Housing Submarkets and the Lattice of Substitution

Gwilym Pryce

Urban Studies (School of Social and Political Sciences),
University of Glasgow, Glasgow G12 8RS, UK
Email: g.pryce@lbss.gla.ac.uk
Webpage: www.gpryce.com

Working Draft as at 27 September 2010

DRAFT – PLEASE DO NOT QUOTE WITHOUT PERMISSION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 their potential for improving prediction accuracy but there is more to submarkets than regression refinement. Submarkets are important because they reflect how housing markets respond to, and interact with, social and spatial processes at the local level. The paper attempts to establish a set of criteria that submarket methodologies should meet in order to investigate the nature and meaning of submarkets in a more robust and purposeful way. Existing approaches are critically evaluated using these criteria, and an alternative 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. The approach is illustrated using data on Glasgow.

JEL classification: R1

Keywords: Housing submarkets; Substitutability; Cross Price Elasticity; House prices; Hedonic; Cluster analysis.

Acknowledgements:

The author is grateful to the UK Department of Communities and Local Government for providing the initial funding, and to GSPC for unrestricted access to their house price data. 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 the responsibility of the author.

1. Introduction

How are market segmentation and social segmentation related? Issues of social cohesion and integration have become central to debates surrounding the nature of a good society (Lees 2008) and have shaped the fabric of policy innovation over the last two decades (Ostendorf *et al* 2001; Lupton and Tunstall 2008). This wider interest has stimulated, and been stimulated by, a considerable volume of theoretical and empirical work in the academic literature regarding the segmentation of communities and the related socio-economic issues of neighbourhood mix, neighbourhood effects, and neighbourhood dynamics (Kearns and Mason 2007; Andersson et al 2007; Galster 2007).

At the same time, research on local housing markets has flourished, and the economic importance of house prices has never been more apparent (Sanders 2008; Kaplan 2009; Goodhart and Hoffmann 2008). Given the likely interconnectedness 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 becomes all the more surprising when one considers the strong links made in the early submarkets literature (Grigsby *et al*, 1963 Rothenburg *et al* 1991).

Submarkets are a potentially powerful framework for conceptualising and calibrating the deep processes that lead to segmentation and shape fundamental structure of cities. Yet, very little attention has been paid in the recent cohorts of submarket papers (see the literature reviews in Watkins' 2001 and Baroussa et al 2007) to these sorts of theoretical possibilities. 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 2003, 2007). 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 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. Methodological innovations have tended to be confined to experimentation with new approaches to (i), (ii) or (iii).

There has, in summary, been a drift away from (a) theoretical research on submarkets and (b) research that integrates with neighbourhood and socio-economic analysis; and 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 Homogenous Attribute Price Vectors (HAPV) as the defining principle by which submarkets are defined and estimated. An important consequence is that submarket estimation methods may not have evolved along lines that are appropriate for some of their most important applications.

The goal of this paper is to open up the debate over submarkets by highlighting weaknesses in the existing HAPV consensus and to demonstrate that there is still much to be done in submarkets research. The paper aims to provide direction and structure to that research by illustrating how submarkets can be conceived within the context of "Substitution Lattices" that connect dwellings in the urban system, and by illustrating how this Lattice could be estimated using the Cross Price Elasticity of Price (CPEP).

The remainder of the paper is structured as follows. In Section 2, 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, leading to the conclusion that HAPV approaches to submarket delineation provide neither a necessary nor a sufficient condition for submarket delineation. An alternative method is offered in Section 3, grounded directly in the notion of substitutability (originally suggested as the basis of submarkets by the fathers of submarkets research, Rapkin *et al.* 1953 and Grigsby 1963). It is proposed that dwellings should be grouped on the basis of substitutability, measured using CPEP, which is shown to have a one-to-one mapping with the Cross Price Elasticity of Demand (CPED). Clustering by CPEP offers a necessary condition for identifying submarkets (any pair of dwellings in the same submarket will have a high CPEP), and it is argued that one is able to reduce the probability of Type II errors (i.e. allocating dwellings to the same submarket when in fact they belong to different ones) by including the geographical distance in the cluster function. Section 4 considers how CPEP could be used to explore the existence and spatiality of submarkets. Section 5 illustrates the method using real data (house sales in Glasgow, Scotland). Section 6 concludes.

2. What do we require from Submarket Estimation Methods?

Consider the following inventory of housing market entities and definitions:

Household:	$b = 1, 2, \dots B$
Dwellings (or blocks of dwellings):	d_i where $i = 1, 2, \dots V$
Attribute vector for dwelling i:	$\mathbf{z}_i = z_{(1)i}, z_{(2)i}, \dots z_{(A)i}$
Attribute price vector for dwelling i:	$\mathbf{P}(\mathbf{z}_i) = P_{(1)i}, P_{(2)\,i}, \dots P_{(A)i}$
Homogenous attribute price vectors for i and j:	$\mathbf{P}(\mathbf{z}_i) = \mathbf{P}(\mathbf{z}_j)$
Euclidean distance between i and j:	D_{ij}
Cross price elasticity of price (CPEP) between i and j:	η_{ij}

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. Where there is more than one submarket ($K > 1$), S_k is defined by some criterion that allocates dwellings to each subset of **M**. The criterion that has come to dominate the literature is that of Homogenous Attribute Price Vectors (HAPV), based on the law of one price. If HAPV is applied as a sufficient condition, it is implicitly assumed that any two dwellings *i* and *j* with the same attribute prices will belong to the same submarket. However, the argument put forward in this paper is that similarity of attribute prices does *not,* in fact, imply that two dwellings are in the same submarket—the equivalence may be coincidental or simply irrelevant to the existence of submarkets.

If HAPV is applied as a necessary condition, it is implicitly assumed that dwellings in the same submarket will have the same attribute prices:

$$i,j \in S_k \qquad \Rightarrow \qquad \mathbf{P}(\mathbf{z}_i) = \mathbf{P}(\mathbf{z}_j)$$

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

$$\mathbf{P}(\mathbf{z}_i) \neq \mathbf{P}(\mathbf{z}_j) \qquad \Rightarrow \qquad i,j \notin S_k$$

This too is problematic. As explained below, HAPV is unlikely to hold as a necessary condition because Transformative Interaction Effects (TIEs) and Many to Many Mappings of Means and Ends (MMMEs) make it perfectly possible to observe different attribute prices on two dwellings even though they belong to the same submarket. Homogenous attribute prices can be a false indicator of common membership of a submarket, and heterogeneous attribute prices can be a false indicator of market segmentation.

These arguments are now developed more fully, woven around a set of proposed criteria for evaluating submarket estimation methods (SEMs), namely: SEMs should be robust to *1. the contingent nature of attribute effects; 2. the continuity of substitutability space; 3. unobserved attribute variation; and 4. non-convexity, non-compactness, and non-contiguity.*

Criterion 1: SEMs should be Robust to the Contingent Nature of Attribute Effects

<u>Rationale</u>*: The existence of Many to Many Mapping of Means and Ends* (*MMMEs) and Transformative Interaction Effects (TIES) implies that properties can be in the same submarket but have different attribute prices.*

MMMEs exist because the same human need can be met in different ways, and the same housing bundle will meet different needs for different people. Consequently, two homes can have very few common attributes, and very divergent attribute prices, yet still be considered close substitutes. An important driver of MMMEs, and one that makes predicting the global set of MMMEs considerably more elusive, is the existence of *Transformative Interaction Effects* (TIEs). TIEs occur when the effect of an attribute (whether geographical or structural) is fundamentally transformed when placed in a particular context, such as being combined with another attribute (either geographical or structural).

To illustrate, consider the options for crossing the English Channel.[1] Suppose ferries, planes and trains are all close substitutes—a change in the price or availability of one will have a large effect on the demand for the others (they are all in the same market for transferring passengers across the English Channel).[2] While each mode of transport has contrasting physical features, the features combine to offer a similar *service* because of TIEs. This leads us to question the meaningfulness of relying on attribute price differences as a gauge of substitutability or, indeed, submarkets.

