Revista Portuguesa de Estudos Regionais, nº 63 2023, 7-25

Artigo submetido a 2 de Junho 2021; versão final aceite a 27 de Novembro de 2021
Paper submitted on June 2, 2021; final version accepted on November 27, 2021

House Price Indexes: An Empirical Exercise Using Spatial Econometrics

Indices de Preços Imobiliários: Um Exercício Empírico com Recurso à Econometria Espacial

Miguel Viegas
mlbv@ua.pt
Research Unit of Governance, Competitiveness and Public Policies, DEGEIT, Universidade de Aveiro

João Marques
jjmarques@ua.pt
Research Unit of Governance, Competitiveness and Public Policies, DCSPT, Universidade de Aveiro

Paulo Batista
pauloricardolb@ua.pt
Research Unit of Governance, Competitiveness and Public Policies, DCSPT, Universidade de Aveiro

Abstract

The housing market occupies a central place within any society. Residential property represents the essential parts of most families' wealth, as well as a substantial part of their monthly expenditure. It also represents a large fraction of private investment, generating significant earnings. However, and despite this importance, the compilation of reliable residential property price indexes (RPPI) is far from satisfactory in Portugal, revealing an important gap in the field of statistical information.

The construction of a residential property price index (RPPI) poses problems arising from the inherent nature of the object concerned. Considering the gap in Portugal regarding the lack of a sufficiently enlarged database allowing the compilation of reliable RPPIs, this article seeks to contribute to this area, proposing a methodology applied to the urban area of Aveiro and Ílhavo by means of a spatial data analysis.

Keywords: Residential property Indexes, dwelling, spatial dependence.

Códigos JEL: R31, R32, R15

Resumo

O mercado imobiliário ocupa um lugar central em qualquer sociedade. A propriedade imobiliária representa a parte essencial da riqueza da maioria das famílias, e origina uma parte substancial de suas despesas mensais. Também representa uma grande fração do investimento privado, gerando ganhos significativos. No entanto, e apesar desta importância, a compilação de índices fiáveis de preços de imóveis residenciais está longe de ser satisfatória em Portugal, revelando uma lacuna importante no domínio da informação estatística.

A construção de um Índice de Preços Residenciais apresenta problemas particulares decorrentes da natureza inerente do objeto em questão. Atendendo à lacuna existente em Portugal de uma base de dados suficientemente alargada que permita a compilação de Índice de Preços Residenciais fiáveis, este artigo procura contribuir nesta área, propondo uma metodologia aplicada à área urbana de Aveiro e Ílhavo através de uma análise de dados espaciais.

Palavras-chave: índice de Preços Residenciais, habitação, dependência espacial.

JEL classification: R31, R32, R15

1. INTRODUCTION

Residential property represents the essential part of most families' wealth, as well as a substantial part of their monthly expenditure. It also represents a large fraction of private investment, generating significant earnings. The construction sector in Portugal currently employs about 10% of the work-force. It is also responsible for about half of the Gross Fixed Capital Formation (INE, 2011). Apart from its importance as an individualized sector, housing construction generates significant multiplier effects in several sectors. Moreover, house price risk has attracted much attention after the U.S. subprime crisis and the ensuing global financial crisis. It was the significant increases in house prices that enhanced the perceived lower risk and consequent lax lending criteria in mortgage markets. Housing and real estate have a decisive influence on the behavior of the global economy. Therefore, it is essential that agents have reliable information about house prices and their evolution, since these determine the value of important assets, the composition of the family budget as well as the profit-ability of real estate investments. Information quality and transparency are key attributes for the proper functioning of the markets. Good residential property price indexes (RPPI) have also numer-ous applications in various fields. They represent a good indicator of the macro-economic activity in a country or a region. They can also be used as an indicator for the monetary policy.[1] Good RPPIs are necessary to evaluate the wealth of families for whom, houses are, in many cases, the most valuable asset. They can, as well, make a valuable contribution in the measurement of inflation as key input for a basket of consumer prices. Finally, they give security and increases the efficiency of the housing market by introducing credible, independent, and timely information. However, and despite this importance, the compilation of reliable RPPIs is far from satisfactory in Portugal, which remains an important gap in the field of statistical information.

Given the need to improve the quality and reliability of RPPIs in Portugal, this article proposes to contribute to this area, using a Time Dummy hedonic price model inserted in a spatial econometric framework applied to the urban area of Aveiro and Ílhavo. Spatial correlation in the residuals of hedonic models has been reported in many empirical studies (Bourassa et al., 2007; Pace et al., 1998). However, from our knowledge, very few RPPIs includes proper mechanism to deal with spatial dependence to offset omitted variable problems. Our model represents a compromise between completeness and parsimony that can be replicated in other localities, covering the various sub-markets, heterogeneity and geographical features and key attributes. Although hedonic models can incorporate effects of neighbourhood or proximity, the literature recognizes the complexity of these effects and the difficulty in finding variables capturing spatial effects (Basu & Thibodeau, 1998). Even in models including neighbourhood and accessibility variables, the residual produced by these hedonic models may exhibit patterns of autocorrelation by misspecification of the model, creating inconsistency and inefficiency in OLS estimators (Anselin, 1988). Therefore, our model processes the spatial dependence by a spatial data analysis, thus seeking to infer the extent to which the estimators calculated from a model of spatial dependence contribute to increasing the quality of our price index.

After the introduction, we proceed with a literature review in section 2. Section 3 describes the database as well as the methodology used in the construction of our RPPI emphasizing the spatial dependence framework. Section 4 is dedicated to the presentation of results and discussion, compa-ring the hedonic model with and without the spatial data analysis. The main findings and conclusion are presented in Section 4, in which a set of recommendations are made, paving the way for a new strategy for regular publication of qualified RPPIs by public authorities.

[1] The Macroeconomic Imbalance Procedure (MIP) is part of the EU's so-called 'six-pack' legislation, which aims to reinforce the moni-toring and surveillance of macroeconomic policies in the EU and the euro area. The MIP is partially based on a scoreboard of fourteen indicators, which includes a House Price Index in order to detect real estate bubbles.

2. LITERATURE REVIEW

The construction of a RPPIs poses problems arising from the inherent nature of the good concerned. Each dwelling constitutes a unique and not reproducible asset with a long durability, which prevents the conceptualization of a fixed basket of goods whose prices are recorded over time. Indeed, in the case of housing, prices are rarely observed due to the sporadic nature of transactions. In fact, real estate assets are highly spread geographically and centralised market in which properties prices can easily be observed does not exist. Their durability makes the quality vary over time due to the depreciation or improvement and renovation activities. Considering these difficulties, it is not possible to construct a perfect index of housing prices. We can only aspire to build an approximation as accurate as possible to the theoretical RPPI (Baumont, 2009; Haurin & Hendershott, 1991).

2.1. The main methods of constructing RPPIs

The literature describes four main methods of constructing RPPIs: stratification or mix adjustment, repeat sales method, the use of property assessment information or appraisal-based method and the hedonic regression method (Hoffmann & Lorenz, 2006). The proper choice depends on the available database, and on the goals, we want to achieve. When the volume of data is large, allowing for sufficient homogeneous grouping of observations, the stratification or mix adjustment may be the most appropriate method. It only requires data about price and location to calculate central price tendency estimations such as average and median prices. This method is among the less data intensive methods and has also the advantage of being easily understood by market agents (Mark & Goldberg, 1984; McDonald & Smith, 2009). The repeat sales method consists in observing the evolution of the price of a specific type of housing over time. The scarcity of data naturally raised strong barriers to its implementation, leading often to biased results (Shiller, 1991; Wang & Zorn, 1997). The appraisal-based method becomes attractive in countries where the government conducts systematic reviews of all properties for tax purposes. Unfortunately, it is difficult to incorporate qualitative changes (depreciation or renovation) and always raises important questions about the reliability of the appraisal methodology.

