income. No research has been done in the UK to date on this issue.

A related topic also of interest is that of the impact of differences in the average loan to value ratios at the local level on the time-to-sale of owner occupied housing. Work by Stein (1995), for example, suggests that households with high loan to value ratios need to make a larger capital gain and they tend to put their dwelling on the market at a higher asking price and as a result are likely to face longer times-to-sale. Since loan to value ratios are related to income, a closely allied question is that of whether lower income groups face greater capital loss and time-to-sale risks.

Added to this complex mix is the effect of future increases local supply on time to sale, and the corollary for local possession rates. For example, if an expansion in housing supply increases the time to sale in an area, this might eliminate one of the possible escape routes for mortgage borrowers facing repayment difficulties. As such, homeowners that would otherwise have been confident of rapidly selling their home (with a view to downsizing or switching to rental accommodation) in the event of mortgage repayment difficulties, in a supply-rich world may actually find themselves unable to sell rapidly and hence facing possession (and possible homelessness) or long term reliance on Income Support for Mortgage Interest. That such households may already be more likely to have high loan to value ratios and (according to Stein) higher reservation prices as a result, the local effect of new supply may be to compound the repossession risks facing vulnerable households. Thus submarkets may not only be defined in terms of product mix or attribute valuation shifts, but also in terms of credit market behaviour, which in turn impinges upon the shape of housing demand. Different credit rationing patterns across space will influence the spatial contours of demand elasticities, which in turn may determine the impact of new supply.

There are also ways in which the structure and nature of submarkets might be expected to shape lending policy. For example, if there are significant differences in price volatility and liquidity (see below) then the corresponding variation in risk across submarkets should, in an efficient market for credit, be reflected in variations in the risk premiums embodied in mortgage rates. Controlling for spatial variations in loan to value ratios, one would expect there to be variation in mortgage rates across submarkets according to risk. On the other hand, if there are adverse selection consequences of risk pricing, as anticipated by Pryce (2003), then such a pattern may not emerge. Risk pricing by submarket, if it exists, would of course have sociological implications if those submarkets facing the highest risk and highest mortgage rates were predominantly occupied by low income families (which might well be the case if the higher price volatility is indeed associated with higher loan to value ratios – one would expect low income groups to have higher debt gearing[2] – and if liquidity is lower in deprived areas[3]).

4.6 Liquidity Bias

Earlier in this chapter the issue of sample selection bias was considered and it was shown how variations in the frequency of trade can cause bias house price indices. However, not only do some properties trade more frequently than others, but the liquidity of housing assets can also vary considerably over time. Fisher et al (2003) note that,

> "During 'up' markets, capital flows into the sector, there is much greater volume of trading, and it is much easier to sell assets. Just the opposite typically occurs in "down" markets. This intertemporal variation in the ease of selling an asset affects the interpretation of transaction prices. An important implication is that transaction-based price indices do not hold constant the liquidity in the market." (p. 270).

Fisher et al (2003) argue that investors are interested in the expected time taken to sell property as well as the price and growth in price. As such, it could be argued that liquidity and knowledge of liquidity will affect investment rates in new construction. Clearly the motivation for developers to actually make use of land released for new residential construction is a prerequisite to the success of any such policy. Crucial to the investment decision is adequate information about expected liquidity and such information may also be a driver of the growth of the private rented sector (other things being equal, landlords will be more likely to expand their portfolio of properties if there high rates of liquidity in the housing market). The argument put forward below is that like many other housing market factors, liquidity will mary not only over time and across regions but across housing submarkets within cities and so there is much to be gained from improving our understanding of the submarket behaviour of housing liquidity.

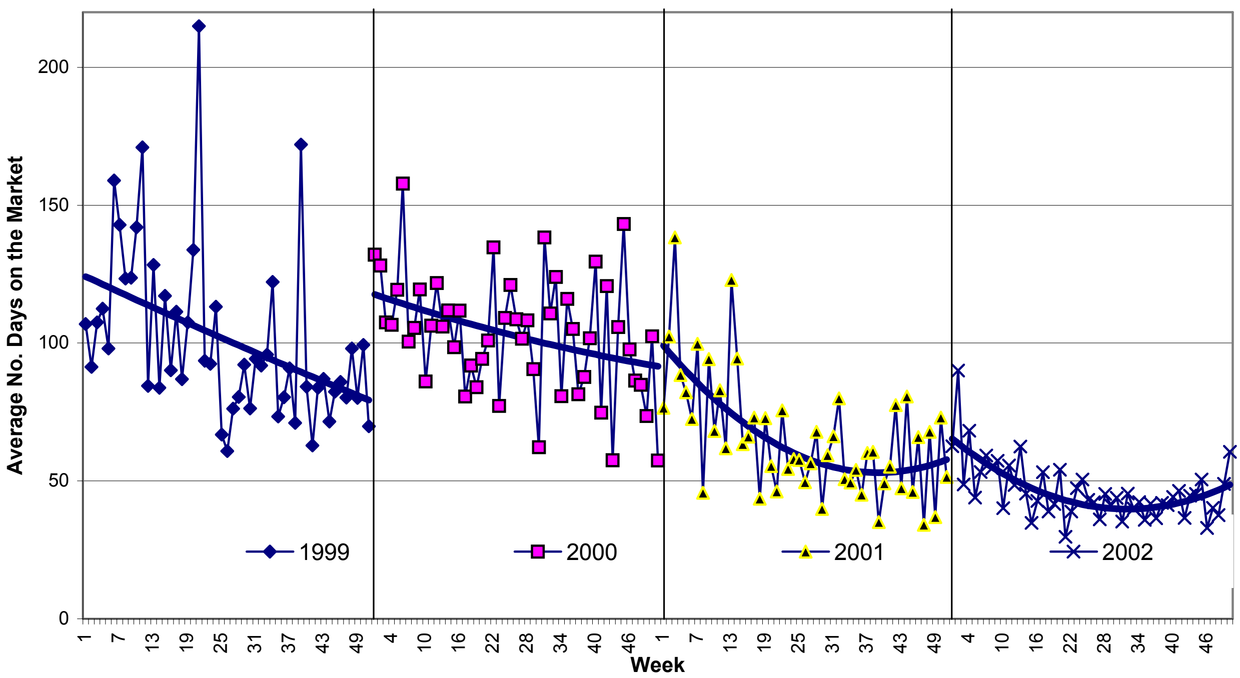
4.6.1 Example from Glasgow data: Variations in liquidity over time and submarket

To illustrate the extent to which the probability of a property selling varies even within a city, analysis from Gibb and Pryce (2004a) based on Glasgow transactions data is summarised below. Consider first how liquidity varies over the housing market cycle. In Figure 1 it can be seen that there is a seasonal effect on liquidity, with properties taking less time to sell in quarter 3. This

[2] Hendershott, Pryce and White (2003) find LTVs in the UK to be negatively correlated with income.
[3] See Pryce and Gibb (2004a,b).

seasonal effect is dominated, however, by cyclical movements in the housing market, to the extent that as the market accelerates in 1999 and 2000, quarter 4 properties actually sell more quickly than quarter 3 properties. In 2002, as the boom approaches its zenith, the seasonal "U" shape in time on the market becomes more apparent, and the overall level of time-to-sale is at its lowest.

Figure 4-1 Average Time on the Market in Glasgow Over Time



4.6.2 Submarket variations in liquidity

An under-researched element in the time-to-sale literature is that of variations between submarkets. One would expect that large, buoyant submarkets with efficient dissemination on properties for sale and good transport facilities available to house-searchers would enjoy greater levels of liquidity (shorter time on the market) than niche properties in less desirable markets with an inefficient estate agency sector and poor transport facilites/remote location. Also of interest here is the question of whether the probability of a property selling increases or decreases the longer it has been on the market. That is, whether or not there is "duration dependence" in time-to-sale. It may be, for example, that after a property has been on the market for two months or so the fact that it has not sold may be a signal to prospective purchasers that the property is of poor quality (in a way not reflected in the estate agent's published description and so reduce the probability that the prospective bidder would actually view the property and/or reduce the probability of bidding. Crucially, this signaling effect may vary by area depending on the information and search structures of each submarket. As a result "duration dependence" may also vary between submarkets.

The customary way of considering duration dependence is to consider the hazard rate (function) h(t). This measures the probability of a dwelling selling in a very small time interval assuming that the property has remained on the market until the beginning of the interval. As such it is the limit of the probability that a property will in the interval t to $t+\Delta t$ per unit of unit width Δt given that the property has remained on the market until time t:

$$h(t) = \lim_{\Delta t \to 0}\left(\frac{P(\text{property sells in the interval } t \text{ to } t+\Delta t \text{ given current TOM})}{\Delta t}\right),$$
$$= \frac{f(t)}{1-F(t)}$$

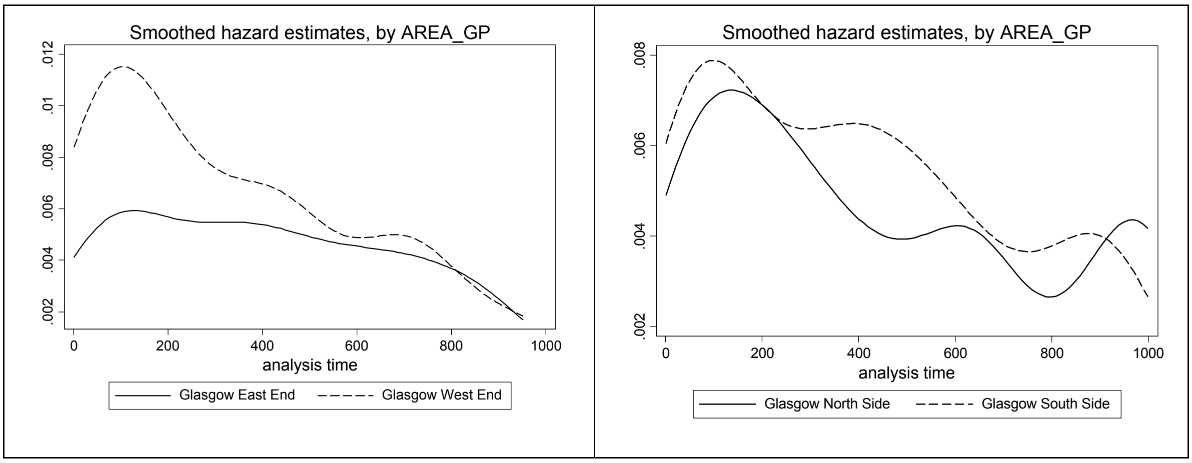
If there are no censored observations, the hazard function can be estimated as:

$$\hat{h}(t) = \frac{\text{number of properties selling in the interval starting } t}{\text{(number of properties still on the market)(interval width)}}$$

The hazard function measures the rate at which risk is accumulated. It can vary from zero (meaning no risk of selling at all) to infinity (meaning the certainty of sale at that moment). "Over time, the hazard rate can increase, decrease, remain constant, or even take on more serpentine shapes. There is a one-to-one relationship between the probability of survival past a certain time and the amount of risk that has been accumulated up to that time, and the hazard rate measures the rate at which risk is accumulated" (Cleves et al p.8). If there is no signaling effect, one would expect that the longer a property stays on the market, the more risk of selling is accumulated and so the cumulative hazard will be always increasing over time on the market. If there is a signalling effect of poor quality, there will be dis-accumulation of risk after a certain time on the market, and the cumulative hazard will dip down at some point over the course of time on the market.

