

Marques, Joao Lourenço; Castro, Eduardo; Bhattacharjee, Arnab; Batista, Paulo

Conference Paper

Spatial heterogeneity across housing sub-markets in an urban area of Portugal

52nd Congress of the European Regional Science Association: "Regions in Motion - Breaking the Path", 21-25 August 2012, Bratislava, Slovakia

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Marques, Joao Lourenço; Castro, Eduardo; Bhattacharjee, Arnab; Batista, Paulo (2012) : Spatial heterogeneity across housing sub-markets in an urban area of Portugal, 52nd Congress of the European Regional Science Association: "Regions in Motion - Breaking the Path", 21-25 August 2012, Bratislava, Slovakia, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/120764>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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

Spatial heterogeneity across housing sub-markets in an urban area of Portugal

João Lourenço Marques¹, Eduardo Castro², Arnab Bhattacharjee³, Paulo Batista¹
jjmarques@ua.pt, University of Aveiro, Portugal

¹ jjmarques@ua.pt, University of Aveiro, Portugal

² ecastro@ua.pt, University of Aveiro, Portugal

³ a.z.bhattacharjee@dundee.ac.uk, University of Dundee, UK

⁴ pauloricardolb@ua.pt, University of Aveiro, Portugal

ABSTRACT

Spatial homogeneity is a strong assumption in the hedonic housing price context; if not analyzed conveniently it can be a potential source of specification errors. Spatial heterogeneity occurs when a territorial segmentation exists in the housing market and, therefore, either the hedonic prices associated with different attributes or the characteristics of the houses are not constant over space. The evidence of recognition of housing sub-markets and the argument that caution should be exercised when interpreting the results of standard hedonic models has been identified early in the literature. Despite the argument that housing sub-markets should be adopted as a working framework, some ambiguity remains about how to deal with this issue. The early empirical works on sub-markets tended to be segmented into two major perspectives: those studies that adopt a supply side determinant focusing on the structural characteristics of dwellings and on neighbourhood characteristics; and those that focus on demand side determinants, such as, on household incomes or other demographic and socioeconomic characteristics. In this case the existence of distinct subgroups of demand is reflected on the territorial segmentation of the hedonic price vector. Thus, the objective of this communication is to present several approaches to define housing sub-markets. The analysis is applied to the urban area of Aveiro-Ílhavo in Portugal where demand and supply side views of defining housing sub-markets will be compared.

Key-words: *Spatial econometrics, spatial heterogeneity, housing market, hedonic model*

1. HEDONIC MODELS IN CONDITIONS OF SPATIAL HETEROGENEITY

The concept of spatial heterogeneity is particularly important in the context of the housing market, which is characterized by being segmented and structured by a complex pattern and inter-related elements, rather than being determined by a single, homogeneous process of spatial organization. Different social groups, with specific interests, preferences and economic capabilities tend to be organized into distinct territorial and socio-economic clusters, not necessarily geographically contiguous. Thus, the spatial heterogeneity, in the context of housing analysis, occurs when a segmentation in the housing market exists; and hence, the prices associated with different hedonic attributes or the housing characteristics are not constant across the territory. For example, it is reasonable to expect that households living in the centre of an urban area may value proximity to central facilities differently from those living on the periphery; likewise the implicit price of an additional bedroom in a leafy neighbourhood in the suburbs is likely to be different from that in the centre.

Contextualizing these concepts in terms of the hedonic model, spatial heterogeneity is related to the lack of stability over space, either by the choice of explanatory variables (H), or by their influence or importance (v) in the explanation of the housing price (P). Thus, the equation that represents variation of the implicit price of each of the attributes is defined as follows:

$$P_{iz} = c_z + \sum_{j=1}^J v_{jz} H_{izj} + \varepsilon_{iz} \quad \text{Eq.1}$$

Where:

P_{iz} is a vector of housing prices i (typically in a logarithmic form) located in the sub-market z . H_{izj} is the matrix containing j attributes for i dwellings, related to intrinsic characteristics and location. v_{jz} are the implicit prices (or shadow prices) of each attribute j in each sub-market z , and can be interpreted as the marginal contribution of each attribute j to the value of a dwelling (these attributes are contingent on the available information and on the perceptions of individuals that make housing decisions). These coefficients in terms of homogeneity are constant for any housing i and are independent of the value of H_{izj} . Assuming heterogeneity conditions these parameters should be estimated independently for each housing sub-market. c_z denotes the intercept; and ε_{iz} is the vector that represents the stochastic error or noise.

Despite the variety of techniques available in the field of spatial econometrics major challenges for empirically analyse spatial housing sub-markets are found (Marques, 2012).

2. SPATIAL HETEROGENEITY

The recognition of housing sub-markets and the argument that caution should be exercised when interpreting the results of standard hedonic models has been identified early in the literature (e.g. Rapkin, Winnick *et al.*, 1953; Grigsby, 1963; Straszheim, 1974; 1975; MacLennan, 1977; Quigley, 1979). Despite the argument that housing sub-markets should be adopted as a working framework, some ambiguity remains about how to deal with this issue. Watkins (2001) illustrates this difficulty suggesting five reasons to explain the failure of sub-market models. The first argument, mentioned by the author, is the difficulty of defining a housing sub-market, because a range of meanings can be adopted; second, even if theoretically there is consensus about the definition, there is little consensus about how sub-markets should be identified in practice; third, there is a large spatial variability regarding how urban areas are examined, making comparison difficult between studies; the fourth reason is related to the previous one, that is, the variability of the time period from which market data are drawn, which affects the market condition; and finally, the variety of statistical tests used to analyse the existence of sub-markets in different studies.

The definition of housing sub-market areas has proved a difficult problem. MacLennan *et al.* (1989) affirm this, saying that a housing market in most urban analysis can be considered as “(...) *a simple theoretical construct with no specific form and often it has no qualitative, temporal or spatial dimensions*”. Nevertheless, some definitions of a housing sub-market can be found in the literature, which are useful for delimiting the concept.

William G. Grigsby, considered a pioneer on the study of neighbourhood changes, pointed out, in his book published in 1963, “Housing Markets and Public Policy”, that sub-markets are distinctive because houses within them are viewed as (more or less)

perfect substitutes by the households. In this way, two dwellings are elements of the same sub-market “(...) *if the degree of substitutability between them is sufficiently great to produce palpable and observable cross-relationships in respect to occupancy, sales, prices and rents, or in other words, whether the units compete with one another as alternatives for the demanders of housing space*” (Grigsby, 1963 p.34). In fact, Grigsby followed the theoretical framework of Rapkin and Winnick (1953) defining housing sub-market as “(...) *the physical area within which all dwelling units are linked together in a chain of substitution*”, considering that “(...) *every dwelling unit within a local housing market may be considered a substitute for every other unit*” (Rapkin, Winnick et al., 1953, p.9-10). More recently the same concept of substitutability has been followed by Goodman and Thibodeau (2007a, p.4) that considered housing markets “(...) *as geographic areas where the price per unit of housing quantity (defined using some index of housing characteristics) is constant*”.

In practice, sub-markets can be analysed at three levels. The first is considered a macro scale approach, which includes works that adopt national areas (or at least large regions, or states) as the unit of analysis. Linneman (1981) and Struyk, Malpezzi *et al.* (1980) are some works that fall into this category. The second is a meso scale level, dealing with a regional/metropolitan approach, more or less coincident with the labour market and comprises works like Malpezzi, Ozanne *et al.* (1980); Goodman and Thibodeau (1998; 2007) and Fingleton (2008). Finally, the third approach is a micro level analysis, which examines sub-markets below the metropolitan area. Several works adopt this level as the unit of analysis to examine sub-markets, for example: Kiel and Zabel (1996); Maclennan and Tu, (1996); Bourrasa et al.,(1999b) and Clapp and Wang (2006).

There is a substantial literature that presents appropriate methods for defining housing markets. The common point of these approaches is the idea of finding areas in which the coefficients of hedonic price equations are similar. The question arises of how to analyse this similarity?

