1. Introduction

The objective of the study is to present hedonic pricing models for the housing market in Lajeado and Montenegro, two mid-size cities located in the south of Brazil. This type of pricing model can be used, for example, in tax valuations, which aim at calculating property taxes, sales taxes and capital gain taxes. In order to build the models, it is necessary to demarcate the market area to be explored, define the variables to be included in the model, obtain a reliable and good-sized data sample and treat data through statistical analyses (Appraisal Institute, 2001). In general, small and mid-size cities have still not adopted regression-based models for the valuation, partly because they have the reputation of needing a difficult equation building. It is important to demonstrate that you can develop simple price models, without having a big staff or elevated costs to collect and analyze data.

The housing market is an important segment of the urban economy. It is a common practice to assess the importance of the goods for the society through their prices. In this context, the preferences of the consumers are basically explained by the prices. The properties offering the highest number of characteristics desired by the buyers will have higher prices, and the properties with less characteristics will have lower prices (Robinson, 1979) (Sheppard, 1999).

A relevant characteristic of the market is the real estate heterogeneity. This makes it more difficult or prevents the direct comparison of its units and suggests the use of models for calculating the price. The property valuation is developed with different methods, where the comparative method, based on the regression analysis, has been considered an option because it allows a good precision and objectivity (Appraisal Institute, 2001) (Pagourtzi et al., 2003).

In addition to the heterogeneity, there are other elements that make the real estate market different from other markets, such as the immobility of the product and the time needed for the project and construction of new units. These features have an impact on the prices when the supply and demand conditions are altered. When the demand increases, the prices increase in the short term, based on the fact that it is impossible to make a spatial relocation of the offer and that it takes time to deal with a higher demand. If the demand decreases, prices go down, because the owner or producer needs to reduce the prices in order to recover potential consumers (Balchin and Kieve, 1986) (Sheppard, 1999). On the other hand, the location, often indicated as the most influencing factor on housing prices, also depends on the preferences of the society. Since the product cannot undergo a spatial relocation, the moves in the surroundings can improve or worsen the conditions of the property (Din et al., 2001) (Robinson, 1979).

Consequently, it is easy to perceive that many attributes have to be considered simultaneously in the analysis of the real estate market, which probably has
different values for each situation. In that case, hedonic pricing models gain importance. These models follow the theory of hedonic pricing, which has important contributions from (Lancaster, 1966), (Griliches, 1971), (Rosen, 1974) and (Goodman, 1978), among others. In this context, a real estate is a “compound good” described by a set of attributes. The created model tries to relate the price with these attributes. The estimate of the model—based on data—generates an equation where the coefficients are called hedonic prices (or implicit prices, because the identification of the prices for each attribute is done indirectly).

Many of these models are described in the literature, with the aim of estimating prices or observing the effects of a particular attribute. Some authors present research surveys in the sector, such as (Ball, 1973), (Smith et al., 1988), (Boyle and Kiel 2001), (Din et al., 2001) and (Malpezzi, 2002). When analyzing their models, we perceive a strong variation in the format and components of the models. However, certain basic characteristics can be verified and grouped into three main groups representing attributes of the own construction (physical elements), the location, and time of the transaction. The physical attributes are related to the size (total surface area, number of rooms, bathrooms and garages), building standard, design, technology and way of constructing, among others. The location aspects represent the neighborhood’s quality conditions and accessibility. In turn, the neighborhood is related to the presence of public services, crime level, educational level and income of the people living in the area. The accessibility refers to the distance or access time to the places that are important for the society. Finally, the information regarding the conditions of the transaction, such as the payment method and the time of the sale, can also affect the prices.

In more specific terms, some authors worked with sustainability aspects. (Sander and Haight, 2012) have studied the value of ecosystems. (Saphores and Li, 2012) measured the effect of green areas. (Cohen et al., 2015) studied the effects of the distance of houses from the water. (Swoboda et al., 2015) worked with price variations due to traffic noise. (Welch et al., 2016) investigated the influence of the access to bike lanes and the proximity of public transport. Other works studied the effects of accessibility and location, with elements regarding the distance from highways (Allen et al., 2015), the accessibility in general (Li et al., 2016) and easy access and safety for people who walk in the neighborhood (Li et al., 2015).

Moreover, housing studies indicate some key variables, such as size (measured in built area or number of rooms), construction quality and location (qualitative variable defined by the researchers or measurement of the distance to important points, such as downtown). In this sense, the contributions of (Helbich et al., 2014), (Nuñez and Surhoff, 2002), (Ramirez and Giraldo, 2013) can be highlighted. Additionally, a number of research works address the same region of this study. (Braga and Alves, 2015) built pricing models including variables that measure socioeconomic factors. Other authors developed models with similar structure and size, like (Delli and Zancan, 2013) and (Palagi et al., 2014).

