Hedonic Price Models with Spatial Effects: an Application to the Housing Market of Belo Horizonte, Brazil*

Paulo Brígido Rocha Macedo**

Summary: 1. Introduction; 2. Literature review; 3. Specification issues; 4. Data description; 5. Empirical results; 6. Conclusion.

Key words: spatial dependence; hedonic price models; Box-Cox transformation; housing market.
JEL code: R1 e R15.

This paper uses both standard and spatial autoregressive hedonic price models (HPM) to analyze sample data from the housing market of Belo Horizonte, Brazil. Among the spatial econometric tools used are diagnostic tests for the detection of spatial dependence and heterogeneity, which provide the means to identify adjacency effects in the determination of housing prices. The study tests also a number of alternative functional forms for both the standard and the spatial HPM using the Box-Cox transformation of the variables analyzed. The results show that spillover (adjacency) are an important source of price variation in the housing market analyzed, and the doublelog specification provides the best fit in describing the relationship between housing prices and attributes.

Este artigo se utiliza de modelos de preços hedônicos (MPH) – tanto na especificação padrão como na especificação com efeitos espaciais, para analisar dados de uma amostra de unidades residenciais em Belo Horizonte, Brasil. Entre os instrumentos de econometria espacial usados estão testes de diagnósticos para a verificação da presença de dependência espacial e heterogeneidade que provêm os meios de identificação de efeitos de adjacência (transbordamento) na determinação dos preços de imóveis. O estudo testa também um número de formas funcionais alternativas para MPH – na especificação padrão, bem como naquela com efeitos espaciais – usando a transformação Box-Cox das variáveis analisadas. Os resultados evidenciam que o transbordamento (adjacência) é uma importante fonte de variação nos preços das unidades residenciais analisadas, e que a especificação (dupla) logarítmica se adequa melhor, em termos de ajustamento estatístico, à descrição da variabilidade dos preços das unidades relativamente a seus respectivos atributos.

*Paper received in June 1996 and approved in Oct. 1997. An earlier version of this paper was presented at the XIV Latin American Meeting of the Econometric Society, 1996.

**Department of Economics, and Center for Regional Development and Planning (Cedeplar), Federal University of Minas Gerais, Brazil.

1. Introduction

First developed in the 60's, hedonic price models (HPM) have frequently been applied to the study of housing markets. In the early 90's, some studies incorporated spatial econometrics in HPM housing models to address methodological issues ignored in the standard approach.

This paper uses both standard and spatial autoregressive hedonic price models to analyze sample data from the housing market of Belo Horizonte, Brazil. Among the spatial econometric tools used are diagnostic tests for the detection of spatial dependence and heterogeneity, which provide the means to identify adjacency effects in the determination of housing prices. The paper also tests a number of alternative functional forms for both the standard and the spatial HPM models, using the Box-Cox transformation of the variables analyzed.

The results show that spillover (adjacency) effects are an important source of price variation in the residential apartment market in Belo Horizonte. The empirical findings support the need to incorporate spatial effects in studies of housing price determination. The importance of incorporating spatial factors makes the use of special spatial econometrics techniques essential to the analysis of economic performance in housing markets.

2. Literature Review

2.1 Hedonic price models

According to Griliches (1971), "the hedonic characteristics approach to the construction of price indexes is based on the empirical hypothesis that the multitude of varieties (or models) of a particular commodity can be comprehended in terms of a much smaller number of basic attributes". Early examples of this approach include the empirical analysis of automobile prices by Griliches (1971) and Dhrymes (1967), the study of the real estate construction market by Bailey, Muth and Nourse (1963), and the analysis of technological change in the computer mainframe industry by Chow (1967). In the 70's empirical applications, as well as theoretical developments, firmly established the hedonic price models (HPM) approach in the literature. A major contribution is that of Rosen (1974) who develops a theoretical framework "based on the hedonic hypothesis that goods are valued for their utility-bearing characteristics". Accordingly, "hedonic prices are defined as the implicit prices of attributes and are revealed to economic agents from observed prices of differentiated products and the specific amounts of characteristics associated with them".

Empirical analyses based on the hedonic approach must address the two following questions first proposed by Griliches (1971):

(a) What are the relevant characteristics?

(b) What is the form of the relationship between prices and characteristics?

