

Furtado, Bernardo; Van Oort, Frank

Conference Paper

Neighborhood weight matrix in a spatial-quantile real estate modeling environment: Evidence from Brazil

50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden

Provided in Cooperation with:

European Regional Science Association (ERSA)

Suggested Citation: Furtado, Bernardo; Van Oort, Frank (2010) : Neighborhood weight matrix in a spatial-quantile real estate modeling environment: Evidence from Brazil, 50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/118893>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Neighborhood weight matrix in a spatial-quantile modeling environment: Evidence from Brazil¹

Bernardo A. Furtado ^{a, b, *}, Frank van Oort ^b

^a *Institute of Applied Economic Research IPEA, Brazil*

*Postal Address: SBS Quadra 1, Edifício BNDES, sala 317
Brasília - DF
70076-900
BRAZIL*

^b *Urban and Regional Research Centre, Utrecht University, The Netherlands*

*Postal Address: P.O. Box 80115
3508 TC, Utrecht
The Netherlands*

*E-mail: Bernardo A. Furtado bernardo.furtado@ipea.gov.br
Frank van Oort f.vanoort@geo.uu.nl*

** corresponding author: Bernardo A. Furtado
Phone: (+55 61) 8131 8810
Fax: (+55 61) 3315 5412*

Abstract: *This paper analysis two central influences on real estate urban market: (a) neighborhood as cognitively-perceived identity units, and (b) heterogeneity of preferences of households. In doing so, two methodological changes are applied. Firstly, a neighborhood spatial matrix is proposed and compared to regularly used matrices and, secondly, a spatial-quantile regression is tested. The results highlight the fact that spatial influence brought to regression models through weight matrix should be carefully used, and that the matrix's choice might carry unobserved and unwanted effects into the estimation. Results also demonstrated that using information of neighborhood identity can optimize the understanding of city's complex influence on real estate markets. Finally, the quantile estimation should always be tested against in real estate estimation, as preferences of households seem to differ significantly for different levels of prices.*

JEL classification: R20, C21

Keywords: neighborhood, weight matrices, real estate market, spatial-quantile regression

¹ We would like to acknowledge the advice and the provision of the estimation program by Professor Zhenlin Yang, School of Economics Singapore Management University and Professor Liangjun Su, Guanghua School of Management, Peking University.

Introduction

Urban real estate price determination has been subject of extensive theoretical and empirical research for quite some time (Sheppard, 1999). Advances in modeling have accordingly emerged (Capello and Nijkamp, 2004). In the late 1990s, previous work on spatial effects (by notably Cliff and Ord, 1973; and Paelinck and Klaassen, 1979) was systematically organized and it became clear that the inclusion of spatial dependence is a crucial precondition for empirical hedonic price studies (Anselin, 1988). In practice, this has resulted in two types of model specifications: (a) specifications considering distance to a Central Business District (CBD) and to other points of interest, such as concentrations of work as the main spatial determinant; and (b) those that apply weight matrices in its most common form of spatial contiguity. We argue in this paper that both conceptualizations do not capture the full influence of urban location on prices. The empirical research approach used intends to shed light on the behavior of the real estate market, in which we specifically considered neighborhood identity and influences of urban fabric separate from influences of the estate itself. We also observe differences in behavior correlated with prices of estates.

First, in an increasingly mobile (Muhammad, 2007), decentralized (Garreau, 1992) and mixed-use (Wheaton, 2004) pattern of urban occupation, distance to CBD loses explanatory power. Moreover, the scale of analysis hampers the specification of what exactly the center of a large metropolis is, making definition arbitrary. Also, proximity relationships may be more relevant to price forming than the distance to the center.

Second, contiguity or distance matrices usually applied in spatial econometric analyses can also be considered abstract and *ad hoc* in nature. The contiguity matrix usually uses Delaunay triangulation to determine the neighbors of a given locational observation. The distance matrix captures a radius of a certain distance (cut-off point) chosen by the researcher to define individuals within reach as neighbors. This procedure might inaccurately include an observation on the other side of a river or a highway as neighbors.

One way to deal with these issues is to use landscape theory, as proposed by the architects Kevin Lynch (1960) and Aldo Rossi (1966). They argue that the perception of the city by its citizens is spatially circumscribed within a known entity. A person relates positive and negative qualities (attributes) to certain spatial portions of the territory that are collectively recognized by a community in urban space (Tversky, 2003). This conceptualization is in accordance with sales practices of real estate agents, who specify the

neighborhood when advertising a sale. Accordingly, the approach hypothesizes that if the attributes are relevant, the neighborhood as a whole is positively (or negatively) valued. In this paper, we treat the neighborhood as a reference for analyzing the determinants of prices of houses. In our analysis, the configuration of neighbors – given by the weight matrix – explicitly considers social identity and cohesion in the choice of who is considered neighbors.

Our analysis applies other methodological improvements over previous studies as well. First, the determinants of housing prices are captured by a limited number of explanatory dimensions (factors) instead of numerous individual variables. Similar variables of the same dimension are weighed together. In addition to contributing to a better interpretation, this is also helpful in dealing with problems of multicollinearity that arise from the complex nature of urban space and the city when applying regression techniques. Principal component analysis (PCA) is used to construct indices of values that summarize the multiplicity of explanatory factors for each neighborhood.

Second, a homogenous treatment of the population of houses and their prices disguises important population-internal heterogeneity. It therefore has become essential in real estate analysis that different patterns of behavior and preferences for various levels of social characteristics are dealt with. As households at different positions on the social ladder might value a single attribute (such as proximity to a large shopping center) differently, segmentation of households should be internalized in the analysis. Methodologically speaking, we apply a quantile regression estimation (Koenker and Bassett, 1978; Buchinsky, 1997). This helps us not miss variation in responses to variables that are present in the full sample (Wooldridge, 2002). The initial study of Zietz (2008) applies this estimation technique, but it is introduced coincident to a spatially-lagged variable in the model rather than in a full spatial-quantile estimation model, such as in the model proposed by Su and Yang (2007). According to Anselin (2002), that procedure would be reasonable only when one considers spatial autocorrelation to be a nuisance. In our analysis, we want to incorporate spatial autocorrelation and quantile regression simultaneously.

Finally, this paper also aims to detail to which extent the inclusion of social heterogeneity and cognitively perceived neighborhoods impact real estate prices. Specifically, we want to test the influence of the following factors:

- (a) In relation to the surface area of the estate:
 - i. Is the positive influence of the estates' surface on prices constant throughout all social levels and neighborhood matrices?
 - ii. Regarding the quality of the estate: for whom is it more important?

- iii. Does the estate's age influence housing prices negatively and uniformly?
- (b) In relation to the neighborhood:
- i. How relevant is the quality of the neighborhood?
 - ii. Is the presence of commerce and services always a positive influence on housing prices?
 - iii. How negative are industry-heavy neighborhoods?
 - iv. Do neighborhoods with innovative services have a positive influence on housing prices?
- (c) In relation to proximity:
- i. Do large shopping malls, proximity to avenues, and arborization all have a positive influence?
 - ii. How does proximity to slums affect real estate prices?

Methodology

Spatial analysis and spatial econometric models

The essential concept to understand when working with spatial analysis is 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 statistical expression that measures this spatial dependence.

Concerning spatial statistical dependence, Anselin (2005) suggests that one first implements traditional models, tests for the presence of spatial autocorrelation, and then, if necessary, uses a spatial model.

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

Equation 1 – Moran's I

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

where W_{ij} is the weights matrix that reflects spatial dependence in 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 ways. These are crucial to the analysis and are discussed in the next section.

However, although Moran's I index is adequate for indicating the presence of spatial autocorrelation, it does not show in the form in which it occurs (Anselin, 2005, p. 197).

Spatial models should be applied when theoretical indications related to the research problem suggest that there is spatial influence among observations. In this paper, for instance, spatial-dependent modeling is considered because the price, quality and availability of services of nearby dwellings influence the values of the dwelling. According to the simple model of spatial process identification proposed by Anselin (2005, p. 199), if there is no detected spatial influence, one can stick to traditional models with more guarantees of their robustness.

