

Day, Brett

Working Paper

Submarket identification in property markets: A hedonic housing price model for Glasgow

CSERGE Working Paper EDM, No. 03-09

Provided in Cooperation with:

The Centre for Social and Economic Research on the Global Environment (CSERGE), University of East Anglia

Suggested Citation: Day, Brett (2003) : Submarket identification in property markets: A hedonic housing price model for Glasgow, CSERGE Working Paper EDM, No. 03-09, University of East Anglia, The Centre for Social and Economic Research on the Global Environment (CSERGE), Norwich

This Version is available at:

<https://hdl.handle.net/10419/80235>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

**SUBMARKET IDENTIFICATION IN
PROPERTY MARKETS:
A HEDONIC HOUSING PRICE
MODEL FOR GLASGOW**

by

Brett Day

CSERGE Working Paper EDM 03-09

**SUBMARKET IDENTIFICATION IN
PROPERTY MARKETS:
A HEDONIC HOUSING PRICE
MODEL FOR GLASGOW**

by

Brett Day

**The Centre for Social and Economic Research
on the Global Environment
School of Environmental Sciences
University of East Anglia,
Norwich NR4 7TJ**

**Contact details:
email: brett.day@uea.ac.uk; Tel: (44) (0)1603 592064**

Acknowledgements:

The support of the Economic and Social Research Council (ESRC) is gratefully acknowledged. This work was part of the interdisciplinary research programme of the ESRC Centre for Social and Economic Research on the Global Environment (CSERGE).

ISSN 0967-8875

Abstract

This paper presents a hedonic housing price model for the city of Glasgow in Scotland. The major innovation of the research is the use of hierarchical clustering techniques to identify property submarkets defined by a combination of property types, locations and socioeconomic characteristics of inhabitants. Separate hedonic price functions are estimated for each submarket and these functions are shown to differ significantly across submarkets. Further, the paper illustrates the use of a generalised moments estimator proposed by Kelejian and Prucha that accounts for spatial autocorrelation in property prices. Spatial autocorrelation is shown to be an important consideration with this data. The principal motivation of the research is to provide an indication of the impact of road traffic noise on the market price of property. In all but one of the submarkets exposure to road traffic noise is shown to have a significant negative impact on property prices.

Key Words:

Hedonics, market segmentation, factor analysis, hierarchical clustering, spatial autocorrelation

1. Introduction

The market in which housing is traded, the property market, differs from that of many other differentiated goods in that it is fundamentally spatial in nature. The property market is not a global or even national market, far from it. Rather, many localised property markets may exist simultaneously even within a single urban conurbation.

Unlike other goods that are not constrained spatially, producers (i.e. property owners) cannot transport their products from one location to another. Indeed, spatial constraints in property markets ensure that, in the short run at least, the supply of properties to each market is extremely inelastic. Consequently, we would not expect market-clearing prices to equalise across property markets. Indeed, the equilibrium hedonic price schedule for any particular housing market will be unique to that market reflecting the specific conditions of supply and demand that exist at that locality (see Day, 2001 for a more detailed explanation).

Since property prices in two different markets may be determined by very different hedonic price functions, a primary concern for hedonic researchers is to ensure that data are drawn from a single property market. Using data from the city of Glasgow in Scotland this paper seeks to identify property submarkets using statistical techniques. Separate hedonic functions are estimated for each submarket and statistical tests used to establish the uniqueness (or otherwise) of each function.

The paper is organised as follows. Section 1 provides a brief theoretical overview of hedonic analysis. We describe the hedonic price function, implicit prices for housing attributes and the mechanism by which these prices are determined in the property market. In particular, attention is drawn to the fundamentally spatial nature of property markets. It is argued that property markets are likely to segment spatially and that identifying these market segments is fundamental in the quest for unbiased estimates of the implicit price of traffic noise avoidance.

Section 2 details the data used in the analysis. GIS allows for the compilation of a rich and diverse set of covariates describing the characteristics of properties in a property market. Unfortunately, the quantity of information provided by the application of GIS complicates the estimation of the hedonic price function. In particular, multi-collinearity is rife in the covariate data, with many variables measuring slightly different dimensions of the same basic characteristic. To overcome this difficulty, Section 3 describes the application of factor analysis to the covariate data. Factor analysis provides a way in which the multitude of

variables available to the analyst can be concentrated into a smaller number of factors that identify the major dimensions of difference and similarity between properties.

In Section 4 we return to the issue of market segmentation. Here we illustrate the use of a statistical technique known as cluster analysis. This technique is used to gather observations into groups of properties displaying similar characteristics; not only in terms of their physical attributes but also in terms of their spatial location. Further statistical tests are applied to these groups of properties in order to identify market segments possessing independent hedonic price functions.

Finally, we turn our attention to the estimation of the hedonic price function in each market segment. Here we also acknowledge the spatial nature of the data. In particular, we test for the existence of spatial autocorrelation. That is, we test the hypothesis that the residuals of the hedonic price regression are correlated for properties located near to each other. This would be the case, for example, if properties in close proximity hold similar values for property characteristics omitted from the analysis. Unsurprisingly, the data show strong evidence of spatial autocorrelation. As a consequence, the final hedonic regressions are estimated using a general method of moments estimator that accounts for spatial autocorrelation.

2. Theory

Housing is an example of a *differentiated good*. Such goods consist of a diversity of products that, while differing in a variety of characteristics, are so closely related in consumers' minds that they are considered as being one commodity. Many other goods, including breakfast cereals, cars, computers and beach holidays also fit this description.

Market forces determine that different varieties of the product command different prices and that these prices depend on the individual products' exact characteristics. For example, properties that have more bedrooms will tend to command a higher price in the market than properties that have fewer bedrooms. Furthermore, the set of prices in the market define a competitive equilibrium. That is, in general, the market will settle on a set of prices for the numerous varieties of the differentiated good that reconcile supply with demand and clear the market. This set of prices can be described by a *hedonic price function*;

$$P = P(\mathbf{z}) \quad (1)$$

where P is the market price of a property and $\mathbf{z} = (z_1, z_2, \dots, z_K)$ is a vector of values describing the quantities of K characteristics of a property's structure, environs and location; what we shall term the property's attributes. The partial derivative of the hedonic price function with respect to one of these attributes, say z_i ;

$$p_{z_i}(z_i; \mathbf{z}_{-i}) = \frac{\partial P(\mathbf{z})}{\partial z_i} \quad (i = 1 \text{ to } K) \quad (2)$$

where \mathbf{z}_{-i} is the vector of all other property attributes, is the *implicit price function*. This function describes the additional amount that must be paid by any household to move to a property with a marginally higher level of characteristic z_i , other things being equal. In the research presented here, the primary objective is to identify the implicit price of road traffic noise. The market in which housing is traded, the property market, differs from that of many other differentiated goods in that it is fundamentally spatial in nature. The property market is not a global or even national market, far from it. Rather, many localised property markets exist simultaneously. The market is said to be segmented.

In fact it is usual to think of the products in a property market as being the set of properties existing in a *particular area*. The consumers of these products are the

households that wish to live in the area. They determine the level of demand in the market. The producers are the *landlords* that own the properties. They determine the level of supply of properties to the market.

Unlike other goods that are not constrained spatially, producers (i.e. landlords) cannot transport their products from one location to another. Indeed, spatial constraints in the property markets ensure that, in the short run at least, the supply of properties to each market is extremely inelastic. Consequently, we would not expect market-clearing prices to equalise across property markets.

The equilibrium hedonic price schedule settled on in a particular property market will reflect many factors on both the demand and supply sides of the market. For example, we would expect a property market in which households are generally better off to be characterised by generally higher levels of willingness to pay for property attributes. In such a market, the implicit prices of property attributes such as ‘peace and quiet’ (i.e. the negative of a property’s exposure to traffic noise) will tend to command higher implicit prices than in an identical market in which households are less wealthy. Likewise, on the supply side, the availability of housing attributes will influence the equilibrium hedonic price schedule. Consider, for example, the price paid for waterfront properties in London and Stockholm. Whilst in both cities such properties command considerable premia, the relatively low availability of “Thames-side” properties in London means that they command highly inflated prices compared to those in Stockholm, a city built upon a series of islands.

As a general result, the equilibrium hedonic price schedule for any particular housing market will be unique to that market reflecting the specific conditions of supply and demand that exist at that locality. There is no theoretical reason to expect implicit prices from hedonic analyses of different property markets to return the same value. Indeed, given the heterogeneous supply and demand conditions in different property markets we would expect them to return different values.

A primary concern in hedonic analysis, therefore, must be to ensure that the data used to estimate a hedonic price function pertains to houses from a single property market. If this is not the case we risk seriously biasing our analysis. Estimates of implicit prices coming from such an analysis may bear little resemblance to the true implicit prices ruling in the individual property markets.

3. Data Description

In empirical applications researchers estimate the hedonic price function of Equation (1) by collecting data on the selling price of houses in a particular property market and regressing these on the attributes of those properties (i.e. the z_i). The data used in this study were collected from the south and north-west of the Scottish city of Glasgow. Using publicly available records¹, the addresses and prices of around 3,500 properties sold in the study area during 1986 were collated.

As a first step, the location of each property was determined. This was achieved using a data base provided by the Ordnance Survey (OS) that provides a unique grid reference for each postal address in the UK² (Martin *et al.*, 1994). The study area and the locations of all the properties in the sample are illustrated in Figure 1.

Now, the houses in a property market are extremely heterogenous structures. They differ, in terms of number of rooms, décor, age, plumbing, central heating, garden size and a myriad of other structural attributes. In the estimation of hedonic price functions researchers would hope to control for the influence of these attributes on property prices. However, as every good estate agent will eagerly inform, another fundamental determinant of property prices is “location, location, location!”. Indeed, a property’s market price will probably contain an element reflecting its location with respect to amenities such as town centres, shops, schools and parks. Further, it would not be altogether surprising to find that property prices reflect the socioeconomic characteristics of the neighbourhood in which the property is located. Moreover, and fundamental to this research program, we might expect environmental characteristics of a property’s location (e.g. exposure to traffic noise) to be capitalised into the market price of the house.

One of the prime motivations of this research was to demonstrate the potential of GIS for compiling data for the estimation of hedonic price functions. Indeed, in this study the vast majority of data on property attributes came from desk-based interrogation of GIS data bases. To structure our discussion, consider Table 1, which provides a categorisation of property attributes.

¹ Fortunately, the selling price of properties in Scotland is public information. Property sales are recorded in the Register of Sasines. Amongst other information the register records the exact postal address of the property and the price at which the property was sold.

