UCLM
UNIVERSIDAD DE CASTILLA-LA MANCHA

Facultad de Ciencias Sociales de Cuenca

Documento de trabajo
SEMINARIO PERMANENTE DE CIENCIAS SOCIALES

MEASURING THE IMPACT OF POLLUTION ON PROPERTY PRICES: OBJECTIVE VS. SUBJECTIVE POLLUTION INDICATORS IN SPATIAL MODELS

José María Montero Lorenzo
Román Mínguez Salido
Gema Fernández-Avilés Calderón

SPCS Documento de trabajo 2011/9
http://www.uclm.es/CU/csociales/DocumentosTrabajo

Facultad de Ciencias Sociales de Cuenca | Avda. de los Alfares, 44 | 16.071–CUENCA
Teléfono (+34) 902 204 100 | Fax (+34) 902 204 130

Autor:
José María Montero Lorenzo
Jose.MLorenzo@uclm.es
Román Mínguez Salido
Roman.Minguez@uclm.es
Gema Fernández-Avilés Calderón
Gema.FAviles@uclm.es

Edita:

Facultad de Ciencias Sociales de Cuenca
Seminario Permanente de Ciencias Sociales
Codirectora: Silvia Valmaña Ochaíta
Codirectora: María Cordente Rodríguez
Secretaria: Pilar Domínguez Martínez
Avda. de los Alfares, 44
16.071–CUENCA
Teléfono (+34) 902 204 100
Fax (+34) 902 204 130
http://www.uclm.es/CU/csociales/DocumentosTrabajo

I.S.S.N.: 1887-3464 (ed. Cd-Rom) 1988-1118 (ed. en línea).
D.L.: CU-532-2005

Impreso en España – Printed in Spain.

MEASURING THE IMPACT OF POLLUTION ON PROPERTY PRICES: OBJECTIVE VS. SUBJECTIVE POLLUTION INDICATORS IN SPATIAL MODELS

José-María Montero Lorenzo, Román Mínguez Salido[1] y Gema Fernández-Avilés Calderón

Universidad de Castilla-La Mancha

ABSTRACT

Much work has been done in the context of the hedonic price theory to estimate the impact of air quality on housing prices. Hedonic specifications have improved enormously compared to the early models and current research even considers the spatial argument as a key factor. However, in the best of cases, empirical research only slightly confirms the hedonic theory. These empirical results go against both common sense and theory, which led us to suspect that the problem is not specification but the way air quality is measured. Research has been conducted using objective measures of air quality, but probably what house buyers include in their utility function is their perception of such quality. Thus, subjective measures are needed.

In this article we propose a kind of spatial hedonic models and compare the results obtained with objective and subjective measures of air quality in Madrid (Spain). Results are quite different and suggest that perceived air quality measures are the variable to be included when applying a hedonic house price model.

Key words: housing prices, environment, subjective measure of pollution, spatial hedonic models, cluster, kriging.

JEL-codes: Q51, R32, C23.

[1] Roman.Minguez@uclm.es

1. INTRODUCTION

According to the theoretical literature, it is reasonable to assume that air pollution enters into the utility function of potential house buyers. It is therefore no surprise that hedonic house price models that incorporate environmental variables among the set of explanatory variables are becoming increasingly popular.

In the context of the hedonic price theory, the traditional approach to this problem has been to use the housing market to infer the implicit prices of these nonmarket goods (see Freeman 1993 for a comprehensive review of property value models for measuring the value of environmental amenities). Under standard assumptions of perfect competition, information and mobility and the maximization of well-behaved preferences, hedonic theory unambiguously predicts that the implicit price function relating housing prices to an environmental amenity will be positively sloped, all else equal.

But the question is that empirical research does not confirm the hedonic theory. Smith and Huang’s (1995) meta-analysis suggested that a one-unit reduction of PM10 (mg/m3) results in a 0.05-0.10 percent increase in property values. Boyle and Kiel (2001) found that air studies produced mixed results and posited that measurement factors are not generally known to homebuyers. Jackson (2001) offered no final observations on the consistency of findings and he called for a more systematic study. Simons and Saginor (2006), who included air pollution together with concentrated animal feeding operations, obtained a different sign in the coefficient of the variable depending on whether the model included positive amenities or not.

In summary, the literature concerning the effects of contamination on property values reveals that the effect of air pollution on property value is far from being conclusive. What is more, there are serious doubts that air pollution significantly affects the price of properties. Additionally, the study type may also generate different results.

Recently a successful line of research has emerged that includes the spatial argument in the hedonic specification. As Straszheim (1988) stated many years ago, it may not be appropriate to assume that the implicit prices of housing attributes are stationary –we use the same term as Carruthers and Clark (2010)– across geographical space, even within a big city. The rationale behind this is that, on the supply side, homes near each other tend to be similar and, on the demand side, homebuyers regularly emulate one another's behavior. The result is a process of spatial interaction among market participants, which at least suggests that the first stage hedonic price function should be modified to include a spatial lag of its dependent variable (Anselin, 1988; Anselin and Bera, 1998). This spatial lag can be interpreted as a flexible fixed effect that absorbs the existing and unobserved spatial correlation structure of supply and/or demand. Recommended literature that considers the spatial argument in the specification of the hedonic model includes the pioneer works of Can (1990), Can (1992), Kim *et al.*, (2003), Theebe (2004), Brasington and Hite (2005), Anselin and Le Gallo (2006), Anselin and Lozano-Gracia (2008) and Osland (2010), among many others.

But again, results are not conclusive. In the best of cases, clean air has a negligible influence on housing prices, which does not fit the hedonic theory.

What is wrong in the preliminary studies? Maybe nothing is wrong but, as stated in Chay and Greenstone (1998), exogenous differences in air quality are extremely difficult to isolate, because the "true" relationship between air pollution and property prices may be obscured in cross-sectional analysis by unobserved determinants of housing prices that co-vary with air pollution. For example, areas with high levels of pollution tend to have well educated populations with higher per capita income and population densities. Economic activity is also a driving force in the determination of property values, but differences across space in the level of activity may also be behind changes in the level of pollution. Of course, the above circumstances lead to a spurious positive relationship between pollution and property values. When Chay and Greenstone use the conventional cross-sectional estimates of the relationship between property values and PM_{10} they conclude that the relationship is weak, unstable and indeterminate.

The other possibility is that the concept of air pollution that enters into the utility function of potential house buyers is perceived pollution rather than objective (measured) pollution. It is important to bear in mind that indirect methods like the hedonic strategy are based on actual transactions and empirical measurements and assume that decision makers possess all the necessary information and always act rationally, attempting to maximize their personal utility. However, when in the process of deciding their location, house buyers weigh up one property or location against another, their choices are not necessarily rational. As Berezansky *et al*. (2010) states, their rationality is essentially bounded by available information, limited processing capacity, errors of judgment, and an inability to foretell the future. Additionally, perceived pollution is not an instantaneous concept like, at least during a short period of time, the price of a dwelling, but one that is created over a long period and needs another long period of time (due to the above mentioned factors) to be modified.

In our opinion, one more question remains to be answered: When measuring air quality in large cities, what do house buyers (and citizens in general) understand as clean air? Our previous research (see Mínguez *et al*., 2010) suggests that, at least in European cities, people decide the location of their house according to a range of factors, including: personal or family income, commercial properties, communications, schools, medical centers, etc. but not according to the level of pollution, as living in the city alone implies a polluted environment. What do they mean by a “clean air” location in a large city? Probably a neighborhood located near parks and open areas. This suspicion has been confirmed by Berezansky *et al*. (2010) in the case of the City of Haifa. This is another important reason in favor of using subjective air pollution variables in spatial housing price hedonic models.

In brief, house buyers do not decide how much they are willing to pay for clean air on the basis of the complete information provided by monitoring stations (they probably do not even know how to interpret it), but rather according to their perceptions, perceptions that are not instantaneous but generated over a long period of time. Additionally, in the case of using environmental interpolated variables as explanatory variables, “objective" and "subjective" air pollution maps could be quite different.

As a consequence, if there is not a strict positive correlation between objective and subjective measurements of pollution, the literature, regardless of whether or not it includes the spatial argument, is not using the right pollution variable or index.

The above statements suggest classifying neighborhoods according to subjective environmental measures (resident perceptions), which could be enormously more informative than objective environmental measures when it comes to accounting for the willingness to pay for clean air or other environmental factors.

We have focused our empirical analysis on Madrid (Spain). There are several important reasons for choosing Madrid as a study case: (i) Most of the empirical research in the literature refers to American cities. This is the main reason; (ii) population is highly concerned with the environment in general and air quality in particular (iii) construction, particularly residential building, is of great importance to the overall economy; and (iii) it can be said that in Madrid there is almost perfect information about air quality all over the city (both excellent ratios monitoring sites/population and monitoring sites/surface), which makes it possible to ignore the problems related to how much is known about air quality variables, because, as is well known, the impact of air quality variables on the hedonic price function depends on how much is known about them (Clark and Allison, 1999).

We constructed a massive data base in 2009 including 11,796 dwellings (after depuration). Apart from price, we have registered more than thirteen core variables for each along with subjective and objective air pollution indicators representative of the level of pollution.

