Housing Market Segmentation: The Theory and Measurement of Submarkets[1]

Gwilym Pryce[a]

[a] *Department of Urban Studies, University of Glasgow, Glasgow G12 8RS, UK*

Working Draft as at 30th April 2009

DRAFT – PLEASE DO NOT QUOTE WITHOUT PERMISSION

Abstract

This paper aims to stimulate a step-change in how and why submarkets are analysed. Recent work on submarkets has focussed on improving prediction accuracy but there is more to submarkets than regression refinement. Submarkets are important because they reflect how housing market respond to, and interact with, social and spatial processes at the local level. This paper attempts to establish a set of criteria that submarket methodologies needs meet in order to investigate nature and meaning of submarkets in a more robust, purposeful and creative way. Existing approaches are critically evaluated using these criteria, and a more suitable methodology is proposed, grounded in the notion of submarkets as a function of substitutability, with a view to helping researchers address a richer set of questions regarding housing submarkets. We illustrate how this approach could be applied using data on Glasgow.

Keywords: Housing submarkets; Substitutability; House prices; Hedonic; Cluster analysis; Neighbourhoods

---

Introduction

How are market segmentation and social segmentation related? Spatial segmentation and the associated issues of social cohesion and integration are central to the debates surrounding the nature of a good society (Chesire 2008; Lees 2008). And these debates have permeated public policy. Social inclusion/exclusion, neighbourhood regeneration, mixed communities – these have all shaped the fabric of policy innovation over the last two decades with varying degrees of success (Ostendorf *et al* 2001; Chesire 2008; Lees 2008; Lupton and Tunstall 2008). This policy interest has stimulated, and been stimulated by, a considerable volume of

[1] *Acknowledgements*: The author is grateful to the UK Department of Communities and Local Government for providing the initial funding to develop practical methods for defining submarkets. This paper represents an extension of that work. Thanks also to George Galster, Chris Leishman and Eric Levin for comments and feedback on the ideas that underpin this work. Any remaining errors are my own.

theoretical and empirical work in the academic literature regarding the segmentation of communities and the related socio-economic issues of neighbourhood mix, neighbourhood effects, neighbourhood dynamics (Blasius *et al* 2007; Kearns and Mason 2007; Andersson et al 2007; Galster 1981a,b , 2007). At the same time, research on local housing markets has flourished, and the importance of house prices in impacting the economy at every spatial scale has never been more apparent (Sanders 2008; Kaplan 2009; Goodhart and Hoffmann 2008).

Given the potentially important interaction between social segmentation and the operation of local housing markets, it is rather surprising, therefore, that the two strands of research have remained relatively detached. The silent divorce between work on neighbourhoods and research on housing submarkets becomes all the more surprising when one considers the strong links made in the early submarkets literature (Strazheim 1975, Grigsby *et al*, 1963 Rothenburg *et al* 1991); and when one reflects on the potentially rich theoretical vein to mined on the two-way interaction of housing market and social processes. Submarkets are a potentially powerful framework for conceptualising and calibrating the deep processes that lead to segmentation (Schelling 1971; Galster and Killen 1995; Galster et al 2000; Meen and Meen 2003) and shape fundamental structure of cities (Maclennan 1982; Brueckner et al 1999).

Very little attention has been paid in the recent cohort of submarket papers to these sorts of theoretical possibilities, however. Issues of urban form, social mix, racial contiguity, amenity access and externality have come to form the backdrop of the empirical studies of submarkets, but are rarely the main object of them. The goal is usually to improve out-of-sample prediction accuracy – important for maximising the performance of mass appraisal models. And to that end, much progress has been made (Baroussa et al 2007; Goodman and

Thibodeau 2003). It could be argued, however, that the pursuit of prediction accuracy has led to submarket innovation being confined to a fairly narrow avenue of empirical refinement. Most recent submarket papers conform to a remarkably similar methodological checklist: (i) cluster of dwellings by physical attributes; (ii) incorporate these clusters into hedonic regression; (iii) verify the existence of submarkets by testing for breaks in the attribute coefficients; though not necessarily in this order. Methodological variations have tended to be confined to different approaches to (i), (ii) or (iii).

We argue, therefore, that there has been a drift away from (a) theoretical research on submarkets and (b) research that integrates with neighbourhood and socio-economic analysis, and that there has been a drift towards (c) submarkets being viewed largely as a means of improving regression fit. These three tendencies have coincided with: (d) the increased importance of automatic valuation methods in the appraisal of properties by lenders and real estate agents, and (e) the adoption of the Law of One Price (LOP) as the defining principle by which submarkets are defined and estimated. The apparent (but questionable – see section 1) theoretical robustness of the LOP approach and the ease with which it can be translated into hedonic regression analysis has, it is argued, stilted innovation. An important consequence is that submarket estimation methods may not have evolved along lines that are appropriate for some of their most important applications. Certainly, the recent government interest in the UK in defining housing market areas extend far beyond econometric refinement . There is a recognition that the location, characteristics, shape, and dynamics of housing submarkets have the potential reveal insights into the likely asymmetries in adjustment to local shocks (new infrastructure, in-migration, increased supply, environmental shocks and hazards), how housing markets interact with related sectors (employment, transport, social mix, crime) and how they relate to the rich tapestry of amenity bundles and consumer behaviour (Barker

2004, 2005; Pryce 2005; Bates 2006). These are issues of importance to policy makers[2] but it is questionable how useful the current approaches offered in the academic literature can be in helping inform that interest.

The main goal of this paper is to open up the debate over submarkets by highlighting weaknesses in the existing LOP/HAPV consensus. We hope to demonstrate that there is still much to be done in submarkets research. Moreover, we seek to provide direction and structure to that research by illustrating how the literature might develop along more creative and theoretically motivated lines.

The paper is structured as follows. In Section 1, an attempt is made to step back from the empirical consensus and ask what qualities would we like a submarket estimation method to have? Criteria are proposed and the existing literature is evaluated accordingly. In Section 2, a new approach to submarkets estimation is presented that goes some way towards addressing these criteria. Section 3 illustrates how our proposed method could be used to identify submarkets in Glasgow.

1) What do we require from Submarket Estimation Methods?

The criteria by which submarket methods should be evaluated fall into three broad categories: a) Theoretical Robustness; b) Methodological Robustness; and c) Versatility and Scope (this is not meant to be a strict categorisation – some criteria fit equally well in more than one category).

[2] Consider, for example, the recent programme of research funded by central government seeking to define housing market boundaries at the sub-regional level for the whole of the UK – Pryce and Evans 2007; NPHAU call 2009.

a) Theoretical Robustness

- *Conceptually Distinct:* LOP/HAPV has become the most commonly applied criterion for defining submarkets (Watkins 2001), based on the principle that competition between sellers will ensure that only one price prevails. In this tradition, the rationale for submarkets is that differences in exposure to externalities and access to amenities lead to shifts in attribute prices. This approach, however, leaves submarkets vulnerable to being defined away. If the utility of a house is a function of its location and structural attributes, why separate location and structural attributes for the purposes of submarket definition? The decision is arbitrary and leads to a very weak definition of submarkets – it allows one to subsume the entire notion of submarkets by simply allowing z_i to include a mixture of location and structural attributes, and interactions between the two (in the notation of Rosen (1974), we are considering a class of commodities – i.e. houses – that are described by n attributes or characteristics, $\mathbf{z} = (z_1, z_2, ..., z_n)$). At one stroke, submarkets have disappeared! If $\mathbf{z}$ includes location characteristics, then interaction between the z's seems perfectly plausible, in which case the fact that proximity to quality schooling interacts with the price per room, for example, does not in itself imply submarkets. It is the demand curve for the dwelling as a whole we are most interested in, but hedonic price functions alone cannot reveal this (Rosen, 1974, p. 50: "Observed marginal hedonic prices … reveal little about underlying supply and demand functions").

- *Robust to the Transformative Interaction Effects:* LOP is a poor criterion for housing submarket analysis, not because LOP is not a sound theoretical principle *per se*, but because true attribute prices are so difficult to measure in housing. This is because of (i) *Transformative Interaction Effects* (TIEs) between attributes, and (ii) the Many to Many Mapping of Means and Ends (MMME) – the same human need can be met in very different ways (there is more than one way to skin a cat), and the same means will meet different needs for different people. Consequently, two goods can have very few common attributes, and very divergent attribute prices, yet still be considered as close substitutes.

Consider the options for crossing the English Channel. Planes, trains, automobiles and ferries are all close substitutes – a change in the price or availability of one can have a large effect on the demand for the others.[3] They are all in the same market – transferring passengers across the Channel. Yet, these four modes of transport have few common attributes and so it would be meaningless to calculate attribute prices. Wings on cars add nothing to their value, whereas wings on planes are essential to their function. The attribute price of wings would reflect this – high value of wings on planes, zero or negative value of wings on cars. The reason is that the utility of wings is transformed when appropriately combined with a jet engine. The interaction effect is transformative – when the two components are combined they become something completely different. Similarly, wheels on ferries are of no value, in contrast to their worth in the functioning of cars and trains, etc. (There are potential parallels in housing: picture windows can be of great value when a house is not overlooked and has a spectacular views, but rather less desirable in a dwelling with little privacy and an unsightly outlook). Divergent attribute prices would lead to the erroneous conclusion that the four modes of transport are distant substitutes, when we know that they are not.

The greater the physical heterogeneity of the goods being compared, the less relevant HAPV is to substitutability – if two goods have four common attributes, but ten that are not, differences/similarities in the attribute prices of the common four attributes are neither here nor there. Unsurprisingly, in the wider economics literature, studies that estimate substitutability do not typically resort to comparing HAPV. To test whether two candy bars are in the same submarket one would not test for equality of attribute prices,

[3] Anguera (2006) records that as the number of Channel Tunnel passengers increased from 0.1 million in 1994 to 6.3 million, the number of ferry passengers fell from 23.7 million to 16.6 million over the same period. See also Szymanski (1996, 2005).

which are essentially irrelevant, or at least potentially misleading. The standard method (with good reason) is to estimate the Cross Price Elasticity of Demand which is based on the rates of change of prices and quantity demanded rather than static estimates of differences in input prices.