² OS ADDRESS-POINT

Figure 1: Location of properties in the sample

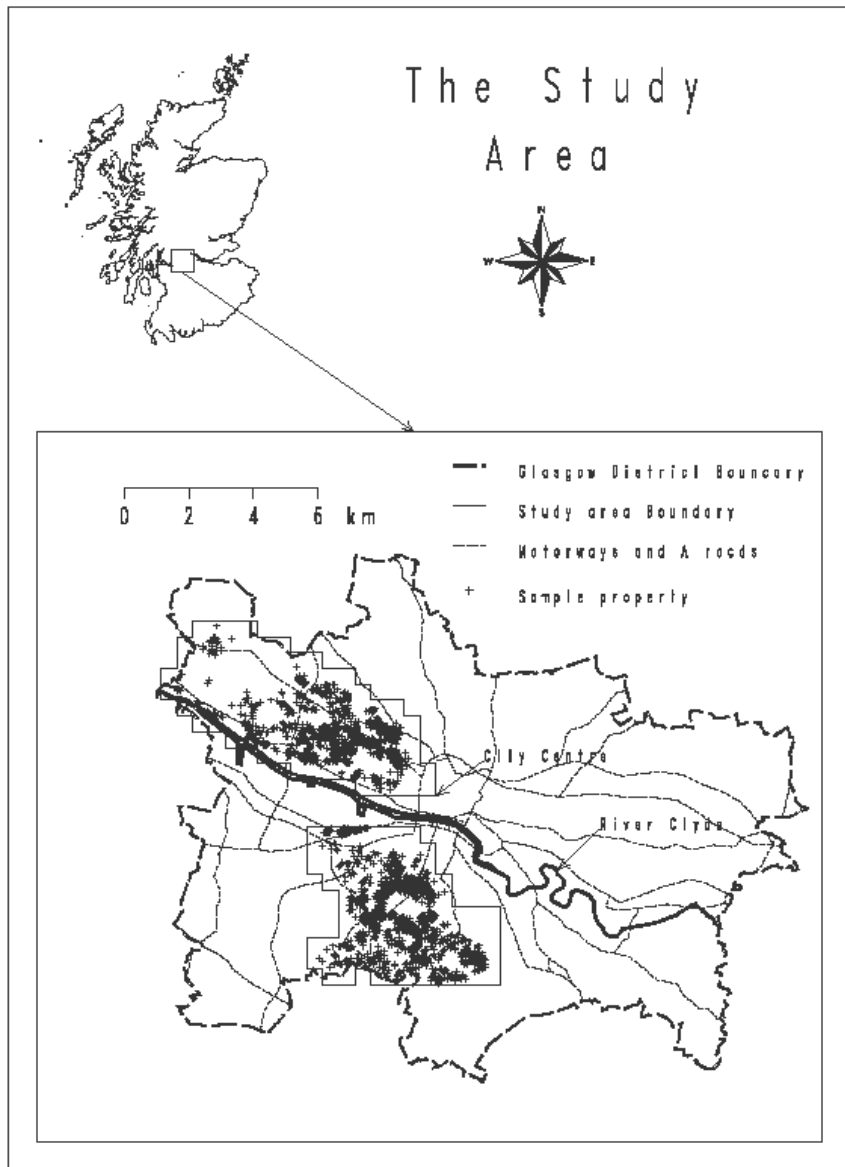


Table 1: Categories and examples of attributes of housing

Attribute Category	Examples of Attributes in this Category
<i>Structural</i>	Number of rooms; presence of garage; size of garden; presence of central heating; etc.
<i>Accessibility</i>	Distance to: bus stop; town centre; school; shopping centre; etc.
<i>Neighbourhood</i>	Average age; race distribution; crime rate; quality of surrounding schools; etc
<i>Environmental</i>	Noise levels; air pollution levels; quality of views from the property; etc.

3.1 Structural attributes:

The OS provides digital maps of the UK recording the locations of ground features, such as buildings, roads and fences³. Given the grid reference of each property in the data set, it was possible to employ GIS to link each property sale observation to features on the ground. Indeed, attributes detailing each property's ground area, size of garden and general shape were calculated in this way. Further, it was possible to identify the general type of each property from this data. Properties were classified as detached, semi-detached, four-blocks, terraced, tenements, flats, subdivided houses or a catch-all category "others".

In matter of fact, in this study the GIS data were supplemented through a brief visual inspection of properties included in the sample. This allowed identification of other structural attributes including, the age of the property, the material from which the property was constructed, the number of storeys and, for flats and tenements, the floor on which the particular property was located. Descriptions of the various structural attributes are provided in Table 2.

Whilst, the GIS greatly simplifies the collection of certain structural attributes for large hedonic property price data sets, it is noticeable that a number of key structural attributes have not been collated. For example, the data does not provide details of the number of rooms of different functions in the property, nor does it indicate whether the property has a garage, central heating or double-glazing. Further, no information is available on the state of repair of property utilities such as plumbing and electricity or, for that matter, the quality of the

³ In this case the digital map used was OS Land-Line.Plus which records ground features with a spatial accuracy of 40 cm (OS, 1996)

internal décor. Since all of these features are likely to be major determinants of property price, their omission should rightly be considered a weakness of the present data set.

Table 2: Structural attributes collected for properties in the data set

Attribute	Description
<i>Floor Area</i>	Calculated as the ground area of the property building on the digital map multiplied by the number of floors of the property determined through visual inspection (m ²).
<i>Garden Size</i>	Calculated as the property's total plot size minus the ground area of the building (m ²).
<i>Shape</i>	Calculated as the ratio of floor perimeter to the square root of area. The greater the ratio, the more complex the property shape.
<i>Storeys</i>	Number of storeys.
<i>Property Type</i>	Series of dummy variables indicating whether the property could be categorised as detached, semi-detached, four-block, terrace, tenement, flat, subdivided house or "other".
<i>Building Material</i>	Dummy variable indicating whether the property was built of locally quarried stone.
<i>Age</i>	Series of dummy variables indicating whether the property was built pre 1919, between 1919 and 1945 or after 1945.
<i>Number of Properties in Building</i>	For flats and tenements, a variable indicating the number of other properties contained within the same building as the target property.
<i>Floor</i>	For flats and tenements, a series of dummy variables indicating whether the property is located in the basement, ground floor, first floor, second floor or on the third or higher floor.

3.2 Accessibility attributes:

Since a property's accessibility attributes are inherently spatial, GIS introduce incredible flexibility and precision into their estimation. For example, it is possible to use GIS to calculate car travel times to important amenities that reflect the actual distance travelled on the road network taking account of road speeds along various road types. In the same way, walking distances from a property to local amenities can be calculated precisely using the network of pedestrian routes (Lake *et al.*, 1998). Descriptions of the accessibility attributes used in the analysis are provided in Table 3.

Table 3: Accessibility attributes collected for properties in the data set

Attribute	Description
<i>Car City Centre</i>	Travel time by car to Glasgow City centre (mins)
<i>Walk Rail</i>	Walking distance to nearest railway station (metres)
<i>Walk Shop</i>	Walking distance to nearest local shop (metres)
<i>Walk School</i>	Walking distance to nearest school (metres)
<i>Walk Park</i>	Walking distance to nearest municipal park (metres)

3.3 Neighbourhood attributes

The census provides a myriad of information on the types of properties and socioeconomic characteristics of the population living in a certain area. As a result census data are ideal for constructing indicators of the attributes of the neighbourhood in which a property is located. Again GIS is a fast and efficient means of matching properties to census data at different spatial scales.

At the most specific level, neighbourhood attributes can be constructed from the smallest unit of the census, the Output Area (OA). Such attributes will reflect the direct neighbourhood of the property. More broadly, census data can be averaged over larger spatial scales such as a postcode district to provide indicators of the attributes of the wider neighbourhood. Descriptions of some of the neighbour attributes collected for the hedonic analysis are provided in Table 4.

Table 4: Neighbourhood attributes collected for properties in the data set

Spatial Scale	Attribute Description
Output Area	% of double income families with no children; % of elderly people living alone; % of ethnic people; % of Irish people; % of multiple earning households; % of households with no access to a car; % of households with no central heating; % of residents not owning their home; % of residents with no exclusive use of WC, bath or shower; % of lone parent households; % of households with > 1 person per room; % of children in households with no earners; % of houses with 1 room; % of houses with 2 rooms; % of houses with 3 rooms; % of houses with 4 rooms; % of houses with 5 rooms; % of houses with 6 rooms; % of houses with 7 or more rooms; % of 2 car households; % of people < 5; % unemployment; % of vacant and second homes; % of residents 16-34 with children < 16; % of residents aged 35-54 with children < 16
Postcode District	% of houses with more than 7 rooms; % of residents born in new commonwealth; % of 2 car households; Number of children per child care worker; % of children aged under 5; Number of children in institutions per 1000 children; % of workers in the construction sector; % of workers in the construction sector; % of households with > 1 person per room; % of double income families with no children; % of people working in the “free” economy; % of full time workers working > 40 hrs per week; % of people with a degree or higher degree; % of people in occupations with gross wages > £23,705; % of homeless people; Ratio of people with limiting long term illness to no. of heal care workers; % of people working in the informational economy; % of Irish people; % houses lacking basic amenities; % of households with working head in social classes IV or V; % of lone pensioner households; % of workers in the manufacturing sector; % of households with working head in social classes I to III; % of residents with a different address 1 year before census; % of multiple earning households; % of recently moving families; % of recently moving pensioners; % of households without access to a car; Number of non manual workers per manual workers; % of 17 year olds not in full time education; % of households not owning their home; % of residents in non self contained accommodation; % of residents aged 35-54 with children < 16; % of children in household with no earners; % of workers in the primary sector; % of men aged 35-54 who are unemployed or on a scheme; % of workers in the service sector; % of single parents; % of lone and cohabiting parents in relation to all families; % unemployment; % of vacant or second homes; % of residents in working class occupations (SOC 4-9); % of working parents with children 0-4; % of residents 16-34 with children < 16; % of 16 and 17 year olds on a government scheme; % of unemployed 16 and 17 year olds

Notice that the OA characteristics include a large number of variables indicating the structural attributes of properties (e.g. percentage of properties with one, two, three, four, five, six, seven or more rooms, percentage of properties with no central heating, percentage of properties without exclusive access to a WC, bath or shower). Since on average in this study an OA consisted of 56 households, these neighbourhood attributes provide reasonable proxies for some of the structural attributes for which data was not available.

3.4 Environmental attributes

In this study, the primary concern was the effect of road traffic noise on property prices. Unfortunately, though hardly surprisingly, there was no specific data immediately available, on the traffic noise exposure of each property. Once again, however, GIS provides a powerful tool for deriving estimates of this environmental attribute.

As a first step, GIS techniques were used to extrapolate point measures of traffic volumes at monitored sites around Glasgow City to all roads in the study area. On the basis of these traffic volumes, the level of noise being emitted from each road was calculated according to formulae specified in CRTN (DOT, 1988) which account for the percentage of heavy vehicles, the speed of the vehicles, the gradient of the road and the road surface. Subsequently the noise exposure at each property was calculated using GIS techniques that accounted for the horizontal and vertical distance between the property and the road, the ground surface separating property from road, sound reflections off neighbouring buildings and the presence of other major roads in the vicinity. Clearly, in the absence of specific data, GIS provide the only means by which reasonably accurate measures of an environmental attribute like exposure to traffic noise can be estimated for large hedonic property price data sets. Further details of the model used to estimate traffic noise are available in Bateman *et al.*, (2001).

A somewhat simpler procedure was employed to determine each property’s exposure to noise pollution from aircraft. A digital map outlining noise contours surrounding Glasgow City Airport was linked to the digital map of property locations. A simple GIS enquiry could then be used to calculate the noise exposure level at each property.

A second set of environmental attributes were collated using GIS; those describing the quality of views from a property. Using digital maps indicating land use and accounting for the topography and location of the property, GIS allowed the researchers to estimate the quantity of land of different types visible from the front and back of the property. The environmental attributes used in the analysis are described in Table 5.

Table 5: Environmental attributes collected for properties in the data set

Attribute	Description
<i>Traffic Noise</i>	Exposure to traffic noise at front of property (dBs)
<i>Aircraft Noise</i>	Exposure to aircraft noise (dBs)
<i>Front View</i>	General measure of open land visible from the front of the property (m ²)
<i>Front Park View</i>	Area of park visible from the front of the property (aggregated with an inverse distance decay that allots greater weight to areas of park near the property than areas of park further away from the property)
<i>Front Industrial View</i>	Area of industrial land use visible from the front of the property (aggregated with an inverse distance decay)
<i>Front Railway View</i>	Area of railway line visible from the front of the property (aggregated with an inverse distance decay)
<i>Front Water View</i>	Area of water visible from the front of the property (aggregated with an inverse distance decay)
<i>Back View</i>	General measure of open land visible from the back of the property (m ²)
<i>Back Park View</i>	Area of park visible from the back of the property (aggregated with an inverse distance decay)
<i>Back Industrial View</i>	Area of industrial land use visible from the back of the property (aggregated with an inverse distance decay)
<i>Back Railway View</i>	Area of railway line visible from the back of the property (aggregated with an inverse distance decay)
<i>Back Water View</i>	Area of water visible from the back of the property (aggregated with an inverse distance decay)

of greater accessibility. As illustrated here, the ability of GIS to calculate large quantities of spatial data rapidly and accurately is a considerable technical advance in the compilation of property attributes for hedonic analysis.

