DOI: 10.7764/RIC.00044.21

ENGLISH VERSION......

The use of geostatistics to estimate missing data in a spatial econometric model of housing prices

El uso de la geoestadística para la estimación de datos faltantes en un modelo econométrico espacial del precio de la vivienda

I. Tamaris Turizo[1*] https://orcid.org/0000-0002-3486-9471
J. Chica Olmo[**] https://orcid.org/0000-0003-2274-7829
R. Cano Guervos[**] https://orcid.org/0000-0003-3425-8723

* Universidad Pedagógica y Tecnológica de Colombia UPTC, COLOMBIA
** Universidad de Granada, Granada, ESPAÑA

Fecha de Recepción: 26/07/2021
Fecha de Aceptación: 18/03/2022
Fecha de Publicación: 02/12/2022
PAG 423 -434

Abstract
Housing prices have been the subject of many studies, and some of them have tried to determine the influencing structural and location factors through hedonic econometric models. One of the main factors considered in the literature on real estate appraisals is the location of the dwellings. For this reason, this study combines the spatial methodologies of geostatistics and spatial econometrics. On the one hand, this work uses geostatistics to estimate missing data to account for the lack of information in the sampled real estate websites. On the other hand, the explanatory factors of prices are determined through spatial econometrics. The combination of both methods facilitates estimating housing prices in Santa Marta (Colombia), solving the problem of missing data. In the modeling, the problems of spatial heteroscedasticity and multicollinearity are corrected. This combination of methods could be of great interest to companyies and public agencies related to real estate activity, which is sustained by the information available on these real estate websites.

Keywords*: Hedonic mode; geostatistics; spatial econometrics; housing prices; missing data; heteroscedasticity*

Resumen
El precio de la vivienda ha sido objeto de estudio en múltiples trabajos, en algunos de los cuales se ha tratado de determinar los factores constructivos y localizativos que influyen en el mismo, aplicando modelos hedónicos econométricos. Uno de los principales factores considerados en la literatura sobre valoración inmobiliaria es la localización de las viviendas, por lo tanto, que se considera pertinente utilizar tanto métodotos geoestadísticos cómo de la economtría espacial. En éste trabajo la geoestadística se utiliza para la estimación de datos faltantes, por la carencia de información en los portales inmobiliarios muestreados. Por otra parte, la econometría espacial se utiliza para determinar los factores explicativos de los precios. La combinación de estos dos métodos ha permitido estimar el precio de la vivienda en Santa Marta (Colombia), resolviendo previamente el problema de los datos faltantes. En el modelado se corrigen los problemas de heteroscedasticidad espacial y multicolinealidad. Esta combinación de métodos puede ser de gran interés para las empresas y organismos públicos relacionados con la actividad inmobiliaria, que se nutren de la información disponible en dichos portales inmobiliarios.

Palabras Claves: Modelo hedónico; geostadística; econometría espacia; precio de la vivienda; datos faltantes; heterocedasticidad

[1] ***Correspondence author:***
Facultad de Ciencias Económicas y Administrativas, Universidad Pedagógica y Tecnológica de Colombia UPTC, ivantamaris@correo.ugr.es

DOI: 10.7764/RIC.00044.21

ENGLISH VERSION...

1. Introduction

When it comes to estimating the price of urban real estate, the theory of hedonic prices (which values property according to the profit generated by its features) is one of the most widely used, as it allows for determining the implicit prices of the features that determine the price of such property. These implicit prices are obtained through the application of the hedonic regression model (Rosen, 1974), in which the coefficients of the explanatory features affecting the price of the property measure the profit they report. However, on many occasions, the application of the regression model causes problems related to multicollinearity (Chinloy, 1977); (Mark, 1980); (Tiwari and Parikh, 1998) and heteroscedasticity (Rehák and Káčer, 2019); (Trojanek and Gluszak, 2018), leading to efficiency issues for the estimated coefficients of the model.

In one of his first works, (Goodman, 1978) used the hedonic price method to estimate housing prices based on the expression $P = f(C)$*, where P represents the price of each dwelling and C is the set of components that are considered to contribute to explain the price. The selling price of the dwelling will depend on the structural quality of these components. Therefore, under equilibrium conditions, the marginal price of each of the components will be equal to their marginal valuation (Rosen, 1974); (Nelson, 1978). The econometric method is useful for estimating such marginal effects and the price of urban housing (Can, 1992).*

The application of hedonic models to determine prices in the housing market has continued to be important, as it is considered an essential tool and a starting point for the analysis of the urban housing market (Can, 1990); (Goodman and Thibodeau, 2003); (Li et al., 2013); (Zhang et al., 2019).

One of the most widely used methods for estimating econometric models is the Ordinary Least Squares (OLS), but this method does not consider spatial effects (Anselin, 1992). In the context of OLS, the model coefficients are estimated under the assumption of no spatial autocorrelation in the disturbances (Zhu et al., 2011). But this assumption is not usually met in the case of hedonic housing price models since prices are not independent of their spatial location (Case et al., 2004);(Wen et al., 2017), leading to inefficient estimates (Anselin, 1988).

Subsequent research developed techniques that aim to overcome the limitations of the OLS method in the presence of spatial autocorrelation by improving the estimate of prices, such as geostatistics (Dubin, 1992); (Chica-Olmo, 1995); (Basu and Thibodeau, 1998); (Yoo and Kyriakidis, 2009); (Gröbel and Thomschke, 2018) and spatial econometrics (Can, 1992); (Pace et al., 1998); (Von Graevenitz and Panduro, 2015); (Wen et al., 2017).

The features that explain housing prices can be classified into construction and location (Chica-Olmo, 2007); (Li et al., 2012); (Li et al., 2019); (Poeta et al., 2019). The construction features most frequently used in hedonic models are housing area, number of bedrooms, number of bathrooms, kitchen, and garage (with positive effects on house price), as well as age (with a negative effect on housing price) (Goodman and Thibodeau, 2003); (Clapp, 2004); (Conroy and Milosch, 2011); (Sah et al., 2015);(Li et al., 2016); (Xiao et al., 2017).