2. Methodology

2.1 Criteria

In order to build pricing models, it is necessary to collect data in the area of interest, and generate the corresponding models through regression analyses. The regression analysis is a technique aimed at associating independent variables to a dependent variable (the market price) by means of an equation. The objective is to develop a numerical model. A hedonic price function can be proposed in the general way shown in (Equation 1):

\[
\text{Price} = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \ldots + \alpha_k x_k + \epsilon
\]  

(1)

Where \( x_1, \ldots, x_k \) are the attributes; \( \alpha_1, \ldots, \alpha_k \) are the coefficients of the equation that represent the relative importance of each attribute for explaining the dependent variable (implicit prices); \( \alpha_0 \) is the intercept of the equation and \( \epsilon \) is the error term.

The statistical analysis establishes the variables that should remain in the model and their relevance for explaining the price. Several conditions (regression assumptions) should be checked to ensure the quality of the generated model, among them, the presence of homoscedasticity, the normality of errors and the linear relationship in (Equation 1), as well as the absence of multicollinearity and outliers. The tests required are a variance analysis of the model with the Fisher-Snedecor F-test and a Student t-test for the significance of the variables. The model’s explanatory power is analyzed with the adjusted coefficient of determination (\( R^2 \)).

Considering that the studied models will probably be used for real estate taxation, the traditional requirements found in the literature were adopted as reference limits for the statistical analysis. The variance was tested with a 1% significance level and the variables were analyzed at a 5% significance level.

2.2 Data Collection and Model Development

2.2.1 Lajeado

Lajeado is a city located 117 km from Porto Alegre, the capital of the state of Rio Grande do Sul, in the south of Brazil. The city has a population of 71,445, according to the 2010 Brazilian Census, and 99.6% lives in the urban area (Lajeado, 2015). The city is located in the valley of the Taquari River, which has 36 municipalities. The data consist of one or two level homes, which were obtained from real estate agents in the city and assessment reports by independent appraisers hired by the Caixa Econômica Federal for the financial study. The city of Lajeado is divided into 27 neighborhoods, among which the studied region included 15 neighborhoods, and 119 data were collected from the local market.

2.2.2 Montenegro

Montenegro is located in the Metropolitan Region of Porto Alegre, and it is 55 km away from the capital. The city is located in the valley of the Cai River, and the region has 20 cities of similar size. It has a population of 59,415, according to the 2010 Census (Montenegro, 2015). The data were
obtained from local real estate agents, and professionals who make appraisals for the Caixa Económica Federal were consulted. The data are distributed in 22 of the 24 neighborhoods of the city. The hedonic models were built using 232 data of house sales.

2.2.3 Studied Variables
The property price was the dependent variable defined for the study. Ten (10) variables were identified based on data collected from the properties themselves, and they were tested through statistical analysis (Table 1). The resulting models presented part of these variables, as indicated below.

Table 1. Variables tested in the model

<table>
<thead>
<tr>
<th>Variables</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>Sale or appraisal price of the property, converted into dollars</td>
</tr>
<tr>
<td>Housing area</td>
<td>Surface built in square meters (m²)</td>
</tr>
<tr>
<td>Bedrooms</td>
<td>(m²) Number of rooms (from 1 to 4)</td>
</tr>
<tr>
<td>Construction standard</td>
<td>Terrible = 1; ruin = 2; regular = 3; good = 4; very good = 5; optimal = 6</td>
</tr>
<tr>
<td>State of conservation</td>
<td>Terrible = 1; ruin = 2; regular = 3; good = 4; very good = 5; optimal = 6</td>
</tr>
<tr>
<td>Land area</td>
<td>Land area in square meters (m²)</td>
</tr>
<tr>
<td>Neighborhood</td>
<td>Ruin = 1; regular = 2; good = 3; very good = 4; optimal = 5</td>
</tr>
<tr>
<td>Date of the band</td>
<td>Transaction or valuation = 1; Ad or sale offer = 2</td>
</tr>
<tr>
<td>Data de la venda</td>
<td>Transaction Year</td>
</tr>
<tr>
<td>Infrastructure</td>
<td>Existence of paving, water network, sewage collection, garbage collection, electricity network (sum, from 1 to 5)</td>
</tr>
<tr>
<td>Urban services</td>
<td>Existence of collective transport, school, health post or hospital (sum, 1 to 3)</td>
</tr>
</tbody>
</table>

Source: Prepared by the authors

The variables of Construction Standard, State of Preservation and Neighborhood were estimated by a qualitative appraisal of the researchers, on site. Infrastructure and Urban Services are variables indicating the number of considered elements that were available in the properties at the time of the sale or appraisal. In relation to urban services, a radius of 0.5 km around the property was considered to verify the offer of these elements. The other variables were obtained directly from the data sources.