The early empirical works on sub-markets tended to be segmented into two perspectives: those studies that adopt a supply side determinant, and those that focus on demand side determinants (Goodman, 1998; Goodman and Thibodeau, 2007). Determinants that may be included on the supply side are: housing characteristics (this dimension includes structural characteristics of dwellings); and neighbourhood characteristics (e.g., public education, public safety, status or racial discrimination). On the other hand, if the focus is on the demand side, the determinants are based on household incomes or other demographic and socioeconomic characteristics. In this case the identification of distinct subgroups of demand is considered crucial to assess preferences and views of the two previous dimensions.

Thus, all three aspects (structural characteristics, spatial characteristics and subgroups of demand) can be used, separately or interactively, to determine sub-markets (Adair, Berry *et al.*, 1996; Maclennan and Tu, 1996).

The treatment of spatial heterogeneity is standard in the econometrics literature. It is commonly assessed with the use of Chow's F test, which examines whether the structural relationship between the dependent and explanatory variables is subject to some kind of change. Assuming the existence of spatial heterogeneity, there remains the problem of identifying and delimiting the various sub-markets. This can be done informally, using a priori knowledge of the geographical area under study, or by employing analytical methods (Nelson, 2008).

Typically pre-existing geographic or administrative boundaries, such as, census tract, zip code district, school district or local political jurisdictions, are used to define sub-markets even if they are considered an inappropriate way to deal with the problem. The reason for using administrative boundaries in some empirical work is typically because of data constraints rather than because of any belief that they are the most appropriate defining concept of the housing market. Works by Straszheim (1975), Goodman (1981), Goetzmann and Spiegel (1997) and Brasington and Hite (2005) are some examples which used political boundaries to define sub-markets. These procedures, based on a priori judgement to define sub-markets, are subject to a *posteriori* validation. Hedonic regressions are computed separately for each sub-market, and F tests then determine whether the resulting reduction in sum of squared residuals is significant. If it is, the sub-markets are assumed to be appropriate (Goodman and Thibodeau, 1998). The criticism to this approach is that the housing segmentation is imposed rather than given by the data (modelled).

In spite of the traditional administrative boundaries, other methods can be applied: hedonic equations to identify sub-markets, for instance. The principle is to use implicit prices of housing to identify areas with similar characteristics and then aggregate those areas into sub-markets. A strategy commonly used is the application of dummy variables to describe each sub-market in the hedonic specification, rather than estimating a separate hedonic equilibrium for each sub-market (Castro, Marques *et al.*, 2011). Once the sub-markets are defined, slope and intercept dummy variables for each sub-market should be included in the model. Significant differences in slope across sub-markets indicate spatial heterogeneity in implicit prices, in other words, the hedonic prices associated with different attributes vary in whole or in part across the sub-markets. Likewise, statistical significance of the intercept dummies indicates unobserved heterogeneity in the fixed effects across different sub-markets. The hedonic equation including dummy variables takes the form:

$$p_i = \sum_{z=1}^Z d_{zi} + \sum_{j=1}^n \sum_{z=1}^Z v_{jzi} h_{ji} d_{zi} + \varepsilon_i \quad \text{Eq.2}$$

The main disadvantage of this method is the large number of implicit prices to be estimated, which requires large sample size. However, if there is lack of spatial homogeneity, there is no simpler alternative. To ignore such heterogeneity would result in specification (omitted variable) bias and a serious failure to account for spatial structure. Examples of the use of hedonic approach to identify space heterogeneity can be found in: Goodman (1981), MacLennan and Tu (1996), Goodman and Thibodeau (1998), Bourassa, et al., (1999a) and Bourassa et al. (2007).

It is also frequent the application of principal component analysis (PCA) and cluster analysis (CA), as a complement of hedonic approach to identify local market areas. The approaches using PCA and CA do not depend on a priori boundary definition, but rely on the structure of data. Factor scores obtained using principal components are used in cluster analysis to find groups of homogeneous observations, which result in sub-markets that do not impose contiguity.

In these multivariate analyses some other techniques may be included: hierarchical models (Goodman and Thibodeau, 1998; Raudenbush and Bryk, 2002; Goodman and Thibodeau, 2003), and mixtures of linear models (Ugarte, Goicoa *et al.*, 2004). Spatial partitions based on socio-economic or environmental characteristics, as shown in Galster (1987), Hårsman et al. (1995) and Schnare et al. (1976; 1980) are simple methods that can be also used to define sub-markets. The non-parametric spatial

statistical methods developed in Clapp et al. (2006), Bhattacharjee and Jensen-Butler (2005) and Bhattacharjee and Holly (2010a; b) are more sophisticated approaches to delineate sub-markets [these methods consider residual spatial autocorrelation - see Bhattacharjee, Castro *et al.*, (2012); Marques (2012)]. The previous methodologies focused on the statistical techniques to determine housing sub-markets, however it can be effected subjectively given expert knowledge, that is, delineated by real estate agents or appraisers (e.g.: Palm, 1978; Michaels and Smith, 1990).

In the next session some of these perspectives will be explored and applied in the context of the housing market of Aveiro and Ílhavo.

3. HOUSING SEGMENTATION FOR AVEIRO-ÍLHAVO

In this section housing sub-markets of Aveiro are empirically defined and analyzed. The study area is located in the Centro Region of Portugal and includes two municipalities; Aveiro and Ílhavo (see Figure 1).

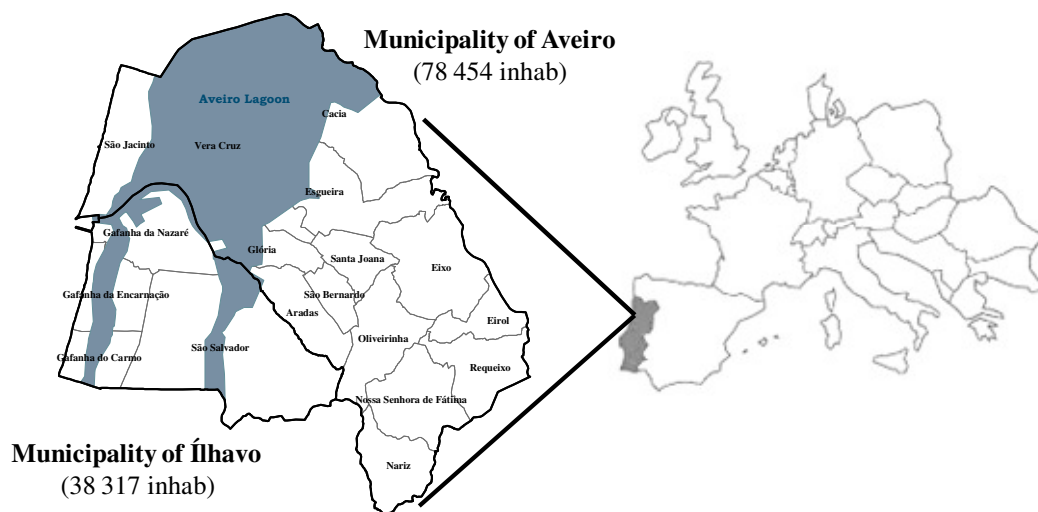


Figure 1 – Location of the study area: Municipalities of Aveiro and Ílhavo

The municipality of Aveiro has a total area of 200 km² and a total population of 78454; the municipality of Ílhavo latter has an area of 75km² and 38317 inhabitants (2011 Census). If the area of the lagoon is removed, the population density is 600 inhabitants per km², a typical value for an urban agglomeration.

The database used for this empirical work is provided by the firm Janela Digital S.A., which owns and manages the real estate portal database CASA SAPO. This portal is the largest site in Portugal of real estate diffusion. Since 2000 (from October 2000 to March 2010) collated about 4 million records of properties available for transaction in Portugal covering all the national territory. For the specific case of Aveiro and Ílhavo 47188 properties entered in the database between 2000 and 2010, as described in the Figure 2. The models used in this empirical work consider a database with 12467 observations, which is a number obtained after data cleaning.