The hedonic regression method is generally regarded as the most appropriate method for constructing price indexes for housing (Diewert et al., 2009). Assuming that all relevant variables are included, this method maximizes the use of available information and adjusts the RPPI to both sampling variation and qualitative changes of housing, be that a consequence of depreciation or renovation (Hill, 2004; Hill & Melser, 2008). The starting point of the hedonic method is simple and based on the assumption that certain goods are demanded in the market not by the good itself but rather by the intensity of the characteristics that define it. Thus, the search of a home is determined not by the house itself but by its attributes: location, rooms, equipment, garage etc. (Maleyre, 1997). The first hedonic regression is usually attributed to (Waugh, 1928) who had the idea of regressing the price of a bundle of asparagus based on three characteristics: the colour, the rode and tip size. Later works (Griliches, 1971; Lancaster, 1966; Rosen, 1974) generalized this approach to the social sciences and in particular to the economy. Concerning real estate and housing market, numerous attributes, attached to the property, have a significant influence on price determination. However, a significant proportion of house prices variability remains unexplained. Part of the unexplained variance may be related to a latent spatial component (Dubé & Legros, 2014). The presence of strong spatial autocorrelation on the hedonic price residuals confirms this hypothesis.

Recently, many applications of hedonic price models have included spatial analysis in order to control for spatial dependence and heterogeneity (Bajat et al., 2018). The presence of autocorrelation and spatial heterogeneity profoundly affects the quality of econometric models (Baumont, 2009) Housing prices are vulnerable to spatial effects for several reasons. The first one concerns the accessibility (to CBD or any other centrality or equipment). The second has to do with the effects of neighbourhood or how housing price capitalizes positive or negative externalities arising from urban dynamics. The third and final reason has to do with shared structural similarity to the scale of neighbourhood that are normally constructed as a whole and at the same time. As such, and considering the potential spatial autocorrelation, several criticisms may be pointed to classic OLS hedonic models, mostly related with the existence of multicollinearity, endogeneity, biasedness and the existence of specification errors. These problems may seriously affect the robustness of the convergence coefficient and produce misleading outcomes (Anselin, 1988).

The hedonic equations incorporate in most cases a set of spatial variables designed to capture some of the spatial effects mentioned above. However, measurement of these spatial effects and the choice of appropriate variables, presents numerous and practical difficulties. In models for areal data or discrete spatial variation, two models widely used are the simultaneously autoregressive model (SAR) and the and conditional autoregressive model (CAR). The SAR model use likelihood methods while the CAR model use hierarchical modelling with Bayesian methods (Banerjee et al., 2014). According to among others, the spatial data analysis with the introduction of the geographical dimension, namely in the presence of spatial autocorrelation, allows not only to capture the spatial effect, but also to improve the estimation and prevision since spatial dependence violates some of the Gauss-Markov assumptions of the OLS estimation (cross section observations are no longer independent) producing inefficient estimators. Two kinds of spatial effects are pointed out in the literature, namely: (i) spatial autocorrelation, revealing that contiguous regions may influence each other's performance through spillover effects and (ii) spatial heterogeneity, whenever the same functional form is erroneously considered for all regions (for comprehensive references about spatial econometric see for instance (Anselin, 1988; Le Gallo, 2002; LeSage & Pace, 2009) To capture the interdependence between locations, it is necessary to consider the relative positions of each location. For this, we must fix exogenously the topology of the space system, building a weight matrix (Le Gallo, 2002). The most widely type of matrix used is the contiguity matrix, in which the term wi,j equal to one if the two locations i and j are contiguous and zero otherwise. It is also possible to use Euclidian distances instead of a binary variable. The weights matrixes are usually normalized with each line sum equal to one.

Spatial econometric offers different possibilities to deal with spatial dependence. A first approach assumes that spatial dependence enters into the right-hand-side of the house price equation. This model, characterized by the presence of spatial lag dependent variable has been extensively used by (Elhorst & Vega, 2015; Gibbons & Overman, 2012; LeSage & Pace, 2009). According to (Elhorst & Vega, 2015; Gibbons & Overman, 2012), this model can be very useful in applied studies because of "its superiority in avoiding identification issues and its flexibility in measuring spillover effects". House price spillovers can also arise from spatially correlated omitted variables. Empirical application using the spatial error model can be found in (Pesaran & Tosetti, 2011; Yang, 2020). Discussions on the best methodologies for choosing the appropriate spatial model can be consulted in (Elhorst, 2014b; Seya et al., 2020).

We found in the literature several methodologies used to address spatial autocorrelation and heterogeneity. (Francke & Van de Minne, 2020) use spatial random effects in hedonic price models to improve prediction accuracy. (Paredes, 2011) applies the quasi-experimental method to Chilean regions. In another article, (Iturra & Paredes, 2013) deals with endogeneity caused by omitted variables with a set of instruments using a generalized method of moments (GMM) procedure. (Chasco & Le Gallo, 2012) used multilevel models to fully capture spatial autocorrelation effects in the errors. They explored a three-level hierarchical model, more capable of explaining data heterogeneity, conclude, however, that more effort should be done to develop alternative tools to deal with spatial autocorrelation. Most authors converge on the need to address spatial dependence but few are the empirical researches. (Wilhelmsson, 2009) proposes a methodology for the construction of a real estate index for Stockholm with a general spatial model incorporates a spatial structure into both the dependent variable and the error term. Spatial dependency and property price index construction are examined by (Se Can & Megbolugbe, 1997). They conclude that the inclusion of a spatial structure is important for both parameter estimates and statistical inference. Trying to overcome the lack of good residential property price statistics (Brunauer et al., 2012) describe the setup of a new RPPI for Austria. However, instead of spatial data analysis, they use a semiparametric model that consider nonlinearity and spatial heterogeneity. More recently (Oust et al., 2019) argue that spatial analysis has been a long-neglected part in the studies of economics. They develop a model applied to residential property transactions in Oslo, Norway between August 2016, and December 2017, combining repeat sales and hedonic regression enhanced with spatial econometric tools. They conclude that this combination resulted in improved accuracy for all hedonic regressions on all metrics.

2.2. The price indices in Portugal

Unfortunately, there is no systematic collection of transaction price by the Portuguese national statistical system. The INE (National Statistics Institute) publishes the monthly "Survey on bank evaluation on housing" trying to build leading indicators of housing prices. The values collected in these surveys are not formal prices. They represent the intention of purchase, thus constituting a reasonable approximation. The information is provided by a set of credit institutions operating in Portugal, considered as the most representative of the market for granting housing loans. The information is transmitted monthly and is used to calculate basic indicators of the average property assessment for a particular typology, $\bar{Y}_{ti}$ in period t, and in the territorial unit i, using the following formula:

$$\bar{Y}_{ti} = \sqrt[n]{\prod_{j=1}^{n} Y_{tij} / A_{tij}}, \qquad (1)$$

Being Y_{tij} the value of the assessment of the property j with the given typology j in period t in the territorial unit i; A_{tij} the corresponding net area of the property j in period t, in the territorial unit i, and finally, n, the number of housing assessments.

The Confidencial Imobiliário (Ci) Price Index has been in place in Portugal since 1988. In 2006 it underwent a profound remodeling. It is based on transaction prices and includes a quality adjustment mechanism that responds to the heterogeneity of the data. In a first step, it estimates a hedonic model based on the prices advertised for each property. In a second step, it calculates the adjustment factor that corresponds to the value of the implicitly observed qualitative attributes. This value is calculated based on the difference between the real advertised price of the property and the estimated price in the model estimated in the first stage. In the third step, a new hedonic model with the same attributes is estimated, using the effective transaction prices as a dependent variable, and adding the adjustment factor calculated in the second step as another explanatory variable. The calculation of changes in the value of transactions is made using the Fisher index, with a regional weighting in the case of the national version of the index.

3. METHODOLOGY AND DATA

The construction of price indices for housing (RPPIs) raises several challenges due to the particularities of the real estate market. The first concerns the data source. The second is about methodological issues and seeks to answer important question as the need to stratify geographic units or to identify sub-markets of similar dwelling, or still, the choice of the most appropriate econometric techniques, the variables to select or the most appropriate functional form. A regular publication of RPPIS implies the existence of a solid and reliable organization capable of collecting the relevant information and building the indexes properly. There are several types of databases. Their distinctions depend on the time of collection which can be done at various stages of the process of buying and selling (advertised prices, banking appraisal or transaction prices).

3.1. Description of the data

The database used in this article covers a geographical area comprising two counties, Aveiro and Ílhavo, with approximately 110 000 inhabitants in an area of about 275 km2. Our database was built from the National Real Estate Portal - Casa Sapo. We collected and further processed 14087 observations covering a period between 2003 and 2010. The database contains pricing announced (which may or may not correspond to a sale). The database covers houses (villas) as well as apartments, new and second-hand and in different locations. The two counties stand almost as an urban continuum. However, in this urban continuum, semi-urban territories coexist with agricultural activity as well as an area of ecological interest belonging to the Natura 2000 network.

