Articles

Estimation of housing prices in Ecuador using hedonic and geostatistical models: A comparative analysis.

Estimación de precios de vivienda en Ecuador mediante modelos hedónicos y geoestadísticos: Un análisis comparative

Livino M. Armijos-Toro[*]
Sergio Castillo-Páez[*]
Fernando Ortega[*]

ABSTRACT

This paper compares the results of two models, one hedonic and the other geostatistical, when obtaining housing price estimates in the Rumiñahui canton, Ecuador. In the first model, different parameterizations of the variables that make up the hedonic model are used to obtain the best predictions. For the geostatistical case, the predictions are composed of a trend function that depends on certain housing characteristics and a spatial error term, which is modeled from a residual variogram. The performance of each model is compared by an analysis of prediction errors from a validation data set. The results indicate a better performance for the geostatistical model, since it considers, in addition to certain inherent characteristics of each house, the effect of its spatial location on the selling price.

Keywords: Hedonic regression, Variogram, Kriging methods, Housing prices.

[*] Master's Degree, Universidad de las Fuerzas Armadas ESPE, Department of Exact Sciences, Riobamba, Ecuador. lmarmijos2@espe.edu.ec. https://orcid.org/0000-0001-8553-536X.
[*] Master's Degree, Universidad de las Fuerzas Armadas ESPE, Department of Exact Sciences, Ecuador. sacastillo@espe.edu.ec. https://orcid.org/0000-0002-5402-2462.
[*] Ph.D, Universidade de Vigo, Vigo, Spain. pgarcia@uvigo.es. https://orcid.org/0000-0003-4542-6630
[*] Master's Degree, Universidad Técnica del Norte, School of Engineering in Applied Sciences, Riobamba, Ecuador, fwortega@utn.edu.ec. https://orcid.org/0000-0002-1545-4182

JOURNAL OF BUSINESS and entrepreneurial studies
ISSN: 2576-0971

Journal of Business and entrepreneurial
January - March Vol. 6 - - 2202 2
http://journalbusinesses.com/index.php/revista
e-ISSN: 2576-0971
journalbusinessentrepreneurial@gmail.com
Receipt: 01 June 2021
Approval: 28 January 2022
Page 57-72

Journal of Business and entrepreneurial studies
ISSN: 2576-0971
http://journalbusinesses.com/index.php/revista

RESUMEN
En el presente trabajo se comparan los resultados de dos modelos, uno hedónico y otro geoestadístico, al momento de obtener estimaciones de precios de vivienda en el cantón Rumiñahui, Ecuador. En el primer modelo, se recurren a distintas parametrizaciones de las variables que conforman el modelo hedónico para obtener las mejores predicciones. Para el caso geoestadístico, las predicciones se componen de una función tendencia que depende de ciertas características de la vivienda y un termino de error espacial, el cual es modelado a partir de un variograma residual. El comportamiento de cada modelo es comparado mediante un análisis de errores de predicción a partir de un conjunto de datos de validación. Los resultados indican un mejor rendimiento para el modelo geoestadístico, pues este considera además de ciertas características inherentes a cada vivienda, el efecto de su ubicación espacial sobre el precio de venta.
Palabras clave: Regresión hedónica, Variograma, Métodos kriging, Precios de vivienda.

INTRODUCTION
33.7% of the population in Ecuador has a qualitative housing deficit (INEC, 2018). The Organic Code of Territorial Organization determines the guidelines to establish the appraisals of assets to the decentralized autonomous governments. However, the appraisal of these assets does not determine their commercial value (sale price), which is determined by supply and demand (Rincón, M. & Campo, J., 2016).
The estimation of the price of the good by means of models is of vital importance. Thus, hedonic regression models are the most common alternatives when estimating the price of housing based on its characteristics and location (Griliches, 1971). In general, hedonic regression models aim to determine the exogenous variables that affect the price and its estimation (Bover & Velilla, 2001).
The exogenous variables that affect price are those associated with the characteristics of housing, its surroundings, proximity to areas of influence, as well as the economic variables of the sector where the housing is located. These characteristics are considered to be: micro-local, macro-local and general (Derycke, 1983).
However, the price of a particular house can also be affected by the price of neighboring houses. In this case the effect of spatial dependence can be analyzed using geostatistical models. The use of spatial kriging prediction methods to obtain housing price estimates has already been widely used, for example in Martínez, Lorenzo and Rubio (2000), Chica (1995), among others. Usually, these models include both information on certain housing characteristics as well as their spatial location to obtain housing price predictions.
In this article, we intend to compare the estimates of housing prices in Cantón Rumiñahui in Ecuador obtained from these two types of models. The following section presents the theoretical hedonic and geostatistical models considered in both cases. Then, in section 3, the modeling and validation data sets used are presented, as well as the estimates obtained for both models, followed by a statistical comparison of the hedonic and geostatistical prediction errors. Finally, the main conclusions and recommendations of the present study are given.

Journal of Business and entrepreneurial studies
ISSN: 2576-0971
http://journalbusinesses.com/index.php/revista

MATERIALS AND METHODS

In this paper, variables were compiled to explain the price per square meter of a house. Montero (2004) presents the methodology used for local government appraisals for the average price per square meter by municipality, community and nationally as a measure that determines the price according to the location of the property. Chica-Olmo et al. (2007) propose the price of housing as an analysis variable and certain construction characteristics of housing and dichotomous variables such as the year in which the measure was taken as explanatory variables, with which they intend to collect the inter-annual variations in prices. Another type of variable that is expected to explain housing prices are macro-local variables. Thus, the following type of variables are expected to be collected for the present work:

- Housing characteristics;
- Location; and,
- Location (proximity) to sectors of interest.

