Real estate market and the relevance of local features in a hedonic prices quantil-spatial analysis – the case of Belo Horizonte – Brazil

Furtado, BERNARDO

CEDEPLAR/UFMG, Centro Universitario UNA

July 2007

Real estate market and the relevance of local features in a hedonic prices quantil-spatial analysis – the case of Belo Horizonte – Brazil

Bernardo Alves Furtado
Ph.D. Candidate in Economics CEDEPLAR/FACE/UFMG
Professor Centro Universitário UNA
furtado@cedeplar.ufmg.br

Abstract

The main motivation of this paper is to identify how relevant the localization of a specific estate is in its market value. Furthermore, it aims to understand better how economic aspects influence and are influenced by urban space. In order to do so, a myriad of concepts is drawn from a variety of fields of science: from geography to architecture, from urbanism to economics, as well as methodologies, which are borrowed from statistics, econometrics and geoprocessing. The proposal of the paper follows the hedonic prices function literature, but suggests that a synthesis of the perception of urban amenities can be expressed by the element of the neighbourhood (as proposed by Lynch, 1997). A number of models are presented, tested and commented. The one with the best fit is the spatial error-lag (Anselin, 1988) specified with a ranking of neighbourhood income. A quantil analysis adds considerably to the understanding of the model.

Key-words: real estate market, neighbourhoods, spatial econometrics, quantil analysis, Belo Horizonte.

JEL Classification: C21, R2, L85.

1 Introduction

The main motivation of this paper is to identify how relevant the location of a specific estate is in its overall market price so that economic relations on urban tissue are better described. In order to do so, a number of concepts that have come from different fields of science: from geography, to architecture; from urbanism, to economics; and methodologies: from statistics, to econometrics to geoprocessing.

Hypothetically, estates with the exact same attributes but located in different neighbourhoods would have its price changed by how much? Hence, the objective of the paper is to identify patterns and behaviour of real estate market for the city of Belo Horizonte meanwhile the methodology is tested and the principles assumed confirmed. In doing so, it follows the approach of Neto (2002) and Hermann and Haddad (2005) for whom the main question is whether the level at which the price is established is determined by its attributes or by its location. In the literature "…the focus of econometric analysis is on the estates' features rather than the locational ones…" (HERMANN & HADDAD, 2005, p. 13). Therefore, the aim is to complement this view and detail all aspects that compose the price.

Traditionally, real estate analysis considers monocentric (or duocentric) cities in which distance to the business district centre (CBD) is an important variable. Furthermore, it considers other measurements such as: distance to the nearest highway, availability of services, presence of green areas and parks, violence and crime indexes, when designing the model. This paper follows a somehow different approach.

It follows the path laid by architect Kevin Lynch (1997) and his proposal of Urban Landscape theory, according to which the perception of the city by its citizens is spatially circumscribed within a known entity. In other words, the person relates positive and negative qualities (attributes) to certain spatial portions of the territory that is somewhat precisely recognized by the community in urban space.

This choice is in accordance to sales practice of real estate which always discriminates the spatial unit of the neighbourhood when advertising a sale. Having considered this approach, the "distance to CBD" factor is embedded within its location and is not measured explicitly[1]. If the attribute is relevant, the neighbourhood as a whole is positively valued (or

[1] From an urban point of view, even the specification of where the center of a city such as Belo Horizonte is located it is rather difficult, given its natural sprawl. All of the centre-south of the city, a large polygonal could be considered the CBD. Measure the distance to this expanded centre is inaccurate and would lead to great error. Furthermore, intra-centre relations might be more relevant to price establishment than distance to an allegedly centre.

otherwise). In short, the point of view of the paper is that the neighbourhood is treated as reference for a number of variables which are weighted together and not separately.

The complex nature of urban space and the city (MONTE-MÓR, 2006) leads to problems of multicollinearity when applying regression techniques that derive from the non-neutrality of space. It is assumed that rigid physical space condition socio-economical issues while it is conditioned by them (SOJA, 1993; BRUECKNER, THISSE *et al.*, 1999)[2].

Principal component analysis (PCA) was central to the problem as it enabled the construction of an index of values of family income which satisfactorily summarized the multiplicity of input factors for each neighbourhood.

The nature of the problem also conditioned the analysis to be spatial and quantilic. That is to say that there were considerable gains in quality of specification when comparing traditional model with the spatial one. The quantilic model, in turn, highlighted the different valuing of location for estates at different level of prices, adding to the proposed analysis.

This article briefly discusses the theory of hedonic prices, urban landscape perception, justifies the use the neighbourhood as a spatial analysis unit; the description of the traditional model, the spatial one along with its spatial weight matrix and the quantilic model. Some limitations of the model are presented. The database of the study case of the city of Belo Horizonte is discussed as well as the analysis of the results and the final considerations.

2 Theoretical references

2.1 Hedonic prices and urban landscape perception: neighbourhoods as units of analysis

Real estate market is peculiar in the sense that every location is different from all others which leads to the monopoly of space (SINGER, 1982). Furthermore, its object, housing, is a heterogeneous good. Hence, Hermann & Haddad argue that the best model to apply is the hedonic price model:

> [Hedonic price model] considers a heterogeneous good as a closed package of attributes and estimates the marginal price of each attribute based on the analysis of the observed value of the good and the quantities of each attribute (2005, p. 239).

Sheppard (1999, p. 1599) presents two arguments to use hedonic price function when applying it to real estate: a) when constructing price indexes that consider change in the quality of consumed goods[3]; and b) as an informational attribute when establishing price levels for heterogeneous goods.

[2] On this point, see also Harvey (1973); Santos (1992) and Soja (1996).
[3] Within this context, a building constructed in a different local is considered a different good.

The discipline that studies urban landscape among architects and urbanists has been greatly formulated by Kevin Lynch and Aldo Rossi (2001)[4] and it specifies that the perception of urban landscape made individually and collectively contributes to the construction of a place's image. Physical and imaginary or symbolic elements[5] (such as highways, train lines, rivers, squares, corners or monuments) segment and label space into unique and identifiable units.

The neighbourhood assumes this function of homogeneous spatial unit upon which attributes are associated to. Following this approach of construction of perceived image, a positive feature, an urban amenity[6], such as presence of green areas or parks, is associated to a certain locality and it is non-linear in space. That means that even tough a park might be closer to a certain building when measured linearly, if the building in question is across the river, in a different neighbourhood, it may be associated to a different environment and not benefit, in people's constructed image, from that amenity presence.

Note that, given the high degree of heterogeneity of urban space, the spatial units referred vary in size. Practically speaking, an irregular occupation of land use, such as a slum, would never be classified in the same spatial unit of the adjacent neighbourhood, although it will devalue it.

2.2 The proposed model and the consequent needs of the modelling process

Following Anselin's recommendation (2005, p. 199), the first model is a traditional one which will enable the use of Moran's I statistic and Lagrange Multiplier (ML) to identify the need of a spatial model.