Hedonic price studies must account for many property attributes. It is a matter of some concern to analysts that omitting important attributes from the hedonic price regression may bias estimates of implicit prices. For example, Harrison and Rubinfeld (1978), in their hedonic pricing study of air quality, computed regressions with and without accessibility attributes. Their results indicated that the implicit price of air pollution changed significantly when accessibility variables were deleted, which implied that without accessibility the parameter on air pollution reflected both disadvantages of greater pollution and advantages

4. Factor Analysis of Neighbourhood Attributes

As illustrated by the myriad of attributes listed in Table 4, the use of GIS to interrogate census data provides a rich source of information on neighbourhood attributes. Indeed, here the problem for hedonic analysis is not one of lack of data on attributes but one of over-abundance. Not surprisingly, many of the attributes listed in Table 4 are highly collinear. For example, at the output area level, the percentage of households not owning cars exhibits high positive correlation with the percentage of households that do not own their own property (correlation coefficient of .70) and high negative correlation with the percentage of households that own two cars (correlation coefficient of -.76).

Whilst each of these neighbourhood attributes might have a bearing on property prices, the presence of such collinearity creates a problem for researchers. As is well known, parameters estimated on highly collinear regressors are difficult to interpret. Parameter estimates may have implausible magnitude or, in the worst case, the wrong sign. Interpretation is further confounded by the fact that individual parameters may exhibit high standard errors and consequently low significance levels.

Moreover, it is not clear that each of these neighbourhood attributes will be independently capitalised into the property market. More likely, households in a market will consider more general indications of the neighbourhood of a property, the wealth of the area, its ethnic makeup, the stage of life of its inhabitants etc.

As a result, we propose condensing the excess of neighbourhood attributes into a more manageable set of indices. Each index picks out a major dimension of difference or similarity between property neighbourhoods. For example one index might indicate the wealth of a neighbourhood, effectively combining the myriad attributes that are indicators of wealth/poverty into one dimension. Subsequently, property neighbourhoods can be scored along each dimension. In our example, poor neighbourhoods would generate low scores on the wealth dimension, whilst affluent neighbourhoods would generate high scores. The procedure by which dimensions are identified and property neighbourhoods are scored along these dimensions is known as *factor analysis*.

We do not intend presenting the intricacies of factor analysis here (for a highly accessible text on the subject see Lindeman *et al.*, 1980). In essence, the procedure seeks to identify major dimensions of association between variables (in our case the attributes of neighbourhoods) such that a smaller set of variables can be defined that approximate the variation shown in the original data.

Since we are interested in patterns of association, it is not surprising that the first step in a factor analysis is to calculate the correlation matrix of the M variables under study. Each row (or column) of this matrix can be thought of as representing a point in M -dimensional space. We use M axes to locate the points in this space, where each axis represents the degree of correlation with one of the M variables and ranges from -1 to 1 . Thus the position of the m^{th} point indicates the nature of the correlation between attribute m and all the other attributes.

If there were no correlation between the attributes, each of the M points would be located on its own axis. Alternatively, when the attributes are correlated, as is the case here, the rows of the correlation matrix form a cloud of points in the space $(-1, 1)^M$. Two attributes showing strong positive correlation will have points located close to each other in this M -space. Likewise, an attribute showing strong negative correlation with these attributes will have a point located near to the mirror image of their points on the opposite side of the origin. Thus, attributes that measure slightly different aspects of one underlying dimension will have points that tend to align themselves along an axis running through the origin.

The first step in a factor analysis is to define an alternative set of axes through this space that capture these patterns of alignment. This is achieved by decomposing the correlation matrix into its eigenvalues and associated eigenvectors. As is well known, the M eigenvectors represent just such a set of alternative orthogonal axes. Let us consider the eigenvector with the highest associated eigenvalue. It transpires that of all the possible axis that could be drawn through the space $(-1, 1)^M$, the axis defined by the first eigenvector picks out the dimension capturing the most variability in the location of the points. That is, of all possible axes, the first eigenvector distinguishes the most significant alignment of points.

To illustrate, in the data on neighbourhood attributes we know that many of the attributes measure slightly different dimensions of the wealth of the inhabitants of that area. For example, for each neighbourhood we have details of the levels of unemployment, the levels of home ownership, the levels of car ownership, and the percentage of households with earnings above the national average. In general, relatively poor neighbourhoods will have high unemployment, but low levels of high earners and similarly low levels of car and home ownership. Conversely, relatively affluent neighbourhoods will contain many high earners coupled with high levels of car and home ownership, but low unemployment. Indeed, if we could plot the correlations in these attribute levels on a four dimensional graph it would not be surprising to find that they lie approximately along a straight line passing through the origin. This axis is the first eigenvector.

It summarises the association between the four attributes into one dimension and this dimension can be interpreted as measuring the wealth of the neighbourhood.

Of course it is unlikely that this one dimension will account for all of the variability in the data. Indeed, the eigenvector with the second highest eigenvalue defines a second axis, orthogonal to the first, best approximating the remaining variation in the points. In the same manner, eigenvectors with successively smaller eigenvalues can be used to better and better approximate the location of the M points. The eigenvectors define the “factors” of factor analysis.

The 3,544 properties in this study came from 1,027 different output areas (OAs) and for each OA our dataset contained details of some 25 attributes. The properties were further grouped into 38 postcode districts (PDs) for which information on 45 attributes was available. Tables 6 and 7 detail the first eight factors for the OA and PD neighbourhood attributes respectively. The second column in these tables provides the eigenvalue of each factor. The third column indicates the percentage of the variation in the location of the M points explained exclusively by that factor. If the attributes were not correlated then each factor would correspond to an original axis and explain $1/M$ of the variation. If the attributes were all perfectly correlated then they would all be aligned along one axis and this axis would explain 100% of the variation. The fourth column provides the cumulative sum of this explained variation.

Table 6: Variation explained by the first ten factors of the output area neighbourhood attributes

Factor	Eigenvalue	Variation Explained by Factor	Cumulative Explained Variation
1	6.73	.38	.38
2	2.92	.17	.55
3	2.03	.12	.67
4	1.55	.09	.76
5	1.14	.07	.82
6	1.00	.06	.88
7	.74	.04	.92
8	.65	.04	.96

Table 7: Variation explained by the first ten factors of the postcode district neighbourhood attributes

Factor	Eigenvalue	Variation Explained by Factor	Cumulative Explained Variation
1	18.19	.40	.40
2	6.71	.14	.55
3	3.77	.08	.64
4	3.05	.06	.70
5	2.25	.04	.75
6	1.59	.04	.79
7	1.42	.03	.82
8	1.30	.03	.85

For both sets of attributes the first factor alone explains around 40% of the variation. This indicates that at both neighbourhood scales many of the attributes are highly correlated (positively or negatively) with a single underlying factor. We shall attempt to give an interpretation as to what these factors describe shortly. Notice that successive factors explain progressively less of the remaining variation. Here we arbitrarily select a cut off point and declare that factors explaining greater than 5% of the variation in the data should be retained in the analysis. As such we select 6 factors at the OA neighbourhood scale that collectively describe some 88% of the variation present in the original 25 variables. Similarly we select 4 factors at the PD scale that collectively describe 70% of the variation in the original 45 variables.

One of the arts of factor analysis is the interpretation of factors. Interpretation of factors is the process of describing the underlying dimension of similarity or difference between the neighbourhood attributes captured by a factor. To explain how this is achieved, observe that if the axis defined by a particular factor is closely aligned with an original attribute axis, then that attribute is important in determining the factor. Conversely if the factor axis is orthogonal to an original axis, the attribute described by that axis plays no part in determining that factor. The degree to which individual attributes contribute to a factor is measured by the factors loadings.⁴ A large positive loading indicates

⁴ Frequently researchers will *rotate* the factor axes to improve the ease with which factors can be interpreted. That is, the subspace defined by the factor axes is not changed, but the orientation of the axes themselves are rotated such that they best align with original axes describing the attributes.

that high values of the original attribute are associated with high values of the factor. Similarly a large negative loading indicates that high values of the original attribute are associated with low values of the factor.

Tables 8 and 9 list the attributes associated with each of the identified factors and provide interpretations of the dimension described by these factors.

The final step in a factor analysis is to define a score for each neighbourhood for each factor. Using the factor loadings a regression-like equation is calculated, the parameters of which indicate how greatly each attribute contributes to each factor. Given the attributes of each neighbourhood, the equation can be used to determine how highly a neighbourhood scores on each factor. In effect, neighbourhoods that exhibit high values for attributes that load positively on a factor receive high scores for that factor whilst neighbourhoods that exhibit high values for attributes that load negatively on that factor receive low scores.

The factors can be used as proxies for the original attributes in regression analysis. As has been demonstrated the factors capture a good proportion of the variation shown in the original neighbourhood attributes. Moreover, the nature of their construction ensures that the factor scores are orthogonal overcoming the problem of collinearity in the original set of attributes.

Table 8: Interpretation of Output Area Factors

Factor	Attributes with High Positive Loading	Attributes with High Negative Loading	Interpretation
1	5 & 6 room properties; households with 2 cars; multiple earning households; house-holds with children	2 & 3 room properties; no central heating; households without a car; non home-owning households, unemployment	This factor loads positively on general indicators of affluence and negatively on general indicators of poverty. It is interpreted as indicating the general <i>wealth of the neighbourhood</i> .
2	Ethnic minorities; households with young children; unemployment; households with >1 person per room	elderly people living alone	This factor loads heavily on the presence of ethnic minorities and picks out attributes that reflect characteristics of such communities. It is interpreted as indicating the <i>ethnicity of the neighbourhood</i> .
3	1 room properties; no exclusive use of WC, bath or shower; unemployment; non home-owning households	multiple earning households; households with children	This factor loads very heavily on one attribute; one-bedroom properties. Other attributes indicate relative poverty and a dearth of families. The factor appears to pick out neighbourhoods characterised by bedsits possibly inhabited by migrant workers. We describe this as the <i>bedsit factor</i> .
4	2 room properties	4 or 5 room properties	This factor is more difficult to interpret. Since it loads very heavily on 2 room properties it would seem to indicate neighbourhoods with small properties probably flats or tenements. We describe this as the <i>tenement factor</i> .
5	Dual income households with no children; multiple earning households	non home-owning households; unemployment; households without a car; households with children;	This factor appears to indicate the presence of relatively affluent, home-owning households, characterised by having two incomes and no children. We describe this as the <i>DINKY (dual income no kids) factor</i> .
6	6 & 7 room properties; multiple earning households; households with children	2 & 3 room properties; properties without central heating; house-holds without a car; non home-owning households	This factor loads heavily on large properties, containing affluent households with children. We describe this as the <i>affluent suburbia factor</i> .