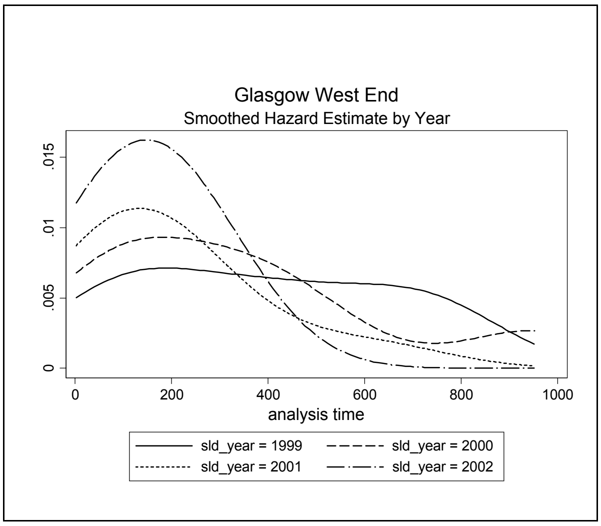
Smoothed hazard function estimates presented below for four different submarkets in Glasgow: East End, West End, North Side, South Side (Pryce and Gibb use the GSPC estate agency definition of submarkets – see Palm 1978 and Michaels and Smith 1990 for precedents to this approach). Large differences in the shape of the hazard curves for the different submarkets are evident, with much higher overall rates of hazard for the West End compared to the other segments, and a much shallower hazard curve for the East End (indicating lower rates of duration dependency in the East End than elsewhere).

Figure 4-2 Comparison of Hazard Functions for Different Submarkets



There are also substantial shifts in the shape and peak of the curve emerge as the market booms. This can be seen in Figure 4-1 where for the West End, the hazard function in 1999 rises slowly until around 200 days and then remains relatively flat until around 700 days. Within twelve months, however, the hazard function has become considerably more peaked and by 2002 a very steep hazard function emerges where the hazard rate for properties not sold after 200 days declines as rapidly as its initial rise. A similar, though less pronounced, picture emerges for the East End, the North Side, South Side and most other areas.

Figure 4-3 Changes in the Hazard Function Over Time in the West End of Glasgow



4.7 Conclusion

The notion that building more houses anywhere will have the same impact on overall real price trends overlooks the heterogenous dynamics of the housing market. House price indices are transactions based. That is, they are dominated by those areas in which properties are frequently traded. In high-turnover areas, a property may well enter the index more than once in a year, whereas a property in a low demand area may never enter the index at all, since the owner is unable (or unwilling) sell the property. He/she may not, for example, be able to attract an offer sufficient to cover outstanding mortgage debt (hence spatial patterns in loan to value and loan to income ratios are potentially important).

What's the connection to housing supply? Well, releasing land for new construction in low demand areas will mean that even if those houses are built and sold, if their resale is so infrequent because of low demand, they will rarely enter the calculation of house price indices, and hence have negiligible effect on house price trends. Perhaps there is a trade-off to be achieved between using new supply to regenerate deprived areas (often associated with low turnover) and alleviating demand pressures in high turnover areas. Here again understanding the nature, influence and location of submarkets becomes paramount. One of the things we need to understand more fully is the relationship between structural breaks in hedonic price equations and the different dynamic behaviour of the submarkets so defined. Again, data is now emerging in some cities (Glasgow and Aberdeen) that allow us to consider market dynamics (frequency of repeat sales, time to sale, inflows and outflows of properties in a given period) in conjunction with submarket delineation.

5 Income and Price Elasticities

Plan:

5.1 Introduction

In chapter 2 the notion of submarkets was defined and discussed in terms of population flows, product groups and homogenous attribute prices. In this chapter, another aspect of submarkets is examined: that of demand elasticities. That is, the responsiveness of quantity demanded to changes in price ("the price elasticity of demand") and income ("the income elasticity of demand"). If demand is highly sensitive to price (income) then it is described as highly price (income) elastic. Conventional economic theory stipulates that price elasticities are determined by the number of close substitutes available. If there are many close substitutes then demand will be highly responsive to price and demand for that good is described as being "price elastic". An increase in price will simply cause consumers to switch to a substitute good. Income elasticities in contrast tend to be determined by the extent to which a good is considered a necessity or luxury. If it is considered a necessity, as income rises there will be only a small increase in the demand for that good and it is described as being "income inelastic". If, on the other hand, it is considered a luxury, then as income rises there will be a large increase in the demand for that good and it is described as being "income elastic".

One of the ways of thinking of submarkets is in terms of a network of substitutes (refer to Grigsby's definition quoted in chapter 2). If a submarket is characterised by a *tight lattice* of dwellings – that is, where there is a high degree of substitutability of dwellings within the submarket – and where dwellings frequently come onto the market and do not sell particularly rapidly when they do, then one would expect a house buyer seeking to locate in that area to be typically faced with a fair range of close substitutes. So the market demand schedule in such an area would typically be elastic (represented by a shallow demand curve). We demonstrate below that land release and an outward shift of new construction in this type of submarket would result in a rather small deflation of prices.

Conversely, in areas where the submarket is characterised by a *loose lattice* of substitutes that infrequently come onto the market and are typically sold rapidly when they do, one would expect demand to be highly inelastic and outward shifts of supply of similar dwellings will have a large impact on price.

Also of interest here is how demand for different types and location of property will shift over time. A crucial determinant is the future trajectory of average income. If real incomes continue to rise (as they have throughout the post war period in the UK) then dwellings with high income elasticities (those considered "luxury" dwellings) will experience a greater outward shift in demand than those with low income elasticities (those considered "necessity" dwellings). As such, unless supply is perfectly elastic (in reality it is nearer perfectly *in*elastic), house prices trends will polarise, with larger, higher quality, attractively located dwellings enjoying greater capital gains than those at the lower end of the desirability scale. The increasing inequality of incomes is likely to exacerbate this trend (see recent work by Geoff Meen).

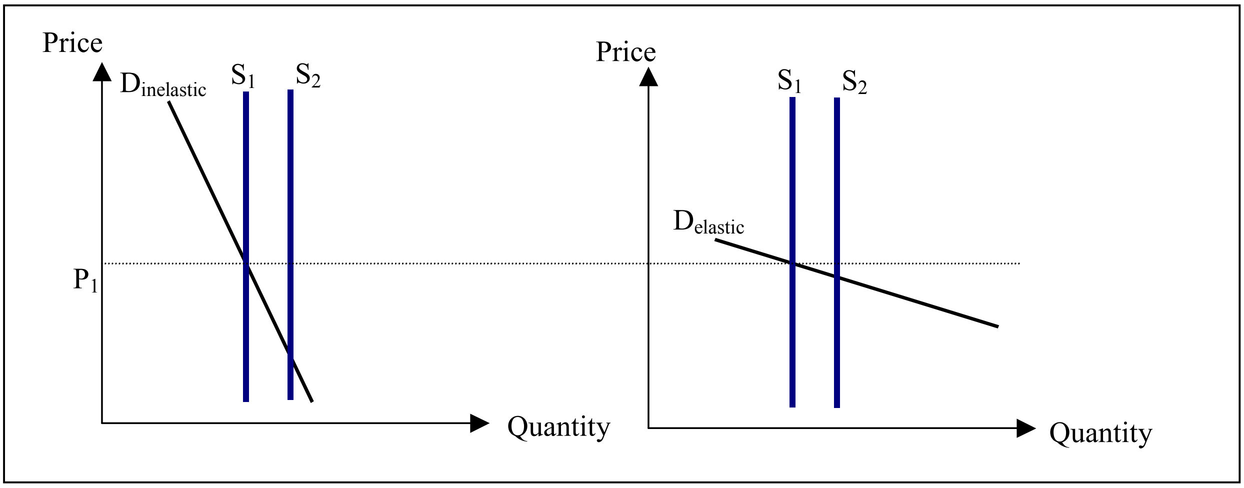
This chapter argues that an understanding of the demand elasticity structure of submarkets is crucial to being able to anticipate the price effects of land release and outward shifts of new supply. The remainder of this chapter elaborates on these theoretical themes and then presents a strategy for estimating demand elasticities at the submarket level.

5.2 Price Elasticity of Demand

5.2.1 What Determines the Price Impact of New Supply?

What will be the price impact of a significant outward shift of new supply? The determining feature, ironically, is not the price elasticity of supply (how responsive is the quantity of new construction to changes in price) but the price elasticity of *demand*. As the following demand and supply diagrams demonstrate (Figure 5-1) and outward shift of supply when demand is inelastic will cause a large downward change in price (left panel). In contrast, when the demand curve is elastic, an outward shift in supply causes a relatively modest change in price (right panel). Typically, macro models of the housing market compute the required increase in the quantity of dwellings to bring about a desired price effect based on an assumed or estimated value of the price elasticity of demand. However, such models tend to assume that the price elasticity of demand is constant over time and across space, whereas Pryce (2001) and others have offered reasons for why this might not be the case. As a result, the type, timing and location of new construction may qualify or amplify its effect on overall house prices. In reality, the housing market may be made up of a rich tapestry of different demand elasticities, with substantial variations across submarkets. Without careful thought and a sound understanding of the demand characteristics of the urban housing market(s), land release and new construction (though as we have seen, the latter does not necessarily follow the former) may not have the anticipated effect on house price inflation. Moreover, it is imperative that we remember that there is not one level of house price inflation – there are significant differences both across regions, between submarkets and between houses with different rates of turnover. This raises the question of which record of house price inflation should be our target, and whether the stabilisation of regional and urban house price differentials should also be a policy goal. If releasing new land in particular areas does more to exacerbate the inequality of house price trends across regions than to stabilise the rate of change in the overall average price of housing, then one has to question whether that policy has been successful. And to understand the regional and submarket variation in the effect of new supply, one has to understand the regional and submarket variation in demand elasticities. At present, we know very little about how such elasticities vary. Indeed, in the last decade, there have only been a handful of studies of housing demand elasticities of any kind in the UK (Rosenthal 1996, Meen 1996, Pain and Westaway 1996, Ermisch et al 1996, Muellbauer and Murphy 1997, Hendershott, Pryce and White 2003).

Figure 5-1 Impact of An Outward Shift of Supply Depends on the Price Elasticity of Demand



5.2.2 Factors that affect the number of effective substitutes & the PED

What causes the price elasticity of demand to vary, and is there a rationale for believing that these causes might have a spatial dimension such that the elasticity varies significantly across submarkets and regions? There are two principle categories of factors that effect the price elasticity of demand. First, the availability and closeness of substitutes, and second, factors that affect the flexibility of consumers to adjust their demand for housing.

Under this first category, there are three main factors that affect the availability of substitutes. These all affect the number of "effective substitutes" facing a household seeking a property:

> "purchasers … do not have perfect information regarding all dwellings for sale, and so buying a house entails a search process. The buyer begins this search process by examining dwelling A, which has a given set of characteristics and price. In order to decide whether or not this is a good purchase, the buyer seeks to examine further dwellings B, C, …, Z which lie in the same price range. The price elasticity of demand is assumed to be determined by the price and availability of known substitutes. Thus, the more dwellings the buyer can survey, and choose from, the more 'effective substitutes' there are available to him, and the more sensitive he is to price. Therefore, if there exists a constraining factor which limits the number of dwellings he can survey, or which increases the cost of surveying further dwellings, then the effect of this factor will be to dampen the price elasticity of demand." (Pryce 2001).