For both cities, the data collection was carried out by a part-time researcher and it took approximately two months. After verifying the data, the analysis of the models was performed in one week.

3. Results and discussion

The tested models follow the basic format presented in (Equation 1). The validation hypothesis for the model was tested with a variance analysis through the F-test and the importance of the variables was analyzed with the Student t-test. The results obtained for the model are shown in (Table 2) and (Table 3).

3.1 Lajeado

Different models were explored, with different configurations in terms of numerical formats and variables included. We herein present the results for the statistical model with best goodness of fit. The coefficient found for the model is $R^2 = 0.938$, which indicates that the model explains around 94% of the variability of real estate prices due to variations of the selected independent variables. The parameter of the Fisher-Snedecor F-test was $F = 285.279$, with $p=4.58 \times 10^{-58}$, which exceeds the requirements and rejects the hypothesis of nonexistent linear relationship, that is, the regression equation can be accepted, indicating a good quality of the model. The coefficients and the individual significance of the independent variables that reached the specified level are shown in (Table 2). The remaining variables (Table 1) were tested, but they did not achieve the 5% level.
It can be concluded that the variables of Surface Area of the House, Rooms, State of Preservation and Neighborhood are acceptable for building the model. The model reached the minimum recommended values in the t-test and F-test, indicating that it is a good quality model and that the selected explanatory variables have an influence on the property prices.

The regression assumptions were analyzed, and no problems were reported. Based on the results obtained in the regression analysis, it can be concluded that it is adequate for the set of data collected. The calculated values allow the (Equation 2), interpreted as a linear regression model that is fitted for determining the value of a house in the city of Lajeado (according to the limits of collected data).

\[
\text{Price} = -13.029.75 +304.89*\text{Land Area} +6.288.45*\text{Bedrooms} +1.155.06*\text{State of conservation} +2.041.53*\text{Neighborhood} \\
\]

The sign of the coefficient of each variable indicates whether the price tends to increase or decrease based on the variation of that variable. The positive coefficient of the variable Surface Area of the House means that the price tends to increase US$304.89 as the surface area increases by 1 m

\[2 \]

Likewise, the probable price variation between two similar properties, with just a difference in the number of rooms, is approximately US$6,300 for each extra room.

\[3.2 \text{ Montenegro} \]

In the same way as in the case of Lajeado, different models were explored for the market of Montenegro and the results for the statistical model with best goodness of fit are presented. The coefficient found for the model is \( R^2 = 0.926 \), indicating an explanation of almost 93% regarding the price variation of the properties, due to the behavior of the independent variables included in the model. The variance test presented an \( F_{\text{calc}} = 399.794 \), with a \( p=5.33*10^{-117} \), which allowed concluding that the error level is very low. The significance of the model’s independent variables, statistically tested with the Student t-test, confirmed the acceptance of the variables (Table 3).

\[
\text{Table 2. Coefficients and significance of the variables – Model estimated for Lajeado} \\
\]

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>( t )</th>
<th>Significance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-13.029.75</td>
<td>3.254.58</td>
<td>-4.00</td>
<td>0.011</td>
</tr>
<tr>
<td>Housing area</td>
<td>304.89</td>
<td>17.40</td>
<td>17.52</td>
<td>7.12*10^{-34}</td>
</tr>
<tr>
<td>Bedrooms</td>
<td>6.288.45</td>
<td>1.621.02</td>
<td>3.88</td>
<td>0.018</td>
</tr>
<tr>
<td>State of conservation</td>
<td>1.155.06</td>
<td>394.63</td>
<td>2.93</td>
<td>0.415</td>
</tr>
<tr>
<td>Neighborhood</td>
<td>2.041.53</td>
<td>538.60</td>
<td>3.79</td>
<td>0.024</td>
</tr>
</tbody>
</table>