For example, whereas wings on trains are of no value, wings on aeroplanes are essential. Presumably, the attribute price of wings would differ dramatically between trains and planes because the utility of wings is transformed when appropriately combined with a jet engine

---

[1] Or *French* Channel, depending on one's perspective…

[2] 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.

and aerodynamic fuselage. Similarly, wheels on ferries will have little or negative value, but vital to the functioning of trains.

Note that, if we were to persist with the application of HAPV as our means of submarket delineation, the divergent attribute prices on wings and wheels across transport modes would lead us to the erroneous conclusion that the modes belong to separate submarkets. The error arises because, although the physical attributes of an aeroplane are quite different to those of a ferry, the *final service bundle* (comfortable and speedy travel) may be very similar.

TIEs of particular relevance to housing are those arising from the impact of geography. Location brings a long list of potential sources of utility—proximity to open space and environmental features, to sources of employment, to retail outlets, to crime, to social networks, to schooling, to religious and cultural centres—all of which can interact with heterogeneous sets of consumer preferences. Demand is shaped not only by the shelter that housing offers but also by fashion, lifestyle aspiration, and the social connotations of dwelling design. The effects of territoriality (Kintrea *et al.* 2008), stigma and area esteem are also potentially important, as are preferences for racial and social mix (Schelling 1971).

Decomposing dwelling prices into the prices of constituent parts is unhelpful when measuring substitutability—what matters is the final service bundle. Different dwellings id different locations can provide remarkably similar levels of utility to a particular consumer despite the structural differences. Empirical estimation is frustrated because the service bundle provided by a house and its location is so multi-faceted—MMMEs and TIEs may be subtle and unforeseen causing different combinations of physical and locations attributes to interact in different ways for different people. Whether attribute prices are observed as homogenous or not becomes irrelevant to the question of substitutability, in the same way that the price of wings on automobiles vs. price of wings on aeroplanes is irrelevant to the cross price elasticity of demand between the two transport modes.

<u>Corollary 1:</u> Cluster methods are theoretically problematic if applied to dwelling attributes (either structural or geographical) rather than behavioural variables.

<u>Rationale</u>: Properties can be in the same attribute cluster group but not be in the same submarket, and they can be in the same submarket but not be in the same attribute cluster group.

Statistical clustering of dwellings by physical attributes (and grouping of attribute variables into product bundles), has become popular in the submarkets literature (e.g. Maclennan and Tu 1996; Leishman 2009), but may not reflect how consumers group them, which may be nuanced and difficult to anticipate, not least because of the interplay between MMMEs and TIEs.

We would like to group properties according to a variable that captures market behaviour (e.g. substitutability), rather than the physical characteristics of the dwelling stock. In transactions data, the only behavioural variable actually measured is the selling price, but this is the dependent variable and is typically excluded from the clustering process. Similarly, grouping variables using factor analysis imposes a structure on the functional form of the hedonic equation which obscures how the market views the interplay of attributes in determining prices.[3]

Note that grouping properties by type of resident is also problematic: while two consumers may have different preference maps, this does not mean they will disagree on whether two properties are close substitutes—they may consider the dwellings to be substitutable but for different reasons. And even if we were able to anticipate individual rankings of substitutability, the outcome at the level of the market is made fundamentally

[3] Greene (1993) questions the usefulness of principal components because "the principal components are not chosen on the basis of any relationship of the regressors to *y*, the variable we are attempting to explain" (p.273).

unpredictable by the *Condorcet paradox*[4] – transitivities that hold at the level of the individual do not necessarily hold in the aggregate.

Ideally, one would like to cluster by market-level substitutability, but this requires a way of measuring the degree of substitution, which has hitherto proved elusive.

Corollary 2: *Spatially correlated errors provide neither a necessary nor a sufficient condition for submarkets*.

Rationale: *Because goods can be close substitutes but have different attributes, there is no guarantee 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 be coterminous with fissures in substitutability.*

Even if one were able to measure, without error, all the physical and amenity differences between dwellings, it remains possible that very different bundles of physical and location attributes could be perceived as close substitutes by consumers—it is the utility of the inseparable and idiosyncratic fusion of attributes that a buyer is purchasing, not the linear sum of components.

Spatially auto-correlated errors as a measure of substitutability may be further distorted by the arbitrary nature of uncaptured nonlinearities. For example, for dwelling i, a given functional form may do well at capturing them; not so for j. So, spatial clusters of errors may not reveal submarket boundaries but patterns of uncaptured nonlinearities which vary across space. Further difficulties arise from the non-monotonic and variable effects of distance in different directions from a given dwelling due to non-convexity, granularity and non-compactness in the shape of submarkets, and the spatial clustering of attribute measurement errors (discussed below).

[4] See, for example, Arrow's (1950) application to the majority voting problem.

Criterion 2: SEMs should be Robust to the Continuity of Substitutability Space

Rationale: *Submarkets can happily exist along a frictionless continuum and so continuous differentiability of the substitution function linking i and j in Cartesian space does not preclude i and j from belonging to separate submarkets.*

Goodman and Thibodeau (1998) argue that attribute price differences persist because of amenity effects. However, this suggests that heterogeneous attribute prices (and hence submarkets) only exist because of omitted amenity variables. In principle, therefore, if one were to construct a model of an efficient market that captures amenity affects (both social and economic), there would be no difference across dwellings in the marginal prices of the structural and locational attributes and their interactions; and no such thing as submarkets!

Essentially, we are asking *Why disentangle location and structural attributes for the purposes of submarket definition?* Using Rosen's (1974) terminology, we are considering a class of commodities—i.e. homes—that are described by A attributes, $\mathbf{z} = (z_{(1)}, z_{(2)}, \ldots, z_{(A)})$. The conventional definition of submarkets is weak because the entire notion of submarkets can be subsumed by simply allowing $\mathbf{z}$ to include a mixture of structural *and* location attributes, along with interactions between the two.

An alternative justification of persistent attribute price differences is to assume that they are caused by market inefficiencies and frictions (such as imperfect competition among households and the inelastic demand and supply of housing service—Schnare and Struyck 1976). While this is feasible, it leads again to a weak definition because it means that submarkets only exist if markets are inefficient or inflexible. This is problematic because variation in substitutability of dwellings could persist, even in a world of perfectly efficient markets. Heterogeneous dwelling types, heterogeneous locations, heterogeneous preferences, MMEs and TIEs, can all exist in a frictionless world, and would cause substitutability to vary across dwellings, even if the supply of each dwelling type was perfectly elastic and all market participants were perfectly informed. Variations in substitutability, therefore, provide a more robust theoretical basis for submarkets because they do not rely on market inefficiency (as in Schnare and Struyck 1976) nor discrete breaks in the land rent surface (as in Fik *et al.* 2003 p.635, 638).

Criterion 3: SEMs should be Robust to Unobserved Attribute Variation

<u>Rationale</u>*: If dwelling characteristics are spatially clustered, but not fully described in our hedonic data, then (a) observed differences in attribute prices can occur even when actual attribute prices are homogenous, and (b) observed spatially correlated errors in hedonic price regressions may simply reflect unobserved attribute heterogeneity.*

Yes, dwellings are heterogeneous, but so are their attributes. If we are to correctly ascribe attribute price differences to submarket effects we must know all variations in attribute quality and quantity between dwellings. That half a tank of petrol costs less than a full tank is no indication of discrepancy in price per unit. Likewise, apparent differences in price per room between tenements and modern flats may reflect unmeasured differences in room size (e.g. tenement rooms have higher ceilings) rather than submarket boundaries. Unfortunately, full information on the quality and quantity of every attribute of every dwelling is rarely, if ever, available; and the measurement errors that result will not be random but correlated with building type, which in turn is likely to be clustered across space. Coefficient shifts in hedonic regressions may therefore be coterminous with the spatial pattern of measurement errors, rather than market segmentation, which precludes the extrapolation of such boundaries from spatially autocorrelated errors (as in Tu *et al.* 2007).

Note, however, that attribute measurement errors would not affect measures of substitutability that are based on the price of the *overall housing bundle*. If two dwellings genuinely belong to the same submarket, one would expect the price of the overall housing bundle to respond in a similar way to demand and supply shocks, irrespective of attribute prices. Focusing on the dynamics of the sale price of the entire housing bundle (which is generally measured with precision) rather than attribute prices (which are not) as the basis for submarket analysis is potentially the most fruitful way for submarket research to develop.