With regard to the first question, the early HPM studies on automobile prices used three car characteristics: size, power, and accessories; Chow's (1967) analysis of the mainframe computer industry had two characteristics: memory capacity and speed of the instruction cycle. Urban housing markets, however, present a much larger number of potentially relevant characteristics. Butler (1982) notes that "data on many of these characteristics are either unavailable or of exceedingly poor quality. Even without data constraints, the intrinsic clustering of characteristics combinations into a relatively small number of configurations leads to considerable multicollinearity in estimates employing a generous selection of the relevant variables". Echoing the warning of Griliches (1971) against "the use of variables which are not direct characteristics of the commodity but an outcome of the market experiment", Butler comments that "this is the case of a number of studies on urban housing markets in which income and other demander characteristics were intended as proxies for neighborhood quality". He analyzes the specification bias cost of employing a simple model (four unit-specific variables) rather than a more extensive model which adds a list of demographic variables. Comparing the two hedonic indexes estimated over the same data base of a single metropolitan area at a given point in time, the empirical findings indicate little practical impact on the specification bias of the more restricted model.

With regard to Griliches' second question, the specification of the functional form in the price-characteristics relationship, a number of HPM studies in the literature use linear, semilog (dependent variable, price, being logarithmic), or doublelog functional forms. These provide straight interpretations of the estimated coefficients, respectively: the implicit marginal characteristic prices prevailing at a given market equilibrium; the percentage change in the commodity price for a unit change in any of its characteristics; the percentage change in the commodity price for a percentage change in any of its characteristics (elasticity). Economic theory has not yet developed criteria for the choice of functional forms, so most researchers view the choice as an empirical question to be decided by the best data fit. Many HPM applications to urban housing markets have used the Box-Cox transformation-of-variables procedure to allow for the possibility that the best functional form could be non-linear.

2.2 Spatial econometrics models

The field of regional science and urban economics addresses issues related to human spatial behavior in cities, regions, and major geopolitical areas. Standard econometric techniques can be used in the statistical analysis of spatial interaction models and the calibration of regional econometric models, but there are specific aspects of spatial data that are beyond the reach of these techniques (Anselin, 1988 and 1992). Anselin (1988) calls such aspects "spatial effects", whose principal types are: "spatial dependence", also called spatial autocorrelation or association, and "spatial heterogeneity". Spatial econometrics models take these spatial effects explicitly into account.

With regard to the idea of spatial dependence, Cliff and Ord (1973) state that if the presence of a phenomenon in one area (district, city, region) changes the likelihood of its presence in neighboring areas, the phenomenon is said to exhibit spatial autocorrelation. In other words, "Everything is related to everything else, but near things are more related than distant things".[1] Spatial dependence may result from: the arbitrariness of borderlines between spatial units of observation such as districts, cities, states; the presence of spatial externalities such as shared neighborhood characteristics which affect housing prices; and/or spillover effects such as the impact of the price of one housing unit on the price of its adjacent neighbors.

The second spatial effect, spatial heterogeneity, refers to spatial variability (structural instability) in the parameters or even in the functional form. For example, a cross-sectional data set with very different spatial units, such as rich areas in a Southern region and poor areas in the North, may exhibit spatial heterogeneity effects. Although standard econometric techniques such as switching regressions can cope with many of the problems of spatial

[1] *Tobler (1979), "first law of geography", quoted in Anselin (1988).*

heterogeneity, there are instances (such as error terms presenting spatial dependence) in which acknowledgement of the underlying spatial structure can improve the efficiency of the estimation procedures.

The first formal treatment of spatial interaction was by Moran (1948), with the introduction of the idea of binary contiguity. The underlying structure is defined by $0-1$ values, with the value 1 assigned to spatial units having a common border (in the case of spatial areal units), or within a critical cut-off distance (in the case of point pattern spatial units). Cliff and Ord (1973) present a more general approach to express the interaction between two spatial units by using a combination of distance measures (inverse distance, or negative exponentials of distance) and a measure of length of their common border. The formal expression is as follows:

$$w_{ij} = [d_{ij}]^{-\alpha} \cdot [\beta_{ij}]^{\gamma}, \tag{1}$$

where d_{ij} stands for the distance between spatial units i and j, β_{ij} denotes the proportion of the interior boundary of unit i in contact with unit j, and α and γ are parameters.