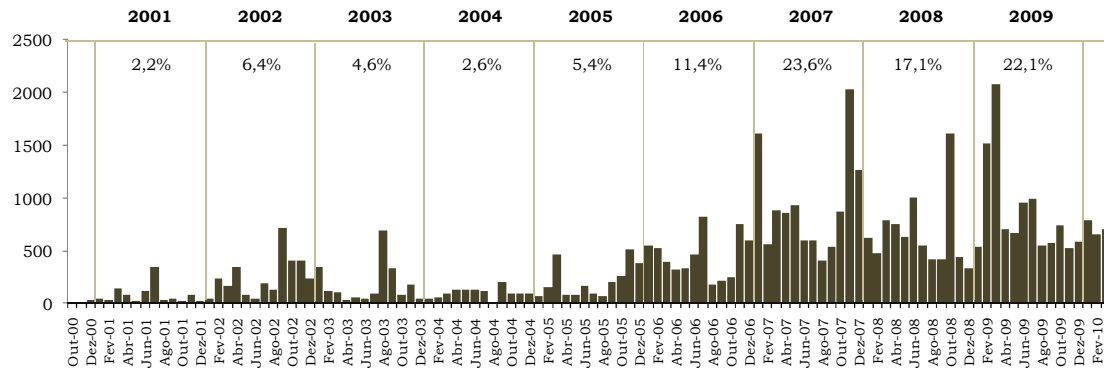


Figure 2 – Number of properties added to the database between 2000 and 2010

For a more detailed description of the main challenges for cleaning the database and procedures used for the construction of new housing attributes (intrinsic and location and neighbourhood) see Marques (2012). The systematization of this information is presented in Table 1 and is used as a starting point for the analysis of sub-markets of the housing stock of Aveiro and Ílhavo.

Table 1 – Descriptive statistics of physical and location housing attributes

	Units	N	Min	Max	Mean	Std. Deviation
Intrinsic physical characteristics						
<i>d</i> Type	(House=1, Flat=0)	12467	0.00	1.00	0.28	0.45
<i>ln</i> Number of bedrooms	(Number)	12467	0.00	2.48	1.23	0.33
<i>d</i> Duplex	(Yes=1; No=0)	12467	0.00	1.00	0.12	0.33
<i>d</i> Preservation: New building	(Yes=1; No=0)	12467	0.00	1.00	0.31	0.46
<i>d</i> Preservation: Under construction	(Yes=1; No=0)	12467	0.00	1.00	0.25	0.43
<i>d</i> Preservation: Restored	(Yes=1; No=0)	12467	0.00	1.00	0.00	0.06
<i>d</i> Preservation: Used building, less than 10 years	(Yes=1; No=0)	12467	0.00	1.00	0.34	0.47
<i>d</i> Preservation: Used building, 10-25 years	(Yes=1; No=0)	12467	0.00	1.00	0.08	0.27
<i>d</i> Preservation: Used building, more than 25 years	(Yes=1; No=0)	12467	0.00	1.00	0.01	0.11
<i>d</i> Preservation: Not restored	(Yes=1; No=0)	12467	0.00	1.00	0.00	0.03
<i>ln</i> Price	(Euros/m2)	12467	5.18	8.65	6.98	0.32
<i>ln</i> Total area	(m2)	12467	3.00	6.40	4.88	0.48
<i>ln</i> Time on the market (TOM)	(Days)	12467	0.00	7.76	5.00	1.64
<i>d</i> Balcony	(Yes=1; No=0)	12467	0.00	1.00	0.39	0.49
<i>d</i> Terrace	(Yes=1; No=0)	12467	0.00	1.00	0.18	0.39
<i>d</i> Provision for garage	(Yes=1; No=0)	12467	0.00	1.00	0.16	0.37
<i>d</i> Garage	(Yes=1; No=0)	12467	0.00	1.00	0.64	0.48
<i>d</i> Central heating	(Yes=1; No=0)	12467	0.00	1.00	0.43	0.50
<i>d</i> Fireplace	(Yes=1; No=0)	12467	0.00	1.00	0.29	0.45
Location and neighbourhood characteristics						
<i>ln</i> Central Amenities	(Min. Dist.-meters)	12467	5.42	11.97	8.02	0.83
<i>ln</i> Local Amenities	(Min. Dist.-meters)	12467	5.04	11.95	7.33	0.63
<i>ln</i> CBD Aveiro	(Min. Dist.-meters)	12467	5.23	11.98	8.08	0.80
<i>ln</i> Local Commerce	(Min. Dist.-meters)	12467	4.07	9.16	6.58	1.15
<i>ln</i> Primary Schools	(Min. Dist.-meters)	12467	3.65	7.59	5.60	0.83
<i>ln</i> Intermediate Schools	(Min. Dist.-meters)	12467	4.38	8.80	6.57	1.01
<i>ln</i> University	(Min. Dist.-meters)	12467	5.46	9.38	8.12	0.63
<i>ln</i> Hospital	(Min. Dist.-meters)	12467	5.39	9.34	7.84	0.88
<i>ln</i> Health Centres	(Min. Dist.-meters)	12467	4.78	9.16	7.15	0.87
<i>ln</i> Pharmacies	(Min. Dist.-meters)	12467	3.60	8.61	5.99	0.95
<i>ln</i> Parks and Gardens	(Min. Dist.-meters)	12467	3.97	8.84	7.04	0.95
<i>ln</i> Rail Station	(Min. Dist.-meters)	12467	4.41	9.22	7.55	0.99
<i>ln</i> Access Node	(Min. Dist.-meters)	12467	5.96	8.62	7.47	0.54
<i>ln</i> Gas Station	(Min. Dist.-meters)	12467	3.37	8.79	6.53	0.96
<i>ln</i> Police	(Min. Dist.-meters)	12467	5.39	11.97	7.84	0.81
<i>p</i> Administration	(Potencial)	12467	2.02	8.71	6.28	1.10
<i>p</i> Culture	(Potencial)	12467	5.24	8.05	6.46	0.69
<i>p</i> Specialised Commerce	(Potencial)	12467	5.31	8.50	6.59	0.72
<i>p</i> Restaurants	(Potencial)	12467	6.92	10.12	8.44	0.64
<i>p</i> Hotels and hostels	(Potencial)	12467	5.79	9.41	7.25	0.69
<i>p</i> Monuments	(Potencial)	12467	7.37	9.90	8.35	0.45
<i>p</i> Banks, ATMs, Post	(Potencial)	12467	6.64	9.80	8.41	0.68
<i>p</i> Sports	(Potencial)	12467	6.39	8.54	7.53	0.44
<i>d</i> Sea/Beaches	(Yes=1; No=0)	12467	0.00	1.00	0.07	0.25

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

The database used in this research, besides the price of the property, considers two main categories of variables to describe each dwelling: i) the intrinsic physical attributes, and ii) the location and neighbourhood of the building. The first group of variables characterizes each dwelling in terms of number of rooms, preservation, age and area etc.. A set of other physical housing characteristics, obtained from a free text field where real estate advertisers describe the property, was used (see Marques (2012)). The second group of attributes is related to the housing location and to the characteristics of

the neighbourhood. Housing location is described by the municipality, parish or zone. This last territorial disaggregation is a relatively homogeneous area, usually smaller than the parishes (see Figure 3). The option of not considering directly these zones to define housing sub-markets is because of low number of observations in each unit of analysis.