HAPV is a poor guide to substitutability and this also alters the interpretation of spatially autocorrelated errors. Because goods can be close substitutes but have different attributes, there is no reason to believe that spatial patterns in "uncaptured non-linear relationships between the dependent and independent variables" (Tu et al 2007 p. 388) in a hedonic regression will have any bearing on where fissures in substitutability lie. Even if one were able to measure, without error, all the physical and amenity differences between dwellings, it is possible that very different bundles of physical and location attributes are actually be perceived to be close substitutes by consumers – it is the utility of the inseparable final bundle in its entirety that a buyer is purchasing, and the whole may be greater than the sum of parts. Any method of defining submarkets should not, therefore, presuppose that attribute prices of dissimilar dwellings reflect substitutability.[4] The case for using spatially autocorrelated errors as a measure of substitutability is further weakened by the arbitrary nature of uncaptured non-linearities. For example, such non-linearities may be captured by our estimation for block *i* but not for block *j*, so only block *j*'s non-linearities are not captured by our particular regression. Because it is a matter of chance which non-linearities happen to be captured, it is also a matter of chance where the spatial autocorrelation in the errors will lie, so we cannot be sure that spatial clusters of such errors will even tell us where HAPV bounds are likely to lie.

Transformative Interaction Effects and the fact that the same need can be achieved by very different means, fundamentally undermines the simple grouping of the housing stock by attributes. Cluster methods have become popular in the submarkets literature,

4

but such approaches are theoretically problematic if they are applied to attributes that occur physically in the housing stock. Clustering dwellings by physical attributes does not necessarily reflect how consumers would group them. In fact, consumer behaviour is entirely absent from such an application – one would ideally like to group according to some market related measure. In transactions data, the only behavioural variable typically measured is selling price, but this is the dependent variables and is excluded from the clustering process. Similarly, grouping variables using factor analysis imposes a structure on the functional form of the hedonic equation that removes the possibility of capturing interactions between individual attributes in determining prices. Locational variables do not interplay with how individual attributes interact with one another, which I think is also implausible. One would like to allow prices, attributes and location to cluster freely.

- *Robust to Continuity:* An issue related to that of definition is continuity. In a LOP/HAPV view of the world, submarkets are often thought of as discrete breaks in the land rent surface (see, for example, Fik *et al* 2003 p.635, 638). If there are no discrete breaks, there are no submarkets. This contrasts strongly with a substitutability approach where submarkets can exist just as legitimately along a continuum as they can in discrete silos. Indeed, one of the facets of submarkets we might be most interested in is to what extent are boundaries thin and precipitous and to what extent are they broad and gradual (revealing, for example, whether there are forces at work that drive housing markets to segment into ever more discrete and specialised neighbourhoods (Cheshire 2007), or whether the forces that dominate (such as preference for social mix) are those that lead to boundary gradation. Evidence for discontinuity may determine whether we represent submarkets as discrete segments rather than a continuous lattice of substitutability which spans, uninterrupted, the urban system. But it will not determine whether or not submarkets exist. There is debate over whether human life begins at conception or whether evolves gradually, but neither philosophy questions whether human life exists at all! We therefore seek a method of identifying and measuring submarkets that does not rely on discontinuities, but does help us calibrate the gradient of boundaries.

- *Logically Robust:* Let P1 be the statement that dwellings *i* and *j* have the same attribute prices, and let P2 be the statement that *i* and *j* are in the same submarket. The logic of submarket testing usually runs as follows:

$$\begin{array}{l} P1 \\ P1 \Rightarrow P2 \\ \hline \therefore P2 \qquad (\textit{modus ponens}) \end{array}$$

The problem is that P1 does not ⇒ p2: similarity of attribute prices does *not,* in fact, imply that two dwellings are in the same submarket. *i* could be located in Paris, and *j* in Glasgow. That they have similar attribute prices at a given point in time is coincidental – the two properties are highly unlikely to fall into the same choice set of any one buyer, *b*, and cannot meaningfully be described as belonging to the same submarket. What we can say, however, is that P2 ⇒ P1, that is, if two dwellings are elements of the same submarket, $i, j \in S_k$, then they will have the same attribute prices:

$$\begin{array}{l} P2 \\ P2 \Rightarrow P1 \\ \hline \therefore P1 \end{array}$$

Application of *modus tollens* allows us to deduce that if *i* and *j* do not have the same attribute prices then they will not be elements of the same submarket:

$$\begin{array}{l} \neg p1 \\ p2 \Rightarrow p1 \\ \hline \therefore \neg p2 \qquad (\textit{modus tollens}) \end{array}$$

where "¬" represents negation. In other words, the Law of One Price provides a necessary but not a sufficient condition for two dwellings to be in the same submarket. However, the same is not true for ¬HAPV, at least not for observed ¬HAPVs. This is because TIEs and MMMEs make it difficult to identify what constitutes an attribute. If one observes HAPV, how does one know that all relevant attributes have been included, and whether one has found a way of truly capturing transformative interactions? The same questions apply to ¬HAPV. In summary, HAPV is not a sufficient condition for the same reason that LOP is not a sufficient condition; and observed HAPV is not a necessary condition for the reasons set out above (TIEs and MMMEs).

- *Robust to Inter-Submarket Migration:* One approach to defining housing market areas is to use the pattern of migration flows (Jones and Mills 1996, Scottish Homes, 1993). Using a threshold of 50% of buyers moving within an area, housing market areas are derived. The approach is problematic, however, because moves do not necessarily indicate coincidence of submarkets between source and destination. As Grigsby (1963, p.40-41) pointed out, people seek to move in order to *change* submarkets, "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, it must be different." (p.40-41). Indeed, one has to ask why a family would face significant transactions costs in order to remain in the same submarket, "… 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" (p.40-41). We therefore seek a method that does not preclude inter-submarket migration.

b) Methodological Robustness

- *Robust to Disequlibria:* HAPV does not cope well with the perpetual disequilibrium of housing markets (in our data, for example, there was no period for which house price change was zero; see also Maclennan's 1981 critique along these lines). One has to assume both that prices of all dwellings adjust simultaneously and that one is able to observe a large number of properties selling at exactly the same point in time (if one includes properties at different time periods, and prices are in flux, then one is averaging over different points in the adjustment process leading one to observe spurious differences or similarities in attribute prices). Disequilibrium fundamentally undermines HAPV as a sufficient condition because the confidence intervals (CIs) of attribute prices in two different submarkets may coincidentally overlap in any single time period. Similarly, it weakens HAPV's potency as a necessary condition because CIs of attribute prices of properties in the same submarket may temporarily be disjoint in any single time period (particularly if sample sizes are large, leading to narrow CIs). Disequilibria also undermines the use of static clustering of spatially autocorrelated errors as a means of defining submarkets (Tu et al 2007). House price levels and attribute price levels observed at snapshots in time are not good measures on which to rest our definition of submarkets. A better approach would be to consider the behaviour of prices over a prolonged period and exploit the fact that, whatever the observed attribute prices in a particular period, one would expect the variations in value of the entire housing bundle to be correlated within submarkets (more on this in the next section).

- *Robust to Attribute Measurement Errors:* An important feature of housing data is that the final sale price can be measured with great precision but that the characteristics of the dwelling and its location cannot. The fundamental problem is that is not only dwellings that are heterogeneous but attributes also. The apparent difference in price per room between tenements and modern flats, for example, may simply reflect the fact that tenement rooms are larger or have higher ceilings. So one is not observing a separate submarket, merely a failure of measurement. Measuring attribute quantity is critical to the LOP/HAPV approach. That half a tank of petrol costs less than a full tank is not an indication that there is a difference in the price per unit. One would like to observe full details on the quality and quantity of each attribute in every dwelling. Unfortunately, such information is rarely available and the measurement errors will not be random but correlated with building type, which in turn tends to be clustered across space. So, in testing for coefficient shifts in hedonic regressions, one may simply end up identifying where the measurement errors occur – either errors of omission (unobserved attributes) or errors of measurement (failure to gauge the true quantity of an attribute) – rather than submarket boundaries. The same point applies to observed clusters of spatially autocorrelated errors (Tu *et al* 2007) – they may reflect clusters of unobserved attributes, measurement errors and non-random errors that arise from functional form (exacerbated by TIEs and MMMEs). Note that, if two dwellings genuinely belong to the same submarket, one would expect the price of the overall housing bundle to respond similarly to demand and supply shocks. In other words, one would expect their price dynamics to be correlated over time, even if there are errors in estimating individual attribute prices. This suggests we should focus on the final sale price (which is measured with precision) rather than attribute prices (which are not) as the basis for submarket analysis.

- *Robust to the Arbitrary Weighting Problem:* Even where structural breaks are statistically significant, the size of the variation in attribute price can vary hugely for different attributes. This is often overlooked when using F-tests and LM tests, but these approaches do not escape the arbitrary weighting problem. What if the average variation in attribute prices across the different attributes is small, but some attribute prices vary hugely? And what if the structural break is statistically significant but small? If some attribute prices are different but others are the same, does one weight each equally, or are some attributes more important? This is made all the more problematic if different dwellings have different *quantities* of certain attributes. And if attributes interact with one another in determining utility, as they almost certainly do, how can we be sure we have correctly identified the marginal prices from such a complex system, if they can be identified at all (see section IV of Rosen, 1974, and the discussion of TIEs and MMMEs above)?

- *Not Reliant on Pre-Determined Boundaries:* The conventional test for violations to the Law of One price is to test for breaks at a point in the sample, which in many early studies involved using adminstrative boundaries. In practice, this amounts to running a Chow Test or LM test for a shift in hedonic coefficients across an administrative (or some other *a priori*) boundary. This may tell us little about the actual sub-structure of the urban housing market because it is only testing for breaks along pre-defined (often non-market) fault-lines, whereas we know that in many places submarket boundaries may have little in common with administrative areas (Bates, 2006). Indeed, Clapp et al (2007) and others have noted that administrative boundaries may systematically cut through submarkets because they are often drawn along railway lines, roads, rivers and other physical landmarks, which may in fact form the locus around which submarkets develop. This problem is compounded by the fact that, when structural break tests are used to identify spatial boundaries, the direction of the break-test might be important. For example, a structural break might exist north-south of a particular point, but not east-west. Because there are, at each point, an infinite number of possible directions along which to test for submarket boundaries, point-by-point testing becomes prohibitively cumbersome.

