The role of space in urban housing market

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

Publication date:
2011

Document Version
Publisher's PDF, also known as Version of record

Link to publication in Discovery Research Portal

Citation for published version (APA):
Marques, J. L., De Castro, E. A., & Bhattacharjee, A. (2011). *The role of space in urban housing market*. Paper presented at 51st European Congress of the Regional Science Association, Barcelona, Spain. http://ideas.repec.org/p/wiw/wiwrsa/ersa11p1420.html

ERSA Congress
51st European Congress of the Regional Science Association
30th August - 3rd September 2011
Barcelona, Spain

The role of space in urban housing market

João Lourenço Marques
E-mail: jjmarques@ua.pt
Eduardo Anselmo de Castro
Department of Social, Juridical and Political Sciences
(Research Unit of Governance, Competitiveness and Public Policy
University of Aveiro, Portugal
Arnab Bhattacharjee
Department of Economic Studies
University of Dundee, UK

ABSTRACT

This paper emphasizes the quantitative analysis of space in relation to hedonic housing price models. Three aspects of space will be highlighted: i) spatial heterogeneity (spatial patterns): hedonic housing amenities may be valued differently in different locations which are related to specific housing sub-markets; ii) spatial dependence (spillovers): the degree by which price increases (or decreases) in a given sub-market is influenced by other sub-markets, or by another property within the same sub-market. iii) spatial scale: the study of heterogeneity and spillovers crucially depends on the level of geographical scale at which submarkets are defined.
In the literature the difficulty of defining sub-markets and understanding the relationship between them is broadly identified, and appropriate methods for defining housing markets are also presented. However, there is not a consensus on which methodologies should be used. As a contribution to understand spatial structure (heterogeneity and spillovers) in urban spaces some empirical results will be presented.
A new methodology to analyse spatial spillovers [rather than an ex ante definition of a spatial weight matrix (W)] will be developed. This procedure based on non-parametric approach will be applied to a rich database. An interesting outcome of this methodology is possibility of finding meaningful values of negative interaction.

1.INTRODUTION

This paper aims to analyze the importance of space in the housing market using a spatial hedonic pricing model applied to the urban and sub-urban areas of Aveiro.

Spatial interactions (spillovers effects), spatial heterogeneity and spatial scale are important aspects to analyse housing market. The first two aspects (heterogeneity and dependence) are widely stressed in the spatial econometrics literature; however, the common practice of representing the spatial interactions, using a weight matrix (W) a priori defined [1-8], has been often expressed as inadequate. The traditional approach to characterize spatial interactions is to define a matrix W, which represents theoretically and ex ante defined forces of interdependence (spatial autocorrelation), usually modelled by functions of distance or contiguity. Since the spatial interactions may be driven by other intangible factors (economic, social etc.) the choice of a spatial weight matrix based purely on geographic distances may be inappropriate. Accordingly, and in line with the notion of abstract space [9], it is presented, in this paper, a methodology to estimate an unknown spatial weight matrix. This approach allows the existence of negative effects on the spatial dependence, which may reflect segmented housing markets or asynchronous cycles of housing demand and supply. Based on a given definition of urban submarkets (or a fixed set of spatial locations) and panel data on these spatial units, some authors [4-5-6-7] have developed several methods to estimate the spatial weights matrix between the submarkets; those methods will be applied in this paper for the urban context of the region of Aveiro. The spatial scale not being so much a purely econometric issue, but rather an important empirical question, has been widely discussed in the literature of urban economics, for example, in [11]. The definition of the most appropriate territorial level (in terms of disaggregation and scope) to capture the relevant aspects of spatial patterns and spatial interactions is a key issue.

Thus, in order to contribute with some answers to the concerns previously presented, a summary presentation of the methodology and the results of empirical application is described.

2.METODOLOGY

The starting point of this analysis is the determination of a hedonic model that expresses the best explanatory power of the value of a dwelling, which is given by the following formulation:

$$\ln p = f(H, v) + \varepsilon$$

Where: p is the vector of the logarithms of household m prices (euros per m^2); v is the vector of hedonic prices, reflecting the weight of attributes in housing price explanation, H is the matrix that quantifies the attributes of dwellings (related to their intrinsic and their location characteristics); and ε is the vector that represents the stochastic component.