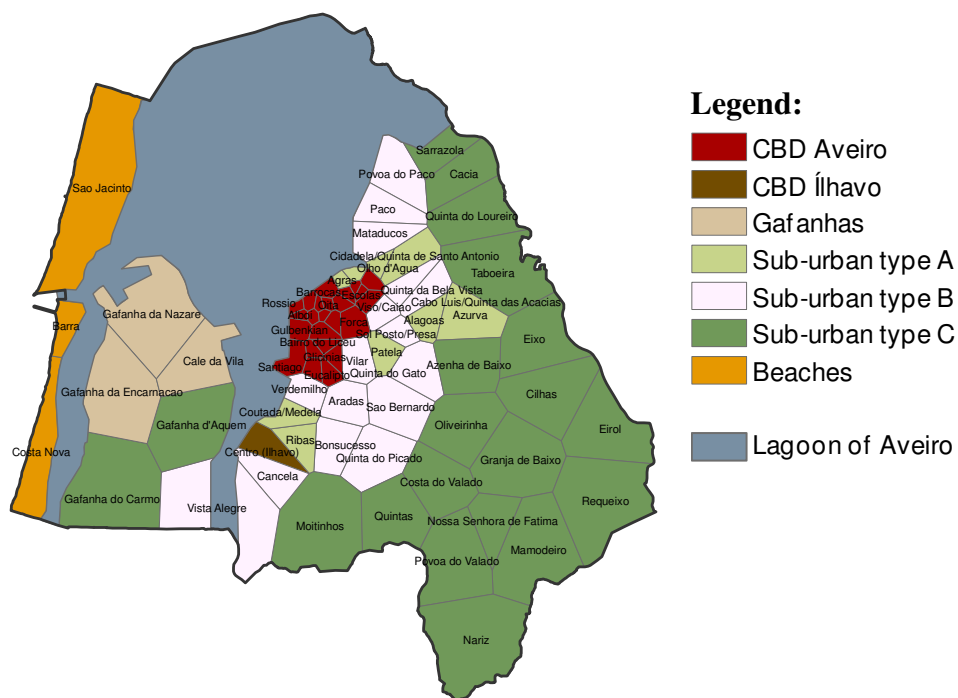
Figure 3 - Housing location by zones

From the previous section emerges the idea that the delimitation of housing sub-markets is generally defined in terms of: i) physical characteristics of housing, ii) geographic aspects, iii) socio-economic characteristics of the neighbourhood, or inductive approaches based on iv) tacit knowledge of estate agents or v) administrative boundaries or geographic pre-existing divisions.

In this exploratory analysis several approaches to assess the consistency of housing sub-markets are applied; and are organized into two groups: i) an inductive approach and ii) an analytical approach. The first approach consists of defining *ex ante* criteria based on the empirical knowledge of the case study, while the second approach uses a spatial clustering analysis applied to different criteria (allowing an identification of the patterns in the data highlighting their similarities and differences).

3.1. INDUCTIVE APPROACH

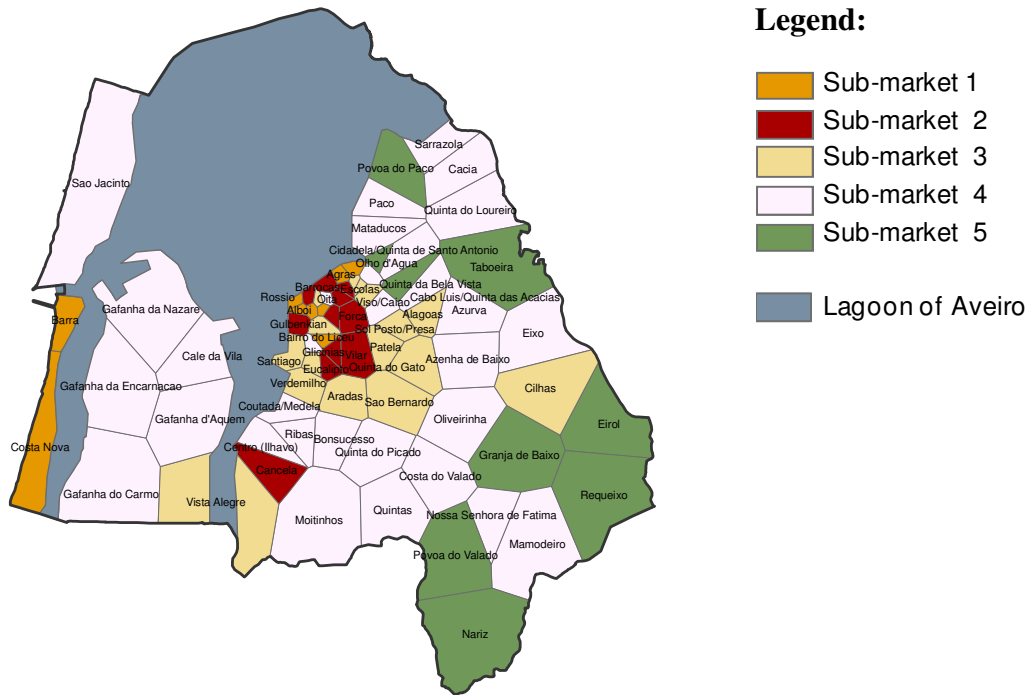
Trying to preserve, whenever possible the spatial contiguity the five following criteria to define and delimit sub-housing markets were considered: administrative boundaries, urban structure, demographic and historical characteristics and urban development (for more detail on how these criteria were applied see Marques, 2012). The result of this housing segmentation is shown in Figure 4.



3.2. ANALITICAL PERSPECTIVE

The results presented below are based on cluster analysis. Thus, a first decision, before applying the cluster analysis is to define the dimensions which are really important to define sub-housing markets. According Watkins (2001) and Bourassa, Cantoni et al. (2008) the alternatives for characterizing and delineating housing sub-markets are: i) spatially stratified housing sub-markets, where geographic and locations attributes are key aspects to delimit housing segmentation; ii) structural characteristics sub-markets, which are based on the similarity of intrinsic housing characteristics; and iii) nested spatial/structural sub-market, which is a hybrid definition that nests dwelling characteristics based sub-markets within spatially defined sub-markets. In this particular case five criteria were used. The first explanatory analysis uses the housing price per square meter (in logarithmic) and the results are shown in Figure 5 and Table 2. The second criterion is based on physical and location housing characteristics. These attributes resulted from the application of a principal components analysis to 43 variables described in Table 1 (19 physical attributes and 24 location and neighbourhood attributes), reducing to five the number housing characteristics; the results of this approach are shown in Figure 6 and Table 3. The third and fourth approaches consider the implicit prices of housing resulting from the application of a hedonic model to explain the value of a property. In one case, the value of each zone, based on the spatial structure of the housing market, is considered; and in another case, the implicit prices of physical and location characteristics of the property is considered. The results of these two perspectives of market segmentation are shown in Figures 7 and 8 and in Tables 4 and 5. Finally, the fifth and final approach aggregates and combines some of the perspectives described above, i.e., the segmentation of the housing market is the result of three main dimensions: the housing characteristics defined by the 5 factors; its importance in the assessment of the property (hedonic coefficients of the regression model) and the price per square meter of a dwelling. The results of this synthesis are shown in Figure 9 and Table 6.

a) Ln price (€/m²)

Figure 5 – Housing sub-markets defined *ex post*: ln price/m²Table 2 – Descriptive statistics of the housing price [LnPrice (€/m²)] by sub-market

Ward Method	Mean	N	Std. Deviation
1	7,386	9	,072
2	7,154	11	,033
3	7,007	16	,048
4	6,797	31	,045
5	6,572	9	,088
Total	6,936	76	,243

Based on the house price, measured in euros per square meter, 5 housing sub-markets are defined. As expected the higher price of housing is located in the sub-market that includes the more urban area, the CBD of Aveiro, in combination with the beaches (Costa Nova and Barra), which have an average price of 1613 euros per m². With the exception of the zone of Ílhavo, housing prices are progressively lower as one moves away to the periphery. The cluster 5 aggregates zones where the prices per square meter are lower (715 euros per m²).

b) Physical and location characteristics

Applying a principal components analysis to the initial variables shown in Table 1, five new dimensions were defined, classified as follows: i) factor 1 - accessibility to the main central area of Aveiro (CBD), ii) factor 2 - accessibility to local services such as health centres, parks and gardens; iii) factor 3 - areas of beaches and accessibility to amenities that characterize the beaches; iv) factor 4 - the size of dwellings, and finally, v) factor 5 - other housing characteristics, such as garage and central heating etc.. See Marques (2012) and Bhattacharjee, Castro et al., (2012) for a detailed description of the principal components analysis used in this explanatory analysis.