Criterion 4: SEMs should not impose nor assume Convexity, Compactness, or Contiguity

Rationale: *If substitutability is granular (non-contiguous), non-convex or non-compact in Cartesian space, Euclidean distance will not adequately describe the spatiality of submarkets.*

A crucial factor that has tended to be overlooked in the debate over the degree to which submarkets are structural or spatial (see review by Watkins 2001) is the role of submarket *shape* which has a potentially important role in qualifying the effect of Euclidean distance in determining or approximating substitutability. If submarkets are elongated, fragmented, or have holes (as in the concentric circles of the access-space model), dwellings can be far apart but still be in the same submarket. This has profound implications for the use of Euclidean distance to approximate submarket effects or compute weights matrices in spatial econometric models. If we use distance to define the spatial weights matrix when submarkets are non-convex then the errors observed will reflect a mixture of errors from attribute and amenity mis-measurement (which will be likely clustered across space—see above), and errors that arise from the failure to account for non-convex patterns of substitutability (which are also likely to be distributed non-randomly and nonlinearly in Cartesian space).[5] We therefore seek a methodology that will do justice to the complexity of submarket spatiality—revealing the shape of submarkets rather than imposing or ignoring it.

Surprisingly, the theoretical processes that fashion the geographical footprint of housing market areas has received little attention compared with those of the firm. 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 creates a tendency for markets to converge to some optimal shape (e.g. the hexagons of Lösch 1940).

Similarly, processes that sort households across space (Schelling 1971; Meen and Meen, 2003) might lead to the shape of submarkets having a systematic component. For example, preference for racial and social homogeneity may minimise submarket perimeter length (because of aversion to living at the boundary) leading to compact convex shapes. However, there may be other factors (the cumulative history of residential planning decisions, local amenities, radial and orbital transport links, heterogeneous preference for mix, etc.) that frustrate such processes. Indeed, the concentric rings of access-space theory would produce highly non-convex sets in Cartesian space. Maclennan (1982, p.23) however, argues that "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 therefore might expect the city to comprise a patchwork of residential enclaves, each with its own core and periphery. Submarkets of the type described by Maclennan may be equally non-convex in Cartesian space, but made up of

[5] A theoretical justification for including spatially lagged dependent variables in hedonic models is to capture the displaced demand, but displaced demand is much more likely to impact on dwellings that are *in the same submarket*. Buyers will switch to alternative dwellings that are close in substitutability space; not necessarily close in Euclidean space

many sets of concentric circles centred on multiple cores, rather than a single sequence centred on the CBD. [6] If we could measure the true configuration of the urban housing market we would presumably be able to verify which of these theories dominates in particular urban contexts, leading to a taxonomy of submarket structures for world cities.

All this does not lead us to reject hedonic methods *per se* which have many useful applications, not least as a means of controlling for the mix of properties coming onto the market when computing measures of house price change (use is made of this feature in the empirical illustration below). It is also worth emphasising that the areas derived using HAPV and attribute cluster methods may be of practical value even if they do not correspond to the theoretical definition of submarkets because they may identify measurement errors in the data (such as unobserved attribute variation) and therefore lead to improvements in prediction accuracy (Bourassa *et al*. 2003). The crucial point is that HAPV does not provide a robust way of defining submarkets. For that we must return to the notion of substitutability as the basis of submarkets and to the development of a tractable method of empirical estimation.

3. Deriving a Substitutability Approach to Submarkets

Early work on submarkets (Rapkin *et al.,* 1953) and Grigsby, 1963) drew directly on the concept of substitutability: “A housing market area is the physical area within which all dwelling units are linked together in a chain of substitution…” (Rapkin *et al*., 1953, pp. 9-10 quoted in Grigsby, 1963, pp. 33-34). It has proven difficult, however, to operationalise this approach empirically. To measure the degree to which two goods are close substitutes, we

[6] Note that many of these possibilities are at odds with the convexity restriction imposed in Clapp and Wang (2006).

ideally seek to estimate the cross price elasticity of demand (CPED) but this requires estimating how the demand for one attribute bundle is affected by the selling price of another. “Observed marginal hedonic prices … reveal little about underlying supply and demand functions” (Rosen, 1974, p. 50). Therefore, some alternative method of approximating CPED, that does not require the decomposition of the housing bundle, seems the most promising way to proceed.

Rothenberg *et al.* (1991) made a concerted effort to measure substitutability but the approach relied heavily on the stability of hedonic coefficient estimates (which is problematic—Maclennan 1982). The alternative to HAPV put forward in this paper attempts to exploit the dynamic nature of the market, and make use of relationships between price *changes* (rather than price levels). Essentially, the Cross Price Elasticity of Price (CPEP) is proffered as a proxy for the Cross Price Elasticity of Demand (CPED), and hence of substitutability.

Proposition 1. *If demand and supply curves are well behaved (sloping 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.*

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;* hence, CPED > 0. Conversely, if *i* and *j* are complements, then CPED < 0. Now consider the following corollary. If *j* is a close substitute, a rise in the price of *i* causes a large increase in the demand for dwelling *j*, and if the supply of homes 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*. That is, $\uparrow p_j \Rightarrow \uparrow Q_{Di} \Rightarrow \uparrow p_i$ (*cet par*).

The argument can be expressed more formally by considering the following equilibrium condition in the market for dwelling type *i*:

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

where **Z** and **W** are vectors of exogenous factors affecting demand Q_D and supply Q_S respectively, and p_i is price of the inseparable housing bundle *i*. By implicit differentiation of [1], the Cross Price Elasticity of Price is derived as:

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

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), it is clear that CPEP will be positive.

Now compare [2] with the formula for CPED (derived again by implicit differentiation of [1]):

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

Again, 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), it is clear that CPED will also 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,

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

where,

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

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 that the CPEP will be monotonically increasing in the CPED,

$$\frac{d\eta_{ij}}{d\varepsilon_{Q_{D_i},p_j}} > 0,$$

and since the CPED is a measure of substitutability, it also follows that CPEP can be interpreted as a proxy. Crucially, however, CPEP does not require us to explicitly decompose the demand function. This is an important advantage because micro economic attempts to isolate housing demand rely heavily on hedonic estimation (e.g. Rothenberg *et al.* 1991).

Instead, we can approximate η_{ij} using the slope coefficient from a regression of π_{ti}, the proportionate change over time in the price of dwelling *i*, on π_{tj}, the proportionate change over time in dwelling *j*:

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

If $\eta_{ij} > 0$, then *i* and *j* are substitutes. CPEP increases with the level of substitutability to the point where $\eta_{ij} = 1$, 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 $CPEP_{ij} > 1$ should occur other than as a result of market friction. For example, there may be contemporaneous overshoot of p_i in response to a change in p_j, possibly as a result of a 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—Levin and Pryce 2007), 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 $\max[E(\eta_{ij})] = 1$.

4. Using CPEP to Understand the Existence and Spatiality of Submarkets

Existence

CPEP 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:

$$\text{if} \quad S_1 = \mathbf{M} \qquad \text{then} \qquad \eta^* = \max[E(CPEP_{ij})] = 1 \;\forall\; i,j \;, \quad \text{where } i,j \in \mathbf{M}$$

We can represent the non-existence of submarkets by plotting substitutability against D_{ij}, the distance between i and j. In η_{ij}, D_{ij} space by a horizontal scattering of points all exactly equal to (or randomly scattered around) η^*, the value representing perfect substitutability. This scenario is depicted graphically in Figure 1.

Figure 1 Single Unified Housing Market ($S_1 = \mathbf{M}$**)**

Spatiality: the Effect of Distance

It would be useful 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 η_{ij} and Euclidean distance D_{ij} between pairs of dwellings (i,j):

$$\phi = \partial \eta_{ij} / \partial f(D_{ij}) \qquad [4]$$

For a simple generic measure of the effect of distance on substitutability, one could assume CPEP to be approximately linear in logged distance: $\eta_{ij} = \alpha + \phi \ln D_{ij}$. If proximity is not an important aspect of substitutability, then one would expect η_{ij} to be unrelated to distance, resulting in a spherical scatter of D_{ij} (measured by $\phi = 0$) on D_{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 disamenities, for example), then one would expect $CPEP_{ij}$ to decline with distance, most probably at a decreasing rate, illustrated in Figure 3.