The database contains a wide variety of attributes. Beside the price, we have access to the areas and number of rooms. Other attributes such as the presence of a garage, elevator or balconies complete the database giving us a wide range of possibilities to build our hedonic model. As can be seen in Table 1our sample contains more apartments than houses and is more concentrated in the years 2007 to 2010.

Table 2 and Table 3 contain information about some of the characteristics of the dwelling of our database: price, area, number of rooms etc. Many observations of the dwelling have been discarded due to the lack of adequately documented. The database contains 10428 observations for apartments and 3659 for houses.

Table 1: database description

Year	2003	2004	2005	2006	2007	2008	2009	2010
Houses	45	1	41	168	875	798	799	932
Apartments	164	17	189	571	2606	1899	2583	2399

Table 2: Descriptive Statistics (apartments).

	N	Minimum	Maximum	Average	S.Dev.
Price	10428	29999,99	480000,00	133010,76	49904,40
Price per m^2	10428	377,95	5272,73	1236,10	377,22
Area	10428	30,00	325,00	111,94	38,82
Rooms	10428	0	5	2,14	0,830

Table 3: Descriptive Statistics (houses).

	N	Minimum	Maximum	Average	S.Dev.
Price	3659	28000,00	900000,00	223100,42	81832,18
Price per m^2 (built)	3659	178,57	3167,94	954,28	307,99
Building area	3659	47,00	690,75	254,39	88,11
Land area	3659	0,00	50000,00	179,42	1475,19
Rooms	3659	1	8	3,88	0,644

3.2. Hedonic Approach

We use the hedonic method, which assumes previously that the price of a house depends on a combination of a limited set of qualitative and quantitative attributes. The relationship between the price per m^2 of housing and their attributes is estimated from a regression based on a set of observations of actual prices, and then used to reconstruct the price of a dwelling in the same sample.

For the hedonic price index, we constructed a cross-sectional hedonic equation with time dummy variable of house prices in the Counties of Aveiro and Ílhavo. The time dummy variable approach to constructing a hedonic house price index has been used frequently in academic studies but, to our knowledge not so much by statistical agencies.[2] One advantage of this approach is its simplicity.[3] The time dummy method is recommended when the database has a large number of characteristics but few transactions for each period. Prices and attributes of all dwellings for several periods are pooled in the same regression and a dummy variable is created for each period. The objective of the exercise is to produce a quality-adjusted price index. As such, the principal interest lies in the dummy parameters which measure the period specific fixed effects on the logarithms of the price level after controlling for the attribute's effects of the dwellings (Hill, 2013). The standard semi-log formulation of our time-dummy equation is as follows:

[2] A few examples can be consulted in (Kestens et al., 2006; Wilhelmsson, 2008)

[3] A concern with the time-dummy method is that when a new period is added to a data set, the price indexes for all periods change. To overcome that problem it is possible to estimate the time-dummy model only over adjacent periods (Triplett, 2004).

$$\ln(p_n^t) = \beta_0 + \sum_{\tau=1}^{T} \delta^\tau D_n^\tau + \sum_{k=1}^{K} \beta_k x_{nk}^t + u_n \qquad (2)$$

D_n^τ represent the time dummies variables and x_{nk} the k attributes of the model for the n observations. The k attribute shadow prices are represented by the parameter β. The price index P_t for period t is derived by exponentiating the estimated parameter δ_t from the hedonic equation (Hill, 2013):

$$\hat{P}_t = \exp(\hat{\delta}_t) \qquad (3)$$

The choice of the relevant explanatory variables to include in the model is crucial. We considered successively some qualitative and quantitative variables. The parsimony criterion implies compromising between the explanatory power of the model and its simplicity and applicability to local, regional, and national contexts. Beside our results, built according to the available data, we looked to other countries indexes.[4] In the present study and considering a multiple set of possible variables, we included in our models the following characteristics:

- Net Surface
- Number of bedrooms
- Presence of storage room
- Garage
- Land area (for housing)
- Garden (for housing)
- Detached (for housing)
- Parish typology
- New / Used

Many other variables were excluded because they were not statistically significant, while others were not included due to lack of availability.

Table 4: Parishes classification

Concelho	Freguesia	Tipo
Aveiro	Aradas	APU
Aveiro	Cacia	AMU
Aveiro	Eirol	AMU
Aveiro	Eixo	AMU
Aveiro	Esgueira	APU
Aveiro	Glória	APU
Aveiro	Nariz	AMU
Aveiro	Nossa Senhora de Fátima	AMU
Aveiro	Oliveirinha	AMU
Aveiro	Requeixo	APR
Aveiro	Santa Joana	APU
Aveiro	São Bernardo	APU
Aveiro	São Jacinto	APR
Aveiro	Vera Cruz	APU
Ílhavo	Gafanha do Carmo	APR
Ílhavo	Gafanha da Encarnação	APU
Ílhavo	Gafanha da Nazaré	APU
Ílhavo	Ílhavo (São Salvador)	APU

[4] We based our choice in several sources, but mainly on the Halifax House Price Indices (Fleming & Nellis, 1984) and the Notaires-INSEE indices published in France (Dubujet, 2000).

Regarding the geographical stratification we included a dummy variable APU using the new version of the Typology of Urban Area (TIPAU) published in 2009 by the National Institute of Statistics applied to parishes. The 2009 TIPAU is a tripartite classification of the parishes of the Portuguese territory in predominantly urban areas (APU), median urban areas (AMU) and predominantly rural areas (APR).[5] Table 2 shows the parishes of both municipalities indicating their typology. The dummy variable APU equal one, for dwellings located in areas predominantly of urban typology, and zero otherwise, joining together the other two remaining typologies (AMU e APR), i.e., the parishes with moderately urban and predominantly rural areas. Many other models have more complex location variables related with accessibility or proximity to central areas or equipment. These are certainly relevant for more precise estimation, but we consider that its application is not feasible for the construction of a national RPPI since its replication would put numerous practical problems.

3.3. Spatial data analysis

Hedonic models are normally based in cross-section data. In the present paper, we focus our attention on how the spatial effects affect the estimation of hedonic price. The presence of autocorrelation and spatial heterogeneity profoundly affects the quality of econometric models (Baumont, 2009).

The hedonic equations incorporate in most cases a set of spatial variables designed to capture some of the spatial effects mentioned above. However, measurement of these spatial effects and the choice of appropriate variables, presents numerous and practical difficulties. According to (Anselin, 1988; LeSage & Pace, 2009) among others, the spatial data analysis with the introduction of the geographical dimension, allows not only to capture the spatial dependence effect, but also to improve the quality of estimation since spatial dependence violates some of the Gauss-Markov assumptions of the OLS estimation (cross section observations are no longer independent) producing inefficient estimators. Two kinds of spatial effects are pointed out in the literature, namely: (i) spatial autocorrelation, revealing that contiguous regions may influence each other's performance through spillover effects and (ii) spatial heterogeneity, whenever the same functional form is erroneously considered for all regions (for comprehensive references about spatial econometric see for instance (Anselin, 1988; Guillain & Le Gallo, 2007; Le Gallo, 2002; LeSage & Pace, 2009).