One distinctive feature of Cliff and Ord's approach, as opposed to Moran's binary contiguity, is the assymmetry of the resulting weights in the former case. Spatial areal units such as counties (or "municípios") are typically suited to have their interaction expressed by expression (1): both the distance between their centers and the relative importance of their common border are taken into account. Point pattern arrangements of spatial units such as store locations or housing units in an urban environment have, as relevant notion of spatial interaction, the distance between any two of them. Examples of distance-based weights matrices are:

(a) the all-or-nothing critical cut-off distance d^* matrix, where any two given units are defined as neighbors if the distance between them is less than d^* (binary contiguity);

(b) the inverse distance matrix ($\alpha = 1$ and $\gamma = 0$ in the expression (1) above);

(c) the squared inverse distance matrix ($\alpha = 2$ and $\gamma = 0$ in the expression (1) above).

The latter two types of weights matrices can also be implemented with a critical cut-off distance, up to which the spatial weights matrix entry follows one of the relationships described above, and beyond which it is assigned the value zero.

Only recently, in the early 90's, have spatial econometrics techniques been used to study hedonic price models, specifically taking into account spatial effects. Can (1990 and 1992) specifies a model in which the price of a housing unit in any location depends not only on its structural and neighborhood characteristics, as in the traditional HPM approach, but on the prices of adjacent units. (The model allows for checking of the strength of the price interdependence.) This approach closely resembles actual characteristics of an urban housing market, where realtors appraise housing units by their relevant individual characteristics and also by the price history of neighboring units.

3. Specification Issues

The consensual view in the HPM literature maintains that the choice of the proper functional form in the price-characteristics relationship is an empirical question given the lack of theoretical basis to anchor any particular specification. Linear, semilog, end doublelog specifications are the most frequently used functional forms; however, the alterative (and more flexible) approach called "Box-Cox procedure" has been often utilized since the early 80's. One example of the Box-Cox technique as a means of searching for the best functional form is the Follain and Jimenez (1985) paper analyzing the demand for housing characteristics in developing countries. Their equation $(FJ, 4)$, reproduced below, represents the price-characteristics relationship:

$$\frac{P^{\lambda} - 1}{\lambda} = \beta_0 + \sum_{j-1}^{m} \beta_j Z_j + u$$

where P is the market price of the housing unit, β is a vector of m regression coefficients, Z is a vector of m housing characteristics, u is a vector of error terms, and λ is a parameter used to transform P to do Box-Cox analysis.

After estimating the parameters of the Box-Cox functional form, FJ express $P_i = \partial P / \partial Z_i$, the unobserved marginal price of the i-th characteristic, as:

$$\hat{P}_i = \hat{\beta}_i \hat{P}^{(1-\hat{\lambda})}$$

derived from their estimated equation $(FJ, 4)$ through algebraic manipulation. It is worth noting that $\lambda = 1$ yields the linear relationship where the estimated P_i is equal to β_i, the only instance in which it does not vary with the observation.

This paper uses the hedonic price relationship in its reduced form, with a general Box-Cox model in two versions: one for the standard HPM approach and another which allows for spatial effects. The model for the standard approach is as follows:

$$\frac{P^{\lambda}-1}{\lambda} = \beta_0 + \sum_{i=1}^{m} \beta_i \left(\frac{Z_i^{\Theta}-1}{\Theta}\right) + u \tag{2}$$

where P is the price of the housing unit, β is a vector of m regression coefficients, Z is a vector of m housing characteristics, u is a vector of error terms, and λ and Θ are parameters used to transform P and Z, respectively, to do Box-Cox analysis.

The expression for P_i is derived analogously as that of FJ:

$$\hat{p} = \left[(1+\hat{\lambda}\hat{\beta}_0) + \hat{\lambda}\sum_{1}^{m}\hat{\beta}_i\,\frac{(Z_i^{\hat{\Theta}}-1)}{\hat{\Theta}}\right] 1/\hat{\lambda}$$

$$\frac{\partial P}{\partial Z_i} = \hat{P}_i = \hat{\beta}_i Z_i^{(\hat{\Theta}-1)} \left[(1+\hat{\lambda}\hat{\beta}_0) + \hat{\lambda}\sum_{1}^{m}\hat{\beta}_i\,\frac{(Z_i^{\hat{\Theta}}-1)}{\hat{\Theta}}\right] 1/\hat{\lambda}$$

$$\hat{P}_i = \hat{\beta}_i Z_i^{(\hat{\Theta}-1)}\hat{p}^{(1-\hat{\lambda})} \tag{3}$$

For any λ, $\Theta = 1$ yields the hedonic equation linear in the characteristics analyzed by FJ. $\lambda = 1$ and $\Theta = 1$ yield the linear hedonic equation whose coefficents are the implicit marginal characteristic price. For any other pair of values of λ and Θ, the estimated value of the unobserved marginal price of the i-th characteristic varies by observation, but a mean value may be computed by averaging the $P_i s$ corresponding to each observation.