Clearly, if there are large number of properties of a similar type available for purchase relative to the number of buyers, then buyers of that type of dwelling will enjoy a large number of close substitutes. So the first factor is the heterogeneity of properties. Note though that substitutability depends as much on location as physical structural attributes. A two bedroom flat in Mayfair is viewed very differently by the market to an identical property in the East End of Glasgow. Neighbourhood and amenity characteristics are crucial and these (along with certain structural factors) can provide barriers to the creation of new substitutes. Historical properties made with traditional materials and techniques and which have enjoyed the enriching ageing process of time, may not have depreciated in value at all but like good wine may be perceived to have improved with age. The ambience and aesthetic of such properties and localities render them inimitable. Developers, however talented, cannot reproduce the look, lifestyle or authenticity of Old Town Edinburgh. The set of potential substitutes in the future as well as the present rests entirely on the existing stock and, as such, the price elasticity of demand for such dwellings may be highly inelastic. Other examples would be properties located within the catchment areas of excellent schools, or adjacent to particularly attractive parkland.

A second factor under this category is the frequency of sale of properties, particularly when inimitability constrains the set of potential substitutes to the existing stock. That there may be a vast supply of such properties in the total housing stock is of little consequence if those properties never come onto the market.

Third, even if there is a high turnover of stock, if properties of a particular ilk or location sell within a day, the physical limits to a housebuyer's search capacity may mean that at any one point in time, he/she only has a limited number of properties to choose from. The probability that another house-seeker will purchase the property in the interim means that each time a potential buyer chooses to view an additional property, there is a chance that a previously viewed property is no longer available. So the number of "effective substitutes" is conditioned by time on the market as well as frequency of sale. There is evidence that both these factors vary considerably not only over time but also across space. Submarkets again are the decisive driver of the demand elasticities.

The second category of factors that affect the price elasticity of demand were set out above as being those that determine the flexibility of consumers to adjust their demand for housing. Even if there are a large number of substitutes available, consumers can only take advantage of bargain prices if they can easily adjust their consumption of housing services. Unfortunately, housing is perhaps the least of goods in its amenability to realising demand changes. To increase ones consumption of housing one typically has to either refurbish/extend one's property, or move house. Both these options entail considerable upheaval (house moving is often listed as one of the most stressful life events, and not least, also one of the most the most costly). Households may wait for years for a property of a particular type/in a particular locale to come on the market, but when it does, to have a chance of being a potential bidder the household has to be able to rapidly sell its own property, have sufficient ready cash available to pay conveyancing and survey fees, and have the capacity to swiftly raise appropriate mortgage finance. So asset liquidity, transactions costs and credit constraints are all potentially inhibiting factors to the demand-adjustment process. As such, one would expect demand to be more price elastic the less prevalent are these barriers. Since these barriers to adjustment vary across household type (previous owners, for example, are likely to be less credit constrained than first time buyers – Hendershott, Pryce and White 2003), one would expect demand elasticities to vary across households and possibly across areas even if property types were identical (there is evidence in the

US, for example, that mortgage pricing and rationing vary by submarket), though there has been no empirical investigation of the existence or extent of this phenomenon in the UK.

5.2.3 To Which submarket will the new supply belong?

If the impact of supply depends on price elasticities, and if price elasticities potentially vary across submarkets, the crucial question for any planner interested in the price effects of new construction, is to which submarket will the new supply belong? Bearing in mind the inertia of location (once a new property has been built on a particular spot, it cannot easily be moved to another!), this decision is effectively irreversible. New construction will have a permanent influence in shaping the nature of the housing stock and will alter the number of effective substitutes of a particular dwelling type for some time to come. Thus if the increase of new supply is large enough it may actually shape the price elasticity of demand for its parent submarket.

To the extent to which planners and government policy control the nature and location of new supply, the state (in the UK at least) has a potentially profound influence on the long term demand elasticity structure of the housing market. If the nature and location of new supply is increasingly of a kind associated with low price elasticities (this may be the case, for example, if there were a predominance of high density, small living space properties among new properties), and if the permanent changes to the housing stock that result, further lower the responsiveness of demand to price for these kinds of properties, then the incremental effect of each new property constructed will diminish over time. More and more new dwellings will have to be constructed to achieve the same effect on house prices. Predictions based on past parameters may grossly underestimate the number of new dwellings needed to stabilise future house price inflation.

Note that there may be good reasons for controlling, and natural limits to, the type and location of stock. In particular, the availability of prime land may inexorably diminish over time. This is analogous to the argument for diminishing returns to agriculture first put forward by Malthus, one of the patriarchs of economics. Because the most fertile land always gets used first, cultivation of successive portions of the natural stock of land will inevitably suffer diminishing returns. As the population grows, a disproportionately large increase in land will be needed to feed each addition to that population. (It was Malthus' tenebrous predictions that gave rise to economics being labelled the "Dismal Science").

5.3 Where will people want to live in future?

Consider two types of housing, where type is differentiated by location. That is we have a type of housing located at L2 which is considered to be highly desirable and prestigious, and another type of housing located at L1 which is considered less desirable and prestigious. Label the former "luxury housing" and the latter "necessity housing". We could either hold attributes constant and assume that it is purely the location effects that determine whether or not the dwelling is considered as *luxury* housing (the aforementioned notion that a two bedroom apartment in Mayfair, London is viewed differently to an equivalent property in Easterhouse, Glasgow). Or we could allow dwelling attributes to vary, and assume that dwellings of a particular type tend to cluster. Either way, the key implication is the same: the two types (however defined) of housing will incur different income elasticities of demand (i.e. different rates of responsiveness of quantity demanded to changes in income).

According to textbook micro economic theory, luxury housing will have a higher income elasticity of demand than the necessity housing. This is because as incomes rise by a given proportion, it seems improbable households will want to consume an equivalent proportional increase in *necessity housing*. If the locale of the necessity housing is of such poor quality, the income elasticity of demand may even be negative, denoting an "inferior" good. To illustrate the implication of this for housing supply, assume for a moment that both types of housing have similar price elasticities and that the price elasticity of supply is zero. In panels (a) and (b) we see the asymmetric effect of a uniform positive income shock for the two housing submarkets. Because the IED is smaller for necessity housing, a given proportionate increase in income will cause the outward shift of demand to be less pronounced. This in turn will result in a much smaller price increase for L1 type housing than L2.

Suppose now that new land release policy causes supply to shift out by equal amounts for the two locations. For L1, the overall effect is a noticable price fall, and for L2 the price rise is ameliorated, but the overall price effect is still positive. If a Pryce and Gibb (2003) type neighbourhood effect emerges for L1 then there may perhaps be a further outward shift of demand, but if not, one might conclude that a better policy option is to release no new land for type L1 houses, and instead allow the land release to occur only for type L2 houses (supply shifts to S'2). Note though, that without the asymmetric supply intervention, continued income rises will cause greater polarisation of houe price trends between L1 and L2. Note further that if it is the luxury dwellings that are more frequently traded, then turnover effect noted above will exacerbate the divergence of price trends as revealed in transaction based indices. The key corollary to grasp, though, is that if an insufficient proportion of the new land is released for luxury type housing, then the overall mitigating impact on rising real price trends will be negligible.

Different Submarket-IEDs Result in Different Price Outcomes

5.3.1 What Causes the Income Elasticity of Demand to vary?

There are three key factors that cause the income elasticity to vary for housing. The first is dwelling size and attributes. This can be as much a qualitative factor as quantitative since the ambience of a property can be a key determinant of its value. Other things being equal, the larger and more attractive the property, the more of a luxury good it will be perceived to be, and the greater the sensitivity of demand to changes in income. Rather predictably, the second driver of income elasticity is location. Once the basic facilities of access to transport and employment have been achieved, households will seek to spend a greater proportion of increases in income on “luxury” location aspects, such as views, access to leisure facilities and status.

The third element, not discussed at length in the report so far, is the possible variation of preferences across households. There is evidence in the US to suggest, for example, households seek racial and social compatibility. Different preferences for rural/urban lifestyles, and for different leisure activities, will also give rise to different perspectives on what is considered ‘luxurious’ and what is considered ‘necessity’ housing. There are also lifecycle factors. Location next to excellent schools may be seen as the height of luxury for families with young children. Not so for a retired couple living alone. Demographic factors will therefore have an important influence on the shape of the price elasticity matrix across households. An ageing population may give rise to a lower average preference for access to the central business districts and greater preference for semi-rural locations.

Growing income inequality will exacerbate the effects of heterogeneous income elasticities for property type and location because it will imply greater variation in the size of the outward shifts of demand depicted in the last set of diagrams. One would expect, under such circumstances, the increased polarisation of prices between the most luxurious housing and the least desirable dwellings. The former will enjoy ever greater increases in price inflation relative to the latter. As such, the “class reproduction” consequences of home ownership are set to intensify. Those on highest incomes will be able to afford the most luxurious housing which will enjoy the greatest levels of capital gains. Higher income, asset-rich households will be able to accumulate capital at an ever increasingly greater rate than low income, asset-poor households. Ironically, the consequence of restrictive planning may be to exacerbate rather than alleviate the most socially divisive aspects of capitalism and as such add to the long list of state failures associated with the allocation of land (see Pennington 2000 and Pryce 2003 “Greening by the Market?”).

Ironically, the only way to counteract this growing disparity in house price inflation is to increase the number of “luxurious” properties being built, and take the pressure out of the top end of the housing market. As noted in chapter 2, whilst there may be a filtering effect whereby poorer households eventually gain access better quality properties previously occupied by the wealthy, the process maybe unpredictable has the inevitably dubious ethical consequence of helping the poor by subsidising the rich. The dilemma is less direct, however, if the policy entailed not a subsidy to luxurious housing, but simply a relaxing of restrictions on the development of housing of any kind, the emancipation of the market to decide what properties should be built. If price gains are greatest at the highest end of the quality scale, then supply will adjust accordingly.

5.4 A Proposed Method for Estimating Demand Elasticities

Appropriately defined submarket boundaries (see chapter 3) will allow the researcher to derive appropriate constant quality price indices for each location. This is important since it determines the denominator of the dependent variable in a Goodman and Kawai type demand function (see below). Income and price elasticity of demand estimates could be computed for separate submarket areas. It is beyond the scope of this report to actually derive submarket boundaries and test for variations in IED (incomd elasticity of demand) and PED (price elasticity of demand) across space. However, some preliminary evidence that these parameters vary across house type is presented below. The estimates are derived from the CML 5% sample using a demand estimation method that could feasibly be applied to submarkets (here it is applied only at the regional level and the variation explored is not spatial but across dwelling types and price brackets).