To capture the interdependence between locations, it is necessary to consider the relative positions of each location. The choice of spatial weights is a key issue in spatial models. In the present work we adopt a standard approach, based on geographical proximity (contiguity). However geometric space is not always adequate to fully embrace the complexities of spatial structures (Bhattacharjee et al., 2012). For the adjacency criterion we use a normalized first order contiguity spatial weights matrix. Formally we define our weight matrix as follows:

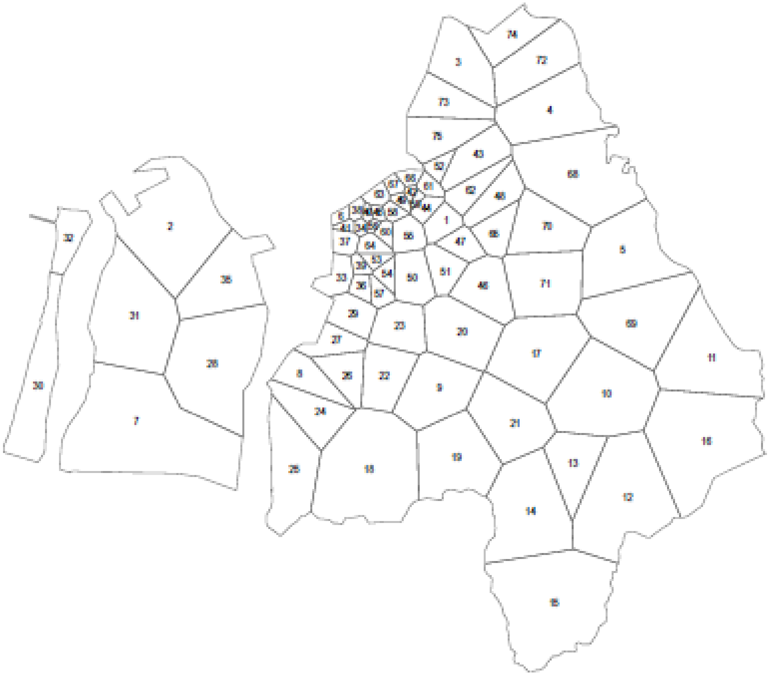
$$W = \begin{cases} w_{ij} = 0 \; if \; i = j \\ w_{ij} = 1 \; if \; d_{ij} = 0 \\ w_{ij} = 0 \; if \; d_{ij} > 0 \end{cases} \qquad (4)$$

This matrix is a square matrix, W, in which the term $w_{i,j}$ equal to one if the two locations i and j are contiguous and zero otherwise (d_{ij} represent the distance between the two regions). The weights matrixes are normalized with each line sum equal to one.

In our sample, the observations are not classified by any geographic information system, i.e., we do not have the coordinates of each individual residence. To circumvent this limitation we have built our contiguity matrix using the same methodology as in (Marques, 2012), delimiting 75 different geographical areas. These zones correspond to homogenous territories, normally smaller than parishes, and represent neighbourhoods, centre or other clusters which similar pattern (Figure 2). The working spatial contiguity matrix, W, appears in Figure 2. As we can see, the matrix W shows a sparse structure with most of the non-zero elements residing near the diagonal.

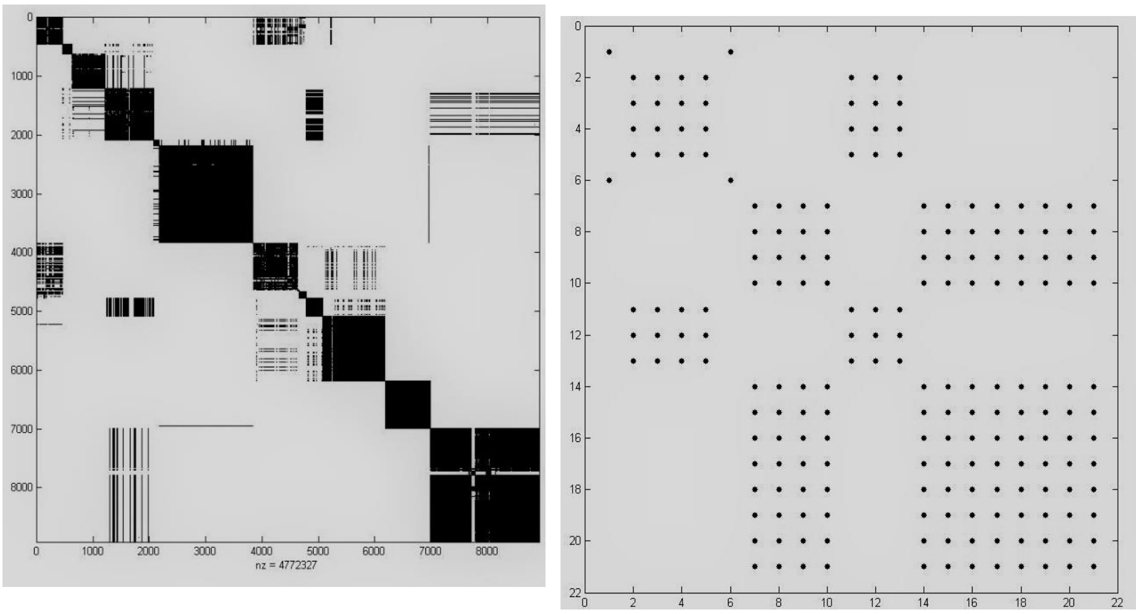
[5] For a detailed description consult www.ine.pt.

Figure 1: Codes of the 75 zones used in spatial weight matrix. Source: (Marques, 2012).



In our sample, the observations are not classified by any geographic information system, i.e., we do not have the coordinates of each individual residence. To circumvent this limitation we have built our contiguity matrix using the same methodology as in (Marques, 2012), delimiting 75 different geographical areas. These zones correspond to homogenous territories, normally smaller than parishes, and represent neighbourhoods, centre or other clusters which similar pattern (Figure 2). The working spatial contiguity matrix, W, appears in Figure 2. As we can see, the matrix W shows a sparse structure with most of the non-zero elements residing near the diagonal.

Figure 2: full 8921×8921 spatial weights (contiguity) matrix (left side), and a partial view (right side).



We use that contiguity matrix to, firstly, measure the degree of spatial autocorrelation of log prices. The spatial autocorrelation is based on the Moran's statistic (Moran's I), which can be represented by the expression:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} x_i x_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j} \quad (5)$$

in which $w_{i,j}$ represents the (i,j) element of the spatial contiguity matrix, W, x_{it} represents the logarithm of price per square meter of apartment i, and n corresponds to the number of observations. Moran's statistic estimates the linear dependence between a variable in a specific location and the mean of the same variable in the neighbourhood.

A positive value of Moran 'I statistic indicates the presence of correlation between the dwelling price and the average price of the surrounding habitations. The Moran scatter plot (Figure 4) depicts the log-price on the horizontal axis with the average values of the neighbouring houses on the vertical axis. A positive value of I implies that most houses remain in the first and second quadrants. The remains quadrants are considered as atypical locations.

There are several mechanisms through which an observation can be influenced by a neighboring observation. (Elhorst, 2014a) establishes a taxonomy of the effects of spatial dependence. Three types of interaction are considered:

- the endogenous effects: the decision variable depends on the decision of neighboring agents.
- exogenous effects: the decision variable depends on attributes of neighboring observations.
- the effects of spatial correlation resulting from unobservable variables or variables not included in the model.

A complete model with all types of spatial interaction takes the following form (Elhorst, 2014a):

$$\boldsymbol{Y = \partial WY + \alpha I_N + X\beta + WX\theta + u \ \ with \ \ u = \lambda Wu + \varepsilon} \quad (6)$$

Where Y represents a vector of dimension N × 1 containing the value of the dependent variable for each unit of the sample (i = 1, ..., N), WY represents the endogenous interaction effect on the dependent variable, I_N represents a vector of dimension N × 1 of ones associated with the constant term, α. X represents an N × K matrix of explanatory variables, β is a vector of dimension K × 1 with the parameters of the model to be estimated, WX represents the effect of exogenous interaction on the explanatory variables, Wu, the interaction effects on the error term, and finally, ε = (ε_1, ..., ε_N) represents a vector of the error term that we assume iid for all i. W represents the neighborhood matrix. It is a non-negative N × N matrix. As for the spatial interaction parameters, we have the term δ associated with the spatial dependence of the lagged dependent variable, λ the term associated with the error autocorrelation and θ, representing a vector of K × 1 parameters associated with each one of the explanatory variables.

Within spatial econometrics, the Spatial Error Model (SEM) (δ = 0, θ = 0 and λ ≠ 0), the Spatial Lag Model (SAR) (δ ≠ 0, θ = 0 and λ = 0) and the Spatial Durbin Model (SDM) (δ ≠ 0, θ ≠ 0 and λ = 0) have been widely used in the so-called spatial hedonic models (LeSage & Pace, 2009). The SEM model implies that the price of a house is affected by a missing variable associated with a neighboring house. In this case, prices are directly affected by the coefficient of each attribute given by the value of the estimated parameters of the model. In the SAR model, house prices are affected by the prices of neighboring houses. The reduced form of the model shows a spatial multiplier $(I - \rho W)^{-1}$ that must be considered when interpreting coefficients. In the SDM model the price of houses is affected not only by the price of neighboring houses, but also by the attributes of neighboring houses. The spatial multiplier takes the form$(I - \rho W)^{-1}(\beta_k + W\theta_k)$. Based on other restrictions on the parameters, we can find other models such as SDEM (δ = 0, θ ≠ 0 and ≠ λ0). However, these models are less used in the literature (Elhorst, 2014a).