After this introduction, section 2 is devoted to both briefly delineating the way we propose to construct APIs based on interpolated objective pollution measures and the construction and comparison of both objective and subjective environmental air quality maps. In section 3 we briefly describe the Spatial Durbin Model, which is the specification that we use in this article to measure the impact of clean air on housing prices. Section 4 is devoted to the case study: Madrid City. First we give some details about air pollution (both objective and subjective measures) and the housing sector in the study area. Second, we comment on both the air quality and housing data sets used in this research. Third, we report the main results obtained from the inclusion of subjective environmental measures in the spatial hedonic specification proposed in Section 3. Finally, some concluding remarks are reported in Section 5.

2 OBJECTIVE AND SUBJECTIVE AIR QUALITY MAPS IN MADRID CITY: THE BIG DIFFERENCE

The objective level of pollution is usually measured by monitoring stations and then the objective pollution estimation approach consists of the following stages:

(i) Measurements of the selected pollutants are collected.

(ii) In some cases, an Air Pollution Index (API) is created (Montero and Fernández-Avilés, 2009a, b; Montero *et al.*, 2010). This is not usually the case. Researchers normally use one or two pollutants, the most visible in smog.

(iii) The values of some specific pollutants or the API are kriged over the area under study or at the points where sampled dwellings are located. This step allows us to match dwelling prices and objective pollution variables.

As is well known, kriging is a univariate procedure which interpolates the values of the target random function at unobserved locations using the available observations of the same random function. This interpolation produces the best linear unbiased estimator and uses the covariance or variogram function (the spatial equivalent of the autocorrelation function in time series analysis) to account for the correlation structure in making interpolative estimates. Kriging techniques are the usual strategy for estimation because they take into account the spatial dependencies of pollution values, which is an extremely important feature to take into account when it comes to estimating pollution variables (see Cressie 1993 for details).

However, kriging, and specifically ordinary kriging, the procedure recommended in Anselin and Lozano-Gracia (2008), has some drawbacks that could make it unsuitable for measuring the impact of air quality on dwelling prices. First, the amount of actual data on pollution is enormously scarce with respect to the number of dwellings in the sample because there used to be between 20 and 30 monitoring stations in the best of cases in large cities. This means that estimates are needed for practically all locations where dwelling prices are sampled. Second, it is an interpolative method and, as a consequence, provides a weighted mean of the values of neighboring locations as estimates. That means that "peak-estimates" cannot be obtained, "peak-estimates", the value of which exceeds the values of neighboring locations. Obviously this feature makes the OK strategy strongly dependent on where monitoring stations are located. And we have to bear in mind that the location of monitoring stations in a large city does not obey statistical criteria (design of an optimal monitoring network) but other criteria that include municipal regulations. And third, in large cities, most of the monitoring stations are located in the city centre, only a few being on the periphery. This implies that in peripheral neighborhoods, OK estimates tend to be near to the mean pollution value and their variance is higher than desirable.

These three drawbacks (among others) make OK estimates too unreliable to be included in a spatial (or not) hedonic housing price model to evaluate the impact of air pollution on dwelling prices and could be one of the reasons why empirical results contradict or, in the best of cases, slightly confirm, the hedonic theory.

In order to form neighborhood clusters with objective measures of pollution, following EU directives, we have interpolated (OK) the values of sulphur dioxide (SO_2), nitrogen oxides (NO_x) —which is a generic term for mono-nitrogen oxides (nitric oxide (NO) and nitrogen dioxide (NO_2))—, carbon monoxide (CO), particulate matter (PM_{10})' and ground-level ozone (O_3), in the non observed neighborhoods. Subsequently, we have created an API for the complete set of neighborhoods in Madrid. This is the inverse procedure to that used in the literature (first computing the index and then kriging it), but it can be shown (Myers 1983) that it leads to lower error variance.

Indeed, let the variables of different pollutants, $X_1, X_2, \ldots, X_K$, be intrinsic stationary stochastic processes of order zero. The two options to linearly estimate an API are:

(i) Elaborating a synthetic index with the *K* environmental variables provided by the *n* monitoring stations, $API(\mathbf{s}_i)$, and after that computing the kriged estimates of this index for the total number of *m* neighborhoods:

$$API^*\left(\mathbf{s}_j\right) = \sum_{j=1}^{m} \lambda_i API(\mathbf{s}_i), \qquad j = 1, \ldots, m$$

for $API(\mathbf{s}_i) = \sum_{k=1}^{K} a_k X_k(\mathbf{s}_i) = \mathbf{A}'\mathbf{X}$, $\mathbf{A}' = (a_1, \cdots, a_K)$ and $\mathbf{X} = \left[X_1(\mathbf{s}_i), \cdots, X_K(\mathbf{s}_i)\right]$ being the vectors of Principal Components or DP2 weights (Montero *et al*., 2009) and variables, respectively.

(ii) Kriging each original variable $X_1(\mathbf{s}),\cdots,X_K(\mathbf{s})$ for the *m* neighborhoods, and then computing the synthetic index of the interpolated variables $X_1^*(\mathbf{s}),\cdots,X_K^*(\mathbf{s})$ as follows:

$$API(\mathbf{s}_j) = \mathbf{A}'\mathbf{X}_j^* = \sum_{k=1}^{K} a_k X_k^*(\mathbf{s}_j) = \sum_{k=1}^{K}\sum_{i=1}^{n} a_k \lambda_i \, X_k(\mathbf{s}_i)$$

Following Myers (1983, pp.634), it can be demonstrated that:

$$Var\left[API^*(\mathbf{s}_j) - API(\mathbf{s}_j)\right] > Var\left[API(\mathbf{s}_j) - API(\mathbf{s}_j)\right]$$

Finally we have classified neighborhoods into four groups. For this purpose, we have conducted a model-based cluster procedure implemented in the MCLUST algorithm. This procedure is based on the assumption that data are generated by normal multivariate distributions with different covariance matrices. That is to say, the data generation processes are a mixture of normal distributions (Fraley and Raftery, 1999, 2002, 2010). The covariance matrices are decomposed in terms of volume, shape and orientation −which allows for the definition of a range of models− and their implied type of distributions, considering a complete range of decomposition possibilities. All the potential models are summarized Table 1 (Fraley and Raftery, 2010).

TABLE 1: Cluster models

Model Identifier	**Distribution**	**Volume**	**Shape**	**Orientation**
EII	Spherical	Equal	Equal	NA
VII	Spherical	Variable	Equal	NA
EEI	Diagonal	Equal	Equal	Coordinate Axes
VEI	Diagonal	Variable	Equal	Coordinate Axes
EVI	Diagonal	Equal	Variable	Coordinate Axes
VVI	Diagonal	Variable	Variable	Coordinate Axes
EEE	Ellipsoidal	Equal	Equal	Equal
EEV	Ellipsoidal	Equal	Equal	Variable
VEV	Ellipsoidal	Variable	Equal	Variable
VVV	Ellipsoidal	Variable	Variable	Variable

The choice of both a specific model and a specific number of groups allows for maximum likelihood estimation of the different group matrices (assuming a mixture of normal distributions). Then, observations can be assigned to a group. To select both the model class and the number of groups, MCLUST algorithms proceeds to maximizing a reparametrization of the BIC criteria where the maximum is taken over all the models and number of potential groups considered.

In our case, the results of applying this methodology for clustering Madrid neighborhoods on the basis of objective measures of SO_2, NO_x, NO_2, CO, PM_{10} and O_3 are summarized in Table 2.

TABLE 2: BIC values for cluster models using objective measures of pollution

Groups	EII	VII	EEI	VEI	EVI	VVI	EEE	EEV	VEV	VVV
1	-911.6	-911.4	-647.9	-647.9	-647.9	-647.9	-79.8	-79.8	-79.8	-79.8
2	-646.7	-617.2	-364.6	-326.7	-341.1	-322.3	-47.3	17.8	13.4	45.8
3	-608.1	-490.6	-271.2	-190.9	-352.6	-94.6	8.2	62.2	215.1	229.8
4	-486.5	-414.5	-215.3	-167.3	-241.3	-98.2	31.4	44.1	117.4	203.9
5	-443.2	-363.7	-215.9	-109.7	-237.7	-46.5	33.1	97.2	143.3	109.6
6	-447.9	-322.2	-250.4	-87.4	-219.7	-43.6	43.9	75.1	74.7	105.5

The maximum of the BIC was found for the VVV model and 3 groups. However, we have opted for the VVV-4 group classification because the value of the BIC is very close to the maximum and facilitates interpretation enormously in view of the well-known facts on pollution in Madrid (see Figure 1a).

On the other hand, the usual way of measuring subjective environmental amenities is annoyance scores (Poor *et al.*, 2001; Jacquemin *et al*., 2007 are recommended references). However, our objective is to rank areas of a city according to both objective and subjective air pollution measurements, and to examine the level of concordance between the neighborhood clusters under the objective and the subjective approach. This is why we prefer census information to survey information, and we subjectively characterized the areas by the percentage of residents that agree or disagree with the environmental feature they were questioned about. The advantages of this approach include non-dimensionality and common range [0-1], which enormously facilitates interpretation. In addition, somehow this approach could be interpreted as a way of averaging the results obtained in surveys based on annoyance scales. It could also be viewed as a consensus measure of air quality.

When it comes to clustering the neighborhoods on the basis on residents' perception of pollution (RP), we use the percentage of residents in the neighborhood who state that the neighborhood is seriously polluted. Following the same procedure as in the objective case, the results of BIC are reported in Table 3.