It is based on the result of principal component analysis that the heterogeneity of housing is then analysed.

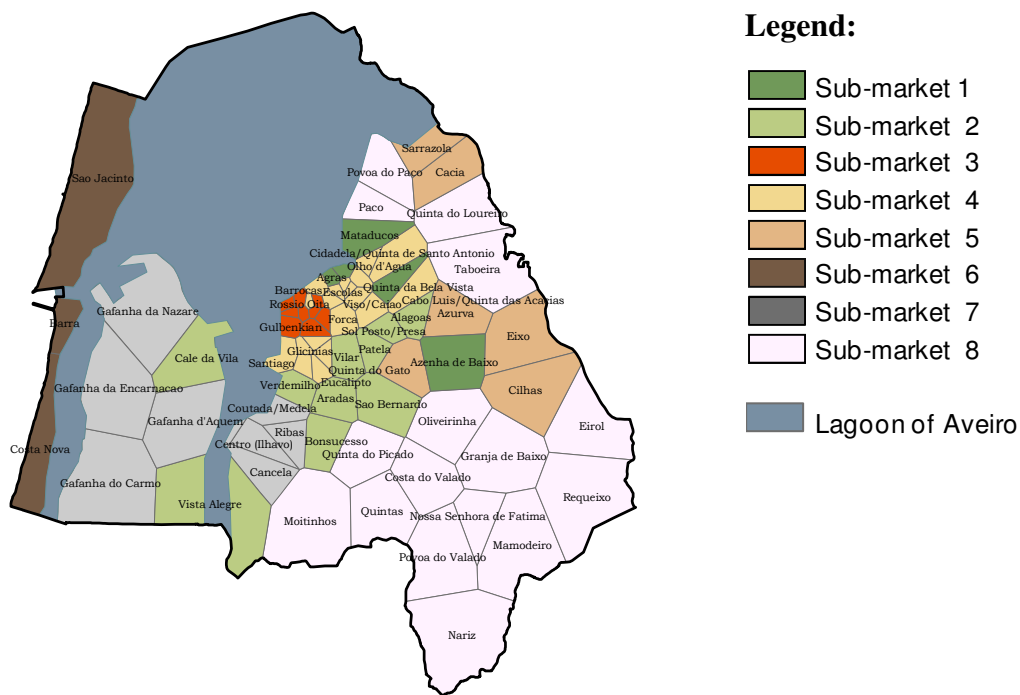


Figure 6 – Housing sub-markets defined *ex post ex post*: factor scores

Table 3 - Descriptive statistics of the housing attributes [F_i] by sub-markets

Ward Method		Factor 1: accessibility to main centralities	Factor 2: accessibility to local services	Factor 3: accessibility to amenities related to beaches	Factor 4: housing dimension	Factor 5: other housing characteristics
1	Mean	-.835	.745	-.122	.951	.560
	N	5	5	5	5	5
	Std.	.450	.286	.466	.686	.344
	Deviation					
2	Mean	.074	.154	-.285	.124	.154
	N	10	10	10	10	10
	Std.	.430	.363	.427	.283	.239
	Deviation					
3	Mean	-2.104	-.573	.564	.298	-.764
	N	9	9	9	9	9
	Std.	.368	.398	.446	.175	.590
	Deviation					
4	Mean	-.946	.025	-.072	-.339	-.356
	N	19	19	19	19	19
	Std.	.357	.560	.415	.229	.428
	Deviation					
5	Mean	.625	1.237	-1.015	-.397	-.034
	N	6	6	6	6	6
	Std.	.369	.514	.310	.312	.193
	Deviation					
6	Mean	1.352	1.182	2.500	-.177	-.284
	N	3	3	3	3	3
	Std.	.923	.133	.839	.882	.570
	Deviation					
7	Mean	.944	-.576	-.034	.381	-.098
	N	8	8	8	8	8
	Std.	.461	.655	.207	.624	.239
	Deviation					
8	Mean	.948	1.443	-.653	.818	-.020
	N	16	16	16	16	16
	Std.	.482	.369	.339	.316	.354
	Deviation					
Total	Mean	-.129	.395	-.119	.203	-.151
	N	76	76	76	76	76
	Std.	1.164	.881	.780	.605	.493
	Deviation					

Analysing Table 3 the dichotomy between two groups of sub-markets is evident: one group (C1, C3, and C4) is located close to the city centre where central amenities are abundant; and the other group is placed more distant from the city centre (C5, C6, C7, and C8). The spatial contiguity is verified in all clusters but not exclusively. Looking in more detail at each sub-market the main characteristics are emphasised.

□ C1: This set of zones belongs to a sub-market characterised by a high level of centrality to the city centre, but are also quite far from local amenities. Regarding the physical attributes, this sub-market has typically big properties with garages, balconies and central heating (housing features).

- C2: This second group of zones is well served in terms of local services and score factors for physical characteristics of properties are positive, meaning that houses located in this sub-market are bigger and have more housing facilities than the average.
- C3: This cluster (sub-market) corresponds to the CBD of Aveiro and is a distinguishable cluster because it has good accessibilities (both local and central) but bad housing facilities (almost all scores of the factor 5 are negative). The high positive values in the scores of factor 4 do not mean that properties are big, but that there are there few small houses. Note that variable housing type is included in this factor.
- C4: This sub-market encompasses zones which typically are considered to be the urban city centre of Aveiro. This sub-market has good accessibilities to beaches, schools and local commerce, has small housing, typically flats with less garages, balconies, terraces and other housing features than the average.
- C5: The fifth housing sub-market belongs to the group which has bad access to the central and local facilities. Additionally, the access to beaches, schools and local commerce is not good, when compared with the average of housing sub-markets for Aveiro.
- C6: This sub-market is very similar with the previous sub-market 5. The differences lie in the easy access to beaches, schools and local commerce. In this cluster, dwellings are typically flats (or small houses) with good additional facilities.
- C7: Distant to central amenities but high level of accessibility to local amenities, beaches, schools and local commerce are aspects that describe this housing sub-market. Regarding the intrinsic housing attributes this sub-market has traditionally larger single houses.
- C8: This last cluster has positive scores in almost all factors, except factor 3. It means that this housing sub-market is distant from central amenities and houses are more spacious than the average. The negative sign for factor 3 is not because is close to the beaches but reflects the access to schools and local commerce.

c) Values of zones

The approach presented below is based on the values of each zone, assuming dummy variables for each zone in the hedonic model, specified as follows:

$$P_{\epsilon/m^2} = N^{\alpha_1} \cdot A^{\alpha_2} \cdot d_T^{\alpha_3} \cdot d_D^{\alpha_4} \cdot \sum_{i=1}^7 d_{C_i}^{\alpha_{C_i}} \cdot \sum_{k=1}^6 d_{F_k}^{\alpha_{F_k}} \cdot \sum_{Z=1}^{76} d_{Z_j}^{\alpha_{Z_j}} \cdot \sum_{m=1}^{114} d_{T_m}^{\alpha_{T_m}} \cdot T_{TOM}^{\alpha_5} \quad \text{Eq. 3}$$

Or

$$\ln P_{\epsilon/m^2} = \alpha_1 \ln N + \alpha_2 \ln A + \alpha_3 \ln d_T + \alpha_4 \ln d_D + \sum_{i=1}^7 \alpha_{C_i} \ln d_{C_i} +$$

$$+ \sum_{k=1}^6 \alpha_{F_k} \ln d_{F_k} + \sum_{Z=1}^{76} \alpha_{Z_j} \ln d_{Z_j} + \sum_{m=1}^{114} \alpha_{T_m} \ln d_{T_m} + \ln T_{TOM}^{\alpha_5}$$

Where: P is the price (to get a more scale neutral dependent variable the normalised measure price of per square meter has been used); N - Number of bedrooms; A- Area (square meters); d_T- Housing type dummy variable (1=house; 0=flat); d_D- Dummy variable for duplex flats (yes =1; no=1); d_{C_i}- Level of preservation dummy variable (C_i=1,..., 7) [d_{C1}=New; d_{C2}=Under construction; d_{C3}=Restored; d_{C4}=Used10 (<10 years); d_{C5}=Used1025 (10-25 years); d_{C6}= Used25 (> 25 years); and d_{C7}=Not restored]; d_{F_k}-

Other physical attributes ($F_k=1,...,6$) [d_{F1} =Balcony; d_{F2} =Terrace; d_{F3} =Garage space; d_{F4} =Garage; d_{F5} =Central heating; and d_{F6} =Fireplace]; d_{Zj} - Dummy variables identifying zones ($Z_j=1,...,76$); D_{du} - Distances to several urban amenities ($d_u=1$ and 23); d_{Tm} - Dummy variable for time (monthly dummies) ($T_m=1, \dots, 114$); T_{TOM} - is the time on the market (TOM) measure in days; and Alfa (α) are parameters which measure the relevant implicit marginal prices of the attributes.