Chica-Olmo et al. (2007) propose hedonic models as a function that seeks to explain an analysis variable (in this case the variable Pxm2T) as a function of a vector of variables (attributes or characteristics) that conceptually influence the analysis variable. In general terms, the model can be expressed as follows:

$$Pxm2T_H = f(\mathbf{I},\mathbf{V},\mathbf{U},\mathbf{Z}) + \varepsilon \,, \qquad [1]$$

where the variable $Pxm2T_H$ is the price of the total square meter of a house (obtained from the proposed hedonic model), which is explained by the arguments described in the function f and a which ε is an independent random variable following a normal distribution with zero mean and variance σ^2 .

In this particular case, vector **I** describes the variables that describe the characteristics of the real estate, such as: land area, construction area, number of bedrooms, number of bathrooms, among others. **V** collects the factors associated with the neighborhood in which the property is located, for example: socioeconomic level and safety. **Z** are the characteristics within the regulatory plan, such as: population, social and cultural growth, which influence the price of housing (Figueroa & Lever, 1992). On the other hand, vector **U** explains whether the property is located near areas of economic influence.

According to Lever (2000) the relationship between the variable under analysis and the explanatory variables are not always linear and usually respond to a logarithmic form. The hedonic regression model must meet the assumptions of error normality, constant variance and error independence. The hedonic model [1] will be used in the present work, using a linear regression model given by:

$$Pxm2T_H = \beta_0 + \rho I + \delta V + \alpha U + \theta Z + \varepsilon \,, \qquad [2]$$

where β_0 is the intercept of the linear model and ρ is the vector of coefficients related to housing characteristics, δ represents the coefficients that explain the relationship of the price of the total square meter with respect to the neighborhood factors. As well

Journal of Business and entrepreneurial studies
ISSN: 2576-0971
http://journalbusinesses.com/index.php/revista

as ∝ is the vector of coefficients related to the proximity of the property to areas of economic influence. Y, θ the coefficients that explain the linear relationship of the price per square meter of housing with respect to the characteristics of the location of the property.

The main assumption in a geostatistical model is that the variable of analysis (in this case, the total price per square meter obtained through this model, $Pxm2T_G$) depends in some way on the special location of the house. For our case, we will express this model as follows:

$$Pxm2T_G = \mu(\boldsymbol{X}) + \varepsilon(\boldsymbol{s}), \quad [3]$$

where $\mu(\boldsymbol{X})$ is a trend function that is related to certain explanatory variables $\boldsymbol{X}$, and$\varepsilon(\boldsymbol{s})$ is an error term with zero mean that depends on the spatial position $\boldsymbol{s}$, whose covariance is given by the function:

$$\text{Cov}[\varepsilon(\boldsymbol{s}),\varepsilon(\boldsymbol{s}+\boldsymbol{u})] = -\sigma^2\gamma(\boldsymbol{u})$$

being the σ^2variance of the errors, and$\gamma(\boldsymbol{u})$ is called the (semi)variogram, which depends only on the distance $\boldsymbol{u}$ between two spatial positions, and is defined by:

$$\gamma(\boldsymbol{u}) = \frac{1}{2}\text{Var}[\varepsilon(\boldsymbol{s}) - \varepsilon(\boldsymbol{s}+\boldsymbol{u})].$$

In the model [3], the trend function captures the effect of certain variables related to housing on the price of housing, which can be estimated by:

$$\hat{\mu}(\boldsymbol{X}) = \boldsymbol{X}\hat{\beta}, \quad [4]$$

where $\hat{\beta}$ the coefficient vector of a linear regression model between $Pxm2T$ and the independent variables $\boldsymbol{X}$.

Likewise, the effect of spatial dependence on the dependent variable is represented through $\varepsilon(\boldsymbol{s})$, specifically through its variogram. The estimation of $\gamma(\boldsymbol{u})$ is done through a process called "Structural Analysis", which consists of the following steps (see e.g. Cressie, 1993, Section 2.6):

1. Estimate the errors from the residuals given by: $\boldsymbol{r}(\boldsymbol{s}) = Pxm2T_G - \hat{\mu}(\boldsymbol{X})$,
2. Obtain a pilot variogram from these residues.
3. Fit a valid variogram model (u) to the pilot variogram. $\hat{\gamma}(\boldsymbol{u})$ to the pilot variogram.

For step 1, it is first necessary to have constructed the linear regression and then to construct the pilot variogram from the corresponding residuals. The empirical or Matheron estimator (Matheron, 1962) is usually used to obtain this first variogram. Subsequently, a valid variogram model is chosen that best fits this pilot variogram, which depends on certain parameters, namely: "nugget effect", "threshold" and "range" of the variogram, represented by c_0, c_1, and a respectively. It is worth mentioning that, while the nugget effect indicates the variability of the errors near the origin (i.e. at very small distances), the threshold expresses the variability at large jumps. Also, under certain conditions, it is true that $c+_0\ \ c=_1 \sigma^2$. On the other hand, the range indicates the

maximum distance up to which the spatial dependence maintains its influence on the analysis variable.