In addition, the locational features are determined by the location of the dwelling in relation to its surroundings (Li et al., 2013); (Simlai, 2014); (Sah et al., 2015); (Chica-Olmo et al., 2018). For this reason, it is presumed that houses close to each other tend to have similar prices. According to Tobler's (1970) geographic law, "everything is related to everything else, but near things are more related than distant things." This explains why housing prices show a high degree of spatial dependence since location factors that strongly influence price, such as accessibility or proximity to externalities, have a similar influence on nearby dwellings. Specifically, location can generate a positive effect if the property is close to an amenity or attraction, but it can also have a negative effect when the property is affected by a negative externality in the neighborhood (Can, 1992), (Zhang et al., 2014). For example, in some cases, factors such as noise, safety or vibrations have an adverse effect on home appraisals (Bowes and Ihlanfeldt, 2001), or other factors such as atmospheric pollution, electromagnetic radiation, noise pollution, taxes, among other factors, also have a negative impact on housing prices (Chay and Greenstone, 2005); (Geng et al., 2015). This spatial dependence on housing prices will also lead to spatial autocorrelation in the disturbances of the econometric model (Can, 1998); (Basu and Thibodeau, 1998).

Regarding Latin America, some studies on hedonic prices have been carried out. They have included construction and location variables in their models. Thus, including location variables has improved the results of the models. For example, in the city of Medellín (Colombia), the results of the model improved after including the spatial component (Urrea et al., 2019). Some studies also showed that structural features had more influence on housing prices than locational features, as shown in the city of Juarez (Mexico) (Fierro et al., 2009) or in the Metropolitan Region of Chile, where age and surface area were the most relevant variables (Sagner, 2011).

However, other characteristics related to the quality of life and income level have also been used to estimate hedonic price models, as in the city of Buenos Aires (Argentina), where the results were consistent and showed a strong correlation between price, income level and neighborhood characteristics (Cruces et al., 2008). Similarly, in the city of Santa Marta

DOI: 10.7764/RIC.00044.21

(Colombia), a strong relationship was found between income level and the demand for this real estate (Garza and Ovalle, 2019).

It is important to consider that public services are not adequately provided in several areas of Latin America. This deficit in services generates a differential that affects housing prices in these areas. Thus, in Machala (Ecuador), a study to assess the price of housing rental concluded that the most relevant variables, those with the greatest impact, were the quality and availability of public services, together with the distance to the city's central park (Zambrano-Monserrate, 2015). In Brazil, it was observed that areas with an adequate water supply and garbage collection services significantly increase housing rental prices (Morais and Cruz, 2015).

In the Latin American region, there have also been studies that considered the ease of access to public transportation through the proximity of the dwelling to an urban transportation station. These studies show that proximity has an impact but in different ways. According to (Rodríguez and Targa, 2004) and (Muñoz-Raskin, 2010), proximity to transport stations positively influences housing values in the case of Bogota (Colombia), while in Buenos Aires (Argentina), the impact of proximity to transport stations on property prices is not so significant (D' Elia et al., 2020). However, in Santa Marta (Colombia), it has been observed that proximity to freight railways has a negative effect on housing prices (Chica-Olmo et al., 2018).

A problem when applying regression models to models involving different information sources is the absence of information on some explanatory variables. Sometimes the sample information is not consistently provided by the different sources. This problem must be addressed by using some statistically valid method. (Anselin, 1992) suggests solving this type of difficulty by applying the Kriging interpolation technique to predict the values at the locations where no observations are available. Therefore, this is a tool of special importance when seeking to improve the estimates in case data are unavailable for all the variables at the sampling points.

In this work, three hypotheses are proposed:

H1. Incorporating missing data information improves the quality and significance of the model variables.

H2. Housing prices in the city of Santa Marta show an increasing spatial trend from the outskirts to the coast.

H3. Housing prices do not necessarily have to be related to some points of interest.

This paper is structured as follows: after the literature review related to hedonic prices and the different methods used for their determination, there is a brief explanation of the spatial methodologies used in this work, such as geostatistics and spatial econometrics. Then, the main results obtained by applying the aforementioned methods are presented. Finally, the conclusions of the study are included.

2. Methodology

2.1 Geostatistics

Classically, hedonic models have been estimated using Ordinary Least Squares (OLS). However, this technique may present problems since it does not consider the presence of spatial dependence, which means that the housing price is not independent of its location in space. In addition, it may present structural instability, which implies that the parameters take different values if the property is located in a certain area or not (Baronoi et al., 2018).

Fortunately, the development of Geographic Information Systems (GIS), spatial econometrics and geostatistics has enabled to consider the presence of spatial dependence.

Through GIS, a cartographic model is generated that shows the relationship between the housing price and its location, i.e., the spatial distribution of housing prices can be observed. According to (Can, 1998), GIS is a powerful tool for obtaining a geographic visualization of dwellings to facilitate spatial analysis and data management.

GISs also contribute to spatial analysis using geostatistical methods, which are extremely useful tools to be applied in real estate. Also, through the application of the geostatistical Kriging interpolation tool, it is possible to extend the estimates to areas where there is a lack of sample information.

Kriging has proven to be a useful tool not only in real estate or mining, where it originated but also in other fields of study. For example, in the area of health, it was possible to identify the geographic disparities of breast cancer in the USA, using Kriging to solve the problem of the lack of homogeneous information caused by a change in the data collection methodology, with almost 70% of missing data (Chien et al., 2013). This geostatistical tool has also been used in meteorology to solve problems of incomplete information. According to (Kilibarda et al., 2014), it is required to use interpolation methods when forecasting daily temperatures from spatio-temporal data. When pixel loss of more than 50% occurs, Kriging regression models are used to obtain better maps resulting from a global geostatistical model.

DOI: 10.7764/RIC.00044.21

Geostatistics has also been used in real estate appraisals with missing data through the Cokriging method. This method was used to estimate the prices of commercial premises in Toledo (Spain), with auxiliary information provided by the prices of surrounding dwellings, which significantly improved the quality of price predictions (Montero-Lorenzo et al., 2009). This methodology has also been applied in the city of Granada when using heterotopic data samples to estimate housing prices (Chica-Olmo, 2007).