TABLE 3: BIC values for cluster models using residents' perceived pollution

Groups	**EII**	**VII**	**EEI**	**VEI**	**EVI**	**VVI**	**EEE**	**EEV**	**VEV**	**VVV**
1	640.6	640.6	681.1	681.1	681.1	681.1	829.9	829.9	829.9	829.9
2	766.2	793.1	767.9	793.7	771.8	802.5	823.3	918.6	846.1	922.9
3	809.1	835.1	803.6	842.7	811.6	880.1	863.2	894.1	919.4	920.9
4	806.3	845.7	795.5	843.9	810.4	854.2	855.4	903.5	900.8	889.6
5	789.2	849.2	783.0	864.8	796.5	835.2	833.0	884.1	880.3	856.0
6	792.6	839.5	798.8	850.3	761.4	841.3	867.6	885.8	850.2	815.1

Again, the best model is the VVV, and two groups record the maximum of BIC. However, we can see that there are no significant differences in BIC for 2, 3 or 4 groups. Taking this into account, we decided to choose the VVV-4 group option to both facilitate the interpretation of the cluster and to compare it to the four group cluster selected when clustering was performed in terms of the objective API variable. Figure 1b reports the result of the clustering process with residents' perceptions.

There are large differences between the two neighborhood classifications. More specifically, Pearson's correlation coefficient between objective and subjective values of pollution in Madrid neighborhoods stands at 0.20 and Spearman's ranks coefficient is 0.39.

When comparing the neighborhood clustering obtained using objective estimates of pollution and perceived (subjective) pollution values (Figures 1a, 1b), it can be noticed that both the "objective" and "subjective" clusters coincide in the peripheral neighborhoods, the most depressed area of Madrid in terms of residential construction. But, precisely in these peripheral areas, only the "subjective" cluster classifies the Southeast area of the city in the most polluted group (red neighborhoods). Indeed, the Southeast area of the city is made up of 'arid' neighborhoods with important construction activity in the years prior to the economic crisis and they are very close to industrial areas. Obviously, these circumstances lead to a high degree of suspended particulates. Objective clustering classifies this area in the best group (green one), the reason being both that the neighborhoods included in it are close to the second best area (the light blue one) and also the shortage of monitoring stations in the Southeastern part of the city. Therefore, it is no surprise that OK estimates are below the mean of the city.

FIGURE 1a: Classification of Madrid neighborhoods according to kriged objective measures of pollution

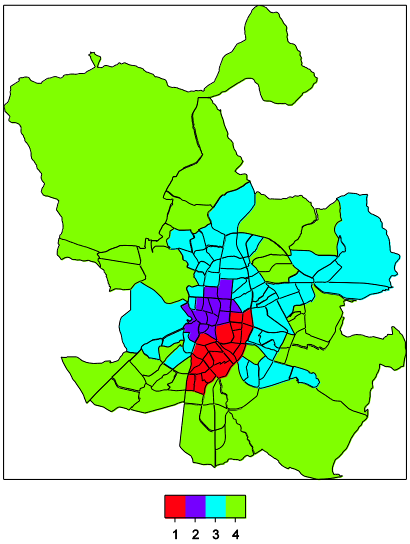
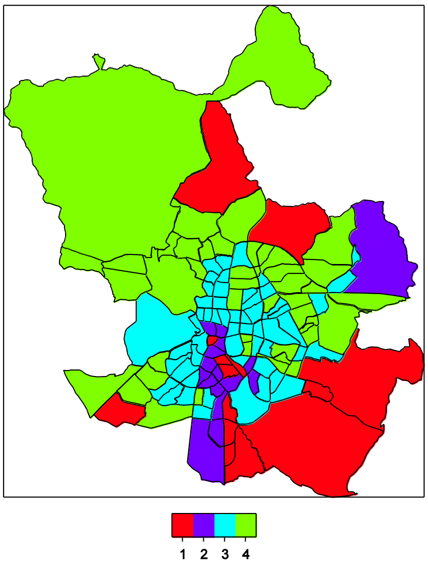


FIGURE 1b: Classification of Madrid neighborhoods according to residents' perception of pollution



Note: 1. Severe problems of pollution; 2. Medium-high problems of pollution; 3. Low levels of pollution; 4. No pollution problems.

Objective clustering classifies most of the neighborhoods next to the main ring road of the city (M30), a very busy road, in the second best group, the reason once again being the influence of the low pollutant values recorded by the neighborhoods outside the ring road (with the exception of the Southern part). However, when residents' perceptions are taken into account, although the majority of these neighborhoods are also included in the second best cluster (the light blue one), the number of them in the purple cluster is not negligible, as could be expected.

As for downtown Madrid, very clear differences are found between "objective" and "subjective" clusters. The best neighborhoods in Madrid City show the highest objective level of pollution, but most of them are perceived as the second best air pollution group of the city. The reason is that residents tend to identify good air quality with living near a large green area. This is precisely the case of the neighborhoods close to the Retiro Park and Casa de Campo, the two largest green areas in Madrid. However, traffic density in these neighborhoods is among the highest in Madrid. Perceived pollution is also lower than interpolated (based on actual values) pollution in the pedestrian areas of the city (the city center).

It can be also noticed that in the "subjective" cluster, some neighborhoods are incrusted in apparently uniform clusters. This situation cannot occur in the "objective" cluster, because the kriging estimator is based on a weighted mean. This circumstance can be clearly appreciated (Figure 1b) in the area next to both the main bus and train station of the city (red neighborhoods –with a high density of traffic and highly floating population– incrusted in an area of purple neighborhoods).

In summary, we can conclude that perceived and objective pollution in Madrid are two different concepts, and since perception is the concept that consumers include in their utility function, perceived pollution could be the appropriate measure to take into account when it comes to estimating the impact of air quality on house prices. Obviously, this finding leads us to encourage decision makers to go beyond environmental initiatives and raise public awareness.

3. SPECIFICATION AND ESTIMATION OF SPATIAL HEDONIC MODELS THAT INCLUDE INTERPOLATIVE VARIABLES

Hedonic models are the usual strategy to estimate the impact of pollution on housing prices. This specification corresponds to the equation:

$$y_i = \alpha + \lambda Pol_i + \mathbf{z}_i^T \boldsymbol{\delta} + \varepsilon_i \qquad \text{i=1,...,n} \tag{3}$$

where y_i represents the log of the price of the i-th dwelling, Pol_i indicates the level of pollution at the location where the i-th dwelling is sited, $\mathbf{z}_i^T = (z_{1i}, z_{2i}, ..., z_{ki})^T$ includes the k individual and areal characteristics of the i-th dwelling, α is the intercept of the equation and ε_i is a random disturbance that is assumed to distribute as $N(0, \sigma_\varepsilon^2)$. The impact of pollution on the housing price, or semi elasticity, is given by $\frac{\partial y_i}{\partial Pol_i} = \lambda$.

As is well known, under the assumptions of homoscedasticity, non autocorrelation and multivariate normal distribution of the vector of random disturbances, the OLS estimation method provides both BLUE estimates of the model parameters and the estimated variance of such parameters. However, when pollution is estimated by a kriged indicator (in our case an API) OLS is not consistent due to the need to interpolate the measured values of pollution at the sites where sampled dwellings are located, causing a potential errors-in-variables problem. The potential "errors in variables" aspect of interpolated air pollution measures (see Anselin and Lozano-Gracia 2008) has been ignored —barring a few exceptions— by the literature. But in our opinion, it is a core aspect when dealing with interpolated environmental measures as regressors. The reason is that the use of spatially interpolated values of pollution results in a prediction error which may be correlated to the overall model disturbance term. And this could lead to simultaneity bias in an OLS regression. In this case, the API would be an endogenous regressor (Anselin and Lozano-Gracia, 2008; Minguez *et al.*, 2010) and the OLS estimation of λ would be non consistent. We therefore used the instrumental variables method to solve this problem. In our case, following Anselin and Lozano-Gracia (2008), we took longitude and latitude coordinates as instruments and subsequently the Two-Stage Least Squares (TSLS) method has been applied for estimating purposes. As a result, we obtain consistent estimates.

However, model (3) does not take into account the spatial argument, that is to say, the existing spatial dependencies among the prices of dwellings. As has been shown in the literature (Anselin, 1988), the omission of spatial effects can result in estimators being inefficient and, what is worse, inconsistent, regardless of the estimation method. In order to capture the existing spatial dependencies in the prices of dwellings, following LeSage and Pace (2009), the specification we propose is the spatial Durbin model (SDM). We chose this model because it is general and robust, since it is the most general spatial model. In fact, the usual spatial specifications, —spatial autoregressive models (SAR) and spatial error models (SEM) — are particular cases of SDM. In addition, SDM provides consistent estimates for the majority of spatially correlated data generating processes.