Figure 2 Non-Spatial submarkets: "Spherical" Scatter of η_{ij} on D_{ij} ($\Rightarrow \phi = 0$)

Figure 3 Spatial submarkets: Downward sloping Scatter of η_{ij} on D_{ij} ($\Rightarrow \phi < 0$)

Why might we expect there to be a scatter, rather than a line, of points in η_{ij}, D_{ij} space? If strong substitutability occurs between distant dwellings because of elongated and non-convex shapes of spatial submarkets, then dwellings at the extreme ends of that submarket may be highly substitutable, but far apart. Also, there may exist scattered clusters of substitutable bundles due to similar dwelling, neighbourhood and amenity combinations occurring at different points in the city (Rothenberg *et al.* 1991 p. 64). Thirdly, there may exist non-causal (i.e. coincidental) contemporaneous movements in distant pairs of inflation time series. This leads to Type II errors: CPEP will lead to some dwellings being allocated to the same submarket when in fact they belong to a different submarkets.

If we are interested in deriving *spatial submarkets* – that is, if we want the areas delineated to reflect proximity between dwellings – we can reduce the probability of Type II errors by screening out spurious correlations between price changes of distant dwellings. Consequently, when deriving submarkets from the substitutability measure, there may be an argument for including an explicit spatial component, particularly since spatial submarkets can then be compiled into non-spatial ones (Clapp and Wang 2006). Further criteria (e.g. similar price bracket) could be added to reduce Type II errors, but this is likely to increase the incidence of Type I errors (i.e. allocating dwellings to different submarkets when in fact they belong to the same submarket).

Lattice of Substitution and the Shape of Submarkets

Having decided on a measure that allows us to gauge the substitutability between a given dwelling $i = a$ and all other dwellings $j \neq a$ in the urban area, we can conceive of this set of bilateral links for the dwelling $i=a$ as a *Lattice of Substitution,* $\boldsymbol{L}_a = \{\eta_{ij}: i = a\}$. This set of relationships can be represented as a digraph, as depicted in Figure 4 where each dwelling represents a node and each cross price elasticity, η_{aj} , represents an edge. We can also think of the lattice as a surface in Cartesian space, plotted for a particular dwelling $i = a$, where the hills of this surface represent dwellings that are considered by the market to be close substitutes to a and the valleys represent dwellings that are not considered close substitutes to a. Note that a separate surface (or digraph) could be plotted for each and every dwelling in the housing stock. The complete family of lattices $\mathbf{L} = \{\boldsymbol{L}_i: i = 1, 2, \ldots, V\}$ therefore fully describes the substitutability set.

Figure 4 Digraph for a First Order Substitution Lattice

Submarkets can then be thought of as schema for summarising and simplifying the structure of the substitution lattice family for an urban area. The simplest derivation of a set submarkets is one that categorises a single *Lattice of Substitution* into groups with similar levels of substitution with respect to a particular dwelling $i = 1$. We label this a *First Order Categorisation* (FOC) and is essentially a matter of identifying contour lines of substitutability with respect to dwelling $i=a$. *Second Order Categorisation* (SOC) entails clustering according to two substitution lattices, $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$:

$$S_1, S_2, \ldots S_K \subseteq \mathbf{M} = \{i: \text{cluster}(\eta_{1j}, \eta_{2j}) \text{ where } \eta_{1j} \in \boldsymbol{L}_1 \text{ and } \eta_{2j} \in \boldsymbol{L}_2\}$$

We might choose randomly the dwelling that constitutes the basis for L_2, or we might be more judicious and deliberately select a dwelling that is not a close substitute to dwelling i = 1. *Third Order Categorisation* would involve clustering according to three substitution lattices L_1 , L_2 and L_3, and so on. Using more than one substitutability lattice offers a means of triangulating our results. Ideally, one would like to perform a V^{th} order Categorisation but this is likely to be computationally prohibitive.

Note that the clustering processes described above do not impose spatiality because dwellings are clustered in substitutability space not Cartesian space. This is important because, while the outcome of this process may well lead to systematic patterns in Cartesian space, any apparent spatiality of clustering outcome will not have been caused by the method and therefore should reflect the true spatiality of housing substitution. One might nevertheless seek to include an explicitly spatial component in how dwellings are grouped which can be achieved by including the Cartesian coordinates (x, y) of j in the cluster algorithm.

5. Empirical Illustration

To illustrate, consider now the application of the CPEP method to 33,680 GSPC residential property transactions in Glasgow, Scotland, all spatially coded[7] and with attribute information, for the period 1999 to 2007. Quarterly house price inflation time series were computed for ten thousand individual dwellings using the following procedure:

(1) ***Estimate a Third Order Taylor Series approximation of the house price surface in Cartesian space for each year.*** Extending the parlance of Fik *et al.* (2003)[8], Time-

[7] The geocoding is accurate to within around ±25m on average.
[8] see also Clapp and Wang (2006).

Location Value Signatures (TLVS), were estimated using flexible functional forms that include interactions between attributes, *x,y* coordinates for the dwelling, area dummies (based on *a priori* information on where likely shifts in the price surface may lie[9]), and quarter dummies for year's surface estimation. Insignificant variables and dummies were then eliminated using a stepwise procedure. Note that each TLVS was estimated independently for each year, allowing coefficients complete freedom to vary over years. Coefficients were also allowed to vary over space through interactions with *x,y* coordinates and area dummies. Dwelling type and attributes are included to control for the mix of properties selling in a particular time period. Quarterly time dummies are also included (and allowed to interact with dwelling type and location). The adjusted R^2 results for all nine TLVS regressions were as follows: 0.73 (1999), 0.73 (2000), 0.76 (2001), 0.71 (2002), 0.63 (2003), 0.58 (2004), 0.61 (2005), 0.64 (2006), 0.71 (2007). Having estimated a TLVS for each year with quarterly slope and intercept dummies, an estimated price, P_i, could then be computed for each i in every time period (quarter): $P_i = TVLS_{i(t)}$.

(2) ***Estimate inflation surfaces for each intervening time period and extracting time series for ten thousand constant quality dwellings.*** By calculating the vertical distance for each i between each successive TLVS ($\Delta P_i = TVLS_{i,(t=2)} - TVLS_{i(t=1)}$), as illustrated in Figure 5, and then computing this as a proportionate change ($\pi_{i(t=2)} = \Delta P_i \, / \, TVLS_{i(t=1)}$) we were able to produce house price inflation surfaces. NB once the inflation surfaces have been estimated, it is possible to read-off inflation values for any point in the geographical space covered by the model. In the event, 10,057 hypothetical dwellings were selected at points where house sales had occurred at some point in our data period to ensure that the dwellings occurred at real life residential locations.

[9] Area dummies are based on realtor jurisdictions, and local authority areas (property taxes, in the form of Council Tax levies, and service provision, vary by local authority).

Figure 6 Computing Price Change for Dwelling *i*

a. Existence of Housing Submarkets

Calculating $\phi = \partial\eta_{ij}/\partial f(D_{ij})$ is not a trivial exercise. If there are 10,057 dwellings for which we have computed inflation time series, then there are 10,057 x 10,057 potential correlations between inflation time series to be estimated, and 10,057 x 10,057 distances to be calculated. Including correlations/distances from i to itself, and those correlations/distances from i to j when D_{ij} has already been calculated, results in over one hundred million pairs of dwelling units, (i,j), for which we need to compute $\ln\eta_{ij}$ and D_{ij}. To make computation more manageable, we estimated η_{ij} , the Cross Price Elasticity of Price, for a random selection of 100,000 pairs, derived from the slope coefficient from regression of π_i on π_j, where π_i is the annual constant quality price inflation time series for i. The scatter plot of η_{ij} (CPEP) against D_{ij} (distance) are plotted against each other in Figure 6. 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. However, as distance increases, it clearly has distance has a diminishing effect in determining substitutability—beyond 8km the slope of the line of best fit does indeed become horizontal.