Table 9: Interpretation of Postcode District Factors

Factor	Attributes with High Positive Loading	Attributes with High Negative Loading	Interpretation
1	unemployment; youth unemployment; middle-aged unemployment; children with unemployed parents; households with no car; non home-owning households; single parents; lower social classes	multiple earning households; dual income households with no children; working in informational economy;	This factor loads heavily on attributes indicating the economic well-being of the inhabitants of a neighbourhood. In particular it picks out differences in the levels of employment and we describe the factor as the <i>unemployment factor</i> .
2	recently moved to neighbourhood; degree level education; workers in service sector; higher social classes	households with children; workers in manufacturing or construction sectors; workers in the "free" economy	This factor picks out attributes describing neighbourhoods characterised by transient inhabitants that are educated, reasonably affluent service sector workers with no children. We describe this factor as the <i>young upwardly-mobile factor</i> .
3	ethnic minorities; households with young children; youth unemployment; households with >1 person per room	pensioners living alone; households with working parents and young children	This factor loads heavily on the presence of ethnic minorities and picks out attributes that reflect characteristics of such communities. We describe this factor as the <i>ethnicity factor</i> .
4	high social class; workers in non-manual jobs; workers in service sector or information economy; full time employed working > 40 hrs per week; large properties; degree level education, households with 2 cars	lower social class; workers in manufacturing or construction sectors; 17 year old school leavers; households with no car;	This factor loads positively on general indicators of affluence and negatively on general indicators of poverty. We describe this as the <i>wealth factor</i> .

5. Market Segmentation

Most hedonic analyses make the implicit assumption that an urban area represents a single property market. Data is collected from the whole urban area and a single hedonic price function is estimated to describe the equilibrium prices. Maintaining this assumption we estimate a hedonic price function for the 3,544 observations in the Glasgow data set. In the absence of an economic model indicating the functional form of the hedonic price function we adopt the usual assumption and regress the logarithm of the property sale price on a linear combination of the regressors described in sections 3 and 4. The estimated model takes the traditional form;

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (3)$$

where \mathbf{y} is an $[N \times 1]$ vector of observations on the dependent variable, \mathbf{X} is an $[N \times K]$ matrix of explanatory variables, $\boldsymbol{\beta}$ is a corresponding $[K \times 1]$ vector of unknown parameters and $\boldsymbol{\varepsilon}$ is an $[N \times 1]$ vector of random error terms with expected value 0 and variance-covariance matrix $\sigma^2 \mathbf{I}$.

Model (3) was estimated using ordinary least squares regression and the results are reported in the first column of Table 10. For the sake of brevity we do not provide a full interpretation of the results of this regression here. However, we note that the structural characteristics of the property are highly significant and influence price in a predictable way; larger properties with bigger gardens fetch higher prices. Detached houses fetch higher prices than semi-detached, semi-detached than terraced, terraced than subdivided houses, subdivided houses than four-blocks, four-blocks than flats, and flats than tenements. Older properties fetch higher prices, as do flats on the first floor as opposed to ground floor or basement. The neighbourhood factors also perform well. At the OA neighbourhood scale, the wealthier a neighbourhood and the higher it scores on the DINKY and suburbia factors, the higher the price of property. Conversely, the more a neighbourhood is characterised by tenements and inhabitants from ethnic minorities, the lower the price of property. At the PD scale, the wealth of the neighbourhood increases the price of a property whilst the level of unemployment in the area deflates property prices. The accessibility variables are not so convincing. Only the variable measuring the distance to walk to the nearest shop is significant. Unfortunately, this parameter has an unexpected sign suggesting that property prices increase with increasing distance from a shop. Likewise, the variables describing the view from a property appear to have little influence on its selling price. Reassuringly, property prices increase with increases in the quantity of open land visible from the front of the property but this is only significant at a 10% level of significance. Finally, and most

importantly in this study, the parameter on exposure to traffic noise is negative and highly significant (t statistic = -3.293). Given the functional form, the parameter on traffic noise can be interpreted as the percentage change in the selling price of a house brought about by a unit change in traffic noise. In this case then, the analysis suggests that the price of a property will fall by 0.24% for each decibel increase in traffic noise⁵.

As argued in section 2, however, this figure may be misleading if the urban area (in this case Glasgow city) does not constitute a single property market. A number of researchers have investigated the existence of submarkets in urban areas (e.g. Straszheim, 1973; Schnare and Struyk, 1976; Ball and Kirwan, 1977; Sonstelie and Portney, 1980; Goodman, 1978; Michaels and Smith, 1990; Allen *et al.*, 1995). These studies have applied different rules by which properties in an urban area are allotted to a particular submarket. Criteria include, locational or political boundaries, characteristics of households (e.g. income and race), property types and classifications based upon the judgement of estate agents. Here we suggest an approach that makes no *a priori* assumptions concerning the criteria defining submarkets, rather the data itself is used to suggest the pattern of market segmentation.

The procedure suggested here is as follows. Properties are grouped into clusters based on their similarity along a multitude of dimensions; locational, structural and socioeconomic. Hedonic functions are estimated for each cluster of properties. Using tests suggested by previous researchers each hedonic function is compared to the other hedonic functions. If, for two clusters of properties, we cannot reject the hypothesis of equality of parameters, then the clusters are merged. Ultimately, the properties are partitioned into a small number of clusters each displaying a unique hedonic price function and these clusters are interpreted as submarkets.

⁵ This measure is known as the Noise Depreciation Sensitivity Index (NDSI) and is the measure that dominates hedonic price studies into the impact of noise on property prices. Typical values range between .10 and 1.30. For a recent review see Bateman *et al.*, (2001).

Table 10: Hedonic price functions for property markets in Glasgow (OLS)

Independent Variable	Full Sample	Submarkets			
		One	Two	Three	Four
<i>Intercept</i>	9.784 (.088) [†]	9.646 (.178) [†]	9.477 (.140) [†]	8.902 (.299) [†]	9.152 (.169) [†]
<i>Log Floor Area</i>	.135 (.012) [†]	.089 (.025) [†]	.156 (.020) [†]	.142 (.041) [†]	.228 (.034) [†]
<i>Garden Size</i>	.183 (.023) [†]	.044 (.206) [#]	.549 (.133) [†]	.121 (.048) [#]	.164 (.042) [†]
<i>Shape</i>	.013 (.002) [†]	.008 (.004) [#]	.009 (.004) [#]	.007 (.005)	-.001 (.003)
<i>Storeys</i>	-.107 (.026) [†]	-	-	-	-.149 (.031) [†]
<i>Detached</i>	B	-	-	-	B
<i>Semi-Detached</i>	-.057 (.030) [*]	-	-	-	-.070 (.020) [†]
<i>Terraced</i>	-.091 (.033) [†]	-	-	-	-.133 (.038) [†]
<i>Subdivided House</i>	-.431 (.060) [†]	-	-	-.040 (.068)	-
<i>Four Block</i>	-.470 (.050) [†]	-	-.013 (.046) [#]	-	-
<i>Flat</i>	-.549 (.050) [†]	.178 (.072) [#]	-.036 (.022)	-.049 (.036)	-.715 (.147) [†]
<i>Tenement</i>	-.575 (.049) [†]	B	B	B	-
<i>Other</i>	-.473 (.067) [†]	.145 (.093)	-.307 (.117) [†]	.028 (.090)	-
<i>Building Material</i>	.040 (.024) [*]	.038 (.057)	-.070 (.092)	.201 (.186)	.085 (.028) [†]
<i>Age (pre 1919)</i>	.062 (.034) [*]	.084 (.076)	.004 (.108)	.152 (.195)	-.006 (.046)
<i>Age (1919-45)</i>	.065 (.031) [#]	-.064 (.085)	-.036 (.082)	.269 (.216)	-.037 (.042)
<i>No. of Properties in Building</i>	-.006 (.001) [†]	.00004 (.004)	-.015 (.004) [†]	-.012 (.005) [#]	.001 (.003)

Independent Variable	Full Sample	Submarkets			
		One	Two	Three	Four
<i>Basement Flat</i>	-.126 (.037) [†]	.050 (.2)	-.080 (.068)	-.153 (.055) [†]	-
<i>Grd^l Floor Flat</i>	B	B	B	B	B
<i>1st Floor Flat</i>	.038 (.011) [†]	.036 (.021)*	.050 (.015) [†]	-.001 (.040)	.065 (.171)
<i>2nd Floor Flat</i>	.013 (.012)	-.011 (.021)	.029 (.015)*	.019 (.039)	-.066 (.150)
<i>3rd Floor Flat</i>	-.006 (.012)	-.033 (.022)	.021 (.016)	-.048 (.050)	-
<i>OA Wealth Factor</i>	.136 (.007) [†]	.166 (.025) [†]	.098 (.014) [†]	.100 (.021) [†]	.127 (.013) [†]
<i>OA Ethnicity Factor</i>	-.056 (.005) [†]	-.058 (.014) [†]	-.040 (.006) [†]	-.065 (.029) [#]	-.100 (.017) [†]
<i>OA Bedsit Factor</i>	.005 (.005)	.020 (.016)	-.006 (.010)	-.002 (.013)	-.028 (.0140) [#]
<i>OA Tenement Factor</i>	-.060 (.004) [†]	-.080 (.008) [†]	-.050 (.006) [†]	-.031 (.023)	-.054 (.017) [†]
<i>OA DINKY Factor</i>	.048 (.004) [†]	.052 (.009) [†]	.056 (.007) [†]	.0004 (.019)	.013 (.011)
<i>OA Suburbia Factor</i>	.080 (.005) [†]	.060 (.015) [†]	.066 (.009) [†]	.044 (.022) [#]	.076 (.014) [†]
<i>PC Unemployment Factor</i>	-.033 (.006) [†]	-.040 (.016) [#]	-.044 (.013) [†]	-.125 (.050) [#]	-.026 (.019)
<i>PC YUPPY Factor</i>	.053 (.006) [†]	.023 (.012)*	.025 (.010) [#]	.014 (.024)	.100 (.015) [†]
<i>PC Ethnicity Factor</i>	.004 (.006)	-.033 (.02)*	.011 (.008)	.019 (.038)	.055 (.017) [†]
<i>PC Wealth Factor</i>	.076 (.005) [†]	.079 (.016) [†]	.086 (.010) [†]	.078 (.029) [†]	.059 (.009) [†]
<i>Walk Rail</i>	.016 (.013)	.0190 (.025)	.057 (.025) [#]	.124 (.062) [#]	-.023 (.024)
<i>Walk Park</i>	-.007 (.028)	-.075 (.042)*	.037 (.025)	-.040 (.083)	-.075 (.034) [#]
<i>Walk Shop</i>	.076 (.028) [†]	.092 (.076)	-.006 (.044)	-.007 (.112)	.148 (.048) [†]