To analyze spatial effects in the hedonic price-characteristic relationship, the following spatial lag model[2] is considered:

$$\frac{P^{\lambda}-1}{\lambda} = \beta_0 + \rho W\left(\frac{P^{\lambda}-1}{\lambda}\right) + \sum_{i=1}^{m} \beta_i \left(\frac{Z_i^{\Theta}-1}{\Theta}\right) + \varepsilon \tag{4}$$

where P is a vector of N observations of the dependent variable, price of a housing unit, β is a vector of m regression coefficients, ρ is the coefficient

[2] *See Anselin (1988) for a full discussion.*

of the spatially lagged dependent variable, W is an $N \times N$ spatial weights matrix, Z is a vector of m housing characteristics, ε is a vector of N error terms, and λ and Θ are parameters used to transform P and Z, respectively, to do Box-Cox analysis.

The spatial weights matrix W plays a role similar to the time-lag operator in a time-series modeling context. It is built in such a way that each row and matching column correspond to an observation pair, ij, and its entry value signals when these observations i and j are considered to be neighbors. The relevant set of neighbors for each observation can be defined as either those that share a border, i.e., simple contiguity (for areal units), or those that are within a critical distance (for point data), as in the urban housing market.

The rejection of the hypothesis, $H_0 : \rho = 0$, implies the existence of adjacency (spillover) effects in the housing market, i.e., the price of one housing unit affects the prices of neighboring units.

If there are no spillover effects but there is spatial dependence in the regression error terms (spatial error model), the OLS estimates remain unbiased but the t- and F- statistics for tests of significance will be biased and the statistical interpretation of the regression model will be incorrect. The spatial error model is specified as follows:

$$P = \sum_{i=0}^{m} \beta_i Z_i + \varepsilon \qquad (5)$$
$$\varepsilon = \delta W \varepsilon + V$$

where P, β_i, Z_i, W and ε have the same meaning as in equation (3), v is a vector of N independent and normally distributed error terms, and δ is the residual spatial autoregression coefficient. The rejection of the hypothesis $\delta = 0$ implies the presence of spatial dependence in the residuals.

4. Data Description

The database analyzed has price and characteristic information for a sample of Belo Horizonte residential apartments lying within a spatial region of approximately 16km^2. The apartments were included in a market survey of residential prices conducted for the Belo Horizonte municipal government in October 1995 by the Instituto de Pesquisas Econômicas e Administrativas

(Ipead) of the Federal University of Minas Gerais. The apartments' characteristics were drawn from the city's property tax data files, which include variables such as apartment area (square meters), age, availability of garage space, local topography, and the level of public services such as piped water, electricity, and garbage collection. Topography is fairly homogeneous for the studied region, with a uniform index assigned to all apartments by city tax assessors, and this characteristic does not affect their relative market value according to realtors. The region is also well-provided with city services, and there is a homogeneous overall index of their availability. For this study, therefore, the sources of price variation are the area of the housing unit in square meters, its age, and the availability of a garage space.

To build the spatial weights matrix, the geographic information necessary was derived from a city map of scale 1cm = 25,000cm. The average distance between any two housing units in the sample is 2.5km, and the maximum distance is 6.5km.

5. Empirical Results

The following section presents four specifications of the standard hedonic price model (HPM): semilog, doublelog, linear, and a nonlinear form estimated by means of the Box-Cox transformation procedure, all of them commonly used in urban housing market studies. The empirical analysis presents also the same specifications with an additional spatially lagged dependent variable to test for spatial effects. Two weights matrices are considered here: the first one has its element w_{ij} set to 1, if the distance between units i and j d_{ij} is less or equal to 1.5km (critical cut-off distance), and zero otherwise; the second one has its element set to $1/d_{ij}$, where d_{ij} is the distance between units i and j. A similar specification of hedonic housing price model including spatial effects is estimated by Can (1992), testing three hypotheses about the spatial interaction of housing units summarized by:

(a) a critical cut-off distance weights matrix;

(b) an inverse distance weights matrix;

(c) an inverse square distance weights matrix.

The Box-Cox procedure is implemented by iterated ordinary least squares, following Spitzer (1981). The systematic grid search checks the residual sum of squares (RSS) for different pairs of the parameters (λ, θ) within the range $[-1.25, 1.25]$, in increments of 0.25, by using the statistical package Limdep.[3] The best fit in this search is given by the pair of parameters $\lambda = -0.25$ (dependent variable) and $\theta = 0$ (independent variables). The Box-Cox best fit estimation with spatial effects uses the Box-Cox transformed variables obtained with the estimated parameters $\lambda = -0.25$ and $\theta = 0$.