5.4.1 Demand estimation method

The following is an extension of Hendershott, Pryce and White (2003) and recently unpublished work by Pat Hendershott and Gwilym Pryce. The first step is to estimate ψ_j, the probability that an unrationed borrower's loan will exceed the tax deductibility ceiling of £30,000, where $j \in J$, $J \subset I$, and J is the set of *a priori* unrationed borrowers (defined as those mortgagors whose loan-to-value and loan-to-income ratios are below the thresholds typically applied by lenders to ration credit), itself a subset of I, the set of all mortgage borrowers in the sample. This probability will later be multiplied by the borrowers marginal income tax to calculate the tax deductible component of interest in the user cost of capital (needed for the housing demand regression). ψ_j is derived from the predicted values of a logit regression of the form:

$$\psi_j = f\,(OO_j, Y^B{}_j, Y^O{}_j, AGE_j, t_i) \qquad [1]$$

where OO_j is a dichotomous variable indicating whether the borrower is a previous owner, $Y^B{}_j$ is the basic income of the borrower, $Y^O{}_j$ is other income, AGE_j is the age of the main borrower, and t is the year in which the mortgage of unconstrained borrower j is transacted. Ermisch, Findlay and Gibb (1996) and Hendershott, Pryce and White (2003) estimate a demand regression of the form,

$$\ln HC_j = \mathrm{f}(\ln MCH_j, \ln Y_j, AGE_j), \qquad [2]$$

where lnHC is the natural log of housing consumption of borrower *i*, and lnMCH is the natural log of the marginal cost of housing, and lnY is the natural log of real total income. *HC* is calculated as follows:

$$HC_i = P_i / P_{Rt}^* \qquad [3]$$

where $i \in \mathrm{I}$, P_i is the price the borrower has paid for the house, and P_{Rt}^* is the constant quality price in submarket *R* for the corresponding time period derived from hedonic regressions of price on housing attributes and quarter dummies run separately for each UK region and each year (approximately 25,000 observations in each year for the UK as a whole). MCH is computed as,

$$MCH_i = UCC_i\,(P_{Rt}^* / \mathrm{RPI}_t) \qquad [4]$$

where RPI is the monthly retail price index and *UCC* is defined as:

$$UCC_i = (1 - \psi_j\,\tau_\mathrm{i})r_t + 0.03 + \gamma\pi_{Rt}^* \qquad [5]$$

where *r* is the interest rate, τ_i is the marginal tax rate above the ceiling and [illegible] is the expected rate of nominal house price inflation.

5.5 Demand estimation results

The demand regression results are listed there for different UK regions. The first regression for each region is run on the lowest quartile of house prices in that region (credit constrained and unconstrained borrowers combined). The second regression in each region is a MLE Heckman sample selection version of the same run only on unconstrained borrowers (this controls for the bias associated from selecting a non-random sub-sample). The third regression is an OLS (ordinary least squares) regression on the upper quartile of house prices in that region, followed by a Heckman version of the same. The fifth and sixth regressions are OLS regressions for properties with 4 and 6 rooms respectively (credit constrained and unconstrained combined). The important result to note is how the income elasticity of demand tends to be higher for more expensive and larger properties.

Note also that if data of the kind employed here (CML Survey of Mortgage Lending data) could be augmented with spatial identifiers, a grid search procedure of the kind applied to hedonics in the section above could be used to map out spatial undulations in PED and IED (an alternative procedure to consider is geographically weighted regression – see the apppendix to chapter 3). But note that hedonic submarket boundaries would need to be identified first in any case in order to derive an appropriate denominator to the dependent variable in the Goodman and Kawai type demand regression (see the description of the estimation method above).

Results of the logistic regression are as follows:

```
Logistic regression                               Number of obs   =      32188
                                                  LR chi2(26)     =   11546.58
                                                  Prob > chi2     =     0.0000
Log likelihood = -14134.047                       Pseudo R2       =     0.2900

------------------------------------------------------------------------------
        ceil | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     prev_oo |   .7728325   .0385126    -5.17   0.000     .7009183    .8521252
  incbasic_K |   1.105677   .0051634    21.51   0.000     1.095603    1.115844
  incother_K |   1.091455   .0045118    21.17   0.000     1.082648    1.100334
    incoth_d |   1.338637   .0647791     6.03   0.000     1.217507    1.471818
    age_lt25 |   .7984305   .2035724    -0.88   0.377     .4844065    1.316026
    age25_34 |    2.75209   .3679418     7.57   0.000     2.117683    3.576551
    age35_44 |   4.014166   .4711758    11.84   0.000     3.189208    5.052519
    age45_54 |   2.168358   .2524847     6.65   0.000     1.725904     2.72424
incK_agelt25 |   1.147501   .0212908     7.42   0.000     1.106522    1.189998
incK_age2~34 |   1.059603   .0073829     8.31   0.000     1.045231    1.074172
incK_age3~44 |   1.017982   .0060547     3.00   0.003     1.006184    1.029919
incK_age4~54 |   1.003383   .0057802     0.59   0.558     .9921178    1.014776
prevoo_ag~25 |   .7247608   .1148487    -2.03   0.042     .5312631    .9887346
prevoo_ag~34 |   1.010288   .0753295     0.14   0.891     .8729262    1.169264
       yr_96 |   .8867997   .0376228    -2.83   0.005     .8160429    .9636916
       yr_97 |   .9421676   .0389262    -1.44   0.149     .8688809    1.021636
       yr_98 |   .9467393   .0412972    -1.25   0.210     .8691617    1.031241
    northern |   .7399019   .0653389    -3.41   0.001     .6223098    .8797143
yorks_humber |   .9893612   .0788678    -0.13   0.893     .8462537    1.156669
   east_mids |   .8626373   .0691751    -1.84   0.065     .7371744    1.009453
  south_east |   1.292535   .0918723     3.61   0.000     1.124449    1.485748
  south_west |   1.107341   .0851236     1.33   0.185     .9524623    1.287405
   west_mids |   1.066086   .0847151     0.81   0.421     .9123314    1.245753
  north_west |   .9351762   .0729811    -0.86   0.390     .8025383    1.089736
    scotland |   .9354093    .076619    -0.82   0.415      .796673    1.098306
greater_lo~n |   1.448908   .1190805     4.51   0.000     1.233342     1.70215
------------------------------------------------------------------------------
```

REGION 1: NORTHERN ENGLAND

purprice <= lower_quar

```
Regression with robust standard errors                 Number of obs =    1081
                                                       F(  3,  1077) =   25.45
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0775
                                                       Root MSE      =  .22826

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .1106689   .0207572     5.33   0.000     .0699398     .151398
         PED |  -.3875073   .0820169    -4.72   0.000    -.5484383   -.2265764
         age |  -.0009458   .0007664    -1.23   0.217    -.0024497     .000558
       _cons |   2.219997   .6631228     3.35   0.001      .918838    3.521156
------------------------------------------------------------------------------

Heckman selection model                         Number of obs      =      1082
(regression model with sample selection)        Censored obs       =       882
                                                Uncensored obs     =       200

                                                Wald chi2(3)       =      3.66
Log likelihood = -437.4205                      Prob > chi2        =    0.3012
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hc_l         |
         IED |  -.0403882   .0416692    -0.97   0.332    -.1220583    .0412819
         PED |    -.38991   .2080067    -1.87   0.061    -.7975957    .0177757
         age |   .0012774   .0023648     0.54   0.589    -.0033575    .0059123
       _cons |   2.711402    1.69034     1.60   0.109    -.6016043    6.024408
-------------+----------------------------------------------------------------
```

purprice >= upper_quar

```
Regression with robust standard errors                 Number of obs =    1049
                                                       F(  3,  1045) =   86.03
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2501
                                                       Root MSE      =  .25313

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .1475349   .0269411     5.48   0.000     .0946701    .2003998
         PED |  -.6751833   .0792821    -8.52   0.000    -.8307535   -.5196131
         age |   .0071515   .0008658     8.26   0.000     .0054527    .0088504
       _cons |   4.958227   .6901352     7.18   0.000     3.604019    6.312436
------------------------------------------------------------------------------

Heckman selection model                         Number of obs      =      3147
(regression model with sample selection)        Censored obs       =      2208
                                                Uncensored obs     =       939
                                                Wald chi2(3)       =    288.96
Log likelihood =  -1679.89                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hc_l         |
         IED |    .199655   .0249151     8.01   0.000     .1508222    .2484878
         PED |  -.4119645   .1053645    -3.91   0.000    -.6184752   -.2054539
         age |  -.0115223   .0010459   -11.02   0.000    -.0135721   -.0094724
       _cons |   3.265569   .8595942     3.80   0.000     1.580795    4.950343
-------------+----------------------------------------------------------------
```

no. rooms == 4

```
Regression with robust standard errors                 Number of obs =     991
                                                       F(  3,   987) =   72.68
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2652
                                                       Root MSE      =  .31417

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .3342591   .0285254    11.72   0.000     .2782816    .3902365
         PED |  -.3661719   .1245791    -2.94   0.003    -.6106422   -.1217016
         age |   .0078883   .0009083     8.68   0.000     .0061058    .0096708
       _cons |    1.02662   .9904276     1.04   0.300    -.9169657    2.970206
------------------------------------------------------------------------------
```

no. rooms == 6

```
Regression with robust standard errors                 Number of obs =    1016
                                                       F(  3,  1012) =  211.16
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4080
                                                       Root MSE      =  .31808

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .4153304   .0304354    13.65   0.000     .3556067     .475054
         PED |  -.6920993   .1154954    -5.99   0.000    -.9187372   -.4654613
         age |   .0099309   .0012937     7.68   0.000     .0073922    .0124695
       _cons |   3.269682   .9794403     3.34   0.001     1.347716    5.191649
------------------------------------------------------------------------------
```

REGION 2: YORKSHIRE AND HUMBERSIDE

purprice <= lower_quar

```
Regression with robust standard errors                 Number of obs =    1913
                                                       F(  3,  1909) =   52.08
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0994
                                                       Root MSE      =  .20045

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .1771302   .0147458    12.01   0.000     .1482106    .2060498
         PED |  -.0097658   .0379751    -0.26   0.797    -.0842429    .0647114
         age |  -.0002871   .0005547    -0.52   0.605    -.0013749    .0008007
       _cons |   -1.04566   .3206917    -3.26   0.001    -1.674603   -.4167168
------------------------------------------------------------------------------
```

```
Heckman selection model                         Number of obs      =      1913
(regression model with sample selection)        Censored obs       =      1580
                                                Uncensored obs     =       333

                                                Wald chi2(3)       =     14.51
Log likelihood = -684.8533                      Prob > chi2        =    0.0023
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hc_l         |
         IED |   .0933796   .0264576     3.53   0.000     .0415237    .1452355
         PED |   .0809549   .0981102     0.83   0.409    -.1113376    .2732473
         age |   .0021228   .0014783     1.44   0.151    -.0007746    .0050202
       _cons |  -1.532655   .8350087    -1.84   0.066    -3.169242    .1039323
-------------+----------------------------------------------------------------
```

purprice >= upper_quar

```
Regression with robust standard errors                Number of obs =    1902
                                                       F(  3,  1898) =  143.38
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2751
                                                       Root MSE      =   .2632

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .2994389    .021864    13.70   0.000      .256559    .3423188
         PED |  -.2202781   .0469325    -4.69   0.000    -.3123227   -.1282334
         age |   .0071855   .0007322     9.81   0.000     .0057496    .0086215
       _cons |   .6722256   .4308805     1.56   0.119    -.1728235    1.517275
------------------------------------------------------------------------------

Heckman selection model                         Number of obs      =      5642
(regression model with sample selection)        Censored obs       =      4069
                                                Uncensored obs     =      1573

                                                Wald chi2(3)       =    509.36
Log likelihood = -2771.219                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hc_l         |
         IED |   .2661784   .0165814    16.05   0.000     .2336794    .2986774
         PED |  -.1326862   .0486447    -2.73   0.006     -.228028   -.0373444
         age |  -.0085032   .0007292   -11.66   0.000    -.0099325    -.007074
       _cons |   .5717134   .4097874     1.40   0.163    -.2314552    1.374882
-------------+----------------------------------------------------------------
```

no. rooms == 4

```
Regression with robust standard errors                Number of obs =    1646
                                                       F(  3,  1642) =  147.84
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3234
                                                       Root MSE      =  .25872

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .3748024   .0200251    18.72   0.000     .3355251    .4140798
         PED |  -.0405143   .0578776    -0.70   0.484    -.1540361    .0730074
         age |   .0067889   .0006916     9.82   0.000     .0054323    .0081455
       _cons |  -1.691153   .4765663    -3.55   0.000    -2.625894   -.7564109
```

no. rooms == 6

```
Regression with robust standard errors                Number of obs =    1687
                                                       F(  3,  1683) =  346.12
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4768
                                                       Root MSE      =  .29022

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .5355532   .0200258    26.74   0.000     .4962751    .5748312
         PED |  -.0098824   .0580969    -0.17   0.865    -.1238321    .1040673
         age |   .0110978   .0009227    12.03   0.000     .0092881    .0129075
       _cons |  -2.642507   .4922797    -5.37   0.000    -3.608052   -1.676962
------------------------------------------------------------------------------
```