The SDM is given by the following matrix equation:

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \alpha\mathbf{i}_n + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma_\varepsilon^2\mathbf{I}_n) \tag{4}$$

where $\mathbf{y}$ is a (nx1) vector including the observations of the logarithms of the house prices, $\mathbf{X}$ is a (nxk) matrix comprising the observations of the individual and areal characteristics associated to each dwelling and other spatial variables such as the *API*, surface, condition, mean mortgage in the neighborhood, etc., $\mathbf{i}_n$ is a (nx1) unit vector for the intercept (removed from $\mathbf{X}$ to avoid problems of exact multicollinearity in the estimation) and $\mathbf{W}$ is the (nxn) spatial weights matrix. Obviously, $\mathbf{Wy}$ and $\mathbf{WX}$ capture the spatial lags corresponding to the dependent and the variables included in $\mathbf{X}$, respectively. On the other hand, ρ is a spatial parameter that measures the existing spatial dependence of the dependent variable, α is the intercept parameter, σ^2 is the variance of the noise under homoskedasticity and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are (kx1) vectors of parameters associated to the independent variables and their lags, respectively. While in the SDM model we impose the restrictions $\rho = 0, \boldsymbol{\gamma} = \mathbf{0}$, the non spatial hedonic model (3) is obtained as a particular case.

As is well known, the specifications that include the spatial lag of the endogenous variable, $\mathbf{Wy}$, as a regressor, produce an endogeneity bias, because the spatial lagged variable is correlated to $\boldsymbol{\varepsilon}$. However, under the assumption of multivariate normal distribution of the disturbances, the estimation of the parameters of the model, $\boldsymbol{\theta} = (\rho, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma_\varepsilon^2)^T$, can be carried out by the ML procedure. For this purpose, following LeSage and Pace (2009), we first re-write (4) as:

$$\mathbf{y} = (\mathbf{I}_n - \rho\mathbf{W})^{-1}[\alpha\mathbf{i}_n + \mathbf{X}\boldsymbol{\beta} + \mathbf{WX}\boldsymbol{\gamma}] + (\mathbf{I}_n - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma_\varepsilon^2\mathbf{I}_n) \qquad (5)$$

But, unfortunately, if $\mathbf{X}$ includes an endogenous regressor (this could be the case of an API) ML estimates are non consistent. This circumstance led us to use the TSLS method for estimation purposes because the estimates it produces are consistent (Anselin and Lozano-Gracia, 2008). In this case, instruments are not only needed for the API but also for $\mathbf{Wy}$. It is common practice in the literature (Kelejian and Robinson, 1993; Kelejian and Prucha, 1998; Anselin, 2007) to take the successive powers of $\mathbf{WX}: \{\mathbf{WX}, \mathbf{W}^2\mathbf{X}, \mathbf{W}^3\mathbf{X}, \ldots, \mathbf{W}^p\mathbf{X}\}$ as instruments for $\mathbf{Wy}$, excluding the API indicator from $\mathbf{X}$ to avoid the endogeneity bias.

It is important to note that spatial spillovers (effects of changes in independent variables on the dependent variable) are not given by any vector of parameters directly in SDM. This is why, once again following LeSage and Page (2009), we express equation (5) as follows:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \sum_{r=1}^{k+1} \begin{pmatrix} S_r(\mathbf{W})_{11} & S_r(\mathbf{W})_{12} & \cdots & S_r(\mathbf{W})_{1n} \\ S_r(\mathbf{W})_{21} & S_r(\mathbf{W})_{22} & \cdots & S_r(\mathbf{W})_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ S_r(\mathbf{W})_{n1} & S_r(\mathbf{W})_{n2} & \cdots & S_r(\mathbf{W})_{nn} \end{pmatrix} \begin{pmatrix} x_{1r} \\ x_{2r} \\ \vdots \\ x_{nr} \end{pmatrix} + \mathbf{V}(\mathbf{W})\mathbf{i}_n\alpha + \mathbf{V}(\mathbf{W})\boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_n) \tag{6}$$

where

$$\mathbf{V}(\mathbf{W}) = (\mathbf{I}_n - \rho\mathbf{W})^{-1}$$
$$\mathbf{S}_r(\mathbf{W}) = \mathbf{V}(\mathbf{W})(\mathbf{I}_n\boldsymbol{\beta}_r + \mathbf{W}\boldsymbol{\gamma}_r)$$

Now, we can compute both the direct and indirect effects, respectively, of a change in x_{ir} and x_{jr} observations on y_i as:

$$\frac{\partial y_i}{\partial x_{ir}} = S_r(\mathbf{W})_{ii} \quad \text{and} \quad \frac{\partial y_i}{\partial x_{jr}} = S_r(\mathbf{W})_{ij}$$

Both impacts are nonlinear functions of the estimated parameters and, in addition, depend on the parameters associated to the regressor X_r as well as on ρ .

Since, in general, the magnitude of the impact of a variable X_r differs across regions, Pace and LeSage (2006) define the Average Direct Impact (ADI), Average Total Impact (ATI) and Average Indirect Impact (AII) of regressor X_r as follows:

$$\begin{aligned} ADI &= n^{-1} trace\left(\mathbf{S}_r(\mathbf{W})\right) \\ ATI &= n^{-1}\mathbf{i}_n^T \left(\mathbf{S}_r(\mathbf{W})\right)\mathbf{i}_n \\ AII &= ADI - ATI \end{aligned} \tag{7}$$

Finally, one of the main advantages of SDM is that if we set some restrictions in this model, it is possible to obtain other well-known spatial models. Setting $\boldsymbol{\gamma} = \mathbf{0}$ leads to the SAR model, and setting $\boldsymbol{\gamma} = -\rho\boldsymbol{\beta}$ the SEM is obtained. Since the SDM framework nests those models, it is robust under different specifications. Another advantage is that once SDM, SAR and SEM have been estimated by ML, we can perform LR tests to select the appropriate specification.

In summary, for comparative purposes, we will estimate the hedonic house prices model using OLS, TSLS and ML when it includes the objective interpolated indicator API, whereas only OLS and ML will be used if pollution is indicated by subjective resident perceptions because in this case there is no reason for suspecting the existence of a potential errors-in-variables problem.

4. CASE STUDY: INCORPORATING A SUBJECTIVE MEASURE OF AIR QUALITY INTO A SPATIAL HEDONIC HOUSING PRICE MODEL FOR MADRID

4.1. Housing market and pollution

Madrid (the capital of Spain) is the third-most populous city in the European Union (pop. 6,271,638 in 2009, 3,213,271 of which live in the city). Like other capitals in the world, Madrid is the city where Government institutions, the Parliament, embassies, main museums, central offices of the most relevant companies, etc. are located. This has made Madrid a large city covering 60,430.76 ha, together with a large peripheral metropolitan area with more than five million inhabitants that it is closely related to. Obviously, these relations imply movement and a large number of trips and regular flows of both population and also goods, etc., which has lead to a complex transportation system.

More specifically, Madrid has both a dense ring road network (M-30, M-40, M-45, M-50) and a dense radial highway network. Both networks have enormously improved accessibility to emerging industrial and high economic activity areas, resulting in competitiveness and dynamism. However, as a negative consequence of the above positive factors, road traffic has become the main source of atmospheric pollution.

In addition, Madrid has the fourth largest European airport and is the centre for train communications (half a thousand trains enter Madrid from the 10 most important Spanish cities, as well as from Paris and Lisbon). Freight transportation by train is also really important in Madrid. Every day 400 trains enter and leave the city, transporting 150,000 tons of commodities. In fact, Madrid has the largest inland maritime customs centre in Europe.

It is therefore no surprise that the number of vehicles in Madrid has increased by 5.6% over the last decade, amounting in 2010 to a total of 1,917,382. This implies

1,202.5 vehicles per km. and 683.5 vehicles per 1,000 inhabitants. Two million drivers enter and leave the city on a daily basis.

So, car pressure is increasing as well as its negative environmental impacts. Nevertheless, air pollution in Madrid can also be attributed to other factors, such as manufacturing and heating systems during winter, among others.

As a result of the economic development of Madrid City and the increase in population, construction, especially residential construction, has become an extremely important industry for the economy of Madrid as a whole. According to the Spanish Regional Accounts, 2009, this sector contributes 8.6% of total GDP. Madrid is the city with the largest housing stock in Spain —11.5% of the total, with a percentage of home ownership of 78.7% (2,275,188 out of 2,890,229) — and is also the main housing market: in 2009 some 53,513 housing transactions were made in Madrid (Spanish Housing Office). The highest housing prices of the country are also registered in Madrid.

The most central districts of the city are well established areas where few new houses are built and large projects are uncommon due to the lack of available land. Supply clearly exceeds demand, with the market for rent growing significantly. In addition, there is only a token presence of State-subsidized housing. Current supply focuses mainly on second-hand homes, the price of which, due to the characteristics of the area itself, has remained unchanged due to their quality and advantageous location. Prices will however tend to decrease as these two arguments lose strength.

It is worth highlighting the districts of: a) Salamanca (prices ranging from 3,750 to 11,549 euro/m2) which currently has a highly variable number of empty dwellings, most of which belong to the second-hand home market. The majority of sales made have affected dwellings in the lower echelon of prices, although the high level of purchasing power required to buy a house in the area does not significantly affect prices, which can rise in some cases; b) Chamartín (prices ranging from 4,225 to 11,183 euro/m^2) has a very small housing stock, generally open blocks with few individual houses, with communal gardens and swimming pools. New house development is minimal due to a lack of building sites, the majority of dwellings being second-hand, some 30 years old and medium to high quality; c) Hortaleza (prices ranging from 2,667 to 6,458 euro/m2), where there are mainly multi-family dwellings, although semi-detached and luxury stand-alone houses are abundant in some areas. These districts are practically new and have grown markedly in recent years. In the vicinity of the M-40 urban freeway, luxury houses are currently being built in closed housing developments with large communal areas.

