Dundee Discussion Papers in Economics 253

Bhattacharjee, Arnab; Anselmo de Castro, Eduardo ; Marques, Joao Lourenco

Dundee Discussion Papers in Economics

Spatial Interactions in Hedonic Pricing Models: The Urban Housing Market of Aveiro, Portugal

Arnab Bhattacharjee, Eduardo Anselmo de Castro & João Lourenço Marques

Department of
Economic Studies,
University of
Dundee,
Dundee.
DD1 4HN

Working Paper
No. 253
June 2011
ISSN:1473-236X

Spatial Interactions in Hedonic Pricing Models: The Urban Housing Market of Aveiro, Portugal [#]

Arnab Bhattacharjee

Economic Studies

University of Dundee, UK

(email: a.z.bhattacharjee@dundee.ac.uk)

Eduardo Anselmo de Castro

and

João Lourenço Marques

Department of Social, Juridical and Political Sciences

University of Aveiro, Portugal

(emails: ecastro@ua.pt and jjmarques@ua.pt)

Abstract

Spatial heterogeneity, spatial dependence and spatial scale constitute key features of spatial analysis of housing markets. However, the common practice of modelling spatial dependence as being generated by spatial interactions through a known spatial weights matrix is often not satisfactory. While existing estimators of spatial weights matrices are based on repeat sales or panel data, this paper takes this approach to a cross-section setting. Specifically, based on an *a priori* definition of housing submarkets and the assumption of a multifactor model, we develop maximum likelihood methodology to estimate hedonic models that facilitate understanding of both spatial heterogeneity and spatial interactions. The methodology, based on statistical orthogonal factor analysis, is applied to the urban housing market of Aveiro, Portugal at two different spatial scales.

Keywords: Spatial econometrics; Spatial heterogeneity; Spatial dependence; Spatial scale; Hedonic pricing; Statistical factor analysis; Spatial weights matrix.

JEL Classification: C21; C13; C14; R12; R31.

[#] The authors are grateful to Glen Bramley, Bernie Fingleton, George Galster, Geoffrey Hewings, Lung-fei Lee, Duncan Maclennan, Steve Malpezzi, Jesús Mur and Gwilym Pryce, and to participants at the 9th Spatial Econometrics and Statistics Workshop (Orléans, 2010) for many helpful comments and suggestions. The usual disclaimer applies.

> *“Euclidean space is defined by its "isotopy" (or homogeneity), a property which guarantees its social and political utility. The reduction to this homogenous Euclidean space, first of nature's space, then of all social space, has conferred a redoubtable power upon it. All the more so since that initial reduction leads easily to another – namely, the reduction of three-dimensional realities to two dimensions (for example, a "plan," a blank sheet of paper, something drawn on paper, a map, or any kind of graphic representation or projection).”* (Lefebvre, 1974 [1991], p.285)

1. INTRODUCTION

This paper examines the role of space in housing markets within the context of a spatial hedonic pricing model applied to the city of Aveiro, Portugal. The above application is based on a methodological contribution, namely maximum likelihood inference for an unknown spatial weights matrix in a pure cross-section setting, for a given *a priori* characterisation of housing submarkets. Specifically, the paper modifies the methodology developed in Bhattacharjee and Jensen-Butler (2004) for estimation of a symmetric spatial weights matrix in a panel data setting, to the current context of cross-section data, admitting a factor structure and structural spatial dependence arising from a spatial error model. The estimated spatial weights matrix is not constrained by restrictive assumptions relating to drivers of spatial diffusion, and offer unique opportunities to understand the nature of interactions in urban space.

Our framework and applications pay special attention to three related but distinct aspects of spatial analyses relating to housing markets – spatial heterogeneity, spatial dependence and spatial scale. The former two spatial effects have been extensively discussed in the spatial econometrics literature; see, for example, Anselin (1988a,b, 1999, 2002). To quote: "*Spatial dependence may be caused by different kinds of spatial spill-over effects, while heteroskedasticity could easily result from the heterogeneity inherent in the delineation of spatial units and from contextual variation over space.*" (Anselin, 1988b:1). Spatial scale is not so much an econometric, but an important empirical issue, and has been discussed widely in the housing economics literature, for example in Malpezzi (2003). Whether an urban scale is the most suitable, or whether the appropriate scale for analysis should be peri-urban (including an urban centre, adjoining suburbs and the countryside), regional or national, depends on both the spatial phenomenon under analysis and the specific spatial context.

In our framework, analysis proceeds as follows. First, a suitable spatial scale is fixed. Next, at the above chosen scale, we begin by segmenting the housing market into submarkets, based on a combination of several criteria: administrative boundaries, hedonic substitutability and socio-cultural segmentation.[1]

Given the above segmentation into submarkets, spatial dependence relates to inferences on spatial weights representing spillovers across different submarkets, and those between houses within the same submarket. For the former, the methodology in Bhattacharjee and Jensen-Butler (2004) can be adapted, provided we can estimate the cross-submarket spatial autocovariance matrix of the errors; for this purpose, we use a methodology based on the cross-section factor model. For the second, we develop maximum likelihood methodology, where we assume for simplicity that the within submarket spatial weights are the same across all submarkets.

Following the spatial econometrics literature (see, for example, Anselin, 1988b, 1999), spatial heterogeneity is used to inform spatially varying coefficients, spatial structural change and heteroscedasticity. In our framework, this is achieved by allowing for heterogeneity across submarkets in intercepts and slopes of the factor-based hedonic housing price model, as well as the error variance.[2]

As mentioned above, the proposed methodology is based on statistical factor analysis on housing and location characteristics. Applied to the housing market of Aveiro segmented into submarkets at two different spatial scales, the method provides a description of urban spatial structure based on spatial heterogeneity, spatial interactions and spatial scale. The resulting spatial model is useful for understanding relative importance of various elements – housing characteristics and access to central and local amenities, as well as interactions within and between housing submarkets – and provides useful inferences on residential location, urban planning and policy. Substantial gains are also obtained with regard to house price prediction.

[1] There is considerable debate in the literature as to which of these alternatives constitute an appropriate criterion, and even whether submarkets are truly spatial entities; see Rothenberg et al. (1991). Here, we abstract from these issues somewhat and assume that our submarkets, at the given spatial scale, have a spatial context which we examine in terms of spatial heterogeneity and spatial dependence.

[2] Inference on cross-submarket heteroscedasticity is a by-product of our methodology. However, we do not focus on this issue in the paper.

In several ways, the proposed framework offers improved understanding of space in housing markets. First, we represent spatial heterogeneity through a factor-based hedonic pricing model estimated at the submarket level. In the process, we follow the literature, beginning with Archer and Williamson (1973) and Davies (1974), on the use of statistical factor analysis to aggregate hedonic characteristics into interpretable behavioural categories. These orthogonal factors are then used as explanatory variables in a hedonic pricing regression model in log-linear form. The model allows for spatial heterogeneity in the form of different preferences for housing and access characteristics in different submarkets and submarket specific fixed effects. Substantial spatial heterogeneity is observed, providing useful interpretation for urban spatial structure.

Second, following the emerging econometric literature on unknown spatial weight matrices (Bhattacharjee and Jensen-Butler, 2005; Bhattacharjee and Holly, 2009, 2011), we develop methodology to estimate spatial interactions within and across housing submarkets under the structural assumption of symmetry. As Anselin and Lozano-Gracia (2008) and Anselin et al. (2010) demonstrate, it is important in hedonic models to allow for interactions modelled by a spatial econometric model. At the same time, there is explicit acknowledgement in the literature of considerable uncertainty regarding real drivers of spatial diffusion, and that assuming spatial weights based on geographic distances or contiguity is far too simplistic (Giacomini and Granger, 2004; Bhattacharjee and Jensen-Butler, 2005). Our work extends the literature on estimation of the spatial weights matrix to a pure cross-section setting under the structural assumption of symmetry; for a critical review of some recent methods in the panel data setting, see Bhattacharjee and Holly (2011).

Third, in conducting analyses and obtaining inferences of heterogeneity and interactions at two different spatial scales, our work emphasizes the importance of spatial scale. In our view, studying spatial structure of housing markets from all the above dimensions is important not only for understanding residential location but also for urban planning and housing policy; see Maclennan (2010) and Bhattacharjee et al. (2010) for further discussion.

Finally, our conceptual framework, combining unknown spatial interactions with unrestricted spatial heterogeneity, has important connections to philosophical views in geography and urban studies. Spatial heterogeneity relates to the idea of distinctive identity of spatial units or "unique spaces" (Hartshorne, 1939) and emphasizes neighbourhood effects. Further, our representation of spatial dependence in terms of an unknown spatial weights matrix is closely wed to the idea of abstract and endogenously produced space in Lefebvre (1974 [1991]). At the same time, our spatial weights matrices acknowledge the notion of distance decay inherent in logical positivism (Schaefer, 1953), while abstracting from the Cartesian characterisation of space emphasized in exogenous and given spatial weights matrices based on distance or contiguity.

The rest of the paper is organised as follows. In section 2, we discuss our methodology and place it within the context of the emerging literature on inferences relating to unknown spatial interactions. Following this, in section 3, we apply our methods to a spatial hedonic price model for the city of Aveiro, using a small but very rich dataset on hedonic characteristics of properties sold through the leading real estate agency in 2007. Next, we extend the analysis to the peri-urban spatial scale including neighbouring parishes with potentially substantial spillovers with Aveiro, based on data for all properties put on the market between 2000 to 2010 (section 4). With both datasets, we find important evidences on spatial heterogeneity and dependence that inform understanding of the urban spatial structure of housing markets in Aveiro, as well as the importance of spatial scale. Section 5 concludes.

2. METHODOLOGY

Our proposed methodology combines spatial hedonic analysis based on orthogonal factors with a method for inferences on unknown spatial weights matrix under the structural constraint of symmetric spatial weights. Below, we discuss our methodology in further detail.

2.1. Hedonic pricing model

Building on the early work of Lancaster (1966) and Rosen (1974), hedonic pricing models continue to be actively used in housing studies. In particular, valuation of access to central and local services and other housing attributes, and construction of price indices based on single sales data, have been addressed through hedonic specifications; see Maclennan (1977) for a classic and critical discussion, and Chattopadhya (1999), Malpezzi (2003) and Palmquist (2005) for recent reviews.

In hedonic pricing models, dwelling unit values (or proxies such as prices or rents) are regressed on a bundle of characteristics of the unit that determine the value:

$$P = f(S, N, L, C, T),$$

where P denotes the value of the house (price, or price per unit area), and S, N, L, C and T denote respectively, structural characteristics of the dwelling (living space, type of construction, tenure, etc.); neighbourhood characteristics (and local amenities); location within the market (or access to employment/ business centre); other characteristics (access to utilities and public services, such as clean water supply, electricity, central heating, etc.); and the time (date, month) when value is observed.

Estimating the hedonic price function using a collection of observed housing values and dwelling unit characteristics yields a set of implicit prices for housing characteristics that are essentially willingness-to-pay estimates. This allows analysis of various upgrading scenarios, targeted to specific subgroups, defined either by socio-economic characteristics or by location. Thus, the model facilitates understanding of residential location, and therefore urban structure, and provides valuable input towards urban planning and housing policy.

Theory provides no guidance for the functional form appropriate for hedonic regression. However, a nonlinear hedonic function is useful for recovering the underlying structural demand curve from estimates of the hedonic relationship (the reduced form). This and several other important advantages motivated Follain and Malpezzi (1980) to recommend the semi-log form; for further discussion, see Follain and Malpezzi (1980) and Malpezzi (2003).

In this paper, we estimate the hedonic model with a small modification to the semi-log form, where logarithm of price per square meter of living space is regressed on logarithm of house area, conditioning on several other hedonic housing characteristics.

2.2. Factor structure and multivariate dimension reduction[3]

There are literally hundreds of potential housing characteristics that could be included on the right hand side of a hedonic regression model. Unfortunately, coefficient estimates are not

[3] We are grateful to an anonymous referee for valuable suggestions that helped us understand better the role of the factor model in our setting.

robust to the omitted variables problem (Butler, 1982; Ozanne and Malpezzi, 1985). However, the same correlation between omitted and included variables that biases individual coefficient estimates often aids better prediction from a "sparse" model (Malpezzi, 2003).

This feature of the hedonic pricing model enhances the possibility of exploiting the factor structure to obtain parsimonious estimates and improved predictions. Several studies, beginning with Kain and Quigley (1970) and Archer and Wilkinson (1973) have taken this approach, and Davies (1974) combined factor analysis with the regression approach. In a critical review, Maclennan (1977) suggests that the researcher needs to make sure that the extracted factors do not reflect solely statistical properties but behavioural collections of housing characteristics.