One of the most cited econometric spatial models in the literature is the spatial lag, which includes a term of influence of neighbors in the regression process (LeSage, 1999):

Equation 2 – Spatial lag model

$$Y = \rho Wy + X\beta + \varepsilon,$$

where Y is the dependent variable; Wy is a spatial lag vector, derived from the choice of spatial weights matrix²; ρ is a spatial auto-regressive coefficient; X is the matrix of independent variables and their coefficients β ; and ε is the error vector.

Another very common model in spatial econometrics is the spatial error model³. When this occurs, the spatial effect implicit in the problem has not been totally included and is thus captured by the error.

Equation 3 – Spatial error model

$$Y = X\beta + u, \text{ and}$$

Equation 4

$$u = \lambda Wu + \varepsilon, \text{ where}$$

λ is the error auto-regressive spatial parameter⁴.

A full spatial model includes spatial autocorrelation as a variable and in the errors. Furthermore, it is a combination of the two presented models (LeSage, 1999).

Equation 5

$$Y = \rho Wy + X\beta + u, \text{ and}$$

² It is not the objective of this paper to discuss the underlying theory. For further details, see Anselin (1988).

³ The spatial error model was originally proposed by Whittle (1954).

⁴ Other spatial models include average moving spatial models, spatial Durbin models, mixed models with spatial lag, and error moving averages of the first degree.

$$u = \lambda Wu + \varepsilon.$$

The tests recommended for identifying spatial models are: (a) the Lagrange Multiplier tests for Lags, for lag-spatial models, and (b) the Lagrange Multiplier test for Error specifications to indicate the use of an error auto-regressive model. Finally, there is also the LM-SARMA test that indicates the need for a more complex model in which there is spatial autocorrelation in both the variable(s) and the errors.

Ordinary Least Squares (OLS) estimations are not appropriate for treating spatial analysis because the literature indicates that in the presence of spatial autocorrelation, OLS results are inconsistent (LeSage, 1999). The suggested alternative is to use a Maximum Likelihood (ML) estimator to estimate a parameter that would most likely be generated in the observed sample⁵ (Anselin, 1988; Anselin and Bera, 1998; Smirnov and Anselin, 2001; Anselin, 2005).

Weight matrices

Weight matrices play an essential role in spatial analysis, represent the influence of neighbors in each observation, and account for how this influence diminishes in space. Many authors insist that different weight matrices may alter results significantly and so it is essential to choose a matrix that is theoretically appropriate. Also, robustness tests for different specifications of weight matrices are recommended (Anselin, 1988; 1999; LeSage, 1999; Arbia, 2006).

The simplest models of weight matrices are models of spatial contiguity that indicate, through a binary relation, whether one is a neighbor. There are two basic types: the *queen* matrix that considers as neighbors all individuals which share a border or have at least a vertex in common; and the *rook* matrix in which only individuals who have a common border are considered to be neighbors. In both cases, it is possible to include higher order matrices so neighbors of neighbors might be considered to influence ‘own’ issues. A strong underlying theory of the phenomenon being studied should indicate the most appropriate choice.

Alternatively, there are matrices built on spatial distance, in which an arbitrary limit is specified and individuals within that limit are considered to be related and dependent on each other. This specification is naturally preferred, which is indicated in cases where the influence can be readily measured. There is also the matrix of the k-nearest neighbors in which neighbors – as the name suggests – are defined as individuals who are closest to the

⁵ According to Anselin, “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).

observation. Once more, this number k is often set in a rather arbitrary manner. In this paper we apply three different matrices, including one that captures a newly proposed neighborhood approach (see below).

Quantile regression analysis

Quantile regression analysis⁶ is based on points taken in regular intervals from the accumulated distribution function⁷ of a stochastic variable. The intervals denote sub-groups of the sample that can be of any order (a hundred, for centis). In this paper, the reference is the one proposed by Buchinsky (1997), who reinforces the need to verify whether the patterns observed in a traditional regression (OLS) repeat themselves in the quantile regression. If not, the quantile analysis adds valuable information to interpretation of the phenomenon. Moreover, it helps determine whether the studied behavior at higher/lower levels of the dependent variable differs. The quantile regression is not particularly sensitive to the presence of outliers (like OLS is) and it enables change analysis “at different points of the distribution”. Buchinsky (1997, p. 90) summarizes the relevance of quantile analysis, stating that “clearly, it is not enough to investigate changes in the mean when the entire shape of the distribution changes dramatically”.

Wooldridge (2002, p. 367) recommends quantile 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.

Spatial-quantile regression analysis

A recent theoretical advance to make a spatial and quantile model was proposed by Su and Yang (2007). Starting from a traditional quantile model,

Equation 6

$$Y_n = X_n \beta_{0\tau} + u_n,$$

⁶ See Buchinsky (1997) and Koenker and Bassett (1978) for details.

⁷ $F(x) = P(X \leq x)$, where $F(x)$ is the probability that the variable X presents value inferior or equal to x such that the probability of X is in an interval (a, b) is $F(b) - F(a)$ if $a \leq b$.

where $\beta_{0\tau}$ is a quantile regressor that can vary with the value of τ . They propose that

Equation 7

$$Y_n = \rho_{0\tau} W_n Y_n + X_n \beta_{0\tau} + u_n,$$

where $\rho_{0\tau}$ is a scalar spatial-lag parameter that varies according to τ (Su and Yang, 2007, p. 4)⁸.

The advantages of using this model is that one can consider different degrees of spatial dependence on the regressor at different points of the distribution.

However, Su and Yang (2007, p. 14) note that given the lack of a distributional assumption, inferences based on the results would demand a method for the calculation of the variance-covariance matrix that would “complicate the matter to a great deal”. A log-likelihood is not calculated for the same reason. In addition, calculation of an R-squared would make no sense when estimating using instrumental variables (Wooldridge, 2003, p. 471).

Su and Yang (2007) show through Monte Carlo experiments that the instrumental variable quantile estimation of spatial autoregressive models perform better than both the quasi maximum likelihood (QML) and the generalized method of moment (GMM). They also show that it is robust to the presence of outliers and heteroskedasticity. However, in terms of comparison with spatial and quantile models, one can only use the coefficient values.

Dataset

The dataset⁹ contains information on actual transaction prices for all 5,512 observations that were negotiated from June to August 2007 in the city of Belo Horizonte. The transactions are linked to a real estate city tax database that includes information on: (a) the property value for tax purposes (which may differ from actual transaction prices); (b) the year of construction of the dwelling; (c) the fraction of parcels ownership¹⁰; (d) finishing and estate quality levels; (e) the real estate use type (see Table 2); (f) a depreciation factor, which considers both quality of estate and years of construction; (g) a typology factor, which considers whether the original building has been subject to usage adaptations; (h) a commercialization factor, which considers whether the estate has been subjected to a change in tax value; (i) the average

⁸ The authors originally used the notation λ . However, to avoid confusion with the parameter for the error it was replaced with ρ .

⁹ The dataset of real estate households was provided by the Finance Department of the City Administration of Belo Horizonte.

¹⁰ For individual houses, this value is always 1. For apartments or multifamily units, the value is proportional to the number of domiciles within the parcel. The value is 0 if there is no building on the parcel.

square meter price of construction given building quality; (j) the built surface; (k) a topography factor; (l) the level of public infrastructure; (m) a pedology factor; (n) a parcel position factor; (o) the land square meter price; (q) the parcel size; (r) the City tax value; and (s) the property's geographic coordinates¹¹.

Descriptive statistics are presented in Table 1. The average price of commercialized estates during the period is around 51,000 Euros¹², with an average built surface of approximately 150 square meters an average age of twenty years. The properties are spatially well distributed within the municipality (Figure 1).

Table 1 – Descriptive basic statistics of estate's sample

Variable	Observations	Minimum	Maximum	Average	Standard-deviation
Transaction value (€)	5512	493	1401899	50934	62999
Fraction of parcels	5512	0	1	0.39	0.40
Built surface (sq. m.)	5512	2	1905	145	110
Annual city tax value (€)	5512	0	8126	694	391
Value of sq. m. construction (€)	5512	0	595	159	97
Value of sq. m. of land (€)	5512	8	1061	86	94
Value of sale based city tax (€)	5512	398	703608	34708	38508
Transaction value by surface (€/sq.m.)	5512	44	13202	362	361
Year of construction	5512	1924	2006	1986	14.5
Latitude (UTM)	5512	7785573	7811839	7798245	5015
Longitude (UTM)	5512	598612	617013	609021	3247

Source: GEAVI/PBH, 2007

The apartment unit typology is the dominant type in sample, accounting for almost 60% of all observations, followed by single family housing units (Table 2).