Independent Variable	Full Sample	Submarkets			
		One	Two	Three	Four
<i>Walk School</i>	.021 (.018)	.091 (.05)*	.033 (.031)	-.019 (.068)	.024 (.034) [†]
<i>Car City Centre</i>	.0001 (.003)	-.015 (.007) [#]	-.006 (.005)	.053 (.018) [†]	.014 (.007) [#]
<i>Front View</i>	.009 (.005)*	.011 (.01)	.006 (.009)	.046 (.029)	.054 (.016) [†]
<i>Back View</i>	-.007 (.006)	-.005 (.01)	.004 (.008)	-.079 (.021) [†]	-.041 (.020) [#]
<i>Front Park View</i>	-.028 (.103)	1.058 (.448) [#]	-.045 (.142)	-.154 (.316)	-.356 (.237)
<i>Back Park View</i>	.091 (.022)	.440 (.471)	-.146 (.265)	1.330 (1.185)	-1.207 (.783)
<i>Front Industrial View</i>	-.123 (.150)	.218 (.330)	-.280 (.185)	2.040 (35.357)	-.363 (.627)
<i>Back Industrial View</i>	.210 (.136)	.203 (.196)	-.340 (.282)	-2.200 (2.982)	.838 (.287) [†]
<i>Front Railway View</i>	.630 (.688)	1.326 (1.042)	.518 (1.245)	-9.723 (9.716)	-1.704 (1.661)
<i>Back Railway View</i>	2.530 (1.041) [#]	.813 (2.147)	2.493 (1.303)*	-9.259 (14.210)	-2.896 (2.766)
<i>Front Water View</i>	-.045 (1.030)	3.972 (2.221)*	.3972 (1.264)	-2.192 (4.393)	-16.428 (4.340) [†]
<i>Back Water View</i>	-3.36 (1.053) [†]	1.779 (2.317)	-2.839 (1.440) [#]	-4.741 (5.366)	-8.321 (2.366) [†]
<i>Aircraft Noise</i>	-.0009 (.001)	1.779 (2.317)*	-	-.035 (.324)	.001 (.001)
<i>Traffic Noise</i>	-.0024 (.0007) [†]	-.0023 (.0015)	-.0046 (.0011) [†]	-.0057 (.0025) [#]	.0038 (.0016) [#]
σ^2 (OLS)	.0430	.0394	.0390	.0571	.0325
Observations	3544	859	1696	370	619
R^2	.778	0.655	0.499	.589	0.696
Adj R^2	.775	0.638	0.487	.538	0.675

B Baseline Category
[†] Significant at the 1% level
[#] Significant at the 5% level
* Significant at the 10% level

The process by which properties are grouped into clusters is known as *cluster analysis*. Cluster analysis divides a dataset into groups (clusters) of observations that are similar to each other. There are two basic approaches to cluster analysis, *partitioning methods* and *hierarchical methods*. With both methods, the researcher determines the M characteristics that are to be used to cluster the observations. Here properties were characterised by their grid reference (longitude and latitude), their proximity to the city centre (car travel time), their selling price, their structural characteristics (dimensions, property type and property age) and the characteristics of the neighbourhood (OA and PD factors). Each observation then can be plotted in M -space according to how highly it scores on each of these M characteristics. Clearly, observations holding similar values for the different characteristics will be located close to each other in this M -space.

With partitioning methods the researcher decides upon the number of clusters *a priori*. Let us denote this number of clusters k . The partitioning algorithm seeks to find k locations in M -space, known as medoids, such that the sum of the distances between each observation and its nearest medoid is minimised. Once the k medoids have been determined the observations are partitioned into clusters by assigning each observation to its nearest medoid.

Hierarchical methods work in a somewhat different manner. With a bottom-up approach, each observation is initially considered as a small cluster by itself. As a first step the two observations lying closest together are merged into a new cluster. At each subsequent step, the two nearest clusters are combined to form one larger cluster. Clusters are merged until one large cluster remains containing all the observations. Alternatively, hierarchical algorithms exist that start with one large cluster and split this into two smaller clusters and continue splitting until each observation forms a cluster of its own. Either way, the final result is a hierarchy of association appearing much like an inverted tree. The researcher can plot this hierarchy and determine which branches of the hierarchy should be treated as separate clusters.

The advantage of hierarchical methods is that they do not impose any *a priori* assumptions on the pattern of association in the observations. The drawback with these methods, however, is that they are computationally burdensome with large data sets.

Here we propose a hybrid method. In the first step, the 3,544 observations were clustered into 100 groups using a partitioning method. In the second step, the average values for the characteristics of each cluster were calculated and these

were analysed using a hierarchical method.⁶ The hierarchical plot of this analysis is presented in Figure 2. The labels at the end of each branch refer to the 100 groups generated in the first step. From the hierarchical plot we have identified eight distinct clusters of observations.

Whilst, the cluster analysis identifies groups of properties that are similar locationally, structurally and in their neighbourhood characteristics, it does not indicate whether these clusters represent separate submarkets. Consequently, separate hedonic price functions were estimated for each cluster of properties using OLS. Following Allen *et al.*, (1995) we compare these regression equations using a Chow test. The Chow test identifies whether there is a significant difference between a pair of regression equations under the null hypothesis that the two models are equivalent. The test statistic is given by;

$$\hat{F} = \frac{(SSR_C - (SSR_1 + SSR_2))}{(SSR_1 + SSR_2)} \times \frac{((N_1 + N_2) - (K_1 + K_2))}{Min(K_1, K_2)} \quad (4)$$

where; SSR_1 , SSR_2 and SSR_C are the sum of squared residuals for the individual models and the combined model and N_1 , N_2 and K_1 , K_2 are the number of observations and number of parameters in the individual models⁷ respectively. The test statistic, \hat{F} , has an F distribution with $Min(K_1, K_2)$, $(N_1 + N_2) - (K_1 + K_2)$ degrees of freedom.

Table 11 reports the results of this test for the 8 clusters of properties. Clearly, in a number of cases, it is not possible to reject the hypothesis of equivalence of hedonic price functions (test scores italicised). We selected the combination of clusters showing the greatest likeness (judged by the Chow test statistic) and combined them into one larger cluster of properties. In this case, cluster 2 and 8 produced the lowest test statistic and hence were merged for the second round. The tests statistics were recalculated using the 6 remaining clusters and the new combined cluster. Again, it was not possible to reject the hypothesis of equivalent hedonic price functions for a number of combinations of clusters. Again the pair of clusters showing the most similarity were merged. This process was repeated until all the test statistics were significant at a 1% level of confidence or greater.

⁶ All cluster analysis was carried out using the excellent "Cluster Package" written in the R programming language.

⁷ Note that not all models contained the same number of parameters. For example, clusters 1, 2, 3 and 4 contained no detached properties such that this indicator variable was not included in the hedonic price functions for these clusters.

Figure 2: Submarket determination using hierarchical cluster analysis

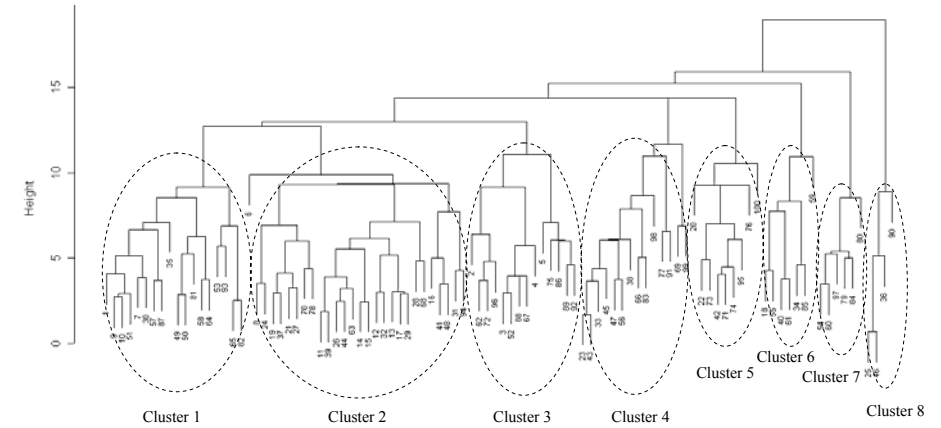


Table 11: Chow test F-Statistics for differences between hedonic price functions of property clusters (italicised scores show cases where hedonic price functions are equivalent)

Cluster	Obs	Cluster						
		1	2	3	4	5	6	7
1	859							
2	1586	3.722 [†]						
3	338	2.487 [†]	3.219 [†]					
4	291	1.847 [†]	2.637 [†]	2.014 [†]				
5	100	2.437 [†]	2.336 [†]	.982	.667			
6	228	2.311 [†]	1.894 [†]	1.408*	1.163	1.446 [#]		
7	32	1.832 [†]	2.489 [†]	1.335	1.894 [†]	1.776 [#]	2.933 [†]	
8	110	.677	.655	.944	.662	1.657 [#]	1.123	4.334 [†]

[†] Significant at the 1% level

[#] Significant at the 5% level

* Significant at the 10% level

The results of the final round of tests are provided in Table 12. Here the 8 original clusters have been reduced to 4 clusters. According to the Chow test, the hedonic price function estimated for each of these clusters of properties is significantly different from the hedonic price function of each of the other clusters of properties. These hedonic price functions are reported in the final four columns of Table 10.

Table 12: Chow test F-statistics for differences between hedonic price functions of property submarkets

Submarket	Obs	Clusters in Submarket	Submarket		
			1	2	3
1	859	1			
2	1696	2,8	3.704 [†]		
3	370	3,7	2.606 [†]	3.346 [†]	
4	619	4,5,6	3.492 [†]	4.183 [†]	2.782 [†]

[†] Significant at the 1% level

An important assumption in applying the Chow test as Allen *et al.*, (1995) have prescribed is that the disturbance variance is the same across regression models. A quick glance across the estimates of σ^2 in table 10 indicates that, in this case, this is unlikely to be true. The disturbance variance from the hedonic price regression for the third cluster of properties is considerably larger than that for the other clusters. When comparing the other models with that estimated for the third property cluster, we must assume that the combined model is heteroscedastic. In these circumstances, it is likely that we shall overestimate the significance of differences in the parameter estimates of the models.

A number of alternative tests suggest themselves. Here we employ a simple Wald test to confirm the results of Table 12. Provided the samples are reasonably large, the Wald test will be valid whether or not the disturbance variances are the same.

Let us denote the sets of parameters from the two models being compared as \mathbf{b}_1 and \mathbf{b}_2 and their accompanying variance matrices \mathbf{V}_1 and \mathbf{V}_2 . Now, under the null hypothesis that the two clusters of properties are independent samples drawn from the same property market we can conclude that \mathbf{b}_1 and \mathbf{b}_2 are normally distributed estimators of the same population parameters, $\boldsymbol{\beta}$. Thus, the expected value of $\mathbf{b}_1 - \mathbf{b}_2$ will be zero and the variance of the difference in the parameter estimates will be $\mathbf{V}_1 + \mathbf{V}_2$ (since the clusters are independent samples). The Wald statistic is given by;

$$W = (\mathbf{b}_1 - \mathbf{b}_2)' (\mathbf{V}_1 + \mathbf{V}_2)^{-1} (\mathbf{b}_1 - \mathbf{b}_2) \quad (5)$$

which has a chi-squared distribution with k degrees of freedom, where k is the number of parameters in common between the two models.⁸ The results of the Wald test are presented in Table 13.

All the test statistics are highly significant confirming the conclusions of Table 12 and supporting the assumption that the four clusters represent properties being traded in distinct property submarkets of the Glasgow urban area.

⁸ Parameters unique to one of the models being tested were dropped from the calculation of the statistic.

Table 13: Wald test Chi-squared statistics for differences between the hedonic price functions of property submarkets

Submarket	Obs	Clusters in Submarket	Submarket		
			1	2	3
1	859	1			
2	1696	2,8	151.4 [†]		
3	370	3,7	91.5 [†]	116.7 [†]	
4	619	4,5,6	277.8 [†]	333.4 [†]	180.7 [†]

[†] Significant at the 1% level

To better understand the character of these four submarkets, Table 14 provides a list of the mean values of some important characteristics of the properties in each submarket. Further, Figure 3 plots the location of the properties in the different submarkets. Notice immediately that whilst the properties in the different submarkets show a good deal of spatial separation, there is also considerable overlap. The submarkets defined by the procedure described here do not result in geographically distinct segmentation of the property market.

Table 14: Mean values of selected characteristics of submarkets

Variable	Submarket			
	1	2	3	4
<i>Log Price</i>	9.934	9.932	10.479	10.714
<i>Log Floor Area</i>	99.64	109.844	196.43	189.221
<i>Garden Size</i>	3.688	26.611	199.936	301.355
<i>Detached</i>162
<i>Semi-Detached</i>368
<i>Terraced</i>449
<i>Subdivided House</i>	.	.	.086	.
<i>Four Block</i>	.	.065	.	.
<i>Flat</i>	.010	.085	.592	.021
<i>Tenement</i>	.983	.848	.295	.
<i>Other</i>	.007	.002	.027	.

