HEDONIC MODELLING OF HOUSING MARKETS USING GEOGRAPHICAL INFORMATION SYSTEM (GIS) AND SPATIAL STATISTICS: A CASE STUDY OF GLASGOW, SCOTLAND

By

DR. SURIATINI ISMAIL*
Fakulti Kejuruteraan & Sains Geoinformasi
Universiti Teknologi Malaysia
81310 Skudai, Johor Bahru, Johor,
Malaysia
e-mail: suriatini@fksg.utm.my

PROFESSOR BRYAN D. MACGREGOR
Department of Property
University of Aberdeen Business School
Edward Wright Building
AB24 3QY Aberdeen
Scotland
e-mail: b.d.macgregor@abdn.ac.uk

ABSTRACT

This paper presents the results of a simultaneous consideration of detailed accessibility measures and spatial autocorrelation in house price hedonic modelling. It illustrates the application of GIS and spatial statistics in the estimation of hedonic models for the entire housing market in Glasgow, Scotland, using 2,715 house prices for 2002 and 61 independent variables. GIS is used in this study to construct spatial variables including detailed accessibility measures, to help detect spatial autocorrelation, and for map visualisation. Spatial statistics are used to test formally and model explicitly the spatial autocorrelation. The results suggest that an individual accessibility measure is more influential than a zonal accessibility measure because the former is able to capture the micro effect of location on house price. Furthermore, the application of spatial statistics can produce more accurate, robust and reliable estimates of implicit prices.

Keywords: Hedonic modelling, house prices, GIS, spatial statistics.

Dr Suriatini Ismail is lecturer at UTM Skudai Malaysia and research fellow in the Aberdeen University Business School. Professor MacGregor is Chair of Property in the University of Aberdeen and Editor of Journal of Property Research. The authors would like to thank David R Green of University of Aberdeen for his GIS expertise lent throughout this study. Thanks are also due to all the data providers and individuals who have involved in this study. * corresponding author.

1 BACKGROUND

The property market consists of residential (or housing), commercial, and agricultural sectors. **Housing markets** constitute a major component of the real estate market. Second-hand units constitute considerably larger portion than the supply of new-build at any time (Leishman, 2003, 115). According to Hwang and Quigley (2004), owner-occupied housing is a substantial fraction of aggregate wealth in most advance countries. Thus, house prices provide important economic indicators. For example, in the UK, the housing sector accounts for slightly more than half of the nation's fixed capital stock (Muth and Goodman, 1989). It occupies about 40 percent of UK urban land (Field and MacGregor, 1987, 54). The importance of house prices as a leading economic indicator (Barret and Blair, 1982, 168) nationally and locally indicates that house price analysis is an important aspect of property economics.

House prices are an important consideration when assessing macroeconomic and financial developments in the UK (Thwaites and Wood, 2003) and other developed countries such as the USA. House price indices are used by the government and private sectors in policy evaluation and implementation. Models of housing prices are commonly estimated on national statistics (Goodman, 1998). This can be based on time series, cross-sectional, or panel (a combination of both) data. For example, in the current UK's economic policy climate, house price appreciation rates are used as a barometer for more general inflationary pressures and are thought to be one of the important indicators consulted by the Monetary Policy Committee (Costello and Watkins, 2002). This normally involves time series analysis. Housing prices also *act as a sensitive barometer for many social phenomena such as crime, congestion, job opportunities, and demographics* (Pace and LeSage, 2004a, 180). This normally involves cross-sectional analysis. Thus, house price analysis is an important element in housing economics.

The use of econometric/economic models including **hedonic modelling** has become an established part of not only the policy framework employed by both the Treasury and the Bank of England (Meen and Meen, 2003), but also of housing market analysis. In his review of hedonic price modelling, Malpezzi (2003, 84) describes hedonic modelling as *having been applied in every permanently inhabited region of the globe.* Indicating the established state of the technique, he concludes that *over the*

past three decades, hedonic estimation has clearly matured from a new technology to become the standard way economists deal with housing heterogeneity (Malpezzi, 2003, 87). Watkins (1998) also notes the dominance of hedonic modelling in the real estate literature.