The regression coefficients for the location attributes, considering the variables of each dummy zone (76 in total) are shown in Figure 7 and Table 4.

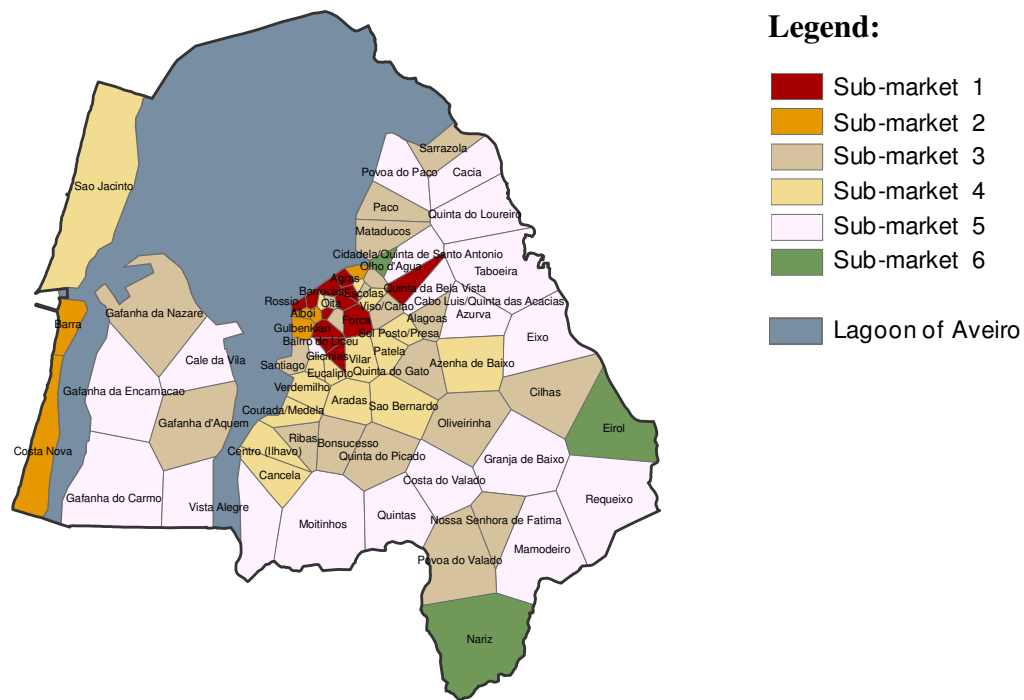


Figure 7 – Housing sub-markets defined *ex post ex post*: value of the zone

Table 4 - Descriptive statistics of the value of the zones [d_{Zj}] by sub-markets

Ward				Std.
Method	Mean	N		Deviation
1	-,404	12		,036
2	-,189	6		,097
3	-,661	20		,030
4	-,532	16		,037
5	-,763	19		,034
6	-,939	3		,037
Total	-,593	76		,186

The results of housing markets segmentation using the coefficients of the value of each zone it are very similar to that using housing prices (€ / m²) (see Figure 5). Considering the zone Nariz as the reference point (where housing prices per square meter are lowest) the values of the remaining zones, when compared to Nariz, are reported. In general, considering constant all physical attributes, the price of housing in the urban areas have

higher land value (as expected), as well as beaches (Costa Nova e Barra) that also appear to be highly valued. For example, the price of a house placed in the Forum zone is 63% more expensive than a house located in Nariz zone, and Barra and Costa Nova is 45%.

d) Values of housing attributes

The results of housing market segmentation using hedonic coefficients of the location of the property (d_{Zj}) and the housing physical characteristics (F_k) are presented. The hedonic model used in this analysis is reported in equation 4 and considers 5 factors (which resulted from principal component analysis; instead of the initial variables) and uses dummy variables (intercept and slope) to define the location value of a house (d_{Zj}) and the value of housing intrinsic attributes [expressed by the factors 4 and 5 (size of dwellings and other housing characteristics, respectively)].

$$P_{\epsilon/m^2} = A^{\alpha_2} \cdot \sum_{L=1}^3 F_L \cdot \sum_{k=1}^2 \sum_{Z=1}^{76} d_{Zj} F_k^{\alpha_{Zj}} \cdot \sum_{Z=1}^{76} d_{Zj}^{\alpha_{Zj}} \cdot \sum_{m=1}^{114} d_{T_m}^{\alpha_{T_m}} \cdot T_{TOM}^{\alpha_5} \quad \text{Eq. 4}$$

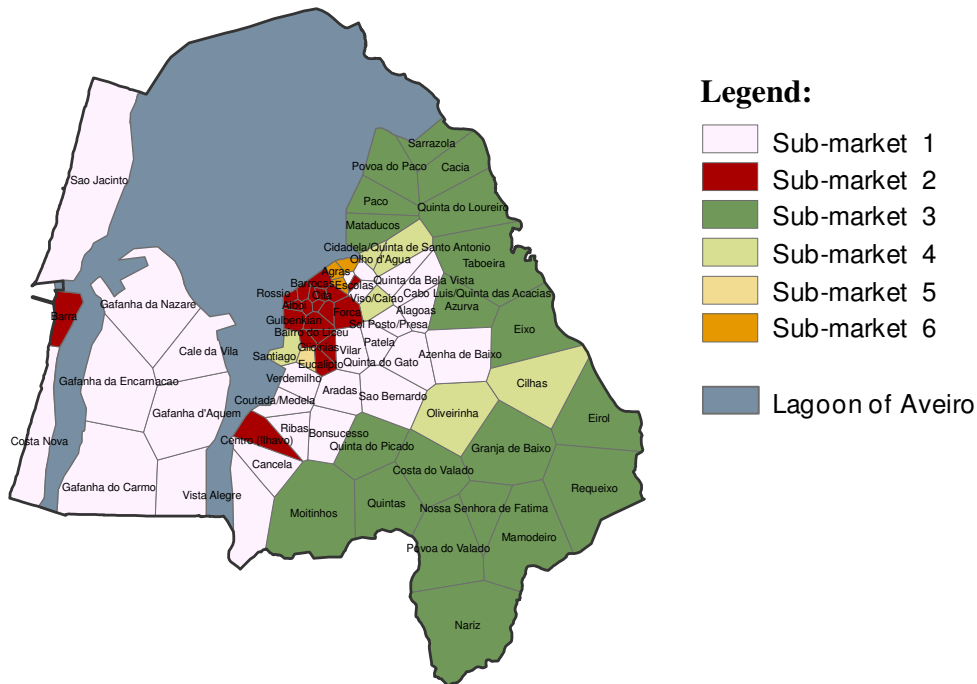


Figure 8 – Housing sub-markets defined *ex post ex post*: implicit prices

Table 5 - Descriptive statistics of the value of housing implicit prices [Zscore(F4), Zscore (F5), Zscore (d)] by sub-markets