In this paper, we employ a hybrid approach, combining factor analysis with regression. Initially, we identify a small collection of leading factors from a large number of potential hedonic characteristics, using factor analysis with varimax orthogonal rotations. For both the datasets used in this paper, these (orthogonal) factors are clearly identified with interpretable collections of housing characteristics, such as structural dimensions, access to utilities, centrality and access to local services. At the second stage, we predict factor values for all properties, including those for which some hedonic characteristics are missing, and use these predicted factors to estimate the hedonic regression model in semi-log form.

In this way, we address the Davies (1974) and Maclennan (1977) critiques. Further, the assumed factor model is absolutely crucial for our proposed inference methods for the spatial weights matrix; we will discuss this issue in section 2.4. In addition, the factor based hedonic housing price model offers three potential advantages.

First, in building a hedonic model based on a small number of factors, rather than a large collection of housing characteristics, we build a parsimonious model with more precise estimation that offers better interpretation of the regression coefficients (implicit prices). More importantly, in including all potential economic factors affecting prices, the factor based regression model is less susceptible to the omitted variables problem. Second, hedonic regression based on factors allows a unique opportunity to reduce missing value problems, where factors can be predicted (imputed) using the information available on only a selection of included characteristics, under the assumption that the missing data are allocated randomly across the properties conditional on the values of observed features. This leads to considerably larger sample sizes for estimation of the hedonic model, with benefits of improved precision in the estimates. Third, the approach based on orthogonal factors is not subject to multicollinearity, and could therefore contribute to higher efficiency, which in turn can lead to better prediction of housing prices; see Malpezzi (2003) for further discussion. Besides, the orthogonality of factors is also important for estimating the spatial weights, which we discuss later in section 2.4.

Finally, and most importantly, the factor based approach is absolutely crucial for our methodology for estimating spatial weights. Thus, it is very useful for studying spatial dependence in urban housing markets driven by an unknown spatial weights matrix.

2.3. Spatial issues in hedonic pricing estimates

The recent literature has discussed the potential bias and loss of efficiency that can result when spatial effects are ignored in the estimation of hedonic models; see, for example, Pace and LeSage (2004), Anselin and Lozano-Gracia (2008) and Anselin et al. (2010). Spatial patterns in the housing markets arise from a combination of spatial heterogeneity and spatial dependence (Anselin, 1988a). Additionally, as discussed before, choice of an appropriate spatial scale is important (Malpezzi, 2003). We now turn to a discussion of spatial issues in the construction of our hedonic pricing models, including all of the three above aspects of space.

2.3.1. Spatial scale and housing submarkets

Definition of submarkets is important at both conceptual and empirical levels. Housing markets are local and diverse, and hedonic price estimation requires careful delineation of these markets (Malpezzi, 2003). The definition of submarkets in practice ranges from the national or regional scale (Linneman, 1981; Mills and Simenauer, 1996), through metropolitan areas (Follain and Malpezzi, 1980), to below the metropolitan level (Straszheim, 1975; Gabriel, 1984; Grigsby et al., 1985; Rothenberg et al., 1991; Maclennan and Tu, 1996; and Bourassa et al., 1999).

Malpezzi (2003) argues that one reason why the metropolitan area is appealing as the unit of analysis is that these areas are usually thought of as labour markets, which may therefore be approximately coincident with housing markets. On the other hand, submarkets below the metropolitan level can be segmented by location (central city/suburb), or by housing quality, or even by race or income levels. Such segmentation facilitates both understanding of residential neighbourhood choice and devising appropriate urban housing policy. However, the empirical literature does not suggest an unambiguous definition of a unique spatial scale.

In this paper, we conduct our analysis at two different spatial scales, both disaggregated to a relatively fine spatial level. In the first, we consider administrative regions (parishes) within the city of Aveiro as submarkets, and pool the suburban area together into a single submarket. This definition aids understanding of spatial heterogeneity and interactions within the urban area, but does not provide satisfactory analysis in terms of spillovers between the city and the suburban area. Second, we extend our analysis to a finer spatial scale within the suburban area, constructing submarkets with careful consideration to the principles of segmentation discussed above. Our analysis reflects some advantages of using a flexible spatial scale, since processes of agglomeration and dispersion operate differently at different scales (Arbia et al., 2009, 2010).

2.3.2. Spatial heterogeneity and neighbourhood effects

The conceptual notion behind spatial submarkets discussed above implies that the price determining (hedonic) mechanism can be heterogeneous over space. This spatial heterogeneity can originate from demand and supply factors, institutional barriers or discrimination, each of which can cause differentials across neighbourhoods in the way housing attributes are valued by consumers and house prices determined (Anselin et al., 2010). However, if spatial heterogeneity is present and ignored, an average price across all submarkets is estimated that ignores submarket heterogeneity. Worse still, the error term of the regression can then be correlated with the included regressors and ordinary least squares (OLS) will produce biased estimates.

The standard urban model in the Alonso-Muth-Mills tradition predicts a generally declining pattern of prices with distance from the centre of the city, though there may be spatial variation in relative preference for centrality. Other models based on localised amenities or multiple centres imply a stronger impact of access to local amenities. Like distances, the implicit prices for dwelling characteristics and size may also vary spatially, reflecting either supply constraints or residential sorting. Follain and Malpezzi (1980), Mozolin (1994), Adair et al. (2000) and Soderberg and Janssen (2001), among others, have examined intra-urban variation in the price of housing using hedonic models. We follow the above line of literature in allowing coefficients in our hedonic pricing model to vary across submarkets, and use the estimated variation to infer on residential neighbourhood choice and urban spatial structure.

2.3.3. Spatial dependence and spatial weights matrix

In contrast to spatial heterogeneity, spatial dependence leads to spatial autocorrelation, implying that prices of nearby houses tend to be more similar than those of houses that are farther apart. Likewise, average price of houses in nearby or related submarkets may be correlated more strongly. A common explanation for spatial autocorrelation is spatial spillovers or other forms of contagion effects. However, incorrectly modelled spatial heterogeneity, measurement problems in explanatory variables, omitted variables, and unmodelled features having a spatial pattern can also lead to spatial autocorrelation (Anselin and Griffith, 1988). Recent empirical literature has addressed issues of bias and loss of efficiency that can result when spatial effects are ignored in the estimation of hedonic models,[4] and the use of spatial econometric models to address spatial autocorrelation is becoming increasingly standard.[5]

The usual approach to the representation of spatial interactions is to define a spatial weights matrix, denoted *W*, which typically represents a theoretical and *a priori* characterisation of the nature and strength of spatial interactions between different submarkets or dwellings.[6] These spatial weights represent patterns of diffusion of prices and unobservables over space, and thereby provide a meaningful and easily interpretable representation of spatial interaction (spatial autocorrelation). The spatial weights are typically modelled as functions of geographic or economic distance. The distance between two spatial units reflects their proximity with respect to prices or unobservables, so that the spatial interaction between a set of units (dwellings) can be represented as a function of the economic distances between them.

[4] See, for example, Pace and LeSage (2004) and Anselin and Lozano-Gracia (2009).
[5] For representative applications using hedonic models in a spatial econometric setting, see Can (1992), Pace and Gilley (1997), Basu and Thibodeau (1998) and Anselin et al. (2010).
[6] For a setting with n spatial units under study, W is an $n \times n$ matrix with zero diagonal elements. The off-diagonal elements are typically either dummy variables for contiguity or inversely proportional to the distance between a pair of units, so that spillovers between a pair of units that are farther apart is lower.

Given a particular choice of the spatial weights matrix, there are two important and distinct ways in which spatial dependence is modelled in spatial regression analysis – the spatial lag model and the spatial error model. In the former, the hedonic regression includes as an additional regressor the spatial lag of the dependent variable y (which in our case is price), represented by Wy, and the regression errors (ε) are completely idiosyncratic. By contrast, in the spatial error model, the regression errors are spatially dependent on their spatial lag, $W\varepsilon$.

The implications of spatial interaction on estimation of these two models are different. In the spatial lag model, the endogenous spatial lag implies that OLS estimates not accounting for spatial interaction would be biased, while in the spatial error model, they will be unbiased but inefficient. However, though different in interpretation, the above two models are very difficult to distinguish empirically (Anselin, 1999, 2002). In line with current practise in the area of spatial econometrics, we first estimate the hedonic pricing model under the spatial error assumption. Next, to judge whether endogenous spatial lags are relevant, we perform a test for spatial lag dependence by nesting the spatial error model within a hybrid model incorporating both spatial lag and spatial error dependence; for more discussion on sequential model selection in the spatial context, see Born and Breitung (2009).

The choice of appropriate spatial weights is a central component of spatial models as it imposes *a priori* a structure of spatial dependence, which may or may not closely correspond to reality. Further, the accuracy of these measures affects severely the estimation of spatial dependence models (Anselin, 2002; Fingleton, 2003). Spatial contiguity or suitable functions of geographic distances are frequent choices. However, spatial data may be anisotropic, where spatial autocorrelation is a function of both distance and the direction separating points in space (Simon, 1997; Gillen et al., 2001). Further, spatial interactions may be driven by other factors, such as trade weights, transport cost, travel time, and socio-cultural distances. The choice typically differs widely across applications, depending not only on the specific economic context but also on availability of data. The problem of choosing spatial weights is a key issue in many applications.

Given the above ambiguities regarding measurement of spatial weights, and in line with the notion that factors different from geographic distance or contiguity may potentially drive spatial interactions, we consider the spatial weights matrix (W) as an unknown symmetric matrix with zero diagonal elements. We allow spatial interactions to be potentially negative, often implying segmented housing markets or asynchronous housing cycles.[7] Based on a given definition of urban submarkets (or a fixed set of spatial locations) and panel data on these spatial units, Bhattacharjee and Jensen-Butler (2005) and Bhattacharjee and Holly (2009, 2011) developed several methods to estimate the spatial weights matrix between the submarkets.[8] Here, we extend the panel estimation methodology in Bhattacharjee and Jensen-Butler (2005) under the structural assumption of symmetric spatial weights to a purely cross-section setting.

Specifically, we consider a cross-section Gaussian factor regression model (Liu and Rubin, 1998) where each housing property i $(=1,\ldots,n)$ belongs to a single submarket M_j, where M_j is one of J mutually exclusive and exhaustive submarkets $M_1, M_2, \ldots, M_J$. The price of the dwelling y_i depends on a $(q\times 1)$ vector of unobserved orthonormal Gaussian factors F_i of housing and locational hedonic characteristics (the $(p\times 1)$ vector x_i, $q<p$), where the effects of the factors potentially vary across the submarkets. The corresponding regression error ε_i is uncorrelated with the factors, but may be spatially related to the errors for other houses through an unknown spatial weights matrix, W. In other words, we consider a spatial error model with a cross-section heterogeneous factor structure across the submarkets, where the effects of the factors are potentially different across submarkets and there may be submarket specific fixed effects:

$$\begin{aligned}
&X_i \mid (F_i,\gamma,\Psi) \overset{indep}{\sim} N_p(\alpha+\gamma' F_i,\Psi), \quad i=1,2,\ldots,n, \\
&\Psi = diag\left(\psi_1^2,\ldots,\psi_p^2\right), F_i \overset{iid}{\sim} N_q(0,I), \\
&y_i = \beta_{0j} + \beta_j' F_i + \varepsilon_i, \underline{\varepsilon} = \lambda W \underline{\varepsilon} + \underline{\upsilon}, \quad i\in M_j \\
&\upsilon_i \sim N\left(0,\sigma_j^{\,2}\right) \text{ independent}, Cov(\upsilon_i,F_i)=0.
\end{aligned} \tag{1}$$

---

[7] An analogy with well-researched urban areas in the USA is illuminative. A city like Chicago where house prices have risen in the CBD but grown even faster would imply positive spatial interaction between the centre and the periphery. By contrast, in Detroit, where prices in the centre have declined at the same time as suburban housing prices have continued to rise, suggests negative spatial interaction. We are grateful to Steve Malpezzi for pointing out this connection with urban structure.

[8] See Bhattacharjee and Holly (2011) for a review and discussion, as well as an application to network interactions within a committee.

Here $\underline{\varepsilon} = \left(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\right)'$, the vector of the random errors, has a spatial error structure with an unknown symmetric spatial weights matrix $W_{(n\times n)}$ having zero diagonal elements, and the zero mean Gaussian idiosyncratic errors (υ_i's) are potentially heteroscedastic across submarkets but independent over the cross-section and uncorrelated with the random factors F_i. Equation (1) describes a simplified version of the cross-section factor model with heterogeneous group effects discussed in Andrews (2005), with additional Gaussian assumptions. These distributional assumptions are useful in our case for drawing inferences by maximum likelihood on the intra-submarket spatial weights.