REGION 3: EAST MIDLANDS

purprice <= lower_quar

```
Regression with robust standard errors                 Number of obs =    1750
                                                       F(  3,  1746) =   34.70
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0782
                                                       Root MSE      =  .20963

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .1342498   .0172425     7.79   0.000     .1004317     .168068
         PED |  -.2221794   .0463573    -4.79   0.000    -.3131011   -.1312576
         age |  -.0003544   .0006483    -0.55   0.585    -.0016258    .0009171
       _cons |   .9068564   .3839642     2.36   0.018     .1537784    1.659934
------------------------------------------------------------------------------

Heckman selection model                         Number of obs      =      1753
(regression model with sample selection)        Censored obs       =      1395
                                                Uncensored obs     =       358

                                                Wald chi2(3)       =     13.14
Log likelihood = -746.2415                      Prob > chi2        =    0.0043
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hc_l         |
         IED |    .068097   .0280878     2.42   0.015     .0130459     .123148
         PED |  -.1624472   .0992009    -1.64   0.102    -.3568774     .031983
         age |   .0024206   .0019282     1.26   0.209    -.0013586    .0061998
       _cons |   .5822032   .8426402     0.69   0.490    -1.069341    2.233748
-------------+----------------------------------------------------------------
```

purprice >= upper_quar

```
Regression with robust standard errors                 Number of obs =    1738
                                                       F(  3,  1734) =  185.61
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3002
                                                       Root MSE      =  .27884

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |    .321739   .0204583    15.73   0.000     .2816135    .3618646
         PED |   -.318302   .0489688    -6.50   0.000    -.4143461   -.2222579
         age |   .0077918   .0007425    10.49   0.000     .0063355    .0092481
       _cons |   1.339455   .4386981     3.05   0.002     .4790225    2.199888
------------------------------------------------------------------------------


Heckman selection model                         Number of obs      =      5189
(regression model with sample selection)        Censored obs       =      3612
                                                Uncensored obs     =      1577

                                                Wald chi2(3)       =    380.41
Log likelihood =  -2720.07                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hc_l         |
         IED |   .2493163    .017749    14.05   0.000     .2145289    .2841036
         PED |  -.1136923   .0488843    -2.33   0.020    -.2095038   -.0178807
         age |  -.0070275   .0007416    -9.48   0.000    -.0084809    -.005574
       _cons |    .489797   .4109059     1.19   0.233    -.3155638    1.295158
-------------+----------------------------------------------------------------
```

no.rooms == 4

```
Regression with robust standard errors                 Number of obs =    1176
                                                       F(  3,  1172) =  104.02
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3312
                                                       Root MSE      =  .25324

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .3336943   .0238289    14.00   0.000     .2869422    .3804464
         PED |  -.2922259   .0656313    -4.45   0.000    -.4209939   -.1634579
         age |   .0064468   .0007685     8.39   0.000     .0049389    .0079547
       _cons |   .5401289   .5504004     0.98   0.327    -.5397511    1.620009
------------------------------------------------------------------------------
```

no. rooms == 6

```
Regression with robust standard errors                 Number of obs =    1630
                                                       F(  3,  1626) =  299.25
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4335
                                                       Root MSE      =  .29828

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .4838743   .0207202    23.35   0.000     .4432331    .5245155
         PED |  -.2993674   .0662529    -4.52   0.000    -.4293175   -.1694174
         age |   .0109507   .0009433    11.61   0.000     .0091005    .0128009
       _cons |  -.1215248   .5465178    -0.22   0.824    -1.193478    .9504282
------------------------------------------------------------------------------
```

REGION 4: EAST ANGLIA

purprice <= lower_quar

```
Regression with robust standard errors                 Number of obs =     994
                                                       F(  3,   990) =   29.56
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1042
                                                       Root MSE      =  .20916

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .0720557   .0277408     2.60   0.010     .0176182    .1264931
         PED |  -.3988189   .0487101    -8.19   0.000    -.4944058    -.303232
         age |  -.0001155   .0007283    -0.16   0.874    -.0015448    .0013138
       _cons |   2.517006    .418701     6.01   0.000     1.695363     3.33865
------------------------------------------------------------------------------

Heckman selection model                         Number of obs      =       995
(regression model with sample selection)        Censored obs       =       778
                                                Uncensored obs     =       217

                                                Wald chi2(3)       =     17.70
Log likelihood = -425.2086                      Prob > chi2        =    0.0005
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hc_l         |
         IED |    .046616   .0340044     1.37   0.170    -.0200313    .1132634
         PED |  -.4122057   .1164547    -3.54   0.000    -.6404527   -.1839586
         age |   .0027427   .0019737     1.39   0.165    -.0011256    .0066111
       _cons |   2.590937   .9781751     2.65   0.008     .6737487    4.508124
-------------+----------------------------------------------------------------
```

purprice >= upper_quar

```
Regression with robust standard errors                 Number of obs =     993
                                                       F(  3,   989) =   79.63
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2314
                                                       Root MSE      =  .27749

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .2332468   .0254333     9.17   0.000     .1833373    .2831562
         PED |  -.2853316   .0501471    -5.69   0.000    -.3837386   -.1869245
         age |    .006802    .000928     7.33   0.000     .0049809    .0086231
       _cons |   1.521032   .4653381     3.27   0.001     .6078687    2.434196
------------------------------------------------------------------------------

Heckman selection model                         Number of obs      =      2982
(regression model with sample selection)        Censored obs       =      2019
                                                Uncensored obs     =       963

                                                Wald chi2(3)       =    259.64
Log likelihood = -1599.751                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hc_l         |
         IED |   .2257438   .0216322    10.44   0.000     .1833456    .2681421
         PED |  -.2316668   .0488746    -4.74   0.000    -.3274593   -.1358743
         age |  -.0066146   .0008926    -7.41   0.000    -.0083641   -.0048651
       _cons |   1.451203   .4256621     3.41   0.001      .616921    2.285486
-------------+----------------------------------------------------------------
```

no. rooms == 4

```
Regression with robust standard errors                 Number of obs =     690
                                                       F(  3,   686) =   46.75
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2920
                                                       Root MSE      =  .26371

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .3245951    .035143     9.24   0.000     .2555943     .393596
         PED |  -.2940531   .0714737    -4.11   0.000    -.4343865   -.1537197
         age |   .0047374   .0010204     4.64   0.000     .0027339    .0067408
       _cons |    .561218   .6288071     0.89   0.372    -.6733995    1.795835
------------------------------------------------------------------------------
```

norooms == 6

```
Regression with robust standard errors                 Number of obs =     867
                                                       F(  3,   863) =  116.14
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3930
                                                       Root MSE      =  .28867

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |    .381045   .0280744    13.57   0.000     .3259428    .4361472
         PED |  -.3281076   .0615912    -5.33   0.000    -.4489936   -.2072216
         age |   .0098341   .0010905     9.02   0.000     .0076937    .0119745
       _cons |   .5761058    .516858     1.11   0.265    -.4383399    1.590552
------------------------------------------------------------------------------
```

REGION 5:GREATER LONDON

purprice <= lower_quar

```
Regression with robust standard errors                 Number of obs =    2856
```

```
                                                      F(  3,  2852) =   58.18
                                                      Prob > F      =  0.0000
                                                      R-squared     =  0.0819
                                                      Root MSE      =  .25946
------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .1546396   .0175586     8.81   0.000     .1202107    .1890685
         PED |  -.2119123   .0317794    -6.67   0.000    -.2742253   -.1495993
         age |  -.0016371   .0005952    -2.75   0.006    -.0028042   -.0004701
       _cons |    .583155   .2947865     1.98   0.048     .0051388    1.161171
------------------------------------------------------------------------------

Heckman selection model                         Number of obs      =      2858
(regression model with sample selection)        Censored obs       =      2259
                                                Uncensored obs     =       599

                                                Wald chi2(3)       =    123.26
Log likelihood =  -1362.28                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hc_l         |
         IED |   .0490244   .0363273     1.35   0.177    -.0221758    .1202245
         PED |   -.223967   .0674854    -3.32   0.001    -.3562358   -.0916981
         age |  -.0123849   .0013015    -9.52   0.000    -.0149358   -.0098341
       _cons |   2.249864   .6246954     3.60   0.000     1.025484    3.474245
-------------+----------------------------------------------------------------
```

purprice >= upper_quar

```
Regression with robust standard errors                Number of obs =    2852
                                                      F(  3,  2848) =  250.10
                                                      Prob > F      =  0.0000
                                                      R-squared     =  0.2726
                                                      Root MSE      =  .32393
------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |     .30482   .0150751    20.22   0.000     .2752607    .3343792
         PED |  -.3090773   .0277596   -11.13   0.000    -.3635083   -.2546464
         age |   .0080516   .0008153     9.88   0.000      .006453    .0096503
       _cons |    1.18287   .2633266     4.49   0.000     .6665398      1.6992
------------------------------------------------------------------------------


Heckman selection model                         Number of obs      =      8641
(regression model with sample selection)        Censored obs       =      6236
                                                Uncensored obs     =      2405

                                                Wald chi2(3)       =    540.15
Log likelihood =  -4847.57                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hc_l         |
         IED |   .2035102   .0169744    11.99   0.000      .170241    .2367793
         PED |  -.2775813     .03459    -8.02   0.000    -.3453764   -.2097862
         age |  -.0094819   .0007622   -12.44   0.000    -.0109757   -.0079881
       _cons |   1.997062   .3258364     6.13   0.000     1.358434    2.635689
-------------+----------------------------------------------------------------
```

no. rooms == 4

```
Regression with robust standard errors                Number of obs =    2809
                                                      F(  3,  2805) =  530.81
                                                      Prob > F      =  0.0000
                                                      R-squared     =  0.4384
```

```
                                                       Root MSE      =  .32732
------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .4848859   .0155264    31.23   0.000     .4544416    .5153301
         PED |   -.370607   .0320779   -11.55   0.000    -.4335058   -.3077083
         age |   .0037729   .0008688     4.34   0.000     .0020693    .0054765
       _cons |   .3679817   .3013224     1.22   0.222    -.2228542    .9588177
------------------------------------------------------------------------------
```

norooms == 6

```
Regression with robust standard errors                 Number of obs =    2074
                                                       F(  3,  2070) =  319.07
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4112
                                                       Root MSE      =  .32558
------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .5081786   .0205318    24.75   0.000     .4679135    .5484437
         PED |  -.2111625   .0377134    -5.60   0.000    -.2851225   -.1372024
         age |   .0060499   .0009884     6.12   0.000     .0041115    .0079882
       _cons |  -.9874116   .3672783    -2.69   0.007    -1.707685   -.2671382
------------------------------------------------------------------------------
```