Source: Prepared by the authors

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>( t )</th>
<th>Significance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-64.580.99</td>
<td>4.713.53</td>
<td>-13.70</td>
<td>1.79*10^{-31}</td>
</tr>
<tr>
<td>Housing area</td>
<td>348.83</td>
<td>18.31</td>
<td>19.05</td>
<td>7.91*10^{-49}</td>
</tr>
<tr>
<td>Bedrooms</td>
<td>3.805.43</td>
<td>1.753.37</td>
<td>2.17</td>
<td>3.103</td>
</tr>
<tr>
<td>State of construction</td>
<td>7.807.67</td>
<td>1.375.57</td>
<td>5.68</td>
<td>4.23*10^{-8}</td>
</tr>
<tr>
<td>State of conservation</td>
<td>5.844.61</td>
<td>1.135.72</td>
<td>5.15</td>
<td>5.79*10^{-7}</td>
</tr>
<tr>
<td>Land area</td>
<td>35.83</td>
<td>4.39</td>
<td>8.16</td>
<td>2.34*10^{-14}</td>
</tr>
<tr>
<td>Neighborhood</td>
<td>8.625.20</td>
<td>836.47</td>
<td>10.31</td>
<td>1.15*10^{-20}</td>
</tr>
</tbody>
</table>

Source: Prepared by the authors

\[\text{Table 3. Coefficients and significance of the variables – Model estimated for Montenegro} \]
The presented model is also acceptable, since it shows a good result in the statistical analysis and reaches the proposed limits. All the variables are fitted for the appraisal models, according to the chosen limits. The regression assumptions were analyzed and no statistical problems were detected. The (Equation 3) refers to the model for properties in Montenegro.

\[
\text{Price} = -64,580,99 +348,83* \text{Land Area} +3,805,43* \text{Bedrooms} +7,807,67* \text{Construction standard} +5,844,61* \text{State of conservation} +35,83* \text{Land Area} +8,625,20* \text{Neighborhood}
\]

The variables show results that are consistent with those expected in terms of their contribution to the pricing process and relative importance. For example, the relevance of the Surface Area of the House is reflected in the coefficient that indicates a variation of almost US$350 per square meter.

4. Discussion and conclusions

El estudio recopiló datos de bienes raíces de dos ciudades brasileñas y utilizó la inferencia estadística para construir modelos hedónicos. Los dos modelos de regresión propuestos se analizaron para verificar sus suposiciones básicas, pruebas de significación y poder explicativo de cada modelo, de manera convencional.

The study collected real estate data from two Brazilian cities, and used statistical inference to build hedonic models. The two regression models proposed were analyzed to verify their basic assumptions, significance tests, and explanatory power of each model, in the conventional way.

The models followed traditional specifications in the study of hedonic models, according to the consulted literature. Both models obtained a good explanatory degree of the prices (R² de 94% for Lajeado y 93% for Montenegro), where the models exceeded by far the minimum values in the F-test (\(P_{\text{Lajeado}} = 4.58*10^{-10}\) and \(P_{\text{Montenegro}} = 5.33*10^{-17}\)). Furthermore, all the variables involved in the model have a significance below 5% (actually, only one is above 1%). This indicates that the models have good-quality relationships with the collected data. Additionally, the analysis of the regression assumptions did not present any problems.

When comparing the coefficients of (Equation 2) and (Equation 3), it is verified that the set of variables included in the final model is not the same. Some attributes repeat themselves, but with a different level of participation. For example, although the influence of the Surface Area of the House is similar (US$ 304.89 in Lajeado and US$ 348.75 in Montenegro), there are important variations in other variables. Regarding the number of rooms, the difference is almost US$2,500 more in Lajeado. The Neighborhood’s influence on the prices is four times higher in Montenegro. These differences are common and reflect the different behavior of the population who buys properties in each of the cities.

Part of the difference in the Neighborhood factor can also be attributed to different criteria used to qualify the regions of each city. The intention was to maintain the criteria, but the more or less detailed knowledge about each locality may affect the measurement.

The results reinforce the need to estimate specific models for each city, and avoid reproducing studies or models made in other localities. On the other hand, it can be observed that models were estimated for an adequate appraisal, with relative easiness. The models were developed by two part-time professionals during a few weeks. The spatial reference is 15 from 27 neighborhoods in Lajeado (55%) and 22 from 24 neighborhoods in Montenegro (92%). In the case of Montenegro, almost the entire city could be evaluated with (Equation 3). In a real case, other models would be needed to assess lands, apartments and commercial properties.

Through the development of models with satisfactory statistical performance, it was possible to estimate the behavior of housing prices in these two cities. Accordingly, the work contributes to clarify and give examples of the process, aiming at their use in small and mid-size cities.

5. Acknowledgements

The authors of this paper express their gratitude to the Brazilian research foundations CNPq, CAPES and FAPERGS for supporting and financing this research.

6. References