According to Hoesli and MacGregor (2000, 64), the hedonic method has been widely used in the USA, and also used in other countries such as Switzerland and Taiwan for constructing **price indices**. Lum (2004) also implies that this technique has been used in several Commonwealth countries including Hong Kong and Malaysia. In the UK, the technique is used in the creation of the Nationwide Anglia Building Society and Halifax Bank of Scotland (HBOS) price indices (Lum, 2004; Watkins, 1998), which are the major sources of regional and national house price data in the UK. The Office of Deputy Prime Minister (ODPM) monthly house price index which was launched in September 2003 is also based on hedonic price (Barker Review, 2004). Nonetheless, Can and Megbolugbe (1997) highlight that a major limitation of currently available house price indices constructed based on hedonic price models is their insensitivity to the geographic location of dwellings within the metropolitan area.

House price hedonic analysis is undertaken by regressing usually, the transaction prices of properties against the corresponding property characteristics (Fletcher *et al*., 2000b), which are categorised as structural, accessibility, and neighbourhood in this study. The accessibility and the neighbourhood characteristics comprise mainly location-related factors. Acknowledging the importance of location, Gallimore *et al.* (1996b, 18) state that *...locationally sensitive models are…statistically defensible means of reviewing values and valuation*. Given that house prices are the most widely used measure of property values, this indicates that house price hedonic modelling should consider spatial elements. In addition, according to Orford (2000), if the hedonic house price function is to generate estimates that properly reflect the implicit price of attributes, the model specification must capture sufficiently the spatial elements at the local market level. Therefore, other than Can and Megbolugbe (1997), Gallimore *et al.* (1996b) and Orford (2000) also indicate the importance of proper consideration of spatial elements in hedonic price modelling.

Technically, improper consideration of spatial elements contributes to a substantial portion of the unexplained variability of price in the hedonic model and leads to problems. Des Rosiers *et al*. (2001) outline three main sources of problems to comprise multicollinearity, heteroscedasticity and spatial autocorrelation. While the first two can happen in both time series and cross sectional data, the last one is specifically related to the cross sectional data. Thus, all the three problems can occur in a cross sectional analysis of house prices. Accordingly, it is important to consider these problems in housing market analysis if the results are not to be invalidated. Given that a cross-sectional analysis of house prices involves geographical information, it is important to give attention to the spatial elements. In considering the spatial elements in house price hedonic modelling, suitable tools are required. Two appropriate tools are Geographical Information System (GIS) and spatial statistics.

The applications of **GIS** in real estate were established in the USA. These have started to develop in the UK since the early 1990s. There is evidence of GIS applications in residential, commercial and rural sectors. Residential has shown to be the sector with most of the identified research, particularly since the late 1990s. GIS is a relevant technology for housing markets analysis as all residential real estate information is inherently spatial because housing is fixed in geographic space (Belsky *et al*., 1998). So, spatial data[1] is one of the features of residential property. GIS has the advantages of efficient data integration and spatial analysis (Hamid, 2002).

Spatial analysis functions differentiate GIS from other data management systems. For example, network analysis can improve the practice of distance measurement from merely the straight-line to road network. It also offers a function to calculate minimum travelling time via a transportation network (Des Rosiers *et al*., 2001). The representation of spatial data and model results within a GIS could lead to an improved understanding both of the attributes being examined and of the procedures used to examine them (Fotheringham and Rogerson, 1994). Thus, GIS is relevant to this study because it can deal with spatial elements efficiently. Nevertheless, GIS is not yet a perfect tool for considering spatial elements in housing market analysis. This is because, GIS is conventionally a tool for data handling and thus, in its generic

[1] In this study, the term "spatial data" refers to the map and attributes describing the map.

form, it would not deal with spatial autocorrelation. Recent real estate studies that use GIS revealed the involvement of spatial statistics to deal specifically with spatial autocorrelation. So, a combination of GIS and spatial statistics can be beneficial for effective hedonic modelling of real estate markets.