The spatial weights matrix *W* is the row-standardised version of $W^0{}_{(n\times n)}$, which is assumed to be symmetric and have a block-structure as follows:

$$W_{ij}^0 = \begin{cases} 0 & \text{if} \quad i = j \\ \omega_0 & \text{if} \quad i,j \in M_k, k = 1,\ldots,J \\ \omega_{kl} & \text{if} \quad i \in M_k, j \in M_l, M_k \bigcap M_l = \phi. \end{cases} \qquad (2)$$

The above weights matrix is unknown and quite general, allowing for unknown but fixed spatial weights between properties in the same submarket, and similarly unknown spatial weights between properties in any pair of submarkets.[9] For identification in the reduced form, it is required that *(I–λW)* is nonsingular. Further structural assumptions are required for identification. Following Bhattacharjee and Jensen-Butler (2005), we assume symmetric spatial weights within and between submarkets – a standard assumption in the spatial econometrics literature. However, spatial weights are allowed to be negative.

In this paper, we assume a spatial error model. Because of endogeneity, estimating the spatial weights matrix under the spatial lag model is a very difficult problem. However, we can perform specification tests against the spatial lag model under the assumption that the same spatial weights matrix *W* describes both spatial lag dependence and error dependence. For this

[9] Note that, since the spatial weights matrix is unknown in our setting, it is necessary to row-standardize *W* to enable identification of both *W* and the autoregressive parameter (λ) in Equation (1). The assumption that the intra-submarket spatial weight is the same across all submarkets is not necessary, but retained here for computational simplicity.

purpose, we nest the above spatial error model within the following model that includes both spatial lag and spatial error, with different autoregressive coefficients:

$$\begin{aligned} &\underline{y} = \rho W \underline{y} + \beta_j' F_i + \varepsilon_i, \qquad i = 1,2,\ldots,n, \quad i \in M_j, \\ &\underline{\varepsilon} = \lambda W \underline{\varepsilon} + \underline{\upsilon}, \qquad \upsilon_i \sim N\left(0, \sigma_j^{\,2}\right) \text{ independent}, Cov(\upsilon_i, F_i) = 0. \end{aligned} \tag{3}$$

Borg and Breitung (2009) propose a regression based test, where at the first stage, the spatial error model (1) is estimated. At the second stage, the test evaluates whether there is any residual spatial dependence that can be explained by spatial lag effect. We use this test to verify whether the spatial error model is adequate for our empirical applications. The test is simple to apply and has several advantages over standard LM tests; see Born and Breitung (2009) for further details.

2.4. Estimation of symmetric spatial weights matrix

As discussed before, our main methodological contribution is to estimate unknown spatial weights within a factor-based cross-section spatial error model. Next, we describe our estimation methodology in three steps: first, the cross-market spatial interaction matrix W^* (defined in (4) below); second, the cross-submarket spatial autocovariance matrix Γ; and finally, the within submarket spatial weight ω_0.

2.4.1. Estimation of cross-market spatial interaction matrix

In the panel data setting, the methodology in Bhattacharjee and Jensen-Butler (2005) is based on a given consistent estimator for the underlying hedonic regression model with spatial errors (1). Based on residuals from the above estimation, a consistent estimator $\hat{\Gamma}$ is first obtained for the $J \times J$ cross-submarket spatial autocovariance matrix

$$\begin{aligned} &\Gamma = \left(I - W^*\right)^{-1} \Sigma \left(I - W^*\right), \\ &W^* = \lambda \begin{bmatrix} 0 & \omega_{12} & \ldots & \omega_{1J} \\ \omega_{12} & 0 & \ldots & \omega_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{1J} & \omega_{2J} & \ldots & 0 \end{bmatrix}, \\ &\text{and } \Sigma = diag\left[\sigma_1^{\,2}, \quad \sigma_2^{\,2}, \quad \ldots \quad , \sigma_J^{\,2}\right]. \end{aligned} \tag{4}$$

Let us assume that such a consistent estimator $\hat{\Gamma}$ has been obtained. Bhattacharjee and Jensen-Butler (2005) show how this estimator $\hat{\Gamma}$ can then be used to estimate the unknown cross-submarket spatial weights matrix W^*. Without any structural constraints on the weights matrix, the estimation problem is only partially identified, up to an orthogonal transformation of interactions. Specifically, they show that the matrix

$$V = \left(I - W^*\right) diag\left[\frac{1}{\sigma_1}, \quad \frac{1}{\sigma_2}, \quad \cdots \quad , \frac{1}{\sigma_J}\right]$$

is consistently estimated, up to an arbitrary orthogonal transformation, by

$$\hat{\Gamma}^{-1/2} = \hat{E}.\hat{\Lambda}^{-1/2}.\hat{E}',$$

where $\hat{E}$ and $\hat{\Lambda}$ contain the eigenvectors and eigenvalues respectively of the estimated spatial autocovariance matrix $\hat{\Gamma}$.[10] In other words, $\hat{\Gamma}^{-1/2}$ is a consistent estimator of VT for some unknown square orthogonal matrix T. Since T is an arbitrary orthogonal matrix, it has precisely $J(J-1)/2$ free elements. Hence, the spatial weights matrix W^* can be precisely estimated only under additional structural constraints. Symmetry of the spatial weights matrix constitutes one set of valid identifying restrictions,[11] which is the structural assumption we make here.

Under the symmetry assumption, Bhattacharjee and Jensen-Butler (2005) describe inference methods and an algorithm for estimating the unknown spatial weights matrix. Estimation requires application of the "gradient projection" algorithm (Jennrich, 2001) which optimises an objective function over the group of orthogonal transformations of a given matrix; standard errors are obtained using the bootstrap.

This method can be readily applied to our spatial hedonic pricing model provided an initial consistent estimator can be found for the cross-submarket spatial autocovariance matrix Γ. In this paper, we propose a maximum likelihood method to estimate this autocovariance matrix.

---

[10] Here, $A^{1/2}$ denotes the symmetric square root of a positive definite matrix A, and $A^{-1/2}$ denotes its inverse. In other words, $A^{-1/2}$ has the same eigenvectors as A, but with the eigenvalues replaced by the reciprocal of the square root of the corresponding eigenvalues of A.

[11] See Bhattacharjee and Holly (2011) for further discussion on partial identification and structural constraints in this context.

2.4.2. Estimation of cross-submarket spatial autocovariance matrix

In the panel data setting, Bhattacharjee and Jensen-Butler (2005) estimated the underlying regression model and obtained residuals, and then estimated $\hat{\Gamma}$ as the simple sample covariance matrix of the cross-market residuals. This step was relatively simple because for each time period, there was a residual uniquely identified with each submarket.

In the current pure cross-section setting, the situation is more complex because *a priori* there is no natural way to associate a house in any one submarket with a corresponding house in any other submarket. For this matching problem, we use an analogy of the current cross-section factor model (Andrews, 2005) with the multifactor error structure of cross-sectionally dependent panel data inherent in the common correlated effects methodology of Pesaran (2006).

In the common correlated effects approach (Pesaran 2006), linear combinations of unobserved common factors are approximated by cross-section averages of the dependent and explanatory variables, which are then included in the panel regression model in addition to the other regressors. The cross-section averages vary over time and not over the cross-section, and represent omitted time-specific common factors. Clearly, an alternative to these cross-section averages would be including a full set of time fixed effects.

Bhattacharjee and Jensen-Butler (2005) use residuals across spatial units for the same time period to estimate the spatial error autocovariance matrix. The multifactor spatial error model provides a clear justification for this approach. Residuals for the same period are matched because the corresponding observations on different spatial units align perfectly along the dimension of the unobserved latent factors – in the panel data setting, the time specific common shocks. Taking this intuition to the pure cross-section setting, it is therefore natural to match housing property i in submarket M_j with the dwelling j in another submarket M_j that bears the closest correspondence in the vector of latent factors; in our case F_i and F_j. Thus, the proposed methodology proceeds as follows.

In the first stage, we estimate a suitable set of orthogonal factors based on hedonic characteristics. Using these estimated factors, we consistently estimate the hedonic regression model in (1) separately for each submarket. This estimation allows for full spatial heterogeneity, but ignores the spatial error structure. Based on these submarket specific regression estimates, we obtain residuals for each property.

We match properties across submarkets in the second stage. Specifically, to the residual for an index dwelling i in submarket M_i, we match the residual for that house j in submarket M_j that has the closest match in the vector of estimated factors; in other words, $j = \arg\min_{j^* \in M_j} \left(F_i - F_{j^*}\right)' \left(F_i - F_{j^*}\right)$.

In the third stage, based on matched residuals across the different submarkets, the cross-submarket spatial autovariance matrix is estimated simply by the sample covariance matrix $\hat{\Gamma}$. Finally, estimation of W^* follows using the Bhattacharjee and Jensen-Butler (2005) methodology outlined in the previous subsection.

The assumption of the multifactor model is crucial for this estimation procedure.[12] First, residuals from an estimated hedonic pricing model would be extremely susceptible to the potential omitted variables problem. In practical terms, it is very difficult to avoid this problem, even if a large number of hedonic characteristics are included in the estimation. By contrast, in estimating the factor model, it is simpler to minimise this problem by including factors corresponding to all notional features that theory and past studies have identified as determinants of house prices. One can then make the reasonable assumption that what remains in the error is uncorrelated with the included factors.

Second, and as discussed above, the factor model is conceptually very closely related to the critical distinction between spatial strong and weak dependence (Pesaran and Tosetti, 2011), and therefore to the common correlated effects approach (Pesaran, 2006). Specifically, in the panel data setting, the theoretical justification for matching residuals corresponding to different spatial units for the same time period is that they match on the strong spatial dependence (or,

---

[12] We thank an anonymous referee whose comments encouraged us to examine the special features of the factor model in the cross section context, thereby improving upon the methodology and its discussion substantially.

unobserved factor) dimension. Matching against estimated factors provides an exact conceptual counterpart to this argument for our cross-section factor model setting.[13]

Third, since our estimated factors are orthogonal by construction, it is straightforward to match two properties in different submarkets by the inner product (sum of squares) of the vector of difference of their corresponding estimated factors.

Finally, under the assumptions of the factor model (1), $\hat{\Gamma}$ estimated as above is the maximum likelihood estimator of the spatial autocovariance matrix (see also Andrews, 2005) under the maintained factor-model and Gaussian error assumption. Since there is a unique relation between Γ and the corresponding W^*, the corresponding cross-submarket symmetric spatial weights matrix is a maximum likelihood estimate as well.

<u>2.4.3. Estimation of within submarket spatial weight</u>

What remains now is to estimate the within submarket spatial weight ω_0. Since dwellings within a submarket are in general located closer to each other than those across different submarkets, it is expected that ω_0 will be large compared to the cross-submarket spatial weights. We propose maximum likelihood to conduct this estimation. Specifically, for any candidate value of ω_0, we construct the corresponding row-standardized spatial weights matrix W, estimate the spatial error model using maximum likelihood, and evaluate the value of the maximised likelihood. In this way, we construct spatial weights matrices using various candidate values for ω_0, estimate the corresponding spatial error models by maximum likelihood using Geoda (Anselin, 2005), and maximise the likelihood over all such candidate values. Standard errors are estimated by numerical approximation to compute the Fisher information at this maximised value for ω_0.[14]

Finally, as discussed earlier, we use the Born and Breitung (2009) regression test to examine the validity of spatial error dependence against a hybrid model including a spatial lag.

[13] See Bhattacharjee and Holly (2011) for further discussion of the conceptual distinction between strong and weak dependence and their link with the spatial weights matrix.
[14] In principle, one can allow the within submarket spatial weights to vary across submarkets. In our empirical exercise, we abstract from this issue for the sake of computational simplicity.

3. EMPIRICAL ANALYSIS AT THE URBAN SPATIAL SCALE

3.1. Data

The proposed methodology is applied to the housing market of Aveiro, a city located in the Centro Region of Portugal. The urban agglomeration of Aveiro includes the municipality with the same name and the neighbouring municipality of Ílhavo and had a population of 114,000 inhabitants in 2006. Our first empirical analysis refers only to the city, which corresponds to 6 of the 14 parishes of the municipality of Aveiro (table 1). The dataset includes 166 properties sold through one of the leading real estate agencies in Aveiro in 2007.[15] The spatial distribution of the properties is presented in figure 1, where each house is indicated by a dot.

This dataset is quite different from the one analysed in the following section. One potential limitation is the reduced number of available observations, covering only the urban and suburban areas of Aveiro municipality. However, this is compensated by availability of data at a finer detail, including specific location of each house (so that exact distances can be computed) and greater detail of hedonic characteristics recorded for each house, but especially the availability of true transaction prices (rather than listing prices as in the second dataset). Most importantly, the small sample size enables us to estimate spatial econometric models with alternate definitions of spatial weights and thereby compare the adequacy of different specifications.