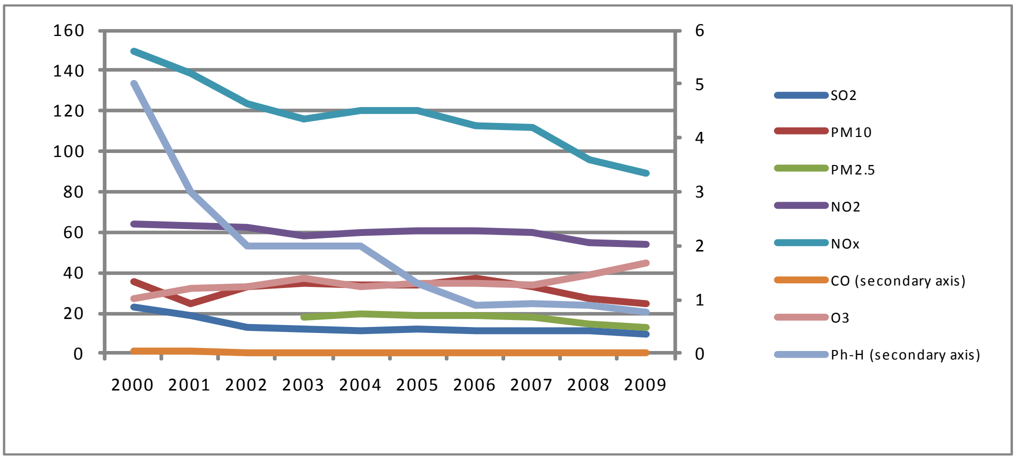
The districts on the periphery are characterized by the balance between public and private housing development. At present, new house prices are decreasing, while this downward trend is much more pronounced in the case of second-hand homes, with very few transactions being made. Building activity has stagnated, focusing on a very small number of developments to replace houses and with little sign of the market picking up. There is a large supply of housing, substantially more than demand, particularly in the second-hand home market, where many home-owners have put their dwellings up for sale due to being unable to meet their mortgage payments.

As for pollution, Madrid City Council started to monitor and control air quality in 1968, initially with a manual system and then, from 1978, with an automatic system of fixed stations. Over the last 40 years, the improvement in industrial processes, combustion systems and the quality of fuels and technological progress in general, together with changes in habits and increased mobility, have given rise to changes in the consideration of the pollutants of greatest concern and, therefore, in the situations to be resolved. Clearly these innovations have brought about significant improvements in air quality with respect to some parameters, but have also given rise to new pollution problems and new aspects to focus on when assessing air quality.

In the last decade, there has been a considerable improvement in sulphur dioxide, carbon monoxide and lead, pollution levels being much lower than those required by law. Nevertheless, as with most large European cities, problems persist with nitrogen dioxide, suspended particles and, especially, tropospheric ozone (see Figure 2).

Following EU directives, in this article we have considered the following six pollutants: SO_2, NO_x —which is a generic term for mono-nitrogen oxides (NO and NO_2)—, CO, PM_{10} and O_3. According to the Department of Environmental Assessment, Control and Quality of Madrid, the mean values of those pollutants in the city in 2009 were as follows (Figure 2): SO_2 (10 micrograms/m^3), suspended particulate matter (25 micrograms/dry standard m^3), CO (0.4 milligrams/m^3), NO_X (89 micrograms/m^3), NO_2 (54 micrograms/m^3) and O_3 (45 micrograms/m^3).

FIGURE 2: Mean values of SO_2, NO_x, NO_2, CO, PM_{10} and O_3



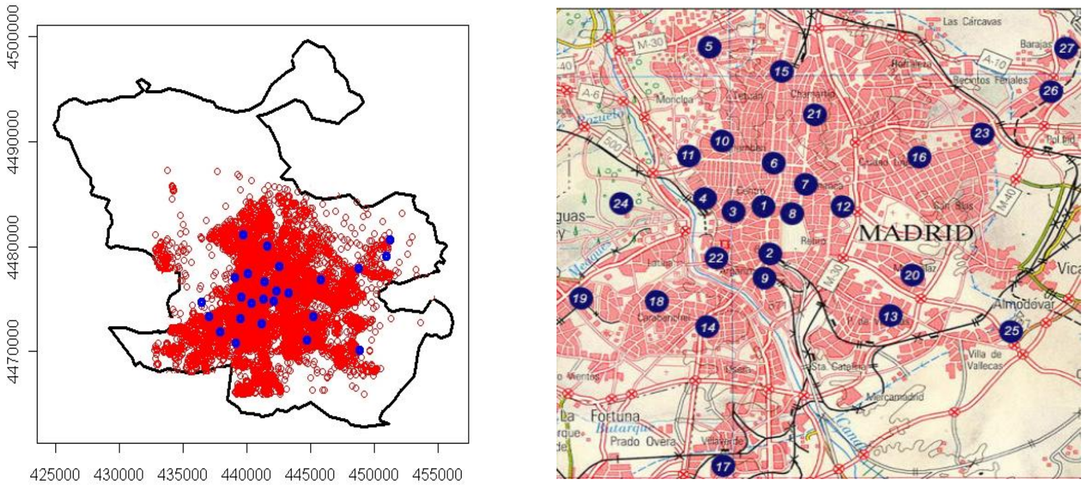
4.2. Data sets

The issue of housing prices remains unresolved in Spain. This is the reason why we have constructed our own database for Madrid. The final database we have created contains information about the price and 33 characteristics of 11,796 owner-occupied single family homes. Figure 3a shows the location of the observed dwellings –along with the location of the monitoring stations– and Figure 3b the location of such stations in the city. Most of them are located in the urban centre and relatively few in peripheral areas. Therefore, the monitoring stations cover the area under study reasonably well, since most of the population of Madrid is concentrated in the urban centre. The database was created from the sales that took place in Madrid in the last quarter of 2009. As far as we know, it is the largest database ever used to analyze the Madrid housing market. It is important to note that the sample accounts for 90% of the sales in that quarter. The list of variables we have used mirrors the usual set used in the literature (see Table A1). Most of them have been codified as categorical to allow for more flexibility in the specification of the model. This allows for nonlinearities between the different levels of each variable.

FIGURE 3: Location of observed houses and air quality monitoring stations

a) Location of houses and monitoring stations

b) Urban map of Madrid including the monitoring stations location



As for the data relative to pollution, they were provided by the Atmosphere Pollution Monitoring System of Madrid. As said above, we deal with six pollutants: SO_2, NO_x, NO_2, CO, PM_{10} and O_3. Note that in the specialized literature, hedonic specifications typically include only O_3 (Banzhaf 2005; Hendrix *et al*., 2005; Anselin and Le Gallo 2006, among others), PM_{10} (Chay and Greenstone 2005; Murthy *et al.*, 2009), or both O_3 and PM_{10} (Anselin and Lozano-Gracia, 2008 is a recent example) since these are the most visible in the form of "smog" and are thought to have the greatest health impact. But a workable approach to environmental data should consider multiple contaminants. Obviously, including six variables in a spatial hedonic house price model is not an easy task, so we decided to incorporate an API that gathers the information contained in such variables.

Pollution measurements were taken in February 2009 at 10a.m, and we have used the monthly average. There were two reasons for this decision: (i) February is the month of the year that records the most pollution; (ii) following the "population affected criterion", 10 a.m. is a critical time.

Another possibility (that used by Anselin and Lozano-Gracia, 2008) is to average the daily maxima during the worst quarter of a particular year. We rejected this option because the spatial structure of dependencies is not the same every hour, and the averaging process could lead to compensate different structures. In any case, the question of when to measure pollution is a core aspect, because the mismatch between the location of monitoring stations and the sites where the houses have been transacted must be overcome by kriging, and the structure of the spatial dependencies of the level of a particular pollutant depends on the temporal argument and is very sensitive to temporal aggregation.

In order to match the housing and pollution databases, we have predicted (kriged) the *API* at the locations corresponding to the observed dwellings accordingly. This alternative reduces the MSE estimated with the usual procedure by more than 7%. In light of the empirical data, we found ordinary kriging to be the best strategy for interpolating the above mentioned values. We used the maximum likelihood method to select the valid semivariograms. These semivariograms and their corresponding estimated parameters are displayed in Table 4. We used the GeoR package (Ribeiro and Diggle 2001) for that purpose as well as carrying out the cross validation procedure and obtaining kriged estimates. Figure 4b reports the interpolated values of the *API* at house sites and Figure 4a the quartile map of house prices. Results suggest a positive correlation between housing prices and the level of pollution.

TABLE 4: Valid Semivariograms for pollutants

Variable	Semivariogram model	Nugget effect	Partial sill	Range	ϕ_A (a)	ϕ_B (b)
CO	Spherical	0.152	0.824	4000.00	0.00	1.00
PM_{10}	Spherical	0.776	0.243	3999.97	0.00	11.32
O_3	Wave	1.040	1.000	1940.40	0.00	1.00
NO_x	Pure Nugget	1.000	0.000	3301.91	6.23	3345.9
NO_2	Pure Nugget	1.000	0.000	1184.26	0.00	1.00
SO_2	Exponential	0.084	1.093	3999.99	0.00	1.07

(a) value (in radians) for the anisotropy angle; (b) value for the anisotropy ratio (always greater than 1).