Table 1 presents the summary statistics of the analyzed variables. The transformed price variables for $\lambda = 0$ and $\lambda = -0.25$ are shown with corresponding mnemonic names $P0$ and $P - 025$, and the transformed exogenous variables area $(\theta = 0)$ and age $(\theta = 0)$ are assigned names $A0$ and $I0$, respectively.

Tables 2 to 5 present the estimates of the four functional forms considered in this work – semilog, doublelog, linear, and Box-Cox best fit – for both the standard hedonic price model and the hedonic regression with spatial effects. All estimations were computed using the software Spacestat (Anselin, 1990), which provides regression diagnostics for multicollinearity, normality, heteroskedasticity, and spatial dependence in the case of the spatial effects model.

The four specifications use the following transformed variables: price, P (parameter λ); area of the housing unit, A (parameter θ); age, I (parameter θ); and the indicator of garage availability, G (a binary 0-1 variable in all specifications). In terms of Box-Cox parameters, values of (λ, θ) equal to (1,1), (0,1), (0,0), (−0.25, 0) correspond to the linear, the semilog, the doublelog, and the "Box-Cox best fit" specifications, respectively. After Box-Cox best fit $(\lambda = -0.25;\ \theta = 0)$, the best statistical results are obtained by the doublelog $(\lambda = 0;\ \theta = 0)$ and semilog $(\lambda = 0;\ \theta = 1)$ specifications, in that order.

The implicit marginal price of the i-th attribute in equation (3) is trivially equal to the estimated regression coefficient β_i for the linear specification correspondent to the set of values $(\lambda = 1,\ \theta = 1)$ of the transformed parameters in the Box-Cox general formulation. For all other possible sets of values of (λ, θ), the implicit marginal price of the i-th attribute does not have a straightforward interpretation because it depends on the sample predicted value of P on the right-hand side of equation (3).

[3] *Software by William H. Greene (1992).*

Table 1

Summary statistics

Variable	Mean	Variance	Standard deviation	Skewness	Kurtosis	Min	Max	Range
P	72330.19	2.450735E+09	49504.9	1.860869	6.188621	21500	245000	223500
A	132.4768	4071.908	63.8115	1.412476	5.311006	50.87	368	317.13
I	18.03774	89.47027	9.458873	-0.2272928	1.93313	1	34	33
$P0$	11.00997	0.329855	0.57433	0.4979435	2.871522	9.975808	12.40901	2.433205
$A0$	4.784981	0.1964776	0.4432579	0.2700276	2.59841	3.92927	5.90808	1.97881
$I0$	2.643905	0.7244319	0.8511357	-1.395625	4.093176	0	3.52636	3.52636
$P-025$	3.742347	0.001282378	0.0358103	0.1530793	2.61091	3.66967	3.82021	0.15054

Notes: P = price in Brazilian reais (R$); A = area of apartment in square meters; I = age; $P0$ = transformed price variable ($\lambda=0$, i.e., $\ln P$); $A0$ = transformed area variable ($\theta=0$, i.e, $\ln A$); $I0$ = transformed age variable ($\theta=0$, i.e., $\ln I$); $P-025$ = transformed price variable ($\lambda=-0.25$).

However, the estimated attribute coefficients have an economic interpretation in the cases of both the semilog specification ($\lambda = 0$, $\theta = 1$ in equation (3)) and the doublelog specification ($\lambda = 0$, $\theta = 0$ in equation (3)): they indicate the percentage change in the housing price with a unit change in any of the attributes for the former one, and represent the percentage change in the housing price with a percentage change in any of the attributes (elasticity of price with respect to the attribute) for the latter one.

Coefficients of the attribute *area* and the dummy variable *garage* are in general significant (the linear hedonic model is one example of poor fit in which the coeficient of *garage* is not significant) and relatively close in magnitude in spatial autoregressive models estimated using the two types of weights matrices. Surprisingly, *age* is not statistically significant, although it has the expected negative sign: its most significant result falls close to the 10% level of significance in the "critical cut-off distance doublelog" specification (table 4).