Table 1: Population and density of housing sample

Parishes	Population		Density	Sample Houses	
Aradas	7,628	15%	854	22	13.3%
Esgueira	12,262	24%	691	42	25.3%
Glória	9,917	19%	1,445	27	16.3%
Santa Joana	8,652	17%	225*	11	6.6%
São Bernardo	4,079	8%	1,037	9	5.4%
Vera Cruz	8,652	17%	1,273	55	33.1%
Total	**51,190**	**100%**	**368**	**166**	**100%**

*including the lagoon area

Population (in number of inhabitants), Density (in inhabitants per km^2)

Figure1: Location of housing sample

The data cover single unit housing (12.3 percent) and flats (87.7 percent), both newly built (11.8 percent) and used (88.2 percent), located in different urban and suburban areas. Data on several physical and location attributes of each house were collected, as well as exact location of each

[15] The name of the agency is withheld because of a confidentiality agreement.

house. Location attributes were defined by geographic distances from each property to several amenities and services available within the city, constructed using Geographic Information System (GIS). These represent the desirability of each neighbourhood.

Table 2: Descriptive statistics of variables, urban scale

	Variable	Units of measurement	N	Minimum	Maximum	Mean	Standard deviation
Internal physical characteristics							
d	Type	(Single unit=1, Flat=0)	166	1.00	2.00	1.13	0.34
d	Duplex	(Yes=1; No=0)	162	1.00	2.00	1.20	0.40
d	Balcony	(Yes=1; No=0)	166	0.00	1.00	0.19	0.40
d	Terrace	(Yes=1; No=0)	166	0.00	1.00	0.10	0.30
d	Provision for garage	(Yes=1; No=0)	166	0.00	1.00	0.59	0.49
d	Cable TV	(Yes=1; No=0)	166	0.00	1.00	0.26	0.44
d	Gas (natural)	(Yes=1; No=0)	166	0.00	1.00	0.38	0.49
	Number of bedrooms	(Number)	165	1.00	5.00	2.32	0.84
d	Build and age	(Used=1, New=0)	165	0.00	1.00	0.88	0.32
	Floors	(Number)	166	1.00	12.00	3.46	2.16
ln	Kitchen area	(m^2)	139	1.70	3.21	2.48	0.31
ln	Living room area	(m^2)	147	2.12	3.35	2.53	0.19
ln	Price	(Euros/m^2)	166	5.98	8.01	7.11	0.34
ln	Total area**	(m^2)	166	3.50	5.52	4.67	0.39
Location characteristics							
ln	Central Amenities	(Min. Dist.-meters)	166	4.51	8.58	7.19	0.74
ln	Local Amenities	(Min. Dist.-meters)	166	8.35	9.26	8.72	0.17
ln	CBD Aveiro	(Min. Dist.-meters)	166	5.54	8.63	7.30	0.68
ln	Local Commerce	(Min. Dist.-meters)	166	3.49	7.96	6.14	0.93
ln	Primary Schools	(Min. Dist.-meters)	166	3.16	6.76	5.48	0.69
ln	High Schools	(Min. Dist.-meters)	166	3.14	8.23	6.39	0.95
ln	University	(Min. Dist.-meters)	166	6.06	8.70	7.49	0.58
ln	Hospital	(Min. Dist.-meters)	166	4.96	8.37	7.08	0.62
ln	Health Centres	(Min. Dist.-meters)	166	5.32	8.60	7.31	0.66
ln	Pharmacies	(Min. Dist.-meters)	166	3.39	7.83	5.86	0.88
ln	Parks and Gardens	(Min. Dist.-meters)	166	5.17	8.20	6.81	0.72
ln	Rail Station	(Min. Dist.-meters)	166	4.88	8.21	6.90	0.70
ln	Access Node	(Min. Dist.-meters)	166	5.41	8.31	7.19	0.51
ln	Gas Station	(Min. Dist.-meters)	166	2.08	7.67	6.07	0.95
ln	Police	(Min. Dist.-meters)	166	3.57	8.41	7.11	0.67
p	Administration	(Potential)	166	5.49	9.09	6.89	0.72
p	Cultural centre	(Potential)	166	6.04	8.66	7.19	0.50
p	Specialised Commerce	(Potential)	166	6.56	8.75	7.71	0.43
p	Restaurants	(Potential)	166	7.80	10.15	8.90	0.54
p	Hotels and hostels	(Potential)	166	5.48	8.15	6.72	0.65
p	Monuments	(Potential)	166	7.95	10.90	8.71	0.48
p	Banks, ATMs, Post	(Potential)	166	7.87	10.19	8.85	0.47
p	Sports	(Potential)	166	7.04	8.81	7.88	0.38

d=dummy variable; ln= in logarithms; p=gravitational potential ** : includes 56 imputed missing values

The dependent variable used in hedonic price models is usually the transaction price. We use a more scale neutral normalised measure – logarithm of price per square meter (p/m^2). Housing prices are explained by a wide set of variables (see table 2). Some location attributes are defined as minimum distances to services such as high schools or pharmacies; the others are defined as gravity type measures of potential, generated by distances to services like restaurants, sport centres or public administration offices. For this purpose, we define the potential (P_i) generated by a given set of services of type S ($S_1, S_2, \ldots, S_n$) at a given location (i) by:

$$P_i(S) = \sum_{j=1}^{n} \frac{S_j}{d_{ij}},$$

where S_j denotes the service located at the point j and d_{ij} the distance between locations i and j (see Stewart, 1947).

Descriptive statistics presented in table 2 reflect large variation in the physical and location attributes across the sample as well as missing value problems. Missing values for total area were imputed, under the standard assumption of conditionally missing at random, using the method of conditional mean imputation based on x's and y; see Little (1992) for further discussion.

The dependent variable in our hedonic regression is price per square meters. All other physical and location attributes (including total area) are treated as explanatory variables. Observations with missing values in the other physical characteristics were omitted from the factor analysis.

3.2. Factor Analysis

The explanatory variables were subjected to factor analysis followed by varimax rotations to extract orthogonal factors with maximum explanatory power. The leading factors were identified by maximum likelihood factor analysis on the original housing attributes. The scree plot suggested five leading factors, which were then optimally rotated by a orthogonal varimax procedure. Taken together, the five factors explain 63.4 percent of the variance of all the data. The extracted orthogonal factor loadings are reported in table 3; for visual clarity, we exclude from the table estimated loadings below the standard cut-off of 0.35. Based on these loadings, predicted factor scores were computed for use in subsequent analysis.

Table 3: Factor loadings for urban scale, varimax rotated (absolute value > 3.5)

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Location attributes					
University	.930				
Central Amenities	.921				
CBD Aveiro	.912				
Parks and Gardens	.876				
Local Amenities	.839	-.413			
Local Commerce	.790				
Hospital	.788				
Health Centres	.778				
High Schools	.733				
Pharmacies	.640	.367			
Police	.580	.426			
Gas Station	.397			.374	
Primary Schools	.391				
Specialised Commerce	-.473	-.814			
Administration	-.784	-.416			
Monuments	-.809				
Banks, ATMs, Post	-.860				
Sports	-.919				
Hotels and hostels	-.923				
Restaurants	-.940				
Cultural centre	-.953				
Railway Station		.785			
Access Node		.593			
Physical attributes					
Gas (natural)			.740		
Cable TV			.736		
Floors			.585		
Type (Single unit=1, Flat=0)			-.473		
Duplex					
Total area				.794	
Number of bedrooms				.749	
Livingroom area					.630
Provision for garage					.575
Terrace					.478
Balcony					.434
Kitchen area					.432
Build and age (Used=1, New=0)					-.362
Percentage of variance explained	**37.60%**	**8.21%**	**6.48%**	**5.65%**	**5.45%**

The five factors provide a clear interpretation in terms of behavioural collections of housing characteristics. Factor 1 corresponds to several indicators of centrality related to the city centre – the loadings being higher for amenities that are closer to the CBD. The loadings are positive on characteristics measured in minimum distances, and negative on those measured in gravitational potential. Factor 2 also describes centrality, in this case negatively related to access to local amenities such as shopping malls, railway stations, supermarkets or motorway connections.

By contrast, factors 3, 4 and 5 represent the internal characteristics of dwellings. Factor 3 is related to a combination of attributes which, in the case of Aveiro, are strongly correlated: being a flat or a single unit housing, having a gas connection and Cable TV infrastructure[16] (high values of the factor correspond to flats with gas and cable TV). Factor 4 represents housing space, combining house size with the number of rooms, while factor 5 refers to additional desirable characteristics such as living room and kitchen area and the provision for garage.

Thus, the extracted factors represent behavioural collections of housing characteristics rather than statistical quantities with ambiguous interpretations (Maclennan, 1977).

3.3. Spatial heterogeneity

Next, ordinary least squares (OLS) regression was used to estimate hedonic pricing models allowing for spatial heterogeneity across the urban submarkets of Aveiro. Predicted orthogonal factors obtained above, including imputations for missing values, were used as explanatory variables. Of the 166 housing properties in the dataset, there were complete data for only 118 houses. The possibility of imputation for missing values in the factors marks one of the advantages of the factor based approach taken in this paper.

The total area of the house was also included as a regressor. Since the dependent variable is logarithm of price per unit area, the coefficient on this regressor (β_s) is expected to lie between zero and negative unity (0 and -1), with the interpretation that $1+\beta_s$ is the price elasticity of house area.

These regression models were estimated for the full sample as well as for each of the four submarkets defined by boundaries of administrative areas (parishes): Submarket 1 (Suburban: São Bernardo, Aradas and Santa Joana); Submarket 2 (Esgueira); Submarket 3 (Glória); and Submarket 4 (Vera Cruz). The final two are the most central areas encompassing the CBD of Aveiro, Glória being mostly residential while Vera Cruz caters to both the residential and service sector. Esgueira is partly urban and partly suburban. The estimated hedonic models reported in table 4 are parsimonious and offer good scope for interpretation, both in terms of individual coefficients and their variation across the submarkets. The variation in price elasticity of housing space and shadow prices of the factors capture our notion of spatial heterogeneity.

[16] Flats tend to be located in areas with high residential density, which in turn generate scale economies for the provision of these infrastructure facilities.

Table 4: Estimated factor based hedonic model with heterogeneity (urban scale)

Explanatory variables	Aggregate model *(All submarkets)*	Submkt. 1 *(Suburban)*	Submkt. 2 *(Esgueira)*	Submkt. 3 *(Glória)*	Submkt. 4 *(Vera Cruz)*
Intercept	**11.49** **(28.64)*****	**12.05** (10.90)***	**10.22** (11.18)***	**10.64** (13.93)***	**11.34** (11.43)***
Log Total area	**-0.94** **(-10.93)*****	-1.05 (-4.66)***	-0.70 (-3.51)***	-0.71 (-4.39)***	-0.90 (-4.19)***
Factor 1 *(Access to city centre)*	**-0.06** **(-3.76)*****	-0.03 (-0.59)	0.01 (0.18)	0.09 (-1.58)	-0.23 (-1.36)
Factor 2 *(Access to local amenities)*	**-0.00** **(-0.13)**	-0.03 (-0.77)	-0.06 (-1.23)	-0.06 (-1.22)	0.26 (1.49)
Factor 3 *(Type of dwelling)*	**-0.05** **(-3.17)*****	-0.09 (-2.14)**	-0.07 (-2.17)**	-0.03 (-0.83)	0.02 (0.31)
Factor 4 *(Housing space)*	**0.20** **(6.49)*****	0.26 (2.25)**	0.05 (-0.52)	0.16 (2.68)**	0.16 (1.63)
Factor 5 *(Additional desirable features)*	**0.21** **(10.92)*****	0.26 (4.49)***	0.27 (8.79)***	0.15 (4.57)***	0.19 (3.65)***
Number of observations	**166**	**42**	**42**	**27**	**55**
Adjusted R^2	**0.583**	**0.587**	**0.736**	**0.587**	**0.332**

*t-statistics in parentheses; *** significant at the 1% level/ ** significant at the 5% level/ * significant at the 10% level*

In the aggregate model, the explained variation (in terms of adjusted R^2) is quite high and all the regressors are highly significant, with the exception of factor 2. The signs of the coefficients agree with *a priori* expectations. The price per square meter decreases with distance to CBD and increases with factor 4 (housing space) and factor 5 (size of living room and kitchen and provision for garage). The negative coefficient on factor 3 implies a single unit housing is preferable even if it implies absence of cable TV or gas infrastructure. The relatively low price elasticity of living area, about 6 percent, conceals heterogeneity across submarkets.

Substantial spatial heterogeneity is observed across the 4 submarkets in terms of shadow prices for different factors related to physical and location characteristics, as well as the price elasticity of space. Analysis by submarkets shows important differences in the explanatory factors across the different areas of the city. Importantly, there is a substantial contrast between Vera Cruz and the other areas, showing that the traditional core of the city has a distinctive housing market. The estimates indicate that the elasticity of housing space is strongest in Glória and Esgueira, and weak in Vera Cruz and the suburban area.