FIGURE 4a: Quartile map of house prices

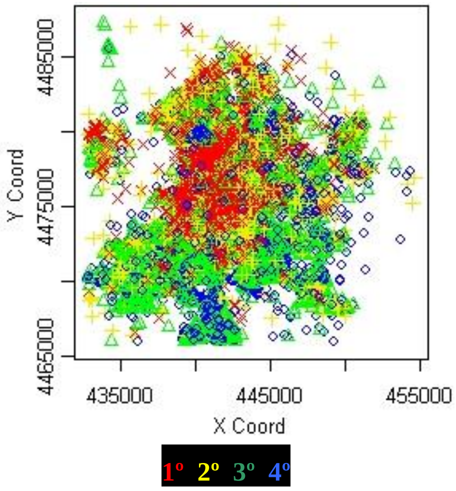
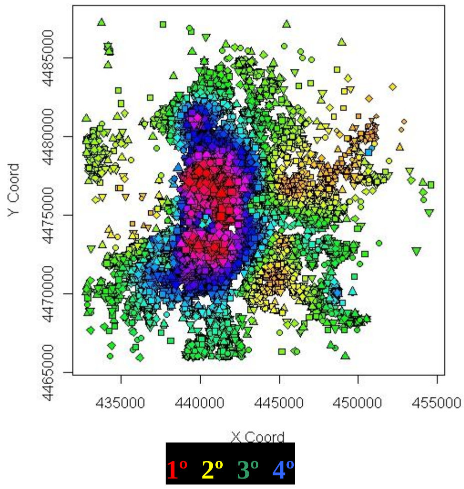
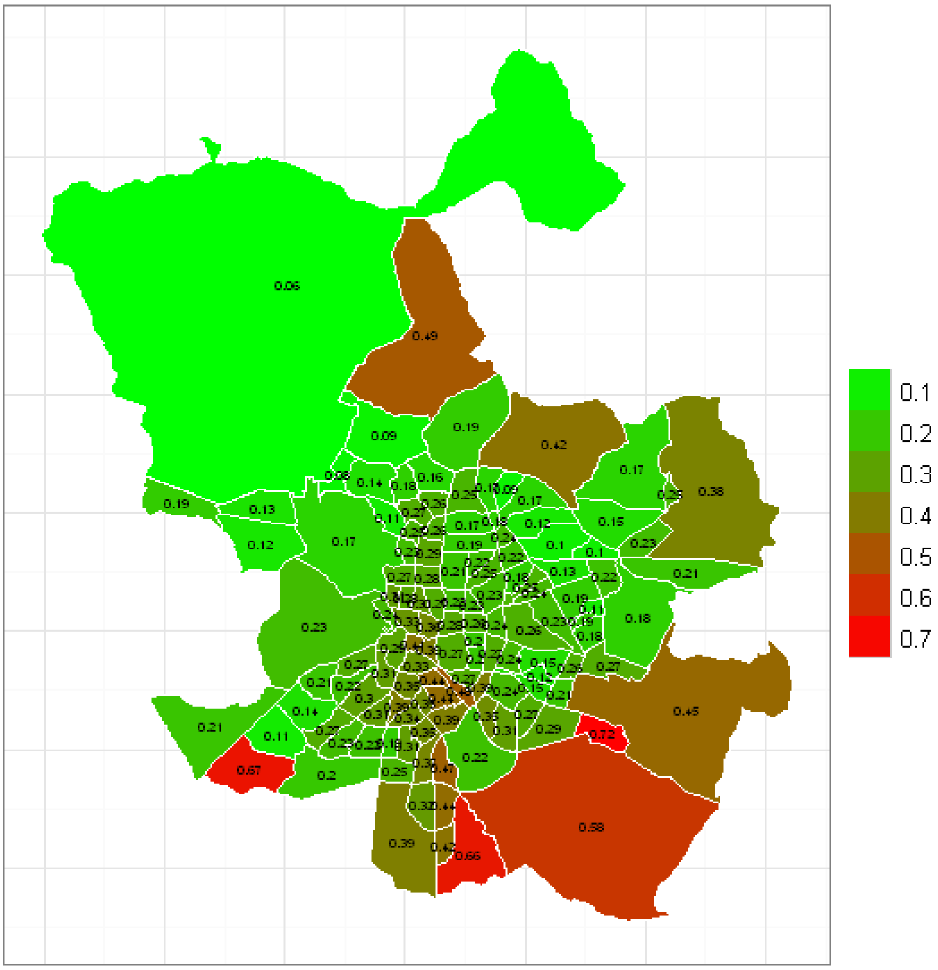


FIGURE 4b: Interpolated (kriged) API values at houses sites



The information relative to the pollution perceived by residents has been taken from the Spanish Population and Housing Census. As stated in section 2, we prefer census information to survey information and we subjectively characterize the areas by the percentage of residents that think that the area where they live has serious pollution problems. The advantages of this approach include exhaustiveness, non-dimensionality and common range [0-1], which enormously facilitates interpretation. In addition, this approach could be interpreted as a way of averaging the results obtained in surveys based on annoyance scales. It could also be viewed as a consensus measure of air quality. Figure 5 reports the percentage of dwellings with pollution problems in every neighborhood.

FIGURE 5: Percentage of dwellings with pollution problems in every neighborhood (Census data)



4.3 Results

The starting point in our analysis is the nonspatial hedonic house price model outlined in equation (3), where the dependent variable is the log price of dwellings and the explanatory variables are the individual and areal characteristics of such dwellings (see Table 1A). The first regressor we have included, apart from the intercept, is the variable indicative of the level of pollution –*API* when we deal with objective measures or *RP* when we are interested in subjective resident perceptions.

Model (3) has been estimated by OLS and TSLS when the pollution indicator is the interpolated *API* based on objective measures[2]. As said above, TSLS estimation

[2] All computations were been made using R (R Development Core Team 2010) and the spdep package (Bivand 2010).

responds to the potential errors-in-variables problem that arises from the use of *API* kriged values as a regressor. Table 5 reveals that the impact of *API* on the log price of houses in the OLS estimation is, surprisingly, positive and marginally significant at 0.05%. However, with TSLS estimation the impact is negative and clearly insignificant (0.05 significance level). Nevertheless, the influence of *API* on housing prices does not appear to be relevant.

TABLE 5: Summary of estimated impacts of pollution on dwelling prices

Pollution indicator: *API*						
Model	**Total Impact**	**t-value**	**Direct Impact**	**t-value**	**Indirect Impact**	**t-value**
OLS (non spatial model)	6.07E-03	2.347	6.07E-03	2.347	-	-
TSLS (non spatial model)	-1.82E-03	-0.2334	-1.82E-03	-0.2334	-	-
SDM (ML)	4.11E-03	1.10544	1.86E-02	0.2524	-1.45E-02	-0.195
SDM (TSLS)	-2.14E-02	-0.29405	-0.084626	-1.0737	6.32E-02	0.577
SAR (ML)	3.00E-03	0.8898	2.30E-03	0.8899	6.95E-04	0.887
SAR (TSLS)	3.65E-03	1.119	2.89E-03	1.1192	7.60E-04	1.114
SEM (ML)	7.04E-03	2.0184	7.04E-03	2.0184	-	-
Pollution indicator: *RP*						
Model	**Total Impact**	**t-value**	**Direct Impact**	**t-value**	**Indirect Impact**	**t-value**
OLS (non spatial model)	-1.17e-01	-4.817	-1.17e-01	-4.817	-	-
SDM (ML)	-1.44E-01	-3.7448	-6.67E-03	-0.4691	-1.37E-01	-3.561
SAR (ML)	-1.49E-01	-4.8263	-1.14E-01	-4.833	-3.46E-02	-4.6523
SEM (ML)	-1.13E-01	-3.5621	-1.13E-01	-3.5621	-	-

In order to ascertain which of the two estimates is the most suitable, we have performed a Hausman-Wu test, the null hypothesis of which is that the complete set of regressors, including *API*, are exogenous (Hausman, 1978). The test statistic is equal to 1.155, with an associated p-value of 0.28, which means the OLS estimation could be consistent. However, we must be cautious when interpreting the results of the test as it depends among other things on the choice of instruments used for API (longitude and latitude in our case).

On the other hand, when resident perceptions are taken as an (subjective) indicator of the level of pollution, not subject to the possible bias of endogeneity as it is not interpolated, the impact of pollution on the nonspatial model is clearly negative and significant, as expected. The rest of variables, both individual and suburb or district, display the expected signs and are statistically significant. The only exceptions –with a significance level of 10%–, although they display the expected signs, are the suburb crime rate and the retired person/population ratio measured on a district scale. As regards the crime rate, on the one hand the score can be explained by the rest of variables; in fact, the regression of the crime rate over the rest of regressors yields an R^2 of almost 60%. Furthermore, both variables can become significant if we include spatial specifications, as we will see later.

In order to test whether it would be more suitable to specify a spatial hedonic model instead of a nonspatial one, ML spatial dependence tests are performed on the residuals of all the foregoing nonspatial regressions (Anselin, 1998, 2007). The null hypothesis of these tests is the absence of spatial dependence in the residuals, as opposed to the alternative hypothesis of the SAR model (ML-lag tests) or SEM model (ML-Err tests). Table 6 shows that the null hypothesis is rejected for any given level of significance and more clearly rejected in the case of the ML-lag tests than with the ML-Err tests (a circumstance that is repeated in the robust versions of the tests compared to erroneous spatial specifications). All the tests are performed with a neighborhood matrix **W** that includes the 6 nearest neighbors (other neighborhood matrices **W** have been tested recorded similar results).