Table 2 – Number of observations by estate type

Typology	Observations	Percentage
Apartment	3233	0.587
House	1588	0.288
Shop unit	301	0.055
Low-quality house	182	0.033
Industry/working shed	113	0.021
Second floor shop unit	40	0.007
Comercial garage unit	28	0.005
Residencial garage unit	27	0.005
Total	5512	1.000

Source: GEAVI/PBH, 2007

The description of how the City calculates the 'construction quality index' is provided in the annex of "Decreto 10925 of 2001". It proposes five levels of quality, P1 being the worst

¹¹ Detailed explanation of the factors is available in the legislative decree (Decreto 10925 of 2001 from City of Belo Horizonte).

¹² The exchange rate was R\$ 2.60 on March 7, 2008.

and P5 being the best, depending on the items present in each unit. The presence of items of quality (such as internet cable availability, window material, or external finishing) sums to certain levels that classify the standard for that estate. Different items have different valuations depending on the type of estate (house, apartment, shop). For low-quality housing, for instance, the maximum standard is P3 (average quality). The building standard dominant in the sample, representing more than half of observations, is the P3 standard (Table 3). In cumulative terms, 86.5% of estates are classified at P3 or lower and only 13.5% of the sample is of higher quality (P4 or P5).

Table 3 – Number of observations by construction quality standards

Construction quality	Observations	Percentage
P1 (lowest)	309	0.06
P2	1330	0.24
P3	3160	0.57
P4	634	0.12
P5 (highest)	79	0.01
Total	5512	1.00

Source: GEAVI/PBH, 2007

Table 4 shows the spatial units that have the highest number of transactions. The 10 spatial units with the highest concentration of transactions account for nearly one third of the total number of transactions (Table 4), and 38 spatial units (out of 161) did not include any commercialized property within the analyzed period. The data show that the Castelo and Buritis neighborhoods have the highest rate of transactions. Traditional areas, such as São Pedro, Santo Antônio, Itapoã, Dona Clara, Sion and Funcionários, also have a high number of observations.

Table 4 – Spatial units with highest number of transactions

Top 10 spatial units in number of transactions	Observations	Percentage
BURITIS/ESTORIL/MANSÕES	268	4.81%
PARQUE URSOLINA DE MELO/CASTELO/RECREIO/ITAMARATI	268	4.81%
JD. ATLÂNTICO/STA. AMÉLIA/CONJUNTO STA. MÔNICA	166	2.98%
SÃO PEDRO/SANTO ANTÔNIO	165	2.96%
GRAJAU/GUTIERREZ	153	2.74%
BARROCA/PRADO	153	2.74%
DONA CLARA/JARDIM SANTA BRANCA/ ITAPOÃ	141	2.53%
CIDADE NOVA/SILVEIRA	139	2.49%
CARMO/SION	132	2.37%
FUNCIONÁRIOS	130	2.33%
<i>General total</i>	5512	100.00%
<i>General average of observations per neighborhood</i>	34.86	0.63%
<i>General standard-deviation</i>	49.22	0.88%

Source: author's elaboration based on GEAVI/PBH, 2007 and FJP, 2007.

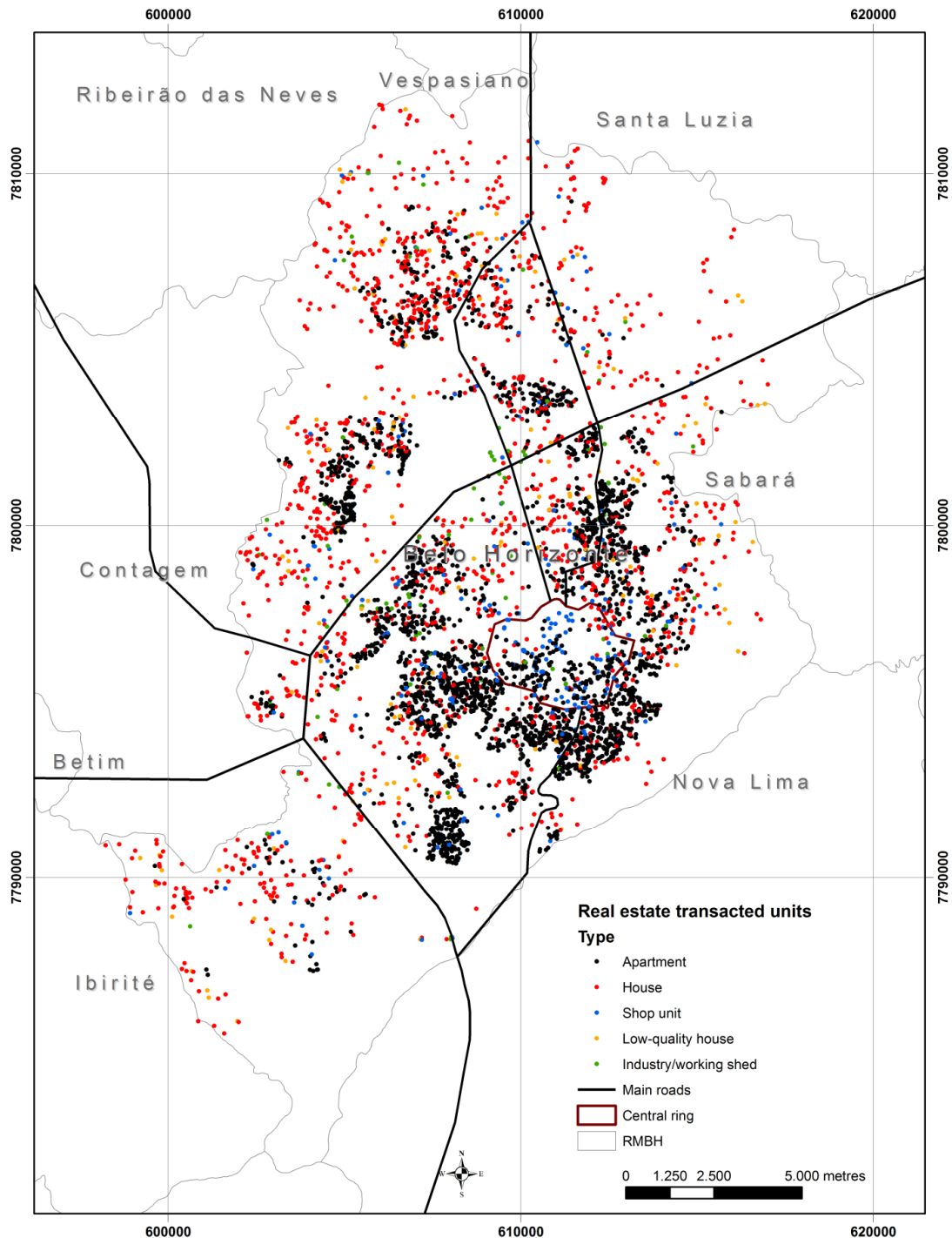


Figure 1 – Commercialized real estate units in Belo Horizonte, 2007, by type
 Source: author’s elaboration based on data provided by GEAVI/PBH, 2007.

Dataset used in the model

The dataset actually implemented in the models is related to each dwelling or to the neighborhood where it is located. In terms of the dwelling, we have: (a) Neperian logarithm of the price of the transaction that actually occurred ($\ln price$); (b) Neperian logarithm of transaction value per surface (in square meters); (c) Neperian logarithm of estate’s surface; (d)

construction quality level, as described above; (e) age of the dwelling, measured in years since construction and three *dummy* variables to indicate whether the observation is an apartment, a house, a shop or none of the above.

In terms of neighborhood, we have the following variables¹³. The first variable (*domic_n*) represents the neighborhood in terms of the average quality of the housing building and the families who inhabit them, in terms of income and education level. The second variable (*Pca_activ*) reflects quantitative information on the number of economic establishments, specifically the availability of services such as bakeries, gas stations, banks, and private offices. Both a high-class area and a relatively poor neighborhood might have low values for this variable. The third variable (*pca_ind_n*) reflects information on the presence of industry, exclusive of other economic activities. That is to say that industry is present at locations with higher values but stands out in relation to other economic activities. The fourth urban attribute of the neighborhood (*pca2_ino_n*) is based on company turnover in sectors that are strong in innovative service provision, specifically commerce, science and education, publicity, and information technology. In order to complement the description of location, the presence of public services – mainly arborization – is depicted by *f_melpub*. The presence of major shopping centers up to a radius of 1,500 meters is represented by *shop_1500*, presence of slums up to a radius of 200 meters is represented by *slum_200*. Finally, the presence of an avenue or arterial roads up to a radius of 200 meters (*aven_200*) (see Table 5).