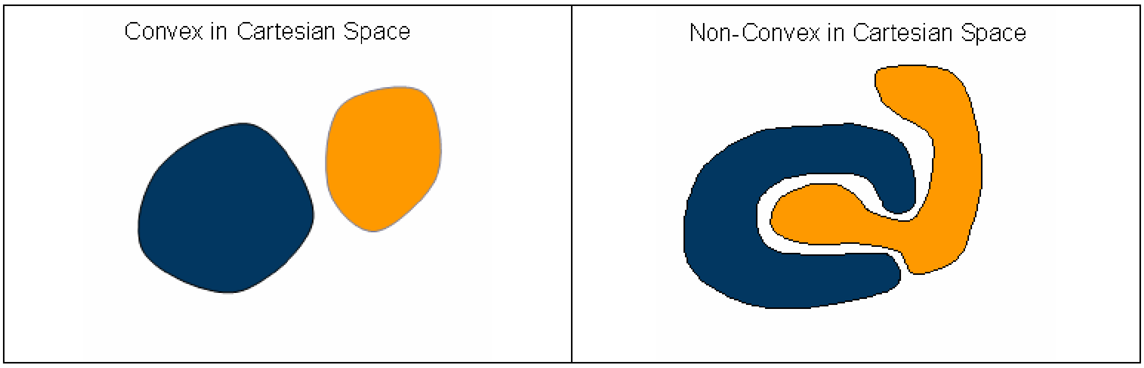
c) Versatility and Scope

- *Quantifies But Does Not Impose Spatiality:* The early literature on submarkets tended to be divided between those studies that adopted a spatial approach to submarkets (Straszheim (1975), Ball and Kirwan (1977), Palm (1978), Sonstelie and Portney (1980), Gabriel (1984)) and those that cluster dwellings by attributes or some other non-spatial criterion (Dale-Johnson (1982), Bajic (1985), Rothenberg *et al* (1991); see review by Watkins 2001). Most recent studies, however, acknowledge that there are both spatial and non-spatial drivers of submarkets and so some form of joint estimation is used (Goodman (1981), Adair *et al* (1996), Maclennan and Tu (1996), Bourassa *et al* (1997), 1999, 2003, 2007; Leishman 2009). Yet, while acknowledging the role of distance, none of these studies explicitly calibrate it. For example, at what distance does proximity cease to have a significant role in binding a property to a particular submarket? We seek a methodology that would allow us not only to permit spatiality in submarkets but also to quantify it.
- *Reveals but Does Not Impose Convexity/Non-Convexity:* The *shape* of submarkets – whether there are sound theoretical reasons to expect market areas to have particular shapes – has attracted little attention among housing researchers. This contrasts with the literature on theory of the firm, where the costs of transporting goods to and from the point of production to the point of consumption, leads one to expect that "there would be forces at work to minimize total transportation costs" (Puu 2003, p.104) which, in turn, creates a tendency to converge to some optimal market shape. Lösch 1940, for example, argued that the optimal shape of a market area for a single isolated firm would be circular. When there are many firms, the optimal shape of an individual market area is determined by a complex set subdivisions, of which the hexagon is the most compact

optimal shape under a variety of conditions. And Christaller 1933 found some empirical support for the hexagonal market area in his study of firms in Southern Germany.

It is beyond the scope of this paper to posit a general theory of optimal submarket shape but we can say that there are good reasons, such as such as Schelling's (1971) seminal chequerboard model, that might lead us to believe that the sorting of households across space will yield regular patterns (see review by Meen and Meen 2003). Schelling type processes, for example, might lead us to expect compact/convex shapes, but there may be other factors (the cumulative history of residential planning decisions, access to employment, schooling, local transport and amenities) that lead to elongated and non-convex submarket shapes. Therefore, the degree of compactness and convexity of a city's submarkets may tell us something about the potency of social sorting mechanisms relative to other forces that might mould more idiosyncratic spatial forms. Note also the implications of heterogeneous preferences – if there is a significant minority who are indifferent to, or actually prefer, racial and social mix, then the implications for boundary length are more ambiguous because there could be a large number of households that are happy to live along submarket boundaries, diluting the tendency for market forces to minimise the boundary length. Note also that the concentric rings of the access-space model are highly non-convex. So the more compact and convex are a city's submarkets, the more that city conforms to the Maclennan (1982) view of the world where large cities are "as much characterisd by residential sectors as they were by residential rings" (Maclennan, 1982, p. 23). This is because "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." (*ibid*). As the city develops, one might expect, therefore, the city to be made up by a patchwork of residential enclaves, each with its own core and periphery.

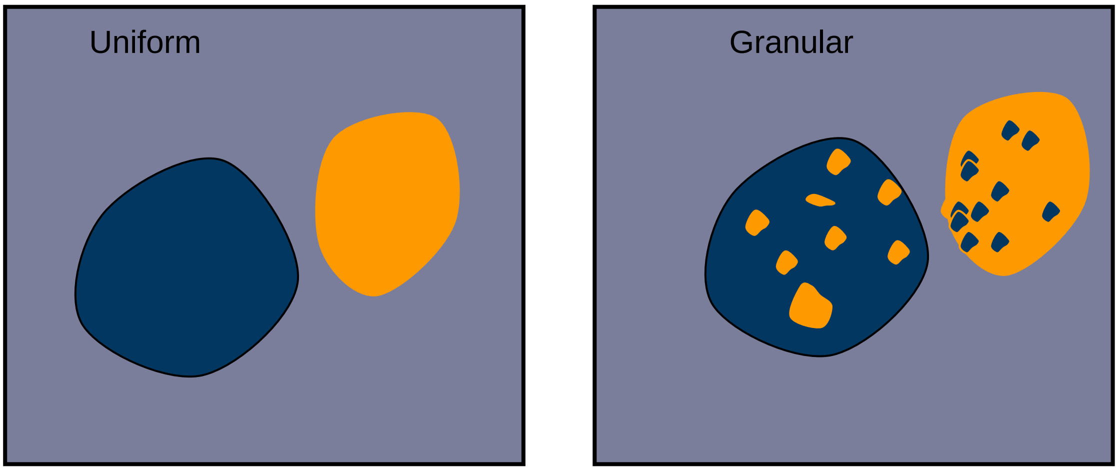
Figure 1 Submarkets as Convex and Non-Convex Sets in Cartesian Space



These possibilities challenge the convexity restriction imposed in Clapp and Wang (2007). We therefore seek a method that allows us to test for and measure convexity, rather than imposing it. Note further that the use of administrative boundaries to approximate submarkets is particularly problematic here because they are often drawn to run down the centre of roads, railways, or rivers (Myers 2004, Clapp and Wang 2007), each of which are sources of amenity and have the potential to form the locus around which submarkets tend to form. Given the topological nature of these physical features, administrative boundaries derived from them could systematically impose erroneous non-compactness and non-convexity.

- *Permits Granularity* Another implication of the Schelling model is that, other things being equal, one might expect submarkets to be relatively uniform. Granularity occurs when individual dwellings (or pockets of dwellings) within a submarket area, are not close substitutes for the majority of dwellings in that submarket. The extent to which submarkets are uniform or granular is itself of interest because it may important aspects of the spatial adjustment process driving how households self-organise. Does the level of granularity vary between cities? And if so, why? To what extent does the granularity within a submarket reflect the random market processes (aberrations that occur as a result of stochastic processes) and to what extent does non-uniformity reflect the intentional design of planners – deliberate decisions to build pockets of affordable housing, for example, or the effect of building a series of small parks or other local amenities placed at carefully chosen locations across an established submarket? The entropy measure of Gonzalez et al (2004 p. 466) allows one to distinguish between granularity of a regular geometric kind (the type that might follow a grid pattern of roads and amenities, for example) and that of a more random, arbitrary nature. If the cumulative sum of planning decisions over time intentionally or unintentionally convincingly mimics a random process (cf Mayo and Sheppard's "stochastic planning controls") then the entropy measure will fail to distinguish between market and planned granularity in price dynamics. Similarly, if market processes, such as those identified by Schelling, lead to regular spatial patterns, both within and without individual submarkets, then the entropy measure will enable us to divine the difference between planned and self-organised spatial patterns. The power of the entropy measure is therefore greatly enhanced by knowledge of the planning system in a particular city. If one knows, for example, that a submarket (or city) has developed entirely without planning, and yet has a low entropy value, then this might be taken to confirm the spatial regularity of market forces. As with preceding measures listed here, the entropy measure could also be calculated for **M** to gauge the regularity of the entire urban system. It could also be applied to the web of identified submarket boundaries, as a way of measuring the degree to which its structure is random rather than regular.

Figure 2 Uniform and Granular Submarkets



- *Useful for Policy Analysis:* One of the most important implications of housing submarkets for planners and policy makers is asymmetric price adjustment to shocks and interventions. If estimating the submarket structure of a city will help predict the different price effects of new construction (Pryce 2005; Bramley and Leishman 2005), tax changes (Berry et al 2003), urban regeneration initiatives (Bates 2006)), or exogenous shocks such as natural hazards (Pryce and Chen 2009), the usefulness of submarkets becomes immediately apparent. It is not clear how well the LOP/HAPV approach helps to predict responses to shocks. Given the problems noted above, LOP/HAPV may do a poor job of identifying the underlying spatial structure of dwelling substitutability in a city and hence be suboptimal way of predicting spatial asymmetries in price responses.

- *Can be used to Explore the Causes of Submarkets:* Returning to the opening question of this paper, how are market segmentation and social segmentation related? Using the LOP/HAPV approach, this can only be explored in a somewhat cumbersome way. One could test for a structural break across areas with different mixes of ethnicity, income etc but this does not yield a quantifiable measure of the contribution of these variables to a dwelling being allocated to a particular submarket. One could include ethnicity and race variables in the hedonic price equation but this leads to identification problems. Where dwelling units vary by the quantity of housing offered, the effect of income and social characteristics on selling price (= expenditure = price × quantity) may simply be capturing the impact on quantity demanded. Ideally, one would like a measure that gauges the substitutability of any two dwellings. This measure would then be amenable to further analysis – such as regression estimation that explores the extent to which differences in the social characteristics between the localities of the two dwellings determines their substitutability.

2) Deriving a Substitutability Approach to Submarkets

In the early work on submarkets by Rapkin (1953) and Grigsby (1963), was the concept of submarkets was drawn directly from the notion of substitutability:

> "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) argued 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 for the demanders of housing space" (Rapkin et al, 1953, p. 10 quoted in Grigsby op cit).

It has been difficult, however, to operationalise a substitutability based approach. The drift towards LOP/HAPV has occurred partly due empirical convenience. Cross Price Elasticity of Demand measures are unwieldy because of the problem of measuring housing demand. Neither is the empirical elusiveness of substitutability solved by applying price-band measures (as in the Rothenburg et al 1991 study) – one still faces the difficulty of separating out quantity effects from submarket effects, and the method becomes vulnerable to the instability of hedonic parameters across time and space. If one groups properties by attribute type (as in Goodman and Thibideau (1998), Maclennan and Tu (1996) and others[5]), one faces theoretical inconsistency due to the fact that the clustering occurs without reference to market behaviour – as noted above, one ends up with product groups are being defined, either arbitrarily through a pre-determined algorithm, or according to the judgement of the researcher.[6]

The alternative to price-band and attribute clustering suggested below exploits the dynamic nature of the market and make use of relationships between price changes (rather than price levels). Essentially, we propose using the Cross Price Elasticity of Price (CPEP) as a proxy for the Cross Price Elasticity of Demand (CPED) and hence of substitutability.