Then, there are several valid variogram models, such as the Circular, Spherical, Exponential or Penta-spherical (these and other models can be seen in Chilès and Delfiner, 1999, section 2.5.1), where each one expresses the functional form of the variogram from the fitted values for c_0, c_1,and a. This fitting is usually done by weighted least squares, minimizing the squared errors between the pilot variogram and the chosen model.

Once the estimator (u) has been obtained, this variogram $\hat{\gamma}(\boldsymbol{u})$, this variogram can be used to construct spatial predictions of the error term at a specific spatial position $\boldsymbol{s_0}$, for which kriging interpolation methods are used (see e.g. Wackernagel, 1998, Chap. 11.). This method uses the spatial information of the variable of analysis, to obtain the simple kriging predictor $\hat{\varepsilon}$ ($\boldsymbol{s_0}$), which is expressed as a weighted sum:

$$\hat{\varepsilon}\ (\boldsymbol{s_0}) = \sum_{i=1}^{n} \lambda_i\ \mathbf{r}(\mathbf{s}),$$

where λ_i are known as the kriging weights and are computed from the fitted variogram $\hat{\gamma}(\boldsymbol{u})$, while $\mathbf{r}(\mathbf{s})$ are the residuals obtained at the n observed sample positions (in our case, $\boldsymbol{s}$ = (longitude, latitude) of the location of each house in the modeling dataset).

From this kriging prediction the house price estimate can be constructed from the model [3], i.e.:

$$\widehat{Pxm2T_G} = \hat{\mu}(\boldsymbol{X}) + \hat{\varepsilon}\ (\boldsymbol{s_0}) \qquad [5]$$

It is worth mentioning that kriging methods can also be used to compare the fits of different models, and thus select the best of them by statistically comparing their prediction errors, either by using a modeling dataset and a validation dataset, or by using cross-validation techniques. These approaches were used in the following section to obtain the valid variogram model, as well as to compare the estimates obtained with the geostatistical model with those corresponding to the hedonic model.

RESULTS

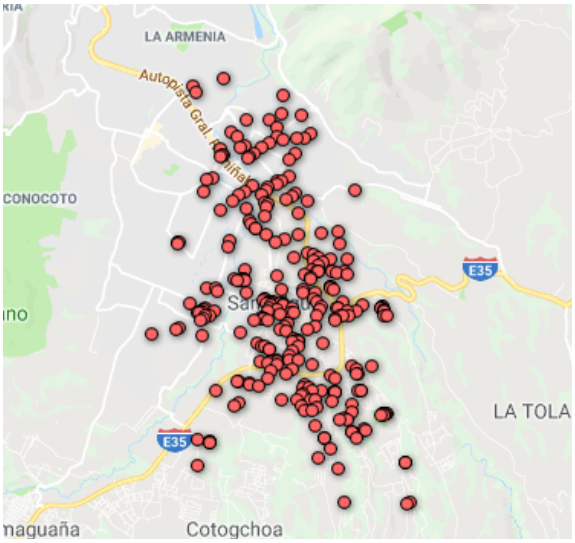
The data used in this paper correspond to information contained in some portals of available housing sales in Ecuador. The information was collected from advertisements made between January 1 and March 20, 2019 in the Rumiñahui canton, province of Pichincha, Ecuador.

From an initial total of 700 ads, first the duplicate records were purged. Then we proceeded to geolocate each ad, filtering those whose confirmed location corresponded to the Rumiñahui canton, and also considered the cases in which the information (characteristics of the property) is complete. Finally, a purified database of 296 records was obtained, with the following variables: Total area (*SupT*), Covered area (*SupCub*), Number of rooms (*NumCua*), Number of bathrooms (*NumBan*), Number of parking spaces (*NumEst*), Age in years (*Anti*), Location: Latitude and Longitude (*Lat*, *Log*), Sales Price (*P*), Number of Floors (*NumPis*), Condominium Membership (*Cond*).

Journal of Business and entrepreneurial studies
ISSN: 2576-0971
http://journalbusinesses.com/index.php/revista

The above variables collect the characteristics of the goods for sale, so then we proceeded to collect external information that may affect the price of housing. The geolocation of: educational units with high school, health facilities (public and private), economic zones (shopping malls, commercial areas) and higher education institutions (universities and institutes) were extracted within the Rumiñahui canton. Then, we proceeded to calculate the minimum distance of each home for sale to each of the types of locations described in this paragraph. Finally, the following variables were obtained for each dwelling: Minimum distance to an educational unit (*DminCE*), Minimum distance to an economic zone (*DminCEC*), Minimum distance to a health establishment (*DminCS*), Minimum distance to a higher education institution (*DminCES*), Minimum distance to a commercial zone (*DminCEC*), Minimum distance to a health establishment (*DminCS*) and Minimum distance to a higher education institution (*DminCES*).
Information was also collected on the location of the house, since the neighborhood where it is located can also influence its price. The neighborhood dummy variables considered are the following: Armenia (*ZArm*), Ilaló (*ZIla*), El Triángulo (*ZTria*), La ESPE (*ZEspe*), Rancho (*ZRanc*), River Mall (*ZRiver*), El Colibrí (*ZColi*), Sangolquí centro (*ZSang*), Panamericana sur (*ZPana*), Selva Alegre (*ZSelva*), Barrio Carlos Gavilánez (*ZGavil*).
Finally, based on the variable *P* and *SupT*, the variable that is the object of this study, total square meter price (*Pxm2T*), is calculated. Figure 1 shows the location of the dwellings that are the subject of this study.

