COMPARISON OF KRIGING, CO-KRIGING AND HEDONIC REGRESSION TO ESTIMATE HOUSING PRICE

Jorge Chica-Olmo
Universidad de Granada. Facultad de Ciencias Económicas y Empresariales. Campus Cartuja. Granada. Spain. jchica@ugr.es

Mario Chica-Olmo
Universidad de Granada. Facultad de Ciencias. Campus de Fuente Nueva. Granada. Spain. mchica@ugr.es

Rafael Cano Cuervos
Universidad de Granada. Facultad de Ciencias Económicas y Empresariales. Campus Cartuja. Granada. Spain. rcano@ugr.es

ABSTRACT: Kriging and co-kriging are geostatistical methods to estimate spatial correlated variables. This method allows spatial estimations to be made and interpolated maps or continuous maps of house price to be created. These maps are interesting for appraisers, real-estate companies and bureaus because they provide an overview of housing prices. Kriging is a univariate method and co-kriging multivariate. Kriging uses one variable of interest (house price) to make estimates at unsampled locations, and co-kriging uses the variable of interest and auxiliary correlated variables. A particular case of co-kriging is when auxiliary variables are observed in a housing set that are higher than the houses for sale (heterotopic data). This work compares the results obtained with the methods kriging, isotopic data co-kriging and heterotopic data co-kriging methods. Also, these results are compared with the hedonic regression. The study area is the city of Granada, Spain. Cross-validation is used to compare the methods.

Key words: kriging, co-kriging, housing price, geostatistics, variogram, hedonic regression.

1. INTRODUCTION

In the recent literature there are two perspectives that consider the spatial autocorrelation of housing prices: econometrics and geostatistics. The regression method is the most widely used to obtain econometric models, while kriging is used most in geostatistics. The regression method is a multivariate method and kriging univariate. It is well known that there are many works in which the regression model has been applied to real-estate appraisal. However, works that apply only kriging itself are few and far between (Chica-Olmo 1992). Within spatial econometrics the combination of the methods of econometrics and geostatistics (we could call this geo-econometrics) allow an examination of the causality among the variables, from the econometrics perspective, and of the spatial dependence, from the geostatistics viewpoint. Some authors, such as Dubin (1992), Chica-Olmo (1995), and Basu & Thibodeau (1998) have obtained good results from combining the regression methods and kriging. These works use the regression method to estimate the parameters of the structure of housing characteristics and the kriging method for modeling the spatial autocorrelation of disturbances. Dubin (1992) estimates hedonic regression and spatial autocorrelation parameters using the maximum likelihood method, Chica-Olmo (1995) uses iterative residual kriging with estimated generalized least-squares (EGLS) and Basu and Thibodeau (1998) use an iterative method and EGLS.

The method employed to analyze house price is, traditionally, hedonic regression. The explanatory variables can be: structural characteristics, neighborhood characteristics and accessibility (Can 1990). Structural characteristics are individual characteristics of the house itself (age, size, bathrooms, etc) and may or may not depend on the location. Neighborhood characteristics and accessibility depend on the location. The spatial autocorrelation or spatial dependence of house price is caused by the characteristics that depend on the location. Usually, housing sale price will be directly related to the sale price of other neighboring houses. Location is probably the most important variable to explain house price. The kriging and co-kriging methods are characterized by the fact that they use the spatial structure of correlation to explain the housing price. These geostatistical methods are used to create interpolated maps or continuous maps. It is important for the appraisal companies, bureaus, investment banking and administration to speed up mass appraisal and to draw up continuous maps of housing price. These maps reflect patterns in the spatial distribution of housing price within a city.

If housing price is spatially correlated, we will be able to apply kriging to estimate at unsampled locations from data on said pricing. If, in addition to this, we have auxiliary variables that are spatially correlated, along with the variable of interest, then co-kriging will increase estimation accuracy. The co-kriging method estimates the value of the variable of interest at an unsampled location from data on said variable and from auxiliary variables in the neighborhood. The spatial correlation is described by a variogram. This variogram expresses the spatial dependence between housing prices at different distances. The cross-dependence between two variables is described by the cross-variogram.

Co-kriging uses data defined on different characteristics. The house prices may be estimated from a combination of house prices and structural characteristics. The co-kriging procedure is an extension of kriging when multivariate data are available (Wackernagel, H. 1995). Co-kriging considers the simple and crossed spatial correlation of the housing price and of the auxiliary variables.

In hedonic regression the number of variables that influence house prices is both large and heterogeneous. Furthermore, identifying all relevant neighborhood characteristics within the city is difficult (Tse 2002). The correlation between the explanatory variables causes serious multicolinearity problems in the regression model. Co-kriging is used when two or more explanatory variables are correlated and spatially intercorrelated.