Besides geostatistics, other research solved the problem of missing data using different methods. (Lesage and Pace, 2004,) in their study on hedonic prices to determine housing values, used a Monte Carlo experiment to improve the sample, managing to reduce the prediction errors by more than 75%. Also, (Kelejian and Prucha, 2010) applied a model with spatial lags to overcome this problem. They worked with the complete sample and obtained better estimators that minimize the disturbance terms associated with the spatial correlation. In addition, (Li and Li, 2018) addressed this problem by using a Cubic Spline Interpolation Model to analyze the negative relationship between the bad odor from landfills and housing prices in Hong Kong.

2.2 Spatial Econometrics

Furthermore, in order to carry out the spatial econometric modeling, an Exploratory Spatial Data Analysis (ESDA) can be performed to contrast the possible presence of spatial autocorrelation (Anselin et al., 2006). Finally, SAR/SEM spatial regression models are estimated, and the Lagrangian Multiplier (LM) is used (Anselin, 1988) to choose the model with the best significance.

In spatial econometrics, when considering the presence of spatial autocorrelation, it is necessary to specify the form and degree of the neighborhood between georeferenced data. For this purpose, the W-weight matrix is used, which helps to establish multidirectional spatial relationships, i.e., the interdependence between regions (Can, 1990); (Corrado and Fingleton, 2011); (Anselin and Rey, 2012).

To carry out the house pricing estimate in this study, we will use spatial econometrics and geostatistics as techniques that rely on spatial effects related to location. On the one hand, the lack of information in some sample observations has been overcome by applying the Kriging spatial interpolator, which helped to enlarge the sample. On the other hand, spatial econometrics has been applied to perform tests to validate assumptions and estimate the effects of price drivers.

When data is searched throughout space, the value will depend on the location, which translates into structural instability in space because the observations are not homogeneous. This aspect implies heterogeneity in the data, so the previous condition implies that heteroscedasticity problems are generated (Anselin, 1988). A proposal to solve this problem is based on the application of spatial models developed by (Cliff and Ord, 1973); (Cliff and Ord, 1981), widely used in the field of spatial modeling (Kelejian and Prucha, 1998), (Kelejian and Prucha, 1999), (Kelejian and Prucha, 2010); (Arraiz et al., 2010), which is based on the following expression:

$$y = \rho W y + X\beta + u$$
$$u = \lambda W u + \varepsilon$$

In this expression, y is a n x1 vector of observations of the explained variable, X is the n x k matrix of observations of k explanatory variables, W is the n x n matrix of known spatial weights, β is a k x 1 vector of regression parameters, λ and ρ are scalar autoregressive parameters, u is a n x 1 vector of regression disturbances which may present spatial autocorrelation and ε are normally and independently distributed disturbances.

If λ and ρ are different from zero, the model is called SARAR. When ρ is different from 0 and λ is equal to zero, it is called SAR (Spatial autoregressive), and when ρ is equal to 0 and λ is different from zero, it is called SEM (Spatial Error Model). Of these three models, SAR and SEM are the most widely used in practice. Lagrange tests (LM) are used (Anselin et al., 1996) to determine the most appropriate model.

3. Data and area of study

This study estimates housing prices in the city of Santa Marta (Colombia), which is the capital of the Magdalena Department. The city has a population of 479.000 inhabitants (according to the 2018 census of the National Administrative Department of Statistics (from the Spanish Departamento Administrativo Nacional de Estadística, DANE). This city has an area of 2,393 km², of which 36.6 km² belong to the urban area. The urban area is known for being monocentric, having its main commercial and business center next to the coastal area on the shores of Santa Marta Bay, which belongs to the Caribbean Sea. According to the DANE census, 75% of the dwellings are single-family homes.

DOI: 10.7764/RIC.00044.21

The information related to the sample data of the properties for sale in Santa Marta was obtained from different real estate agency websites. Specifically, the real estate websites consulted were Metrocuadrado, Safe, Coldwell Banker and Habitat Kasa. Data collection focused on the urban perimeter of the city of Santa Marta, in particular, on second-hand single-family dwellings for sale. The prices offered for these properties were obtained from these websites. The construction or structural features observed were the price of the dwelling (RV), expressed in millions of Colombian pesos; number of bedrooms (HABITAC); number of bathrooms (BANOS); constructed area in square meters (M2); age, in years (ANTIG); the dichotomous variables GARAGE: 1 if the dwelling has a garage and 0 if it does not; and PISCINA: 1 if the dwelling has a swimming pool and 0 if it does not. This cross-sectional information, obtained from real estate websites during the first half of 2016, represents approximately 24% of the dwellings offered in that period, as approximately 900 single-family properties were for sale.

In addition, the following location variables were considered: distance to the nearest university (DUNIV), distance in km to the nearest supermarket (DSUPERMDO), distance in km to the nearest shopping center (DCTROCCIAL), distance in km to the center (DISTC) and the UTM coordinates of each dwelling in deviations from the mean (COORDXM and COORDYM). These coordinates allow considering the increasing spatial trend of prices (Chica-Olmo, 1995); (Chica-Olmo, 2007).

The information is cross-sectional and corresponds to dwellings located in the urban perimeter of the city of Santa Marta (between the 1st and 66th lanes and between the 4th and 50th streets approximately). It is important to mention that the information provided by the different real estate websites is heterogeneous. In particular, not all the websites provided the age of the property. In short, within the sample of 211 dwellings, only complete data was available for 142 dwellings. Figure 2 shows the location of the sample dwellings and the price. This figure shows that the highest prices are found close to the center of the city, which is located near the seaport. There is also a spatial drift determined by a plane that goes from the coast (with high values) to the city outskirts (with lower values). It is also observed that there are points of interest such as universities, supermarkets or shopping centers, both in the downtown area, where values are high and, in the outskirts, where values are low.