The construction of the model followed a choice in which synthetic variables were chosen instead of the traditional listing of an expressive number of them that could properly represent the bundle. The objective of this choice was to reduce problems of multicollinearity whereas enabling access to a wider database[7]. The results indicate that this choice was reasonable.

The model proposed considers that the price of the estate, the dependent variable, would be explained by a) its area, since it is expected that the larger the estate, the more expensive it is ; b) its age, and in this case the expected coefficient is negative – meaning that

[4] To see other approaches, see the work of Cullen (1961) and Yi-Fu. Tuan (1980; 1983).
[5] The five analytical elements proposed by Lynch (1997) are: roads, limits, marks, neighbourhoods and knots.
[6] For the formal derivation of an urban amenity model in New Urban Economics, see Brueckner, Thisse e Zenou (1999).
[7] Actually, econometric models are theoretically intended to represent behaviour on average.

the older the estate, the cheaper it is; c) its territorial city tax (IPTU) that serves well as a proxy to the construction standard features of the estate, such as finishing quality; d) its building service fee (COND) – since all observations are of apartment buildings – and the fee is a good indicator of services of the building. Actually, it reflects the presence of porters; expenditure on elevators, pool maintenance, garden care[8]; e) its number of available parking spaces[9].

The control of urban local characteristics was tested in two different ways. Firstly, an index was provided by a research institute which has based it on family income information and it has four levels: popular, average, luxury and high luxury. This scaled standard serves as reference for real estate brokers along with a price by square meter. Initial results clearly demonstrated that this specification was not detailed enough which led to the construction of a more specific index that will be discussed in the next section.

2.3 Multivariate analysis – principal components

Principal component analysis – PCA – is a statistics technique which has been largely used in data compression. According to Mingoti[10]:

> …its main objective is to explain the structure of the variance and co-variance of a stochastic vector (…) through the construction of linear combinations of the original variables. This linear combinations are called principal components and are not correlated among each other (2005, p. 59).

Be $X = (X_1, X_2, \ldots, X_p)$ a random vector; $\sum_{pxp}$ its matrix of co-variance with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$, the eigenvalue of the matrix and $e_1, e_2, \ldots, e_p$, the eigenvectors normalized. The j[th] principal component of the matrix $\sum_{pxp}$ can be defined as:

Equation 1 – Definition of principal component

$$Y_j = e'_j X = e_{j1}X_1 + e_{j2}X_2 + \cdots + e_{jp}X_p \qquad \text{(MINGOTI, 2005, p. 60-61)} \qquad (1)$$

The construction of the components implies that the first component is always the most representative in terms of total variance and so forth. The numerical values of the principal components are called scores and can be used in regression analysis (MINGOTI, 2005, p. 60).

The procedures – made automatically by the software ArcGIS or numerous other statistical commercial packages – are as follows: the data is standardized, the matrix of

[8] In fact, a regression with the service fee (COND) as dependent variable and presence of pool, presence of common building room, presence of ballroom and location showed a good fit with an adjusted R^2 of 0.6175.
[9] These variables were validated in a meeting where there were real estate brokers representatives who represented one fifth of all the transactions of the city of Belo Horizonte.
[10] See also Johnson and Wichern (1998).

correlation is calculated and so are the eigenvectors and eigenvalues. Those will enable the construction of the values for each observation according to equation 1.

One of the advantages of PCA technique is that it is not necessary to assume the normality of the distribution. A disadvantage, however, is that some studies have shown that changes on the scale of the variables might alter the results (NAIK & KHATTREE, 1996). The standardization of the data helps minimize this issue.

Principal component analysis was used within the context of this paper to make an index of relevance of the neighbourhood based on the income of families supplied by the results of census tract (IBGE, 2000). principal component analysis was used by Hermann and Haddad (2005) in Brazil and by Can and Megbolugebe (1997) in the United States. Specially when analysing real estate market

2.4 Spatial analysis and spatial econometric models

The essential concept to understand when working with spatial analysis is that of spatial dependence. It is rooted in what is commonly called the First Law of Geography of Tobler, according to which “Everything is related to everything else, but near things are more related than distant things” (TOBLER, 1970, p. 236). Spatial autocorrelation, in turn, is the expression that measures this spatial dependence.

When it comes to spatial econometrics, Anselin (2005) suggests that one first implements traditional models and then tests for the presence of spatial autocorrelation and, therefore, whether it is necessary to use a spatial model.

The first indication of the presence of spatial dependence is Moran’s statistic:

Equation 2 – Moran’s I

$$I = \frac{\sum_{j=1}^{n} \sum_{i=1}^{n} w_{ij} \left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}, \text{ where} \qquad (2)$$

W_{ij} is the weights matrix which brings the spatial considerations into the formula; n is the number of variables; x_i is the variable of interest.

The weights matrix can be constructed in a number of different ways and they will be discussed in the next section.

It is important to notice, however, that although Moran’s I index is adequate to indicate the presence of spatial autocorrelation, it does not show in what form it occurs (ANSELIN, 2005, p. 197).

Spatial models should be applied when there are theoretical indications derived from the problem in question that there is spatial influence among observations. On this paper, for instance, the spatial case should be considered if there is an indication that the price, quality and offer of services of nearby estates influence the values of the estate that is being analysed. According to the simple model of spatial process identification proposed by Anselin (2005, p.199), if there is not spatial influence detected, one can return to traditional models with more guarantees of its fitness.

One of the most cited econometric spatial models in the literature is the spatial lag, which includes a term of influence of neighbours in the regression process:

Equation 3 – Spatial lag model

$$Y = \rho Wy + X\beta + \varepsilon, \text{ where} \qquad (3)$$

Y is the dependent variable; Wy is the spatial lag vector, derived from the choice of spatial weights matrix[11]; ρ is the spatial auto-regressive coefficient; X is the matrix of independent variables and their coefficients β and ε are the error vector.

Another very common model in spatial econometrics is the spatial error lag[12] that is capable of capturing a "…spatial effect that has not been included in the model…" (ALMEIDA, 2004, p. 61). When this occurs, the spatial effect implicit in the problem has not been totally included and is captured by the error.

Equation 4 – Spatial error model

$$Y = X\beta + u, \text{ and} \qquad (4)$$

$$u = \lambda Wu + \varepsilon, \text{ where} \qquad (5)$$

λ is the error auto-regressive spatial parameter[13].

The tests recommended to identify spatial models are: a) the Lagrange Multiplier Lag, when it is a lag-spatial model and b) the Lagrange Multiplier Error to indicate the use of error auto-regressive model. Finally, there is also LM-SARMA which would indicate the necessity of a more complex model in which there is spatial autocorrelation as a variable and in the errors, a combination of the two presented models.