TABLE 6: Tests for spatial error dependence

OLS residuals with *API*		
LMerr = 468.5008	d.f. = 1	p-value < 2.2e-16
LMlag = 481.628	d.f. = 1	p-value < 2.2e-16
RLMerr = 54.812	d.f. = 1	p-value = 1.327e-13
RLMlag = 67.9392	d.f. = 1	p-value = 2.220e-16
TSLS residuals with *API*		
LMerr = 472.433	d.f. = 1	p-value < 2.2e-16
LMlag = 492.273	d.f. = 1	p-value < 2.2e-16
RLMerr = 52.573	d.f. = 1	p-value = 4.146e-13
RLMlag = 72.413	d.f. = 1	p-value = 2.220e-16

In the first place, these results suggest the need to include spatial dependence in the hedonic regression model. In the second place, it seems that, in regard to spatial specifications, the models with a spatial lag (SAR or SDM) fit the data better than SEM.

Due to being general and robust (LeSage and Page, 2009), the first spatial model we propose is SDM, described by equation (4). This model is estimated by ML and TSLS when the API is included as an (objective) indicator of pollution, and by ML when the variable *RP* is included as an (subjective) indicator of pollution. Table 5 reveals that, for SMD, the total impacts of the *API* on house prices are not significant in any cases and display alternating signs (positive in ML and negative in TSLS). In contrast, the total impact of the subjective indicator of pollution, estimated by ML, is clearly negative and significant, recording a similar absolute value to that of the nonspatial regression (-0.144 now compared to -0.118 in OLS). Furthermore, as the parameter ρ estimated is positive and significant with a value of 0.26, the absolute value of the indirect impact estimated is considerable (-0.137) and clearly significant. It is worth highlighting that the value of the parameter ρ in the previous case (when the pollution indicator is *RP*) is also similar to the spatial parameter of the SDM model with *API* estimated by ML. However, when estimated by TSLS, the value of the parameter rises to 0.98. As a result, we must be careful when interpreting this model as there is some evidence of overparameterization, which makes it difficult to estimate the spatial parameters and separate total impact into direct and indirect impact estimated by TSLS. In addition, as we saw previously with the Hausman-Wu test, there is no empirical evidence that TSLS is suitable in this scenario. In any case, as total impact is not significant in any cases for the *API* (either with TSLS or ML) and clearly significant and negative for *RP*, it seems obvious that the latter is the most appropriate variable to use as an indicator of pollution and, moreover, the spatial specification is a clear improvement on the nonspatial specification.

In order to ascertain whether the data allow a simpler spatial model than SDM to be specified, two LR tests are performed with null hypotheses from the SAR model (equation 8) and the SEM model (equation 9), respectively. In both cases, as detailed at the end of section 3, the alternative hypothesis is SDM, which is nested in both as particular cases. Table 7 shows the results of these tests for the subjective indicator of pollution and, in both cases, the null hypothesis is clearly rejected indicating that SDM is always preferred to alternative models. In addition, the AIC is lower in this model than in all the others and, likewise, it also minimizes the value of residual standard deviation. Nevertheless, it is worth highlighting that, as can be observed in Table 5, the total impacts of the subjective indicator of pollution are always negative, significant and similar in value (from -0.128 for SEM to -0.149 for SAR) and the spatial dependence parameter ρ is always positive, significant and similar in size (0.24 for SAR, 0.26 for SDM and 0.29 for SEM). Furthermore, these estimations of ρ are similar to those obtained for the SAR and SEM models including the *API* variable as the objective pollution variable but, as in the case of the nonspatial OLS regression and the SDM spatial regression, the total impact of the *API* is not significant either when estimated by ML or TSLS. All the results are available upon request.

TABLE 7: Model selection with subjective pollution indicator (*RP*)

Model	AIC	LR(H_1: SDM)	logL(·)	Parameters	σ	ρ or λ
OLS	-2947.8	-	-	35	0.2132	-
SDM	-3498.9	-	1820.438	71	0.20631	0.26168
SAR	-3346.4	220.4952	1710.19	37	0.20843	0.23906
SEM	-3319.9	246.9634	1696.956	37	0.20817	0.29539

In summary, it seems obvious that the measure of pollution people take into account in their utility function when assessing a house is subjective rather than objective pollution. In this sense, taking the value obtained in the SDM as the measure of total impact, an increase of ten percentage points in the number of residents that consider the neighborhood where the dwelling is located has pollution problems would reduce the price of houses in that neighborhood by 1.44%, which is economically significant and in line with the estimates obtained in Berezansky *et al*. (2010) for the city of Haifa.

5. CONCLUSIONS

The literature on hedonic housing prices has failed to measure the impact of pollution (or clean air) on the price of dwellings. Unlike economic theory, empirical research based on hedonic models reveals that the effect of air pollution on property value is far from conclusive. What is more, there are serious doubts that air pollution significantly affects property prices. Additionally, the type of study undertaken may also generate different results.

We have no doubts that the hedonic house price model is the appropriate instrument for estimating the effect of pollution on property prices, but we are convinced that the concept of air pollution that enters into the utility function of potential house buyers is perceived pollution rather than objective (measured) pollution. It is important to bear in mind that indirect methods like the hedonic strategy are based on actual transactions and empirical measurements, and assume that decision makers possess all the necessary information and always act rationally, attempting to maximize their personal utility. However, when deciding their location, house buyers weigh up one property or location against the other, their choices not necessarily being rational.

We are also convinced that spatial strategies are needed when dealing with the prices of dwellings because it may not be appropriate to assume that the implicit prices of housing attributes are stationary. But, even using spatial hedonic strategies, results regarding the impact of measured pollution on property prices are not conclusive. In the best of cases, clean air has a negligible influence on house prices, which contradicts the hedonic theory. Of course, this finding reinforces our hypothesis that objective pollution is not the appropriate indicator to measure such an influence.

In addition, interpolative methods for the estimation of pollution at observed house sites not only lead to an errors-in-variables problem, but also to very smooth predictions that could be far from realistic.

The above reasons led us to compare the spatial hedonic house price model including interpolated values of pollution based on values measured at monitoring stations –currently the usual model in the literature for the purpose we pursue– to the same model including resident perceptions of pollution (a subjective indicator).

On the basis of a massive database in Madrid –a large city with excellent monitoring site/population and monitoring site/surface ratios– we first classified Madrid neighborhoods according to their interpolated *API* (objective indicator) and *RP* (subjective indicator). Not surprisingly, the classification maps were found to be very different. More specifically, Pearson's correlation coefficient between objective and subjective values of pollution in Madrid neighborhoods stands at 0.20 and Spearman's ranks coefficient is 0.39.

The first finding we obtained is that the spatial argument is clearly relevant when it comes to estimating the impact of pollution on property prices. We found that the SDM was not only a general spatial hedonic strategy, but also the most appropriate for estimating the impact of pollution on property prices. We put both SDM to work, the one including *API* as a pollution indicator and the one including *RP* as a pollution variable, results confirmed our suspicions: While the model with *API* as a regressor does not confirm the hedonic theory, the one including the subjective indicator RP does. This is the second core finding in this paper. More specifically, according to the latter model, an increase of ten percentage points in the number of residents that consider that the neighborhood has pollution problems will reduce the price of the houses in the neighborhood by approximately 1.5%. This finding agrees with the results obtained in Berezansky *et al.* (2010), the only research we know using subjective indicators of pollution. Berezansky *et al.* (2010) found a 1% drop in housing prices due to a 1-scale worsening of perceived air pollution (on a 5-point scale), statistically significant at the 1% level.

Finally we encourage researchers to use hedonic spatial models including subjective indicators of pollution to confirm the results obtained for Madrid and Haifa. Notwithstanding, these results should be confirmed using other subjective indicators (based on annoyance surveys, for example). This could be a challenging avenue of research.

ACKNOWLEDGMENTS

This work has been partially funded by Junta de Comunidades de Castilla-La Mancha, under FEDER research project POII10-0250-6975. We would like to thank Felipe D. Carrillo (United States Fish and Wildlife Service) for his valuable help in mapping work.

REFERENCES

ANSELIN, L. (1988). *Spatial Econometrics: Methods and Models*. Boston, MA: Kluwer Academic Publishers.

ANSELIN, L. (2007). "Spatial Econometrics". In T. Mills and K. Patterson (eds.), *Palgrave Handbook of Econometrics*. Vol 1: Econometric Theory. UK: Palgrave-MacMillan.

ANSELIN, L. and BERA, A.K. (1998). "Spatial Dependence in Linear Regression Models with an Introduction to spatial Econometrics". In A. Ullah and D.A.E. Giles (eds.), *Handbook of Applied Economic Statistics*. New York: Marcel Dekker.

ANSELIN, L. and LE GALLO, J. (2006). "Interpolation of Air Quality Measures in Hedonic House Price Models: Spatial Aspects", *Spatial Economic Analysis*, 1(1), pp. 31-52

ANSELIN, L. and LOZANO-GRACIA, N. (2008). "Errors in variables and spatial effects in hedonic house price models of ambient air quality", *Empirical Economics*, 34, pp. 5-34.