Ward	Method	Zscore(F4)	Zscore(F5)	Zscore(d)
1	Mean	-.137	-.325	.094
	N	26	26	26
	Std. Deviation	.321	.731	.326
2	Mean	.431	.325	1.204
	N	20	20	20
	Std. Deviation	.616	.561	.377
3	Mean	-.425	-.447	-1.236
	N	20	20	20
	Std. Deviation	.587	.479	.319
4	Mean	.071	2.402	-.591
	N	6	6	6
	Std. Deviation	1.321	.823	.986
5	Mean	-4.716	.131	.853
	N	1	1	1
	Std. Deviation	.	.	.
6	Mean	2.572	-1.219	.291
	N	3	3	3
	Std. Deviation	.000	.000	.000
Total	Mean	.000	.000	.000
	N	76	76	76
	Std. Deviation	1.000	1.000	1.000

An analysis of the results summarized in Table 6 allows the characterization each of the sub-market as follows: The sub-market 1 includes all dwellings where some physical attributes, not related to the housing dimension (garage, terrace etc.), are not highly valued; mainly located in the outskirts of the sub-urban area, e.g., in the beaches (Costa Nova and São Jacinto) and in Gafanhas. The sub-market 2 is essentially characterized by being highly valued in terms of location. The other two housing attributes tend to have positive values, i.e., housing dimension and other housing characteristics (such as garages, terraces and balconies) are highly valued. The sub-market 3 is the most rural zone of the study area. It is characterized by dwellings which are neither valued in terms of intrinsic nor extrinsic attributes. The sub-market 4 is where dwellings (preferably houses), equipped with garages, terraces and garages are highly valued. The sub-market 5 includes only one zone: Vila Jovem / Santiago; being a social neighbourhood characterized with basic housing elements, the other housing features are highly valued. Finally, the sub-market 6 is characterized by the values of housing physical characteristics, i.e. housing dimension is highly appreciated and other housing features are highly devalued.

e) Synthesis

This final explanatory analysis combines the previous approaches and gathers in the same cluster analysis three key elements: housing characteristics (physical and location), the valuation of these characteristics, and finally, the value of the property represented by its price, measured in euros per m². The results of this analysis are presented below.

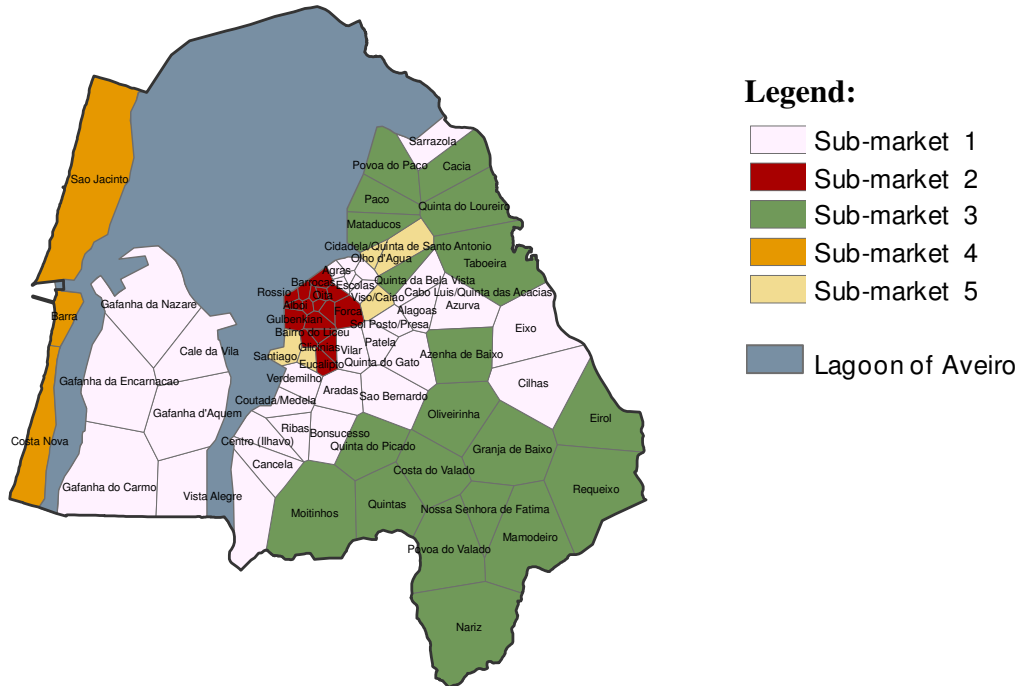
Figure 9 – Housing sub-markets defined *ex post ex post*: synthesis

Table 6 - Descriptive statistics of the housing characteristics [F4, F5, d], value of housing implicit prices [Zscore(F4), Zscore (F5), Zscore (d)] and housing prices [P] by sub-markets

		[value of each housing attribute]			(Price)	[housing characteristics]				
Ward	Method	Zf4	Zf5	Zd		F1	F2	F3	F4	F5
1	Mean	.145	-.240	-.023	.061	.207	-.207	-.327	-.350	.395
	N	31	31	31	31	31	31	31	31	31
	Std. Deviation	.883	.732	.637	.637	.711	.838	.570	.830	.551
2	Mean	.445	.377	1.265	.979	-1.318	-1.031	.650	-.234	-.812
	N	17	17	17	17	17	17	17	17	17
	Std. Deviation	.655	.593	.302	.819	.504	.425	.564	.521	1.157
3	Mean	-.330	-.394	-1.147	-1.007	.701	1.072	-.599	1.097	.430
	N	20	20	20	20	20	20	20	20	20
	Std. Deviation	.672	.647	.503	.524	.655	.468	.555	.702	.827
4	Mean	.004	-.877	.475	1.203	1.272	.893	3.357	-.628	-.270
	N	3	3	3	3	3	3	3	3	3
	Std. Deviation	.213	1.061	.665	1.303	.793	.150	1.075	1.457	1.155
5	Mean	-1.092	2.307	.147	-.402	-.367	-.032	.201	-1.042	-1.249
	N	5	5	5	5	5	5	5	5	5
	Std. Deviation	2.445	1.310	.586	.633	.181	.725	.284	.491	.515
Total	Mean	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	76	76	76	76	76	76	76	76	76
	Std. Deviation	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

In this synthesis, combining multiple criteria and perspectives for delimiting housing markets, five distinct segments in urban and sub-urban area of Aveiro and Ílhavo are considered. A brief analysis of the data presented in Table 7. The sub-market 1 includes zones located in adjacent areas to the urban centre of Aveiro, with mean values of the criteria used in the analysis. The sub-market 2 is the traditional urban centre of Aveiro and is characterized by zones with high housing prices resulted from their proximity to equipment and services (with predominantly urban features: dZ, F1 and F2). In the sub-market 3 are located zones predominantly semi-rural, characterized by having low housing prices, mainly because of its inaccessibility to urban services and equipment. Households that buy houses in this sub-market do not consider important the location (Zd). The sub-market 4 is defined by beaches zones (Costa Nova, Barra São Jacinto), which is the most expensive market segment of the entire study area. Finally, the sub-market 5 organized by 5 zones (Santiago, Vila Jovem, Viso/caiã, Olho d'agua e Cidadela/Sto António) is constituted by small houses (compared to the average) with no other additional housing features; the first aspect is largely depreciated, while the second does not. Aspects such as the existence of a garage, terrace, etc. are highly valued.

4. CONCLUSION

The main objective of this paper was to contribute to the discussion of the definition of housing sub-markets, both in terms of its conceptualization, and in terms of how to define it empirically. The importance of this issue is widely recognized in the literature. There are several reasons, to understand how urban areas are segmented: i) from a purely statistical perspective it increases the prediction accuracy of the estimated hedonic model; ii) analytically it enables researchers to better understand the spatial variation in housing prices; iii) an accurate assignment of properties to sub-markets improves lenders' and investors' abilities to price the risk associated with financing homeownership; and finally, iv) providing sub-market boundary information to housing consumers reduces their search costs. Moreover, in many cases adequate treatment of spatial heterogeneity could considerably reduce the presence of spatial dependence effects, even though the two problems are theoretically distinct.