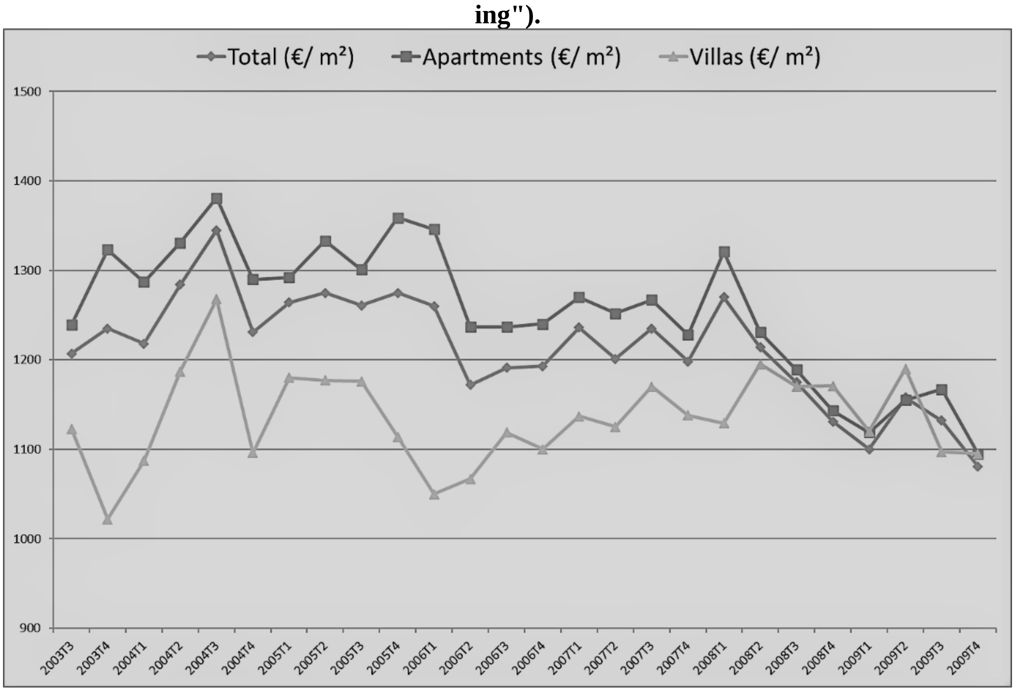
We found two approaches in the literature to select the best hedonic spatial model (see (Maslianskaïa-Pautrel & Baumont, 2016) for a comprehensive description). The first approach starts from the simplest model towards the most complete model, while the second approach does the reverse, evolving from the most complete model to the simplest model. In the bottom-up approach, Lagrange Multiplier Tests are applied, in their robust versions (Anselin et al., 1996) to the non-spatial model. In the top-down version, Spatial Durbin Model serves as a starting point. The most adapted model is then deduced by applying the Likelihood Ratio Test, step by step, on the various parameters of the more general model. (Elhorst, 2010) proposes a mixed approach. It first applies the Lagrange Multiplier Tests to the residue of the OLS model. If spatial autocorrelation is detected in the lag-dependent variable and in the error, the SDM model is directly estimated, and the Likelihood Ratio Tests are applied to compare it with the SAR and SEM restricted models.

4. RESULTS

Figure 3 exhibits the house price index of the INE (National Statistics Institute) "Survey on bank evaluation on housing". The average values of inquiries to bank assessment on housing published by the INE are measured in euros per m^2. The calculations are made for different typologies and for various levels of geographical disaggregation. Prices exist for all NUTSIII (regional level), for the two major metropolitan areas of Lisbon and Porto and, finally, also to urban areas and counties that include median-sized cities. Thus, we used from the INE database the average values of bank assessment on housing comprising the municipality of Aveiro in a quarterly series covering the period from the third quarter 2003 and 4th quarter 2009. The average values correspond to the two submarket apartments and townhouses, and their aggregation are made according to the method explained above (Section 2, equation 1). The area does not correspond exactly to our sample since it does not include the county of Ílhavo.

We can observe in Figure 3 a period of high prices between 2003 and 2005, with a peak in the third quarter of 2004. From there, we have a first decline in the first quarter of 2006 followed by a period of stability until 2007. The years 2008 and 2009 are marked by a downward trend in prices with a slight fall in the second quarter of 2009. This pattern is followed in a more or less similar way for both sub-markets (villas and apartments). Construction areas are normally higher in villas. Moreover, villas tend to be located far from the center. Therefore, villa's price per square meter is normally lower when compared with apartments.

Figure 3: Evolution of house price index published by the INE ("Survey on bank evaluation on housing").



For our exercise we will focus our attention on the apartment submarket, thus excluding the villa's market. With this emphasis, we favour the quality of observations and the homogeneity of the sample. Moreover, we will use only the data from 2005 to 2009, comprehending a total of 8162 observations. All estimations are carried out in Matlab, using the econometric tools made by J.P Elhorst and J. LeSage.[6]

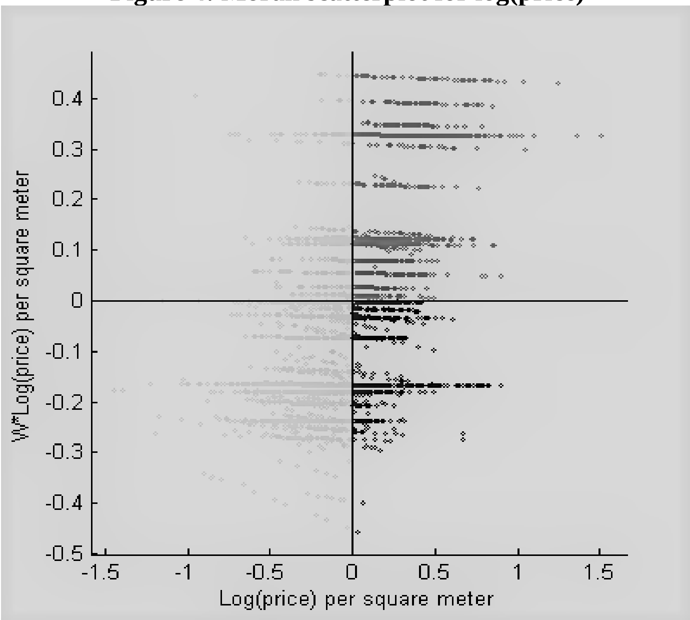
[6] All calculation are based on (Elhorst, 2014a) and (LeSage, 1999). We use the authors Toolbox function available respectively at http://www.regroningen.nl/elhorst/software.shtml and http://www.econ.utoledo.edu.

We start by measuring the global autocorrelation of the dependent variable, i.e., the logarithm of prices using the Moran's I Statistic. As it can be seen in Table 4, prices are positively auto correlated within space, with a Moran's I of 0.49 highly significant. This means that dwellings tend to group themselves according to price, with the segments more expensive concentrated in certain specific areas and the lower segments in other different areas.

Table 4: Moran's I statistic for the log of price per meter.

Variable	Moran's I	Moran's I stat.	P value
Log (price)	0.4916	45.85	0.0000

Figure 4: Moran scatterplot for log(price)



The Moran scatter plot (Figure 4), in which the average value of neighboring dwellings is plotted against the value of central dwelling, shows, as expected, that most of observation are in the quadrant HH and LL. We have 2717 observation in the HH quadrant, 841 in the LH quadrant, 3323 in the LL quadrant and 1281 in the HL quadrant, which means that 74% of apartments are in the HH and LL quadrant. The remaining observations (2122) are in the atypical quadrants. The Moran's I Statistic give us valuable indications on house pricing to concentrate and form clusters (Arbia, 2001). However, it tells us nothing about the spatial location of these specific manifestations of agglomeration. Thus, these global indexes, if relevant, can be an invitation to explore other local measures of agglomeration as statistical LISA (Local Indicator of Spatial Association) which decomposes the Moran's I Statistic to identify the individual contribution of each local site (in our case, each municipality).