(Figure 1) shows the outline that summarizes the development of the work. First, a base model (Mod 0) was estimated in which regression was performed with the data from the original sample, with the distinction that the age variable was not found in all the dwellings. After that, Kriging was applied to estimate the missing data for age, and a new model (Mod 1) was obtained. The orthogonality method was then used to solve the multicollinearity problem, obtaining a new model (Mod 2). Finally, based on (Anselin et al., 1996), two spatial SEM models are estimated, one where heteroscedasticity is not controlled (Mod 3) and the other where this issue is controlled (Mod 4.) Therefore, the assumptions for the validity of the model are fulfilled.

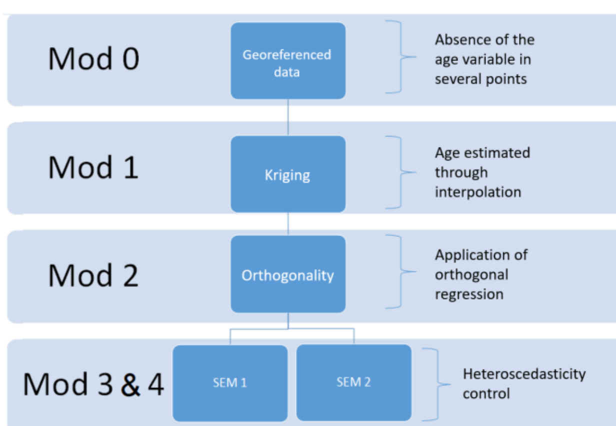

***Figure 1.** Outline of the research process*

DOI: 10.7764/RIC.00044.21

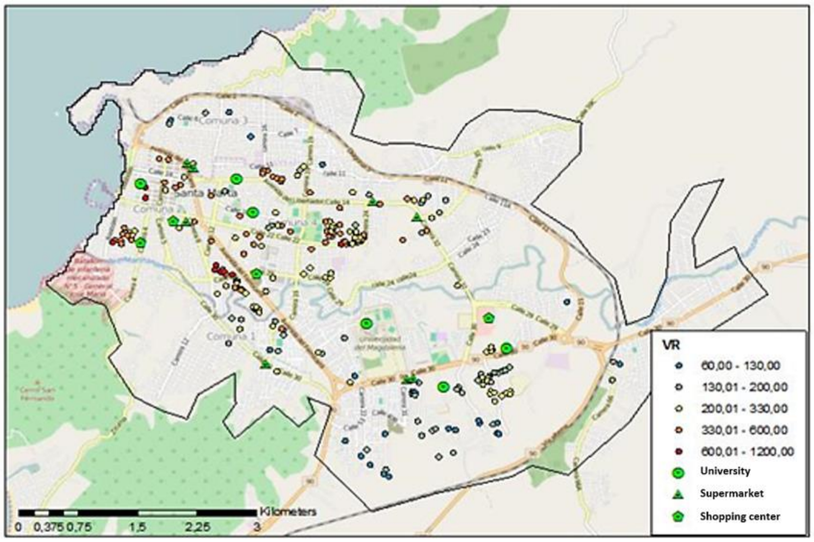

Figure 2*. Spatial distribution of housing prices and some sites of interest in the city of Santa Marta*

To overcome the problem of the lack of information on age, the Kriging method was used, which consists of an estimate through interpolation of the weighted average value that requires experimental and structural information (Chica-Olmo, 1994). That is to say, a statistical prediction of a value for the unsampled sites is obtained from the available data (Felus, 2001). Therefore, this method will estimate the missing data for the age variable. (Figure 3) shows the isovalue map of age obtained through Kriging.

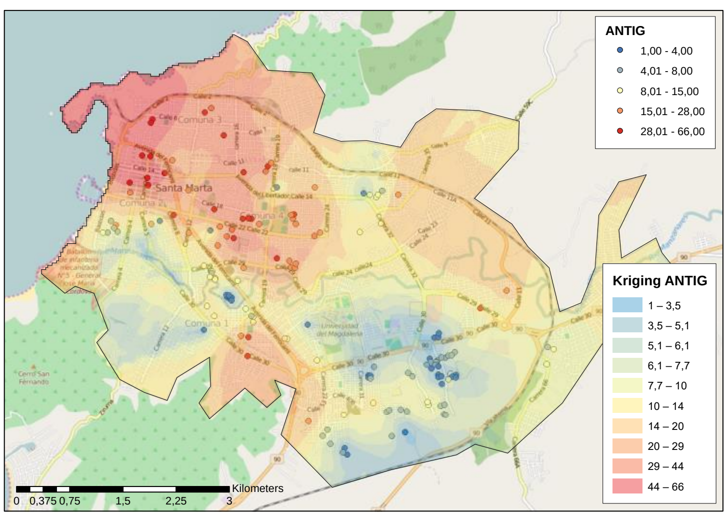

Figure 3*. Spatial distribution of the age variable of dwellings in Santa Marta and their estimates obtained by kriging*

(Table 1) shows the descriptive statistics of the construction and locational features. It should be noted that the variable ANTIG represents the age of sampled dwellings and the estimate by Kriging of the dwellings without available values. From the results of this table, it can be summarized that 87% of the properties have a garage, 15% have access to a swimming pool, and, approximately, on average, have 3 bedrooms, 3 bathrooms, 156 m^2 and are 16 years old. It is noteworthy that the average number of bathrooms is 3, equal to the average number of bedrooms. It is also striking that there is one property with 6 bathrooms, as it is a luxurious property of about 615 m^2 in an exclusive neighborhood of the city.