It is known that it is necessary to have the value of the explanatory variables of the house to be appraised, in order to carry out predictions with the regression method; but this is not necessary when kriging or co-kriging are applied directly to house prices. To predict a housing price, $y(s_0)$, at an unsampled location, s_0, using the least-squares-estimator of the regression model, it is necessary to know the correspondent value of that housing's explanatory variables $x_1(s_0),\ldots x_k(s_0)$. However, in order to apply co-kriging we do not need to know the value of the explanatory variables in location s_0, since they are also estimated using the other housing in the neighborhood. This provides co-kriging with the ability to carry out predictions of the price of the housing at any point on the map, taking into account price and the housing characteristics in its neighborhood. In this way, it is possible to obtain interpolated maps of the estimated price of the housing with similar characteristics to those of its neighborhood.

Another important characteristic of co-kriging in the field of real-estate evaluation is that it can be applied when the house price and the explanatory variables have not been sampled in the same housing (heterotopic data). For example when we have two samples: a sold housing sample (for which we know its price and characteristics) and another not-for-sale housing sample, for which we know its characteristics but not its price.

 ISBN 978-84-8439-519-5

The regression model does not have this characteristic. In order to apply the regression model we need to know the price and the explanatory variables of all the housing samples (isotopic data).

The next section presents a brief summary of the kriging and co-kriging methods. Section 3 deals with the spatial correlation of the data and presents the results. The last section gives the conclusions.

2. KRIGING AND CO-KRIGING

Let us assume that the data $Z(s_1),\ldots,Z(s_n)$ are a particular realization of a stationary process that satisfies the model (Cressie 1991):

$$Z(s) = \mu + \delta(s)$$

where: μ is the unknown constant-mean and $\delta(s)$ is the spatially autocorrelated error, intrinsically stationary. In this case the ordinary kriging (co-kriging) method is applied to the data to make the spatial prediction.

2.1. Ordinary Kriging

Kriging is an estimation method that uses the spatial dependence of a second-order stationary variable.

The spatial dependence between Z and the separation vector distance (h) and the direction (θ) is expressed by a variogram (semivariogram or direct-variogram) $\gamma(h)$. An unbiased estimator of the variogram is (Matheron, 1965):

$$\hat{\gamma}(h_\theta) = \frac{1}{2N(h_\theta)} \sum_{i=1}^{N(h_\theta)} [Z(s_i + h_\theta) - Z(s_i)]^2$$

where $(s_i + h_\theta)$ and (s_i) are locations and $N(h_\theta)$ is the number of h_θ distant point-pairs.

The empiric variogram computed on different directions on the map are checked to find directional influences (anisotropy). It is necessary to adjust a model[1] to the empiric variogram to carry out estimations with the kriging method. We used the exponential model:

$$\gamma(h) = \begin{cases} C_0 + C\left[1 - \exp\left(-\frac{h}{a}\right)\right] & h > 0 \\ 0 & h = 0 \end{cases}$$

The model fitted depends on three parameters: nugget effect (C_0), range (a) and sill ($C_0 + C$), where (C) is a partial sill. The nugget effect is a measure of spatial continuity; range is the distance where the model levels out and sill is the value that the variogram model attains at the range. The exponential model (Figure 1) reaches the sill asymptotically, with the practical range (a) defined as that distance at which the variogram value is 95% of the sill (Isaaks and Srivastava 1989).

[1] See e.g. Cressie (1991) the models of variogram and methods used to carry out the fit.

Kriging is a univariate method that provides the best linear unbiased estimator (BLUE):

$$\hat{Z}(s_o)=\sum_{i=1}^{n}\lambda_i Z(s_i)$$

An unsampled house price $Z(s_o)$, located in s_o, is estimated with kriging from neighboring house prices, $Z(s_i)$, weighted with λ_i. In order that the kriging estimator be unbiased it has to be true that:

$$\sum_{i=1}^{n}\lambda_i=1$$

The weights are selected to minimize the variance of error:

$$\mathrm{Var}[\hat{Z}(s_o)-Z(s_o)]$$

by resolving the ordinary kriging system:

$$\begin{bmatrix}\Gamma & \mathbf{1}\\ \mathbf{1}' & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{\lambda}\\ \mu\end{bmatrix}=\begin{bmatrix}\boldsymbol{\gamma}_0\\ 1\end{bmatrix}$$

where Γ is a symmetric matrix formed by $\gamma(s_i-s_j)$ for i,j = 1...n; $\boldsymbol{\lambda}$ is the weights vector; **1** the vector of ones; $\boldsymbol{\gamma}_0$ is the vector formed by $\gamma(s_0-s_i)$ for i = 1...n and μ is the Lagrange multiplier. The estimation variance of the kriging is:

$$\sigma_k^2(s_0)=\sum_{i=1}^{n}\lambda_i\gamma(s_0-s_i)+\mu_1$$

2.2. Ordinary Co-kriging

When dealing with different variables, each variable is measured at the different points on the map. The location of points can be equal for all the variables (isotopy), equal for some variables (partial heterotopy) or different for all the variables (complete heterotopy). In partial heterotopy, co-kriging is interesting when the auxiliary variables are available at more points than the main variable (Wackernagel 1995).