While they are not statistically significant in any submarket,[17] access to both the centre and to local amenities is valued relatively highly in Vera Cruz. The suburban area has higher concentration of relatively larger detached houses without modern infrastructural facilities (like cable TV and gas); hence, housing space is the more important discriminator between dwellings. Likewise, detached (single unit) housing attracts additional value in the Esgueira submarket. By contrast, infrastructure is valued relatively more in Vera Cruz, even if the dwellings that are available are largely flats. The effect of factor 4 is the most heterogeneous across the submarkets, indicating that the importance ascribed to the number of rooms differs from area to area. The additional desirable features (factor 5 – living room and kitchen area and the provision for garage) attracts similar premium in all the 4 submarkets.

We evaluated forecast performance of the various models by cross-validation analysis, that is, by comparing each observation against the predicted value based on leave-one-out sample estimates omitting the index dwelling. In line with arguments in Malpezzi (2003), our factor based model generated better predictions compared to a model with a full set of hedonic characteristics. Based on the 118 houses with full data, the estimated factor hedonic model without imputed factors has a cross-validation mean squared error (MSE) that is 16 percent lower than that of a model with full hedonics included. The cross-validation MSE using predicted factors is 30 percent higher, but based on a substantially larger sample of 166 observations. On the whole, the factor based hedonic model has good predictive performance.

3.4. Spatial interaction – Distance and contiguity based spatial weights

We now turn to an examination of spatial interaction between submarkets. First, we take the standard approach in spatial econometrics by constructing spatial weights based on distances and contiguity. Before estimating the hedonic price models with spatial effects, we explore whether properties with similar square meter prices were more spatially clustered than normally expected, using Moran's I test. This test statistic for the presence of spatial dependence is:

[17] Given the small sample sizes in each submarket, it is not surprising that many regression coefficients are not statistically significant. The estimates indicate that, despite small sample sizes, it is important to allow for spatial heterogeneity. Further, the limitation of sample size is counterbalanced by the benefits of estimating spatial econometric models (spatial error and spatial lag models) by maximum likelihood, which is almost computationally impossible on large datasets.

$$I = \frac{n}{S} \frac{\sum_i \sum_j w_{ij}(x_i - \mu)(x_j - \mu)}{\sum_i (x_i - \mu)^2}, \tag{5}$$

where n is the number of observations; x_i and x_j denote the observed prices (€/m^2) at locations i and j respectively; μ denotes the mean price; w_{ij} the *(i,j)*-th element of the spatial weights matrix W, and S denotes the sum of all spatial weights: $S = \sum_i \sum_j w_{ij}$.

The above statistic is often used to analyse global spatial autocorrelation and is known to depend strongly on the assumed specification of the spatial weights matrix (Anselin, 1995). However, as discussed above, the choice of spatial weights in applications is often arbitrary and determined subjectively by the researcher, and there is usually very little formal evidence supporting such choice (Anselin, 2002).

To ensure robustness with regard to choice of the spatial weights matrix, we explored several specifications: binary weights based on distances between houses ranging from within 100, 500, 1000, 1500, 3000 and 5000 meters, as well as rook and queen contiguity. Table 5 reports the estimated Moran's I statistic[18] for these seven different specifications. Results for contiguity are visually illustrated in figure 2. The four quadrants in the figure provide a classification of different types of spatial autocorrelation: high-high (upper right) or low-low (lower left) for positive spatial autocorrelation; and high-low (lower right) or low-high (upper left), for negative spatial autocorrelation. Positive spatial autocorrelation implies that a high (low) value in the current location is surrounded by high (low) values in neighbouring observations. The slope of the best-fitting regression line is Moran's I statistic (Anselin, 2005). While Moran's I index is useful for detecting the presence of spatial autocorrelation, it does not indicate the precise structure of spatial interactions (Anselin, 2005).

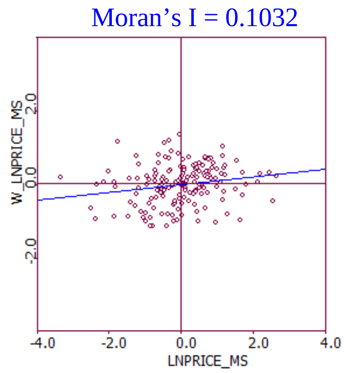
Based on the above measures of contiguity or distance, Moran's statistics tend to be positive but are not significant, showing little evidence of spatial autocorrelation in housing prices. Taking this evidence on face value, we would be tempted to conclude that geographically adjacent observations have little or no influence on house prices – a rather unsatisfactory observation.

[18] The value of the Moran's I statistic ranges from 1 (perfectly positive spatial autocorrelation) to –1 (perfectly negative spatial autocorrelation), a value near zero indicating no spatial autocorrelation.

Table 5: Moran's I test for 7 weighting matrices

Distances	Square meter price (€/m²)
d100	0.1669
d500	0.0952
d1000	0.0954
d1500	0.1001
d3000	-0.0533
d5000	0.2263
Queen/Rook	0.1032

Figure 2: Moran scatter plot for residuals (contiguity weight matrix)



Results was produced using GEODA software Anselin (2005)

Next, we used the GEODA software (Anselin, 2005; Anselin et al., 2006) to perform a series of Lagrange Multipliers (LM) tests (Anselin, 2005) for both spatial lag dependence and spatial error dependence. Specifically, in addition to OLS, we estimated by maximum likelihood (ML) alternative spatial regression models and investigated whether a spatial error or a spatial lag model, or indeed a model without spatial effects, best fit the data. We report the results for a spatial weights matrix based on rook and queen contiguity are reported in table 6; results for other specifications of spatial weights are similar.

Table 6: No spatial dependence, Spatial lag and Spatial error dependence estimates

Variables	No spatial dependence (OLS estimation)	Spatial lag model (ML estimation)	Spatial error model (ML estimation)
Intercept	11.49 (28.64)***	11.31 (14.66)***	11.55 (29.28)***
log Total area	-0.94 (-10.93)***	-0.94 (-11.18)***	-0.95 (-11.26)***
Factor 1	-0.06 (-3.76)***	-0.06 (-3.28)***	-0.06 (-3.42)***
Factor 2	-0.00 (-0.13)	-0.00 (-0.17)	-0.00 (-0.13)
Factor 3	-0.05 (-3.17)***	-0.05 (-3.18)***	-0.05 (-3.08)***
Factor 4	0.20 (6.49)***	0.20 (6.66)***	0.21 (6.60)***
Factor 5	0.21 (10.92)***	0.21 (11.06)***	0.22 (11.19)***
Lagrange Multiplier (lag)	0.08 (*p-value* 0.77)		
Robust LM (lag)	0.27 (*p-value* 0.61)		
Lagrange Multiplier (error)	0.67 (*p-value* 0.41)		
Robust LM (error)	0.86 (*p-value* 0.35)		
Lagrange Multiplier	0.94 (*p-value* 0.63)		
Number of observations	166	166	166
R^2	0.598	0.598	0.600
Log likelihood	20.404	20.442	20.753
Lag coefficient(Rho)		0.026 (*p-value* 0.78)	
Lag coefficient (Lambda)			0.109 (*p-value* 0.37)

*t-/z-statistics in parentheses; *** significant at the 1% level/ ** significant at the 5% level/ * significant at the 10% level*

Like the Moran's I statistics, we find no evidence of spatial dependence. This is despite the fact that we have not accounted for spatial heterogeneity in these estimates – a feature that can contribute to spatial dependence. Neither the LM-error (*p-value* 0.41) nor the LM-Lag (*p-value*

0.77) models are significant. The null hypothesis of both tests, which is the lack of spatial dependence, cannot be rejected at the 5 percent significance level.

Therefore, spatial dependence is either absent or not related to the geographical notions of distances and contiguity considered in the above seven specifications. This highlights an important limitation of spatial econometric methods for studying hedonic pricing models, arising from the treatment of spatial dependence as the outcome of spillover processes which are dependent on previously fixed and arbitrary spatial weights matrices (see also Giacomini and Granger, 2004). In other words, such evidence may have generated from ill-specified definitions of spatial weights, which is an issue we will investigate further in this paper.

3.5. Estimated spatial weights matrix

As discussed before, this paper extends an emerging area of research that takes a more general view on the nature and strength of spatial diffusion and cross section interaction. Moving away from the usual practice of *ex ante* definition of spatial interactions, we estimate the unknown spatial weights matrix that is consistent with an observed pattern of spatial dependence and is therefore suitable for interpretation. Specifically, as proposed in section 2, we use an extension of the Bhattacharjee and Jensen-Butler (2005) estimator to the pure cross-section setting and obtain estimates of a symmetric spatial weights matrix under a spatial error model. The symmetry assumption adopted in this work is in line with the traditional practice in housing studies, and is a natural consequence of defining spatial weights based on distances.

Table 7: Cross-Submarket Spatial Error Autocovariance and Autocorrelation matrix
(Variances reported on the diagonal, autocorrelations below diagonal)

Submarkets	1 (Suburb)	2 (Esgueira)	3 (Glória)	4 (Vera Cruz)
1 (Suburb)	**0.057**			
2 (Esgueira)	−0.042	**0.033**		
3 (Glória)	0.085	0.142	**0.050**	
4 (Vera Cruz)	−0.150	0.031	−0.079	**0.045**

The first step is to estimate the spatial autocovariance matrix of residuals across the four submarkets. As discussed in section 2, we use residuals across the four submarkets, matched by factors, to construct the cross-submarket error spatial autocovariance and autocorrelation matrix (table 7). In contrast to the results above based on *a priori* fixed spatial weights, significant spatial autocorrelation can be observed between some submarkets.

Table 8: Cross-Submarket Estimated Symmetric Spatial Interaction Matrix

Submarkets	**1** (Suburb)	**2** (Esgueira)	**3** (Glória)	**4** (Vera Cruz)
1 (Suburb)	**0.00**			
2 (Esgueira)	−0.024	**0.00**		
3 (Glória)	0.041***	0.074***	**0.00**	
4 (Vera Cruz)	−0.072***	0.017	−0.037	**0.00**

**** significant at the 1% level/ ** significant at the 5% level/ * significant at the 10% level*

Table 8 reports the corresponding estimated symmetric spatial weights matrix for cross-submarket interactions. Results are consistent with the spatial structure of Aveiro, showing that Vera Cruz has a highly significant negative interaction with the suburban area, while Glória has a highly significant positive interaction with both the suburban area and Esgueira.

These observations can be explained by the urban geography of Aveiro. Vera Cruz represents a distinct housing market in the CBD of Aveiro and draws its housing demand from a population quite different from the inhabitants in large detached houses in the suburban area. Such segmented markets imply that negative spatial interactions are likely between these two submarkets. On the other hand, Glória and Esgueira are largely residential submarkets close to the centre and are likely to offer positive spillovers, and likewise for Glória and the suburban area which are contiguous.

However, the best test for how these spatial weights relate to the urban housing market in Aveiro would come from spatial econometric models based on these, to which we turn next.

3.6. Spatial models with estimated spatial weights matrix

Finally, we estimate the within submarket spatial weight ω_0 and test the validity of the spatial error model with the spatial weights matrix estimated as above. The estimates for the factor based hedonic model (table 9) and for a corresponding model based on selected hedonic characteristics (table 10) are reported below.

The factor based model provides inferences on spatial heterogeneity that fall along the lines of the results discussed above. As expected, the estimated within submarket spatial weight, $\hat{\omega}_0 = 0.345$, is much larger than any of the estimated cross-submarket spatial weights. With the corresponding estimated spatial weights matrix $\hat{W}^0$, the data supports spatial dependence through the spatial error model, but evidence in favour of the spatial lag model is not statistically significant. This inference is markedly different from the previous case where contiguity and distance based spatial weights were used.

Table 9: Estimated factor based hedonic spatial dependence models

Variables	**Spatial error model**	**Spatial lag model**
Intercept	10.9248 (31.10)***	11.4615 (6.94)***
Total area (in logarithms)	–0.8128 (-10.85)***	–0.8510 (-9.19)***
Factor 1 (Access to city centre)		
x Suburban	–0.0527 (-1.20)	–0.0478 (-1.18)
x Esgueira	–0.1142 (-3.14)***	–0.1072 (-2.09)**
x Glória	–0.0581 (-1.46)	–0.0230 (-0.50)
x Vera Cruz	–0.2695 (-2.38)**	–0.2596 (-1.95)**
Factor 2 (Access to local amenities)		
x Suburban	–0.0204 (-0.74)	–0.0211 (-0.86)
x Esgueira	–0.0077 (-0.18)	–0.0141 (-0.27)
x Glória	–0.0076 (-0.15)	–0.0253 (-0.48)
x Vera Cruz	0.2846 (2.20)**	0.2716 (1.96)**
Factor 3 (Type of dwelling)		
x Suburban	–0.0831 (-1.93)*	–0.0834 (-2.48)**
x Esgueira	–0.0684 (-2.09)**	–0.0675 (-1.77)*
x Glória	–0.0111 (-0.27)	–0.0135 (-0.33)
x Vera Cruz	0.0309 (0.65)	0.0275 (0.53)
Factor 4 (Housing space)		
x Suburban	0.1756 (3.69)***	0.1876 (3.67)***
x Esgueira	0.0625 (1.26)	0.0809 (1.37)
x Glória	0.1736 (3.93)***	0.1898 (4.02)***
x Vera Cruz	0.1307 (2.51)**	0.1458 (2.46)**
Factor 5 (Additional desirable features)		
x Suburban	0.2379 (4.22)***	0.2409 (5.40)***
x Esgueira	0.2791 (9.05)***	0.2796 (7.78)***
x Glória	0.1782 (5.54)***	0.1773 (5.46)***
x Vera Cruz	0.1749 (5.44)***	0.1815 (5.14)***
Intra-market spatial wt., ω_0	0.345 (6.95)***	0.345 (6.95)***
Spat. err. autoregression, λ	–15.7176 (-6.01)***	$\rho =$ –0.0500 (-0.23)
Number of observations	166	166
Wald (LR) test – no spatial effect	36.08*** (24.72***)	0.051 (0.052)
Reg. test (spat. lag): $\rho = 0$	t = 0.03	

z-statistics in parentheses; *** significant at the 1% level/ ** significant at the 5% level/ * significant at the 10% level

Further, once spatial interaction through spatial error model is accounted for, the Born and Breitung (2009) regression test fails to reject the null hypothesis of no spatial lag. Thus, the spatial error model is appropriate for this application. This observation has important implications for inference and interpretation of spatial dependence in this application.