Note that for the quantification of matrix H, instead of the original independent variables, aggregate indicators of attributes are considered in this study, resulting from a principal components analysis (section 3.2 and 4.2). The use of factor scores in this application is useful because: i) since that hundreds of attributes to characterize a dwelling can be used, this multivariate technique retains the fundamental dimensions of the features considered essential to the model, leading to a more parsimonious estimated model and good scope for interpretation, ii) it also allows the imputation of missing values from observed values, iii) the factors are by its nature orthogonal and for this reason avoid multicollinearity problems, and finally, iv) it has the advantage of being crucial to the proposed methodology, the estimation of the unknown spatial weights matrices.

In this analysis three different aspects of space are analysed quantitatively: i) *spatial heterogeneity*, ii) *spatial dependence* (or spillover effects of spatial interaction), and iii) *spatial scale*. For this purpose two different databases have been used: a small dataset with 166 observations (section 3), covering only the urban area of Aveiro; and a much bigger dataset extended to the municipalities of Aveiro and Ílhavo (section 4).

i) **Spatial heterogeneity** (or spatial patterns) is related with the market segmentation of the housing characteristics. Parameters that are estimated by the regression model (v) are not constant across space (j) leading to structural differences in various housing markets that are expressed as follows:

$$\ln p_j = f(H_j, v) + \varepsilon_j$$

- Market segmentation for the first empirical analysis (subsection 3.3) is based on the administrative boundaries considering the limits of parishes as a criteria to define housing submarkets. In this sense, four segmented markets have been defined (Figure 1): Vera Cruz, Gloria, Esgueira and suburban area (which encompasses the parishes of São Bernardo, Santa Joana and Aradas).
- For the second analysis (subsection 4.3), which covers a wider range of territory, a set of criteria and principles reflecting several kinds of dimensions have been considered, such as, urban infrastructure, demographic and historical characteristics and urban socio-economic development. This analysis resulted into seven housing submarkets (Figure 4).

ii) In turn, **spatial dependence** is associated with interaction effects between submarket or single houses, i.e., when the hedonic prices of a house in a particular location depends on other observations located elsewhere. The functional specification of this spatial autocorrelation model can be found in the vast literature on spatial econometrics (see for example: [1-14]), and is the following:

$$\ln p = \rho W_1 p + Hv + \lambda W_2 \varepsilon + u$$

Where: W_1 and W_2 is a matrix of spatial weights, measuring the interaction between neighbouring sites; $W_1 p$ and $W_2 \varepsilon$ are spatial autoregressive components (spatial lag dependence term, and spatial error dependence, respectively); ρ and λ are the estimated spatial autoregressive coefficients that capture the influence of the average unit located nearby; μ is the vector of error terms [1-2]. The choice of spatial weights is a central issue in many applications of spatial interaction.

- At this level, and considering the spatial weights matrix as the term that defines the spatial dependence, two approaches have been adopted. The first approach (sub-section 3.4) largely follows the traditional assumption of an *ad hoc* matrix W, using distances and contiguity criteria. Hence, global tests for spatial autocorrelation (Moran's index) and more specific tests of spatial autocorrelation, such as spatial error dependence (SED) and spatial lag dependence (SLD) have been performed using GEODA software [3].
- The second approach (sub-section 3.4 and 4.4) adopts a nonparametric technique, which estimates the weight matrix not considering any initial restriction. Instead of using a predefined matrix W, the unknown weight matrix is estimated using statistical inference methods (see [6] for a detailed description of the assumptions and methodology development). The advantage of this method, when compared with the traditional approach, is that it does not consider restrictive assumptions concerning the effects of spatial dependence, providing unique opportunities for understanding the nature of interactions.

iii) Finally, we have the **spatial scale** which is closely related to the vertical spatiality of each of the aspects described above. The idea behind spatial scale is that both the spatial heterogeneity and dependence are strongly conditioned by the level of specification in which each phenomena is analysed.

- The use of two databases (section 3 and 4), with quite different levels of detail, allows of the comparison the previous analyses and the investigation of results robustness (spatial heterogeneity and spatial dependence).

The remainder of this paper is the presentation of the most important results of the analysis and its interpretation.

3.EMPIRICAL ANALYSIS 1: URBAN HOUSING MARKET OF AVEIRO

3.1. INITIAL DATA

The above approach is applied to the housing market in Aveiro (see [12] for more details on this database), a city located in the Centro Region of Portugal. The urban agglomeration of Aveiro includes the municipality with the same name and the neighbour municipality of Ílhavo and has a population of 114,000 inhabitants (2006). The present empirical analysis only refers to the city, which corresponds to 6 of the 14 parishes of the municipality of Aveiro (figure 1). The dataset includes 166 properties sold through one of the leading real estate agencies in Aveiro in 2007. The spatial distribution of the properties is presented in figure 1, where each house is indicated by a dot.