Figure 1 Scatter plot of Cross Price Elasticities and Distance Between Dwellings with Fractional Polynomial line of best fit

Spatiality: the effect of distance

Given the downward sloping relationship between substitutability over short distances, there is *prima facie* evidence in Figure 6 that submarkets have a spatial component: the value of ϕ for the system as a whole is negative, $\phi = \partial\eta_{ij}/ \partial \ln D_{ij}$ = **-.017** (Robust CI = -.0179825, -.0164733; R^2 = 0.02, n = 100,000). The use of logged distance in computing ϕ was imposed as a simplification. When we run a linear spline regression of η_{ij} on D_{ij} we find that that the slope declines in absolute terms. Up to 1km the sllope = $-.45$ (robust CI = -.4879, -.418; Adj R-squared = 0.0785). That is, for every 1km increase in distance between dwellings, the cross price elasticity, η_{ij}, falls by 0.45 units. From 1km to <2km, $\partial\eta_{ij}/\partial\ln D_{ij} = -.35$ (robust CI = -.3691, -.3323; Adj R-squared = 0.0785). From 2 to <4km, $\partial\eta_{ij}/\partial\ln D_{ij} = -.01$ (robust CI = -.0128, -.0037; Adj R-squared = 0.0785). Beyond 4km, the distance effect on substitutability becomes negligible, $\partial\eta_{ij}/\partial\ln D_{ij} \approx 0$ (robust CI = -.0050, -.0030; Adj R-squared = 0.0785).

Nevertheless, it is clear from the very low R^2 values associated with Figure 6 that the substitutability between dwellings also has a large non-spatial component—at least in terms of the capacity of simple Euclidean distances to capture spatiality (even in the spline regression, 92% of the variation in CPEP is due to factors other than Euclidean distance). This provides support for the non-spatial conception of submarkets (an important theme in Rothenberg *et al.* 1991) and an imperative to further explore the shape of submarkets—the existence of convexity, granularity and non-compactness renders distance an incomplete measure of submarket spatiality.

Spatiality: is there variation in SM shape and granularity?

When we plot the substitution lattice for randomly selected dwelling d_{120} (Figure 7) we see clear evidence of non-convex spatiality simply by looking at the shape of contour lines. In terms of allocating dwellings to discrete submarkets based on more than one Substitution Lattice, visualising submarkets requires the application of cluster methods. Two further dwellings were selected (d_{9206} and d_{3247}) as the basis for two further substitution lattices, chosen on the basis of (a) being a distant substitute of d_{120} ; (b) being physically distant in terms of location from d_{120}.

Figure 7 Substitutability Lattice for Dwelling d_{120} Plotted as a Contour Map

Two questions arise at this point: (1) *How many clusters (submarkets) should we opt for?* In the context of a substitution lattice approach, we are clustering along a continuum and so the number of groups is otherwise arbitrary. And (2) *Should we include an explicit spatial component in the cluster algorithm?* The answer to both is that "appropriate definition of submarkets depends on the use to which they are put" (Bourassa *et al.* 2003 p.12). Policy makers, for example, may seek to carve the city up into a few large contiguous submarkets, in which case, one would set the cluster algorithm to derive several clusters and include an explicit spatial component (in order to encourage contiguity).

For purposes of creating area dummies for use in hedonic prediction accuracy, one is likely to prefer many clusters, and again including a spatial component in the cluster algorithm because the more spatially specific the dummies, the more likely one is to account for the unmeasured attribute and amenity variation (see Criterion 3 above). This is illustrated below Figure 8 which plots the Adjusted R^2 and log likelihood results of a series of grid searches over number of clusters for a variety of cluster functions ((i) FOC = First Order Categorisation as described above, (ii) SOC = Second Order Categorisation: $S_1, S_2, \ldots S_s \subseteq \mathbf{M} = \{i\text{: cluster}(\eta_{120j}, \eta_{9206j})$; (iii) TOC = Third Order Categorisation: $S_1, S_2, \ldots S_s \subseteq \mathbf{M} = \{i\text{: cluster}(\eta_{120j}, \eta_{9206j}, \eta_{3247j})$; and (iv) TOsC = Third Order Categorisation with explicit spatial component: S_1, S_2, … $S_s \subseteq \mathbf{M} = \{i$: cluster($\eta_{120j}$, η_{9206j}, η_{3247j}, x, y)). Including an explicit spatial component helps reduce the effect of spurious correlation between price movements of distant dwellings. Note that improvement of hedonic prediction accuracy does not imply better understanding of submarkets—merely that one has done a better job of capturing unobserved measurement errors.

Figure 8 Adj R^2 and LL Diagnostics

If, however, the goal is to identify the shape of submarkets (to ascertain convexity etc.) one should probably avoid including an explicit spatial component in the cluster algorithm. This is because most cluster functions tend to produce convex clusters even when the true clusters are non-convex sets. Omitting the spatial component will not prevent the imposition of convexity in CPEP space by the cluster algorithm, but this will not cause the submarkets to be convex in Cartesian space as demonstrated in Figure 9 which shows the result of a Third Order Categorisation. We can see the effect of including a spatial component to the cluster function by comparing Figure 9 with Figure 10 (which plots the results of a cluster function that includes x, y coordinates)—not only does this tend to produce submarkets made up of contiguous elements but it also tends to produce submarkets that are relatively convex (which also may also be a desirable feature in some applications). The number of clusters could be selected on the basis of dendrograms which, in this empirical example, tended to suggest clusters of between eight and ten submarkets. Dendrograms, however, do not always reveal a clear cut point, so other criteria could be considered, such as the spatial scale of other variables of interest (such as racial segmentation).

What, then, can we conclude from the submarket maps of Figure 10, which have been derived using a method that does not impose spatiality, contiguity or convexity in Cartesian space? First, they tell us that submarkets are indeed a spatial phenomenon. While the submarkets appear fragmented across space, the scattering is not aspatial, but grouped into local sub-clusters. Second, we can see that the spatial effect is not a simple function of distance—there is evidence of non-convexity, non-compactness and non-contiguity (granularity). Third, there is no evidence of the concentric rings predicted by the standard urban economic model – if anything, the clustering of dwellings is more radial than orbital.

Figure 9 Third Order Categorisation

Figure 10 Spatial Third Order Categorisation

6. Conclusion

This paper has sought to clarify the relationship between the theoretical notion of substitutability and the empirical criteria for partitioning housing markets, and to develop a method that leads to a theoretically consistent criterion based on substitutability. The Cross Price Elasticity of Price (CPEP) was offered as a measure of substitutability that is robust to transformative interaction effects and attribute measurement errors (because dwellings are not decomposed into attributes for purposes of determining substitutability; they are instead treated as inseparable bundles). CPEP does not rely on discontinuity in house price surfaces (because substitutability – and hence submarkets – can exist along a continuum) and does not impose or assume convexity, compactness or contiguity (dwellings are clustered in substitutability space rather than Cartesian space, hence revealing, rather than imposing, the geographical pattern of market areas).

It is hoped that the CPEP method will stimulate new avenues of submarket research.[10] For example, CPEP opens up the possibility of investigating what determines substitutability—the extent to which substitution is driven by the physical characteristics of dwellings and the extent to which it driven by the racial and social make-up of the neighbourhood (this might involve regressing $CPEP_{ij}$ on differences in the dwelling characteristics of i and j and differences in neighbourhood characteristics of i and j). It is also hoped that CPEP-based clustering will stimulate the exploration of the nature and causes of submarket shape and the extent to which there are forces that minimise boundary length.

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.

Anguera, R. (2006) The Channel Tunnel—an ex post economic evaluation, *Transportation Research Part A*, 40: 291-315.

Arrow, K.J., (1950) "A Difficulty in the Concept of Social Welfare", *Journal of Political Economy*, 584: 328–346.

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., M. Hoesli and V.S. Peng (2003) Do housing submarkets really matter?, *Journal of Housing Economics*, 12: 12-28.

[10] Computer code used for estimating CPEP based submarkets and more details on the practical steps are available on request from the author

Clapp, J. M. and Wang, Y. (2006) Defining neighborhood boundaries: Are census tracts obsolete? *Journal of Urban Economics*, 59: 259-284.