It is important to highlight that Ordinary Least Squares (OLS) are not adequate to treat spatial analysis as the literature (ALMEIDA, 2004) indicates that when in the presence of

[11] It is not the objective of this paper do discuss the underlying theory. For further details, please consult Anselin (1988) among others.
[12] Error spatial model was originally proposed by Whittle (1954).
[13] Other spatial models include: average moving spatial model, spatial Durbin model, mixed model with spatial lag and error moving average of first degree (ALMEIDA, 2004, cap. 4).

spatial autocorrelation, OLS results are inconsistent. The suggested alternative is to use a Maximum Likelihood (ML) estimator as the parameter that most likely generated the observed sample. This suggestion is in Anselin (2005) and it follows methodological development of Anselin (1988), Anselin and Bera (1998) and Smirnov and Anselin (2001)[14].

2.5 Weight matrixes

Weight matrixes play essential role in spatial analysis and they are both the representatives of neighbours' influence in each observation and responsible to account for how this influence diminishes in space. Many authors insist that different weight matrixes may alter results significantly and it is reasonable to choose wisely the matrix and test the results for different specifications.

The simplest models of weight matrixes are those of spatial contiguity that indicate, through a binary relation, the fact of being neighbours or not. There are two basic types among them: the so-called queen matrix that considers as neighbours all of those which share a border or have even a vertex in common and those called rook in which only those that have a common border are considered to be neighbours. In both cases, it is possible to include higher matrix orders so that neighbours of neighbours might be considered neighbours. A strong underlying theory of the phenomenon being studied should contribute to indicate the most adequate choice.

Alternatively, there are the matrixes built on spatial distance, in which a certain ad hoc limit is specified and those within that limit are considered to be neighbours. This one is specially indicated on those cases where the influence can be readily measured[15]. There is also the matrix of the k-nearest neighbours in which neighbours – as the name suggests – are defined as those that are closest to the observation. Once more, this number k is set in an ad hoc manner.

2.6 Quantilic analysis

The quantilic analysis[16] is based on/upon points taken in regular intervals from the accumulated distribution function[17] of a stochastic variable. The intervals denote sub-groups

[14] According to, "however, due to the simultaneity implied by the spatial nature of the dependence, these procedures [estimated general least square] are not applicable in the spatial case and a full maximum likelihood estimation must be carried out" (1992, p. 214). The results of the generalized moments (two-step) presented very similar results (differences of the order of 0,01), except for the PCA index, which difference was higher (0,1).
[15] For example in the case of a mosquito flying autonomy.
[16] See Buchinsky (1997) for further details and the original article of Koenker & Bassett (1978).
[17] $F(x) = \mathrm{P}(X \leq x)$, where F(x) is the probability that the variable X presents value inferior or equal to x. So that the probability of X is in an interval (a, b) is $F(b) - F(a)$ if $a \leq b$.

of the sample that can be of any order (a hundred, for centis; 20 for quintis). In this paper, the theoretical reference is the one proposed by Buchinsky (1997) in which the author reinforces the necessity to verify if the patterns observed in a traditional regression (OLS) repeat themselves in the quantis. Otherwise, the quantilic analysis adds valuable information to the interpretation of the phenomenon. Actually, it helps determine whether the behaviour at higher/lower levels of the dependent variable differs. Quantilic regression is not so sensitive to presence of outliers (as OLS is) and it enables change analysis “at different points of the distribution”. The author summarizes the relevance of quantilic analysis stating that “clearly, it is not enough to investigate changes in the mean when the entire shape of the distribution changes dramatically” (*op.cit.*, p. 90)[18].

Wooldridge also recommends quantilic regression when the sample is heterogeneous:

> Median regression is a special case of quantile regression, where we model quantiles in the distribution of y given x. For example, in addition to the median, we can estimate how the first and third quartiles in the distribution of y given x change with x. Except for the median (which leads to Least Absolute Deviations), the objective function that identifies a conditional quantile is asymmetric about zero. See, for example, Koenker and Bassett (1978) and Manski (1988, Section 4.2.4). Buchinsky (1994) applies quantile regression methods to examine factors affecting the distribution of wages in the United States over time (2002, p. 367).

2.7 Limitations of the models

This paper is exploratory and intends only to contribute in the understanding of the spatial dynamics of the real estate market of Belo Horizonte.

The main limitation is that the information collected on the price of the estates came from real estate brokers and are just their sale price. Probably they are different from real prices paid at the end of negotiation[19]. This fact is described in the literature and it is not responsible for major problems because when dependent variables are measured with error, we have:

Equation 5 – Model when dependent variable is measured with error

$$Y^* = \alpha + \beta X + \varepsilon, \text{ where } Y^* \text{ is the variable measured with error} \quad (6)$$

Equation 6 – Variable transformation

$$Y = Y^* + u, \text{ and } Y \text{ is the correct variable} \quad (7)$$

In this case, the following arrangements can be made:

[18] This statement will be confirmed by the results of the model.

[19] Data of actual sold prices are being collected in the Tax Department of the City and will be presented in future work.

Equation 7 – Model's adjustment

$$Y - u = \alpha + \beta X + \varepsilon \qquad (8)$$

Equation 8 – Model with error incorporation

$$Y = \alpha + \beta X + (\varepsilon + u) \qquad (9)$$

Equation 9 – New model with adjustment

$$Y = \alpha + \beta X + w, \qquad (9)$$

The measurement errors are incorporated in the errors and one has the original model again, in which the OLS estimators are unbiased, consistent and efficient.

Another relevant limitation is the correct indication of the "neighbourhood" given to a certain estate that has been offered for sale. It would be tempting for the broker to advertise the estate in a closer, more valued, neighbourhood and therefore reach great market value. Against this practice, it is assumed that the broker would damage his or her credibility if systematically providing inaccurate locations' name. Furthermore, the similarity of nearby neighbourhoods and the great number of different ones included and the size of the sample (510 observations) all contribute to diminish these problems.

The spatial analysis in turn suffers from the fact that the estates are not in their exact coordinate since full addresses were not available. Instead they are located in the centre of the neighbourhood to which they belong. Hence, the analysis made does not differentiate effects at the neighbourhood scale level, but that of the city. This limitation is in accordance with the proposed theoretical view that supports the idea that spatial attributes are perceived at the level of neighbourhoods and not at micro-urban level.

The author understands that these limitations do not compromise the exploratory results achieved.[20].

3 Specification of the models for the case of Belo Horizonte

3.1 General database

The database was composed of the collection of data publicly available by real estate brokers of Belo Horizonte. The actual compiling was done by internship students[21] at the internet site of the association of brokers' companies called Rede NetImóveis. The site[22] interface allows the selection by city, neighbourhood and type. Once selected the

[20] Another paper, with information on specific location of each estate is under deveopment.

[21] Cristiane Nobre Prudente made the collect for two-bedroom apartments from 6th to 14th October 2005. Leila Luiza made the collect for three-bedroom apartments on 05th October 2005. Vladimir Augusto made the collect for four-bedroom apartments from 12th to 18th October 2005. The author himself made the collection of houses and penthouses between 20th and 29th October 2005.

[22] <http://www.netimoveis.com.br>.

observations, their characteristics were automatically computed into tables. Overall, data was collected about 3196 estates for sale in October 2005 in 156 different neighbourhoods[23]. There is an implicit bias due to the area where the association works. This reduces the generalization of the results even though covers great part of the city and guarantees good exploratory results with strong indication of behaviour.