The presence of spillover (adjacency) effects on housing prices is strong in all specifications for both the critical cut-off distance weights matrix and the inverse distance weights matrix, but the first one seems to describe better (better fit) the spatial interaction among the housing units in the sample. Rejection of the null hypothesis ($H_0 : \rho = 0$ in equation (4)) of no spatial lag dependence is highly significant for all specifications under either assumption of spatial interaction ("critical cut-off distance" or "inverse distance"), and non-rejection of the null hypothesis of no spatial error dependence ($H_0 : \lambda = 0$ in equation (5)) is also common to all specifications.

Comparison of the statistically significant coefficients estimated using semilog (table 3), doublelog (table 4), and Box-Cox best fit (table 5) specifications (standard HPM and HPM with spatial lag) indicates how spatial effects change the way attributes affect the prices of housing units. For example, looking at the hedonic model and its critical cut-off distance counterpart in the semilog specification, figures show that the percentage change in the housing price with a unit change in *area* is reduced from 0.00716 to 0.00661 after controling for spatial effects.

The marginal prices of attributes in the Box-Cox best fit (BCBF) specification depends on the sample predicted values of the price P (equation (3)) and the estimated coefficients have no straightforward economic interpretation as in the case of the semilog or doublelog specifications. However, the functional form BCBF parameters ($\lambda = -0.25$, $\theta = 0$) "signal" (by being "close" to) the doublelog specification ($\lambda = 0$, $\theta = 0$) as a better way of modeling the relationship between housing prices and attributes than the other often employed linear ($\lambda = 1$, $\theta = 1$) and semilog ($\lambda = 0$, $\theta = 1$) specifications. It is worth mentioning once more that the coefficients of BCBF and doublelog are not comparable: table 1 shows that the range of values of the variable $P - 025$ ($\lambda = -0.25$) is much narrower than the one of the variable $P0(\lambda = 0)$, 0.15 and 2.43 respectively.

Table 2

Linear hedonic model (LHM-OLS) and linear spatial autoregressive specification (LSAS-MLE) dependent variable = P (price)[a]

Variable	LHM	LSAS, critical distance matrix[b]	LSAS, inverse distance matrix[c]
Constant	-13425.6	-27461.9	-28700
	(-0.795)	(-1.875)	(-1.841)
W_P (spatial	–	0.02105	0.00727
lag price)		(4.023)	(3.247)
A (area)	635.23	588.09	596.26
	(10.134)	(10.785)	(10.592)
I (age)	-239.98	-300.63	-141.04
	(-0.516)	(-0.766)	(-0.344)
G (garage,	8272.3	8161.4	6850.8
dummy)	(0.920)	(1.075)	(0.868)
R^2	0.7535	–	–
LIK (value of likelihood FN)	-611.02	-604.33	-606.27
Test for spatial lag dependence[d]	–	13.383	9.496
		(0.00025)	(0.00206)
Test for spatial error dependence[e]	–	0.03093	0.45242
		(0.86040)	(0.50119)

[a] Number of observations = 53; values for each variable are the coefficient estimates; T-statistics (in parentheses) based on OLS variance estimates for LHM; asymptotic T-statistics (in parentheses) based on maximum likelihood estimation (MLE) for the two types of weights matrix chosen in LSAS.

[b] Weights matrix elements w_{ij} defined as follows: $w_{ij} = 1$ if $d_{ij} \leq 1.5$km (critical cut-off distance), $w_{ij} = 0$ otherwise.

[c] Weights matrix elements w_{ij} such that $w_{ij} = 1/d_{ij}$.

[d] Values in parentheses are the probability levels for the likelihood ratio test of the null hypothesis of no spatial dependence.

[e] Values in parentheses are the probability levels for the lagrange multiplier test of the null hypothesis of no spatial dependence in the error structure.

Table 3

Semilog hedonic model (SLHM-OLS) and semilog spatial autoregressive specification (SLSAS-MLE) dependent variable = LNP (log-price)[a]

Variable	SLHM	SLSAS, critical distance matrix[b]	SLSAS, inverse distance matrix[c]
Constant	9.9455	9.6717	9.6114
	(53.9953)	(60.4496)	(50.8293)
W_P (spatial lag log-price)	–	0.00263	0.00092
		(4.678)	(3.367)
A (area)	0.00716	0.00661	0.00683
	(10.4782)	(11.7224)	(11.3036)
I (age)	-0.00237	-0.00537	-0.00081
	(-0.4669)	(-1.0896)	(-0.18203)
G (garage, dummy)	0.22036	0.23327	0.21699
	(2.2470)	(2.9428)	(2.5364)
R^2	0.7821	–	–
LIK (value of likelihood FN)	-5.4390	3.7925	-0.2810
Test for spatial lag dependence[d]	–	18.463	10.316
		(0.00002)	(0.00132)
Test for spatial error dependence[e]	–	0.39043	0.34677
		(0.53207)	(0.55595)

[a] Number of observations = 53; values for each variable are the coefficient estimates; T-statistics (in parentheses) based on OLS variance estimates for SLHM; asymptotic T-statistics (in parentheses) based on maximum likelihood estimation (MLE) for the two types of weights matrix chosen in SLSAS.