DOI: 10.7764/RIC.00044.21

Table 1. *Descriptive statistics*

Variable	Mean	St. Dev.	Var. Coef.	Min	Max
VR	287.737	167.544	0.609	60	1200
GARAJE	0.872	0.335	0.410	0	1
HABITAC	3.242	0.739	0.277	2	6
BANOS	2.938	0.952	0.350	1	6
M2	156.545	93.116	0.672	50	615
PISCINA	0.152	0.360	2.536	0	1
DUNIV	0.858	0.399	0.465	0.160	1.923
DSUPERMDO	0.789	0.386	0.489	0.063	2.601
DCTROCCIAL	1.024	0.543	0.530	0.110	2.465
COORDXM	79.475	1,428.291	17.972	-2,707.738	3,672.262
COORDYM	-55.599	1,160.799	-20.878	-2,557.409	2,296.591
ANTIG	16.380	12.389	0.756	1	66

2. Results

This section presents the main results obtained from the methodology proposed in this study. First, we show the results of the model estimated by OLS, whose endogenous variable is the logarithm of housing prices and whose explanatory variables are the construction and locational features of the dwellings (Mod0-OLS, Table 2). To estimate this model, the 142 dwellings with all the information available were used. This model shows that all the construction variables are significant except ANTIG and PISCINA. As for the location variables, no variable related to the distance to points of interest (distance to the university, supermarkets or shopping centers) is significant, although the variables that include the distance to the city center and the spatial trend of prices, collected by the geographic coordinates are significant.

To solve the problem of missing data corresponding to the ANTIG variable, the Kriging method has been applied, and the missing values of this variable (69 dwellings) have been replaced by the estimates obtained using this method. Thus, a sample of 211 data was available. From these new data, Mod1-MCO was estimated (Table 2). In this model, the significance of the ANTIG variable and the quality of the model improved. Regarding the verification of the hypotheses of the regression model, serious multicollinearity problems were detected with the DISTC variable, as it presents an IVF (variance inflator factor) of 159, which is much higher than the maximum recommended value of 10 (Kennedy, 1992), with 116,111 being the number of condition. To solve this multicollinearity problem, (García et al., 2020) propose the use of orthogonal regression, estimating the following auxiliary model:

$$DISTC_i = \alpha_1 + \alpha_2 COORDXM_i + \alpha_3 COORDYM_i + u_i$$

This methodology to solve the multicollinearity problem has been applied in models 2, 3 and 4. The residuals of the auxiliary model collect the information provided by DISTC that COORDXM and CORDYM do not explain. Model 2 (Mod2-Ort, (Table 2)) includes these residuals as an explanatory variable, which has been called DISTCRES, representing the effect of the distance to the city center, which is not determined by the geographic coordinates of the dwellings. This model no longer presents serious multicollinearity problems since the highest IVF of this model is 3.124, which corresponds to the DCTROCCIAL variable, with a condition number of 23.565. For this reason, the orthogonalized variable DISTCRES will be maintained in subsequent models. Furthermore, this model complies with the normality assumption since the Jarque-Bera test statistic (which asymptotically follows a chi-square distribution with 2 degrees of freedom) takes the value of 1.093, corresponding to a p-value of 0.579. However, this model presents problems of spatial autocorrelation in the disturbances (Moran's I of the errors = 0.246, p-value < 0.000), and of heteroscedasticity (Breusch-Pagan test = 57.938, p-value < 0.000).

We followed the criteria proposed by (Anselin et al., 1996) to determine whether a SAR or SEM model is more appropriate. The LMlag, RLMlag, LM-Lerr and RLMerr statistics were obtained for the analysis of the presence of spatial

ENGLISH VERSION...

autocorrelation. The results of the LM tests (Table 3) show the spatial dependence in terms of the error, so the SEM specification was chosen. The results of the SEM model estimates are shown in (Table 2) (Mod3-SEM1). However, the issue of heteroscedasticity persists in this model (Breusch-Pagan test = 59.351, p-value < 0.000). In this situation, (Kelejian and Purcha, 1998) propose applying two-stage least squares to control heteroskedasticity, alternating the generalized moments method and the instrumental variables method.

(Table 2) shows the estimates of the SEM model controlling for the heteroscedasticity issue. The results of this model indicate that all the structural variables are significant at 95%, except for the PISCINA variable. It is also observed that the location variables of distances to different points of interest (such as universities, supermarkets or shopping centers) are not significant. This is an expected result, considering that these types of points of interest are distributed throughout the city, so they are located near expensive housing (located in the center of the city) and near more affordable housing (located in areas farther away from the center) see (Figure 2). On the other hand, variables that account for the effect of large-scale price variation, such as the coordinates of the dwellings or distance to the city center, are significant. This allows us to contrast the effect noted above, that there is a monotonic decreasing effect in housing prices from the coast to the outskirts. It is also observed that the coefficient λ is significantly different from zero, which indicates that housing prices depend not only on the variables specified in this model but also on other location variables not included.

***Table 2**. Mod0-moc: model 0 before kriging. Mod1-mco: model 1 calculated by ols after kriging. Mod2-ort: orthogonal regression model. Mod3-sem1: sem model. Mod4-sem2: sem model controlling for heteroscedasticity*

	Mod0-MCO	Mod1-MCO	Mod2-Ort	Mod3-SEM1	Mod4-SEM2
Constant	6.4195***	6.259***	4.565***	4.561***	4.598***
Construction. Variable					
GARAJE	0.1745**	0.214***	0.214***	0.210 ***	0.221***
HABITAC	0.0876**	0.076***	0.076***	0.080***	0.067**
BANOS	0.1386***	0.125***	0.125***	0.115***	0.122***
M2	0.0020***	0.002***	0.002***	0.002***	0.002***
PISCINA	0.0562	0.049	0.049	0.064	0.067
ANTIG	−0.0037	-0.004 **	-0.004**	-0.004*	-0.005**
Location Variable					
DUNIV	0.0000	-0.062	-0.062	-0.070	-0.074
DSUPERMDO	0.0000	0.004	0.004	0.007	0.009
DCTROCCIAL	−0.0001	-0.098*	-0.098*	-0.054	-0.064
COORDXM	0.00045***	0.00045***	-0.00004*	-0.00005**	-0.00004*
COORDYM	−0.00014*	-0.00014*	0.00010***	0.00010***	0.00012***
DISTC	−0.0006***	-0.00058***			
DISTCRES			-0.00058***	-0.00072***	-0.00064***
λ				0.4250***	0.2150**
R Square	0.760	0.819	0.819	0.838	0.840
N	142	211	211	211	211

Note: "" not significant, "" significant at 10%, "**" significant at 5%, "***" significant at 1%*

Table 3. Lm statistics

Statistics	
LMerr	24.634 (0.000)

DOI: 10.7764/RIC.00044.21

RLMerr	6.774 (0.009)
LMlag	20.182 (0.000)
RLMlag	2.322 (0.128)

Note: p-values between brackets

3. Conclusions

It is important to mention that although the starting point was a sample with incomplete information due to the lack of information on the age of 69 dwellings, it was possible to extend the sample and improve the quality of the model and the significance of the age variable (H1) thanks to the spatial estimate of the age of the dwelling, using the Kriging method. In addition, this work has included location variables. Among these, the distance to the center and the spatial trend were highly significant (H2).