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

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.

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

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

Greene, W. (1993) *Econometric Analysis*, Macmillan, New York

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

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.

Kintrea, K., Jon Bannister, Jon Pickering, Maggie Reid and Naofumi Suzuki (2008) *Young People and Territoriality in British Cities*, York: Joseph Rowntree Foundation.

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

Leishman, C. (2009) Spatial change and the structure of urban housing submarkets, *Housing Studies*, 24: 563-585.

Levin, E. and Pryce, G. (2007) A Statistical Explanation for Extreme Bids in the House Market, *Urban Studies*, 44: 2339-2355.

Lösch, A. (1940) *Die raumliche Ordnung der Wirtschaft*, Fischer: Jena (cited in Puu, op cit.)

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

Maclennan, D. (1982) *Housing Economics*, Longman, London

Maclennan, D. and Tu, Y., (1996) Economic perspectives on the structure of local housing systems. *Housing Studies*, 11: 387–406.

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.

Puu, T. (2003) *Mathematical Location and Land Use Theory: An Introduction*, 2nd Ed, Springer: Berlin

Rapkin, C., Winnick, L. and Blank, D. (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*, 82:34–55.

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.

Schelling, T. (1971) Dynamic models of segregation, *Journal of Mathematical Sociology*, 1: 143–186.

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

Tu, Y., Sun, H. and Yu, S.M. (2007) Spatial Autocorrelations and Urban Housing Market Segmentation, *Journal of Real Estate Finance and Economics*, 34: 385-406.

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

Figure 2 Single Unified Housing Market (S_1 **= M)**

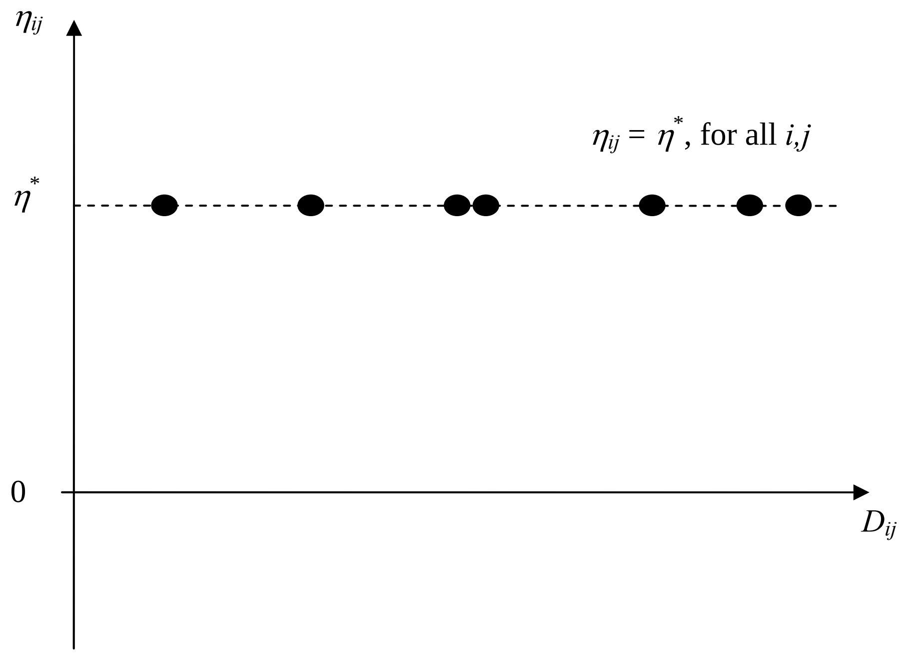


Figure 3 Non-Spatial submarkets: "Spherical" Scatter of η_{ij} on D_{ij} ($\Rightarrow \phi = 0$)

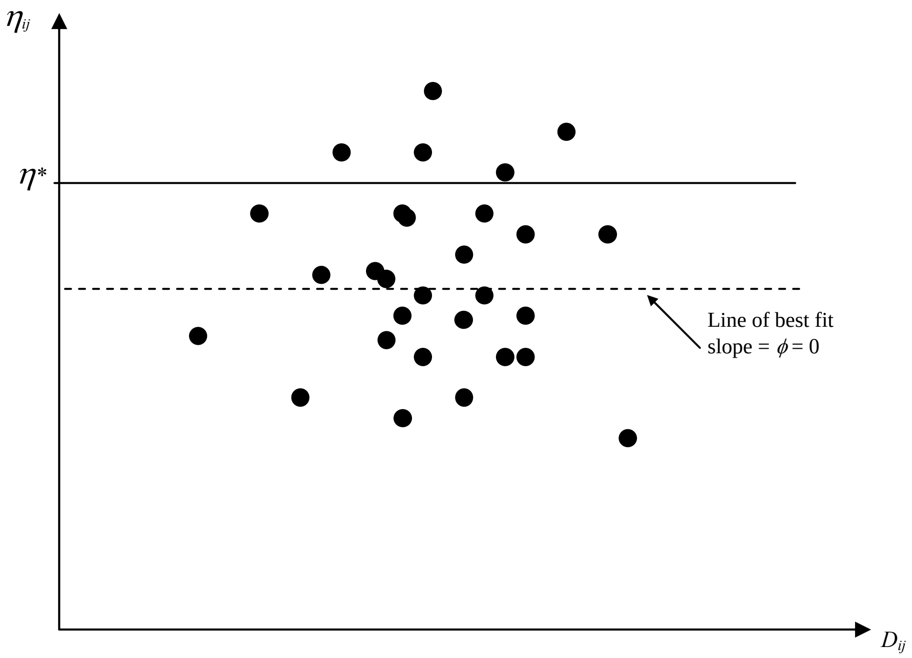


Figure 4 Spatial submarkets: Downward sloping Scatter of η_{ij} on D_{ij} ($\Rightarrow \phi < 0$)