Proposition: *If demand and supply curves slope downwards and upwards respectively, the cross price elasticity of price will have a strictly positive, one to one, relationship with the cross price elasticity of demand.*

Consider the following equilibrium condition in the market for dwelling type *i*:

$$Q_{Si}(p_i, \boldsymbol{W}) - Q_{Di}(p_i, p_j, \boldsymbol{Z}) = 0 \qquad [1]$$

[5] Researchers typically apply principle component, cluster or factor analysis to bunch properties into product groups on the basis of physical characteristics. Dwellings within a particular group are viewed as substitutes. Hedonic price regressions are then run on each product group separately leading to improved regression fit and prediction accuracy. 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 submarkets2" (p. 160). Further examples include Dale-Johnson (1982) and Goodman and Thibodeau (1998).

[6] This is a general problem associated with cluster analysis and principle components – Greene (1993) in his classic econometrics text, for example, questions the usefulness of principle components because "the principle components are not chosen on the basis of any relationship of the regressors to *y*, the variable we are attempting to explain" (p.273).

where **Z** and **W** are vectors of exogenous factors affecting demand and supply respectively. By implicit differentiation of [1], the Cross Price Elasticity of Price is given by:

$$\eta_{p_i,p_j} = \left(\frac{dp_i}{dp_j}\right)\left(\frac{p_j}{p_i}\right) = \left(\frac{\partial Q_{Di} / \partial p_j}{\left(\partial Q_{Si} / \partial p_i\right) - \left(\partial Q_{Di} / \partial p_i\right)}\right)\left(\frac{p_j}{p_i}\right) \qquad [2]$$

Note that, provided all prices are positive (p_i, $p_j > 0$), the demand curve for i is downward sloping ($\partial Q_{Di}/\partial p_i < 0$), the supply is upward sloping ($\partial Q_{Si}/\partial p_i > 0$), and i and j are substitutes rather than complements ($\partial Q_{Di}/\partial p_j > 0$), then the CPEP will be positive.

We can also derive the CPED for good i (again by implicit differentiation of [1]):

$$\eta_{Q_{Di},p_j} = \left(\frac{dQ_{Di}}{dp_j}\right)\left(\frac{p_j}{Q_{Di}}\right) = \left(\frac{\partial Q_{Di}}{\partial p_j}\right)\left(\frac{p_j}{Q_{Di}}\right) \qquad [3]$$

Similarly, provided prices and quantity are positive (p_i, $Q_{Di} > 0$), demand slopes downwards, and i and j are substitutes rather than complements ($\partial Q_{Di}/\partial p_j > 0$), then the CPEP will be positive. Rearranging [3] in terms of the numerator partial derivative we get $\partial Q_{Di}/\partial p_j = (Q_{Di} / p_j)\ \eta_{QDi,pj}$. Substituting this expression into [2] we obtain CPEP as a function of CPED,

$$\eta_{p_i,p_j} = \theta . \eta_{Q_{D_i},p_j}$$

where,

$$\theta = \frac{Q_{Di} / p_i}{\left(\partial Q_{Si} / \partial p_i\right) - \left(\partial Q_{Di} / \partial p_i\right)}$$

The numerator will always be positive, as will the denominator so long as the demand and supply curves for dwelling i slope downward and upward respectively. It follows, therefore, that the CPEP will be monotonically increasing in the CPED,

$$\frac{d\eta_{p_i,p_j}}{d\eta_{Q_{D_i},p_j}} > 0.$$

Since the CPED is a measure of substitutability, it follows that CPEP can also be interpreted as a way of measuring substitutability. Crucially, however, CPEP does not require us to explicitly model demand, which as mentioned earlier, is worth avoiding in the case of housing where observed selling prices are actually a measure of expenditure (price $\times$ quantity) rather than the unit price.

Intuitively, the CPEP approach to substitutability can be understood as follows. Dwellings *i* and *j* are substitutes if a rise in the price of *j* leads to an increase in the demand for good *i*: CPED > 0. Conversely, if CPED < 0, then *i* and *j* are complements. Moreover, a rise in the price of *i* will cause a large increase in the demand for *j*, if *j* is a close substitute. If the supply of houses is less than perfectly elastic, the short run effect of the increase in demand for *j* will be an increase in the price of *i*. Other things being equal: $\uparrow p_j \Rightarrow \uparrow Q_{Di} \Rightarrow \uparrow p_i$.

For a given level of inelastic supply, the more closely two goods are considered by consumers to be substitutes, the more closely will their contemporaneous price changes be correlated. Correlation alone does not capture magnitude of effect, however. Changes over time in the price of *i* can be closely correlated with changes in the price of *j* even if the effect is very small. We can use γ_{ij} as approximation for CPEP, where γ_{ij} is the slope coefficient from a regression of π_{ti}, the proportionate change over time in the price of dwelling *i* , on π_j, the proportionate change over time in dwelling *j*:

$$\eta_{p_i,p_j} = \frac{dp_i / p_i}{dp_j / p_j} \approx \frac{\partial \pi_{ti}}{\partial \pi_{tj}} = \gamma_{ij} .$$

If $\gamma_{ij} > 0$, then *i* and *j* are substitutes. CPEP increases with the level of substitutability to the point where $\gamma_{ij} = 1$, which indicates that *i* and *j* are perfect substitutes and proportionate changes in the price of *i* are always matched by proportionate changes in the price of *j*. If $CPEP_{ij} < 0$ then *i* and *j* are complements. There is no obvious reason why $CPE_{ij} > 1$ should occur other than as a result of market friction. For example, the apparent contemporaneous overshoot of p_i in response to a change in p_j may in fact be the lagged response to changes in p_j from an earlier period, or it may simply reflect idiosyncrasies in the transactions process (such as extreme bids – see Levin and Pryce 2007 and Smith *et al* 2006), which can be counted as white noise. In the long run, and in the absence of market frictions, however, it is implausible that CPEP would be greater than unity, so $\max[E(\gamma_{ij})] = 1$.

3) Using CPEP to Understand the Existence and Spatiality of Submarkets

a) Existence

Consider the following inventory of housing market entities:

Sellers/existing residents:	$a_1, a_2, \ldots a_a \in A$
Buyers:	$b_1, b_2, \ldots b_b \in B$
Dwellings (or blocks of dwellings):	$i = 1, 2, \ldots V \in D$
Submarkets:	$S_1, S_2, \ldots S_s \subseteq \mathbf{M}$

M is the family of submarkets that make up the urban housing market as a whole. Each dwelling is an element of a submarket and of the wider housing market:

$$i \in S_k \Rightarrow i \in \mathbf{M}$$

If there are no submarkets, only a single uniform housing market, then $s = 1$ and,

$$S_1 = \mathbf{M} \qquad [1]$$

For $s > 1$, S_i is defined by some criterion that allocates dwellings to different subsets of **M**. It is assumed that this criterion leads to the partitioning of **M** so that it can be described as a composite of separate, but inter-connected, submarkets. **M** equals the union of all submarkets,

$$\mathbf{M} = \bigcup_{k}^{n} S_k = S_1 \cup S_2 \cup \ldots \cup S_n \qquad [2]$$

Because **M** is a partitioned set, all submarkets are disjoint, and a house cannot be a member of more than one submarket:

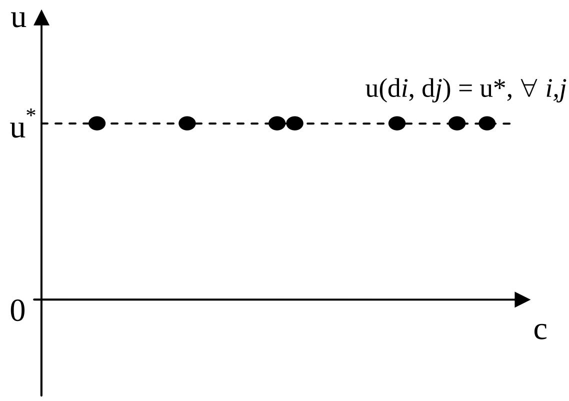
$$S_k \cap S_l = \varnothing \quad \forall\ k \neq l \qquad [3]$$

CPEP also leads to a natural test for the existence of submarkets. If CPEP = 1 for all pairs of dwellings, then all dwellings are perfect substitutes and there is no market segmentation:

$$\text{if} \quad S_1 = \mathbf{M} \qquad \text{then} \quad \max[\mathrm{E}(CPE_{ij})] = 1\ \forall\ i,j \ , \quad \text{where } i,j \in \mathbf{M}$$

In u, c space, where u is our measure of substitutability (in this case, $u = \mathrm{E}(CPE_{ij})$) and c is Cartesian distance between for each pair of dwellings i, j in M, we can represent the non-existence of submarkets by a horizontal scattering of points equal to u^*, where u^* the value of u representing perfect substitutability (in this case, $u^* = \max[\mathrm{E}(CPE_{ij})] = 1$). This scenario is depicted graphically in Figure 3.

Figure 3 Single Unified Housing Market (S_1 = M)



b) Spatiality

Submarket classification methods are often distinguished as being either *Spatial* or *Non-Spatial*. The latter can, in fact, be constituted as an aggregation of the former –non-spatial submarkets can be defined as a higher-level grouping of spatial submarkets. To illustrate, define $S_i \subseteq \mathbf{S}$ as representing a spatial submarket such that, each of the elements of S_i are contiguous in Cartesian space to at least one other S_i. Using a non-spatial criteria of defining submarkets leads to the disjoint grouping of spatial submarkets S_i into a smaller number of larger disjoint sets N_k,

Non-spatial submarkets: $N_1, N_2, \ldots N_\nu \subseteq \mathbf{M}$

where $\nu \leq s$. A particular non-spatial submarket N_k, merely groups together certain non-contiguous spatial submarkets, S_{ij},

$$N_k = \bigcup_{ij} S_{ij}, \forall ij \qquad \text{where } S_i \text{ and } S_j \text{ are non-contiguous subsets of } \mathbf{S}.$$

The urban housing market then constitutes the union of mutually exclusive non-spatial submarkets:

$$M = \bigcup_{k}^{\nu} N_k, \text{ where } N_i \cap N_k = \varnothing \quad \forall\ \mathrm{i} \neq \mathrm{k}$$

The methodological implication is that if one can first identify the set of spatial submarkets, one can always test whether any non-contiguous pair of spatial submarkets actually belong to a common non-spatial submarket, N_k. One should therefore start by identifying spatial submarkets even if one is ultimately interested in non-spatial submarkets. This rule breaks down, however, if all (or most) spatial submarkets are singletons (i.e. in a world where there is no spatial clustering by submarket) properties in different submarkets are randomly scattered across the urban plane.