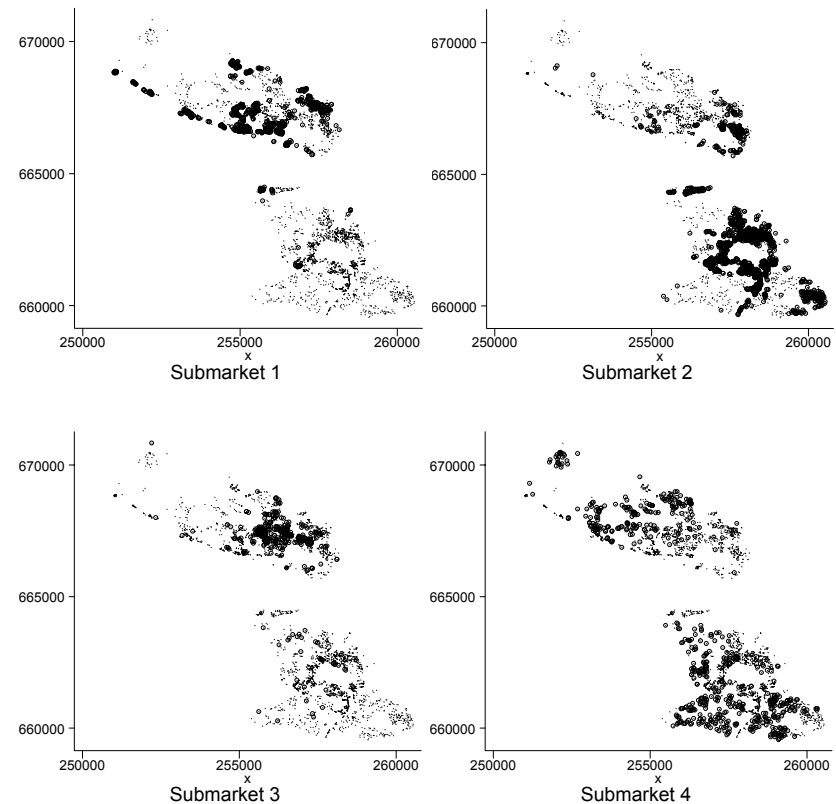
Variable	Submarket			
	1	2	3	4
<i>OA Wealth Factor</i>	-.517	-.413	.259	1.093
<i>OA Ethnicity Factor</i>	-.209	.308	-.411	-.134
<i>OA DINKY Factor</i>	.048	.109	.285	.162
<i>OA Suburbia Factor</i>	-.508	-.389	.182	.743
<i>PC Unemployment Factor</i>	.163	-.729	-.163	-.934
<i>PC YUPPY Factor</i>	.450	.222	1.248	-.471
<i>PC Ethnicity Factor</i>	-.358	.579	-.339	-.121
<i>PC Wealth Factor</i>	.364	-.42	1.212	.388
<i>Traffic Noise</i>	65.403	65.261	64.225	62.715

Submarkets 1 and 2 are characterised by relatively small and cheap properties. The properties in these two submarkets are mostly located in tenement blocks, with a smattering of flats and other property types. Not surprisingly, these submarkets are typified by neighbourhoods that score very poorly on the wealth factor and suburbia factor. At first glance little distinguishes these two submarkets. Notice from Figure 3, however, that there is considerable geographical separation between submarket one and submarket two. Submarket one is concentrated in the north-west of the city whilst submarket two is mostly concentrated in the southern part of the city. Moreover, one characteristic appears to define the difference between these two submarkets; the ethnicity of their inhabitants. In comparison with all the other submarkets, submarket 2 scores very highly on the factors indicating the ethnic makeup of the property neighbourhoods. It would appear that two submarkets exist in Glasgow for similar types of properties fetching similar prices. These submarkets are geographically distinct and are inhabited by two different ethnic groups, submarket 1 by ethnically Scottish residents, submarket 2 by members of the ethnic minorities.

Submarket 3 exhibits a greater diversity of property types than the first two submarkets. Properties in this submarket comprise a variety of flats, tenements and subdivided houses. The properties are larger, have larger gardens and command a considerably higher price than those in the first two submarkets. The neighbourhoods containing properties in submarket 3 are relatively wealthy with low unemployment. Whilst submarket 3 scores relatively highly on the

suburbia factor, this is not a submarket typified by green-berged, leafy avenues. Rather the clue to the identity of this submarket lies in the very high scores on the DINKY and YUPPY factors. Submarket 3 appears to represent a market consisting of urban dwelling young professionals living mainly in flats to the north of the city centre.

Figure 3: Location of properties in different submarkets



Submarket 4 is the most distinct of the four markets. The properties in this submarket are, in the main, detached, semi-detached or terraced houses. Not surprisingly, they are considerably larger than properties in the other

submarkets, have larger gardens and command much higher prices. This is the affluent suburbia submarket. Neighbourhoods containing properties in these submarkets are wealthy and score highly on the suburbia factor indicating the presence of large properties inhabited by high (frequently dual-) earning families. Table 15 summarises the interpretations of the submarkets.

An important consequence of the existence of property submarkets in Glasgow is that estimates of implicit prices derived from the parameters of the pooled data hedonic price function are likely to be biased. To compare the parameters across submarkets we employ the Tiao-Goldberger (Tiao and Goldberger, 1962) test as suggested by Michaels and Smith (1990).

Table 15: Descriptions and interpretations of the submarkets in Glasgow

Submarket		Brief Description
Number	Name	
1	White Tenements	Tenements and some flats, located to the north-west of Glasgow City centre, in relatively poor ethnically Scottish neighbourhoods with high unemployment.
2	Ethnic Minority Tenements	Tenements and some flats, located to the south of Glasgow City centre, in relatively poor neighbourhoods distinguished by high concentrations of ethnic minorities.
3	Urban Young Professionals	Mostly flats, tenements and subdivided houses located in relatively affluent areas to the north of the city centre, in neighbourhoods characterised by relatively young urban dwelling professionals.
4	Affluent Suburbia	Detached, semi-detached and terraced houses mostly located on the fringes of the city in suburban areas. Properties are large, expensive and with large gardens and are located in relatively wealthy neighbourhoods.

The Tiao-Goldberger test considers the null hypothesis that the coefficients of a particular regressor take the same value in each of the models. In effect, the test works by comparing each parameter estimate to the weighted sum of parameter estimates across all four models. The sum of the squared differences between

each parameter and this average value form the core of the test statistic. Following the notation of Michaels and Smith (1990), the Tiao-Goldberger test statistic is given by;

$$F^{TG} = \frac{\sum_{j=1}^L (b_{ji} - \bar{b}_i)^2 / P_{ji}}{\sum_{j=1}^L SSR_j} \times \frac{\sum_{j=1}^L (T_j - K_j)}{(L-1)} \quad (6)$$

where

$$\bar{b}_i = \frac{\sum_{j=1}^L (b_{ji} / P_{ji})}{\sum_{j=1}^L (1/P_{ji})}$$

and L is the number of models; b_{ji} is the OLS estimate of the i^{th} parameter in the j^{th} independent model; P_{ji} is the diagonal element of the i^{th} parameter of $(X'X)^{-1}$ in the j^{th} model; SSR_j is the sum of squared residuals for the j^{th} model; T_j is the number of observations used to estimate the j^{th} model and K_j is the number of parameters in the j^{th} model. The test statistic has an F distribution with $(L-1), \sum_{j=1}^L (T_j - K_j)$ degrees of freedom.

Table 16 reports the parameter estimates and the Tiao-Goldberger test statistics for a number of parameters that were estimated across all four submarkets.

For many of the variables listed in Table 16 we can reject the hypothesis of stability of parameter estimates across submarkets. For example, observe the parameter estimates on the log of floor area. The Tiao-Goldberger statistic is significant at a 5% level of confidence and we conclude that the parameters across the four submarkets are different. On closer inspection it is clear that the implicit price of floor area is greatest in submarket 4 (affluent suburbia); households purchasing properties in this submarket pay more for relatively larger properties than households in the other submarkets. We must interpret this result with caution. As described in section 2, the market clearing implicit prices are the result of both demand and supply conditions in that submarket. Thus, if we were to assume that the supply of properties of different floor areas were the same in all submarkets, then it would be reasonable to conclude that there was greater demand for larger properties in submarket 4. Alternatively, if

we were to assume that demand for larger properties was constant over all submarkets, we would be forced to conclude that the availability of larger properties was restricted in submarket 4. Intuition suggests that the differences in the implicit price of floor area across submarkets are more likely to be driven by demand side factors, though clearly the truth will lie somewhere between the two extremes.

Table 16: Tiao-Golberger test F statistics for selected parameters from the four hedonic price functions of property submarkets

Parameter	Submarket				F^{TG}
	White Tenements	Ethnic Minority Tenements	Urban Young Professionals	Affluent Suburbia	
<i>Intercept</i>	9.646 [†]	9.477 [†]	8.902 [†]	9.152 [†]	2.616 [#]
<i>Log Floor Area</i>	.089	.156 [†]	.142 [†]	.228 [†]	3.467 [#]
<i>Garden Size</i>	.044 [#]	.549 [†]	.121 [#]	.164 [†]	3.448 [#]
<i>OA Wealth Factor</i>	.166 [†]	.098 [†]	.100 [†]	.127 [†]	2.274 [*]
<i>OA Ethnicity Factor</i>	-.058 [†]	-.040 [†]	-.065 [#]	-.100 [†]	3.299 [#]
<i>OA Tenement Factor</i>	-.080 [†]	-.050 [†]	-.031	-.054 [†]	3.551 [#]
<i>OA DINKY Factor</i>	.052 [†]	.056 [†]	.0004	.013	6.168 [†]
<i>OA Suburbia Factor</i>	.060 [†]	.066 [†]	.044 [#]	.076 [†]	.601
<i>PC Unemployment Factor</i>	-.040 [#]	-.044 [†]	-.125 [#]	-.026	1.534
<i>PC YUPPY Factor</i>	.023 [*]	.025 [#]	.014	.100 [†]	6.13 [†]
<i>PC Ethnicity Factor</i>	-.033	.011	.019	.055 [†]	3.449 [#]
<i>PC Wealth Factor</i>	.079 [†]	.086 [†]	.078 [†]	.059 [†]	1.214

Parameter	Submarket				F^{TG}
	White Tenements	Ethnic Minority Tenements	Urban Young Professionals	Affluent Suburbia	
<i>Walk Rail</i>	.0190	.057 [#]	.124 [#]	-.023	2.908 [#]
<i>Walk Park</i>	-.075 [*]	.037	-.040	-.075 [#]	3.023 [#]
<i>Walk Shop</i>	.092	-.006	-.007	.148 [†]	1.834
<i>Walk School</i>	.091	.033	-.019	.024 [†]	.742
<i>Car City Centre</i>	-.015 [#]	-.006	.053 [†]	.014 [#]	7.763 [†]
<i>Front View</i>	.011	.006	.046	.054 [†]	2.614 [#]
<i>Back View</i>	-.005	.004	-.079 [†]	-.041 [#]	6.699 [†]
<i>Traffic Noise</i>	-.0023	-.0046 [†]	-.0057 [#]	.0038 [#]	6.361 [†]