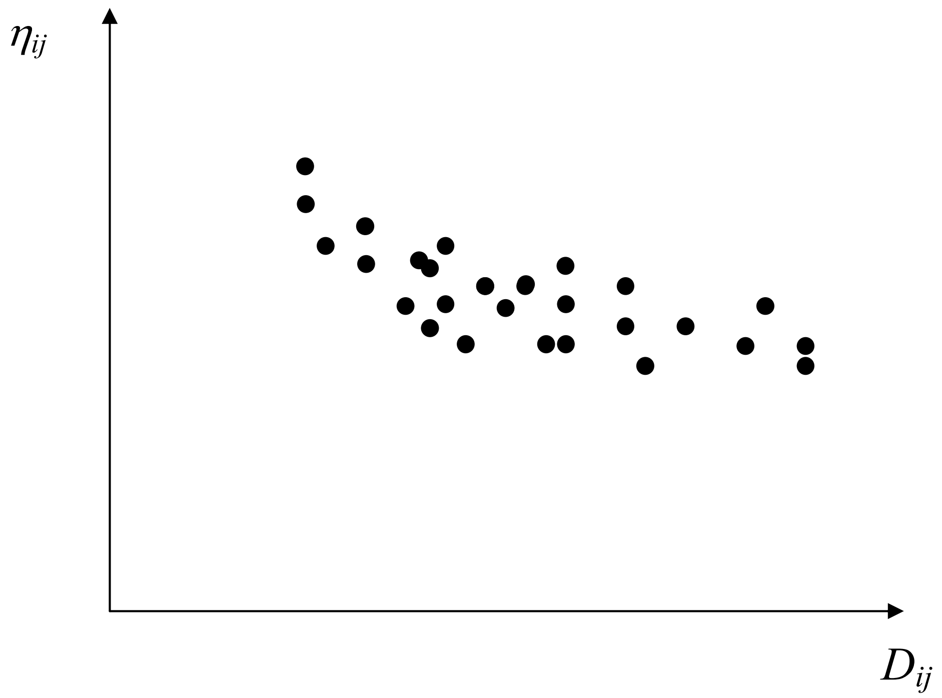


Figure 5 Digraph for a First Order Substitution Lattice

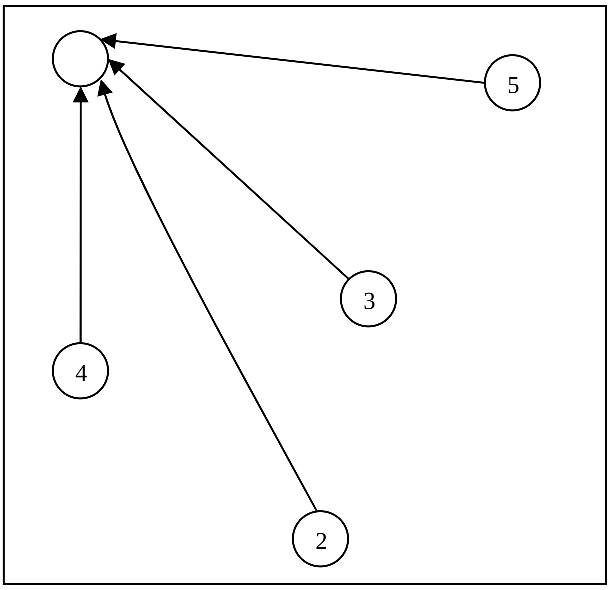


Figure 6 Computing Price Change for Dwelling *i*

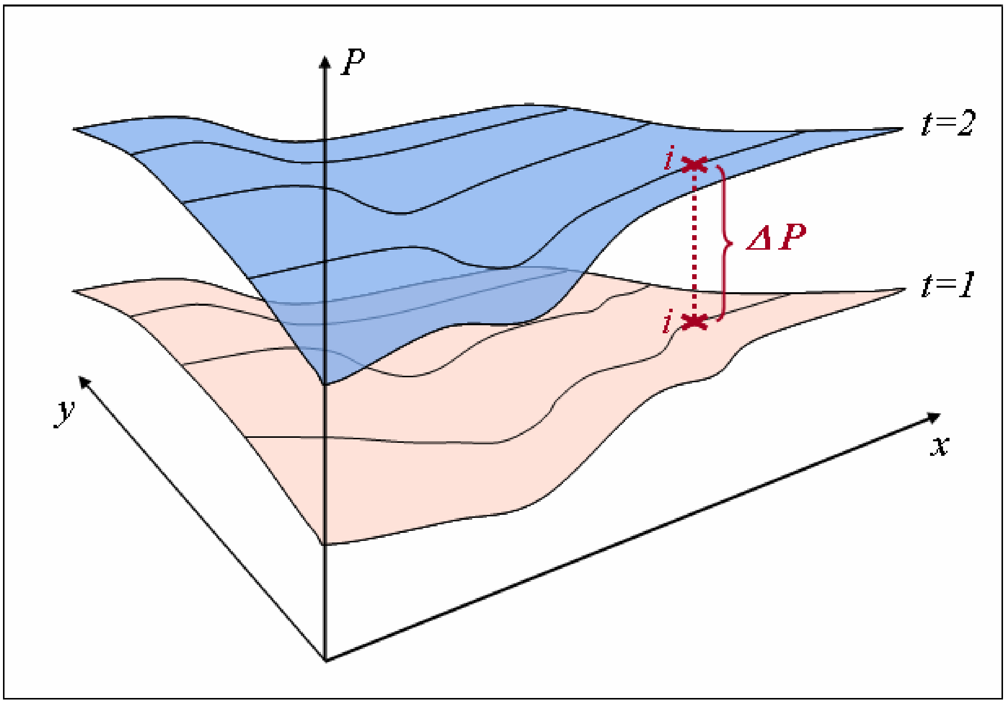


Figure 7 Scatter plot of Cross Price Elasticities and Distance Between Dwellings with Fractional Polynomial line of best fit

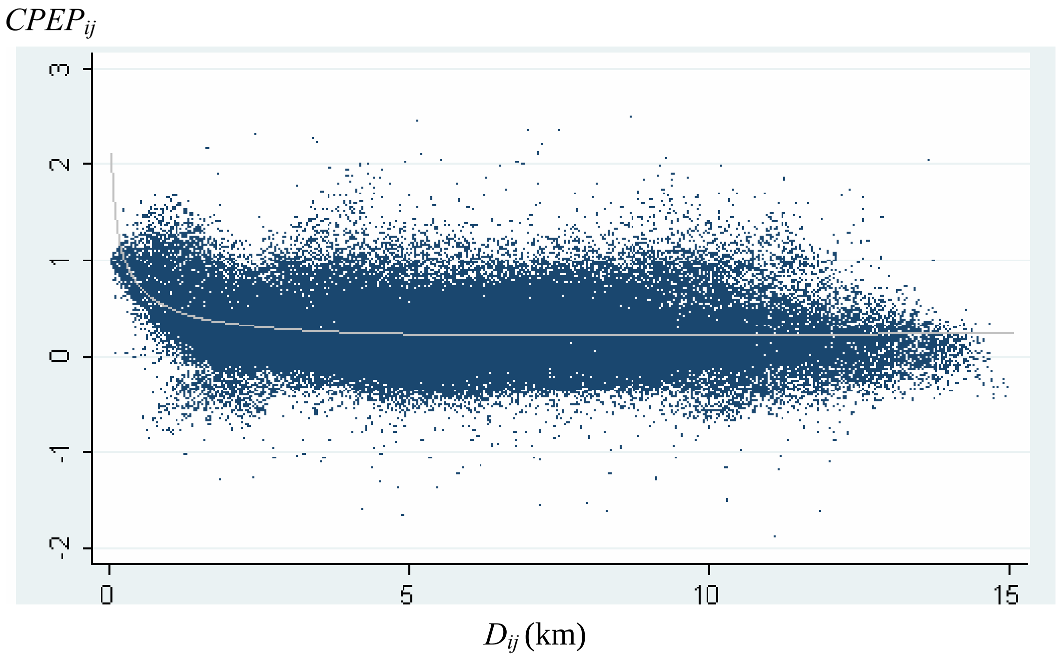
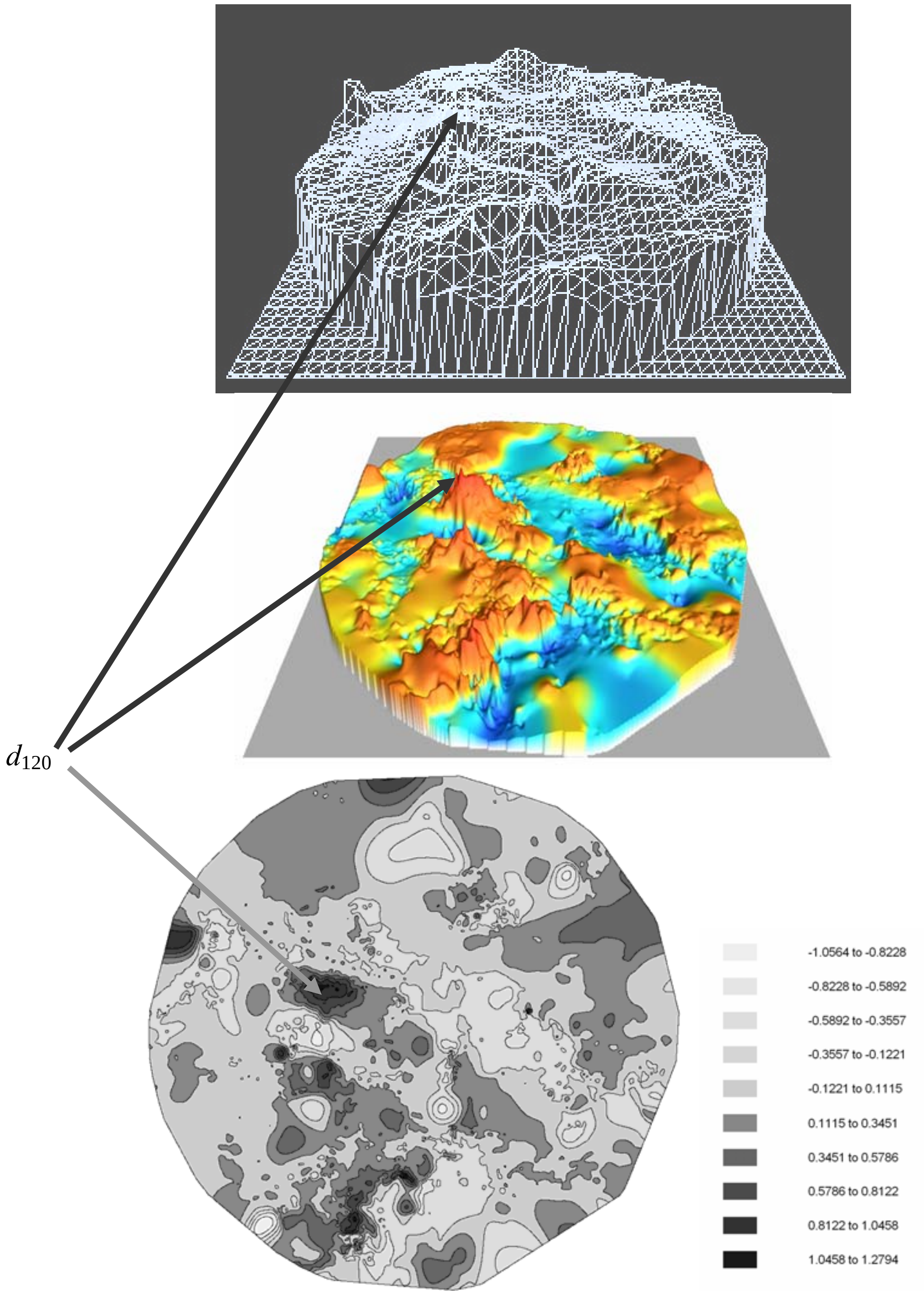