The literature shows that the study of spatial aspects of hedonic modelling falls under the umbrella of **spatial econometrics[2],** a sub-field of spatial statistics (Anselin, 1988). Anselin (1988, 7) defines spatial econometrics as *the collection of techniques that deal with the peculiarities caused by space in the statistical analysis of regional science models*. According to him, the emphasis on the model as the starting point differentiates spatial econometrics from the broader field of spatial statistics, although they share a common methodology framework. However, this study does not differentiate the two terms and uses them interchangeably. Spatial econometrics (and/or spatial statistics) are relevant to this study because it explicitly accounts for the influence of space in real estate modelling (Wilhelmsson, 2002a). The spatial effects are of two types, namely spatial autocorrelation and spatial heterogeneity[3]. Spatial autocorrelation is a weaker expression for spatial dependence (Wilhelmsson, 2002a). In his review on spatial effects and real estate, Wilhelmsson (2002a) states that before 1990, the problems of the existence of spatial effects have been ignored in real estate analysis. However, they seem to be gaining more attention from researchers in the past few years (Anselin, 2002).

Pace and Barry (1997a) assert that regression is perhaps the most often used technique in statistics. Pointing towards the established state of regression analysis, Kim *et al.*, (2003) note that much research has been carried out to solve specific econometric issues pertaining to hedonic regression such as functional form, identification and statistical efficiency. However, while the applications of classical statistics in real estate research date back to the early 1970s, spatial statistics were an

[2] Anselin (1988) outlines five characteristics of spatial econometrics proposed by Paelink and Klaaseen (1979, 5-9) as follows:

- The role of spatial interdependence in spatial models
- The asymmetry in spatial relations
- The importance of explanatory factors located in other spaces
- Differentiation between ex post and ex ante interaction
- Explicit modelling of space

[3] Spatial heterogeneity refers to the variation in the relationship under study across space (Patton and McErlean, 2003) or the systematic variation in the behaviour of a given process across space (Can, 1990). It usually leads to heteroscedastic error terms, thus violating the assumption of homoscedasticity in the classical regression model (Can, 1990).

addition to the statistics literature only ten years later (Cressie, 1989). More importantly, it was only in the late 1990s that the use of spatial statistics started to gain the attention of many researchers[4].

However, very few studies come from the UK. Similarly, the importance of spatial dependency on the efficiency and consistency of hedonic model estimates has only very recently started to receive some attention (Kim *et al.,* 2003). Cressie (1989) believes that spatial prediction is just as important as temporal prediction. However, Anselin and Bera (1998) state that generally, econometric theory and practice have been dominated by a focus on the time dimension. They criticise that *in stark contrast to the voluminous literature on serial dependence over time, there is scant attention paid to its counterpart in cross sectional data, spatial autocorrelation* (Anselin and Bera, 1998, 237). In the UK, the consideration of spatial dependence in the housing market studies is not obvious. Day (2003) considers spatial autocorrelation in his study but provides no spatial hedonic model for the entire market of GCC[5].

On one hand, examining spatial dependency as a hedonic problem could portray it as a methodological disadvantage. On the other hand, it can give information on spatial pattern structure and process (Overmars *et al*., 2003) when explicitly specified in a spatial model. Spatial models are generally specified as *linear regression models with spatial interdependence taking the form of a linear additive relationship of observations on neighbours* (Wilhelmsson, 2002a, 95). This is based on the *first law of geography* (Tobler, 1970), which states that *everything is related to everything else, but closer things more so*.

Therefore, data that are close together are usually more correlated than data that are far apart (Cressie, 1989). Based on this, Anselin and Bera (1998, 240) suggest that spatial dependence is a rule rather than an exception. Supporting this, Bowen *et al.*

[4] For example, Pace *et al.* (forthcoming), Wilhelmsson (2004), Tu *et al.* (2004), Dawkins (2004), Day (2003), Cano-Quervos *et al.* (2003), Brasington (2002), Besner (2002), Bowen *et al.* (2002), Deddis (2002), Tse (2002), Wilhelmsson (2002a), Paez *et al.* (2001), Quercia *et al.* (2000), Gillen *et al.* (2001), Pearson (2001), Carter and Haloupek (2000), Deddis *et al.* (2000), Figueroa (1999), Dubin *et al.* (1999), Dubin (1998, 1992, 1988), Can and Megbolugbe (1997), Can (1992, 1990), Wiltshaw (1996), Olmo (1995), Pace *et al.* (1998a, 1998b), Pace and Gilley (1997), Pace and Barry (1997), Pace (1997), and Basu and Thibodeau (1998).

[5] Day (2003) uses General Method of Moment (GMM), which Bell and Bockstael (2000) contend to be less effective than the Maximum Likelihood approach adopted in this study.