Information collected for each estate includes (table 1) a) neighbourhood; b) neighbourhood classification according to IPEAD/UFMG (see page 5, item 2.2)[24]; c) price (in reals R$); d) area (in square meters); e) type [house, apartment, penthouse]; f) number of bedrooms; g) number of bathrooms; h) number of parking spaces; i) number of pools in the building; j) presence of sports court in the building; k) presence of garden or porch; l) age of the estate, m) fee paid for city tax (IPTU), [monthly, in reals] and n) fee for building services (monthly, in reals).

As it will be shown, not all observations contain information on all variables. The main model was sampled with 510 observations in which all of them contained the needed variables. This reduction was essential due to the small number of observations that contained information on IPTU (city tax) considered to be fundamental to the model. The detailing of the model is made in the sequence.

Concerning houses and penthouses only price and area were available. In relation to the age of the estate, the information was unclear and it was not possible to discern f "0" meant a new launching or absence of the information. Therefore, aiming at the quality of the database, all analyses were made for estates that were at least one-year old (802 observations). The study in practice is of used real estate market.

[23] The city of Belo Horizonte has a total of 267 neighbourhoods according to the City.
[24] This is the first variable that intends to represent the neighbourhood. In the second model it will be replaced by the index constructed using PCA.

Table 1 – Number of observations for each variable considered

Variable	Observations
Area (surface)	3119
Neighbourhood	3196
Bathrooms	1875
IPEAD classification	3152
Age	802
IPTU (city tax)	541
Gardens	1923
Number of bedrooms	1923
Number of parking spaces	1900
Pool	1923
Price	3195
Price by square meter	3079
Sports court	1922
Type	3196
Building services fee (COND)	816
Porch	1579

Source: author elaboration

It is easy to see that there sales advertisements are biased as 71% of the sample is of estates available at neighbourhoods classified as luxurious (3) or high luxurious (4) by IPEAD. Hence, on average the neighbourhood classification is 2.92 while if one considered all 267 neighbourhoods of the city the value would be 1.64[25]. This is probably due to the formalization process that is higher among more valued estates. Sales of cheaper apartments would be more likely to be negotiated without brokers' interference. As a result, average price of all observations (3,196) is fairly high for Brazilian standards at R$ 245,537 (US$ 126,565)[26]. The average area of flats is 183.98 m^2. Therefore, the average price of one square metre is R$ 1,447 (US$ 745.88). Among the researched flats, 73% have one or two parking space and the average age is 12.75 years. Average City tax (IPTU) is R$ 85 (US$ 43.81) and average building fee service (COND) is R$ 305 (US$ 157.21).

3.2 Sample description

When making the models only a) neighbourhood; b) classification of IPEAD (model 1) and PCA index (model 2); c) area; d) City tax (IPTU); e) building service fee (COND) and f) age of the building were used. The estates that contained information on all those variables amounted to 510 observations (table 2). They are located in 71 different neighbourhoods; have an average price of R$ 184,000 (US$ 94,845) and 121 square metres, which make an

[25] Standard-deviation of 0.92 for the sample of 156 neighbourhoods and 0.99 for all 267 neighbourhoods.
[26] Exchange rate of 24 May 2007 of R$ 1.94 to buy US$ 1.

average price of the square metre of R$ 1,522 (US$ 784.54). Average age of the estate is 14 years, it is located in a neighbourhood classified by IPEAD as luxurious and it pays on average R$ 300 (US$ 154.64) of building service fee and R$ 90 (US$46.39) of City tax.

Table 2 – Sample description

	Average	Standard-deviation	Maximum	Minimum
IPEAD classification	3.14	0.96	4	1
Price (R$)	184 024	161 139	1 200 000	30 000
Area (m2)	120.88	57.69	450	40
Age (years)	14.31	10.05	50	1
City tax (IPTU monthly)	88.10	73.76	400	12.76
Building service fee (COND)	299.08	242.66	1 500	25
Number of parking spaces	1.75	0.95	6	0

Source: author elaboration

3.3 PCA income index construction and geoprocessing – model 2

The index constructed by principal components methodology used ten income variables taken straight from census tract database[27] (available at IBGE – Brazilian Institute of Geography and Statistics). They were: v0145 (no income); v0136 to v0144 (that corresponds to the following levels of income: 0 to 0.5 minimum wage salaries; 0.5 to 1; 1 to 2; 2 to 3; 3 to 5; 5 to 10; 10 to 15; 15 to 20 and above 20). In the linear combination that resulted – the principal component – the three variables that corresponded to the higher income levels (above 10 minimum-wages) were positively included with coefficients of: 0.287; 0.3178 and 0.2502; those with income between 5 and 10 minimum-wages were multiplied by 0.0209 (close to nothing) and the remaining levels of income were added with negative coefficients of around 0.4. ArcGIS software presents results in raster format[28].

Afterwards it was necessary to use neighbourhoods' layer in ArcGIS provided by the City of Belo Horizonte so that the average per neighbourhood values could be extracted from the raster. In other words, a spatial analysis technique was applied to aggregate values at the level of the wished spatial unit. In practice what happened was the transformation of information initially available at census tract level into neighbourhood ones.

Note also that, in this case, there was a transformation and a synthesis of spatial variables in which information available at census tract level was aggregated and presented by neighbourhoods. The process allowed the construction of an income rank of all 286

[27] See Brasil (2003).

[28] Raster is a matrix georeferenced format used to form an image in which to every squared cell, a pixel, is associated a certain value. To further details check Druck and colleagues (2004) or geoprocessing and spatial analysis manuals, such as: Lillesand, Thomas M., & Ralph W. Kiefer. **Remote Sensing and Image Processing**. John Wiley and Sons, 1987 and Richards, John A. **Remote Sensing Digital Image Analysis: An Introduction**. Berlin: Springer–Verlag. 1986.

neighbourhoods of Belo Horizonte[29]. This way, all neighbourhoods have their specific value at the ranking and the representability of the neighbourhood is much more precise than the simple division in four categories supplied by IPEAD's classification.

In the regression analysis in which the PCA index was used, the value to each specific neighbourhood was applied to the observation. For the model with 510 observations the following basic statistics were calculated (table 3):

Table 3 – Basic statistics for variable PCA index

variable	obs.	average	standard-deviation	minimum	maximum	unique values
PCA index	510	0.5963	0.1737	0.2440	0.9164	67

Source: author elaboration.

3.4 Spatial division: Neighbourhoods

The City of Belo Horizonte applies in its municipal planning the subdivision called Regional (9) and Planning Units (81). Brazilian equivalent of Census Bureau – IBGE –divides the municipality in 2,561 census tract (2000) in turn . Although officially recognized, neighbourhood division is not usually used for planning, statistical areas or analysis. Real estate market, however, has traditionally used neighbourhoods' names to qualify the estate being advertised and sold. It is also used as reference by the citizens who associate them with positive and negative attributes.