Figure 8 Substitutability Lattice for d_{120} Plotted as Surface Plot & Contour Map



Note: Plotted in Cartesian Space for Glasgow

Figure 8 Adj R^2 and LL Diagnostics

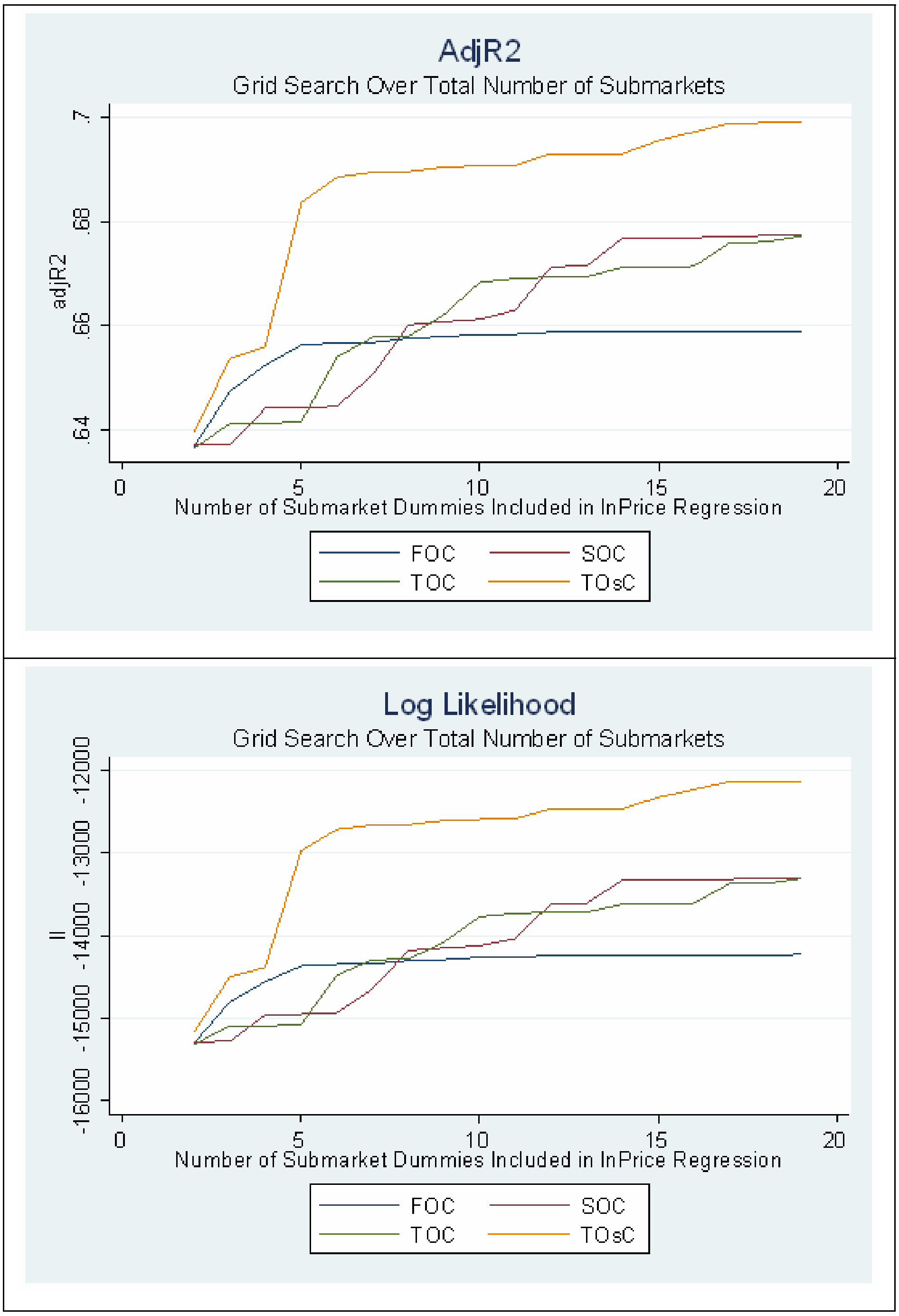


Figure 9 Third Order Categorisation

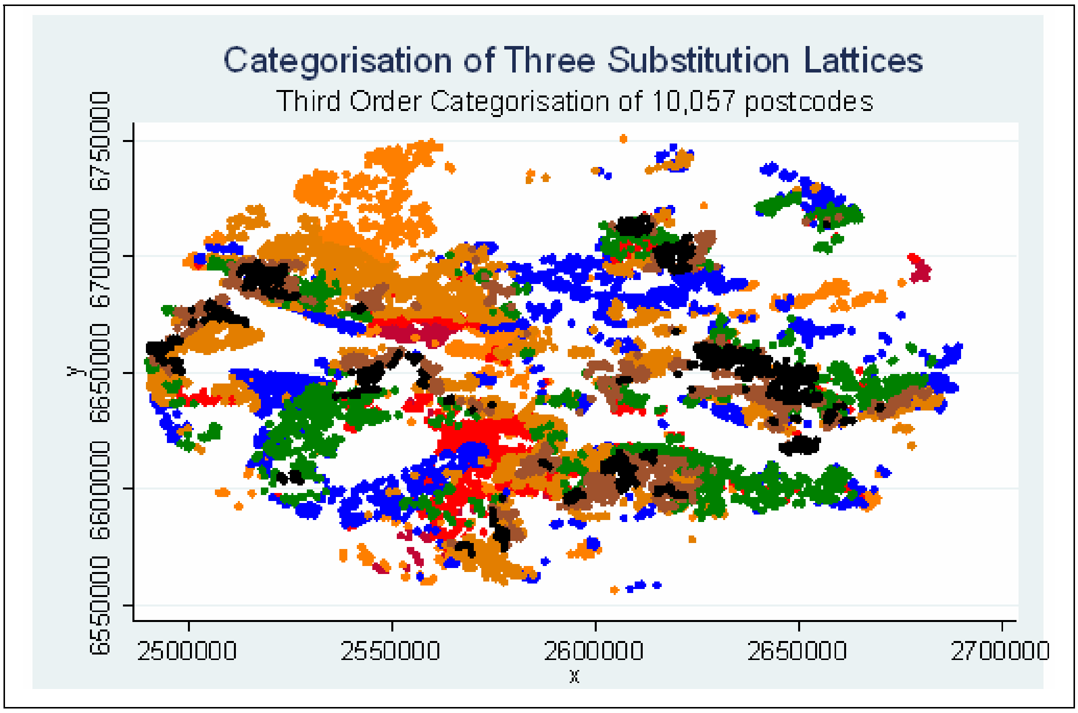


Figure 10 Spatial Third Order Categorisation