It would be useful, therefore, to have an overall measure of the spatiality of the entire submarket system. Using our price-dynamic approach to measuring substitutability, a global indicator of spatiality for an urban area is given by gradient ϕ of the relationship between CPE_{ij} and the log of Euclidean distance c_{ij} between pairs of dwellings (i,j):

$\phi = \partial CPE_{ij} / \partial c_{ij}$

Substitutability is assumed to be approximately linear in logged distance: $CPE_{ij} = a + bc_{ij}$. If proximity is not an important aspect of substitutability, then one would expect CPE_{ij} to be unrelated to distance, resulting in a spherical scatter of u_{ij} (measured by ϕ) on c_{ij} as in Figure 2. On the other hand, if proximity is an important determinant of substitutability (due to access to the same amenities and dis-amenities, for example), then one would expect CPE_{ij} to decline with distance, most probably at a decreasing rate, illustrated in Figure 3.

Figure 4 Non-Spatial submarkets: "Spherical" Scatter of u_{ij} on c_{ij}

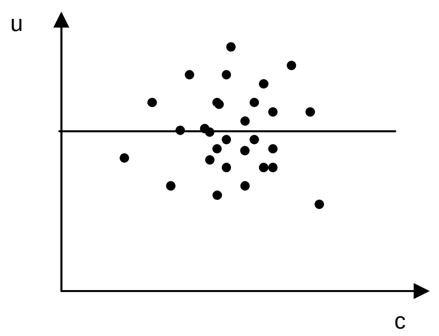
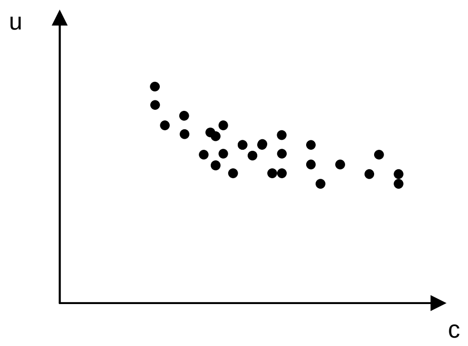


Figure 5 Spatial submarkets: Subsitutability u between d_i and d_j falls with distance c



Below are three hypotheses on why we might expect there to be a scatter, rather than a line, of points in Figure 3:

> <u>*A. Non-compact Submarket Shapes:*</u> Strong substitutability occurs between distant dwellings because of elongated and non-convex shapes of spatial submarkets. For example, a submarket could be the shape of long strip or crescent, following the path of a major road, railway or view. Points at the extreme ends of that submarket may be highly correlated but far apart. Therefore, the shape of submarkets is potentially important in conditioning the distance effect on substitutability.

B. Scattered clusters of substitutable bundles: There exist equivalent combinations of spatial amenities at different points in the city lead to distant clusters of dwellings being close substitutes.

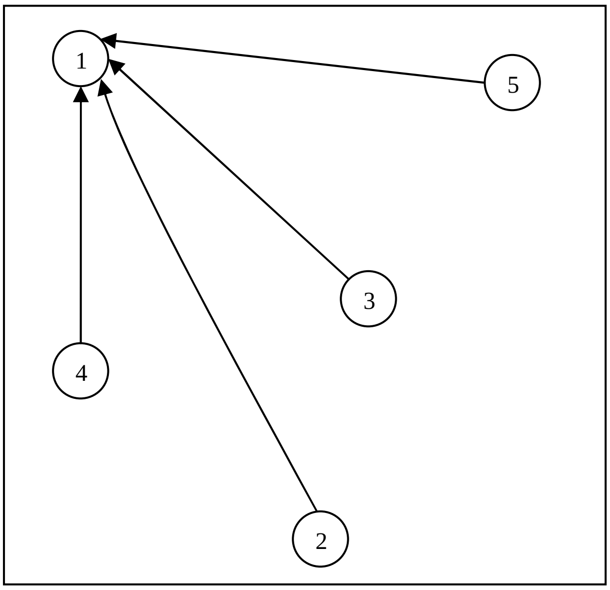
C. Coincidental correlations between distant dwellings: There is likely to exist non-causal (i.e. spurious) corresponding contemporaneous movements between distant pairs of inflation time series.

Hypothesis A motivates the case for exploring the shape of submarkets. *Hypothesis B* suggests that, once spatial submarkets have been defined, one should test whether non-contiguous segments are in fact part of the same submarket. *Hypothesis C* implies that one should include distance as an supplementary necessary condition for grouping dwellings into submarkets (that is, one should specify spatial submarkets before testing *Hypothesis B*). While these three hypotheses are mutually exclusive for a particular pair of dwellings, the same is not true not for the system as a whole which could exhibit all three properties simultaneously.

c) Categorising Dwellings into Non-Spatial Submarkets

If we find from applying the above analysis that distance plays no significant role in the substitutability of dwellings, how then can we categorise a set of dwellings into submarkets? The method proposed entails estimating a *Substitutability Lattice* for a given dwelling *i*. Consider a particular dwelling i = 1, 2, 3, 4, 5. Computing SL_1, the *Substitutability Lattice* for dwelling $i = 1$ would involve estimating CPEP_{1j} for all j not equal to 1. This is depicted in the digraph below (Figure 6) where each dwelling represents a node and each CPEP_{1j} represents an edge.

Figure 6 Digraph for a First Order Substitution Lattice



We can plot a *First Order Substitution Lattice* in Cartesian space, were we can think of it as a map of the substitutability of all dwellings in the city relative to dwelling 1. The simplest categorisation based on CPEP would therefore be to cluster our estimate of all bilateral values of γ with respect to a single dwelling, in this case i = 1. We label this *First Order Categorisation* and is defined as follows:

$$S_1, S_2, \ldots S_s \subseteq \mathbf{M} = \{i: \text{cluster}(\gamma_{1j}) \text{ where } \gamma_{1j} \in SL_1\}$$

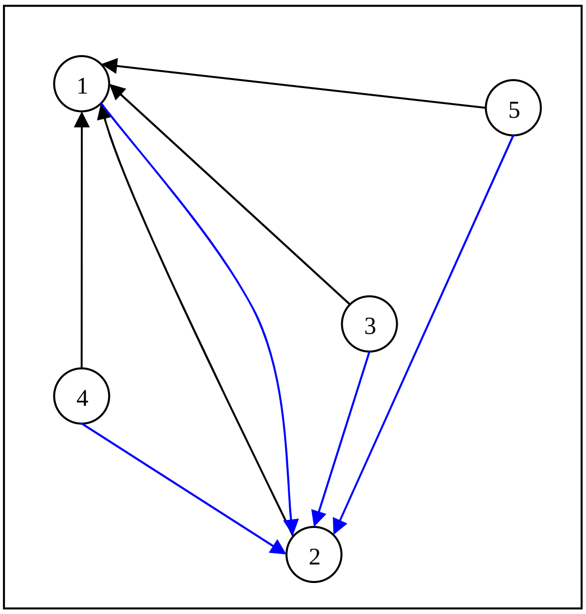
which is essentially a matter of identifying contour lines of substitutability with respect to dwelling i. If I plotted this lattice for my own dwelling, it would tell me immediately where across the city are the dwellings considered closest substitutes to my home.

Second Order Categorisation entails clustering according to a *Second Order Substitution Lattice*, comprised of two *First Order Substitution Lattices*, one for dwelling 1, and one for a second dwelling, i = 2:

$$S_1, S_2, \ldots S_s \subseteq \mathbf{M} = \{i: \text{cluster}(\gamma_{1j}, \gamma_{2j}) \text{ where } \gamma_{1j} \in SL_1 \text{ and } \gamma_{2j} \in SL_2\}$$

A *Second Order Substitution Lattice* is illustrated as a digraph below (). We might choose randomly the dwelling that constitutes the basis for SL_2 , or we might be more judicious and select a dwelling that is not a close substitute to dwelling 1. This would offer a means of triangulating our results.

Figure 7 Digraph for a Second Order Substitution Lattice



Third Order Categorisation would involve clustering according to a *Third Order Substitution Lattices* (cluster($\gamma_{1j}, \gamma_{2j}, \gamma_{3j}$)), and so on.

d) *Explaining Substitutability and Submarkets*

$CPEP_{ij}$ offers a potentially a powerful way to estimate the determination of substitutability and submarkets. Once computed, it could be incorporated as the dependent variable in a regression where the explanatory variables are the differences in various location and social attributes between *i* and *j*. That is, we could estimate a regression of the form:

$$CPE_{ij} = f(RD_{ij}, ID_{ij}, ED_{ij}, CD_{ij}, AD_{ij}) \text{ for all } i \text{ no equal to } j$$

where,

RD_{ij} = difference in the racial profile of the locality of *i* and the locality of *j*

ID_{ij} = difference in the income profile of the locality of *i* and the locality of *j*

ED_{ij} = difference in education profile of the locality of *i* and the locality of *j*

CD_{ij} = difference in the crime profile of *i* and *j*

ADij = difference in the age profile of *i* and *j*

4) Empirical Application

Work on the empirical application is still in progress. We present our initial findings below.

a) Data

Data are based on 34,120 dwelling transactions for 10,057 blocks of dwellings in a 30km radius of the centre of Glasgow, Scotland, supplied by GSPC, a consortium of over 200 realtors (for background information on the Scottish housing market and selling process see Pryce and Gibb 2006, Smith et al 2006, Levin and Pryce 2007). The period of the data (1999 to 2007) is one of a booming market. Summary statistics are provided in Table 1.