Finally, we use a spatial econometric methodology to estimate the time dummy hedonic price model (Equation (2)). To choose the most appropriate spatial model, we applied the methodology proposed by (Elhorst, 2010). We estimated the OLS model and applied the robust LM tests (Anselin et al., 1996). Both tests (LM lag and LM error) confirm the presence of spatial dependence in the lagged dependent variable and in the error term. Therefore, the non-consideration of spatial dependence between the observations leads to biased and inconsistent estimators (LeSage & Pace, 2009). Based on these results, we estimate the SDM model and compare its results with the SAR and SEM models using the Likelihood Ratio Test to determine which model best fits the data. All results can be seen in Table 5.

We first compare SDM and SAR models. The LR-test of the SDM versus SAR model assumes a value of 2576.7, exceeding the critical value of 11.07 for a 95% significance level and 5 degrees of freedom. Secondly, we compare the SDM model with the SEM model. The LR-test statistic of the SDM model versus the SEM model assumes a value of 17.82, also exceeding the critical value of 11.07 for the same degrees of freedom and level of significance. According to the test, the SDM model performs better comparing with SAR and SEM models. Moreover, the SDM parameters are in general highly significant and with the correct signal. We therefore chose to distinguish SDM as the model that best fits the data.

Table 5: Estimation results and spatial tests (Source: own calculations based on Sapo-casa database)

Models	**OLS model**	**SAR**	**SEM**	**SDM**
Estimation				
Obs	8162	8162	8162	8162
log-likelihood	1032.1	4604.25	5888.72	5897.63
Const.	9.111794 (p=0.0000)	6.624933 (p=0.0000)	9.153512 (p=0.0000)	7.735180 (p=0.0000)
log(area)	-0.548759 (p=0.0000)	-0.529134 (p=0.0000)	-0.536364 (p=0.0000)	-0.476261 (p=0.0000)
Rooms	0.090233 (p=0.0000)	0.093107 (p=0.0000)	0.096875 (p=0.0000)	0.081607 (p=0.0000)
Garage	0.017429 (p=0.0006)	0.026519 (p=0.0000)	0.049877 (p=0.0000)	0.050609 (p=0.0000)
New	0.221423 (p=0.0000)	0.195516 (p=0.0000)	0.180144 (p=0.0000)	0.181860 (p=0.0000)
APU	0.214557 (p=0.0000)	0.132102 (p=0.0000)	0.096899 (p=0.0000)	0.173189 (p=0.0000)
Y2006	-0.013881 (p=0.0000)	-0.018385 (p=0.0271)	-0.027311 (p=0.0000)	-0.025449 (p=0.0017)
Y2007	0.003524 (p=0.0423)	-0.000221 (p=0.0974)	-0.006671 (p=0.0987)	-0.017351 (p=0.0223)
Y2008	-0.007849 (p=0.0036)	-0.014007 (p=0.0883)	-0.031135 (p=0.0000)	-0.031853 (p=0.0000)
Y2009	-0.011720 (p=0.0008)	-0.027409 (p=0.0020)	-0.062411 (p=0.0000)	-0.049826 (p=0.0000)
W.log(area)				-0.566897 (p=0.0000)
W.Rooms				0.151958 (p=0.0000)
W.Garage				-0.080487 (p=0.0000)
W.New				0.074093 (p=0.0000)
W.APU				-0.058548 (p=0.0571)
ρ		0.350981 (p=0.0000)		0.494964 (p=0.0000)
λ			0.90001 (p=0.0000)	

Direct and indirect effects

When spatial dependence effects are included in the model, the interpretation of coefficients associated with independent variables must be adapted. Those coefficients, associated to each independent variable, have two effects. The direct effect stems from the impact associated with an observation within the geographical unit (including the spillover feedback). The indirect effect stems from the impact associated with an observation located in the neighbourhood of the geographical unit under study.

In a cross-sectoral model, no role is assigned to the time factor. The estimated model is interpreted as a steady-state equilibrium model. However, the effects of spatial diffusion take time. Therefore, the interpretation of the coefficients must be done with caution. The indirect effects can be interpreted as a movement towards a steady state equilibrium (LeSage, 2008). The direct effects represent the partial derivative of the dependent variable with respect to the explanatory variables. Unlike the OLS model, this partial derivative cannot be read directly from the value of the parameter estimators. It must be corrected by the previously mentioned spatial multiplier.

Table 6 shows the decomposition of the direct and indirect effects of the SDM model comparing the results with the same model without correction for spatial autocorrelation (baseline model). The direct effects do not differ much from the estimated coefficients, still revealing non-negligible spillover effects. According to (Anselin & Lozano-Gracia, 2009) the direct effect is the only correct measure for pecuniary externality. According to the estimation results, and focusing our attention on the direct effects, the logarithm of the price per meter of housing varies negatively with the logarithm of the total area. On the other hand, the number of rooms, the existence of a garage, first-hand purchase and insertion in a predominantly urban area positively influence the log price of housing.

Table 6: Estimation of the baseline model (OLS without correction for spatial autocorrelation) and SDM model with direct and indirect effects.

Variables	Baseline	**SDM**	
		Direct effects	**Indirect effects**
Const.	9.111794 (p=0.0000)	7.743578 (p=0.0000)	
log(area)	-0.548759 (p=0.0000)	-0.478024 (p=0.0000)	-0.460327 (p=0.0000)
Rooms	0.090233 (p=0.0000)	0.081950 (p=0.0000)	0.078904 (p=0.0000)
Garage	0.017429 (p=0.0006)	0.050833 (p=0.0000)	0.048960 (p=0.0000)
New	0.221423 (p=0.0000)	0.182193 (p=0.0000)	0.175468 (p=0.0000)
APU	0.214557 (p=0.0000)	0.174469 (p=0.0000)	0.168170 (p=0.0000)
Y2006	-0.013881 (p=0.0000)	-0.025745 (p=0.0014)	-0.024806 (p=0.0016)
Y2007	0.003524 (p=0.0423)	-0.017655 (p=0.0220)	-0.017007 (p=0.0223)
Y2008	-0.007849 (p=0.0036)	-0.032251 (p=0.0000)	-0.031075 (p=0.0000)
Y2009	-0.011720 (p=0.0008)	-0.050460 (p=0.0000)	-0.048619 (p=0.0000)

With the model estimation, we proceed with the construction of two RPPIs based on the annual dummy variables. The first uses the coefficients estimated from the OLS model. The second uses the direct effects associated with the annual dummy variables estimated in the SDM model. Both series can be seen in Table 8. A simple Paired-Samples T Test procedure reject the null hypotheses of equality between the two series, with a significance value of 0,084. Moreover, the standard errors of the coefficients on the dummy variables for time are lower for the SDM estimation compared with the OLS, which may be used as a criteria for our index evaluation (Eurostat, 2017).[7]

Both models display a similar trend, although with important differences in the price level. As it can be seen in Figure 5, prices decrease from 2007 onward. The two RPPI series run almost parallel. After a break in 2006, we see a slight recovery in 2007, followed by a further break in 2008 and 2009. However, the RPPI based on the SDM spatial model presents a more pessimistic scenario. The falls are more pronounced, especially in 2006 and 2009 and the 2007 recovery is less significant.

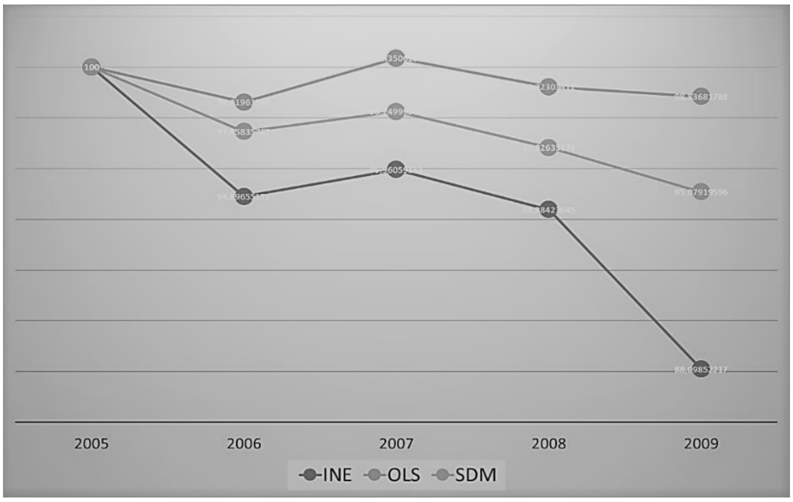
Table 7: RPPOs evolution, with (SDM) and without spatial dependence (baseline OLS).