REGION 6: SOUTH EAST

purprice <= lower_quar

```
Regression with robust standard errors                 Number of obs =    5644
                                                       F(  3,  5640) =  145.03
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1031
                                                       Root MSE      =  .24013
------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .1938501   .0106333    18.23   0.000     .1730047    .2146955
         PED |  -.1196656   .0238083    -5.03   0.000    -.1663391   -.0729921
         age |    .000226   .0003197     0.71   0.480    -.0004008    .0008528
       _cons |  -.4113535   .2135822    -1.93   0.054    -.8300567    .0073498
------------------------------------------------------------------------------

Heckman selection model                         Number of obs      =      5657
(regression model with sample selection)        Censored obs       =      4411
                                                Uncensored obs     =      1246

                                                Wald chi2(3)       =    388.94
Log likelihood = -2482.099                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hc_l         |
         IED |   .0930294   .0196816     4.73   0.000     .0544542    .1316047
         PED |   -.120822   .0422194    -2.86   0.004    -.2035705   -.0380734
         age |  -.0111889   .0006364   -17.58   0.000    -.0124363   -.0099415
       _cons |   1.036682   .3806903     2.72   0.006     .2905431    1.782822
-------------+----------------------------------------------------------------
```

purprice >= upper_quar

```
Regression with robust standard errors                 Number of obs =    5611
                                                       F(  3,  5607) =  517.10
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2883
                                                       Root MSE      =  .30139
```

```
------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .3475013   .0111973    31.03   0.000     .3255503    .3694524
         PED |  -.1737023   .0238613    -7.28   0.000    -.2204797   -.1269249
         age |   .0098229   .0004825    20.36   0.000     .0088769    .0107688
       _cons |   -.164284   .2252269    -0.73   0.466    -.6058159     .277248
------------------------------------------------------------------------------

Heckman selection model                         Number of obs      =     16871
(regression model with sample selection)        Censored obs       =     11690
                                                Uncensored obs     =      5181
                                                Wald chi2(3)       =   1465.73
Log likelihood = -9405.932                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hc_l         |
         IED |   .2309066   .0108299    21.32   0.000     .2096803    .2521329
         PED |  -.2090934   .0260132    -8.04   0.000    -.2600784   -.1581084
         age |  -.0087513   .0004523   -19.35   0.000    -.0096378   -.0078648
       _cons |   1.254743   .2386978     5.26   0.000      .786904    1.722582
-------------+----------------------------------------------------------------
```

no.rooms == 4

```
Regression with robust standard errors                 Number of obs =    4189
                                                       F(  3,  4185) =  377.52
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3261
                                                       Root MSE      =  .28667
------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |    .378475   .0133248    28.40   0.000     .3523513    .4045988
         PED |  -.2288018   .0305128    -7.50   0.000    -.2886231   -.1689805
         age |     .00658   .0004414    14.91   0.000     .0057147    .0074454
       _cons |  -.3890892   .2737313    -1.42   0.155    -.9257479    .1475696
------------------------------------------------------------------------------
```

norooms == 6

```
Regression with robust standard errors                 Number of obs =    4608
                                                       F(  3,  4604) =  907.27
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4293
                                                       Root MSE      =  .30573
------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .4858899   .0122347    39.71   0.000     .4619039    .5098758
         PED |  -.2605653   .0300213    -8.68   0.000    -.3194214   -.2017092
         age |   .0090209   .0005664    15.93   0.000     .0079105    .0101312
       _cons |  -.5680924    .277312    -2.05   0.041    -1.111757    -.024428
------------------------------------------------------------------------------
```

REGION 7: SOUTH WEST

purprice <= lower_quar

```
Regression with robust standard errors                 Number of obs =    2292
                                                       F(  3,  2288) =   71.34
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1232
                                                       Root MSE      =   .2035
------------------------------------------------------------------------------
```

```
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .1250089   .0156924     7.97   0.000     .0942362    .1557817
         PED |  -.4473455   .0568502    -7.87   0.000    -.5588288   -.3358622
         age |  -.0000974   .0004519    -0.22   0.829    -.0009836    .0007888
       _cons |    2.69447   .4846185     5.56   0.000     1.744133    3.644808
------------------------------------------------------------------------------


Heckman selection model                         Number of obs      =      2297
(regression model with sample selection)        Censored obs       =      1766
                                                Uncensored obs     =       531
                                                Wald chi2(3)       =    165.07
Log likelihood = -921.9325                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hc_l         |
         IED |  -.0068588   .0265054    -0.26   0.796    -.0588085    .0450909
         PED |  -.3918597   .0748382    -5.24   0.000    -.5385399   -.2451795
         age |  -.0089434   .0008002   -11.18   0.000    -.0105118   -.0073749
       _cons |   3.633144   .6388754     5.69   0.000     2.380971    4.885316
-------------+----------------------------------------------------------------
```

purprice >= upper_quar

```
Regression with robust standard errors                 Number of obs =    2266
                                                       F(  3,  2262) =  246.27
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2954
                                                       Root MSE      =  .27267
------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .2124361   .0199459    10.65   0.000     .1733221    .2515502
         PED |  -.6315267   .0505877   -12.48   0.000    -.7307299   -.5323235
         age |   .0095947   .0006143    15.62   0.000     .0083901    .0107993
       _cons |     4.2325   .4695381     9.01   0.000     3.311729     5.15327
------------------------------------------------------------------------------

Heckman selection model                         Number of obs      =      6770
(regression model with sample selection)        Censored obs       =      4619
                                                Uncensored obs     =      2151

                                                Wald chi2(3)       =    486.60
Log likelihood = -3627.918                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hc_l         |
         IED |    .149207   .0156382     9.54   0.000     .1185568    .1798573
         PED |  -.4499059   .0552265    -8.15   0.000    -.5581478    -.341664
         age |  -.0061243   .0005967   -10.26   0.000    -.0072938   -.0049548
       _cons |   3.532254   .4718803     7.49   0.000     2.607385    4.457122
-------------+----------------------------------------------------------------
```

norooms == 4

```
Regression with robust standard errors                 Number of obs =    1840
                                                       F(  3,  1836) =  176.60
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3335
                                                       Root MSE      =  .25232
------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .3235039   .0190791    16.96   0.000     .2860848    .3609229
```

```
         PED |  -.4257777   .0658347    -6.47   0.000    -.5548964    -.296659
         age |   .0059679   .0005463    10.92   0.000     .0048965    .0070394
       _cons |   1.595368   .5613101     2.84   0.005     .4944944    2.696241
------------------------------------------------------------------------------
```

norooms == 6

```
Regression with robust standard errors                 Number of obs =    2067
                                                       F(  3,  2063) =  414.94
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4265
                                                       Root MSE      =  .27512
------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .3571889   .0168817    21.16   0.000     .3240819    .3902958
         PED |  -.6297809   .0585675   -10.75   0.000    -.7446384   -.5149233
         age |   .0111983   .0006335    17.68   0.000     .0099559    .0124407
       _cons |   3.054953   .5048376     6.05   0.000     2.064908    4.044997
------------------------------------------------------------------------------
```

REGION 8: WEST MIDLANDS

purprice <= lower_quar

```
Regression with robust standard errors                 Number of obs =    1865
                                                       F(  3,  1861) =   46.50
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.1012
                                                       Root MSE      =  .19783
------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .1342823   .0168734     7.96   0.000     .1011895    .1673752
         PED |   -.278531   .0551406    -5.05   0.000     -.386675    -.170387
         age |  -.0009215   .0006072    -1.52   0.129    -.0021124    .0002693
       _cons |   1.309062   .4699931     2.79   0.005     .3872928    2.230831
------------------------------------------------------------------------------

Heckman selection model                         Number of obs      =      1866
(regression model with sample selection)        Censored obs       =      1481
                                                Uncensored obs     =       385
                                                Wald chi2(3)       =    157.14
Log likelihood = -727.1943                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hc_l         |
         IED |   .0520433   .0297011     1.75   0.080    -.0061699    .1102564
         PED |  -.3152483   .0819866    -3.85   0.000    -.4759391   -.1545575
         age |   -.011864    .001124   -10.55   0.000    -.0140671   -.0096609
       _cons |   2.875708   .6904479     4.16   0.000     1.522455    4.228961
-------------+----------------------------------------------------------------
```

purprice >= upper_quar

```
Regression with robust standard errors                 Number of obs =    1863
                                                       F(  3,  1859) =  232.35
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3415
                                                       Root MSE      =  .26097
------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .3088582   .0210446    14.68   0.000     .2675847    .3501316
         PED |  -.4525657    .062688    -7.22   0.000     -.575512   -.3296195
```

```
         age |   .0089811   .0006932    12.96   0.000     .0076216    .0103405
       _cons |   2.395024   .5623496     4.26   0.000     1.292121    3.497927
------------------------------------------------------------------------------

Heckman selection model                         Number of obs      =      5605
(regression model with sample selection)        Censored obs       =      3911
                                                Uncensored obs     =      1694

                                                Wald chi2(3)       =    476.84
Log likelihood = -2944.383                      Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hc_l         |
         IED |   .2740487   .0176761    15.50   0.000     .2394043    .3086932
         PED |  -.1400555   .0677043    -2.07   0.039    -.2727535   -.0073576
         age |  -.0070979   .0007466    -9.51   0.000    -.0085611   -.0056346
       _cons |   .5395614   .5705963     0.95   0.344    -.5787868    1.657909
-------------+----------------------------------------------------------------
```

no. rooms == 4

```
Regression with robust standard errors                 Number of obs =    1232
                                                       F(  3,  1228) =   87.08
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3110
                                                       Root MSE      =  .25822

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .3234151   .0260451    12.42   0.000     .2723172    .3745129
         PED |  -.2735463   .0981253    -2.79   0.005    -.4660582   -.0810345
         age |   .0085736    .000808    10.61   0.000     .0069883    .0101589
       _cons |   .3268279   .8217084     0.40   0.691     -1.28528    1.938936
------------------------------------------------------------------------------
```

norooms == 6

```
Regression with robust standard errors                 Number of obs =    1861
                                                       F(  3,  1857) =  397.57
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4771
                                                       Root MSE      =  .28335

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .4647668   .0204101    22.77   0.000     .4247376     .504796
         PED |  -.5025921   .0755128    -6.66   0.000     -.650691   -.3544932
         age |   .0105022   .0008526    12.32   0.000     .0088301    .0121744
       _cons |   1.551752   .6455064     2.40   0.016     .2857571    2.817746
------------------------------------------------------------------------------
```