Table 1 Descriptive Statistics

Continuous and Count Variables:

variable	mean	sd	p50	min	max	N
sellingp	100708.60	67903.95	85000.00	4000.00	1975000.00	34,120
bedrooms	2.11	0.87	2.00	0.00	11.00	34,120
publicro	1.23	0.51	1.00	0.00	15.00	34,110
nbathrms	1.00	0.05	1.00	1.00	4.00	34,120
deprivtn	4.84	1.68	4.15	2.35	13.29	34,120
nearneighb1	3.09	7.60	0.84	0.00	445.00	34,120
footprint	158.72	253.21	127.00	21.00	6007.00	34,120
elevation	34.41	23.75	28.00	0.00	200.00	34,120
a_roads	426.12	372.97	330.17	1.45	2415.42	34,120
b_roads	871.22	743.47	640.24	3.39	3199.66	34,120
coastline	2645.22	1971.87	2184.74	2.63	10641.88	34,120
dlua	662.57	491.28	538.58	0.27	2468.20	34,120
lakes	4034.28	1570.22	3958.52	92.45	7487.73	34,120
motorways	2087.23	1323.39	1898.02	25.63	8870.33	34,120
rail_stati	767.87	518.10	637.96	21.95	5003.94	34,120
rivers_lar	6137.80	3391.86	5961.36	98.78	13640.43	34,120
rivers_med	1988.40	1439.43	1650.62	0.34	6300.25	34,120
rivers_sma	1914.40	1424.84	1624.25	0.06	5366.01	34,120
woodland	3382.06	1363.05	3601.16	5.17	6971.49	34,120

Binary Variables:

variable	mean	sd	N
hous_all	0.263	0.440	34,120
gch_d	0.607	0.488	34,120
flt1st_d	0.121	0.326	34,120
flt2nd_d	0.095	0.294	34,120
flt3rd_d	0.012	0.109	34,120
fltlwr_d	0.031	0.173	34,120
fltmdr_d	0.013	0.113	34,120
fltupp_d	0.039	0.194	34,120
convsn_d	0.016	0.124	34,120
bundet_d	0.018	0.134	34,120
bunsd_d	0.016	0.127	34,120
bunter_d	0.001	0.038	34,120
vildet_d	0.032	0.176	34,120
vilsd_d	0.128	0.334	34,120
othcot_d	0.002	0.046	34,120
garden_d	0.630	0.483	34,120
garage_d	0.220	0.414	34,120
needsupg	0.009	0.093	34,120
spacious	0.227	0.419	34,120
alarm	0.039	0.194	34,120
mature	0.020	0.140	34,120
bay	0.240	0.427	34,120
victrian	0.004	0.063	34,120
luxury	0.027	0.161	34,120
stone	0.149	0.356	34,120
trad	0.161	0.367	34,120
parking	0.111	0.314	34,120
ensuite	0.035	0.185	34,120
views	0.047	0.211	34,120
conservy	0.016	0.126	34,120
driveway	0.044	0.205	34,120

b) Estimating Price Dynamics for Each Postal Unit

I extend the model of Fik *et al* (2003)[7] to create a series of Time Location Value Signatures – essentially a set of inflation surfaces for Glasgow, one for each time period. These surfaces are constructed as follows:

(1) Estimate a Third Order Taylor Series approximation of the house price surface separately for each year. In the parlance of Fik et al, these would be called Location Value Signature (LVS, Fik *et al* 2003) – a constant quantity price surface – for each year using a flexible functional form that includes x, y interactions with attributes, quarter dummies, and area dummies based on a priori information on where the

[7] Similarly, Clapp and Wang (2006) "control for large and medium scale variation with a polynomial lattitude and longitude and spatial dummy variables".

shifts in the price surface may lie.[8] These include realtor jurisdictions, and local authority areas (property taxes and service provision vary by local authority). Insignificant variables and dummies are eliminated using a general-to-specific refinement procedure. Note that the LVS was estimated independently for each year, allowing coefficients to vary over time. Coefficients on attributes are allowed to vary over space through interactions with x,y coordinates and with area dummies. I use this method to control for the mix of properties coming onto the market rather than to test for HAPVs.

(2) Construct an inflation surface for each intervening time period by calculating the vertical distance between each LVS as a proportion of the base period value in each case.

Having created a series of surfaces of annual inflation (one for each quarter since 2000 q1), it was possible to extract a time series of the estimated constant quality price inflation series for any point in the geographical space covered by the model (i.e. Glasgow). An infinite number of points could have been chosen, either randomly, or at points along a regular grid, or based on the location of actual dwellings (or blocks of dwellings). I opted for the latter of these three on the basis that we are most interested in the price dynamics of actual residential locations. Inflation time series were therefore created for each centroid of the 10,057 postcode units (blocks of around fifteen residences)[9] within 30km of Glasgow.

To illustrate the results of the TLVS regressions, we present the results for 2007 in Table 2. The R^2 results for all nine regressions are listed in Table 3.

[8] Bourassa and Hoesli (2007) find that the prediction accuracy of simple OLS hedonic regression with *a priori* submarket boundaries compares very favourably with geostatistical estimation; in fact, the “absolute errors are lower than those for the geostatistical models without the submarket dummies”.
[9] This represents a much higher spatial resolution for a city wide analysis than previous UK research. Watkins (200??), for example, uses postcode sectors which contain, on average, around 30,000?? dwellings.

Table 2 Summary Information on TLVS Regressions

Sample Regression Results: for 2007 (dependent variable = sellingp_ln):

Source	SS	df	MS			
				Number of obs	=	4,332
				F(54, 4277)	=	195.91
Model	710.5116	54	13.15762	Prob > F	=	0.000
Residual	287.2523	4277	0.067162	R-squared	=	0.712
				Adj R-squared	=	0.709
Total	997.7639	4331	0.230377	Root MSE	=	0.259

	Coefficient	SE	t	sig	CI lower	CI upper
_cons	-549.274	152.0371	-3.61	0.000	-847.345	-251.202
bay	0.09696	0.010198	9.51	0.000	0.076967	0.116953
bedrooms	0.195341	0.005713	34.19	0.000	0.184141	0.206541
bunsd_d	0.141379	0.031956	4.42	0.000	0.078728	0.20403
coastline	8.63E-05	1.01E-05	8.57	0.000	6.66E-05	0.000106
convsn_d	0.346679	0.028538	12.15	0.000	0.290731	0.402627
deprivtn	-0.01681	0.003805	-4.42	0.000	-0.02427	-0.00935
dlua	8.05E-05	1.18E-05	6.82	0.000	5.73E-05	0.000104
ensuite	0.10484	0.019053	5.5	0.000	0.067486	0.142193
footprint	5.27E-05	1.41E-05	3.74	0.000	2.51E-05	8.03E-05
garage_d	0.079364	0.012325	6.44	0.000	0.0552	0.103528
garden_d	0.04866	0.009608	5.06	0.000	0.029823	0.067496
gch_d	0.052703	0.008398	6.28	0.000	0.036238	0.069168
hous_all	0.194207	0.014164	13.71	0.000	0.166438	0.221976
lakes	4.22E-05	6.14E-06	6.87	0.000	3.02E-05	5.43E-05
luxury	0.144521	0.022944	6.3	0.000	0.099539	0.189502
motorways	0.000058	7.18E-06	8.08	0.000	4.39E-05	7.21E-05
nearneighb1	0.010732	0.001114	9.63	0.000	0.008547	0.012917
needsupg	-0.12113	0.032186	-3.76	0.000	-0.18423	-0.05803
parking	0.04006	0.013954	2.87	0.004	0.012704	0.067416
publicro	0.155838	0.009765	15.96	0.000	0.136694	0.174982
rail_stati	-4.3E-05	1.06E-05	-4.02	0.000	-6.3E-05	-2.2E-05
rivers_lar	-3E-05	1.08E-05	-2.81	0.005	-5.2E-05	-9.21E-06
rivers_sma	6.86E-05	7.00E-06	9.8	0.000	5.49E-05	8.23E-05
t2x2_AREA_~7	-1.85162	0.225045	-8.23	0.000	-2.29283	-1.41042
t2x2_AREA~11	-2.64046	0.616594	-4.28	0.000	-3.8493	-1.43162
t2x2y2_AR~12	0.204347	0.09529	2.14	0.032	0.017529	0.391165
t2x2y2_ARE~4	-9.07E-06	1.59E-06	-5.71	0.000	-1.2E-05	-5.95E-06
t2x2y2_ARE~6	0.068258	0.005029	13.57	0.000	0.058399	0.078116
t2xy_AREA_~5	0.013749	0.002062	6.67	0.000	0.009707	0.01779
t2xy_AREA_~6	-2.34781	0.172558	-13.61	0.000	-2.68612	-2.00951
t2xy_AREA_~7	1.436028	0.174004	8.25	0.000	1.09489	1.777165
t2xy_AREA~11	2.102381	0.491383	4.28	0.000	1.139015	3.065746
t2xy_AREA~12	-6.81129	3.160023	-2.16	0.031	-13.0066	-0.616
t2y2_AREA_~5	-0.00549	0.000805	-6.82	0.000	-0.00707	-0.00391
t2y2_AREA_~7	-0.27845	0.033632	-8.28	0.000	-0.34439	-0.21251
t2y2_AREA~11	-0.41851	0.097892	-4.28	0.000	-0.61043	-0.22659
t2y2_AREA~12	1.264735	0.584097	2.17	0.030	0.119602	2.409868
t3x3y	3.35E-06	6.57E-07	5.1	0.000	2.06E-06	4.64E-06
trad	0.081937	0.012542	6.53	0.000	0.057348	0.106526
tx2y2_ARE~10	-8.9E-05	2.96E-05	-3	0.003	-0.00015	-3.1E-05
tx2y2_AREA~3	-0.01705	0.004168	-4.09	0.000	-0.02522	-0.00888
tx2y2_AREA~8	-0.01919	0.001507	-12.73	0.000	-0.02214	-0.01623
txy_AREA_G~3	0.447413	0.117229	3.82	0.000	0.217583	0.677244
txy_AREA_G~6	19.73941	1.469123	13.44	0.000	16.85917	22.61966
txy_AREA_G~8	0.635544	0.041731	15.23	0.000	0.553729	0.717359
ty2_AREA_G~3	-0.05943	0.02144	-2.77	0.006	-0.10146	-0.0174
ty2_AREA_G~8	-0.11789	0.007516	-15.69	0.000	-0.13262	-0.10315
woodland	2.86E-05	7.51E-06	3.81	0.000	1.39E-05	4.33E-05
x	-3.49836	0.666022	-5.25	0.000	-4.80411	-2.19261
x2_AREA_G~12	632.9326	284.2068	2.23	0.026	75.73978	1190.125
x2y2_AREA_~6	-4.82695	0.36425	-13.25	0.000	-5.54107	-4.11283
x2y2_AREA~12	-14.104	6.333576	-2.23	0.026	-26.5211	-1.68687
y2	38.43204	10.33561	3.72	0.000	18.16888	58.69521
y3	-3.84566	1.036157	-3.71	0.000	-5.87707	-1.81426

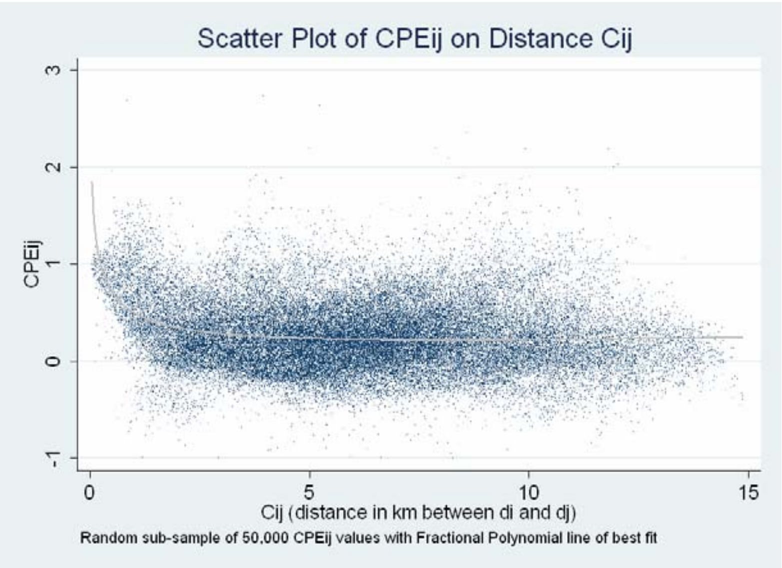
Table 3 R^2 Results for each TVLS:

R^2 1999 = .72937087
R^2 2000 = .72922658
R^2 2001 = .76130094
R^2 2002 = .70703397
R^2 2003 = .63359069
R^2 2004 = .58362324
R^2 2005 = .60579499
R^2 2006 = .63729338
R^2 2007 = .70846909

c) Existence and Spatiality of Housing Submarkets

Calculating $\theta = \partial CPE_{ij} / \partial c_{ij}$ is not a trivial exercise. If there are 10,057 blocks of dwellings, then there are 10,057 x 10,057 potential correlations between inflation time series, and 10,057 x 10,057 distances to be calculated. Ignoring correlations/distances from *i* to itself, and those correlations/distances from *i* to *j* when c_{ij} has already been calculated, leaves around fifty million pairs of dwelling units, (*i*,*j*), for which we need to compute CPE_{ij} and c_{ij}. CPE_{ij} each of these pairs was calculated as the slope coefficient from regression of π_{ti} on π_{tj} where π_{ti} is the annual constant quality price inflation time series for *i*.