Figure 1- Housing locations of the sample 1

The data covers single-family homes (12.3%) and flats (87.7%), both new (11.8%) and used (88.2%), which are located in different urban and suburban areas. The variables collected are representative of physical and location attributes of dwellings. The choice of independent variables was somewhat limited by data confidentiality issues. Neighbourhood characteristics were defined by geographical distances from each property to the several facilities and services available within the city. We used Geographic Information System (GIS) to construct these location attributes. Descriptive statistics presented in table 1 reflect large variation in the attributes and substantial missing value problems.

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. In general, the potential (Pi) generated by a given set of services (S) in a given point (i) is:

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

Where, Sj is the service located in point j and dij is the distance between points i and j (see [15]).

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/m2). Housing prices are explained by a wide set of variables, some expressed in logarithms, related either to location characteristics or to internal physical attributes (see table 1).

Table 1 – Descriptive statistics of variables

		Units	N	Min	Max	Mean	Std. Deviation
Internal physical characteristics							
d	Type	(House=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	CATV	(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	Conservation	(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	Livingroom 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 characeteristics							
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	(Potencial)	166	5.49	9.09	6.89	0.72
p	Culture	(Potencial)	166	6.04	8.66	7.19	0.50
p	Specialised Commerce	(Potencial)	166	6.56	8.75	7.71	0.43
p	Restaurants	(Potencial)	166	7.80	10.15	8.90	0.54
p	Hotels and hostels	(Potencial)	166	5.48	8.15	6.72	0.65
p	Monuments	(Potencial)	166	7.95	10.90	8.71	0.48
p	Banks, ATMs, Post	(Potencial)	166	7.87	10.19	8.85	0.47
p	Sports	(Potencial)	166	7.04	8.81	7.88	0.38

d=variável dummy; ln= in logaritmo; p= potential gravitational

As variáveis apresentadas no quadro anterior foram submetidas a uma análise factorial, através do método de rotação Varimax, de modo a estimar factores ortogonais com o máximo poder explicativo.

3.2. DIMENSIONAL REDUTION OF DATA: FACTORIAL ANALYSIS

Principal components analysis was used to extract a set of orthogonal factors from the original housing attributes. The scree plot suggested five leading factors, which were then re-estimated by a rotated orthogonal varimax procedure. Taken together, the five factors explain 63.4% of the variance of all data. The extracted 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 2 – Factor loadings

Factor loadings (abs. value >3)					
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Culture	-.953				
Restaurants	-.940				
University	.930				
Hotels and hostels	-.923				
Central Amenities	.921				
Sports	-.919				
CBD Aveiro	.912				
Parks and Gardens	.876				
Banks, ATMs, Post	-.860				
Local Amenities	.839	-.413			
Monuments	-.809				
Local Commerce	.790				
Hospital	.788				
Administration	-.784	-.416			
Health Centres	.778				
High Schools	.733				
Pharmacies	.640	.367			
Police	.580	.426			
Gas Station	.397			.374	
Primary Schools	.391				
Specialised Commerce	-.473	-.814			
Railway Station		.785			
Access Node		.593			
Gas (natural)			.740		
CATV			.736		
Floors			.585		
Type (House=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
Conservation (Used=1, New=0)					-.362
Percentage of variance	37.60%	8.21%	6.48%	5.65%	5.45%

The five factors provide a clear interpretation. Factor 1 arises from several indicators of centrality related to the city centre (according to the loading signs the higher the score, the bigger the distance to CBD). Factor 2 also describes centrality, in this case related to spatial elements such as shopping malls, railway stations, hypermarkets or motorway connections (the higher the score, the lower the centrality). 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 particular case of Aveiro, interact strongly with each other: being a flat or a detached house, being connected to gas and CATV infrastructure (high values of the factor correspond to flats with gas and CATV); factor 4 combines housing size with number of rooms; factor 5 refers to additional elements such as the area of living room and kitchen or the existence of garage.

3.3. MARKET SEGMENTATION: 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 the explanatory variables for logarithm of price per unit area. The area of dwellings was also included as an additional regressor.