It was also found that some location factors, such as distance to certain points of interest, were not significant in this case (H3), probably because these points of interest are located either in areas with higher prices or with lower prices.

In addition, the problem of heteroscedasticity was controlled by applying the generalized moments method and instrumental variables. All structural variables were also significant at 95%, except PISCINA. It was also found that the Mod4-SEM2 model obtained better results than the conventional hedonic price model. This model showed advantages because it considers the interaction between the characteristics of the dwelling and its location. Therefore, according to the results, when the data obey a heterogeneous spatial structure, spatial econometric models are preferable, which have been fundamental in the study of housing price estimates within the hedonic price approach.

The results obtained in this study may be relevant for the real estate or construction sector because these results provide information on the most influential attributes of housing prices. These include, for example, construction factors such as surface area, the existence of a garage, number of bathrooms, age, and number of rooms or location factors such as distance to the city center or the decreasing spatial trend of prices from the coast to the outskirts of the city. The study also shows that incorporating missing information improves the significance of some variables in the model, for example, the age of the dwellings. The results obtained can serve as a reference for the city appraisers when assigning a price to a property. These results can also be useful for developing social real estate projects related to the District Mayor's Office.

As future lines of work, we propose the geostatistical modeling of housing prices considering the above variables and complementing it with the addition of variables with neighborhood characteristics, such as proximity to parks, recreational areas and schools to determine whether they would be significant and to what extent they would improve the level of significance of the model. Therefore, it can be considered that housing prices not only depend on the variables specified in this model but also on other location variables that are not included in the model, such as places of interest in the city, which could have an impact on the price.

4. References

Anselin, L. (1988). Spatial Econometrics: Methods and Models. Dordrecht: Kluwer Academic Publishers.

Anselin, L. (1992). Space and applied econometrics. *Regional Science and Urban Economics. 22*, 307-316. https://doi.org/10.1016/0166-0462(92)90031-U

Anselin, L., Sridharan, S., y Gholston, S. (2006). Using exploratory spatial data analysis to leverage social indicator databases: the discovery of interesting patterns. *Social Indicators Research. 82*, 287-309. https://doi.org/10.1007/s11205-006-9034-x

Anselin, L. y Rey, S. (2012). Spatial econometrics in an age of CyberGIScience. *International Journal of Geographical Information Science, 26*(12), 2211-2226. https://doi.org/10.1080/13658816.2012.664276

Anselin, L., Bera, A. K., Florax, R., y 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

Arraiz, I., Drukker, M.D., Kelejian, H.H. y Prucha, I.R. (2010). A spatial Cliff-Ord-type Model with Heteroskedastic Innovations: Small and Large Sample Results. *Journal of Regional Science, 50(2)*, 592-614. https://doi.org/10.1111/j.1467-9787.2009.00618.x

Basu, S. y Thibodeau T. (1998) Analysis of Spatial Autocorrelation in House Prices. *Journal of Real State Finance and* **Economics, 17(1) 61-85.**

Baronoi, A., Bianco, A., y Rabanal, C. (2018). Una introducción a la econometría espacial. Recuperado de: http://www.econometricos.com.ar/wp-content/uploads/2012/11/Espacial.pdf

Bowes, D., y Ihlanfeldt, K., (2001). Identifying the impacts of rail transit stations on residential property values. *Journal of Urban Economics 50*, 1–25. https://doi.org/10.1006/juec.2001.2214

Can, A. (1990). The measurement of neighborhood dynamics in urban house prices. *Economic Geography*, *66*(3), 254-272. https://doi.org/10.2307/143400

Can, A. (1992). Specification and estimation of hedonic housing price models. *Regional Science and Urban Economics, 22*, 453-474. https://doi.org/10.1016/0166-0462(92)90039-4

Can, A. (1998). GIS and Spatial Analysis of Housing and Mortgage Markets. *Journal of Housing Research, 9(1)*, 61-86. https://doi.org/10.1080/10835547.1998.12091927

Case, B., Clapp, J., Dubin, R., y Rodriguez, M. (2004). Modeling Spatial and Temporal House Price Patterns: a comparison of four models. *Journal of Real State, Finance and Economics,* 29(2), 167-191. https://doi.org/10.1023/B:REAL.0000035309.60607.53

Chay, K., y Greenstone, M. (2005). Does Air Quality Matter? Evidence from the Housing Market. *Journal of Political Economy, 113*(2), 376-424. https://doi.org/10.1086/427462

Chica-Olmo, J. (1994). Teoría de las variables regionalizadas. Una aplicación economica espacial y valoración inmobiliaria. Granada: servicio de publicaciones de la Universidad de Granada.

Chica-Olmo, J. (1995). Spatial Estimation of Housing Prices and Locational Rents. *Urban Studies, 32*(8), 1331-1344. https://doi.org/10.1080/00420989550012492

Chica-Olmo, J. (2007). Prediction of Housing Location Price by a Multivariate Spatial Method: Cokriging. *The Journal of Real State Research, 29*(1), 91-114.