[†] Significant at the 1% level

[#] Significant at the 5% level

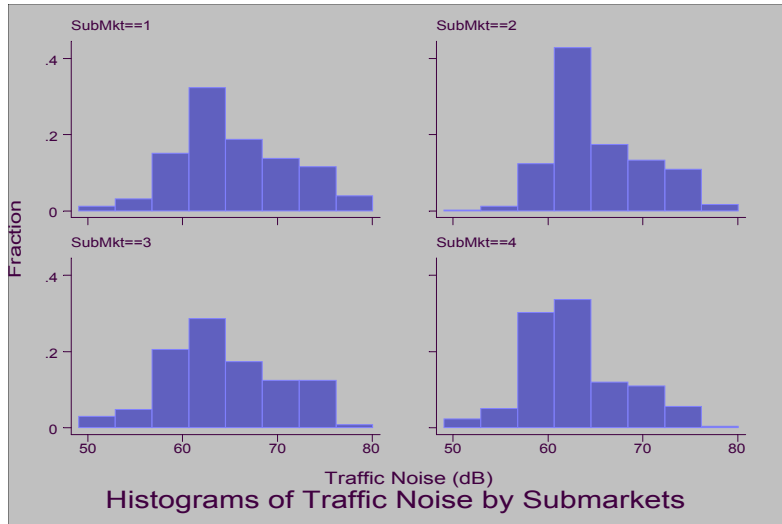
^{*} Significant at the 10% level

Of most interest in this research are the differences in the implicit price of traffic noise. From Table 16 we observe that the Tiao-Golberger statistic indicates that the hypothesis of equality of parameter estimates can be rejected at the 1% level of confidence. The price of properties are most adversely affected by increasing exposure to traffic noise in submarket 3 (urban young professionals), followed by submarket 2 (ethnic minority tenements), then submarket 1 (white tenements). Contradicting expectations, exposure to road traffic noise appears to increase the price of properties in submarket 4 (affluent suburbia) and this is significant at a 5% level of confidence. To understand better the supply conditions, Figure 4 provides histograms depicting the traffic noise exposure of properties (in the sample) traded in the four submarkets.

Notice in particular the relative abundance of low noise exposure properties in submarket 4. One possible explanation of the greater price paid for traffic noise avoidance in submarkets 1 to 3 might be related to the fact that quiet properties are relatively more difficult to come by in these submarkets

Of course this does not explain the unexpected sign on the traffic noise variable in the hedonic price function for submarket 4. We conclude that, in this submarket at least, traffic noise must be proxying for some other variable that positively influences traffic prices. We do not follow-up on this contention further here.

Figure 4: Distribution of properties with different exposures to traffic noise in the four submarkets



To ensure that the noise parameter estimate for submarket 4 was not the sole cause of the high Tiao-Golberger test statistic score in Table 16, we reapply the test using only data from submarkets 1 to 3. Once again, the test statistic is large ($F^{TG} = 5.178$) leading us to reject the hypothesis of equality of the noise parameter estimates across these three submarkets at a 1% level of confidence.

6. Spatial Hedonic Regression

So far, our statistical analysis of the hedonic price functions in the separate submarkets has ignored the spatial organisation of the data. In effect, we have assumed that the observations of property sales are independent such that we can glean no information on the selling price of a property from the selling price of other properties. Of course, this is hardly likely to be the case. Properties that are located near to each other in space are also likely to share common environmental, accessibility, neighbourhood and perhaps even structural characteristics. Even once we account for the values of known covariates, omitted variables are likely to induce spatial dependence among the errors.

If hedonic residuals are spatially correlated, the parameter estimates from an OLS regression will be inefficient and will produce biased estimates of the standard errors of the parameter estimates. In the case where the residuals are positively spatially correlated, as is to be expected with hedonic property price regressions, OLS will underestimate the population residual variance and the resulting t -statistics will be biased upwards. Whilst OLS parameter estimates remain unbiased, ignoring spatial autocorrelation may lead to erroneously high significance being attached to the influence of property attributes on selling prices.

Over recent years, the existence of spatial autocorrelation has received a great deal of attention in the hedonic literature (e.g. Dubin, 1992; Can, 1992; Pace and Gilley, 1997; Basu and Thibodeau, 1998; Bell and Bockstael, 2000). In the main, researchers have focused on the spatial error dependence model that can be expressed as;

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (7a)$$

$$\text{where } \boldsymbol{\varepsilon} = \lambda \mathbf{W}\boldsymbol{\varepsilon} + \mathbf{u} \quad (7b)$$

where \mathbf{y} , \mathbf{X} and $\boldsymbol{\beta}$ are defined as in (3) but $\boldsymbol{\varepsilon}$ is now an $[N \times 1]$ vector of random error terms with mean 0 and a non-spherical variance-covariance matrix $\sigma^2(\mathbf{I} - \lambda \mathbf{W})^{-1}(\mathbf{I} - \lambda \mathbf{W}')^{-1}$. The nature of the spatial autocorrelation is defined by equation (7b). Here \mathbf{W} is an $[N \times N]$ weighting matrix, λ is the error dependence parameter to be estimated and \mathbf{u} is the usual $[N \times 1]$ vector of random error terms with expected value 0 and variance-covariance matrix $\sigma^2 \mathbf{I}$.

The error in the spatial error dependence model, therefore, is made up of two parts; a purely random element and an element containing a weighted sum of

the errors on nearby properties. The association between one property and another is contained in the weighting matrix, W . The diagonal elements of the weighting matrix are zero since, clearly, the error for an observation cannot be used to explain itself. The off-diagonal elements of the matrix represent the potential spatial dependence between observations. Thus if the ij^{th} element of the weighting matrix, w_{ij} , is zero, we are assuming that there is no correlation in the errors of the i^{th} and j^{th} observations. Conversely if w_{ij} takes on a non-zero value we are assuming that there is correlation in the errors of these two observations.

The researcher must stipulate the nature of dependence between observations by defining the weights matrix in advance of estimation. Here we experimented with a variety of weights matrices but the final specification of (7) used a weights matrix in which it was assumed that properties separated by more than 100 metres were unrelated. Moreover the non-zero elements of the weights matrix were defined as the inverse of the squared distance between properties. This format for the weights matrix allows for the error terms to be more closely correlated with the error terms of close neighbours than with the error terms of more distant neighbours.

Following normal procedure, W was row standardised such that each row's elements were made to sum to one. When W is row standardised, the product $W\epsilon$ equals $\sum_j w_{ij}\epsilon_j$, and has an intuitive interpretation; it is simply a vector of weighted averages of the errors of neighbouring observations. As Bell and Bockstael (2000) point out, row standardisation is undertaken to simplify estimation of the model. There is usually no underlying economic story supporting the procedure. Moreover, the spatial dependence parameter λ estimated on a row standardised weights matrix must be interpreted with caution. In particular, λ in this case is not directly equivalent to an autocorrelation coefficient.

The characteristics of the weights matrices constructed for the property sales observations in the four submarkets are detailed in columns 2 to 4 of Table 17. Even with a relatively restrictive 100 metre cut-off, the majority of properties are associated with other properties in the same submarket. In submarket 2, for example, only 46 properties out of the 1,696 observations were further than 100 metres from another property in the sample. On average in this submarket, each property was located within 100 metres of 16 other properties in the sample. Notice that the number of associations in submarket 4 is somewhat lower than in the other submarkets and that this is not entirely explained by sample size. One explanation of this observation is that properties in suburbia are more greatly dispersed than those in the other submarkets.

The final columns of Table 17 report two tests of spatial dependence based on the Lagrange multiplier principle that can be calculated using OLS residuals and the spatial weights matrices⁹. The first, suggested by Burridge (1980), tests the null hypothesis that $\lambda = 0$. That is, the hypothesis that there is no spatial dependence between error terms. The test statistics reported in column 5 are chi-squared distributed with one degree of freedom. The null hypothesis is rejected with a high degree of confidence in all four submarkets, supporting the contention that the OLS residuals exhibit spatial autocorrelation.

An alternative model of spatial dependence stipulates a dependent variable that is functionally related to the value of the dependent variable of neighbouring observations¹⁰. In this case our model would include spatially lagged values of the dependent variable rather than the error term. This model is variously known as substantive spatial dependence, structural dependence or spatial autoregression. A test of this model (suggested by Anselin; 1988) is provided in the final column of Table 17. In all but one submarket we cannot reject the null hypothesis of no substantive spatial dependence. Again the data would seem to suggest that the correct model is that presented in (7).

Table 17: Characteristics of the spatial weights matrices and tests of spatial dependence

Submarket	Characteristics of the Spatial Weights Matrix			Tests of Spatial Dependence	
	Obs.	Unassociated Obs.	Average Association nsper Obs.	LM test for Spatial Error Dependence	LM test for Substantive Spatial Dependence
1. White Tenements	859	28	13.84	20.50 [†]	1.26
2. Ethnic Minority Tenements	1696	46	16.02	65.67 [†]	4.32 [#]
3. Urban Young Professionals	370	49	5.35	3.98 [#]	.02
4. Affluent Suburbia	619	116	2.18	18.56 [†]	.43

⁹ Computational details can be found in Anselin and Hudak (1992).

¹⁰ For an excellent and accessible of this and other models of spatial dependence see Anselin (1993).

The spatial dependence model in (7) can be estimated using maximum likelihood (ML) techniques. However, for large samples this may be computationally prohibitive. Instead we follow Bell and Bockstael (2000) and use the generalised moments (GM) estimator developed by Kelejian and Prucha (1999). As Bell and Bockstael (2000) describe, whilst this estimator may not be as efficient as the ML estimator it possesses two advantages. First, the calculation of the estimator is fairly straightforward even with extremely large samples. And second, the GM estimator is consistent even when the error terms u are not normal. The parameter estimates from the GM estimator of model (7) applied to the data for the four property submarkets in Glasgow are listed in Table 18.

As expected, the parameter estimates do not significantly differ from those reported in Table 10. Again we focus on the implicit price of traffic noise avoidance.

- In submarket 1 the parameter becomes more negative, falling from a value -.0023 in the OLS model to -.0030 in the spatial dependence model. Where the OLS result proved statistically insignificant, we can reject the hypothesis that the parameter in the spatial dependence model is actually zero with over 93% confidence.
- There is little change in the estimates for submarket 2 between the two models. The parameter is -.0046 in the OLS model and -.0047 in the spatial dependence model. Both are highly statistically significant.
- In submarket 3 the actual parameter value changes little between the two models (-.0057 in the OLS model, -.0058 in the spatial dependence model). However, the standard error in the spatial dependence model is somewhat smaller. Indeed, in the spatial dependence model we can reject the hypothesis of a zero value for the noise parameter with over 99% confidence.
- In submarket 4, the parameter on traffic noise took a value of .0038 in the OLS model. The unexpectedly positive value for this parameter was significant at the 95% level of confidence. Reassuringly, the parameter value falls to a value of .0031 in the spatial dependence model and is no longer significant at the 95% confidence level.

It would seem that, in comparison to the OLS mode, the spatial dependence model returns estimates of the parameter on traffic noise that are more in line with prior expectations.