The regression models were estimated for the full sample as well as for each of the four submarkets defined by boundaries of administrative areas (parishes): Area 1 (Suburban: São

Bernardo, Aradas and Santa Joana); Area 2 (Esgueira); Area 3 (Glória); and Area 4 (Vera Cruz). The last two are the most central areas, being Glória mostly residential while Vera Cruz is both residential and service oriented, encompassing the CBD of Aveiro. Esgueira is partly urban and partly suburban.

The estimated hedonic models are reported in table 3. The dataset has 166 housing properties but only has complete data for 118. The missing values could be estimated under reasonable assumptions because initial variables were converted into factors. This is one of the advantages of hedonic pricing models based on factor analysis. Furthermore, the estimated models are parsimonious and offer good scope for interpretation.

Table 3 – The estimated coefficients of the hedonic model using the factors

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

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

In the aggregated model, independent variables explain 58.3% of the price variance, and all the regressors are highly significant, with the exception of factor 2. The signs of all coefficients have a logical explanation. The price per square meter decreases with area and with distance to CBD and increases with factor 5 (size of living room and kitchen and availability of garage). Because the contribution of factor 4 is controlled for area, the positive sign of the coefficient means that the higher the number of rooms the higher the price. Factor 3 coefficient means that local demand prefers detached houses even if this implies absence of CATV or gas infrastructure.

Substantial spatial heterogeneity is observed across the 4 submarkets in terms of shadow prices for different factors related to physical and location characteristics. Analysis by submarkets shows important and interesting differences in the explanatory factors across the several areas of the city. First of all 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. Looking at each explanatory variable we can see that the effect of area is similar and highly significant everywhere, stronger in suburban and weaker in Glória and Esgueira. Suburban is more specialised in big detached houses which face higher decreasing returns to size. The coefficient of factor 1 shows that distances to CBD are not significant in any submarket but highly significant in the aggregated model; then we can conclude that distance to CBD discriminates the four areas but is not important to discriminate houses inside each area. Factor 2 is generally not significant, showing that centralities related to this factor don't provide any marginal value. This means that in Aveiro, proximity to shopping malls or hypermarkets does not increase the value of properties. Factor 3 is only significant for areas 2 and 3: detached houses only have an added value in the outer parts of the city. Factor 4 provides heterogeneous results, showing that the importance ascribed to the number of rooms differs from area to area. Factor 5 provides similar results everywhere.

3.4. INTERACTION BETWEEN MARKETS: SPATIAL DEPENDENCE

We now turn to a study of spatial interaction between submarkets. First, we take the standard approach in spatial econometrics by constructing spatial weights based on distances and contiguity.

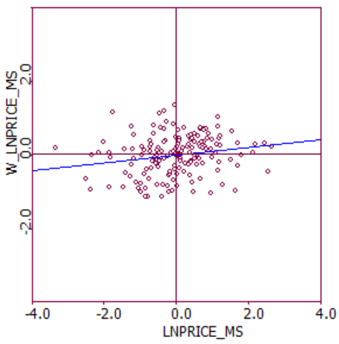
Before estimating the hedonic price models with spatial effects, we explore whether properties with similar square meter price were more spatially clustered than normally expected, using Moran's I test (table 4 e figure 2).

Table 4.- Moran's I test for 7 weighting matrices

Distances	Square meter price (€/m^2)
d100	0.1669
d500	0.0952
d1000	0.0954
d1500	0.1001
d3000	-0.0533
d5000	0.2263
Queen/Rook	0.1032

Figura 2.- Moran scatter plot for residuals (contiguity weight matrix)

Moran's I = 0.1032



GEODA software [3]

As discussed above, the choice of spatial weights matrix (W) is often arbitrary and subjectively and ad hoc defined.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 [1]. Table 4 reports the results from the Moran's I test for these seven different specifications. Results for contiguity are 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 (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 [3].

Table 5 – OLS, SLD and SED model estimates

	Ordinary Least Squares estimation	Spatial lag model – ML estimation	Spatial error model – ML estimation
Variable	Coefficient		
Constant	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		
R-squared	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)

Once again, there is no evidence of spatial dependence, even though spatial heterogeneity is not accounted for in the model (see table 7). For all seven weighting matrices, 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 ruled out. 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.
As discussed before, there is an emerging area of research that takes a nonparametric 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, such views are the basis for new methods of estimation of unknown spatial weights which are consistent with an observed pattern of spatial dependence and can be therefore subject to interpretation. Specifically, we use the modification to the pure cross-section setting proposed by [4] in order to 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 market studies, and is a natural consequence of defining spatial weights based on distances.