¹³ See chapter 5 of Furtado (2009) for a full description of the indices calculus.

Table 5 – Basic statistics of variables used in the model

Variable	Variable code	Observations	Minimum	Maximum	Average	Standard-deviation
Estate's attributes						
Log of transaction value - log price	<i>lnprice</i>	5512	7.1554	15.1089	11.4697	0.7472
Log of transaction value per surface	<i>lnpricearea</i>	5512	4.7305	10.4436	6.6891	0.5063
Log of estate's surface	<i>lnarea</i>	5512	0.8796	7.5522	4.7806	0.6288
Construction quality levels	<i>standard</i>	5512	1	5	2.7903	0.7707
Age of estate (in years of construction)	<i>age</i>	5512	1	83	21.1852	14.5245
Dummy apartment	<i>d_ap</i>	5512	0	1	0.5865	0.4925
Dummy house	<i>d_house</i>	5512	0	1	0.2881	0.4529
Dummy shop	<i>d_shop</i>	5512	0	1	0.0546	0.2272
Location's attributes						
Normalized index of characteristics of the family and domicile by spatial units (see chapter 5) [0 worst;1 best]	<i>domic_n</i>	5512	0	1	0.2287	0.1899
Normalized index (PCA) of level of activities. Includes services, number of establishments	<i>pca_activ</i>	5512	0	1	0.1497	0.1789
Principal component that isolates industry activities in relation to services and commerce	<i>pca_ind_n</i>	5512	0	1	0.2463	0.2463
Second component of PCA analysis that concentrates innovative sectors and education services	<i>pca2_ino_n</i>	5512	0	1	0.5401	0.1232
Presence of public services. 98% of those without services refer to lack of arborization	<i>f_melpub2</i>	5512	0 (43%)	1 (57%)	0.5664	0.4956
Presence of major shopping mall within 1500 m	<i>shop_1500</i>	5512	0 (78%)	1 (22%)	0.2172	0.4124
Presence of slums within 200 m	<i>slum_200</i>	5512	0 (84%)	1 (16%)	0.1626	0.3690
Presence of an avenue or arterial road within 200 m	<i>aven_200</i>	5512	0 (48%)	1 (52%)	0.5185	0.4997

Dataset details related to the neighborhood characteristics are available on request.

Model, diagnoses, tests, weight matrix alternatives, and results

The basic model to be applied has the log of the price – the actual transaction value – as the dependent variable and the summarizing characteristics of the estate and the neighborhood, according to Table 5. The estimating equation is:

$$\ln Price = \alpha + \beta_1 \ln Area + \beta_2 Standard + \beta_3 Age + \beta_4 D_{ap} + \beta_5 Domic_n + \beta_6 PCA_{activ} + \beta_7 PCA_{innovation} + \beta_8 Factor_{melpub} + \beta_9 Shop_{1500} + \beta_{10} Slum_{200} + \beta_{11} Aven_{200} + \varepsilon.$$

The first step concerns the test for spatial dependence. According to Anselin (Anselin, 1988, pp. 103-104; Anselin et al., 2006), this can be done using the LM error statistic for spatial correlation in residuals and the LM lag statistic for omitted spatial lags. However, to implement the tests, a weight matrix, that is, the exogenously determined spatial dependence of the data has to be constructed. As suggested before, different definitions of the matrix might generate different, even contradictory results. We therefore test for different conceptualizations of matrices to confirm the robustness of the matrices chosen.

Weight matrices

Three matrices are used. The most traditionally defined is the *contiguity matrix*, in which all neighbors are calculated based on Delaunay triangles. In this case, distance is not important and neighbors are all the closest individuals to a certain observation. Another matrix typically applied is the *distance matrix*, which considers the distance to neighbors instead of a binary matrix (neighbor, non-neighbor). In the default alternative (Anselin, 2005) all observations have at least one neighbor, preventing an island (no neighbor) situation from happening. For the dataset, this maximum distance was 1,057 meters. The third weight matrix to be considered is called the *neighborhood matrix* and is defined as observations within the same neighborhood – the spatial unit (UDH) – are considered neighbors. The idea is that this matrix captures the notion of belonging to a certain identity-based neighborhood¹⁴.

All matrices are row-standardized.

Visually, we present the matrices in Figure 1, Figure 2 and Figure 3. Every line indicates a connection among nodes (house transactions). The contiguity matrix, most widely used in the literature, considers all the closest neighbors, independently of distances in the spatial configuration. This implies some deformations in the spatial structure tested. The number of connections is smaller when compared to other matrices. The distance matrix is almost as dense as the neighborhood matrix; however, its links encounter no limits, such as railways or rivers. The neighborhood matrix depicts the idea of belonging, in which observations are clearly separated in different groups, which further conveys the idea that the influence of observations of one group on another is less important.

¹⁴ The k-nearest-neighbor, in which the k number of nearest neighbors are considered, is not evaluated because it lacks theoretical (Smirnov and Anselin, 2001; Anselin, 2002) and empirical intuition. If one thinks spatially, one's five nearest neighbors might have a totally different configuration and distance from one another and should not be considered part of the same pattern. Moreover, Anselin (2005) suggests that results have not proven to be consistent using this matrix and maximum likelihood methodology.

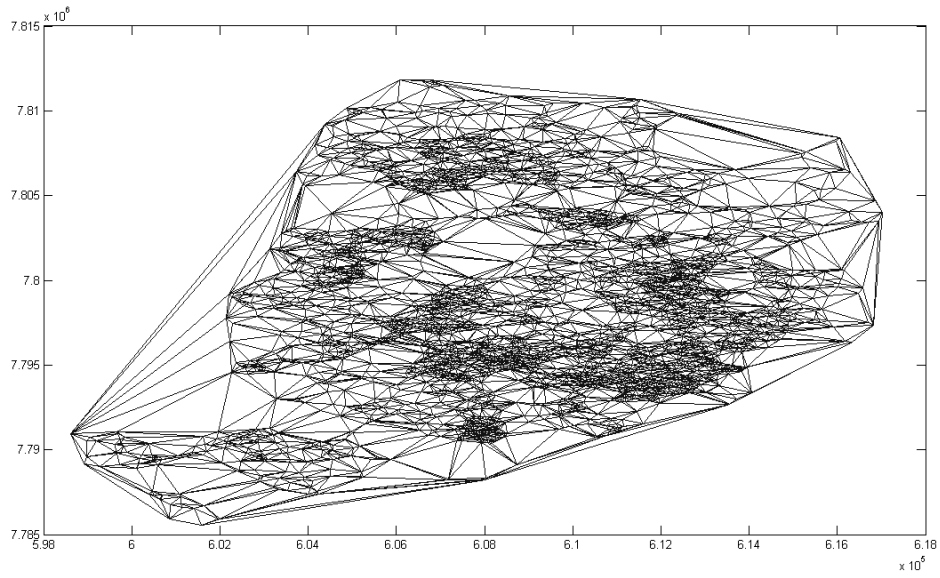


Figure 1 – Visualization of matrix of contiguity

Source: elaborated by the author based on the dataset (GEAVI/PBH, 2007)