[b] Weights matrix elements w_{ij} defined as follows: $w_{ij} = 1$ if $d_{ij} \leq 1.5$km (critical cut-off distance), $w_{ij} = 0$ otherwise.

[c] Weights matrix elements w_{ij} such that $w_{ij} = 1/d_{ij}$.

[d] Values in parentheses are the probability levels for the likelihood ratio test of the null hypothesis of no spatial dependence.

[e] Values in parentheses are the probability levels for the lagrange multiplier test of the null hypothesis of no spatial dependence in the error structure.

Table 4

Doublelog hedonic model (DLHM-OLS) and doublelog spatial autoregressive specification (DLSAS-MLE) dependent variable = LNP (log-price)[a]

Variable	DLHM	DLSAS, critical distance matrix[b]	DLSAS, inverse distance matrix[c]
Constant	5.9988 (12.3152)	5.9778 (15.1453)	5.8326 (13.3730)
W_P (spatial lag log-price)	–	0.00239 (4.639)	0.00079 (3.019)
LNA (log-area)	1.05133 (11.7851)	0.98316 (13.3344)	1.00937 (12.5520)
LNI (log-age)	-0.05374 (-1.1153)	-0.05922 (-1.5163)	-0.02937 (-0.6745)
G (garage, dummy)	0.17104 (1.9747)	0.19919 (2.8276)	0.17454 (2.2688)
R^2	0.8170	–	–
LIK (value of likelihood FN)	-0.8015	8.2706	3.4096
Test for spatial lag dependence[d]	–	18.144 (0.00002)	8.422 (0.00371)
Test for spatial error dependence[e]	–	0.00643 (0.93607)	0.28329 (0.59459)

[a] Number of observations = 53; values for each variable are the coefficient estimates; T-statistics (in parentheses) based on OLS variance estimates for DLHM; asymptotic T-statistics (in parentheses) based on maximum likelihood estimation (MLE) for the two types of weights matrix chosen in DLSAS.

[b] Weights matrix elements w_{ij} defined as follows: $w_{ij} = 1$ if $d_{ij} \leq 1.5$km (critical cut-off distance), $w_{ij} = 0$ otherwise.

[c] Weights matrix elements w_{ij} such that $w_{ij} = 1/d_{ij}$.

[d] Values in parentheses are the probability levels for the likelihood ratio test of the null hypothesis of no spatial dependence.

[e] Values in parentheses are the probability levels for the lagrange multiplier test of the null hypothesis of no spatial dependence in the error structure.

Table 5

Box-Cox best fit hedonic model (BCBFHM-OLS) and
Box-Cox best fit spatial autoregressive specification (BCBFSAS-MLE)
dependent variable = Box-Cox transformed variable P (BCBF-price)[a]

Variable	BCBFHM	BCBFSAS, critical distance matrix[b]	BCBFSAS, inverse distance matrix[c]
Constant	3.4279 (114.81)	3.4252 (138.38)	3.417 (125.83)
W_P (spatial lag BCBF-price)	–	0.00042 (4.296)	0.00014 (2.752)
A (BCBF-area)	0.06547 (11.9728)	0.06177 (13.3927)	0.06322 (12.6999)
I (BCBF-age)	-0.00290 (-0.9827)	-0.00329 (-1.3435)	-0.00144 (-0.5327)
G (garage, dummy)	0.01235 (2.3256)	0.01404 (3.1786)	0.01256 (2.6311)
R^2	0.8232	–	–
LIK (value of likelihood FN)	147.179	155.053	150.729
Test for spatial lag dependence[d]	–	15.749 (0.00007)	7.212 (0.00724)
Test for spatial error dependence[e]	–	0.06376 (0.80065)	0.1287 (0.71976)

[a] Number of observations = 53; values for each variable are the coefficient estimates; T-statistics (in parentheses) based on OLS variance estimates for BCBFHM; asymptotic T-statistics (in parentheses) based on maximum likelihood estimation (MLE) for the two types of weights matrix chosen in BCBFSAS; Box-Cox optimal parameters are -0.025 for P, and 0 for both A and I, i.e., their respective logs.