The first step of this exercise is to estimate the spatial autocovariance matrix of the residuals for the four sub-markets (Table 6). To estimate the spatial error autocovariance the following procedures should be done:

- Choose a housing submarket, lets say submarket Zj
    - Search another dwelling in other sub-market which has similar characteristics (factor scores), i.e., the identification of a "twin" dwelling in Z_i with $i \neq j$ (the minimum distance Euclidean has been used for this matching process).
- Consider the residual of the "twin" properties
- Calculate the autocovariance matrix

In contrast to the results provided by the traditional method (Table 5) values with highly significant autocorrelation have been obtained.

Table 6: Cross-Submarket Spatial Error Autocovariance and Autocorrelation matrix

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

Table 7 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 suburban while Glória has a highly significant positive interaction with both Suburb and Esgueira..

Table 7: 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

4. EMPIRICAL ANALYSIS 2:
URBAN AND SUBURBAN HOUSING MARKET OF AVEIRO AND ÍLHAVO

4.1. INITIAL DATA

The empirical analysis presented in section 3 was a good test for our non parametric spatial interaction model. In this section we will extend the analysis to the housing market of the urban area Aveiro/Ílhavo, using data for several years. This database covers a more heterogeneous area with more submarkets which, rather than merely providing evidence of spatial interaction, enables a richer interpretation of the interaction pattern and of its underlying causes.

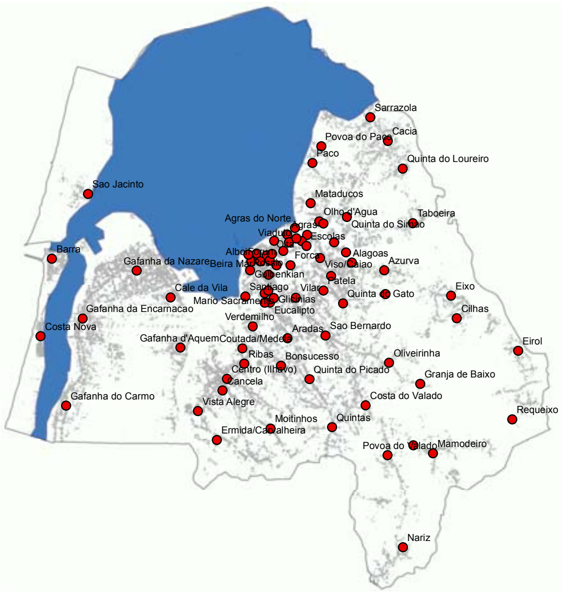


Figure 3. Georreferenciation of each zone

Seven submarkets were selected, using a combination of criteria in line with [10-11]: administrative boundaries, urban structure, demographic features and history of urban development. Spatial contiguity of submarkets was not always considered.
A short description of the selected submarkets follows.

- Aveiro inner city: it is 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: it is the administrative centre of a separate municipality and corresponds to a weaker form of the centrality provided by Aveiro.
- Gafanhas: corresponds to a mixture of residential and industrial areas, including also the most important port of Centro Region. The residential market mixes houses located in older and consolidated settlements with detached houses spread in semi-urban areas. There is a marked bias toward working class and lower middle class residences.
- Beaches: 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: 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 are absorbing people coming from the Aveiro inner city and looking for more affordable prices. 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 urban inhabitants referred to above.
- Suburban Type B: isolated new houses or blocks, typical of Suburban type A, are mixed with old rural settlements. The weight of urban incomers, relative to traditional social groups, is lower than in Suburban Type A.
- Suburban Type C: Equivalent to Suburban type B but with a higher proportion of old rural settlements and traditional social groups.