Year	2005	2006	2007	2008	2009
OLS	100,00	98,62	100,35	99,22	98,84
SDM	100,00	97,46	98,25	96,83	95,08
% Change (OLS)	-	-1,38%	1,76%	-1,12%	-0,39%
% Change (SDM)	-	-2,54%	0,81%	-1,45%	-1,80%

Figure 5 compares the evolution of the two real property indexes, with (SDM) and without correction for autocorrelation (OLS) with the only indicator available in Portugal covering the evolution of real estate prices during the same period. Figure 6 compares the evolution of the two real estate indices, with (SDM) and without spatial autocorrelation correction (OLS), with the INE index that covers the evolution of real estate prices during the same period in the same geographic area of our study. According to the "Survey on bank evaluation on housing bank inquiries", the price drop between 2007 and 2009 was 10.8%. This period corresponds to the subprime crisis that affected the global economy. In the same interval, the decrease in the index built from the OLS model is only 2%. On the other hand, the index based on the SDM model reveals a drop of 5%, much closer to reality represented by the housing bank inquiries. Admitting that the "Survey on bank evaluation on housing bank inquiries" represents a good approximation of the sector's reality, we can conclude that the SDM RPPI better reflects the evolution of the real estate market compared to the classic OLS RPPI.

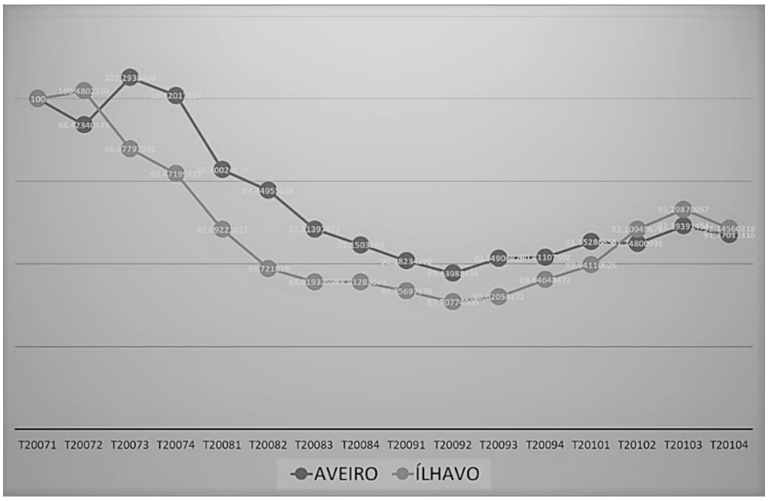
Figure 6 illustrates the quarterly evolution of the Confidencial Imobiliário Index between 2007 and 2010. Like the INE index, the Ci shows a 10% drop in prices between 2007 and 2009. Between 2009 and 2010, it shows a recovery. Also here, the SDM index that incorporates the correction for spatial autocorrelation is closer to the Confidential Index than the OLS Index, indicating the relevance of incorporating the elements of spatial econometrics.

Figure 5: RPPIs graph with (SDM) and without spatial effects (OLS) compared with the evolution of average prices obtained from the "Survey on Bank evaluation on Housing" published by the INE (Portuguese Statistic Institute).



[7] For longer time series, (Guo et al., 2014) use the index returns volatility and the first order autocorrelation to compare indexes.

Figure 6: Evolution of the Confidencial Imobiliário Index between 2008 and 2010 (Quarterly data provided by Confidencial Imobiliário).



5. CONCLUSION

This article builds a reliable RPPI using a spatial hedonic model. Property prices exhibits strong spatial correlation. The spatial correlation affects the validity of the OLS estimates used in the classic hedonic price models used in the literature. Therefore, the need to find tools to model spatial dependence has become obvious. According to ours results, the RPPI estimated from a SDM model applied to the period between 2005 and 2009 performs better compared to the classic OLS. Moreover, the results are closer to the real estate indicators available at INE in the same period.

Quantity and quality of available data are decisive for the choice of the type of price index. The methods, based on the first moments of the price distribution (or median), are very attractive for its simplicity and allow, provided there is sufficient degrees of freedom, to obtain disaggregated prices for various sub-markets, therefore, revealing good and precise information to the market. When the number of observations is scarce, hedonic methods represent a valid alternative, provided that the available observations are accompanied by qualitative information about the characteristics of the property traded.

In line with the literature, home prices show a high degree of spatial correlation not captured by traditional models. In this sense, the results from spatial econometric models show significant differences when compared with the results obtained with the traditional OLS model, thus raising relevant questions about the consistency of these estimators. Even in full hedonic models incorporating a wider set of variables intended to capture the territorial spatial effects, the literature reveals frequently the presence of auto correlation in the residuals (Baumont, 2009). This finding should have, in our opinion, a significant impact in the construction of a methodological framework for the development of RPPIs with national coverage, even considering the necessity of these methods being parsimonious and replicable in different territories. Given the difficulty of harmonizing the development of territorial variables impacting on house prices (CBO, services, equipment, infrastructure etc.) it seems that the tools of spatial econometrics will be an instrument to take increasingly into account in this matter.

Still considering the contribution to a methodological framework for the construction of real estate price indices or RPPIs, it is clear the need to find, upstream, the best mechanisms for the creation of a database that feeds the estimation of indexes. In this sense, the literature and the various practices point to four data sources: the notarial sources for transactions registrations, the fiscal institutions that assess the assets for tax purposes, the real estate agencies that promote business and finally the banking that provides credit for the purchase of housing. Any of these solutions have advantages and disadvantages. Considering the need to monitor the evolution of the real estate price level, eventually integrated in a wider framework of building a broader indicator of prices evolution, the crucial observations consist fundamentally of the trade register, if possible, performed in real time. In this case, the recording notary shall certainly be the more comprehensive mechanism, in that not all sales need credit and many of these transactions are performed outside the domain of real estate agencies or without the bank intermediation. If the objective is to assess property wealth for national accounting or fiscal purpose or even to assess solvency of institutions or agents, the databases of national or regional fiscal institutions are certainly a source unavoidable.

Whatever the solutions or protocol to adopt to operationalize the constitution and updating of a real estate database, the involvement of public institutions, either through the INE, or the academic institutions, seems of utmost importance in a context where new technological solutions related to georeferencing data open new perspectives to econometricians. Therefore, and considering the results that reinforce the importance of spatial effects in the distribution of prices, it is no longer possible to think about the creation and updating of real estate databases without including the issue of developing, in parallel, a geographical information system that allows to incorporate the spatial component in econometric models' routines based on georeferenced data. The present study, despite its limited scope, demonstrates the relevance of spatial effects and raises the need for more studies with more recent data with greater geographical area.

REFERENCES

Anselin, L. (1988). *Spatial econometrics: methods and models* (Vol. 4). Springer.

Anselin, L., Bera, A. K., Florax, R., & Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, *26*(1), 77–104. https://doi.org/10.1016/0166-0462(95)02111-6

Anselin, L., & Lozano-Gracia, N. (2009). *Spatial Hedonic Models BT - Palgrave Handbook of Econometrics: Volume 2: Applied Econometrics* (T. C. Mills & K. Patterson (eds.); pp. 1213–1250). Palgrave Macmillan UK. https://doi.org/10.1057/9780230244405_26

Arbia, G. (2001). The role of spatial effects in the empirical analysis of regional concentration. *Journal of Geographical Systems*, *3*(3), 271–281.

Bajat, B., Kilibarda, M., Pejović, M., & Petrović, M. S. (2018). Spatial Hedonic Modeling of Housing Prices Using Auxiliary Maps. In *Spatial Analysis and Location Modeling in Urban and Regional Systems* (pp. 97–122). Springer.

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. CRC press.

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

Baumont, C. (2009). Spatial effects of urban public policies on housing values*. *Papers in Regional Science*, *88*(2), 301–326.

Bhattacharjee, A., Castro, E., & Marques, J. (2012). Spatial interactions in hedonic pricing models: the urban housing market of Aveiro, Portugal. *Spatial Economic Analysis*, *7*(1), 133–167.