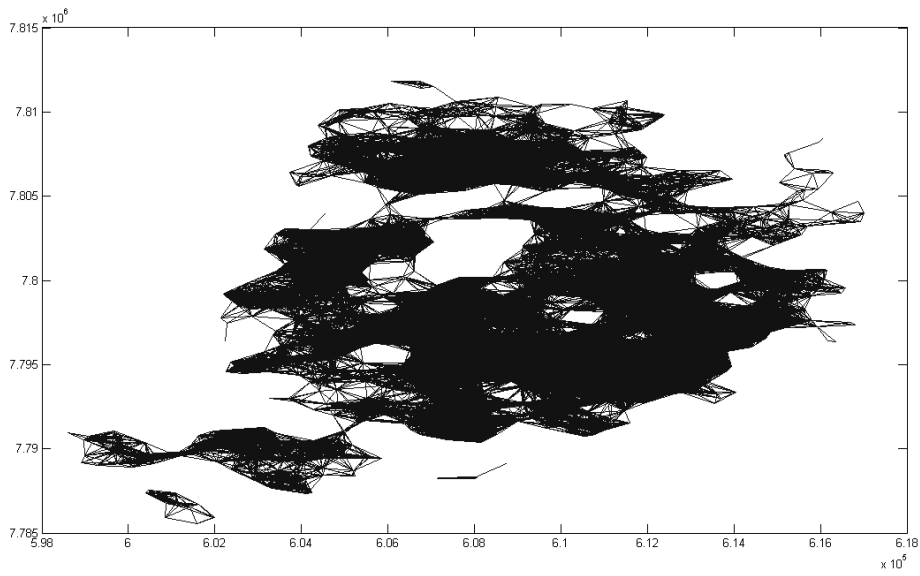


Figure 2 – Visualization of matrix of distance

Source: elaborated by the author based on the dataset (GEAVI/PBH, 2007)

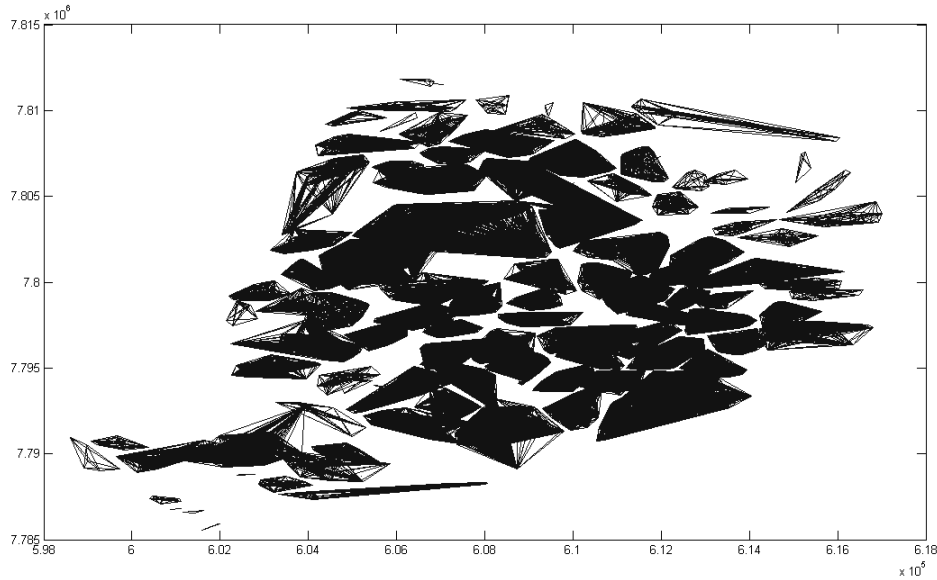


Figure 3 – Visualization of matrix based on neighborhood

Source: elaborated by the author based on the dataset (GEAVI/PBH, 2007)

Interpretation of results and use of matrices

This section discusses the outcomes of the spatial, quantile and spatial-quantile models.

As expected, all results (Table 1, Table 2) confirm the strong spatial dependence of the problem, and therefore the need for spatial modeling. The OLS estimation is presented as a comparative parameter.

Table 1 – Moran's I statistic as confirmation of spatial dependence

	W contiguity	W distance	W neighborhood
<i>Moran's I</i>	0.104	0.046	0.072
<i>Moran's I-statistic</i>	14.307	19.658	26.143
<i>Marginal Probability</i>	0.000	0.000	0.000
<i>mean</i>	-0.001	-0.001	-0.001
<i>standard deviation</i>	0.007	0.002	0.003

Table 2 – Tests for spatial autocorrelation

Anselin's tests for spatial autocorrelation				
	LM ERROR (chi 0.01 17.611)		Robust (chi 0.01 6.64)	
	value	prob.	value	prob.
W contiguity	198.767	0.000	203.083	0.000
W distance	351.104	0.000	151.690	0.000
W neighborhood	626.101	0.000	561.710	0.000
	LM LAG (chi 0.01 6.64)		Robust (chi 0.01 6.64)	
	value	prob.	value	prob.
W contiguity	3.551	0.060	7.867	0.005
W distance	326.606	0.000	127.192	0.000
W neighborhood	97.383	0.000	32.991	0.000

OLS and spatial

The results presented in Table 3 show that the coefficients are statistically significant and that there is a high degree of consistency among the different (weight-matrix) estimates. The observed differences help enhance understanding of the model, by providing insight into the spatial structure. Although the values for the R-squared are also close to one another, the best fit is for the spatial estimates of the neighborhood matrix with a value of 0.689. The log-likelihood is even better, at 2,028 in that model.

Table 3 – OLS and spatial model results

	OLS	Spatial model		
		W contiguity	W distance	W Neighborhood
Lnprice				
<i>Lnarea</i>	0.772 (77.93)**	0.765 (81.17)**	0.764 (79.13)**	0.758 (79.02)**
<i>Standard</i>	0.144 (15.23)**	0.118 (13.26)**	0.113 (12.9)**	0.110 (11.61)**
<i>Age</i>	-0.003 (-6.95)**	-0.004 (-7.62)**	-0.005 (-9.72)**	-0.004 (-8.94)**
<i>D_ap</i>	-0.083 (-5.99)**	-0.111 (-7.75)**	-0.129 (-9.39)**	-0.149 (-10.3)**
<i>Domic_n</i>	0.365 (10.76)**	0.350 (8.8)**	0.239 (7.16)**	0.266 (4.44)**
<i>Pca_activ</i>	0.343 (8.11)**	0.403 (7.67)**	0.232 (5.37)**	0.499 (5.97)**
<i>Pca_ind_n</i>	-0.278 (-5.89)**	-0.269 (-4.79)**	-0.151 (-3.09)**	-0.248 (-2.53)*
<i>Pca2_ino_n</i>	0.270 (5.26)**	0.288 (5.37)**	0.158 (3.19)**	0.429 (4.78)**
<i>Shop_1500</i>	0.170 (10.14)**	0.176 (8.38)**	0.079 (4.65)**	0.113 (5.47)**
<i>Slum_200</i>	-0.073 (-4.54)**	-0.054 (-2.94)**	-0.081 (-5.1)**	-0.063 (-3.72)**
<i>Aven_200</i>	0.061 (5.11)**	0.056 (4.15)**	0.045 (3.84)**	0.038 (3.05)**
<i>F_melpub2</i>	0.108 (8.23)**	0.077 (5.78)**	0.067 (5.21)**	0.061 (4.61)**
<i>Constant</i>	6.974 (126.51)**	7.161 (561.19)**	3.271 (20.24)**	6.779 (646.43)**
<i>rho</i>		-0.005 (-3.00)**	0.355 (27.2)**	0.037 (5.69)**
<i>lambda</i>		0.313 (65.69)**	0.070 (8)**	0.594 (172.64)**
Observations	5512	5512	5512	5512
Log-likelihood		1968.5	1983.7	2028.1
R-squared	0.666	0.683	0.680	0.689

Absolute value of t statistics in parentheses

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

The presence of different spatial structures represented by the weight matrices should be kept in mind when interpreting the results. Given the results of the tests in Table 2, we

consider the OLS results to be only indicative of order and parameters, because it is clear that the lack of spatial information renders this model inferior. The estimation with the contiguity matrix includes a weak spatial structure. The construction of the matrix, through utilization of Delaunay triangulation (Voronoi tessellation), defines only the closest observations as neighbors.

The results estimated with the distance and neighborhood matrices, on the other hand, include neighbors in a somewhat more broadly defined manner. The distance matrix includes every observation within a given radius, and the neighborhood matrix includes all neighbors who are within the same homogeneous spatial unit. The difference between the two is that, while in the distance matrix, one includes neighbors within an abstract physical distance, in the neighborhood matrix, one considers identity in establishing neighborhood membership. Viewing the problem from an urban perspective in which features such as rivers, rails or roads divide space and in which being on one side or the other clearly matters, the neighborhood matrix seems more plausible, practically speaking. This is confirmed in the results, which indicate that this estimation fits the data best. Furthermore, the estimation with the contiguity matrix seems to perform slightly better than the distance matrix if one considers the value of the R-squared statistic.