Figure 1. Location of houses - Cantón Rumiñahui.



Of the total number of records obtained, the data set was divided into a modeling data set and a validation data set. The between set represents 90% (267 records) of the total data, which were randomly selected. The validation set (29 records) was used to compare the predictions of the hedonic and geospatial models. Based on the model [2]

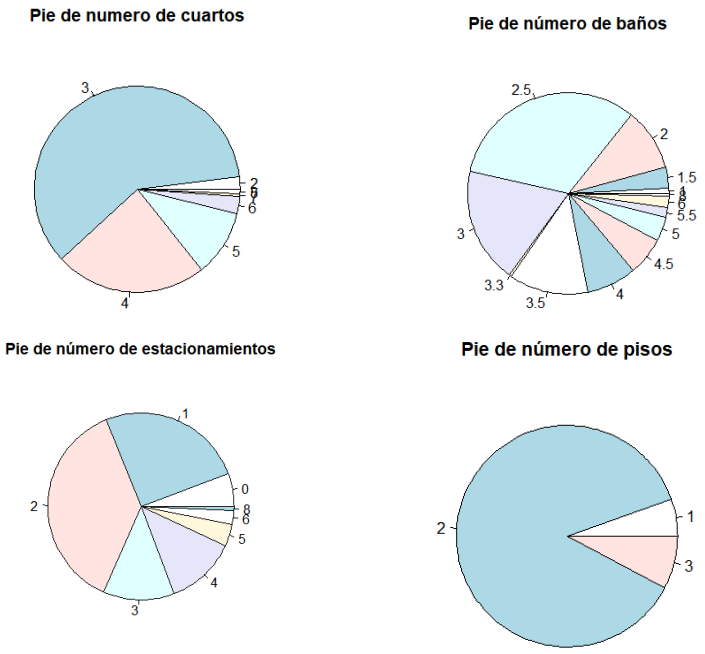
Journal of Business and entrepreneurial studies
ISSN: 2576-0971
http://journalbusinesses.com/index.php/revista

presented above, a linear regression model was estimated to explain the price of the total square meter of housing, considering the variables that describe the characteristics of the housing, the variables of minimum distance to zones of influence, as well as the variables of the location of each house (zones). The variables corresponding to socioeconomic factors of the location (*V*) are not considered in the linear model since the sector of analysis is only 139 km^2 (canton Rumiñahui), so there is no updated information on this type of variables at this level of disaggregation. Finally, the model used for the estimation is:

$$\widehat{Pxm2T}_H = \beta_0 + \rho_1 NumC + \rho_2 NumBan + \rho_3 NumEst + \rho_4 Anti + \rho_5 NumPis + \alpha_1 DminCE + \alpha_2 DminCS + \alpha_3 DminCES + \alpha_4 DminCEC + \theta_1 ZArm + \theta_2 ZIla + \theta_3 ZTria + \theta_4 ZEspe + \theta_5 ZRanc + \theta_6 ZRiver + \theta_7 ZColi + \theta_8 ZSang + \theta_9 ZPana + \theta_{10} ZSelva + \theta_{11} ZGavil + \varepsilon ,$$

After performing several regression models, it is necessary to transform the variables *NumCua*, *NumBan*, *NumEst* and *NumPis* into dichotomous variables in order to comply with the assumptions described in the previous section. For the transformation, it is considered that these categories are homogeneously distributed, as shown in Figure 2.

Figure 2. Pie charts of variables to be transformed.



Journal of Business and entrepreneurial studies
ISSN: 2576-0971
http://journalbusinesses.com/index.php/revista

Therefore, the proposed coding is shown in Table 1.

Table 1. *Proposed coding of variables*

Dummy variable	Assignment to dummy variable	
	1	**0**
DEst	NumEst >2	opposite case
Dpis	NumPis >1	opposite case
Dcua	NumCua >2	opposite case
Dban	NumBan >3	opposite case
Ant	Ant >0	opposite case

In addition to this, the minimum distance variables were standardized in order to prevent their dimension from influencing the model. With these transformations the hedonic regression models are made. This is presented below.

$$\widehat{Pxm2T}_H = \beta_0 + \rho_1 DCua + \rho_2 DBan + \rho_3 DEst + \rho_4 Ant + \rho_5 DPis + \alpha_1 DminCE + \alpha_2 DminCS + \alpha_3 DminCES + \alpha_4 DminCEC + \theta_1 ZArm + \theta_2 ZIla + \theta_3 ZTria + \theta_4 ZEspe + \theta_5 ZRanc + \theta_6 ZRiver + \theta_7 ZColi + \theta_8 ZSang + \theta_9 ZPana + \theta_{10} ZSelva + \theta_{11} ZGavil + \varepsilon ,$$

Then, several linear estimation models were performed, selecting the one that presented the best goodness of fit, as well as compliance with the assumptions of normality, homoscedasticity and correlation, Table 2 shows a summary of the four best models and the one selected.