Chica-Olmo, J., Cano, R., y Tamaris, I., (2018) Determination of buffer zone for negative externalities: Effect on housing prices. *The Geographical Journal, 185(2)*, 222-236. https://doi.org/10.1111/geoj.12289

Chien, LC., Yu, HL., y Schootman, M., (2013). Efficient mapping and geographic disparities in breast cancer mortality at the county-level by race and age in the U.S. *Spatial and Spatio-temporal Epidemiology, 5*, 27-37. http://dx.doi.org/10.1016/j.sste.2013.03.002

Chinloy, P.T. (1977). Hedonic Price and Depreciation Indexes for Residential Housing: A Longitudinal Approach. *Journal of Urban Economics, 4*, 469-482. https://doi.org/10.1016/0094-1190(77)90007-9

Clapp, J., (2004). A Semiparametric Method for Estimating Local House Price Indices. *Real State Economics, 32(1)*, 127-160. https://doi.org/10.1111/j.1080-8620.2004.00086.x

Clif, A. y Ord, J. (1973). Spatial autocorrelation. London: Pion. https://doi.org/10.1177/030913259501900205

Clif, A. y Ord, J. (1981). Spatial process-models and aplications. London: Pion.

Conroy, S.J., y Milosh, J.L. (2011). An Estimation of the Coastal Premium for Residential Housing Prices in San Diego County. *Journal of Real Estate Finance and Economics, 44*(2), 211-228. https://doi.org/10.1007/s11146-009-9195-x

Corrado, L y Fingleton, B. (2011). Where is the economics in spatial econometrics? *Journal of Regional Science, 52*(2), 210-239. https://doi.org/10.1111/j.1467-9787.2011.00726.x

Cruces, G., Ham, A., y Tetaz, M. (2008). *Quality of Life in Buenos Aires Neighborhoods: Hedonic Price Regressions and the Life Satisfaction Approach*. Research network working paper # R-559. Recuperado de https://publications.iadb.org/en/publication/10739/quality-life-buenos-aires-neighborhoods-hedonic-price-regressions-and-life

D' Elia, V., Conte, M., y León, S. (2020). Bus rapid transit and property values in Buenos Aires: Combined spatial hedonic pricing and propensity score techniques. *Research in Transportation Economics, 80.* https://doi.org/10.1016/j.retrec.2020.100814.

Dubin, R. (1992). Spatial autocorrelation and neighborhood quality. *Regional Science and Urban Economics, 22*, 433-452. https://doi.org/10.1016/0166-0462(92)90038-3

Felus, Y. (2001). New methods for spatial statistics in geograpchic information systems. *Degree of Doctor of philosohy*. The ohio university state, Estados Unidos.

Fierro, K.P., Fullerton, T. M., y Donjuan-Callejo, K.E. (2009). Housing Attribute Preferences in a Northern Mexico Metropolitan Economy. *Atlantic Economic Journal, 37(2)*, 159-172. https://doi.org/10.1007/s11293-009-9174-x

García, C. B., Salmerón, R., García, C., y García, J. (2020). Residualization: Justification, properties and application. *Journal of Applied Statistics, 47(11)*, 1990-2010. https://doi.org/10.1080/02664763.2019.1701638

Garza, N., y Ovalle M.C. (2019). Tourism and housing prices in Santa Marta, Colombia: *Spatial determinants and interactions. Habitat International, 87*, 36-43. https://doi.org/10.1016/j.habitatint.2019.04.001

Geng, B., Bao, H., y Liang, Y. (2015) A study of the effect of a high-speed rail station on spatial variations in housing price based on the hedonic model. *Habitat International, (49)*, 333-339. http://dx.doi.org/10.1016/j.habitatint.2015.06.005

Goodman, A.C. (1978). Hedonic Prices, Price Indices and Housing Markets. Journal of Urban Economics, *5*, 471-484. https://doi.org/10.1016/0094-1190(78)90004-9

Goodman, A.C. y Thibodeau, T.G., (2003). Housing market segmentation and hedonic prediction accuracy. *Journal of Housing Economics. 12*, 181-201. https://doi.org/10.1016/S1051-1377(03)00031-7

Gröbel, S. y Thomschke, L. (2018). Hedonic pricing and the spatial structure of housing data – an application to Berlin. *Journal of Property Research, 35*(3), 185-208. https://doi.org/10.1080/09599916.2018.1510428

Kelejian, H.H., y Prucha, I.R. (1998). A Generalized Spatial Two Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances, *Journal of Real Estate Finance and Economics*, 17, 99–121. https://doi.org/10.1023/A:1007707430416

Kelejian, H.H. y Prucha, I.R. (1999). A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model, *International Economic Review*, 40, 509–533. https://doi.org/10.1111/1468-2354.00027

Kelejian, H. H., y Prucha, I. R., (2010). Spatial models with spatially lagged dependent variables and incomplete data. *Journal of Geographical Systems, 12(3)*, 241-257. https://doi.org/10.1007/s10109-010-0109-5

Kennedy, P. (1992). A Guide to Econometrics. Cambridge: MIT Press.

Kilibarda, M., Heng, T., Heuvelink, G., Gräler, B., Pebesma, E., Tadic, M., y Bajat, B. (2014) Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. *Journal of Geophysical Research: Atmospheres, 119*, 2294-2313. https://doi.org/10.1002/2013JD020803

Lesage, J. P. y Pace R. K., (2004). Models for Spatially Dependent Missing Data. *Journal of Real State Finance and Economics, 29(2)*, 233-254. https://doi.org/10.1023/B:REAL.0000035312.82241.e4

Li, Y., He, L., y Jiao, J., (2012). 6th International Association for China Planning Conference, IACP. *A Spatial Analysis of Housing Prices in Chinese Coastal Cities, A Case Study of The City of Xiamen*, China. Wuhan, China.

Li, Y., He, L., Xu, W., Wang, H., y He, Z., (2013). Using GIS and Hedonic in the modelling of spatial variation of housing price in Xiamen city. *International review for spatial planning and sustainable development, 1*(2), 29-42. http://dx.doi.org/10.14246/irspsd.1. 4 29

Li, H., Wei, Y.D., Yu, Z., y Tian, G., (2016) Amenity, accessibility and housing values in metropolitan USA: A study of Salt Lake County, Utah. *Cities, 59*, 113-125. http://dx.doi.org/10.1016/j.cities.2016.07.001

Li, R., y Li, H. (2018). Have Housing Prices Gone with the SmellyWind? Big Data Analysis on Landfill in Hong Kong. *Sustainability, 10(2)*, 34. https://doi.org/10.3390/su10020341

Li, H., Wei, Y.D., Wu, Y., y Tian, G., (2019). Analyzing housing prices in Shanghai with open data: Amenity, accessibility and urban structure. *Cities, 91*, 165-179. https://doi.org/10.1016/j.cities.2018.11.016

Mark J.H. (1980). A Preference Approach to Measuring the Impact of Environmental Externalities. *Land Economics, 56*(1), 103-116. https://doi.org/10.2307/3145833

Montero-Lorenzo, J. M., Larras-Iribas, B., y Páez, A., (2009). Estimating commercial property prices: an application of cokriging with housing prices as ancillary information. *Journal of Geographical Systems, 11(4)*, 407-425. https://doi.org/10.1007/s10109-009-0095-7

Morais, M., y Cruz, B. (2015). Demand for housing and urban services in brazil: a hedonic approach. Institute for Applied Economic Research.