In particular, it would appear that the cost of not accounting for spatial dependence lies mainly in efficiency. Predictions using the spatial model provide a cross-validation MSE that is 11 percent lower than the OLS hedonic factor based model with full spatial heterogeneity. This suggests substantial efficiency gains in terms of prediction.

Table 10: Estimated hedonic spatial models using housing characteristics

Variables	**Spatial error model**	**Spatial lag model**
Intercept	9.9553 (13.97)***	42.1409 (3.58)***
Total area (in logarithms)	–0.7703 (-10.03)***	–0.6850 (-9.08)***
Factor 1 (Log – Distance from CBD)		
x Suburban	0.0099 (0.11)	–0.0581 (-0.80)
x Esgueira	–0.1039 (-1.56)	–0.1863 (-2.55)**
x Glória	–0.1091 (-1.09)	–0.1288 (-1.85)*
x Vera Cruz	–0.0276 (-0.35)	0.0657 (0.99)
Factor 2 (Log – Distance from superstore)		
x Suburban	–0.0342 (-0.61)	–0.0086 (-0.19)
x Esgueira	0.0447 (0.72)	–0.0007 (-0.01)
x Glória	0.0872 (1.25)	0.0788 (1.39)
x Vera Cruz	–0.0140 (-0.29)	–0.0137 (-0.30)
Factor 3 (Dummy – House, not flat)		
x Suburban	0.3884 (3.71)***	0.3395 (4.18)***
x Esgueira	0.5792 (6.07)***	0.5724 (6.62)***
x Glória	0.3772 (1.41)	0.3139 (1.48)
x Vera Cruz	0.4568 (3.87)***	0.4659 (4.07)***
Factor 4 (Number of bedrooms)		
x Suburban	0.0827 (1.16)	0.0773 (1.39)
x Esgueira	0.1503 (2.91)***	0.1352 (2.72)***
x Glória	0.1947 (4.05)***	0.1670 (3.72)***
x Vera Cruz	0.2093 (4.99)***	0.1948 (4.75)***
Factor 5 (Dummy – provision for garage)		
x Suburban	0.2512 (2.83)***	0.2119 (3.08)***
x Esgueira	0.2179 (3.46)***	0.1828 (2.84)***
x Glória	0.3698 (3.53)***	0.3317 (4.06)***
x Vera Cruz	0.2089 (2.75)***	0.1643 (2.26)**
Intra-market spatial wt., ω_0	0.472 (3.69)***	0.472 (3.69)***
Spat. err. autoregression, λ	–16.5915 (-5.29)***	$\rho =$ –4.5554 (-2.74)***
Number of observations	166	166
Wald (LR) test – no spatial effect	28.02*** (25.02***)	7.52*** (11.00***)
Reg. test (spat. lag): $\rho = 0$	$t = 0.00$	

*z-statistics in parentheses; *** significant at the 1% level/ ** significant at the 5% level/ * significant at the 10% level*

Inferences were largely similar when selected hedonic characteristics were used. In this case, both the spatial error model and the spatial lag model were individually supported by the data. However, the Born and Breitung (2009) regression test fail to reject the null hypothesis that there is no spatial lag evidence after spatial error dependence is accounted for.

Overall, our analysis points to important benefits in terms of understanding the spatial nature of the housing market in Aveiro at a finer urban scale. Both issues relating to spatial heterogeneity and spatial interactions are prominent in the urban submarkets defined by administrative boundaries. The empirical analysis provides a nice illustration of the proposed framework and methods for understanding spatial heterogeneity and dependence. In particular, whereas traditional measures of spatial weights based on distance or contiguity create a counterintuitive illusion of insignificant spatial dependence, the estimated spatial weights matrix is useful for modelling interactions and spillovers more appropriately. The sample size was somewhat limited, but provided the opportunity for ML estimation of spatial econometric models.

However, some inferences from our analysis are weak, particularly those related to factors representing access and centrality. On the one hand coefficients are not significant in most cases (see tables 4 and 9). On the other hand the restriction of the analysis to a limited number of parishes does not provide a general understanding of the housing market of Aveiro and its connection with the spatially heterogeneous social and economic conditions.

The consideration of the urban agglomeration corresponding to the municipalities of Aveiro and Ílhavo is more in line with the concept of metropolitan area discussed in Malpezzi (2003). Hence, following arguments in Maclennan and Tu (1996), we have thus extended our analysis to a broader spatial scale, and defined submarkets specifically segregated by location, housing quality and income levels.

4. ANALYSIS AT THE PERI-URBAN SCALE

Now, we extend our analysis to the housing market of the peri-urban area combining the municipalities of Aveiro and Ílhavo and the surrounding areas, using data for several years. This dataset covers a more heterogeneous area with more submarkets which, in addition to providing inferences on spatial dependence, enables a richer interpretation of the interaction pattern and of its underlying drivers.

Seven submarkets were selected, using a combination of criteria in line with Maclennan and Tu (1996) and Malpezzi (2003): administrative boundaries, urban structure, demographic features and history of urban development. Spatial contiguity of submarkets was generally preserved, but not always.[19] A short description of the selected submarkets is as follows.

- *Aveiro inner city*: the core of Aveiro city, including the administrative and service centre, as well as high density housing. This area has a higher concentration of more affluent residents.
- *City of Ílhavo*: the administrative centre of a separate municipality and corresponds to the second centre of the urban agglomeration of Aveiro.
- *Gafanhas*: a mixture of residential and industrial areas, and includes the most important port of Centro Region. The residential market combines older and consolidated settlements with detached houses spread in semi-urban areas. There is a marked predominance of working class and lower middle class residents.
- *Beaches*: an area with a high population density, corresponding to a strip of land stretching between the sea and the lagoon. Most houses are either second residences or used for rent in the high season.
- *Suburban Type A*: *a* group of small areas not very far away from Aveiro inner city. New planned residential areas dominate, being either blocks of flats or clusters of detached houses; these areas attract people from the Aveiro inner city looking for more affordable housing. Traditional social groups of people owning a small agricultural property and working either in manufacturing or in low skill service jobs have been gradually substituted by the above urban inhabitants.
- *Suburban Type B:* a combination of isolated new houses or blocks, typical of Suburban type A, with old rural settlements. The proportion of urban incomers, relative to traditional social groups, is lower than in Suburban Type A.
- *Suburban Type C*: Similar to Suburban type B but with a higher proportion of old rural settlements and traditional social groups.

The database used was provided by the biggest real estate agency of Portugal (Casa Sapo/ Janela Digital) and includes a set of hedonic and location characteristics similar to, but less detailed than, those presented in section 3. However it covers all the municipalities of Aveiro and Ílhavo and includes a much larger number of cases (12,476 dwellings) sold over a time span of about 10 years (October 2000 to March 2010).

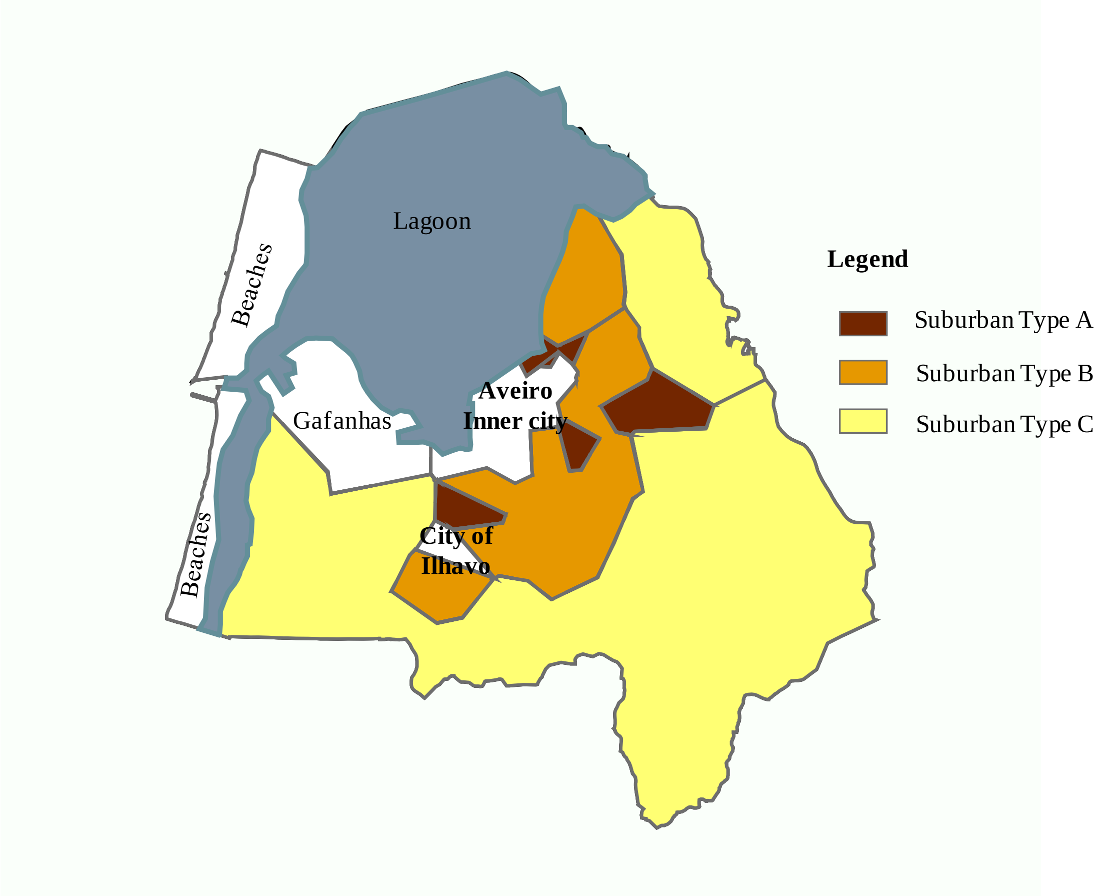
[19] Rothenberg et al. (1991) define submarkets in terms of bundle "quality" (that is, close hedonic substitutability), and these sets of close substitute units may or may not have any spatial content.