Table 2. *Comparison of worked models*

Model Dear	Goodness of fit	Normality	Homocedasticity	Correlation
	R^2	Lilliefors	Bartlett	Durbin-Watson
MH1	0.2539	0.09467	0.33	0.662
MH2	0.2538	0.09725	0.38	**0.754**
MH3	0.3082	0.2056	**0.41**	0.484
MH4	**0.3078**	**0.1348**	**0.41**	0.41

Model MH1 considers as one of its influential variables the fact that the house is located in the Colibri area, which ends up negatively affecting the price per square meter. While for models MH2 and MH3 a transformation different from the one proposed in Table 1 was considered, but which in the end obtain the same

Journal of Business and entrepreneurial studies
ISSN: 2576-0971
http://journalbusinesses.com/index.php/revista

significant variables as those present in model MH4. Finally, the selected model is MH4, described in the following equation:

$$\widehat{Pxm2T}_{\mathrm{H}} = 326.721 + 76.06DCua - 71.973\ DEst - 42.333\ DBan - 56.346\ Ant\ + 162.124\ DPis\ + 70.695\ Cond - 29.888\ DMinCS\ + 8.985\ DMinCES\ + 254.642\ ZArm\ + 76.648\ ZIla\ + 111.181\ ZTria\ + 1.927\ ZEspe\ + 54.294\ ZPana\ - 64.677\ ZGavil$$

[6]

From the results obtained in equation [6] it is explained that the price of the total square meter of a house in the Rumiñahui canton. Regarding the variables related to the characteristics of the good (*I*), having more than two rooms is the variable that most explains the price of housing, while the fact that the house has more than two parking spaces and more than one floor, negatively affects the price, on the other hand, the fact that the house is not brand new has a positive influence on the price per square meter as well as the fact that the house is located within an urbanization. Regarding the variables that describe the proximity of the house to zones of influence (*U*), being close to a health center has a negative influence on the price, as does being close to a higher education center (university or high school). Finally, the zoning variables (*Z*), when the property is located in the areas of the Universidad de las Fuerzas Armadas ESPE, has the most positive effect on the price per square meter, followed by the areas of Armenia, El Triángulo, Ilaló and Panamericana respectively; on the other hand, when the property is located in the area of Gavilánez, the effect on the price is negative.
It is important to note that the R^2 of the model explained by equation [6] is 0.3078, so it does not capture much of the existing variability. On the other hand, the proposed model meets all the assumptions of normality, linearity, homoscedasticity and independence, this is shown in Table 3.

Journal of Business and entrepreneurial studies
ISSN: 2576-0971
http://journalbusinesses.com/index.php/revista

Table 3. *Diagnostic tests of the hedonic model* $\widehat{Pxm2T}_H$.

Diagnostic tests for regression	P-value
Normality of residuals (Lilliefors)	0.135
Autocorrelation (Durbin Watson Test)	0.410
Homoscedasticity (Bartlet)	0.410

As indicated in section 2.3, the first step is to obtain the estimation of the trend function μ(***X***) by means of a multiple linear regression, in order to obtain [4]. For this purpose, we selected those quantitative variables of information on the dwellings that presented the highest values of linear correlation with respect to $Pxm2T_G$. The variables finally selected were: Number of rooms (*NumCua*), Number of bathrooms (*NumBan*), Number of parking spaces (*NumEst*), Years of age (*Anti*) and Number of floors (*NumPis*). Therefore, the model [3] initially proposed would be:

$$Pxm2T_G = + \beta_0 \text{NumCua}\beta_1 + \text{NumBan}\beta_2 + \beta_3 \text{NumEst} + \text{Anti}\beta_4 + \text{NumPis}\beta_5 + \varepsilon \ (\boldsymbol{s}),$$

However, after estimating the model it was found that the *NumCua* and *NumBan* variables were not significant, so they were removed and the following estimated trend function was obtained:

$$\hat{\mu}(\boldsymbol{X}) = 540.02 - 30.47\ \text{NumEst} - 8.07\ \text{Anti} + 86.21\ \text{NumPis} \qquad [7]$$

It is interesting to note that the median price is directly related to the number of apartments, while it is inversely proportional to the age of the house, which is to be expected. Regarding the number of parking spaces, the negative sign could indicate that having too many parking spaces available is not considered attractive for determining the average price of a house. It should also be noted that, although this model does not capture much price variability ($R^2 = 0.19$), this behavior is very similar to that obtained in the hedonic model. However, in the geostatistical approach, much of the residual variability can be captured in the spatial dependence. On the other hand, this regression model meets all the assumptions of normality, linearity, homoscedasticity and independence, as shown in the following table:

Table 4. *Diagnostic tests of the model* $\hat{\mu}(\boldsymbol{X})$.