BANZHAF, H.S. (2005). "Green Price Indices", *Journal of Environmental Economics and Management*, 49 (2), pp. 262-280.

BEREZANSKY, B.; PORTNOV, B.A. and BARZILAI, B. (2010). "Objective vs. Perceived Air Pollution as a Factor of Housing Pricing: A Case Study of the Greater Haifa Metropolitan Area", *Journal of Real Estate Literature*, 18 (1), pp. 99-124.

BIVAND, R. (2010). "spdep Spatial dependence: weighting schemes, statistics and models". R package version 0.5-24. Available in: http://CRAN.R-project.org/package=spdep

BOYLE, M.A. and KIEL, K.A. (2001). "Survey of House Price Hedonic Studies of the Impact of Environmental Externalities", *Journal of Real Estate Literature*, 9, pp. 117-144.

BRASINGTON, D.M. and HITE, D. (2005). "Demand for Environmental Quality: A Spatial Hedonic Analysis", Regional Science and Urban Economics, 35, pp. 57-82.

CAN, A. (1990). "The Measurement of Neighborhood Dynamics in Urban House Prices", *Economic Geography*, 66, pp. 254-272.

CAN, A. (1992). "Specification and Estimation of Hedonic Housing Price Models", *Regional Science and Urban Economics*, 22, pp. 453-474.

CARRUTHERS, J.I. and CLARK, D.E. (2010). "Valuing Environmental Quality: A Space-based Strategy", *Journal of Regional Science*, doi. 10.1111/j.1467-9787.2010.00662.x

CHAY, K.Y. and GREENSTONE, M. (2005). "Does air quality matter? Evidence from the housing market", *Journal of Political Economy*, 113(2), pp. 376-424.

CLARK, D.E. and ALLISON, T. (1999). "Spent Nuclear Fuel and Residential Property Values: The Influence of Proximity, Visual Cues and Public Information", *Papers in Regional Science*, 78, pp. 403-421.

CRESSIE, N. (1993). *Statistics for Spatial Data*. New York: Wiley.

FRALEY, C. and RAFTERY, E. (1999). "MCLUST: Software for model-based cluster analysis", *Journal of Classification*, 16, pp. 297-306.

FRALEY, C. and RAFTERY, E. (2002). "Model-based clustering, discriminant analysis and density estimation", *Journal of the American Statistical Association*, 97, pp. 611-631.

FRALEY, C. and RAFTERY, E. (2010). "MCLUST Version 3 for R: Normal Mixture Modeling and Model-Based Clustering", Technical Report, nº 504. Department of Statistics. University of Washington.

FREEMAN, III A.M. (1993). *The Measurement of Environmental and Resource Values. Theory and Methods*. Washington, DC: Resources for the Future.

HAUSMAN, J.A. (1978). "Specification Tests in Econometrics", *Econometrica*, 46, pp. 1251-1271.

HENDRIX, M.E.; HARTLEY, P.R. and OSHERSON, D. (2005). “Real Estate Values and Air Pollution: Measured Levels and Subjective Expectations”. Discussion Paper, Rice University.

JACKSON, T. (2001). “The Effects of Environmental Contamination on Real Estate: A Literature Review”, *Journal of Real Estate Literature*, 9(2), pp. 93-116.

JACQUEMIN, B.; SUNYER, J.; FORSBERG, B.; GÖTSCHI, T.; BAYER-OGLESBY, L.; ACKERMANN-LIEBRICH, U.; DE MARCO, R.; HEINRICH, J.; JARVIS, D.; TORÉN, K. and KÜNZLI, N. (2007). “Annoyance due to air pollution in Europe”, *International Journal of Epidemiology*, 36, pp. 809-820.

KELEJIAN, H.H. and PRUCHA, I. (1998). “A generalized spatial two stage least squares procedures for estimating a spatial autoregressive model with autoregressive disturbances”, *Journal of Real Estate Finance and Economics*, 17, pp. 99-121.

KELEJIAN, H.H. and ROBINSON, I. (1993). “A suggested method of estimation for spatial interdependent models with autocorrelated errors and an application to a country expenditure model”, *Papers in Regional Science*, 72, pp. 297-312.

LESAGE, J. and PAGE, R.K. (2009). *Introduction to Spatial Econometrics*. New York: Chapman & Hall.

MÍNGUEZ, R.; FERNÁNDEZ-AVILÉS, G. and MONTERO, J.M. (2010). “Does air pollution affect the price of housing? A Joint Geostatistics and Spatial Econometric Perspective”, In M- Mladenov, K. Psarris, N. Mastoraskis, A. Caballero & G. Vachtsevanos (eds.), *Development, Energy, Environment, Economics*, (pp. 333-337).

MONTERO, J.M. and FERNÁNDEZ-AVILÉS, G. (2009a). “Mixed environmental quality indexes for hedonic housing price models: An alternative with area-to-point kriging with external drift”, *AESTIMUM*, 54, pp. 73-94.

MONTERO, J.M. and FERNÁNDEZ-AVILÉS, G. (2009b). “How to include environmental quality kriged indexes in hedonic housing price models”, *The Polish Journal of Environmental Studies*, 18(5B), pp. 149-158.

MONTERO, J.M.; LARRAZ, B. and CHASCO, C. (2009). "Building a spatio-temporal Environmental Quality Index: the case of Madrid", *Statistica Applicata*, 21(2), pp. 133-151.

MONTERO, J.M.; LARRAZ, B. and CHASCO, C. (2010). "Building an environmental quality index for a big city: a spatial interpolation approach combined with a distance indicator", *Journal of Geographical Systems*, 12(4), pp. 435-459.

MURTHY, M.N.; GULATI, S.C. and BANERJEE, A. (2009). "Hedonic property prices and valuation of benefits from reducing urban air pollution in India". Delhi Discussion Papers, 61, Institute of Economic Growth: Delhi, India

MYERS, D.E. (1983). "Estimation of linear combinations and cokriging", *Mathematical Geology*, 15, pp. 633-637.

PACE, R.K. and LESAGE, J. (2006). "Interpreting Spatial Econometric Models", paper presented at the Regional Science Association International North American meetings, Toronto.

OSLAND, L. (2010). "An Application of Spatial Econometrics in Relation to Hedonic House Price Modeling", *Journal of Real Estate Research*, 32(3), pp. 289-320.

POOR, J.; BOYLE, K.J.; TAYLOR, L.O. and BOUCHARD, R. (2001). "Objective versus Subjective Measures of Environmental Quality in Hedonic Property Value Models", *Land Economics*, 77(4), pp. 482-493.

R Development Core Team (2010). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Available in http://www.R-project.org

RIBEIRO, P.J.Jr. and DIGGLE, P.J. (2001). "geoR: a package for geostatistical analysis", *R-NEWS*, 1(2), pp. 14-18. Available in: http://CRAN.R-project.org/doc/Rnews/

SIMONS, R. and SAGINOR, J. (2006). "A meta–analysis of the effect of environmental contamination and positive amenities on residential real estate", *Journal of Real Estate Research*, 28(1), pp. 71-104.

SMITH, V.K. and HUANG, J.C. (1995). "Can markets value air quality? A meta-analysis of hedonic property value models", *Journal of Political Economy*, 103, pp. 209-227.

STRASZHEIM, M. (1974). "Hedonic Estimation of Housing Market Prices: A Further Comment", *Review of Economics & Statistics*, 56, pp. 404-406.

THEEBE, M.A.J. (2004). "Planes, Trains, and Automobiles: The Impact of Traffic Noise on House Prices", *The Journal of Real Estate Finance and Economics*, 28, pp. 209-235.

APPENDIX

TABLE A1: List of variables

Variable name	**Description**
Dependent variable	
Price	House price
Variable of interest	
API	Objective Air Pollution Indicator
RP (resident perceptions)	Subjective Air Pollution Indicator
Coordinates	
Coordinate x	Longitude
Coordinate y	Latitude
House characteristics	
Good condition	Indicator variable for good condition
Flat	Indicator variable for flats
Studio-apartment	Indicator variable for studios
Top-floor flat	Indicator variable for top-floor flats
House	Indicator variable for houses
Age	Age of the house
Ground level	Indicator variable for ground level
Floor.1st	Indicator variable for 1st floor
Floor.2nd - 3rd	Indicator variable for 2nd and 3rd floors
Floor.4th - 5th	Indicator variable for 4th and 5th floors
Floor.6th or more	Indicator variable for 6th floor or higher
Baths	Number of bathrooms
Garage	Indicator variable for parking space
Elevator	Indicator variable for elevator
Air conditioning	Indicator variable for central air conditioning
Swimming pool	Indicator variable for swimming pool
Monthly mortgage	Monthly mortgage
Areal characteristics	
M.30	Indicator for houses inside the M-30
M.30.2	Indicator for houses close to the M-30
Shopping area	Indicator for houses in the shopping area
Historical quarter	Indicator for houses in the historical quarter
Built up area	Number of square meters of built-up area
Density pop. distr.	Population density in the district
Retired (% distr.)	Percentage of retired people in the district
Children (% distr.)	Percentage of children under 14 years
Immigrants (% distr.)	Percentage of immigrants in the district
Crime	Crime rate in the district
Trees (% Ha. distr.)	Trees per Ha. in the district
Mortgage reference area	Mean mortgage in the area