Figure 8 Scatter plot of β coefficients from fifty million regressions of CPE_{ij} on c_{ij}



The values of CPE_{ij} (our measure of sustainability) and c_{ij} (distance) are plotted against each other for c_{ij} < 8km in Figure 7 (beyond 8km the slope of the line of best fit becomes horizontal).

The graph indicates that not all dwelling units are perfect substitutes (the values of CPEP do not lie along the horizontal line of unity as in Figure 1), neither is the substitutability of dwellings entirely aspatial (indicated by a spherical scatter plot, as in Figure 2). There is evidence, therefore, for spatial submarkets: the value of gamma for the system as a whole is negative,

$$\phi = \partial CPE_{ij} / \partial c_{ij} = -.0176 \qquad \text{(Robust CI = [-.0177, -.0175] } R^2 = 0.023\text{, n = 6592000)}$$

However, the effect of distance declines rapidly as depicted by the non-linear shape of the graph. When we run a regression of CPE_{ij} on c_{ij} for c_{ij} < 8km we find that

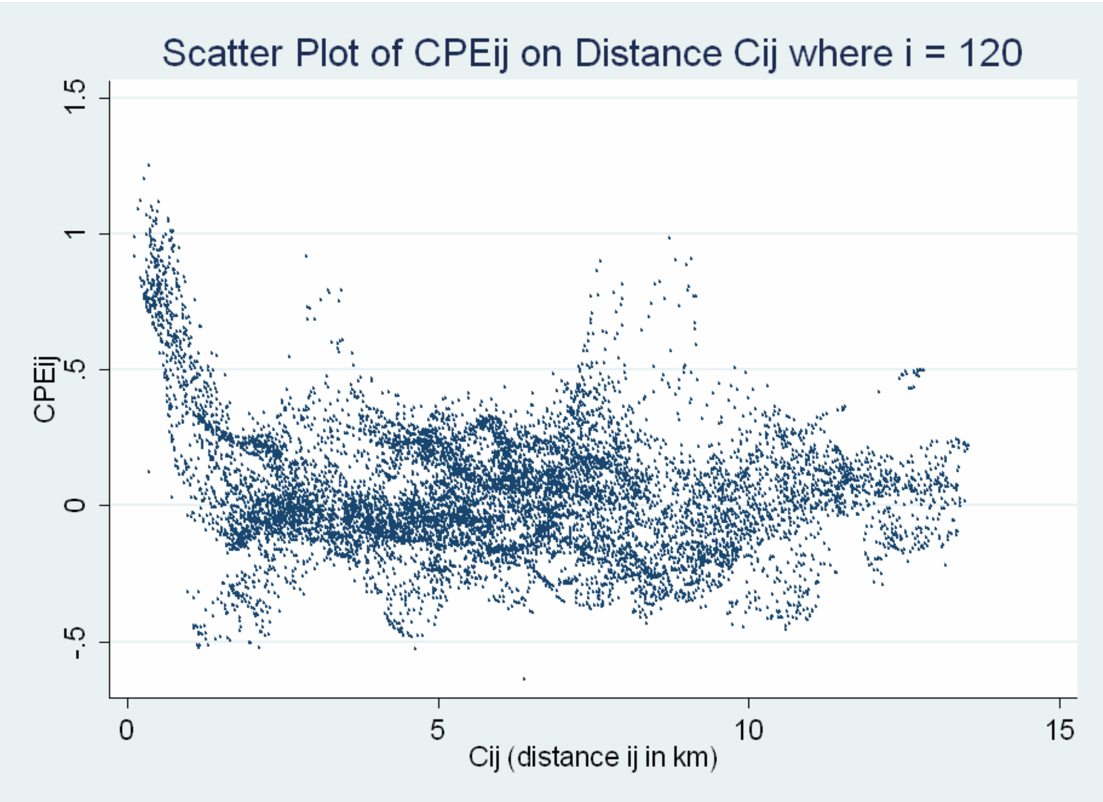
```
Number of obs = 6592000       R-squared     =  0.0805
-----------------------------------------------------------------------------
           b   |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------+-------------------------------------------------------------
   0 to <1 km  |  -.4426043   .0025781  -171.68   0.000    -.4476572   -.4375513
   1 to <2 km  |  -.3492354   .0009798  -356.44   0.000    -.3511558    -.347315
   2 to <4 km  |  -.0105645   .0001758   -60.10   0.000    -.0109091     -.01022
   >=4 km      |  -.0030142   .0000687   -43.90   0.000    -.0031487   -.0028796
       _cons   |   1.057743   .0021665   488.23   0.000     1.053497     1.06199
-----------------------------------------------------------------------------
```

from zero to <1km, $\phi = -.44$. That is, for every 1km increase in distance between dwellings, the cross price elasticity, CPE, falls by 0.44. From 1km to <2km, $\phi = -.35$. From 2 to <4km, $\phi = -.01$. Beyond 4km the distance effect on substitutability becomes negligible.

Nevertheless, it is clear from the very low R^2 values associated with Figure 7 that the substitutability between postcodes has a large non-spatial component – at least in terms of the simple distance measure: 95% of the variation in CPEP is due to factors other than Euclidean distance. Verification that there exist in our data *some* pairs of dwelling units where there appears to be a tenuous connection between substitutability (as measured by CPEP) and distance. This provides an imperative to further explore *Hypotheses A* (the existence of non-compact submarkets), *B* (the existence of non-contiguous submarkets) and *C* (the need for distance as a supplementary necessary condition for clustering).

2.2 *CPE-Distance Plots for Individual Dwellings:*

Existence of Spatial idiosynchresies is reinforced when one plots CPEP-Distance for an individual dwelling. The figure below shows evidence of a small cluster at around 3km with CPEP > 0.5.



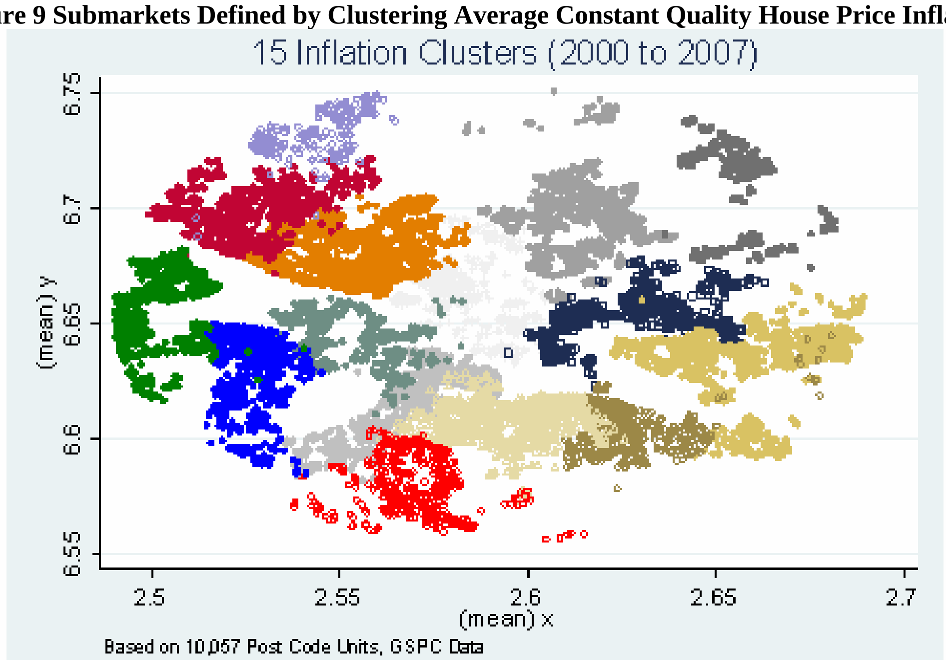
For i = 110 there appears to be a ridge of points between 4 and 7 km with similar levels of substituability with $d_{i=110}$. True for i = 100.

d) Defining spatial submarkets

As a precursor to lattice based categorisation, we take the following three exploratory steps based on the TVLS estimation to delineate submarkets: (1) summarise the time series of inflation for each *i*; (2) capture the distance effect if our estimates of γ above imply a *de facto* spatial dimension to substitutability (which they do); (3) decide on a cluster method.

With regard to (1), the simplest approach is to calculate the average rate of constant quantity house price inflation, π_i^*, for all time periods for each *i*. This is the approach presented below as an approximation of clustering by CPEP for a representative dwelling (which is the next step of the paper). Regarding (2) and (3), d_i are clustered using Ward's method applied to the 10,057 postcode units using latitude, longitude, and π_i^* the estimated rate of constant quality house price inflation. Cluster analysis does not yield a unique number of clusters, though inspection of the dendrogram suggests that 15 groups would be a sensible cut point. Colour coded centroids of these clusters are plotted in Figure 8.

Figure 9 Submarkets Defined by Clustering Average Constant Quality House Price Inflation



The standard deviation of average annual house price inflation for the period 2000 to 2007 is 21% of the mean for the submarket system as a whole. Note that the variation within submarkets varies hugely. The coefficient of variation in submarket S_4, for example, is 6% of the mean, compared to 23% in S_1. 76 blocks of dwellings (postcodes) were allocated to very small submarkets (less than 70) and so were omitted from the calculations.