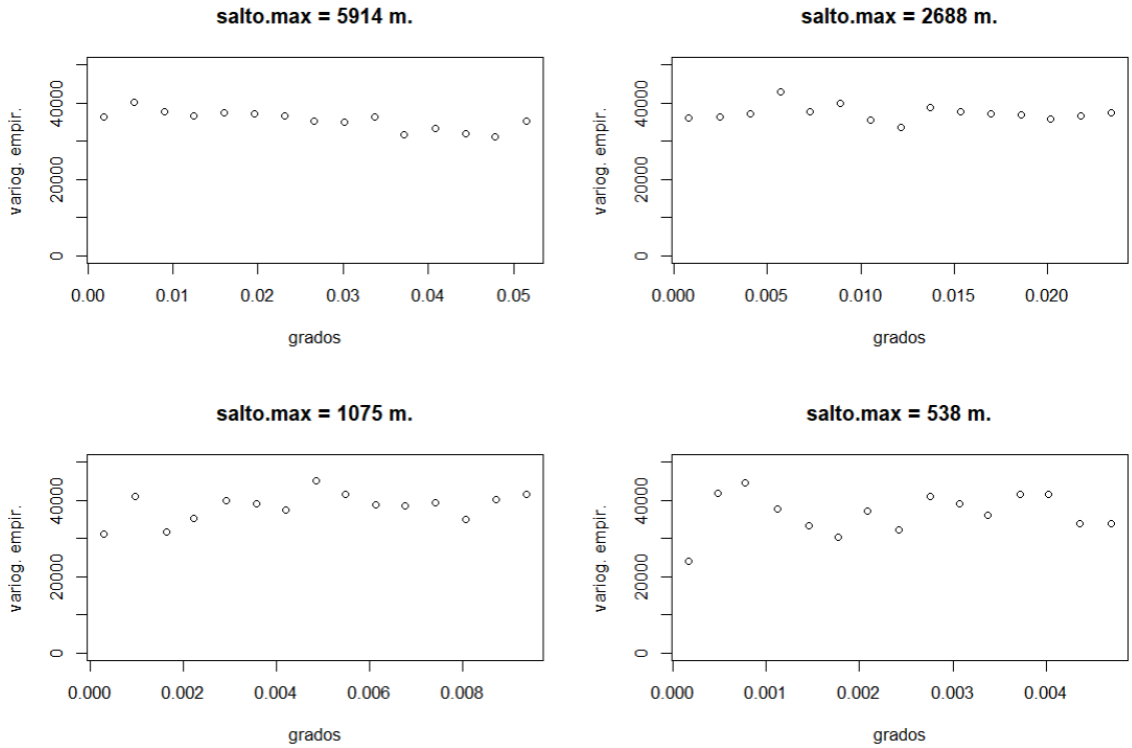
Diagnostic tests for regression	P-value
Normality of residuals (Shapiro-Wilks test)	0.222
Autocorrelation (Durbin Watson Test)	0.922
Homoscedasticity (Harrison-McCabe test)	0.503
Linearity (RESET Test)	0.088

The next step was to resort to Structural Analysis to estimate the variogram. Prior to this, the test based on Moran's Index (Gaetan and Guyón, 2010, section 5.2.1) was used, which showed that there was a significant spatial dependence between the residuals (P-value = 0.018). Based on this result, the empirical estimator was applied to obtain the

Journal of Business and entrepreneurial studies
ISSN: 2576-0971
http://journalbusinesses.com/index.php/revista

pilot variogram. Thanks to this variogram it was also feasible to identify from which maximum distance a greater spatial dependence could be evidenced. For this purpose, 4 maximum distances were considered for the ***u*** jumps: starting from the criterion proposed by Journel and Huijbregts (1978, section 4.1.1.) which is equivalent to 55% of the maximum sample distance (0.0532° or 5914 meters approx.), and was reduced to 25% (2688 m), 10% (1075 m) and 5% (538 m), the last one being the one in which a relevant spatial dependence is observed (See Figure 3).

Figure 3. Empirical variograms with different maximum distances for ***u***.



From these pilot estimates, four valid variogram models were considered for fitting: Circular, Spherical, Exponential and Penta-spherical. The fitting process was performed by weighted least squares, and in each case the corresponding nugget effect, threshold and range estimates were obtained. Cross-validation techniques based on kriging methods were also used to obtain different error measures, such as the mean error (ME) and the standardized mean square error (SMSE) as shown in Table 5.

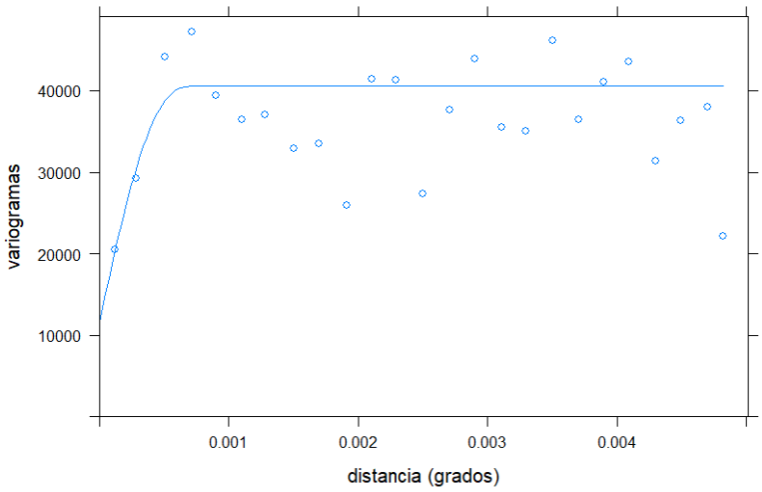
Journal of Business and entrepreneurial studies
ISSN: 2576-0971
http://journalbusinesses.com/index.php/revista

Table 5. *Parametric model estimates of variogram (u) and cross-validation error measures.* $\hat{\gamma}(u)$ *and error measures by cross-validation.*

Model $\hat{\gamma}(u)$	c_0	c_1	a	EM	ECME
Circular	12411.50	28046.97	60.01	0.69	0.938
Spherical	12035.58	28493.73	68.28	0.68	0.938
Exponential	0	45742.3	34.24	0.87	0.906
Penta-spherical	11655.78	28911.32	82.44	0.67	0.939

Considering the results of the previous table, it can be deduced that the Penta-spherical model provides better results, since its EM is lower and its ECME is closer to 1. It should be noted that, according to the range of this variogram, the maximum distance over which the spatial dependence has an effect on the termε (.) of the model [3] is approximately 82. 44 meters around each house located at spatial position s. The values obtained for the parametric variogram parameters can be seen in Figure 4.The values obtained for the parametric variogram parameters can be seen in Figure 4. Therefore, this model was finally selected as a valid variogram to perform the kriging predictions and the final estimates of the geostatistical model given by [7] and [8]. $\widehat{Pxm2T}_G$given by [7] and [5].