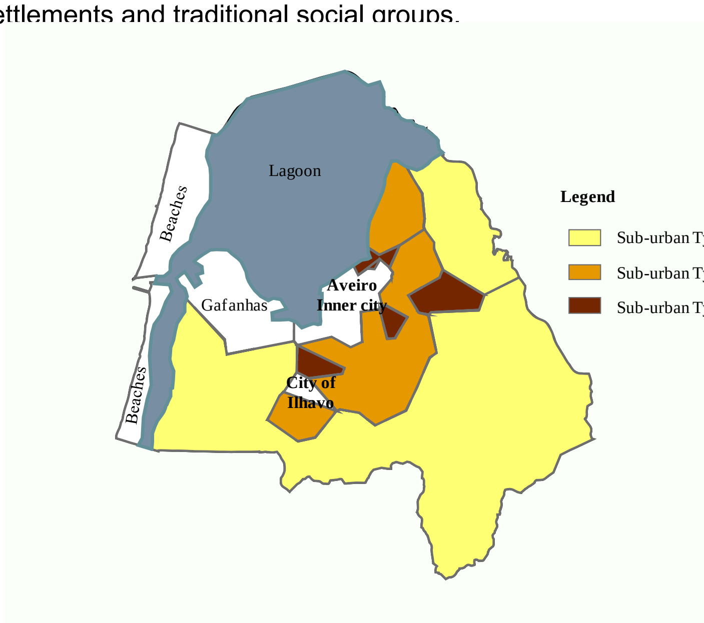


Figure 4. **eiro and Ílhavo**

The database used was provided by the biggest real estate agency of Portugal (Casa Sapo/ Janela Digital) and includes a set of variables similar to those presented in section 3. However it covers all the municipalities of Ílhavo and Aveiro and provides a much bigger number of cases (12476 dwellings) spread in a time span of 10 years.
An important difference relative to section 3 is that data refers to listing prices, rather than selling prices. We compensate for this effect including, as a control variable, the logarithm of time on the market (the time passed between the moment when the house starts to be advertised and the moment when it is removed from the site). In addition, we include yearly fixed effects to control for the aggregate cyclical and political factors. Another important difference compared to the other database, is the location of the property. In this particular situation houses are not individually georreferenciated but only the neighbourhood is known (Figure 3). Similarly to what was done for the analysis presented in section 3, GIS tools were

used to compute the proximity to a number of central and local facilities. The used variables are listed in table 8.

Table 8 – Variables included in the analysis

		Units	N	Min	Max	Mean	Std. Deviation
		Internal physical characteristics					
d	Type	(House=1, Flat=0)	12467	0.00	1.00	0.28	0.45
ln	Number of bedrooms	(Number)	12467	0.00	2.48	1.23	0.33
d	Duplex	(Yes=1; No=0)	12467	0.00	1.00	0.12	0.33
d	Preservation: New building	(Yes=1; No=0)	12467	0.00	1.00	0.31	0.46
d	Preservation: Under construction	(Yes=1; No=0)	12467	0.00	1.00	0.25	0.43
d	Preservation: Restored	(Yes=1; No=0)	12467	0.00	1.00	0.00	0.06
d	Preservation: Used building, less than 10 years	(Yes=1; No=0)	12467	0.00	1.00	0.34	0.47
d	Preservation: Used building, 10-25 years	(Yes=1; No=0)	12467	0.00	1.00	0.08	0.27
d	Preservation: Used building, more than 25 years	(Yes=1; No=0)	12467	0.00	1.00	0.01	0.11
d	Preservation: Not restored	(Yes=1; No=0)	12467	0.00	1.00	0.00	0.03
ln	Price	(Euros/m^2)	12467	5.18	8.65	6.98	0.32
ln	Total area	(m^2)	12467	3.00	6.40	4.88	0.48
ln	Time on the market (TOM)	(Days)	12467	0.00	7.76	5.00	1.64
d	Balcony	(Yes=1; No=0)	12467	0.00	1.00	0.39	0.49
d	Terrace	(Yes=1; No=0)	12467	0.00	1.00	0.18	0.39
d	Provision for garage	(Yes=1; No=0)	12467	0.00	1.00	0.16	0.37
d	Garage	(Yes=1; No=0)	12467	0.00	1.00	0.64	0.48
d	Central heating	(Yes=1; No=0)	12467	0.00	1.00	0.43	0.50
d	Fireplace	(Yes=1; No=0)	12467	0.00	1.00	0.29	0.45
		Location characteristics					
ln	Central Amenities	(Min. Dist.-meters)	12467	5.42	11.97	8.02	0.83
ln	Local Amenities	(Min. Dist.-meters)	12467	5.04	11.95	7.33	0.63
ln	CBD Aveiro	(Min. Dist.-meters)	12467	5.23	11.98	8.08	0.80
ln	Local Commerce	(Min. Dist.-meters)	12467	4.07	9.16	6.58	1.15
ln	Primary Schools	(Min. Dist.-meters)	12467	3.65	7.59	5.60	0.83
ln	Intermediate Schools	(Min. Dist.-meters)	12467	4.38	8.80	6.57	1.01
ln	University	(Min. Dist.-meters)	12467	5.46	9.38	8.12	0.63
ln	Hospital	(Min. Dist.-meters)	12467	5.39	9.34	7.84	0.88
ln	Health Centres	(Min. Dist.-meters)	12467	4.78	9.16	7.15	0.87
ln	Pharmacies	(Min. Dist.-meters)	12467	3.60	8.61	5.99	0.95
ln	Parks and Gardens	(Min. Dist.-meters)	12467	3.97	8.84	7.04	0.95
ln	Rail Station	(Min. Dist.-meters)	12467	4.41	9.22	7.55	0.99
ln	Access Node	(Min. Dist.-meters)	12467	5.96	8.62	7.47	0.54
ln	Gas Station	(Min. Dist.-meters)	12467	3.37	8.79	6.53	0.96
ln	Police	(Min. Dist.-meters)	12467	5.39	11.97	7.84	0.81
p	Administration	(Potencial)	12467	2.02	8.71	6.28	1.10
p	Culture	(Potencial)	12467	5.24	8.05	6.46	0.69
p	Specialised Commerce	(Potencial)	12467	5.31	8.50	6.59	0.72
p	Restaurants	(Potencial)	12467	6.92	10.12	8.44	0.64
p	Hotels and hostels	(Potencial)	12467	5.79	9.41	7.25	0.69
p	Monuments	(Potencial)	12467	7.37	9.90	8.35	0.45
p	Banks, ATMs, Post	(Potencial)	12467	6.64	9.80	8.41	0.68
p	Sports	(Potencial)	12467	6.39	8.54	7.53	0.44
d	Sea/Beaches	(Yes=1; No=0)	12467	0.00	1.00	0.07	0.25