Table 11: Descriptive statistics of variables, peri-urban scale

	Variable	Units of measurement	N	Minimum	Maximum	Mean	Standard deviation
Internal physical characteristics							
d	Type	(Single unit=1, Flat=0)	12,467	0.00	1.00	0.28	0.45
ln	Number of bedrooms	(Number)	12,467	0.00	2.48	1.23	0.33
d	Duplex	(Yes=1; No=0)	12,467	0.00	1.00	0.12	0.33
d	Build and age: New building	(Yes=1; No=0)	12,467	0.00	1.00	0.31	0.46
d	Build and age: Under construction	(Yes=1; No=0)	12,467	0.00	1.00	0.25	0.43
d	Build and age: Restored	(Yes=1; No=0)	12,467	0.00	1.00	0.00	0.06
d	Build and age: Used building, less than 10 years	(Yes=1; No=0)	12,467	0.00	1.00	0.34	0.47
d	Build and age: Used building, 10-25 years	(Yes=1; No=0)	12,467	0.00	1.00	0.08	0.27
d	Build and age: Used building, more than 25 years	(Yes=1; No=0)	12,467	0.00	1.00	0.01	0.11
d	Build and age: Not restored	(Yes=1; No=0)	12,467	0.00	1.00	0.00	0.03
ln	Price	(Euros/m^2)	12,467	5.18	8.65	6.98	0.32
ln	Total area	(m^2)	12,467	3.00	6.40	4.88	0.48
ln	Time on the market (TOM)	(Days)	12,467	0.00	7.76	5.00	1.64
d	Balcony	(Yes=1; No=0)	12,467	0.00	1.00	0.39	0.49
d	Terrace	(Yes=1; No=0)	12,467	0.00	1.00	0.18	0.39
d	Provision for garage	(Yes=1; No=0)	12,467	0.00	1.00	0.16	0.37
d	Garage	(Yes=1; No=0)	12,467	0.00	1.00	0.64	0.48
d	Central heating	(Yes=1; No=0)	12,467	0.00	1.00	0.43	0.50
d	Fireplace	(Yes=1; No=0)	12,467	0.00	1.00	0.29	0.45
Location characteristics							
ln	Central Amenities	(Min. Dist.-meters)	12,467	5.42	11.97	8.02	0.83
ln	Local Amenities	(Min. Dist.-meters)	12,467	5.04	11.95	7.33	0.63
ln	CBD Aveiro	(Min. Dist.-meters)	12,467	5.23	11.98	8.08	0.80
ln	Local Commerce	(Min. Dist.-meters)	12,467	4.07	9.16	6.58	1.15
ln	Primary Schools	(Min. Dist.-meters)	12,467	3.65	7.59	5.60	0.83
ln	Intermediate Schools	(Min. Dist.-meters)	12,467	4.38	8.80	6.57	1.01
ln	University	(Min. Dist.-meters)	12,467	5.46	9.38	8.12	0.63
ln	Hospital	(Min. Dist.-meters)	12,467	5.39	9.34	7.84	0.88
ln	Health Centres	(Min. Dist.-meters)	12,467	4.78	9.16	7.15	0.87
ln	Pharmacies	(Min. Dist.-meters)	12,467	3.60	8.61	5.99	0.95
ln	Parks and Gardens	(Min. Dist.-meters)	12,467	3.97	8.84	7.04	0.95
ln	Rail Station	(Min. Dist.-meters)	12,467	4.41	9.22	7.55	0.99
ln	Access Node	(Min. Dist.-meters)	12,467	5.96	8.62	7.47	0.54
ln	Gas Station	(Min. Dist.-meters)	12,467	3.37	8.79	6.53	0.96
ln	Police	(Min. Dist.-meters)	12,467	5.39	11.97	7.84	0.81
p	Administration	(Potential)	12,467	2.02	8.71	6.28	1.10
p	Culture	(Potential)	12,467	5.24	8.05	6.46	0.69
p	Specialised Commerce	(Potential)	12,467	5.31	8.50	6.59	0.72
p	Restaurants	(Potential)	12,467	6.92	10.12	8.44	0.64
p	Hotels and hostels	(Potencial)	12,467	5.79	9.41	7.25	0.69
p	Monuments	(Potential)	12,467	7.37	9.90	8.35	0.45
p	Banks, ATMs, Post offices	(Potential)	12,467	6.64	9.80	8.41	0.68
p	Sports	(Potencial)	12,467	6.39	8.54	7.53	0.44
d	Sea/Beaches	(Yes=1; No=0)	12,467	0.00	1.00	0.07	0.25

d=dummy variable; ln= in logarithms; p=gravitational potential

Figure 3: Housing submarkets for Aveiro-Ílhavo at the peri-urban spatial scale



An important difference relative to section 3 is that the price data here refers to listing prices, rather than selling prices. We compensate for the wedge between listing and selling prices by including in our regressions the logarithm of time on the market (in days). In addition, we include time (yearly) fixed effects to control for aggregate cyclical and political factors.

Similar to the analysis presented in section 3, GIS tools were used to compute the proximity of each house to a number of central and local amenities. The variables and descriptive statistics are reported in table 11.

Following an identical methodology, we use maximum likelihood factor analysis with orthogonal varimax rotation. Following the arguments in Davies (1974) and Maclennan (1977), we verify that the resulting five leading factors align well with housing characteristics related to behavioural patterns (table 12). The factors are as follows:

- Factor 1: Access to the centre or central amenities.

- Factor 2: Access to local services and amenities (health centres, parks/gardens, etc.), also implying proximity to the traditional local centres within the area under study.[20]
- Factor 3: Access to beaches, schools and local commerce.
- Factor 4: Housing space.
- Factor 5: Additional facilities (garage, balcony, central heating, etc.).

Table 12: Factor loadings at peri-urban scale, varimax rotated (absolute value > 3.5)

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Location attributes					
Centrality, Central Amenities	.913				
CBD Aveiro	.907				
Hospital	.853				
University	.851	.368			
Police	.818				
Railway Station	.646		.521		
Access Node	.460				
Restaurants	-.702		.548		
Culture	-.752				
Sports	-.819	-.376			
Hotels and Hostels	-.844		.443		
Monuments	-.889				
Specialised Commerce	-.924				
Health Centres		.878			
Parks and Gardens		.858			
Gas Station	.432	.520			
Intermediate Schools	.494	.518			
Pharmacies	.363	.399			
Administration	-.563	-.601			
Banks, ATMs, Post	-.421	-.759			
Sea/Beaches			.849		
Primary Schools	.373		.690		
Local Commerce		.390	-.785		
Physical attributes					
Total area				.815	
Type (Single unit=1; Flat=0)	.353			.759	
Number of rooms				.753	
Build and age: Used building, less than 10 years				-.446	
Garage					.779
Balcony					.614
Central Heating					.575
Fireplace					.458
Provision for garage					.427
Percentage of variance explained	**25.02%**	**10.10%**	**8.03%**	**5.88%**	**4.91%**

[20] It should be noted that, the consolidation of a single urban area corresponding to the municipalities of Aveiro and Ílhavo was built on a territory previously organised as a set of small urban and rural clusters, each with its own provision of small scale services. Factor 2 reflects the proximity to such local centres.

The estimated hedonic model with spatial heterogeneity based on factors is reported in table 13. The results show substantial heterogeneity across the submarkets. Several important observations follow.

Table 13: Estimated factor based hedonic model with heterogeneity (peri-urban spatial scale)

Variables	**Aggregate model**	**CBD Aveiro**	**CBD Ílhavo**	**Gafanhas**	**Suburban Type A**	**Suburban Type B**	**Suburban Type C**	**Beaches**
Intercept	**9.890**	9.786	10.638	10.560	10.567	10.016	10.375	15.122
	(236.87)***	(101.66)***	(55.36)***	(72.19)***	(86.53)***	(115.73)***	(89.63)***	(-16.56)***
Log total area	**-0.598**	-0.571	-0.685	-0.761	-0.762	-0.614	-0.693	-0.871
	(-70.79)***	(-30.14)***	(-22.20)***	(-29.30)***	(-29.83)***	(-34.61)***	(-29.28)***	(-25.66)***
Log Time on the	**0.005**	0.006	0.016	0.003	0.011	0.004	-0.003	-0.007
market	**(3.69)*****	(2.10)**	(3.99)***	(0.98)	(3.19)***	(1.30)	(-0.90)	(-1.53)
Factor 1	**-0.043**	-0.036	-0.164	0.099	-0.144	-0.025	0.001	-1.761
(Access to city centre)	**(-19.77)*****	(-3.65)***	(-1.57)	(2.29)**	-(6.34)***	(-2.27)**	(0.13)	(-4.46)***
Factor 2	**0.027**	0.010	0.180	0.042	-0.079	-0.098	-0.029	-0.146
(Access, local amenities)	**(14.65)*****	(0.97)	(6.19)***	(2.04)**	(-7.06)***	(-7.58)***	(-2.17)**	(-0.84)
Factor 3	**0.077**	-0.016	-0.214	0.015	-0.120	-0.016	-0.005	-0.745
(Access to beaches)	**(38.21)*****	(-1.62)	(-2.78)***	(0.32)	(-4.31)***	(-1.29)	(-0.51)	(-5.48)***
Factor 4	**0.154**	0.199	0.217	0.209	0.242	0.162	0.171	0.211
(Housing space)	**(40.12)*****	(19.51)***	(15.64)***	(21.25)***	(20.65)***	(20.01)***	(15.60)***	(7.34)***
Factor 5	**0.043**	0.061	0.044	0.028	0.038	0.025	0.019	-0.002
(Additional facilities)	**(21.34)*****	(15.13)***	(6.17)***	(5.15)***	(7.21)***	(5.92)***	(3.28)***	(-0.09)
Time fixed effects	**YES**	YES	YES	YES	YES	YES	YES	YES
Number of obs.	**12,467**	**3,296**	**1,188**	**1,765**	**1,421**	**2,480**	**1,512**	**805**
Adjusted R^2	**0.572**	**0.359**	**0.459**	**0.483**	**0.557**	**0.498**	**0.484**	**0.557**

t-statistics in parentheses; *** significant at the 1% level/ ** significant at the 5% level/ * significant at the 10% level

First, Beaches is quite a distinct housing market from the others, in terms of estimated factor prices that are very different from the rest of the submarkets. Second, and in particular, the price elasticity of house area is the least for Beaches (0.129) and highest for the inner city of Aveiro (0.429). This implies that the size of houses designed for holidays and weekend purposes is not particularly valued, while the demand in the most affluent area (CBD of Aveiro) is considerably more sensitive to size.

Third, while the general model shows that prices increase with access to city centre, there is large variation across the different submarkets. In the CBD of Aveiro or suburban areas close to the city, the negative value attached to poor access to city centre is highly significant, while access is most valuable in the Beaches, which is the submarket located farthest from the centre. The same does not apply for the more remote Suburban Type C or submarkets such as Ílhavo or Gafanhas. This is explained by the different social profiles of inhabitants in these areas.

Fourth, access to local facilities has a heterogeneous effect on prices. However, by contrast to access to the centre, it is valued significantly, with the expected signs, only in Suburban Type A, B and C submarkets. This means that proximity to local centres is valued in the suburban areas, but not in CBD Aveiro, and even negatively valued in Ílhavo and Gafanhas. This is because local centres in the more urbanised locations tend to produce negative externalities such as noise or lack of parking space, while in the suburban areas they tend to be associated to better urban layouts and access to local amenities, different from that of unqualified suburban sprawl.

Fifth, additional facilities such as garage, balcony and central heating are positively valued with high significance everywhere, except in beaches, where such attributes do not matter. Finally, living space is positively valued, and in largely equal measure, across all the 7 submarkets.

In general, spatial heterogeneity is in line with the urban geography of Aveiro and reflects the dynamics of urban development, and its analysis is important to understand the spatial nature of the urban housing market and to provide guidelines for urban planning and housing policy. The unique character of the housing submarket in the Beaches is related to its evolution as the destination for second homes and rental properties for holiday-makers. Likewise, the high price elasticity for house area in the centre of Aveiro reflects scarcity rents. In turn, this shortage of housing space in the centre has led to migration from the city to the suburban areas, which have a larger price sensitivity to access. It would thus appear that further development of quality housing and good local amenities and access to the centre would make the suburban areas both affordable and desirable for the urban population.

The spatial structure of the urban agglomeration of Aveiro is also prominent in the analysis of spatial interaction based on the estimated cross-submarket symmetric spatial weights matrix (table 14). The first striking conclusion is that spatial interaction is significant for 17 out of 21 cells of the matrix. The main drivers of spatial interactions are common patterns of response to stochastic shocks; if for example, houses with particular characteristics become preferable for given social groups, we expect to obtain positive interactions between places with similar social structures and negative interactions for places where contrasted social groups dominate. Positive interactions between submarkets are also related to spillovers between these areas in unobservable housing characteristics, which in turn are prominent for areas contiguous in geographical location. Further, housing preferences similar across all submarkets are expected to generate an overall pattern of positive interaction.

Table 14: Estimated Symmetric Spatial Interaction Matrix (peri-urban spatial scale)

Submarkets	**CBD Aveiro**	**CBD Ílhavo**	**Gafanhas**	**Suburban Type A**	**Suburban Type B**	**Suburban Type C**	**Beaches**
CBD Aveiro	**0.00**						
CBD Ílhavo	0.0231**	**0.00**					
Gafanhas	–0.0089	0.0521***	**0.00**				
Suburban Type A	0.0415***	0.0495***	–0.0725***	**0.00**			
Suburban Type B	–0.0190***	0.0047	–0.0404***	0.0189***	**0.00**		
Suburban Type C	0.0227***	0.0984***	0.0263**	–0.0309**	0.0427***	**0.00**	
Beaches	0.0674***	0.0012	0.0328**	0.0062	0.0274**	0.0406***	**0.00**

*** *significant at the 1% level/* ** *significant at the 5% level/* * *significant at the 10% level*

As expected, contiguity or distance explains a number of the significant positive spatial weights across submarkets in Aveiro. These include: spatial weights between Beaches, Gafanhas and Suburban Type C; and between Suburban Type A and Suburban Type C on the one hand and CBD Aveiro, CBD Ílhavo and Suburban Type B on the other.