[b] Weights matrix elements w_{ij} defined as follows: $w_{ij} = 1$ if $d_{ij} \leq 1.5$km (critical cut-off distance), $w_{ij} = 0$ otherwise.

[c] Weights matrix elements w_{ij} such that $w_{ij} = 1/d_{ij}$.

[d] Values in parentheses are the probability levels for the likelihood ratio test of the null hypothesis of no spatial dependence.

[e] Values in parentheses are the probability levels for the lagrange multiplier test of the null hypothesis of no spatial dependence in the error structure.

The fact that the housing market appraises residential units taking into consideration the historical evolution of prices in their near neighborhood, in addition to other attributes, lends support to the critical cut-off distance weights matrix as a way of describing the spatial interaction in the market. It is worth recalling that the sample analyzed has an average distance between units equal to 2.5km, a maximum distance between units equal to 6.5km, and the critical cut-off distance is set to 1.5km.

6. Conclusion

This study examines sources of price variation in a sample of residential apartments of Belo Horizonte, Brazil, using both standard and spatial autoregressive hedonic price models. The latter specification checks for the possibility that the market appraises a residential unit taking into consideration the historical evolution of prices in its near neighborhood as well as other characteristics. The Box-Cox transformation procedure is used to choose the best non-linear functional form.

The results show that spillover (adjacency) effects are an important source of price variation in this housing market. Rejection of the null hypothesis of no spatial lag dependence is highly significant for all specifications under any of the assumptions of spatial interaction ("critical cut-off distance" or "inverse distance"). For the sake of comparison, other specifications often employed in analyzing housing markets are also estimated, including the linear, semilog and doublelog forms. Among these functional forms, the results suggest that the doublelog specification describes better the relationship between housing prices and attributes than the other ones.

These empirical findings support the need to incorporate spatial effects in studies of housing price determination. The importance of incorporating spatial factors makes the use of special spatial econometrics techniques essential to the analysis of economic performance in housing markets.

References

Anselin, Luc. *Spatial econometrics: methods and models.* Dordrecht, Kluwer Academic, 1988.

________. *What is special about spatial data? Alternative perspectives on spatial data analysis.* Ucla-Santa Barbara Geography Department, National

Center for Geographic Information & Analysis (NCGIA), 1989. (Technical Paper, 89-4.)

________. SpaceStat, a program for the statistical analysis of spatial data. *Manual of the preliminary beta release.* 1990.

________. *Spatial data analysis with GIS: an introduction to application in the social sciences.* Ucla-Santa Barbara Geography Department, National Center for Geographic Information & Analysis (NCGIA), 1992. (Technical Report, 92-10.)

Bailey, M. J.; Muth, R. F. & Nourse, H. O. A regression method for real estate price index construction. *Journal of the American Statistical Association, 58*:933-42, 1963.

Butler, Richard V. The specification of hedonic indexes for urban housing. *Land Economics, 58*(1):96-108, Feb. 1982.

Can, Ayse. The measurement of neighborhood dynamics in urban house prices. *Economic Geography, 66*(3):254-72, 1990.

________. Specification and estimation of hedonic housing price models. *Regional Science and Urban Economics, 22*(3):453-74, 1992.

Chow, Gregory C. Technological change and the demand for computers. *American Economic Review, 57*:1.117-30, 1967.

Cliff, A. D. & Ord, J. K. *Spatial autocorrelation.* London, Pion, 1973.

Dhrymes, P. J. On the measurement of price and quality changes in some consumer capital goods. *American Economic Review, 57*(2):501-18, 1967.

Follain, James R. & Jimenez, Emmanuel. Estimating the demand for housing characteristics. *Regional Science and Urban Economics, 15*:77-107, 1984.

________ & ________. The demand for housing characteristics in developing countries. *Urban Studies, 22*:421-32, 1985.

Greene, William H. *Limdep version 6.0.* Econometric Software, 1992.

Griliches, Zvi (ed.). *Price indexes and quality change.* Cambridge, Harvard University Press, 1971.

Megbolugbe, Isaac F. A hedonic index model: the housing market of Jos, Nigeria. *Urban Studies, 26*(5):486-94, Oct. 1989.

Rosen, Sherwin. Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy, 82*:34-55, Jan./Feb. 1974.

Spitzer, John J. A primer on Box-Cox estimation. *The Review of Economics and Statistics, 64*:307-13, 1981.