In terms of the influence of estate-surface on value, there is little variation between the estimates, and on average, an increase of 1% in individual living area will increase prices about 0.76%. However, there is a slight decrease from the estimations of the OLS to the neighborhood matrix, suggesting that as model explanation power improves, less value is connected to the building-surface coefficient. When it comes to the quality of the estate, represented by the variable *standard*, it is clear that some of its importance is due to location instead of individually defined attributes, when comparing OLS results with other estimations. Therefore, if one considers the OLS estimate, moving upwards one category within the standard levels (from P3 to P4, for instance), there would be an increase in estate prices that is higher than average on spatial estimates¹⁵. The influence of the age of the estates is largely similar, indicating that aging one year reduces prices around 0.5%, on average.

¹⁵ Because the variable *standard* is ordinal, the recommended procedure (Wooldridge, 2002, p. 214) would be to have dummy variables assigned to all categories but one. This would more accurately capture the differentiated impact from moving from a specific standard to another. However, for the sake of clarity and simplicity and because the standard plays the role of the control variable, only one dummy variable that represents an average influence is applied in the model. As a robustness check, an estimation with all variables was performed and little difference was observed in the parameters. The results are available on request. Finally, the quantile approach implicitly considers differentiation in quality of the estate and, by the construction of the standard variable criteria by the government, it is not possible to have an estate of standard 1 that, at the same time, is embedded in the highest quantile of price. The results would reflect this, which would make it more difficult to interpret locational variables (which are the focus of our study).

There are much larger differences over model specifications concerning the estimates of urban attributes. It seems, however, that the results of the estimates with the neighborhood matrix are again most plausible.

We used the information gathered thus far to illustrate the impact on estate prices when considering moving into a better neighborhood or a worse neighborhood. When comparing an estate located in an emerging middle-class neighborhood such as *Buritis* with an estate with exactly the same attributes but in an upper-class traditional neighborhood such as *Lourdes* or with a traditional, stable neighborhood such as *Caiçara*, we find the results presented in Table 4.

Table 4 – Illustration of results' interpretation

Neighborhoods/Values	<i>Domic_n</i>	<i>Pca_activ</i>	<i>Pca_ind_n</i>	<i>Pca2_ino_n</i>	
Lourdes (upper-class)	0.95	1.00	0.12	1.00	
Buritis (trendy, emerging, middle-class)	0.67	0.09	0.32	0.86	
Caiçara (traditional low-middle-class)	0.63	0.23	0.54	0.59	
Difference (moving into upper-class from Buritis)	0.28	0.91	-0.20	0.14	
Difference (moving into lower-class from Buritis)	-0.04	0.13	0.22	-0.26	
<i>Coefficients for different estimations</i>	<i>Domic_n</i>	<i>Pca_activ</i>	<i>Pca_ind_n</i>	<i>Pca2_ino_n</i>	
OLS	0.37	0.34	-0.28	0.27	
W contiguity	0.35	0.40	-0.27	0.29	
W distance	0.24	0.23	-0.15	0.16	
W Neighborhood	0.27	0.50	-0.25	0.43	
<i>Impact on price for apartments for upward move (%)</i>					Sum
OLS	0.10	0.31	0.06	0.04	0.51
W contiguity	0.10	0.37	0.05	0.04	0.56
W distance	0.07	0.21	0.03	0.02	0.33
W Neighborhood	0.07	0.45	0.05	0.06	0.64
<i>Average</i>	0.08	0.34	0.05	0.04	0.51
<i>Impact on price for apartments for downward move (%)</i>					Sum
OLS	-0.02	0.05	-0.06	-0.07	-0.10
W contiguity	-0.01	0.05	-0.06	-0.08	-0.10
W distance	-0.01	0.03	-0.03	-0.04	-0.05
W Neighborhood	-0.01	0.07	-0.05	-0.11	-0.11
<i>Average</i>	-0.01	0.05	-0.05	-0.08	-0.09

Source: elaborated by the author based on results of Furtado (2009) and regressions

Note that a change of neighborhoods is not a linear move. For example, even though domiciles and families in Buritis and Caiçara are only slightly different in value (0.67 to 0.63), and Caiçara has an older but more stable availability of services (*pca_activ* from 0.23 to 0.09), Buritis has a lower negative impact from the presence of heavy industry (0.32 to 0.54) and, on the contrary, a higher presence of dynamic services and commerce (0.86 to 0.59). In this case, moving into a somewhat lower-class or less dynamic neighborhood, means a decrease in price an average of 9%.

Comparing the matrices' impact, if one considers the matrix of neighborhood embeddedness, than this difference is 2 percentage points higher than the average. If, on the

other hand, the distance matrix is used, the decrease in prices is downplayed to a mere 5%, compared to an average of 9%.

A movement from a middle-class neighborhood to nearly the best neighborhood in terms of presence of economic activities, innovative sectors, domiciles and family characteristics would mean an increase of 51%, on average, of prices of apartments. Once again, the distance matrix tends to diminish differences, while the neighborhood matrix amplifies them.

The other variables considered come from the socioeconomic characterization of the urban fabric and aspects (a) and (b)¹⁶ as defined by Galster (2001). There is not much difference in how they impact estimates. All models show (as expected) positive effects on evaluation of being near major shopping centers, but this is less so when considering the neighborhood matrix in contrast to OLS models or a contiguity matrix. Again, the negative impact of the proximity of slums and the positive impact of being near main avenues are captured. The presence of public services (mainly arborization) presents the lowest value in the estimates when the neighbourhood matrix is considered. This is probably because having a pleasant, green neighborhood might be valued as an attribute of the neighborhood itself, which is accounted for by using the neighborhood matrix.

Quantile Regression Results

The results of the quantile regression are insightful for decomposing the variance within different quantile, but they should be viewed with caution because the tests clearly demonstrate the need for a spatial model.

In general, the analysis in Table 5 clearly shows that the impact of the various factors differs from the lowest considered quantile (at .1) to the highest (at .9). All present significance. The results and signs are conform to intuition.

¹⁶ (a), the quality of public service offered within, and (b), proximity (such as accessibility to shopping or employment centers), as described in Furtado (2009).

Table 5 – Results for quantile regression

	q .1	q .25	q .5	q .75	q .9
	Lnprice				
<i>Lnarea</i>	0.909 (79.15)**	0.895 (106.42)**	0.806 (101.37)**	0.761 (64.15)**	0.667 (34.78)**
<i>Standard</i>	0.221 (29.96)**	0.192 (28.41)**	0.201 (26.44)**	0.170 (14.29)**	0.181 (9.85)**
<i>Age</i>	-0.011 (32.31)**	-0.009 (27.92)**	-0.005 (14.07)**	-0.002 (3.83)**	-0.003 (3.41)**
<i>D_ap</i>	0.201 (16.66)**	0.101 (9.54)**	-0.061 (5.45)**	-0.181 (11.73)**	-0.440 (20.77)**
<i>Domic_n</i>	0.240 (10.34)**	0.385 (17.73)**	0.351 (12.91)**	0.250 (5.36)**	0.263 (3.54)**
<i>Pca_activ</i>	0.149 (4.15)**	0.272 (8.42)**	0.400 (11.77)**	0.517 (10.86)**	0.528 (7.70)**
<i>Pca_ind_n</i>	0.001 -0.030	-0.110 (3.23)**	-0.230 (6.08)**	-0.377 (6.61)**	-0.461 (5.44)**
<i>Pca2_ino_n</i>	0.173 (4.03)**	0.089 (2.30)*	0.181 (4.38)**	0.205 (3.50)**	0.209 (2.54)*
<i>Shop_1500</i>	0.090 (6.20)**	0.117 (9.06)**	0.154 (11.45)**	0.181 (9.56)**	0.208 (7.79)**
<i>Slum_200</i>	-0.088 (6.70)**	-0.048 (3.99)**	-0.074 (5.73)**	-0.091 (4.95)**	-0.096 (3.77)**
<i>Aven200</i>	0.015 -1.590	0.024 (2.74)**	0.030 (3.09)**	0.056 (4.13)**	0.094 (5.00)**
<i>f_melpub2</i>	0.048 (4.51)**	0.058 (6.02)**	0.079 (7.52)**	0.099 (6.56)**	0.151 (7.05)**
<i>Constant</i>	5.844 (116.71)**	6.100 (149.09)**	6.713 (151.67)**	7.349 (107.07)**	8.168 (75.35)**
Observations	5512	5512	5512	5512	5512
Pseudo-R2	0.502	0.475	0.455	0.464	0.472

Absolute value of t statistics in parentheses

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

The size of the apartment is a much more important factor for valuation in the cheapest segment of the city *vis-à-vis* more expensive parts of the city. It appears that as apartments become larger and more expensive (as in higher quality city segments), surface area becomes proportionally less important (the marginal utility of more surface area decreases with general improvements in quality). The same interpretation applies to the quality of the apartment. Most likely, more luxurious apartments already have a high-quality standard. It is clear that houses are preferred in the high-priced market, whereas apartments are more valued in the low-priced market.