(2001) stress that spatial diagnostics need to be included as part of the test model-fitting procedure for hedonic house price applications. Anselin[6] (1998) contends that, despite widespread recognition by both theorists and practitioners of the complex roles of location and spatial interaction and the resulting geographically segmented nature of real estate markets, an explicit spatial treatment of these markets in empirical research is still in its infancy. Bowen *et al*. (2001, 467) note that *many applications of hedonic housing price models have not included recent advances in spatial analysis that control for spatial dependence and heterogeneity*. This provides an opportunity for real estate research.

Realising the lack of evidence of simultaneous consideration of spatial elements in hedonic price modelling, particularly in the UK, this paper focuses on the simultaneous consideration of detailed accessibility measures and spatial autocorrelation in a case study of Glasgow, Scotland. The next section of this paper describes the study area and the hedonic data involved. This is followed by the results of hedonic modelling and discussion. The final section concludes the paper by highlighting the importance of individual accessibility measures and the benefits of applying spatial statistics in hedonic price modelling.

2 THE STUDY AREA AND THE DATA

Glasgow was chosen as the main study area for its sufficient size for a meaningful housing market study, complex accessibility conditions, availability of previous studies based on the same area, which can serve as a guideline, and availability of data by the time the research was scheduled to commence the empirical investigation. The selection of the study area boundaries of Glasgow City Council (GCC) has considered the theoretical and practical aspects[7]. The theoretical aspects include three criteria of prominent quantitative research namely reliability, replicability and validity (Bryman, 2001), as well as the housing market economics and evidence from real estate literature. The GCC area has a wide range of housing, is a socially heterogeneous city and has been the area that researchers concentrate on

[6] According to Anselin (1998), early efforts to implement spatial regression models in urban and real estate analysis include Griffith (1981), and Anselin and Can (1986) which focused on urban density functions as well as Dubin (1988; 1992) and Can (1990; 1992) in the context of hedonic models for house prices. He follows on to state that these studies were characterised by the use of fairly small datasets (in contrast to more "mainstream" microeconomic cross-sectional analyses) and a focus on methodological issues.

[7] The practical aspects include computing issues, location of information and familiarity with the study area.

in previous studies of Glasgow housing markets. These support that GCC is a valid and appropriate area for a housing market study that focuses on the issue of neglected spatial elements in hedonic modelling.

Based on the GCC area, four main groups of **data were gathered** for this study. These are house prices, structural characteristics, neighbourhood characteristics and accessibility measures[8]. Most of the data used in this study have been obtained from government agencies. The literature suggests that data sources reflect data quality. Thus, this study has used data of UK government quality. Although the analysis also involved the 1991 census based data, which are relatively outdated, this is not thought to give an adverse effect on the whole findings because no drastic change has been reported about the population of Glasgow City Council as per comparison between the 1991 and 2001 censuses.

The following stage of **data preparation** verified, cleaned, and converted the data as necessary into the formats suitable for further analysis[9]. This stage has made ready all the hedonic variables among which are several newly GIS constructed spatial variables. Most importantly, the prepared data include the detailed accessibility measures and the spatial weight matrix needed for spatial hedonic modelling. Having the empirical data gathered and prepared, the final hedonic datasets contain 2,715 sale prices as the dependent variable and 61 independent variables. The details of the data and their sources are as in Appendix 1.

Descriptive statistics show that structurally, the dataset is dominated by flats (75%) followed by atttached (22%) and detached (3%) properties. Thus, there is a possibility for flats to influence the hedonic models. The dependent variable is normally distributed when log of selling prices are used. The independent variables also have a reasonable variability in values based on their standard deviation (Description of 61 variables are as in Appendix 2. Simple descriptive statistics of the variables are as in Appendix 3).

[8] This study considers zonal and individual accessibility measures. Data for the former were obtained from David Simmonds Consultancy (DSC) with consent from The Scottish Executive. Data for the latter were constructed using GIS.

[9] Since further analysis were carried out in SPSS 11.5.1, ArcView 3.2 and Matlab 6.5.1 the relevant data were to be in *.sav, shapefiles* and *.mat* formats respectively. The application of GIS and spatial statistics software is summarised in Appendix 5.