d=variável dummy; ln= in logaritmo; p= potential gravitational

4.2. DIMENSIONAL REDUCTION OF DATA: FACTORIAL ANALYSIS

Following an identical methodology, we use a principal component's based factor analysis with orthogonal varimax rotation. The five leading factors align well with housing characteristics related to behavioural patterns (table 9). The factors are as follows:

- Factor 1 explains 25.02% of variance and defines dwellings' access to CBD type centralities.
- Factor 2 explains 10.10 % of variance and defines dwellings' access to special services and amenities such as health centres and parks/gardens.
- Factor 3 explains 8.03 % of variance and is related to location in beaches and to the types of accessibility which characterize beaches.
- Factor 4 explains 5.88 % of variance and is associated to the size of dwellings.
- Factor 5 explains 4.91 % of variance and is related to availability of additional elements such as garage or central heating.

Table 9 – Factor loadings

Factor loadings (abs. value >3)					
Attributes	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Specialised Commerce	-.924				
Centrality, Central Amenities	.913				
CBD Aveiro	.907				
Monuments	-.889				
Hospital	.853				
University	.851	.368			
Hotels and Hostels	-.844		.443		
Sports	-.819	-.376			
Police	.818				
Culture	-.752				
Restaurants	-.702		.548		
Rail Station	.646		.521		
Access Node	.460				
Health Centres		.878			
Parks and Gardens		.858			
Banks, ATMs, Post	-.421	-.759			
Administration	-.563	-.601			
Gas Station	.432	.520			
Intermediate Schools	.494	.518			
Pharmacies	.363	.399			
Sea/Beaches			.849		
Local Commerce		.390	-.785		
Primary Schools	.373		.690		
Centrality, Local Amenities					
Preservation: Used building, 10-25 years					
Total area				.815	
Type (House=1; Flat=0)	.353			.759	
Number of rooms				.753	
Preservation: Used building, less than 10 years				-.446	
Preservation: Under construction					
Preservation: New building					
Garage					.779
Balcony					.614
Central Heating					.575
Fireplace					.458
Provision for garage					.427
Terrace					
Duplex					
Preservation: Used building, more than 25 years					
Preservation: Restored					
Preservation: Not restored					
Total Variance Explained	25.02%	10.10%	8.03%	5.88%	4.91%

Extraction Method: Principal Component Analysis. Rotation Method: Varimax

4.3. MARKET SEGMENTATION: SPATIAL HETEROGENEITY

The estimated hedonic model with spatial heterogeneity based on factors is reported in table 14. The results show substantial heterogeneity across the submarkets. In particular the submarket beaches have coefficients which are particular by different from the rest of the area. Rather than exautive description of tables we provide a set of interesting examples:

- The minimum coefficient of log of total area is for Beaches while the highest is for Inner City of Aveiro. This means that the size of houses designed for holidays and weekend purposes is not particularly valued, whilst the demand in the most affluent area (Inner city of Aveiro) is considerable more sensitive to size.
- While the general model shows that prices increase with access to city centre, submarket's models tell different stories. In places like Aveiro or suburban areas close to the city, the negative value attached to poor access to city centre is highly significant, while the same does not apply in the more remote Suburban Type C or places such as Ílhavo and Gafanhas.
- The size of the living room or the provision of garage are positively valued with highly significance everywhere, except in beaches, where such attributes do not matter.