REGION 9: NORTH WEST

purprice <= lower_quar

```
Regression with robust standard errors                 Number of obs =    2229
                                                       F(  3,  2225) =   45.75
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0881
                                                       Root MSE      =  .19799

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .1611128   .0152822    10.54   0.000     .1311439    .1910817
```

```
         PED |  -.0610271   .0433851    -1.41   0.160    -.1461066    .0240524
         age |  -.0006229   .0004847    -1.29   0.199    -.0015735    .0003277
       _cons |  -.6038128   .3691507    -1.64   0.102    -1.327729     .120103
------------------------------------------------------------------------------

Heckman selection model                         Number of obs      =      2230
(regression model with sample selection)        Censored obs       =      1842
                                                Uncensored obs     =       388
                                                Wald chi2(3)       =     30.68
Log likelihood = -856.6889                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hc_l         |
         IED |   .1396453   .0285046     4.90   0.000     .0837772    .1955133
         PED |  -.0342842   .1169826    -0.29   0.769    -.2635658    .1949974
         age |   .0016587   .0020521     0.81   0.419    -.0023635    .0056808
       _cons |   -.856824   .9701547    -0.88   0.377    -2.758292    1.044644
-------------+----------------------------------------------------------------
```

purprice >= upper_quar

```
Regression with robust standard errors                 Number of obs =    2225
                                                       F(  3,  2221) =  268.00
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3295
                                                       Root MSE      =  .27233
------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .3189694    .017483    18.24   0.000     .2846846    .3532542
         PED |  -.3698801   .0486508    -7.60   0.000    -.4652859   -.2744744
         age |   .0082317   .0006553    12.56   0.000     .0069467    .0095168
       _cons |   1.664833   .4342704     3.83   0.000     .8132143    2.516451
------------------------------------------------------------------------------


Heckman selection model                         Number of obs      =      6667
(regression model with sample selection)        Censored obs       =      4824
                                                Uncensored obs     =      1843

                                                Wald chi2(3)       =    574.26
Log likelihood = -3330.536                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hc_l         |
         IED |   .2867486   .0168327    17.04   0.000     .2537572    .3197401
         PED |  -.1464589   .0550584    -2.66   0.008    -.2543714   -.0385464
         age |  -.0077195   .0007236   -10.67   0.000    -.0091377   -.0063014
       _cons |   .5155134   .4697999     1.10   0.273    -.4052775    1.436304
-------------+----------------------------------------------------------------
```

norooms == 4

```
Regression with robust standard errors                 Number of obs =    1632
                                                       F(  3,  1628) =  159.01
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3538
                                                       Root MSE      =  .27673
------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .4031269   .0221922    18.17   0.000     .3595987    .4466552
         PED |  -.2366082    .072833    -3.25   0.001    -.3794645   -.0937519
         age |   .0070024   .0007385     9.48   0.000     .0055539    .0084508
       _cons |  -.3331529   .6125639    -0.54   0.587    -1.534649    .8683436
------------------------------------------------------------------------------
```

norooms == 6

```
Regression with robust standard errors                 Number of obs =    2298
                                                       F(  3,  2294) =  403.73
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4232
                                                       Root MSE      =  .29957

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .4620284   .0189028    24.44   0.000       .42496    .4990967
         PED |  -.3371585   .0580119    -5.81   0.000    -.4509198   -.2233972
         age |   .0101463   .0007878    12.88   0.000     .0086015    .0116911
       _cons |   .2358306   .5056719     0.47   0.641    -.7557913    1.227452
------------------------------------------------------------------------------
```

REGION 11: SCOTLAND

purprice <= lower_qua

```
Regression with robust standard errors                 Number of obs =    1532
                                                       F(  3,  1528) =   27.13
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0773
                                                       Root MSE      =  .23545

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .1523121   .0188225     8.09   0.000     .1153914    .1892327
         PED |   .0083085   .0425516     0.20   0.845    -.0751571    .0917742
         age |  -.0026657   .0007797    -3.42   0.001    -.0041951   -.0011363
       _cons |  -.9015887   .3584497    -2.52   0.012    -1.604694   -.1984833
------------------------------------------------------------------------------

Heckman selection model                         Number of obs      =      1535
(regression model with sample selection)        Censored obs       =      1231
                                                Uncensored obs     =       304
                                                Wald chi2(3)       =    100.00
Log likelihood = -551.5004                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hc_l         |
         IED |   .0256277   .0310136     0.83   0.409    -.0351578    .0864132
         PED |  -.0880664   .0716148    -1.23   0.219    -.2284287    .0522959
         age |  -.0097936   .0010041    -9.75   0.000    -.0117615   -.0078257
       _cons |   1.154372   .6034113     1.91   0.056    -.0282922    2.337037
-------------+----------------------------------------------------------------
```

purprice >= upper_qua

```
Regression with robust standard errors                 Number of obs =    1527
                                                       F(  3,  1523) =   93.42
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2094
                                                       Root MSE      =  .27723

------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .3073861   .0214573    14.33   0.000     .2652972     .349475
         PED |  -.0858232   .0448014    -1.92   0.056    -.1737022    .0020558
         age |   .0048048   .0008885     5.41   0.000     .0030621    .0065476
       _cons |  -.2180824   .4032092    -0.54   0.589    -1.008987    .5728217
------------------------------------------------------------------------------
```

```
Heckman selection model                         Number of obs      =      4570
(regression model with sample selection)        Censored obs       =      3066
                                                Uncensored obs     =      1504

                                                Wald chi2(3)       =    428.63
Log likelihood = -2548.351                      Prob > chi2        =    0.0000
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
hc_l         |
         IED |   .2466049   .0186274    13.24   0.000     .2100959    .2831139
         PED |  -.0906214   .0441116    -2.05   0.040    -.1770785   -.0041643
         age |  -.0086442   .0007589   -11.39   0.000    -.0101316   -.0071569
       _cons |   .4388869   .3843051     1.14   0.253    -.3143373    1.192111
-------------+----------------------------------------------------------------
```

norooms == 4

```
Regression with robust standard errors                 Number of obs =    1653
                                                       F(  3,  1649) =  171.65
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2905
                                                       Root MSE      =  .31457
------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .4482013    .020846    21.50   0.000     .4073139    .4890887
         PED |   .0078954   .0575836     0.14   0.891    -.1050492    .1208401
         age |   .0029198   .0009289     3.14   0.002      .001098    .0047417
       _cons |  -2.168921   .4817682    -4.50   0.000    -3.113863   -1.223979
------------------------------------------------------------------------------
```

norooms == 6

```
Regression with robust standard errors                 Number of obs =     856
                                                       F(  3,   852) =   92.84
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2972
                                                       Root MSE      =  .29631
------------------------------------------------------------------------------
             |               Robust
        hc_l |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         IED |   .4085415   .0298694    13.68   0.000     .3499152    .4671678
         PED |  -.1675963   .0522169    -3.21   0.001    -.2700851   -.0651075
         age |   .0060743   .0012775     4.75   0.000      .003567    .0085817
       _cons |  -.4553372    .472483    -0.96   0.335    -1.382704    .4720299
------------------------------------------------------------------------------
```

6 Conclusion and Recommendations

Plan:

6.1 Summary

This report has presented a variety of perspectives on the possible impact of new supply. One vista presents new dwellings, affordable only by high income groups, setting in motion a domino filtering effect throughout the housing system. The homes vacated by the wealthy fall in value and so become affordable to the next income tier. Adjustment continues until the lowest tier of housing becomes vacant and is eventually demolished. In a tenure system dominated by homeownership and where 'half the poor' are in owner occupation, this outcome may not be entirely desirable. Those on lowest incomes may not be able to sell their low quality properties making them immobile and their homes vulnerable to possession in the event of a downturn in their financial circumstances. There are, however, a number of questionable assumptions associated with the filtering model. First, it assumes that the highest income households have a preference for new construction, when in actual fact the reverse may be true. Listed buildings, dwellings of particular character that tie them to the ambience of a by-gone age, may be of hold special value in the market place. Such properties are intrinsically inimitable. If new houses are not of a particularly high standard they will be viewed as poor substitutes, and older luxury dwellings will become increasingly scarce as average incomes continue to rise. The demand curve will become steeper over time increasing the price effect of supply shifts in that sub-market. It might be that new housing is actually tailored primarily to of middle to low income where there may exist a plethora of substitutes, reflected in a shallow demand curve, dampening the price impact of shifts in supply.

Yet an other perspective of the supply story emphasises the role of turnover in determining the effect on macro house price indices. So for example if new dwellings do indeed have shallow demand curves, the low- micro price effect may be offset if such dwellings are also of the ilk that tend to be re-sold relatively frequently. The compensating role played by the peculiar way in which macro price indices are calculated may be purely illusionary reflecting computation bias rather than any true economic benefit. For example, if length of stay at least in part reflects household satisfaction with the dwelling and its surroundings, then the goal of supply policy should (in part) be to increase the number of properties which people will be happy to inhabit long tern.

These insights into the role of supply are complicated further by neighborhood effects that new supply can itself cause. Perhaps locating new construction adjacent to deprived areas can engender and upward regeneration cycle as more affluent households move in, local economic activity rises, schools improve further rounds of investment result. New supply has the potential to shape the character and perception of the neighbourhood and so may actually result in a rise in house prices, at least at the very local level.

6.2 The Role of Planning

The complexity that has emerged from this overview of the connections of city submarkets and dynamics to housing supply at first glance seems bewildering. A pessimistic response to this is to opt for the status quo of ignorance – cities and their housing markets are too complex to understand so let's continue to make planning and investment decisions from a position of agnosticism or on purely political lines. Less pessimistic, and certainly more radical, is the view taken by Portugali (2000; reviewed in Meen and Meen 2003). Portugali's view of the planning system is as a top-down hierarchical system, where targets are set at the national level and fed down through the various branches of the rigid hierarchical tree without any interactions between branches at the same spatial level. Cities, in contrast, can be viewed as lattices "in which there is no central organising authority, but there are overlaps between systems at similar spatial levels" (Meen and Meen 2003 p. 931). This implies that planning authorities would do better to work *with* the fluidity of the city rather than attempt to manage and structure it.

There is, however, a middle road and it rests very heavily on a concerted effort at all levels of the planning system to advance our common understanding of the economic forces that drive city development and its success. After surveying the exotic array of non-linear and self-organising models of socio-economic interaction in the housing market that have been developed from the science of biological and physical systems, Meen and Meen (2003) conclude that,

> '… the problem is not intractable because … the analogies between biological or physical systems and housing markets are not exact …additional information is available on real housing systems that generates a degree of predictability and controllability… locations are not featureless, implying that some areas are more likely to act as basins of attraction than others. There is considerable empirical evidence at the national level that house price movements are partly predictable (see Meen, 2002). At the regional level in the UK, house prices exhibit a ripple effect, suggesting that future price movements in one region can be predicted from past changes in another. In addition, spatial contiguity in price movements occurs at the local level" (Meen and Meen, 2003, pp. 931-2).

And that is the view taken here. Whilst there are examples of rapid decline or advancement, what is perhaps most apparent (at least anecdotally) is the persistence of submarkets over time. The West End has been a desirable locale within Glasgow for more than a hundred years and it remains so. Edinburgh's New Town has similarly remained a relatively prosperous and attractive locale since the time it was built. Kensington in London, and Didley in Manchester, are further examples among many of spatial inertia rather than dynamism. There is much to be gained, therefore, even from cross-sectional snapshots of submarket structures, and even more to be gained from an investigation into the demand elasticity and dynamic behaviour of the urban system within these housing segments.

6.3 A Suggested Programme of Research

Our knowledge base be at a low starting point but there are very real improvements within our grasp. Whilst the submarket structure of most UK cities may be something of an unknown, it does not have to remain so. The coincidence of recent data opportunities in the UK with developments in the methods available for examining submarkets give rise to a potential sea change in our capacity to understand the role of submarkets in determining the outcome of new supply. These opportunities can be synthesised into four phases of proposed research:

1. Delineation of submarkets for 6 UK cities

I speak here of the application of the grid-search approach to the location of submarket boundaries described in chapter **2** allied with geographically weighted regression approaches of the kind developed by Stewart Fotheringham and others. A variety of cities could be selected. It is recommended that at least one city from the three constitutional regions be selected (I suggest Cardiff, London, and Glasgow), supplemented by two other industrial conurbations (Birmingham and Manchester) plus a non-industrial city (such as York).

2. Modelling the Dynamic Properties of Submarkets

This includes an investigation into (i) the movement of submarket boundaries over time using Nationwide and CML data; (ii) variations in the rate of turnover across space and time; (iii) submarket house price volatility; and (iv) variation in liquidity between submarkets. A more ambitious version of this project would seek to establish the interactions between submarkets and the existence of ripple effects within the urban system using techniques analogous to those applied to regional level studies of such effects (see Andrew and Meen 1998).