Muñoz-Raskin, R. (2010). Walking Accessibility to bus rapid transit: Does it affect property values? The case of Bogota, Colombia. *Policy Transport, 17*, 72-84. https://doi.org/10.1016/j.tranpol.2009.11.002

Nelson, J. P. (1978). Residential Choice, Hedonic Prices, and the Demand for Urban Air Quality. *Journal of Urban Economics, 5*, 357-369. https://doi.org/10.1016/0094-1190(78)90016-5

Pace, R. K., Barry, R., y Sirmans, C. F. (1998). Spatial Statistics and Real Estate. *Journal of Real Estate Finance and Economics, 17*(1), 5-13. https://doi.org/10.1023/B:REAL.0000035307.99686.fb

Poeta, S., Gerhardt, T., y Stumpf Gonzalez, M., (2019). Hedonic price analysis of single-family housing. *Revista Ingeniería de Construcción RIC, 34(2)*, 215-220. http://dx.doi.org/10.4067/S0718-50732019000200215

Rodríguez, D., y Targa, F. (2004). Value of Accessibility to Bogota's Bus Rapid Transit System. *Transport Reviews, 24(5)*, 587-610. https://doi.org/10.1080/0144164042000195081

ENGLISH VERSION..........

Rehák, Š., y Káčer, M. (2019). Estimating price gradient in Bratislava with different distance measurements. *Journal of European Real Estate Research*, *12(2)*, 190-206. https://doi.org/10.1108/JERER-10-2018-0045

Rosen, S. (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *The Journal of Political Economy*, *2*(1), 34-55.

Sagner, A. (2011). Determinantes del precio de viviendas en la región metropolitana de Chile. *Trimestre económico. 78(312)*, 813-839.

Sah, V., Conroy, S., y Narwold, A., (2015). Estimating School Proximity Effects on Housing Prices: the Importance of Robust Spatial Controls in Hedonic Estimations. *The Journal of Real State Finance and Economics*, *53(1)*, 50-76. https://doi.org/10.1007/s11146-015-9520-5

Simlai, P. (2014). Estimation of variance of housing prices using spatial conditional heteroskedasticity (SARCH) model with an application to Boston housing price data. *The Quarterly Review of Economics and Finance*, *54(1)*, 17-30. http://dx.doi.org/10.1016/j.qref.2013.07.001

Tiwari, P. y Parikh, J. (1998). Affordability, housing demand and housing policy in urban India. *Urban Studies*, *35*(11), 2111-2119. https://doi.org/10.1080/0042098984033

Tobler, W.R. (1970). A computer model simulation of urban growth in the Detroit region. *Economic Geography 46(2)*, 234–240.

Trojanek, R., y Gluszak, M. (2018). Spatial and time effect of subway on property prices. *Journal of Housing and the Built Environment*, *33(2)*, 359-384. https://doi.org/10.1007/s10901-017-9569-y

Urrea, M., Cardenas, N., y García G. (2019). *Un modelo espacial de precios hedónicos y calidad de vecindario: una aplicación para Medellín, Colombia* (tesis de pregrado) Universidad Eafit, Colombia.

Von Graevenitz, K. y Panduro, T.E. (2015). An Alternative to the Standard Spatial Econometric Approaches in Hedonic House Price Models. *Land Economics*, *91*(2), 386-409. https://doi.org/10.3368/le.91.2.386

Wen, H. Jin, Y. y Zang L. (2017). Spatial heterogeneity in implicit housing prices: evidence from Hangzhou, *China. International Journal of Strategic Property*, *21*(1), 15-18. http://dx.doi.org/10.3846/1648715X.2016.1247021

Xiao, Y., Chen, X., Li, Q., Yu, X., Chen, J., y Guo, J., (2017), Exploring Determinants of Housing Prices in Beijing: An Enhanced Hedonic Regression with Open Access POI Data. *International Journal of Geo-information. 6(11)*, 358. https://doi.org/10.3390/ijgi6110358

Yoo, E.-H. y, Kyriakidis, P.C. (2009). Area-to-point Kriging in spatial hedonic pricing models. *Journal of Geographical Systems*, *11*, 381-406. https://doi.org/10.1007/s10109-009-0090-z

Zambrano-Monserrate, M, (2015). Formación de los precios de alquiler de viviendas en Machala (Ecuador): análisis mediante el método de precios hedónicos. *Cuadernos de economía, 39*, 12-22. http://dx.doi.org/10.1016/j.cesjef.2015.10.002

Zhang, Y., Zheng, S., y Sun, C. (2014). Proceedings of the 17th International Symposium on Advancement of Construction Management and Real Estate, *What's the Effect of Urban Villages on Commercial Housing Price? An Analysis Based on Second-Hand Housing Transactions in Beijing*. Shenzhen, China.

Zhang, L., Zhou, J., Hui, E. C. M. y Wen, H. (2019). The effects of a shopping mall on housing prices: A case study in Hangzhou. *International Journal of Strategic Property Managemen*, *23*(1), 65-80. https://doi.org/10.3846/ijspm.2019.6360

Zhu, B. Füss, R y Rottke, N. B. (2011). The Predictive Power of Anisotropic Spatial Correlation Modeling in Housing Prices. *Journal of Real Estate Finance and Economics, 42(4)*, 542-565. https://doi.org/10.1007/s11146-009-9209-8