However, the spatial weights between some pairs of contiguous regions are not statistically significant or even negative (for example, between CBD Aveiro and Suburban Type B), and some other significant weights relate to non-contiguous regions. In other words, many significant spatial weights appear to be driven by reasons other than geographic distance or contiguity. Specifically, for some of these submarkets, positive spillovers appear to be related to a combination of the core-periphery relationship and socio-cultural distances. Examples include: CBD Aveiro and CBD Ílhavo; Beaches and CBD Aveiro; and CBD Ílhavo and Gafanhas.

Finally, table 14 indicates significant negative spatial interactions between CBD Aveiro and Suburban Type B, and between Suburban Type A and Suburban Type C. Apparently, both of these are related to market segmentation, where each submarket is attractive to different segments of the population.

Admittedly, some of the above explanations are tentative, and would require further research to confirm and interpret. These developments are beyond the scope of the paper. However, what we clearly show is that the spatial weights matrix, estimated based on our methodology, combined with the analysis of spatial heterogeneity, provides a very rich set of information which can be the basis for detailed analysis and for the disclosing of the causes underlying the observed spatial patterns.[21]

Finally, the above analysis at a larger spatial scale, in combination with previous analysis (based on central parishes), provides some insights about the importance of spatial scale. Largely focusing at the urban scale, our previous analyses provided useful inferences with regard to spatial heterogeneity and interactions across parishes. However, understanding of spillovers between the urban and suburban parishes was somewhat limited by the fact that the suburban area contained a heterogenous mix of neighbourhoods. This issue was addressed in the current analysis by dividing the suburban area into various notional submarkets that segregate the varieties of living space (Lefebvre, 1974 [1991]) in a more useful way. In this larger spatial scale too, very interesting inferences are drawn relating to spatial heterogeneity and interactions. This highlights the fact that, with regard to study housing submarkets, a single scale may not always be adequate (Whitehead, 2003).

5. Conclusion

In summary, our work here puts the connection between urban spaces and housing markets in a new framework and develops methodology for understanding urban housing markets in terms of three distinct but interconnected features of space – spatial heterogeneity, spatial dependence and spatial scale. Our methodology relies on factor based hedonic pricing analysis and offers many advantages in terms of interpretation, improved prediction and the facility to develop understanding of spatial interactions in more general terms.

Applied to the study of housing submarkets in the city of Aveiro, Portugal, the methodology offers a unique understanding of spatial aspects of the housing market. This is important for understanding neighbourhood choice, housing preferences, and the evolution of urban spatial structure. The implications of such studies on place based urban planning and housing policy that is informed by a clear understanding of the links between space and housing is a subject of ongoing and further studies; see Maclennan (2010) and Bhattacharjee et al. (2010).

---

[21] Paradoxically, the main reason mitigating against more formal analysis of spatial structure using the estimated spatial weights matrix is large sample size. Specifically, current methods do not allow for ML based inferences in spatial econometric models when sample size is large. Suitable methodology for large sample applications, based perhaps on regularisation or subsampling, is planned for the future.

References

Adair, A., McGreal, S., Smyth, A., Cooper, J. and Ryley, T. (2000). House prices and accessibility: the testing of relationships within the Belfast urban area. *Housing Studies* **15**(5), 699-716.

Andrews, D.W.K. (2005). Cross-section regression with common shocks. *Econometrica* **73**, 1551–1585.

Anselin, L. (1988a). *Spatial Econometrics, Methods and Models*. Kluwer: Dordrecht.

Anselin, L. (1988b). Lagrange Multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis* **20** (1), 1-17.

Anselin, L. (1990). *Spatial dependence and spatial structural instability in applied regression analysis*. *Journal of Regional Science* **30**, 185-207.

Anselin, L. (1995). Local indicators of spatial association – LISA. *Geographical Analysis* **27**, 93-115.

Anselin, L. (1999). Spatial econometrics. In: Baltagi, B.H. (Ed.) *A Companion to Theoretical Econometrics*, Blackwell: Oxford, 310-330.

Anselin, L. (2002). Under the hood: issues in the specification and interpretation of spatial regression models. *Agricultural Economics* **27**, 247-267.

Anselin, L. (2005). *Exploring Spatial Data with GeoDaTM: A Workbook.* https://www.geoda.uiuc.edu/pdf/geodaworkbook.pdf

Anselin, L. and Florax, R. (1995). (Eds.) *New Directions in Spatial Econometrics*. Springer-Verlag: Berlin.

Anselin, L. and Griffith, D.A. (1988). Do spatial effects really matter in regression analysis. *Papers, Regional Science Association* **65**, 11-34.

Anselin, L. and Lozano-Gracia, N. (2008). Errors in variables and spatial effects in hedonic house price models of ambient air quality. *Empirical Economics* **34**, 5–34.

Anselin, L., Lozano-Gracia, N., Deichmann, U. and Lall, S. (2010). Valuing access to water – a spatial hedonic approach, with an application to Bangalore, India. *Spatial Economic Analysis* **5**(2), 161-179.

Anselin, L., Syabri, I. and Kho, Y. (2006). GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis* **38**, 5–22.

Arbia, G., Espa, G., Giuliani, D. and Mazzitelli, A. (2009). Detecting the existence of space–time clustering of firms. *Regional Science and Urban Economics*, Forthcoming.

Arbia, G., Espa, G., Giuliani, D. and Mazzitelli, A. (2010). Cluster of firms on an inhomogeneous space: the case of high-tech manufacturing industries in Milan (Italy). *Mimeo*.

Archer, C.A. and Wilkinson, R.K. (1973). Measuring the determinants of relative house prices. *Environment and Planning* **5**, 357-367.

Basu, S. and Thibodeau, T. (1998). Analysis of spatial autocorrelation in house prices. *Journal of Real Estate Finance and Economics* **17**(1), 61–85.

Bhattacharjee, A., Castro, E.A. and Marques, J. (2010). Space and planning: some suggestions for positive theory and practice. *Mimeo*.

Bhattacharjee, A. and Holly, S. (2009). Understanding interactions in social networks and committees. Working Paper **1004**, Centre for Dynamic Macroeconomic Analysis, University of St Andrews, UK.

Bhattacharjee, A. and Holly, S. (2010). Structural interactions in spatial panels. *Empirical Economics*, Forthcoming.

Bhattacharjee, A. and Jensen-Butler, C. (2005) Estimation of spatial weights matrix in a spatial error model, with an application to diffusion in housing demand. CRIEFF Discussion Paper No. **0519**, University of St Andrews, UK.

Blundell, R., Chen, X. and Kristensen, D. (2007). Semi-nonparametric IV estimation of shape-invariant Engel curves. *Econometrica* **75**(6), 1613-1669.

Born, B. and Breitung, J. (2009). Simple regression based tests for spatial dependence. Bonn Economics Discussion Papers **23_2009**, University of Bonn.

Bourassa, S.C., Hamelink, F., Hoesli, M. and MacGregor, B.D.. Defining housing submarkets. *Journal of Housing Economics* **8**, 160-184.

Boyle, K. and Kiel, K. (2001). A survey of house price hedonic studies of the impact of environmental externalities. *Journal of Real Estate Literature* **9**(2), 117-144.

Butler, R.V. (1982). The specification of hedonic indexes for urban housing. *Land Economics* **58**(1), 96-108.

Can, A. and Megbolugbe, I. (1997). Spatial dependence and house price index construction. *Journal of Real Estate Finance and Economics* **14**, 203-222.

Casseti, E. (1972). Generating models by the expansion method: applications to geography research. *Geographical Analysis* **4**, 81-91.

Chattopadhyay, S. (1999). Estimating the demand for air quality: new evidence based on the Chicago housing market. *Land Economics* **75**, 22-38.

Davies, G. (1974). An econometric analysis of residential amenity. *Urban Studies* **11**(2), 217-226.

Fingleton, B. (2003). Externalities, economic geography and spatial econometrics: conceptual and modeling developments. *International Regional Science Review* **26**, 197-207.

Follain, J.R. and Malpezzi, S. (1980). *Dissecting Housing Value and Rent*. The Urban Institute: Washington, D.C..

Gabriel, S.A. (1984). A note of housing market segmentation in an Israeli development town. *Urban Studies*, 189-94.

Giacomini, R. and Granger, C.W.J. (2004). Aggregation of space-time processes. *Journal of Econometrics* **118**, 7-26.

Gillen, K., Thibodeau, T. and Wachter, S. (2001). Anisotropic autocorrelation in house prices. *Journal of Real Estate Finance and Economics* **23**(1), 5-30.

Grigsby, W.G., Baratz, M. and Maclennan, D. (1985). *The Dynamics of Neighborhood Change and Decline*. University of Pennsylvania Press.

Hall, P. and Horowitz, J. (2005). Nonparametric methods for inference in the presence of instrumental variables. *Annals of Statistics* **33**, 2904-2929.

Hartshorne, R. (1939). *The Nature of Geography: A Critical Survey of Current Thought in the Light of the Past*. Rand McNally: Chicago.

Jennrich, R.I. (2001). A simple general procedure for orthogonal rotation. *Psychometrika* **66**, 289-306.

Kain, J. and Quigley, J. (1970). Measuring the value of housing quality. *Journal of the American Statistical Association* **65**(330), 532-548.

Lancaster, K.J. (1966). A new approach to consumer theory. *Journal of Political Economy* **74**, 132-157.

Lefebvre, H. (1974 [1991]). *The Production of Space*. (Trans.) Nicholson-Smith, D., Blackwell: Oxford (UK).

Linneman, P. (1981). The demand for residence site characteristics. *Journal of Urban Economics*, 129-148.

Little, R.J.A. (1992). Regression with missing x's: a review. *Journal of the American Statistical Association* **87**(420), 1227-1237.

Liu, C. and Rubin, D.B. (1998). Maximum likelihood estimation of factor analysis using the ECME algorithm with complete and incomplete data. *Statistica Sinica* **8**, 729-747.

Maclennan, D. (1977). Some thoughts on the nature and purpose of hedonic price functions. *Urban Studies* **14**, 59-71.

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

Maclennan, D. (2010). Place, planning and economic policies: still missing links. *Mimeo*.

Maclennan, D. and Tu, Y. (1996). Economic perspectives on the structure of local housing systems. *Housing Studies* **11**(3), 387-406.

Malpezzi, S. (2003). Hedonic pricing models: a selective and applied review. Chapter 5, In: Gibb, K. and O'Sullivan, A. (Eds.), *Housing Economics and Public Policy: Essays in Honour of Duncan Maclennan*. Blackwell Science: Oxford (UK), 67-89.

Mills, E.S. and Simenauer, R. (1996). New hedonic estimates of regional constant quality housing prices. *Journal of Urban Economics* **39**(2), 209-215.

Mozolin, M. (1994). The geography of housing values in the transformation to a market economy – a case study of Moscow. *Urban Geography* **15**(2), 107-127.

Muller, J and Loomis, J. (2008). Spatial dependence in hedonic property models: do different corrections for spatial dependence result in economically significant difference in estimating implicit prices?. *Journal Agriculture and Research Economics* **33**(2), 212-231.

Ozanne, L. and Malpezzi, S. (1985). The efficacy of hedonic estimation with the annual housing survey: evidence from the demand experiment. *Journal of Economic and Social Measurement* **13**(2), 153-172.

Pace, R.K. and LeSage, J. (2004). Spatial econometrics and real estate. *Journal of Real Estate Finance and Economics* **29**, 147-148.

Palmquist, R.B. (2005). Property value models. Chapter 16, In: Ma¨ler, K. and Vincent, J. (Eds.) *Handbook of Environmental Economics Vol. 2*, North-Holland: Amsterdam, 763-813.

Pesaran, M.H. (2006). Estimation and inference in large heterogenous panels with multifactor error structure. *Econometrica* **74**, 967-1012.

Pesaran, M.H. and Tosetti, E. (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics* **161**, 182-202.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation pure competition. *Journal of Political Economy* **82**, 34-55.

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

Schaefer, F.K. (1953). Exceptionalism in geography: a methodological examination. *Annals of the Association of American Geographers* **43**, 226-249.

Simon, G. (1997). An angular version of spatial correlations, with exact significance tests. *Geographical Analysis* **29**(3), 267-278.

Soderberg, B. and Janssen, C. (2001). Estimating distance gradients for apartment properties. *Urban Studies*, **38**(1), 61-79.

Stewart, J. (1947). Suggested Principles of "Social Physics". *Science* **106**(2748), 179-180.

Straszheim, M. (1975). An econometric analysis of the housing market. NBER: New York.

Taylor, L. (2008). Theoretical foundations and empirical developments in hedonic modeling. In: Baranzini, A., Ramirez, J., Schaerer, C. and Thalmann, P. (Eds.), *Hedonic Methods in Housing Markets: Pricing Environmental Amenities and Segregation*. Springer.

Whitehead, C.M.E. (2003). The economics of social housing. In: Gibb, K. and O'Sullivan, A. (Eds.), *Housing Economics and Public Policy: Essays in Honour of Duncan Maclennan*. Blackwell Science: Oxford (UK), 135-152.