Table 18: Hedonic price functions for property markets in Glasgow estimated using GM estimator accounting for spatial autocorrelation

Independent Variable	Submarkets			
	One	Two	Three	Four
<i>Intercept</i>	9.776 (.186) [†]	9.588 (.143) [†]	8.883 (.267) [†]	9.168 (.172) [†]
<i>Log Floor Area</i>	.079 (.024) [†]	.140 (.020) [†]	.120 (.038) [†]	.236 (.033) [†]
<i>Garden Size</i>	-.020 (.204)	.541 (.135) [†]	.122 (.043) [†]	.131 (.041) [†]
<i>Shape</i>	.008 (.004) [#]	.008 (.003) [#]	.009 (.005) [*]	-.002 (.003) [†]
<i>Storeys</i>	-	-	-	-.157 (.031) [†]
<i>Detached</i>	-	-	-	B
<i>Semi-Detached</i>	-	-	-	.067 (.021) [†]
<i>Terraced</i>	-	-	-	-.142 (.038) [†]
<i>Subdivided House</i>	-	-	-.034 (.062)	-
<i>Four Block</i>	-	-.012 (.044)	-	-
<i>Flat</i>	.201 (.069) [†]	-.037 (.022) [*]	-.034 (.033)	-.797 (.140) [†]
<i>Tenement</i>	B	B	B	-
<i>Other</i>	.122 (.086)	-.256 (.111) [#]	.031 (.083)	-
<i>Building Material</i>	.027 (.064)	-.029 (.099)	.233 (.169)	.100 (.030) [†]
<i>Age (pre 1919)</i>	.077 (.085)	-.051 (.113)	.170 (.177)	.005 (.047)
<i>Age (1919-45)</i>	-.063 (.094)	-.027 (.082)	.298 (.199)	-.021 (.043)
<i>No. of Properties in Building</i>	.001 (.004)	-.012 (.004) [†]	-.010 (.005) [#]	.003 (.002)

Independent Variable	Submarkets			
	One	Two	Three	Four
<i>Basement Flat</i>	-.023 (.191)	-.089 (.065)	-.162 (.051) [†]	-
<i>Grdth Floor Flat</i>	B	B	B	B
<i>1st Floor Flat</i>	.036 (.02)*	.045 (.014) [†]	-.007 (.037)	.136 (.146)
<i>2nd Floor Flat</i>	-.006 (.02)	.026 (.014)	.013 (.036)	-.059 (.141)
<i>3rd Floor Flat</i>	-.024 (.021)	.018 (.015)	-.065 (.047)	
<i>OA Wealth Factor</i>	.173 (.026) [†]	.105 (.016) [†]	.103 (.018) [†]	.124 (.013) [†]
<i>OA Ethnicity Factor</i>	-.046 (.015) [†]	-.036 (.007) [†]	-.062 (.026) [†]	-.102 (.018) [†]
<i>OA Bedsit Factor</i>	.010 (.017)	-.002 (.011)	-.004 (.012)	-.034 (.014) [#]
<i>OA Tenement Factor</i>	-.075 (.009) [†]	-.046 (.006) [†]	-.036 (.02)*	-.051 (.017) [†]
<i>OA DINKY Factor</i>	.051 (.009) [†]	.053 (.007) [†]	-.004 (.017)	.022 (.011) [#]
<i>OA Suburbia Factor</i>	.065 (.015) [†]	.067 (.010) [†]	.056 (.020) [†]	.070 (.014) [†]
<i>PC Unemployment Factor</i>	-.039 (.018) [#]	-.044 (.015) [†]	-.122 (.043) [†]	-.025 (.020)
<i>PC YUPPY Factor</i>	.025 (.014)*	.032 (.012) [†]	.016 (.021)	.092 (.016) [†]
<i>PC Ethnicity Factor</i>	-.039 (.023)*	.010 (.009)	.032 (.033)	.055 (.019) [†]
<i>PC Wealth Factor</i>	.085 (.017) [†]	.091 (.012) [†]	.086 (.026) [†]	.065 (.01) [†]
<i>Walk Rail</i>	.008 (.029)	.05 (.031)	.137 (.054) [#]	-.037 (.027)
<i>Walk Park</i>	-.082 (.048)*	.037 (.03)	-.050 (.071)	-.078 (.037) [#]
<i>Walk Shop</i>	.102 (.086)	. (.053)	-.014 (.097)	.14 (.052) [†]
<i>Walk School</i>	.088 (.057)	.04 (.038)	-.023 (.059)	.02 (.038)

Independent Variable	Submarkets			
	One	Two	Three	Four
<i>Car City Centre</i>	-.015 (.007) [#]	-.007 (.006)	.058 (.016) [†]	.016 (.007) [#]
<i>Front View</i>	.014 (.010)	.007 (.009)	.049 (.025) [#]	.053 (.017) [†]
<i>Back View</i>	-.003 (.011)	.006 (.009)	-.083 (.019) [#]	-.038 (.02) [#]
<i>Front Park View</i>	.968 (.44) [#]	-.098 (.152)	-.094 (.28)	-.283 (.237)
<i>Back Park View</i>	.144 (.517)	-.21 (.262)	.887 (1.113)	-1.363 (.736)*
<i>Front Industrial View</i>	.045 (.356)	-.225 (.197)	-3.015 (33.318)	-.451 (.637)
<i>Back Industrial View</i>	.199 (.198)	-.361 (.287)	-2.194 (2.805)	.863 (.291) [†]
<i>Front Railway View</i>	1.207 (1.044)	.81 (1.374)	-10.293 (9.005)	-1.608 (1.592)
<i>Back Railway View</i>	.443 (2.158)	2.123 (1.413)	-15.104 (13.25)	-2.212 (2.795)
<i>Front Water View</i>	4.831 (2.362) [#]	.337 (1.251)	-2.597 (4.021)	-14.959 (4.138) [†]
<i>Back Water View</i>	1.386 (2.411)	-2.271 (1.541)	-4.206 (4.893)	-6.153 (2.38) [†]
<i>Aircraft Noise</i>	-.004 (.002)*		.009 (.305)	.001 (.002)
<i>Traffic Noise</i>	-.0030 (.0016)*	-.0047 (.0012) [†]	-.0058 (.0022) [†]	.0031 (.0016)*
ρ	.214 (.000) [†]	.249 (.000) [†]	-.144 (.000) [†]	.223 (.000) [†]
σ^2	.0363	.0363	.0496	.0285
Observations	859	1696	370	619
R^2	0.637	0.488	0.607	0.702

B Baseline Category
[†] Significant at the 1% level
[#] Significant at the 5% level
* Significant at the 10% level

6. Conclusions

Property markets are essentially spatial in nature and this fact should not be ignored in hedonic analyses of property prices. This case study has illustrated some important analytical techniques that can be used in hedonic analysis of property markets that take account of spatial considerations. First and foremost, GIS proves to be an extremely powerful tool for compiling data for the estimation of hedonic price functions. Researchers can collect data sets rich in information on the structural, accessibility, neighbourhood and environmental characteristics of properties from the comfort of their own desk.

Further, theory suggests that the property market will not be a homogenous entity. Rather, it will be characterised by segmentation, with the implicit prices of property characteristics differing across market segments according to the conditions of supply and demand for characteristics prevailing in each market segment. In the past researchers have imposed a number of criteria (geographical, structural or socioeconomic) in order to define these submarkets. Here we propose an alternative approach based on cluster analysis that allows the data themselves to dictate the pattern of market segmentation. The results of this analysis suggest property submarkets with intuitively appealing interpretations. These submarkets are not defined by one criteria but a combination of spatial, structural and socioeconomic characteristics.

Finally, we illustrate the application of an estimation technique that explicitly allows for the spatial relation between properties in the sample. In general, the results of the hedonic analysis in the separate submarkets concord with prior expectations. In particular, the implicit price for the avoidance of traffic noise is negative and significant in three out of four submarkets. Moreover, these implicit prices are shown to be statistically different supporting the contention that separate hedonic price schedules rule in the different submarkets.

References

- Allen, M.T., Springer, T.M., and Waller, N.G. (1995). "Implicit pricing across residential rental markets", *Journal of Real Estate Finance and Economics*, 11, pp 137-151.
- Anselin, L., (1988). "Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity", *Geographical Analysis*, 20, pp 1 – 17.
- Anselin, L., and Hudak, S., (1992). "Spatial econometrics in practice: A review of software options", *Regional Science and Urban Economics*, 2, pp 509 – 536.
- Anselin, L., (1993). "Discrete space autoregressive models" pp 454-68 in M. F. Goodchild, B. O. Parks and L. T. Steyaert (eds), *Environmental Modelling with GIS*, Oxford University Press, Oxford.
- Bateman, I.J., Day, B.H., Lake, I., and Lovett, A.A., (2001). *The Effect of Road Traffic on Residential Property Value: A Literature Review and Hedonic Pricing Study*, Edinburgh: Scottish Executive and The Stationary Office.
- Ball, M.J. and Kirwan, R.M., (1977). "Accessibility and supply constraints in the urban housing market", *Urban Studies*, 14, pp 11-32.
- Basu, S. and Thibodeau, T.G., (1998). "Analysis of spatial autocorrelation in housing prices", *Journal of Real Estate Finance and Economics*, 17, pp 61-95.
- Bell, K. and Bockstael, N.E., (2000). "Applying the generalised method of moments approach to spatial problems involving micro-level data", *Review of Economics and Statistics*, 82(1), pp 72-82.
- Burridge, P., (1980). "Testing for spatial autocorrelation among regression residuals", *Environment and Planning A*, 13, pp 795-800.
- Can, A., "Specification and estimation of hedonic housing price models," *Regional Science and Urban Economics*, 22, pp 453-474.
- Day, B.H., (2001). "The theory of hedonic markets: Obtaining welfare measures for changes in environmental quality using hedonic market data", *Report to EU Working Group on Noise*, CSERGE, University College London, UK.
- Department of Transport (Welsh Office) (1988). *Calculation of Road Traffic Noise*. London: HMSO.
- Dubin, R.A., (1992). "Spatial autocorrelation and neighbourhood quality", *Regional Science and Urban Economics*, 22, pp 433-452
- Goodman, A.C. (1978). "Hedonic prices, price indexes and housing markets", *Journal of Urban Economics*, 5, pp 471-484.
- Harrison D. and Rubinfeld, D.R., (1978). "Hedonic housing prices and the demand for clean air", *Journal of Environmental Economics and Management*, 5, pp 81-102.

- Kelejian, H.H., and Prucha, I.R., (1999). "A generalised moments estimator for the autoregressive parameter in a spatial model", *International Economic Review*, 40(2), pp 509-533.
- Lake, I.R., Lovett, A.A., Bateman, I.J., and Langford, I.H., (1998) 'Modelling environmental influences on property prices in an urban environment', *Computers Environment and Urban Systems*, 22, pp. 121-136.
- Lindeman, R.H., Merenda, P.F., and Gold, R., (1980). "Chapter 8: Factor Analysis", in *Introduction to Bivariate and Multivariate Analysis*, Scott, Foresman and co.
- Martin, D., Longley, P., and Higgs, D., (1994). "The use of GIS in the analysis of diverse urban datasets", *Computers, Environment and Urban Systems*, 18(1), pp 55-56.
- Michaels, R. G., and V. K. Smith (1990). "Market-segmentation and valuing amenities with hedonic models - the case of hazardous-waste sites", *Journal of Urban Economics*, 28 (2), pp 223-242.
- Ordnance Survey (1996). *Digital Map Data and Customised Services*, Catalogue produced by the Ordnance Survey, Southampton.
- Pace, R.K., and Gilley, O.W., (1997). "Using the spatial configuration of data to improve estimation", *Journal of Real Estate Finance and Economics*, 14(3), pp 333-340.
- Schnare, A., and Struyk, R., (1976). "Segmentation in Urban Housing Markets", *Journal of Urban Economics*, 4, pp 146-166.
- Sonstelie, J.C., and Portney, P.R., (1980). "Gross rents and market values: Testing the implications of Trebout's hypothesis", *Journal of Urban Economics*, 7, pp 102-118.
- Straszheim, M.R. (1973). "Estimation of the demand for urban housing services from household interview data", *Review of Economics and Statistics*, 55, pp 1-8.
- Tiao, G.C., and Goldberger, A. (1962). "Testing equality of individual regression coefficients", WEBH Paper 6201, University of Wisconsin, Social Systems Research Institute.