Figure 4. Empirical variogram (circles) and fitted Penta-spherical variogram model (line).

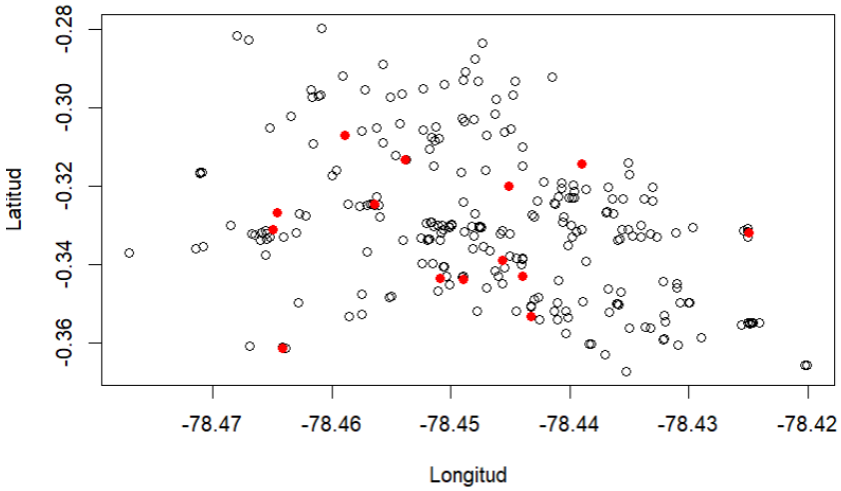


Comparative analysis of hedonic and geostatistical model prediction errors. In order to analyze the behavior of both models when predicting housing prices, the actual *Pxm2T* values of the validation dataset (indicated at the end of section 3.1) were compared with the respective hedonic and geostatistical predictions. $\widehat{Pxm2T}_H$ and

Journal of Business and entrepreneurial studies
ISSN: 2576-0971
http://journalbusinesses.com/index.php/revista

geostatistical $\widehat{Pxm2T}_G$given by [6] and [5] respectively. The spatial distribution of the *Pxm2T* sample data considered for validation is shown in red in Figure 5, while the other points correspond to the values that were used to obtain the predictions corresponding to each model. It is worth mentioning that some points of the original validation set were removed due to their extreme behavior.

Figure 5. Spatial distribution of validation (red) and modeling (black) data.



Specifically, the absolute, quadratic and relative prediction errors were obtained for each estimated model. For example, for the hedonic case they were calculated respectively:

$$err_{abs} = | Pxm2T - \widehat{Pxm2T}_H |,$$

$$err^2 = (Pxm2T - \widehat{Pxm2T}_H)^2, \text{ y}$$

$$err_{rel} = (Pxm2T - \widehat{Pxm2T}_H) / \widehat{Pxm2T}_H.$$

Summary statistics for each of these errors for the models considered are presented in Table 6.

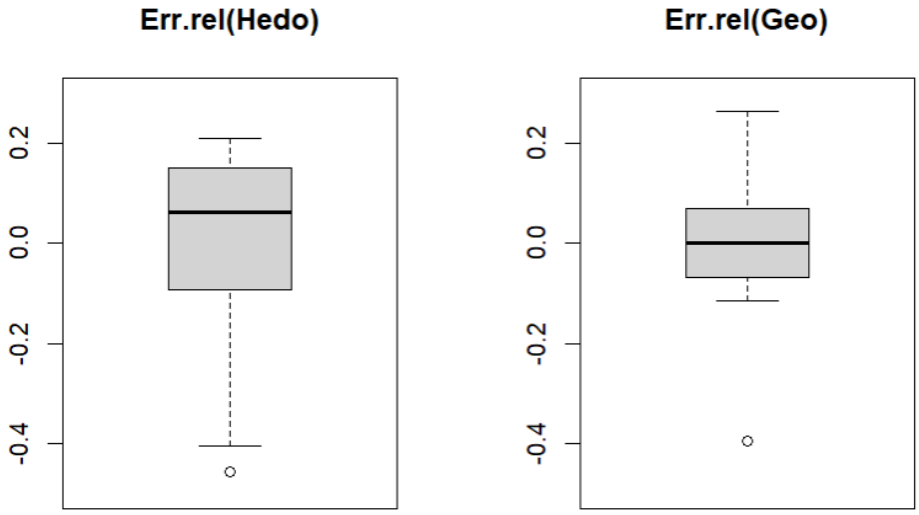
Table 6. *Summary statistics of prediction errors for the hedonic and geostatistical models from the validation data set***.**

Estimated model	err_{abs}		err^2		err_{rel}	
	Media	Desv.St.	Media	Desv.St.	Media	Desv.St.
$\widehat{Pxm2T}_H$	95.70	58.51	12353.30	15202.69	0.01	0.21
$\widehat{Pxm2T}_G$	67.89	75.94	9993.60	15776.49	-0.01	0.16

As can be seen, the average errors are better for the geostatistical model, while similar values are observed in terms of their standard deviation, giving some advantage in general to the prediction $\widehat{Pxm2T}_G$ over its counterpart hedonic version. However, the fact that the dispersion is slightly higher in the geostatistical prediction may be due to the fact that certain validation points are located at the spatial boundary of the data (see

Figure 3). This usually generates an effect on the kriging prediction variance, which in turn makes the predictions less accurate the closer the house is to the edge of the observation region. It is expected that, if the data near the border are removed, the prediction variability will be reduced. $\widehat{Pxm2T_G}$ is reduced.
However, in spite of this boundary effect, it is verified that the geostatistical predictions generally perform better than the values obtained by the hedonic model. This can be corroborated, for example, in the box plots of the relative errors presented in Figure 6.