Spatial heterogeneity in the context of housing sub-market is a complex subject that encompasses different techniques, forms and philosophies of understanding spatial segmentation. This work attempted to present an explanatory analysis where some of these multiple perspectives were analysed, being clear that the analysis of housing segmentation should be defined in a unified manner.

Acknowledgments

The authors acknowledge the support provided by the Research Unit for Governance, Competitiveness and Public Policies Research Unit of the University of Aveiro (GOVCOPP) and by Foundation for Science and Technology (FCT) for financing the research project DONUT (PTDC /AURURB / 100592/2008), as well to the Operational Programme 'Thematic Factors of Competitiveness' (COMPETE) of the EU Community Support Framework (QCA III – European Commission) and European Community Fund (FEDER). The authors are also grateful to Janela Digital, for providing housing data.

REFERENCES

- Adair, A., Berry, J. and McGreal, W. (1996). *Hedonic Modelling, Housing Submarkets and Residential Valuation*. Journal of Property Research **13**(1): 67-83.
- Bhattacharjee, A., Castro, E. and Marques, J. (2012). *Spatial Interactions in Hedonic Pricing Models: The Urban Housing Market of Aveiro, Portugal* Spatial Economic Analysis **7**(1): 133-167.
- Bhattacharjee, A. and Holly, S. (2010a). *Structural Interactions in Spatial Panels*. Centre for Dynamic Macroeconomic Analysis.
- Bhattacharjee, A. and Holly, S. (2010b). *Understanding Interactions in Social Networks and Committees*. Faculty of Economics, University of Cambridge.
- 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*. Centre for Research into Industry, Enterprise, Finance and the Firm.
- Bourassa, S., Cantoni, E. and Hoesli, M. (2007). *Spatial Dependence, Housing Submarkets, and House Price Prediction*. The Journal of Real Estate Finance and Economics **35**(2): 143-160.
- Bourassa, S., Cantoni, E. and Hoesli, M. (2008). *Predicting House Prices with Spatial Dependence: Impacts of Alternative Submarket Definitions*. Swiss Finance Institute.
- Bourassa, S., Hamelink, F. and Hoesli, M. (1999a). *Defining Housing Markets*. Journal of Housing Economics **8**: 160–183.

- Bourassa, S., Hamelink, F., Hoesli, M. and MacGregor, B. (1999b). *Defining Housing Submarkets*. Journal of Housing Economics **8**(2): 160-183.
- Brasington, D. and Hite, D. (2005). *Demand for Environmental Quality: A Spatial Hedonic Analysis*. Regional Science and Urban Economics **35**(1): 57-82.
- Castro, E., Marques, J. and Batista, P. (2011). *Estimação De Uma Superfície Hedónica De Preços Para O Mercado Habitacional Em Portugal*. 17º Conferência da APDR, Bragança/Zamora.
- Clapp, J. and Wang, Y. (2006). *Defining Neighborhood Boundaries: Are Census Tracts Obsolete?* Journal of Urban Economics **59**(2): 259-284.
- Fingleton, B. (2008). *Housing Supply, Housing Demand, and Affordability*. Urban Study **45**(8): 1545-1563.
- Galster, G. (1987). *Residential Segregation and Interracial Economic Disparities: A Simultaneous-Equations Approach*. Journal of Urban Economics **21**(1): 22-44.
- Goetzmann, W. and Spiegel, M. (1997). *A Spatial Model of Housing Returns and Neighborhood Substitutability*. Yale School of Management.
- Goodman, A. (1981). *Housing Submarkets within Urban Areas: Definitions and Evidence*. Journal of Regional Science **21**: 175-185.
- Goodman, A. and Thibodeau, T. (1998). *Housing Market Segmentation*. Journal of Housing Economics **7**(2): 121-143.
- Goodman, A. C. and Thibodeau, T. G. (2007). *The Spatial Proximity of Metropolitan Area Housing Submarkets*. Real Estate Economics **35**(2): 209-232.
- Goodman, C. (1998). *Andrew Court and the Invention of Hedonic Price Analysis*. Journal of Urban Economics **44**(2): 291-298.
- Goodman, C. and Thibodeau, G. (2003). *Housing Market Segmentation and Hedonic Prediction Accuracy*. Journal of Housing Economics **12**(3): 181-201.
- Grigsby, W. (1963). *Housing Markets and Public Policy*, Philadelphia: University of Pennsylvania Press.
- Hårsman, B. and Quigley, J. (1995). *The Spatial Segregation of Ethnic and Demographic Groups: Comparative Evidence from Stockholm and San Francisco*. Journal of Urban Economics **37**(1): 1-16.
- Kiel, K. and Zabel, J. (1996). *House Price Differentials in U.S. Cities: Household and Neighborhood Racial Effects*. Journal of Housing Economics **5**(2): 143-165.
- Linneman, P. (1981). *The Demand for Residence Site Characteristics*. Journal of Urban Economics **9**(2): 129-148.
- MacLennan, D. (1977). *Some Thoughts on the Nature and Purpose of House Price Studies*. Urban Studies **14**(1): 59-71.
- MacLennan, D., Munro, M. and Wood, G. (1989). *Housing Choice and the Structure of Urban Housing Markets*. In *Between State and Market: Housing in the Post-Industrial Era*. B. Turner, J. Kemeny and L. Lundqvist, CÖthenburg, Almquist and Hicksell.
- MacLennan, D. and Tu, Y. (1996). *Economic Perspectives on the Structure of Local Housing Systems*. Housing Studies **11**(3): 387-406.
- Malpezzi, S., Ozanne, L. and Thibodeau, T. (1980). *Characteristic Prices of Housing in 59 Smsas*. The Urban Institute, USA.
- Marques, J. (2012). *The Notion of Space in Urban Housing Markets*. In Department of Social Sciences, Policies and Territory, University of Aveiro. **PhD thesis in Social Sciences**.
- Michaels, R. and Smith, V. (1990). *Market Segmentation and Valuing Amenities with Hedonic Models: The Case of Hazardous Waste Sites*. Journal of Urban Economics **28**(2): 223-242.
- Nelson, J. (2008). *Hedonic Property Value Studies of Transportation Noise: Aircraft and Road Traffic* In Hedonic Methods in Housing Markets: Pricing Environmental Amenities and Segregation. Springer: 300.
- Palm, R. (1978). *Spatial Segmentation of the Urban Housing Market*. Journal of Economic Geography **54**: 210-221.

- Quigley, J. (1979). *What Have We Learned About Housing Markets?* In Current Issues in Urban Economics. P. Mieszkowski, M. Straszheim and J. Hopkins, Press: 388-426.
- Rapkin, C., Winnick, L. and Blank, D. (1953). *Housing Market Analysis. A Study of Theory and Methods*, Housing and Home Finance Agency. Division of Housing Research.
- Raudenbush, S. and Bryk, A. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods (Advanced Quantitative Techniques in the Social Sciences)*, Sage Publications Inc.
- Schnare, A. B. (1980). *Trends in Residential Segregation by Race: 1960-1970*. Journal of Urban Economics **7**(3): 293-301.
- Schnare, A. B. and Struyk, R. J. (1976). *Segmentation in Urban Housing Markets*. Journal of Urban Economics **3**(2): 146-166.
- Straszheim, M. (1974). *Hedonic Estimation of Housing Market Prices: A Further Comment*. The Review of Economics and Statistics **56**(3): 404-406.
- Straszheim, M. (1975). *An Econometric Analysis of the Urban Housing Market* N. B. o. E. Research, Cambridge MA.
- Struyk, R., Malpezzi, S. and Wann, F. (1980). *The Performance Funding System Inflation Factor: Analysis of Predictive Ability of Candidate Series and Models*. Urban Institute Working Paper.
- Ugarte, M., Goicoa, T. and Miltino, A. (2004). *Searching for Housing Submarkets Using Mixtures of Linear Models*. In Spatial and Spatiotemporal Econometrics. Advances in Econometrics. J. LeSage and R. Pace, Oxford: Elsevier.
- Watkins, C. (2001). *The Definition and Identification of Housing Submarket*. Environment and Planning A **33**(12): 2235-2253.