S_k	Mean	SD	Coefficient of Variation	Min	Max	N
1	0.066	0.015	0.233	0.027	0.104	735
2	0.082	0.017	0.211	0.013	0.125	764
3	0.102	0.021	0.209	0.030	0.136	284
4	0.119	0.008	0.064	0.095	0.150	677
5	0.102	0.012	0.114	0.056	0.153	861
6	0.078	0.012	0.158	0.012	0.128	453
7	0.088	0.009	0.106	0.053	0.106	769
8	0.097	0.009	0.091	0.066	0.129	950
9	0.119	0.010	0.085	0.089	0.149	408
10	0.089	0.008	0.088	0.056	0.124	1,404
11	0.099	0.008	0.083	0.067	0.131	773
12	0.076	0.011	0.148	0.023	0.098	216
13	0.123	0.011	0.088	0.096	0.159	476
14	0.104	0.009	0.084	0.089	0.150	619
15	0.095	0.012	0.124	0.063	0.123	592
All Submarkets	0.095	0.019	0.195	0.012	0.159	9,981

5) Conclusion

There is more to submarkets than improving R^2. Submarkets provide a potentially powerful framework for understanding and modelling important policy and theoretical issues of urban form, social mix, racial contiguity, amenity access, externality and asymmetric price adjustment. The first goal of this paper was to establish what we would require of a methodology for it to facilitate the analysis of these topics. The received Law of One Price/hedonic approach, while appropriate for optimising mass appraisal accuracy, when evaluated by these broader set of criteria, did not fair well as a means of identifying the submarket system of a city or region. Whilst important improvements have been made in recent years in the sophistication and precision of hedonic regression prediction, this has not led to innovation in methods or theory that help us understand the wider implications of submarkets.

The second goal of this paper has been to develop an alternative approach to estimating submarkets, one that is grounded in the notion of substitutability as the defining concept of submarket analysis, and one that will potentially lead to a richer set of measures of the nature of submarkets.

Evaluating the CPEP approach using the criteria listed in section one of the paper, we conclude the following:

- *Re Theoretical Robustness*: In principle, a CPEP based measure of substitutability provides a conceptually robust basis on which to derive submarkets, one that is robust to transformative interaction effects (dwellings are treated as inseparable entities), not dependent on discontinuity, and robust to inter-submarket migration.

- *Re Methodological Robustness:* Suppose prices are in a perpetual process of convergence towards equilibrium but never quite reaching it due to intermittent demand and supply shocks. Provided the price adjustment process follows the substitutability as we would expect from the above discussion of CPEP and CPED – i.e. prices of close substitutes are likely to adjust to repeated shocks in more similar ways than those of distant substitutes – the CPEP approach should be fairly robust to disequilbria. The approach is unaffected by attribute measurement errors and the weighting problem (since attribute prices are not used), unless attribute information is used to control for the mix of dwellings coming onto the market. CPEP is potentiall free from the problems of administrative boundaries since submarkets can be freely grouped across space.

- *Re Versatility and Scope:* more importantly, the CPEP approach has the potential to help quantify spatiality and non-convexity, and granularity. Because CPEP is grounded in the relative dynamics of dwellings it is also an idea method for identifying asymmetries in price response to exogenous shocks and policy interventions. Finally, it has the potential to explore the causes of submarkets.

It is hoped that this paper will stimulate a step-change in submarkets research and encourage more work on the integrates our understanding of social and market segmentation.

6) References

Andersson, R; Musterd, S; Galster, G., (2007) What mix matters? Exploring the relationships between individuals' incomes and different measures of their neighbourhood context, *Housing Studies*, 22, 637-660.

Bates, L.K. (2006) Does Neighborhood Really Matter? Comparing Historically Defined Neighborhood Boundaries with Housing Submarkets, *Journal of Planning Education and Research*, Vol. 26, No. 1, 5-17.

Berry, J., McGreal, S. Stevenson, S., Young, J. and Webb, J.R. (2003) Estimation of apartment submarkets in Dublin, Ireland, *Journal of Real Estate Research*

Blasius, J; Friedrichs, J; Galster, G. (2007) Introduction: Frontiers of quantifying neighbourhood effects, *Housing Studies*, 22, 627-636

Bourassa S.C., Hamelink F., Hoesli M., MacGregor B.D. (1999) Defining Housing Submarkets, *Journal of Housing Economics*, 8(2), pp. 160-183(24)

Bourassa, … "Ethnicity, endogeneity and housing tenure choice," *Journal of Real Estate Finance and Economics*, vol. 20, no. 3 (2000), pp. 323-341.

Bourassa, S.C., Cantoni, E. and Hoesli, M. (2007) Spatial Dependence, Housing Submarkets, and House Price Prediction, *The Journal of Real Estate Finance and Economics*, 35(2) 143-160.

Bourassa, S.C., Hoesli, M. and Peng, V. (2003) Do housing submarkets really matter?, *Journal of Housing Economics*, Volume 12, Issue 1, March 2003, Pages 12-28.

Bramley, G. and Leishman, C. (2005) Planning and housing supply in two-speed Britain: modelling local market outcomes, *Urban Studies*, 42(12), 2213-2244.

Chen, Z., Cho, S., Poudyal, N., Roberts, R. (2007) "Forecasting Housing Prices under Different Submarket Assumptions" paper presented at the American Agricultural Economics Association Annual Meeting, Portland, OR, July 29-August 1, 2007

Dale-Johnston (1982) An Alternative Approach to Market Segmentation using Hedonic Price Data, *Journal of Urban Economics*, 11, pp 311-332.

Fik, T.J., Ling, D.C., and G.F. Mulligan (2003) Modelling spatial variation in housing prices, *Real Estate Economics*, 31, 623-646

Fletcher, M., Gallimore, P. and Mangan, J. (2000) The modelling of housing submarkets, *Journal of Property Investment & Finance*, 18(4) 473 – 487.

Galster, G. C. (2007) Neighbourhood social mix as a goal of housing policy: a theoretical analysis, *European Journal of Housing Policy*, 7, pp. 19-43.

Galster, G. C. and Killen, S. (1995) The geography of metropolitan opportunity: a reconnaissance and conceptual framework, *Housing Policy Debate*, 6(1), pp. 7–44.

Galster, G. C., Quercia, R. G. and Cortes, A. (2000) Identifying neighbourhood thresholds: an empirical exploration, *Housing Policy Debate*, 11(3), pp. 701–732.

Gonzalez, R. C., Woods, R. E., and Eddins, S. L. (2004) *Digital Image Processing using Matlab*, Pearson, Prentice-Hall, New Jersey.

Goodhart, C. and Hofmann, B. (2008) House Prices, Money, Credit and the Macroeconomy", *Oxford Review of Economic Policy*, 24(1), 180-205.

Goodman (1981) Housing Submarkets within Urban Areas: Definitions and Evidence, *Journal of Regional Science*, 21, pp 175-185.

Goodman, A. 1981. Housing Submarkets Within Urban Areas: Definitions and Evidence. *Journal of Regional Science* 21: 175–185.

Goodman, A. and Thibodeau, T. G. (2007) The Spatial Proximity of Metropolitan Area Housing Submarkets, *Real Estate Economics*, Volume 35 Issue 2 Page 209-232

Goodman, A.C. and T.G. Thibodeau. 1998. Housing Market Segmentation. *Journal of Housing Economics* 7: 121–143.

Goodman, A.C. and T.G. Thibodeau. 2003. Housing Market Segmentation and Hedonic Prediction Accuracy. *Journal of Housing Economics* 12: 181–201.

Grigsby, W. 1963. *Housing Markets and Public Policy*. Philadelphia: University of Pennsylvania Press.

Jones, C., C. Leishman and C. Watkins. 2003. Structural Change in a Local Urban Housing Market. *Environment and Planning* A 35(7): 1315–1326.

Kaplan, D.H. (2009) An Analysis of the Relationship Between Housing Foreclosures, Lending Practices, and Neighborhood Ecology: Evidence from a Distressed County, *Professional Geographer*, 61, 101-120

Kearns, A; Mason, P (2007) Mixed tenure communities and neighbourhood quality, *Housing Studies*, 22, 661-691.

Lees, L. (2008) Gentrification and Social Mixing: Towards an Inclusive Urban Renaissance?, *Urban Studies*, Vol. 45, No. 12, 2449-2470

Lupton, R. and Tunstall, R. (2008) Neighbourhood regeneration through mixed communities: a 'social justice dilemma'? *Journal of Education Policy*, 23, 105-117.

Michaels, R.G. and V.K. Smith. 1990. Market Segmentation andValuing Amenities with Hedonic Models: The Case of HazardousWaste Sites. *Journal of Urban Economics* 28: 223–242.

Micheals & Smith (1990) Market Segmentation and Valuing Amenities with Hedonic Models: the Case of Hazardous Waste Sites, *Journal of Urban Economics*, 28, pp 223-242.

Ostendorf, W., Musterd, S., and De Vos, S. (2001) Social Mix and the Neighbourhood Effect. Policy Ambitions and Empirical Evidence, *Housing Studies*, 16, 3, 371-380.

Palm, R. 1978. Spatial Segmentation of the Urban Housing Market. *Economic Geography* 54: 210–221.

Pryce, G. (2004) *The Micro and Macro Effects of the Location of New Housing Supply*, London: Office of the Deputy Prime Minister.

Pryce, G. and Evans, G. (2007) Identifying Submarkets at the Sub-Regional Level in England, London: Department of Communities and Local Government.

Pryce, G. and Gibb, K. (2006) 'Submarket Dynamics of Time to Sale', *Real Estate Economics*, 34(3), pp.377-415.

Rapkin, C., L.Winnickand and D.M. Blank. 1953. *Housing Market Analysis.*Washington, DC: U.S. Housing and Home Finance Agency.

Rosen, S. (1974) Product differentiation in pure competition, *Journal of Political Economy*, 82(1):34–55.

Rothenberg, Galster, Butler and Pitkin (1991) *The Maze of Urban Housing Markets: Theory, Evidence, and Policy*, The University of Chicago Press, Chicago.

Sanders, A. (2008) The subprime crisis and its role in the financial crisis, *Journal of Housing Economics*, 17, 254-261.

Schnare, A. and R. Struyk. 1976. Segmentation in Urban Housing Markets. *Journal of Urban Economics* 3: 146–166.

Smith, S. J., Munro, M. and Christie, H. (2006) Performing (housing) markets, *Urban Studies*, 43, pp. 81–98

Straszheim, M. (1975) *An Econometric Analysis of the Urban Housing Market*, National Bureau of Economic Research, New York.

Watkins, C. 2001 The Definition and Identification of Housing Submarkets. *Environment and Planning A* 33: 2235–2253.