In terms of scale, neighbourhoods are adequate because they are smaller than Regional units, but not as small as the census tract for which the information was not available[30].

3.5 Model 1 – using neighbourhood classification of IPEAD

In the OLS traditional analysis, as well as in the spatial one, the functional form applied was semi-logarithmic, formally:

Equation 10 – Model 1

$$lnPrice = \alpha + \beta_1 lnArea + \beta_2 lnIPTU + \beta_3 lnCond + \beta_4 lnAge + \beta_5 IPEAD + \beta_6 Parking + \varepsilon, \quad (10)$$

where:

lnPrice is the natural logarithm (ln) of the price in Reals; lnArea is the logarithm of the area of the flat in square metres; lnIPTU is the natural logarithm of the monthly City tax in Reals; lnCond is the natural logarithm of the monthly building services fee paid in Reals; lnAge is the natural logarithm of the age of the flat in years; IPEAD is the value (between 1 and 4) of

[29] This table is available upon request of the author.
[30] Addresses of the flats were not availble. Only the neighbourhood in which they were located

the position of the neighbourhood on IPEAD's classification and Parking is the number of parking spaces (scalar) per flat.

3.6 Model 2 – using the PCA index

For model 2, the equation is the same of 10, except for the item IPEAD that is replaced by the PCA index, described above (table 3).

4 Tests and results

Some basic tests will be presented to warrant the quality of the proposed models. The semi-logarithm model is common in the literature and is used by Neto (2002) and Hermann and Haddad (2005) in Brazilian analysis of real estate market and it have the added feature of bringing the residues close to normal. Indeed, Shapiro-Wilk test indicates normality rejecting the null hypothesis (H_0). In order to test the presence of outliers, besides visual inspection, an inter-quartile range test was applied and its results did not account for any severe outlier. Breusch-Pagan/Cook-Weisbert tests indicated the variance of the residues is homogeneous and there is no heteroskedasticity.

Multicollinearity of the data was checked using the variance inflation factor – VIF, which rule of thumb indicates that results above 10 denote high multicollinearity. For the models proposed there was no indication of multicollinearity for the OLS regression made (the highest result was lnCOND 3.10 and the average was 2.31). A matrix of correlation was also made and analysed and the highest correlation with price – the dependent variable – were the values of City tax (IPTU) at 0.7880 and building service fee at 0.7951.

Specification tests were also conducted and they indicated that there was good fit and adjustment. Ramsey test, for instance, accepted the hypothesis that the model does not contain omitted variables.

The chosen weight matrix was the spatial distance given that the spatial configuration of the sample is organized in the following way: a) there is a varying number of observations by neighbourhoods; b) all observations were located in the geographical centre of the neighbourhood (as there was no addresses available); c) there are neighbourhoods of varying size and therefore some are closer to the next one them others. That is why, the matrix that better captured the neighbours influence was the spatial distance considered as the ad hoc limit that guaranteed no observation was left without a neighbour.

Furthermore, given the non-contiguity of the neighbourhoods, queen and rook weight matrixes were not recommended[31]. Matrixes of the k-nearest neighbours are not symmetric and have not been proved to be adequate for spatial lag estimation (ANSELIN, 2005, p. 220)[32].

Results of the traditional analysis (table 4) confirm that, although location is important, the attributes of the estate are determinant for price formation. All coefficients are significant and have the expected signal. According to the proposed model, the surface of the estate (ln area) is the most impactante in price formation. A raise of 1% in the surface of a flat would trigger a rise of 0.67 % in sales price. Services available in the building which make the monthly fee higher is the second factor to mostly influence the price, indicating that a raise of 1% in the fee leads to an increase in price of approximately 0.27%. Concerning the age of the estate, as expected, when there is an increase of 10% at the age of the building, final price reduces nearly 1%. Finally, one more parking space raises the price in 14% on average.

Considering the variable responsible for spatial control, the neighbourhood ranking classification of IPEAD, the results indicate that a change of one category, for example, from average to luxurious, would raise final sales price by 11%.

Table 4 – Results of model 1, OLS analysis

Estatistics: Adjusted R^2 = 0.8896; F(6, 503) = 684,79; 510 obs.				
Variables	**Coefficientes**	**Standard-deviation**	**t of student**	**PROB. > |T|**
LN AREA	0.6719	0.0437394	15.36	0.000
LN IPTU	0.1450	0.0254278	5.7	0.000
LN COND	0.2728	0.0248425	10.98	0.000
LN AGE	-0.0912	0.118035	-7.73	0.000
IPEAD	0.1126	0.0149485	7.53	0.000
PARKING	0.1385	0.0170757	8.11	0.000

Source: author elaboration, software STATA, available at CEDEPLAR/UFMG laboratory.

4.1 Model 1 – spatial diagnoses

Before validating the results presented it is necessary, as recommended by the literature, to test for spatial models.

The statistic Moran's I indicated the presence of spatial autocorrelation, as expected (figure 1), and the results of the spatial diagnosis (table 5) suggest the use of spatial error model (see equations 4 and 5).

[31] However, tests were run with all of them.
[32] Yet, tests revealed close results, regardless of the matrix used which confirms the robustness of the model.

The expected value of Moran's statistic equals – [1/(n-1)]. For this paper, since n = 510, was there no spatial auto-correlation, Moran's I value would be expected to be around – 0.0020. Hence, the result of 0.3754 indicates strong correlation with observations concentrated at the second quarter (high-high) and fourth one (low-low). Furthermore, the analysis of pseudo-significance, using the process of permutations (9999), has shown that the statistic is highly significant (p = 0.0001), with an average of – 0.0021 and standard-deviation of 0.0099. The envelope method[33] also confirmed the validity of the results.

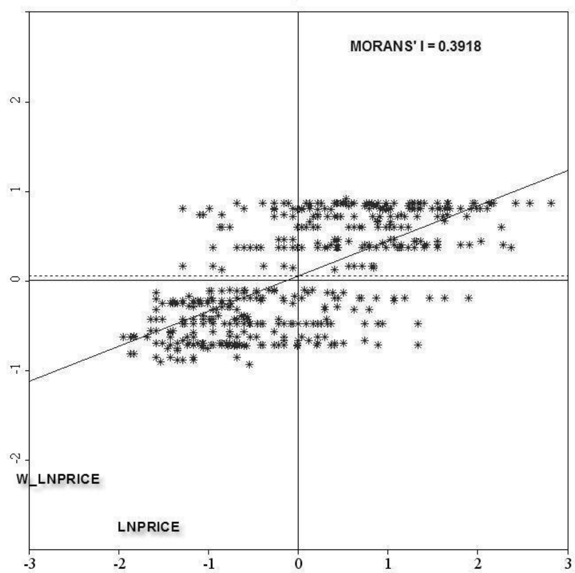

Figure 1 – Moran's I statistic for variable lnprice
Source: author elaboration, software GeoDa.