3. Identifying spatial variation in regeneration effects of new supply

This would entail an analysis of the impact of new supply on the prices of second hand dwellings in the proximity using separate analysis by submarket. An alternative and complementary approach would be to apply geographically weighted regression techniques with a view to deriving contours of the price effects of new supply.

4. Demand Elasticity Structure of 6 UK cities

Hendershott, Pryce and White (2003) have demonstrated that income and price elasticities of demand can in principle be estimated from CML data. This project would estimate these demand elasticities at submarket level for the selected cities and infer from these estimates the effect of new supply. Again, a combination of grid-search and geographically weighted regression techniques will be applied.

The only necessary sequence to these projects in that project 1 should precede the remaining projects because submarkets have to be appropriately defined before they can be used as the basis for analysis of regeneration effects, dynamics, and demand elasticities. Projects 2, 3 and 4 could conceivably be executed in any order, though there might be a case for project 2 to come before the remaining projects since it will add substantially to our understanding of the persistence and evolutionary nature of submarkets, which will itself add to our understanding of how regeneration and elasticities should be analysed. The duration of these projects will depend on the number of cities considered and the depth and sophistication of the analysis carried out. By applying a consistent set of analyses to submarkets of different cities, the result will be a comprehensive set of information whereby cities and submarkets can be compared on a national scale and where future changes can be gauged in a systematic way.

Since the great majority of this analysis will be based on data available for all UK cities, the approach could subsequently be applied to further cities and conurbations. It is hoped that by making the results freely available to local authorities, landlords and developers, the information will be widely disseminated. This process could be assisted by the development of a dedicated website providing submarket information on the cities considered, a site that could be augmented and up dated as subsequent cities are subjected to the analyses proposed.

7 References

Adair, A.S., Berry, J.N. and McGreal, W.S., 1996. Hedonic modelling, housing submarkets and residential valuation. *Journal of Property Research*, 13, pp. 67–83.

Afifi, A.A. and Clark, V., 1990. Computer-Aided Multivariate Analysis. , Chapman & Hall, London.

Allen, M.T., Springer, T.M. and Waller, N.G., 1995. Implicit pricing across residential submarkets. *Journal of Real Estate Finance and Economics*. 11, pp. 137–151.

Bajic, V., 1985. Housing-market segmentation and demand for housing attributes: some empirical findings. *AREUEA Journal*, 13, pp. 58–75.

Banerjee, S., Gelfand, A.E. and Sirmans, C.F. (2003) Directional Rates of Change under Spatial Process Models, *Journal of the American Statistical Association*, 98, 464, 946-954.

Bourassa SC, Hoesli M, Peng VS (2003) Do housing submarkets really matter? *Journal Of Housing Economics* 12 (1): 12-28

Bourassa, S.C., Hamelink, F., Hoesli, M. and MacGregor, B.D., 1999. Defining housing submarkets. *Journal of Housing Economics*, 8, pp. 160–183.

Bramley, G., Pryce, G. & Satsangi, M. (1999) *'The Supply Responsiveness of the Private Rented Sector: An International Comparison'*, Department of the Environment, Transport and the Regions: London.

Crosson, S.T., Dannis, C.G. and Thibodeau, T.G., 1996. Regression analysis: a cost-effective approach for the valuation of commercial property portfolios, *Real Estate Finance*, 13 4, pp. 20–28.

Dale-Johnson, D., 1982. An alternative approach to housing market segmentation using hedonic pricing data. J. Urban Econ. 11, pp. 311–332.

Deng, Y., Fisher, J.D., Sanders, A.B. and Smith, B. (2004) "Estimation of NOI Growth, Volatility and Clustering by MSA", Preliminary Draft.

Fik TJ, Ling DC, Mulligan GF (2003) Modeling spatial variation in housing prices: A variable interaction approach, REAL ESTATE ECONOMICS, 31 (4): 623-646

Fisher J, Gatzlaff D, Geltner D, Haurin D (2003) Controlling for the impact of variable liquidity in commercial real estate price indices REAL ESTATE ECONOMICS 31 (2): 269-303 SUM 2003

Galster, G.C., 1987. Residential segregation and interracial economic disparities: A simultaneous-equations approach. J. Urban Econ. 21, pp. 22–44. Abstract-EconLit | Abstract-GEOBASE

Gatzlaff DH, Haurin DR (1997) Sample selection bias and repeat-sales index estimates JOURNAL OF REAL ESTATE FINANCE AND ECONOMICS 14 (1-2): 33-50 JAN-MAR 1997

Gatzlaff DH, Haurin DR (1998) Sample selection and biases in local house value indices JOURNAL OF URBAN ECONOMICS 43 (2): 199-222 MAR

Gatzlaff, D.H. and D. R. Haurin (1994) Measuring changes in local house prices: an empirical investigation of alternative methodologies, *Journal of Urban Economics*, 35, 221-244.

Gatzlaff, D.H. and Haurin, D.R., 1997. Sample selection bias and repeat-sales index estimates. J. Real Estate Finance Econ. 14, pp. 33–50.

Gatzlaff, D.H. and Haurin, D.R., 1998. Sample selection and biases in local house value indices. J. Urban Econ. 43, pp. 199–222.

Gelfand, A.E., Kim, H.J. Sirmans, C.F. and Banerjee, S. (2003) "Spatial Modelling with Spatially Varying Coefficient Processes", *Journal of the American Statistical Association*, 98, 462, 387-396.

Goodman AC, Thibodeau TG (2003) Housing market segmentation and hedonic prediction accuracy, JOURNAL OF HOUSING ECONOMICS,12 (3): 181-201

Goodman, A.C. and Dubin, R.A., 1990. Sample stratification with non-nested alternatives: theory and a hedonic example. Rev. Econ. Stat. 72, pp. 168–173.

Goodman, A.C. and Kawai, M., 1982. Permanent income, hedonic prices, and demand for housing: new evidence. J. Urban Econ. 25, pp. 81–102.
Goodman, A.C. and Thibodeau, T.G., 1998. Housing market segmentation. J. Housing Econ. 7, pp. 121–143.
Hamilton, T.W., 1998. Real estate market segmentation. Assessment J. 5, pp.56–69.
Harsman, B. and Quigley, J.M., 1995. The spatial segregation of ethnic and demographic groups: comparative evidence from Stockholm and San Francisco. J. Urban Econ. 37, pp. 1–16.
Heckman, J. "Sample Selection Bias a specification error", *Econometrica*, 47, 153-161
Hendershott, P., Pryce, G. and White, M. (2003) 'Household Leverage and the Deductibility of Home Mortgage Interest: Evidence from UK New Borrowers', *Journal of Housing Research*.
Hoyt, H. (1939) The structure and growth of residential neighbourhoods in American Cities. Government Printing Office, Washington, DC.
Hwang, Min and Quigley J. (2004) Selectivity, quality Adjustment and Mean Reversion in the Measurement of House Values, *The Journal of Real Estate Finance and Economics*, 28, 2/3, 161-178.
Kauko T, Hooimeijer P, Hakfoort J (2002) Capturing housing market segmentation: An alternative approach based on neural network modelling HOUSING STUDIES 17 (6): 875-894
Kemp, P. and Pryce, G. (2002) '*Evaluating the Mortgage Safety Net*', Council for Mortgage Lenders.
Kennedy, P., 1994. A Guide to Econometrics. (third ed.),, MIT Press, Cambridge, MA.
Lamont O. and Stein J.C. (1999), "Leverage and House Price Dynamics in U.S. Cities." *Rand Journal of Economics* 30(3): 498-514.
Maclennan, D. and Tu, Y., 1996. Economic perspectives on the structure of local housing systems. Housing Stud. 11, pp. 387–406.
MacQueen J.B., 1967. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 281–297
Mark, J. and Goldberg, M.A., 1988. Multiple regression analysis and mass assessment: a review of the issues. Appraisal J. 56 1, pp. 89–109.
Meen G.P. (2001), "Social Interactions in Local Housing Markets: A Research Programme for Quantitative Housing Analysis", *mimeo.*
Meen, G. and Andrew, M. (1998) "Modelling Regional House Prices: A Review of the Literature", Report Prepared for the Department of the Environment, Transport and the Regions, London.
Michaels, R.G. and Smith, V.K., 1990. Market segmentation and valuing amenities with hedonic models: the case of hazardous waste sites. J. Urban Econ. 28, pp. 223–242. Abstract-EconLit
Mills, E.S. and Hamilton, B.W., 1994. Urban Economics. (fifth ed.),Addison-Wesley, Reading, MA.
Pace, R.K. and Gilley, O.W., 1989. Appraisal across jurisdictions using bayesian estimation with bootstrapped priors for secondary mortgage market applications. Property Tax J. 8 1, pp. 27–42.
Palm, R., 1976. Urban social geography from the perspective of the real estate salesman. Center for Real Estate and Urban Economics Research Report No. 38, University of California, Berkeley
Palm, R., 1978. Spatial segmentation of the urban housing market. Econ. Geogr. 54, pp. 210–221.
Portugali, J. (2000) *Self-organisation and the City*. Berlin: Springer.
Pryce and Gibb (2004a) Graphical analysis of TOM by Glasgow Submarket
Pryce and Gibb (2004b) Bootstrapped stratified proportional hazard estimation
Pryce, G. & Keoghan, M. (2001) 'Determinants of Mortgage Protection Insurance Take-Up', *Housing Studies*, 16, 2, 179-198.

Pryce, G. (1999) 'Construction Elasticities and Land Availability: A Two Stage Least Squares Model of Housing Supply Using the Variable Elasticity Approach', *Urban Studies*, 36, 13, 2283-2304.

Pryce, G. (2003) 'Greening by the market? Distortions caused by fiscal incentives to build on brownfield land', *Housing Studies*, 18(4).

Pryce, G. (2004) "Endogenous Delineation of Submarkets", presentation at Urban and Regional Economics Seminar, Edinburgh.

Pryce, G. (2004) "Endogenous Delineation of Submarkets", presentation at Urban and Regional Economics Seminar, Edinburgh.

Pryce, G. and Gibb, K. (2003) "*Modelling Time to Sale of Second Hand Owner Occupied Residential Properties in Glasgow*", paper presented to ERES Conference, Krakow.

Rapkin, C., Winnick, L. and Blank, D. M. (1953) Housing Market Analysis Washington: US Housing and Home Finance Agency.

Rothenberg, J., Galster, G.C., Butler, R.V. and Pitkin, J.R., 1991. The Maze of Urban Housing Markets. , University of Chicago Press, Chicago.

Schnare, A.B. and Struyk, R.J., 1976. Segmentation in urban housing markets. J. Urban Econ. 3, pp. 146–166.

Schnare, A.B., 1980. Trends in residential segregation by race: 1960–1970. J. Urban Econ. 7, pp. 293–301.

Stein, J. C. (1995), "Price and Trading Volume in the Housing Market", Quarterly Journal of Economics, 1102):379-406.

Stein, J.C. (1995) Prices And Trading Volume In The Housing-Market - A Model With Down-Payment Effects, Quarterly Journal of Economics, 110 (2): 379-406

Sweeney, J.L. (1974) "A Commodity Hierarchy Model of the Rental Housing Market," Journal of Urban Economics, 1:288-323.

Taylor, M., Rubin, G. and Lynford, L., 2000. Submarkets matter! Applying market information to asset-specific decisions. Real Estate Fin. 17 3, pp. 7–26.

Thibodeau TG (2003) Marking single-family property values to market REAL ESTATE ECONOMICS 31 (1): 1-22

Watkins, C.A. (2001) "The definition and identification of housing submarkets", Environment and Planning A, 33, 2235-2253.