The quality of neighborhood has a mixed influence on estate prices (see previous section), but appears to have a similar relationship with prices across all quantiles. Still, the neighborhood appears to matter less for prices of the cheapest apartments. The preference for economic activities, however, increases steadily over quantiles, and reaches its highest value

in relation to pricing of the most expensive apartments. The same pattern can be observed for the relation of the intensity of innovative sectors and education with prices in the sorted quantiles, but in a less steep manner.

Being within the 1,500 meters of large shopping centers appears to be more important for apartments within and above median prices, and less so for apartments at the lowest quantile. The proximity to slums presents negative values for all price ranges. The same is true for the presence of a nearby avenue, with greater importance for pricing of estates in the upper range.

Spatial-quantile (IVQR)

The results of the spatial-quantile model (Table 6) confirm the indications of both spatial and quantile estimations and should be considered the best possible estimate, according to the theoretical discussion provided by Su and Yang (2007).

Table 6 – Results for instrumental variables spatial-quantile regression

Variables	W. Contiguity					W. Distance					W. Neighborhood				
	Quantiles					Quantiles					Quantiles				
	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9
<i>constant</i>	6.193	6.591	7.231	7.231	8.735	6.884	7.077	7.467	7.807	7.879	6.01	6.112	6.911	7.486	8.106
<i>lnarea</i>	0.916	0.877	0.783	0.783	0.629	0.914	0.876	0.778	0.698	0.639	0.910	0.876	0.770	0.703	0.616
<i>standard</i>	0.209	0.176	0.169	0.169	0.127	0.189	0.177	0.157	0.123	0.118	0.195	0.181	0.154	0.119	0.132
<i>age</i>	-0.012	-0.010	-0.007	-0.007	-0.006	-0.013	-0.010	-0.008	-0.006	-0.006	-0.013	-0.010	-0.008	-0.006	-0.006
<i>d_ap</i>	0.147	0.043	-0.115	-0.115	-0.489	0.146	0.043	-0.153	-0.314	-0.525	0.157	0.037	-0.155	-0.304	-0.482
<i>domic_n</i>	0.029	0.093	0.126	0.126	0.199	0.104	0.204	0.130	0.175	0.242	-0.018	0.040	0.138	0.103	0.243
<i>pca_activ</i>	0.141	0.331	0.676	0.676	0.654	0.039	0.048	0.238	0.350	0.225	-0.008	0.109	0.039	0.192	0.257
<i>pca_ind_n</i>	0.051	-0.114	-0.371	-0.371	-0.327	0.002	-0.040	-0.168	-0.275	-0.372	0.143	0.119	0.048	0.148	-0.251
<i>pca2_ino_n</i>	0.002	0.011	0.127	0.127	0.192	0.263	0.068	0.155	0.150	0.323	0.226	0.387	0.314	0.321	0.693
<i>shop_1500</i>	0.005	0.041	0.118	0.118	0.231	0.011	0.012	0.036	-0.007	-0.007	0.041	0.030	0.055	0.083	0.121
<i>slum_200</i>	-0.038	-0.002	-0.012	-0.012	-0.054	-0.099	-0.055	-0.065	-0.112	-0.112	-0.090	-0.049	-0.075	-0.098	-0.083
<i>aven_200</i>	-0.006	-0.002	0.018	0.018	0.033	0.007	0.010	0.023	0.039	0.054	0.005	0.011	0.015	0.035	0.035
<i>f_melpub2</i>	0.025	0.028	0.048	0.048	0.076	0.036	0.025	0.037	0.043	0.047	0.031	0.033	0.045	0.041	0.047
<i>rho</i>	-0.119	-0.126	-0.172	-0.172	-0.217	-0.092	-0.107	-0.078	-0.085	-0.034	-0.094	-0.059	-0.094	-0.152	-0.237

For all weight matrices tested, the estimates vary little over the estate's attributes and show, in general, the same trend. The coefficient for estate-level surface *area*, for instance, consistently decreases in importance from the cheapest to the most expensive estates, and is slightly smaller in the case of estimation using the neighborhood matrix. The same trend can also be observed in relation to the quality of the apartment (*standard*) and the dummy variable that characterizes the estate as an apartment. The age of the estate also has approximately identical coefficients for all matrices as well.

The increase in importance of the quality of the domiciles and neighborhoods from the lowest to the highest quantiles manifests in specifications with different weight matrices. However, its effect is strongest for the neighborhood matrix.

In relation to the level of activities (*pca_activ*), the strength of the coefficients differs substantially between specifications using the matrix of contiguity compared to that of the neighborhood. Clearly, because the matrix of contiguity itself does not account for centrality, the coefficients partly express this aspect as well. For the models that use the matrix of neighborhood on the other hand, part of the effect of centrality is already embedded in the neighbourhood, so that the coefficient expresses, to a larger extent, the isolated advantage of having a higher service level.

The variable that measures the dominant industrial activity (*pca_ind_n*) consistently shows that repulsiveness is an increasing function of the price of the estate, and it is nearly zero for the lowest quantile. This does not hold for the model using the neighborhood matrix. In this case, the negative effect of the industry only lowers prices for the most expensive estates.

In terms of capturing the innovative activities effect, the model using the neighborhood matrix presents the highest significant values, with innovation being much more valued (and used) by individuals seeking more expensive apartments.

The proximity to large shopping centers is more expressed in the valuation of expensive estates in both the contiguity and neighborhood specifications, whereas in the model using the distance matrix, it has a slightly negative effect. This is probably because the variable itself is also based on distance.

All models confirm the negative valuation of proximity to slums (*slum_200*) and the positive valuation of the presence of arborization (*f_melpub2*).

The negative ρ for all models with different matrices and quantiles suggests that there is still heterogeneity in observations within each neighborhood framework. This is further confirmed by the analysis of spatial autocorrelation measured by Moran's I index (for the neighborhood matrix, the 0.5 quantile, for instance, has a statistic of 0.0723, which is statistically significant at the 1% level) and interpolation of the residuals (Figure 1). The variation of the residuals is not prominent across the city (left side of the figure), but it is present in the overall, aggregated plot of observations. The figure on the right side indicates that the difference of prices might be spatially attributed to distances of about 100 meters. The framework of the cognitively perceived urban space cannot capture this level of detail because citizens cannot identify, classify and name urban spatial units at this scale. Perceptions would overlap and be too fuzzy to be considered as valid general variables. Furthermore, the idiosyncrasies seem to be attached to the level of each observation. Consequently, what is not accounted for in the model might be an architectural style, for instance, or other particularities

of the dwelling itself. The larger negative values of ρ are observed with the higher quantiles of the neighborhood matrix, which are more likely to have unaccounted, unique characteristics.