The results of Log-likelihood, Akaike Information Criterion[34] and Schwarz Criterion are used to enable comparison among spatial models. In practice, this means that for Log-likelihood the highest the number, the best the fitness. For the criteria the opposite is true, meaning that the smallest the value, the best the fitness.

Table 5 – Results of tests of model 1 and diagnosis for spatial dependence

[33] See Anselin (2005).
[34] AIC = -2L + 2K, where L is Log-likelihood and K is the number of parameters used in the model. For further details, see Anselin, 1988.

Tests	Values	Prob.
Log-likelihood	-19.8565	
Akaike Information Criterion	53.7129	
Schwarz Criterion	83.3538	
Multicollinearity Condition Number	55.50054	
Test Jarque-Bera	6.47195	0.0393219
Test Breusch-Pagan	19.19366	0.0038488
Koenker-Basset	15.49534	0.0167349
Test White	173.9153	0.0000000
Diagnostics for spatial dependence		
Lagrange Multiplier (lag)	13.6356271	0.0002219
Robust LM (lag)	6.0663148	0.0137784
Lagrange Multiplier (error)	30.2931197	0.0000000
Robust LM (error)	22.7238075	0.0000019
Lagrange Multiplier (SARMA)	36.3594345	0.0000000

Source: author elaboration, software GeoDa.

Other indications provided by the results are the multicollinearity condition number, which rule of thumb suggests that numbers higher than 30 are problematic. The result of the model was 55.63, even though the composition of the model was made so that it could be reduced to a minimum. Jarque-Bera test confirms the normality of the errors and Breusch-Pagan e Koenker-Basset tests reject the presence of heteroskedasticity. White test presents different information suggesting its presence. Anselin (2005, p. 177) reminds us that heteroskedasticity tests are very sensitive to the presence of spatial dependence and that for large samples Breusch-Pagan and Koenker-Basset are adequate[35].

Spatial dependence diagnostics also shows that "… in the rare instance that both [lag model and error model] would be highly significant, choose the model with the largest value for the test statistic …" (ANSELIN, 2005, p. 218)[36].

Table 6 contains the results of spatial error model. The values of the coefficients, which were relatively stable throughout the exercise, are a little different in magnitude. Elasticity price-building service reduces its influence slightly from 0.27 to 0.247 and there is a small elevation on the coefficient of parking spaces. This means that when taking into account spatial influence the services provided by the building decrease its importance favouring specific attributes of the apartment itself (such as quality of finishing) and the locational factor. In other words, effects of the neighbourhood were being captured by the building services coefficient. In fact, the spatial model presents a much better adjustment *vis-à-vis* the

[35] For safety, model 2 was estimated using an heterocedastic error model using Feasible or Estimated Generalized Least Squares (FGLS or EGLS). Results were very close to those of table 8.

[36] Once again, ir order to check the robustness of the model a spatial lag regression was applied but the statistics for log-likelihood and the criterion presented values that indicated a worst fitness.

traditional OLS model as captured by the values of the log-likelihood (from -19.85 to -11.96); AIC (from 53.71 to 37.92) and SC (from 83.35 to 67.56).

It is important to highlight, however, that the presence of a significant lambda indicates that the specification of the model is not complete and that important spatial characteristics are not being modelled. This leads to the conclusion that IPEAD's neighbourhood classification by itself is not enough to explain the price variance among neighbourhoods which induces the search for a more detailed description of the neighbourhoods which is the main objective of this model.

Table 6 – Results of model 1 with spatial error

Tests	Values	Prob.
Log-likelihood	-11.963751	
Akaike Information Criterion	37.9275	
Schwarz Criterion	67.568376	
Tests Breusch-Pagan	24.1121	0.000498
Spatial dependence diagnosis		
Likelihood Ratio Test	15.78545	0.0000709

Statistics: $R^2 = 0.895085$; $\sigma^2 = 0.06881$; 510 obs.				
Variables	**Coefficients**	**Standard-deviation**	**Value of Z**	**Prob.**
LN AREA	0.6696382	0.04242995	15.78221	0.0000000
LN IPTU	0.1439533	0.02478086	5.809051	0.0000000
LN COND	0.2476281	0.02483174	9.972242	0.0000000
LN AGE	-0.09782459	0.01155754	-8.464135	0.0000000
IPEAD	0.1184579	0.01871408	6.329881	0.0000000
PARKING	0.12417038	0.01703637	8.317725	0.0000000
LAMBDA	0.4896931	0.1178465	4.155349	0.0000325

Source: author elaboration, software GeoDa.

The next step is to detail information at the neighbourhood level so that the power of explanation of the model increases.

4.2 Model 2 – classic analysis and spatial diagnosis

A second model was implemented in order to better capture neighbourhood information. As it was described, IPEAD neighbourhood classification ranking had only four categories. Using the PCA index, every neighbourhood will have their specific value.

As it can be seen in table 7, this model has a better adjustment than the previous one[37], having all significant coefficients and slightly different. It is clear that as spatial configuration is better depicted – in this case, with a more faithful index – in general, the coefficients have their values reduced.

Once more, the spatial dependence diagnosis recommends a spatial error model that is applied in the sequence.

[37] Expressed by *log-likelihood*, AIC and SC indexes.

Table 7 – Results OLS model 2 and spatial diagnosis

Statistics: R^2 adjusted = 0.900637; F (7, 503) = 769.938; Prob F = 0; 510 obs.				
Variables	**Coefficients**	**Standard-deviation**	**t of student**	**Prob.**
LN AREA	0.669	0.041387	16.16218	0.0000000
LN IPTU	0.133	0.024149	5.488307	0.0000001
LN COND	0.236	0.024070	9.804033	0.0000000
LN AGE	-0.094	0.011202	-8.364138	0.0000000
PCA INDEX	0.882	0.080971	10.89738	0.0000000
PARKING	0.140	0.016166	8.63495	0.0000000

Tests	Values	Prob.
Log-likelihood	6.93388	
Akaike Information Criterion	0.132235	
Schwarz Criterion	29.7731	
Multicollinearity Condition Number	55.39	
Test Jarque-Bera	8.713438	0.0128204
Test Breusch-Pagan	14.51	0.0244301
Koenker-Basset	11.33058	0.0786813
Test White	147.1872	0.0000000
Spatial dependence diagnosis		
Lagrange Multiplier (lag)	0.1405153	0.7077693
Robust LM (lag)	2.4064662	0.1208349
Lagrange Multiplier (error)	**21.9410448**	**0.0000028**
Robust LM (error)	24.2069957	0.0000009
Lagrange Multiplier (SARMA)	24.347511	0.0000052

Source: author elaboration, software GeoDa.

4.3 Model 2 – spatial analysis

Spatial error model, using PCA index, presents the best adjustment among all tested. There is a further movement of coefficients reduction and a raise in the Parking value. Lambda is still significant albeit with a smaller z-value. This strongly suggests that the model is better capturing spatial influence[38].

[38] Model SARMA was also tested using a generalized moment method two-step and iterated. But, the results were consistently similar with difference of the order of 0.005. **Wy** was not significant at 5%. The biggest difference occured in the values of the coefficients of PCA index (-0.04); building service fees (+0.03) and Parking (-0.015).