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.

Table 10 - Estimated factor based hedonic model, selected submarkets (peri-urban spatial scale)

	Aggregate model	CBD Aveiro	CBD Ílhavo	Gafanhas	Suburban Type A	Suburban Type B	Suburban Type C	Beaches
Constant	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 TOM	0.005	0.006	0.016	0.003	0.011	0.004	-0.003	-0.007
	(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
(Health Cent., Parks/ Gard)	(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
(Beaches)	(38.21)***	(-1,62)	(-2,78)***	(0,32)	(-4,31)***	(-1,29)	(-0,51)	(-5,48)***
Factor 4	0.150	0.199	0.217	0.209	0.242	0.162	0.171	0.211
(Size of dwelling)	(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
(Special dwelling charact.)	(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.	12467	3296	1188	1765	1421	2480	1512	805
Adjusted R Square	0.572	0.359	0.459	0.483	0.557	0.498	0.484	0.557

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

4.4. INTERACTION BETWEEN MARKETS: SPATIAL DEPENDENCE

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 15). 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 (very big living rooms and terraces) become fashionable 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. On the other hand, temporary fashions affecting all type of houses are expected to generate an overall pattern of positive interaction. For example, if in a given year the size of the kitchen tends to be more valued, those houses sold in this year, which have big kitchens will have positive error terms in all submarkets; conversely, houses with small kitchens will have negative error terms. Though this effect cannot be observed by time fixed effects, which only control for inflation, it creates a pattern of positive interaction in almost all cells. Does the pattern shown in table 15 reflect this general sensitivity to short term fashions? Do the few cases where negative interactions are detected reflect market segmentation? The development of such interpretation is beyond the scope of the paper. What the paper clearly shows is that the spatial interaction matrix, calculated accordingly to 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.

Table 11 - 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

Finally, the above analysis at a larger spatial scale, in combination with previous analysis (based on central parishes), provides some useful insights about the importance of spatial scale. Largely focusing at the urban scale, our previous analyses provided useful inferences with regard to 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 [9] that better segregate. 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 [16].

5. CONCLUSIONS

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 interaction 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, our 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 [7-10].

ACKNOWLEDGEMENTS

The authors are grateful to the Portuguese Science and Technology Foundation (FCT) for the financial support (Project: Drivers of housing demand in portuguese urban system - DONUT; and reference: PTDC/AUR-URB/100592/2008) and Janela Digital for providing the database.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Anselin, L., *Spatial econometrics: methods and models*. Kluwer Academic; Boston (1988).

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

3. Anselin, L., Syabri, I. Kho, Y., *GeoDa: An Introduction to Spatial Data Analysis*. Geographical Analysis; Vol. 38 (1), pp. 5-22, (2005).

4. Bhattacharjee, A. and Jensen-Butler, C., *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, (2005).

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

6. Bhattacharjee, A. and Holly, S., *Structural interactions in spatial panels*. Empirical Economics, Springer, vol. 40(1), pages 69-94, (2011).

7. Bhattacharjee, A. Castro, E. Marques, J., Spatial Interactions in Hedonic Pricing Models: The Urban Housing Market of Aveiro, Portugal, MIMEO.

8. Le Sage, J. and Pace, R., *Introduction to Spatial Econometrics*. Chapman & Hall/CRC, (2009).

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

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

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

12. Marques J., Castro E., Bhattacharjee A., *A localização urbana na valorização residencial: Modelos de autocorrelação espacial*. Actas do XV Encontro da APDR, (2009).

13. Marques, J., Castro E., Pinho, C., Batista, P., *O mercado habitacional - uma análise econométrica espacial*. Actas do XVI Encontro Nacional da APDR, (2010).

14. Marques, J., Castro, E., *Modelação do mercado de habitação*, Desafios emergentes para o Desenvolvimento Regional; Viegas, J.; Dentinho, T. (Eds.); Principia, (2010).

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

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