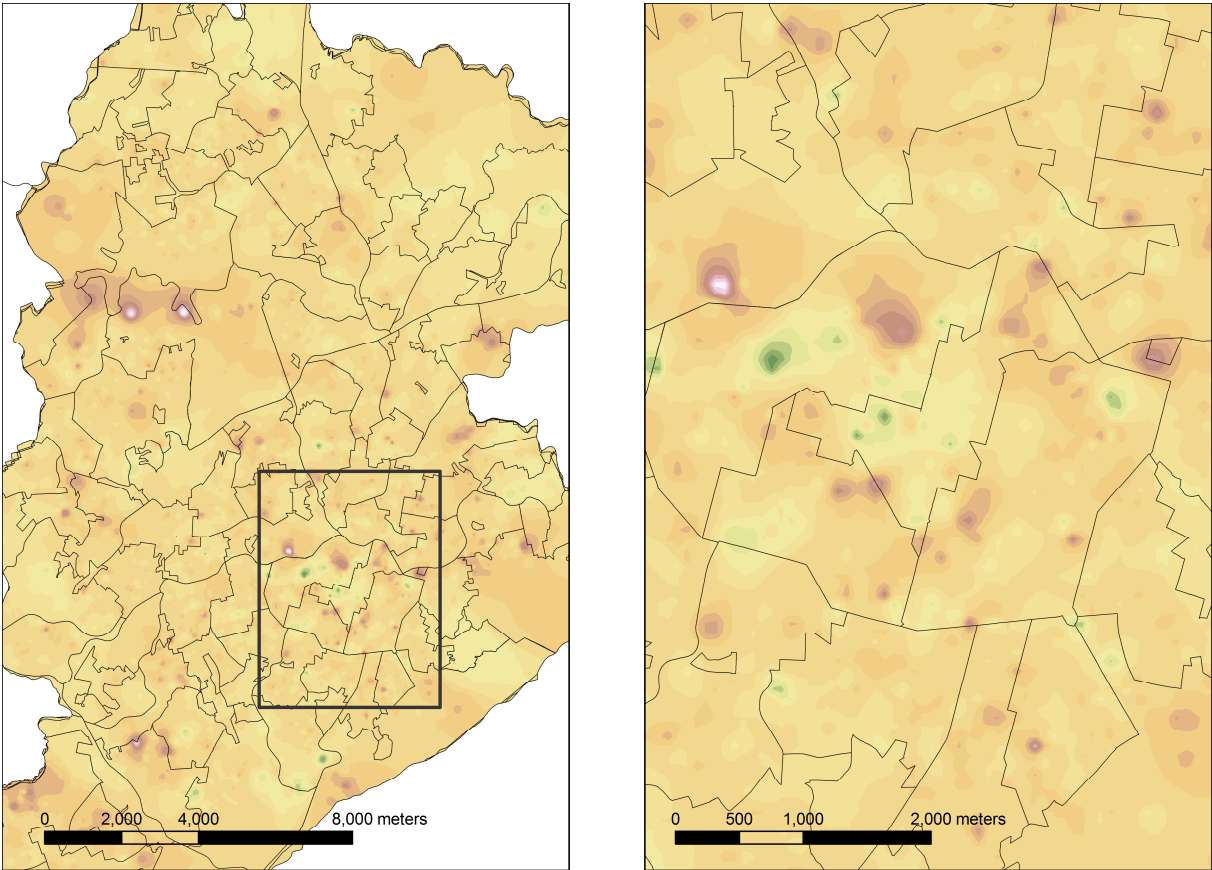


Figure 1 – Spatial illustration of distribution of residues of IVQR estimation, neighborhood matrix, 0.5 quantile

Concluding remarks

Our results confirm that we cannot evaluate real estate markets without taking spatial dependencies into account. The analyses also show that urban complexity embedded within cognitively-perceived neighborhoods can be included in estimation modeling in detail. The methodology applied allowed us to describe economic relations between city space and housing prices better than one could if only measuring distance to CBD.

The factors used are easy to interpret. These dimensions, in turn, made it easier to distinguish among the preferences of socially-heterogeneous actors. Moreover, these factors and the quantile estimation methodology explicitly show that those who can afford the most

expensive estates value variables¹⁷ in a considerably different manner from those seeking cheaper estates.

The different matrices also confirmed that the choice of matrix should always be made cautiously. The results show that the neighborhood matrix: (a) is the best fit to the data; (b) has the strongest theoretical support; (c) the best at reproducing a cognitive description of the city (see Figure 3); and (d) captures the idiosyncrasies of large urban metropolitan complexities without the need to restrict spatial influence to an arbitrarily-decided central place (i.e., the Central Business District).

Finally, this discussion and the results in the previous section determined the extent to which the various factors considered in the introduction – both at the estate level and at the neighborhood level – influence housing prices.

References

- Anselin, L., 1988. *Spatial econometrics: methods and models*, Kluwer Academic, Dordrecht.
- Anselin, L. 1992. *SpaceStat tutorial: a workbook for using SpaceStat in the analysis of spatial data*. Urbana-Champaign: University of Illinois, 263.
- Anselin, L., 1999. *Spatial econometrics*, University of Texas at Dallas
- Anselin, L. *Under the hood. Issues in the specification and interpretation of spatial regression models*. Regional Economics Applications Laboratory, University of Illinois at Urbana-Champaign 2002 [cited.
- Anselin, L., 2005. *Exploring spatial data with GeoDaTM: a workbook*, Spatial Analysis Laboratory, Department of Geography, University of Illinois, Urbana-Champaign, IL.
- Anselin, L., Bera, A. 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In *Handbook of Applied Economic Statistics*, editor A. Ullah and D. E. A. Giles, New York: Marcel Dekker, 237-289.
- Anselin, L., Syabri, I., Kho, Y., 2006. *GeoDa: An Introduction to Spatial Data Analysis*, *Geographical Analysis* 38, 5-22.
- Arbia, G., 2006. *Spatial econometrics: statistical foundations and applications to regional convergence* Springer, Berlin.
- Buchinsky, M., 1997. Recent advances in quantile regression models, *Journal of Human Resources* 33, 88-126.
- Capello, R., Nijkamp, P. 2004. The theoretical and methodological toolbox of urban economics: from and towards where? In *Urban dynamics and growth: advances in urban economics*, editor R. Capello and P. Nijkamp, Amsterdam: Elsevier, 1-27.
- Cliff, A., Ord, J. K., 1973. *Spatial Autocorrelation*, Pion, London.
- Furtado, B. A., 2009. *Modeling social heterogeneity, neighborhoods and local influences on urban real estate prices: spatial dynamic analyses in the Belo Horizonte metropolitan area, Brazil*, Faculteit Geowetenschappen Universiteit Utrecht, Utrecht, The Netherlands.
- Garreau, J., 1992. *Edge City : Life on the New Frontier* Random House Inc.
- Koenker, R., Bassett, G., 1978. *Regression Quantiles*, *Econometrica* 46, 33-50.

¹⁷ All variables, except proximity to slums (see Table 6 – Results for instrumental variables spatial-quantile regression).

- LeSage, J., 1999. The theory and practice of spatial econometrics, Department of Economics, University of Toledo
- Lynch, K., Camargo, J. L., 1960. The image of the city, Martins Fontes, São Paulo.
- Muhammad, S. 2007, Future urbanization patterns: in the Netherlands, under the influence of information and communication technologies Doctorate, Faculty of geosciences, Utrecht University, Utrecht.
- Paelinck, J., Klaassen, L., 1979. Spatial Econometrics, Saxon House, Farnborough.
- Rossi, A., 1966. Architecture of the city (L'architettura della città), Padova/MIT Press
- Sheppard, S. 1999. Hedonic analysis of housing markets. In *Handbook of regional and urban economics*, editor P. Cheshire and E. S. Mills, Amsterdam: North-Holland, 1595-1636.
- Smirnov, O., Anselin, L., 2001. Fast maximum likelihood estimation of very large spatial autoregressive models: A characteristic polynomial approach, *Computational Statistics and Data Analysis* 35, 301-319.
- Su, L., Yang, Z. 2007, Instrumental Variable Quantile Estimation of Spatial Autoregressive Models. Unpublished paper presented at I World Conference on Spatial Econometrics, Cambridge, UK.
- Tobler, W. R., 1970. A computer movie simulating urban growth in the Detroit region, *Economic Geography* 46, 234-240.
- Tversky, B., 2003. Structures of mental spaces: how people think about space, *Environment and Behavior* 35, 66-80.
- Wheaton, W. C., 2004. Commuting, congestion, and employment dispersal in cities with mixed land use, *Journal of Urban Economics* 55, 417-438.
- Whittle, P., 1954. On stationary processes in the plane, *Biometrika* 41, 434-449.
- Wooldridge, J. M., 2002. *Econometric analysis of cross section and panel data*, MIT Press, Cambridge, Mass.
- Wooldridge, J. M., 2003. *Introductory econometrics: a modern approach*, 2ed. ed Mason : Thomson South-Western
- Zietz, J., Zietz, E. N., Sirmans, G. S., 2008. Determinants of house prices: a quantile regression approach, *Journal of Real Estate Finance and Economics* 37, (forthcoming).