Table 8 – Results spatial error– model 2

	Values	Prob.
Log-likelihood	13.139767	
Akaike Information Criterion	-12.2795	
Schwarz Criterion	17.36134	
Testes Breusch-Pagan	20.24231	0.0001294
Spatial dependence diagnostics		
Likelihood Ratio Test	12.41177	0.0004266

Statistics: R^2 = 0.904840; σ^2 = 0.055220; 510 obs.

Variables	Coefficients	Standard-deviation	Z Value	Prob.
LN AREA	0.66793	0.04035	16.55151	0.0000000
LN IPTU	0.13369	0.02363	5.657413	0.0000000
LN COND	0.22244	0.02389	9.311671	0.0000000
LN AGE	-0.09897	0.01101	-8.991497	0.0000000
PCA INDEX	0.97956	0.09882	9.912638	0.0000000
PARKING	0.14497	0.01622	8.939214	0.0000005
LAMBDA	0.46669	0.12187	3.82951	0.0001284

Source: author elaboration, software GeoDa.

Numerically, to illustrate the model, it can be said that, on average, a raise of 1% in the flat's surface would cause a raise of around 0.67% in its final price. Another relevant factor is the quality of services offered in the building that is approximately measured by the fee paid (COND). The model indicates that a raise of 1% in this fee makes the apartment 0.22 % more expensive. Age of the estate, as expected, reduces sales price in a proportion that a 10% increase in age leads to a reduction of 1% in the final price. One extra parking space raises the price of the flat by 14.5%, on average. A worsening of building construction quality of 1%, approximately, measured by City-tax (IPTU) would reduce the price 0.13%. And finally, a flat with the same characteristics but located in a different neighbourhood, for instance moving from downtown (PCA index 0.456219) into a newly affluent neighbourhood, called Buritis (0.567284), would make it more expensive by 11%.

The results also enable comparison among observations which might help likely buyers in doubt. Take an apartment from the sample: a three bedroom apartment located in Buritis with two parking spaces, 6-years old and 110 sq. m. of surface (table 9) with current price at R$ 115,000 (US$ 59,278). Had it been located downtown, its price would have dropped to R$ 102,350 (US$ 52,758). Had it been 10 sq. m. larger, its value would have risen to R$ 121,983 (US$ 62,878). One more parking space would take the price upwards to R$ 131,672 (US$ 67,812).

Table 9 – Illustrative numerical comparison

Price R$	R$ 115,000 (US$ 59,278)	Had the same characteristics, but...	New price
Neigbourhood	BURITIS (0.567284)	*was located downtown (0.456219)*	**R$ 102,350 (US$ 52.578)**
Area (sq. m.)	110	*had 120 sq. m.*	**R$ 121.983 (US$ 62.878)**
Number of parking	2	*had 3 parking spaces*	**R$ 131.672 (US$ 67.812)**
Age (years)	6	*was 7 years old*	**R$ 113.103 (US$ 58.301)**
City-tax R$ (IPTU)	73	*had a City-tax of IPTU de R$ 64*	**R$ 113,105 (US$ 58.302)**
Building fee R$ (COND)	120	*had a building-fee of R$ 132*	**R$ 117.558 (US$ 60.596)**

Source: author elaboration.

4.4 Model 2 – quantilic analysis

Model 2 with quantilic analysis is only exploratory as it is not modelled spatially[39]. However, it does contribute to the understanding of the variables chosen as it clearly shows that coefficient values are distinctly different from different quantis.

The results presented (table 9) are significant[40] and consistent with the theoretical analysis. Note that the importance of the size of the apartment decreases as its price rises (figure 2). Parking spaces and age are comparatively less important for those estates with intermediate prices. The highlight of the quantilic analysis is the behaviour of the PCA index that represents the location in the composition of final prices. It rises consistently as it does the price of the estate. More expensive apartments pay proportionately more (in relation to cheaper ones) for more privileged location[41].

[39] The possibility of making a spatial and quantilic analysis simultaneously is still under investigation.
[40] Except for the values of City-tax (IPTU) for quantil 10%.
[41] Note that at the highest price (q9) 1.104 the value is much higher than that at the inferior one (q1) 0.638. Actually, the parameter is always crescent.

Table 10 – Model 2 – quantilic analysis

	(q.1)	(q.25)	(q.5)	(q.75)	(q.9)
	lnprice				
lncond	0.305	0.228	0.213	0.259	0.252
	(6.68)**	(7.09)**	(8.21)**	(6.88)**	(6.00)**
lnarea	0.674	0.637	0.636	0.634	0.593
	(10.19)**	(11.02)**	(14.33)**	(11.28)**	(9.97)**
lniptu	0.079	0.153	0.197	0.157	0.09
	-1.71	(4.37)**	(7.55)**	(3.89)**	(1.99)*
ln age	-0.105	-0.06	-0.071	-0.097	-0.095
	(4.95)**	(3.75)**	(5.92)**	(5.85)**	(6.36)**
parking	0.169	0.154	0.12	0.122	0.164
	(4.69)**	(6.30)**	(6.91)**	(5.53)**	(7.51)**
PCA index	0.638	0.827	1.008	1.022	1.104
	(5.00)**	(7.93)**	(10.98)**	(7.57)**	(7.18)**
Constant	6.011	6.345	6.34	6.39	7.013
	(27.23)**	(33.80)**	(43.23)**	(34.76)**	(37.65)**

510 obs.

Absolute value of t in parenthesis

* significant at 5%; ** significant at 1%

Source: author elaboration, *software* Stata, laboratory of CEDEPLAR/UFMG

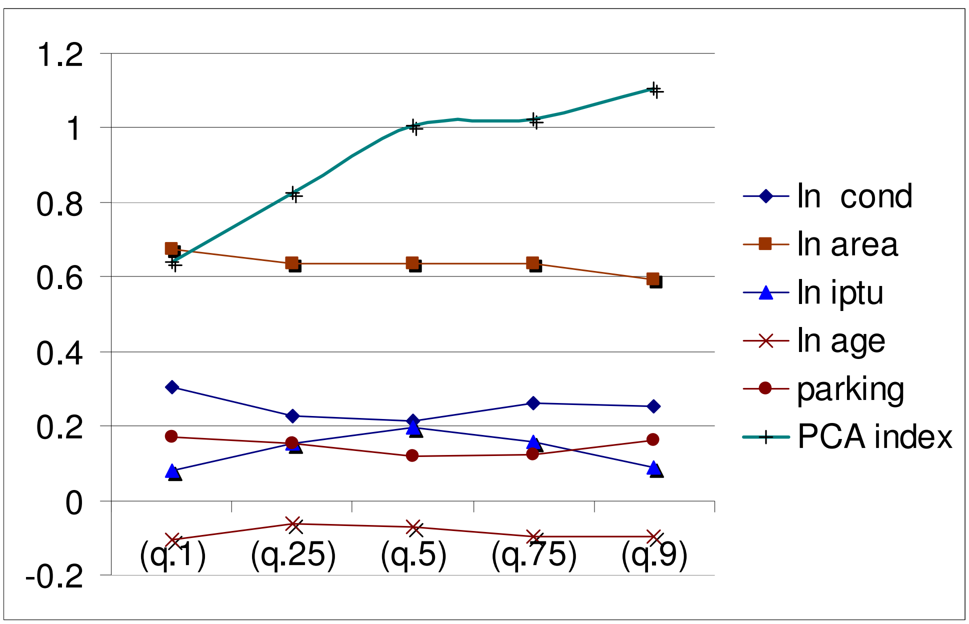

Figure 2 – evolution of the coefficients of the quantilic analysis
Source: author elaboration, based on the results of table 9.