Figure 6. Box plots of relative errors obtained for $\widehat{Pxm2T_H}$ (left) and $\widehat{Pxm2T_G}$ (right).



CONCLUSIONS

In general, although both the hedonic model and the geostatistical model allow obtaining acceptable predictions for the data, the adjustment of the part related to the variables inherent to housing is quite poor (R2 is very low in both cases). This is due to the housing characteristics of the dwellings inhabited in the sector, since there are several factors, from demographic, geographic and cultural characteristics that can affect the value of the dwelling, but that were not considered at the time of modeling since they are beyond the focus of this work. Due to the fact that the research locality (Rumiñahui canton) has a total of 129 km2, it is not possible to obtain socioeconomic information at this level of disaggregation.
The price per square meter of a house in the canton of Rumiñahui is influenced by the characteristics of the property (I), the proximity to areas of interest within the canton (U) and the area where the property is located (Z). In terms of characteristics, a house with more than two rooms within a complex (urbanized) positively explains the price of the property. On the other hand, the proximity of the house to higher education centers

is more desirable (in terms of price improvement) than being close to a health center. The sectors where housing is most appreciated are the University of the Armed Forces ESPE and Armenia in the Rumiñahui canton.

However, it is observed that the prediction errors of the geostatistical model show a better performance compared to its hedonic versions. This derives from the fact that part of the variability that is not picked up by the trend function is captured by the spatial dependence of the residuals through the residual variogram.

Another advantage of the geostatistical model is that the variogram estimates can explain the extent of spatial dependence. For example, the rank a obtained indicates that the price of a specific house is somehow related to the corresponding value of houses located up to approximately 83 meters around.

Regarding the estimation of the variogram, it is necessary to indicate that several authors have indicated the presence of biases when estimating the variogram from the residuals, and in turn they propose to correct this effect by means of different procedures such as iterative algorithms or non-parametric approximations. However, if any of these procedures are applied to the geostatistical estimates obtained in this work, it is expected that, by correcting for this bias, their error statistics will improve even more, which highlights the importance of considering the information on the spatial location of housing when estimating the price of housing.

REFERENCES

Bover, O. and Velilla, P. (2001). Hedonic house prices without characteristic: the case of new multiunit housing. Economic Studies, 73, Bank of Spain.

Camelo, M. and Campo, J. (2016). Analysis of housing policy in Bogotá: An approach from supply and demand. Revista Finanzas y Política Económica, 8 (1), 105,122.

Chica-Olmo, Jorge and Cano-Guervos, Rafael & Olmo, Mario. (2007). Spatio-temporal hedonic model and variographic analysis of housing prices. Geofocus: International Journal of Geographic Information Science and Technology, ISSN 1578-5157, No. 7, 2007.

Cressie, N. (1993). Statistics for Spatial Data. Rev. ed. John Wiley & Sons.

Chica, J. (1995). Spatial estimation of housing prices and locational rents. Urban studies, 32(8), 1331-1344.

Chilès, J.P. and P. Delfiner (1999). Geostatistics: modeling spatial uncertainty. Wiley, New York.

Derycke, P.H. (1983). Economía y Planificación Urbana. Instituto de Estudios de Administración Local, Madrid.

Figueroa Benavides, E. and Lever D., G. (1992-06). Determinants of housing prices in Santiago: A hedonic estimation. Available at https://repositorio.uchile.cl/handle/2250/128244

Gaetan, C., & Guyon, X. (2010). Spatial statistics and modeling (Vol. 90). New York: Springer.

Griliches, Z. (1971). Price indices and quality change. Harvard U.P. Cambrige.

INEC, (2018). Encuesta de Empleo Desempleo y Subempleo Urbano. Retrieved www.ecuadorencifras.gob.ec/enemdu-2018/.

Journel, A.G. and C.J. Huijbregts (1978). Mining Geostatistics. Academic Press, New York.

Lever, G. (2000). Determinants of housing prices in Santiago: A Hedonic estimation. Paper. Santiago, Chile: Editorial.

Martínez, M. G., Lorenzo, J. M. M. M., & Rubio, N. G. (2000). Kriging methodology for regional economic analysis: Estimating the housing price in Albacete. International Advances in Economic Research, 6(3), 438-450.

Matheron, G. (1962). Traite de geostatistique appliquee, Tome I. Memories du Bureau de Recherches Geologiques et Minieres, 14. Editions Bureau de Recherches Geologiques et Minieres, Paris.

Montero, J. (2004). The average price of the square meter of free housing: A methodological approach from the perspective of Geostatistics. Estudios de Economía Aplicada, 22(3),675-693.

Wackernagel, H. (1998). Multivariate Geostatistics. 2nd ed. Springer, Berlin.