Bourassa, S. C., Cantoni, E., & Hoesli, M. (2007). Spatial dependence, housing submarkets, and house price prediction. *The Journal of Real Estate Finance and Economics*, *35*(2), 143–160.

Brunauer, W., Feilmayr, W., & Wagner, K. (2012). A new residential property price index for Austria. *Statistiken–Daten Und Analysen Q*, *3*, 90–102.

Chasco, C., & Le Gallo, J. (2012). Hierarchy and spatial autocorrelation effects in hedonic models. *Economics Bulletin*, *32*(2), 1474–1480. https://www.researchgate.net/publication/254407054_Hierarchy_and_spatial_autocorrelation_effects_in_hedonic_models

Diewert, W. E., Heravi, S., & Silver, M. (2009). Hedonic imputation versus time dummy hedonic indexes. In *Price index concepts and measurement* (pp. 161–196). University of Chicago Press.

Dubé, J., & Legros, D. (2014). Spatial econometrics and the hedonic pricing model: what about the temporal dimension? *Journal of Property Research*, *31*(4), 333–359.

Dubujet, F. (2000). Les indices de prix Notaires-Insee. *Unpublished Manuscript*, *23*, 2000.

Elhorst, J. P. (2010). Applied spatial econometrics: raising the bar. *Spatial Economic Analysis*, *5*(1), 9–28.

Elhorst, J. P. (2014a). *Spatial econometrics: from cross-sectional data to spatial panels* (Vol. 479). Springer.

Elhorst, J. P. (2014b). Spatial panel data models. In *Spatial econometrics* (pp. 37–93). Springer.

Elhorst, J. P., & Vega, S. H. (2015). The SLX model. *Journal of Regional Science*, *55*(3), 339–363.

Eurostat. (2017). Commercial property price indicators: sources, methods and issues. In *Eurostat Methodologies & working papers* (2017th ed.). European Commission. https://doi.org/10.2785/050176

Fleming, M. C., & Nellis, J. G. (1984). *The Halifax house price index: technical details*. Halifax Building Society.

Francke, M., & Van de Minne, A. (2020). Modeling unobserved heterogeneity in hedonic price models. *Real Estate Economics*. https://doi.org/10.1111/1540-6229.12320

Gibbons, S., & Overman, H. G. (2012). Mostly pointless spatial econometrics? *Journal of Regional Science*, *52*(2), 172–191.

Griliches, Z. (1971). Hedonic price indexes for automobiles: an econometric analysis of quality change. *Price Indexes and Quality Change*, 55–87.

Guillain, R., & Le Gallo, J. (2007). Agglomeration and dispersion of economic activities in Paris and its surroundings: An exploratory spatial data analysis. *Regional Economics Applications Laboratory Discussion Paper*.

Guo, X., Zheng, S., Geltner, D., & Liu, H. (2014). A new approach for constructing home price indices: The pseudo repeat sales model and its application in China. *Journal of Housing Economics*, *25*, 20–38. https://doi.org/10.1016/j.jhe.2014.01.005

Haurin, D. R., & Hendershott, P. H. (1991). House price indexes: issues and results. *Real Estate Economics*, *19*(3), 259–269.

Hill, R. J. (2004). Constructing price indexes across space and time: the case of the European Union. *American Economic Review*, 1379–1410.

Hill, R. J. (2013). Hedonic price indexes for residential housing: A survey, evaluation and taxonomy. *Journal of Economic Surveys*, *27*(5), 879–914.

Hill, R. J., & Melser, D. (2008). Hedonic imputation and the price index problem: an application to housing. *Economic Inquiry*, *46*(4), 593–609.

Hoffmann, J., & Lorenz, A. (2006). Real Estate Price Indices for Germany: past, present and future. *Present and Future (November 30, 2006)*.

Iturra, V., & Paredes, D. (2013). Construction of a Spatial Housing Price Index by Estimating an Almost Ideal Demand System. *Tijdschrift Voor Economische En Sociale Geografie*, *105*, 301–314. https://doi.org/10.1111/tesg.12067

Kestens, Y., Thériault, M., & Des Rosiers, F. (2006). Heterogeneity in hedonic modelling of house prices: looking at buyers' household profiles. *Journal of Geographical Systems*, *8*(1), 61–96.

Lancaster, K. J. (1966). A new approach to consumer theory. *The Journal of Political Economy*, *74*(2), 132–157.

Le Gallo, J. (2002). Econométrie spatiale: l'autocorrélation spatiale dans les modèles de régression linéaire. *Economie & Prévision*, *4*, 139–157.

LeSage, J. P. (1999). *Spatial econometrics*. Regional Research Institute, West Virginia University.

LeSage, J. P. (2008). An introduction to spatial econometrics. *Revue d'économie Industrielle*, *3*, 19–44.

LeSage, J. P., & Pace, K. (2009). Spatial Econometrics Introduction to. In *Systems Engineering*.

Maleyre, I. (1997). L'approche hédonique des marchés immobiliers. *Etudes Foncières*, *76*, 22–29.

Mark, J. H., & Goldberg, M. A. (1984). Alternative housing price indices: an evaluation. *Real Estate Economics*, *12*(1), 30–49.

Marques, J. J. L. (2012). The notion of space in urban housing markets. *Ph.D. Thesis. University of Aveiro: Portugal*.

Maslianskaïa-Pautrel, M., & Baumont, C. (2016). Environmental spillovers and their impacts on housing prices: A spatial hedonic analysis. *Revue d'Economie Politique*, *126*(5), 921–945. https://doi.org/10.3917/redp.265.0921

McDonald, C., & Smith, M. (2009). *Developing stratified housing price measures for New Zealand*. Reserve Bank of New Zealand.

Oust, A., Hansen, S. N., & Pettrem, T. R. (2019). Combining Property Price Predictions from Repeat Sales and Spatially Enhanced Hedonic Regressions. *The Journal of Real Estate Finance and*

Economics, 1–25.

Pace, R. K., Barry, R., Clapp, J. M., & Rodriquez, M. (1998). Spatiotemporal autoregressive models of neighborhood effects. *The Journal of Real Estate Finance and Economics*, *17*(1), 15–33.

Paredes, D. J. C. (2011). A methodology to compute regional housing price index using matching estimator methods. *Annals of Regional Science*, *46*(1), 139–157. https://doi.org/10.1007/s00168-009-0346-z

Pesaran, M. H., & Tosetti, E. (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics*, *161*(2), 182–202.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *The Journal of Political Economy*, *82*(1), 34–55.

Se Can, A., & Megbolugbe, I. (1997). Spatial dependence and house price index construction. *The Journal of Real Estate Finance and Economics*, *14*(1–2), 203–222.

Seya, H., Yoshida, T., & Yamagata, Y. (2020). Spatial econometric models. In *Spatial Analysis Using Big Data* (pp. 113–158). Elsevier.

Shiller, R. J. (1991). Arithmetic repeat sales price estimators. *Journal of Housing Economics*, *1*(1), 110–126.

Triplett, J. (2004). *Handbook on hedonic indexes and quality adjustments in price indexes: Special application to information technology products*.

Wang, F. T., & Zorn, P. M. (1997). Estimating house price growth with repeat sales data: what's the aim of the game? *Journal of Housing Economics*, *6*(2), 93–118.

Waugh, F. V. (1928). Quality factors influencing vegetable prices. *Journal of Farm Economics*, *10*(2), 185–196.

Wilhelmsson, M. (2008). House price depreciation rates and level of maintenance. *Journal of Housing Economics*, *17*(1), 88–101.

Wilhelmsson, M. (2009). Construction and updating of property price index series. *Property Management*.

Yang, C. F. (2020). Common factors and spatial dependence: an application to US house prices. *Econometric Reviews*, 1–37.

Acknowledgments: The authors would like to thank the reviewers for their valuable contributions, comments, and suggestions. This work was financially supported by the Research Unit on Gov[1]ernance, Competitiveness and Public Policies (UIDB/04058/2020) + (UIDP/04058/2020), funded by national funds through FCT—Fundação para a Ciência e a Tecnologia, and has been supported by Portuguese national funds through the Foundation for Science and Technology, FCT, I.P., in the context of the JUST_PLAN project (PTDC/GES-OUT/2662/20)