5 Final considerations

The discussion of real estate within national parameters is extremely relevant considering the importance and weight of housing in families' budgets. However, current estimations used by brokers and real estate agencies are highly discretionary. It is interesting for citizens, urban actors and agents that the connection between price and advantages provided by them be better delineated. Simultaneously, the availability of georeferenced data in quality and quantity unimaginable before, at the scale level of the building may now be measured and included in models.

It should be pointed however that although this paper presents good indications of relevant variables and consistency of the estimated parameters, for its complexity, the model is still incomplete. Advances can be made when data from the City concerning real transactions that take place, instead of estimated sales price are used. Temporal analysis can also indicate change of behaviour in time that will help clarify the influence of changes in urban patterns and qualify it. These changes could be public or private urban equipment being supplied, shopping areas being spread or change of average composition of neighbourhoods.

Finally, this paper fulfils its objective of "identifying patterns and characteristics of real estate prices in Belo Horizonte. It presents strong indications of spatial dependence and that a simple 4-level category of neighbourhoods is not enough to explain spatial differentiation. Furthermore, it highlights the importance of the size of the apartment, its type and number of parking spaces in the price and allows for a good comparison among different neighbourhoods. Throughout the paper there has been good consistency of coefficients estimated.

6 References

ALMEIDA, E. S. **Curso de econometria espacial aplicada**. Piracicaba: ESALQP-USP. 2004

ANSELIN, L. **Spatial econometrics: methods and models**. Dordrecht: Kluwer Academic. 1988. 284 p.

ANSELIN, L. **SpaceStat tutorial: a workbook for using SpaceStat in the analysis of spatial data**. University of Illinois. Urbana-Champaign, p.263. 1992

ANSELIN, L. **Exploring spatial data with GeoDaTM: a workbook**. Urbana-Champaign, IL: Spatial Analysis Laboratory, Department of Geography, University of Illinois. 2005. 244 p.

ANSELIN, L. e BERA, A. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: A. Ullah e D. E. A. Giles (Ed.). **Handbook of Applied Economic Statistics**. New York: Marcel Dekker, 1998, p.237-289.

BRASIL. Censo demográfico 2000: agregado por setores censitários dos resultados do universo. Documentação do arquivo. M. D. Planejamento e I. B. D. G. E. Estatística**:** 157 p. 2003.

BRUECKNER, J. K., THISSE, J.-F., *et al.* Why is central Paris rich and downtown Detroit poor?: An amenity-based theory. **European Economic Review**, v.43, n.1, p.91-107. 1999.

BUCHINSKY, M. Recent advances in quantile regression models. **Journal of Human Resources**, v.33, n.1, p.88-126. 1997.

CAN, A. e MEGBOLUGBE, I. Spatial Dependence and House Price Index Construction. **Journal of Real Estate Finance and Economics**, v.14, p.203-222. 1997.

CULLEN, G. **Townscape**. London. 1961. 315 p.

DRUCK, S., CARVALHO, M. S., *et al.* **Análise espacial de dados geográficos**. Brasília: Embrapa. 2004. 209 p.

HARVEY, D. **Social justice and city**. Baltimore: Johns Hopkins University. 1973. 336 p.

HERMANN, B. M. e HADDAD, E. A. Mercado Imobiliário e Amenidades Urbanas: A View Through the Window. **Estudos Econômicos** v.35, n.2, Abril-junho, p.237-269. 2005.

JOHNSON, R. e WICHERN, D. **Applied multivariate statistical analysis**: Prentice Hall. 1998. 799 p.

KOENKER, R. e BASSETT, G. Regression Quantiles. **Econometrica**, v.46, n.1, January, p.33-50. 1978.

LYNCH, K. e CAMARGO, J. L. **A imagem da cidade**. São Paulo: Martins Sources. 1997. 227 p.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada**. Belo Horizonte: Editora UFMG'. 2005. 297 p.

MONTE-MÓR, R. L. D. M. O que é o urbano no mundo contemporâneo. **Textos para discussão**. 281**:** 14 p. 2006.

NAIK, D. N. e KHATTREE, R. Revisiting olympic track records: some practical considerations in the principal component analysis. **The American Statistician**, v.50, n.2, p.140-144. 1996.

NETO, E. **Estimação do preço hedônico: uma aplicação para o mercado imobiliário da cidade do Rio de Janeiro**. (dissertação de mestrado). EGFV, Rio de Janeiro, 2002.

ROSSI, A. **A arquitetura da cidade**. São Paulo: Martins Sources. 2001. 309 p.

SANTOS, M. **Espaço e método**. São Paulo: Nobel. 1992. 88 p.

SHEPPARD, S. Hedonic analysis of housing markets. In: P. Cheshire e E. S. Mills (Ed.). **Handbook of regional and urban economics**. Amsterdam: North-Holland, v.3, 1999, p.1595-1636.

SINGER, P. O Uso do Solo na Economia Capitalista. In: E. Maricato (Ed.). **A produção capitalista da casa e da cidade no Brasil industrial**. São Paulo: Alfa-ômega, 1982, p.21-36.

SMIRNOV, O. e ANSELIN, L. Fast maximum likelihood estimation of very large spatial autoregressive models: A characteristic polynomial approach. **Computational Statistics and Data Analysis**, v.35, p.301-319. 2001.

SOJA, E. W. **Geografias pós-modernas: a reafirmação do espaço na teoria social critica**. Rio de Janeiro: Jorge Zahar. 1993. 324 p.

SOJA, E. W. **Third space: expanding the geographical imagination**. Oxford: Blackwell Publishers. 1996. 354 p.

TOBLER, W. R. A computer movie simulating urban growth in the Detroit region. **Economic Geography**, v.46, p.234-240. 1970.

TUAN, Y. **Topofilia: um estudo de percepção, atitudes e valores do meio ambiente**. São Paulo. 1980. 288 p.

TUAN, Y. **Espaço e lugar: a perspectiva da experiência**. São Paulo: DIFEL. 1983. 250 p.

WHITTLE, P. On stationary processes in the plane. **Biometrika**, v.41, p.434-449. 1954.

WOOLDRIDGE, J. M. **Econometric analysis of cross section and panel data**. Cambridge, Mass.: MIT Press. 2002. 752 p.