The direct-variogram or auto-variogram measures spatial dependence for one variable. The cross-dependence between two variables is measured with the cross-variogram. The cross-variogram estimator is (Matheron 1970):

$$\hat{\gamma}_{jk}(h_\theta)=\frac{1}{2N(h_\theta)}\sum_{i=1}^{N(h_\theta)}[Z_j(s_i+h_\theta)-Z_j(s_i)][Z_k(s_i+h_\theta)-Z_k(s_i)]$$

where $N(h_\theta)$ represents the number of h distant point pairs, where variables Z_j and Z_k are measured. The cross-variogram can only be calculated when variables are measured in the same locations (isotopy and partial heterotopy). The cross-variogram can be negative, which indicates a negative correlation between Z_j and Z_k. (Journel, A.G. and Huijbregts Ch.J. 1978).

The objective is to predict the value of housing price $Z_1(s_o)$ (variable of interest), unsampled site s_o, from the variables $Z_j(s_i)$ (auxiliary variables) sampled sites s_1, s_2, ... s_n. Let us consider for simplicity[2] only two variables, Z_1 (housing price) and auxiliary variable Z_2 (housing age), with

[2] For a generalization see e.g. Wackernagel (1995)

 ISBN 978-84-8439-519-5

number of samples n_1 and n_2, respectively, not necessarily equal. The ordinary co-kriging estimator is a weighted average of observed values of the variables Z:

$$\hat{Z}_1(s_o) = \sum_{i=1}^{n_1} \lambda_{1i} Z_1(s_{1i}) + \sum_{j=1}^{n_2} \lambda_{2j} Z_2(s_{2j})$$

where n_1 and n_2 are the nearest houses to the housing s_o; λ_{1i} and λ_{2j} are the weights associated to each sampling point.

In order that the co-kriging estimator be unbiased two restrictions must be true:

$$\sum_{i=1}^{n_1} \lambda_{1i} = 1 \text{ and } \sum_{j=1}^{n_2} \lambda_{2j} = 0$$

The weights are determined to minimize the variance of error:

$$\mathrm{Var}[\hat{Z}_1(s_o) - Z_1(s_o)]$$

by resolving the co-kriging system:

$$\underbrace{\begin{bmatrix} \Gamma_{11} & \Gamma_{12} & \mathbf{1} & \mathbf{0} \\ \Gamma_{21} & \Gamma_{22} & \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} & 0 & 0 \\ \mathbf{0} & \mathbf{1} & 0 & 0 \end{bmatrix}}_{\Gamma} \underbrace{\begin{bmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \\ \mu_1 \\ \mu_2 \end{bmatrix}}_{\lambda} = \underbrace{\begin{bmatrix} \boldsymbol{\gamma}_{10} \\ \boldsymbol{\gamma}_{20} \\ 1 \\ 0 \end{bmatrix}}_{\gamma}$$

where Γ_{11} and Γ_{22} are direct-variogram matrixes formed by $\gamma_{11}(s_{1i} - s_{1j})$ for i,j = $1...n_1$ and $\gamma_{22}(s_{2i} - s_{2j})$ for i,j = $1...n_2$; Γ_{12} and Γ_{21} are the cross-variogram matrixes formed by $\gamma_{12}(s_{1i} - s_{2j})$ for i = $1...n_1$, j = $1...n_2$ and $\gamma_{21}(s_{2i} - s_{1j})$ for i = $1...n_2$, j = $1...n_1$; $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ are weights vectors; **1** vector of ones; $\boldsymbol{\gamma}_{10}$ and $\boldsymbol{\gamma}_{20}$ are vectors formed by $\gamma_{11}(s_{10} - s_{1j})$ for j = $1...n_1$ and $\gamma_{12}(s_{10} - s_{2j})$ for j = $1...n_2$ and μ_1 and μ_2 are Lagrange multipliers. The estimation variance of the co-kriging is:

$$\sigma^2_{ck}(s_0) = \sum_{i=1}^{n_1} \lambda_{1i}\gamma_{11}(s_{10} - s_{1i}) + \sum_{j=1}^{n_2} \lambda_{2j}\gamma_{12}(s_{10} - s_{2j}) + \mu_1$$

This variance is a measure of the uncertainty of the estimation of Z_1 in s_0.
Where the primary and secondary variables exist at all data locations (isotopic data) and the direct-variograms and cross-variograms are alike, co-kriging is similar to kriging (Isaaks and Srivastava 1989). When all variables are uncorrelated co-kriging is equivalent to kriging.

2.3. Drift and variable mean

Let us now assume that the data $Z(s_1),\dots,Z(s_n)$ are a particular realization of a non-stationary process which satisfies the model:

$$Z(s) = \mu(s) + \delta(s)$$

where: $\mu(s)$ is the deterministic mean structure, called the trend or drift and $\delta(s)$ is spatially autocorrelated error, intrinsically stationary. The drift represents a surface of the housing price with "a large-scale varying mean" and fluctuations on the surface is due to $\delta(s)$ "small-scale variation". In this case, there are different methods to estimate the drift: universal kriging, generalized covariance and residual kriging (Neuman, S. P. And Jacobson, E. A. 1984). Residual kriging or the detrend method consists of carrying out a polynomial least squares regression of the data to estimate $\mu(s)$, and to estimate $\delta(s)$ we apply simple kriging of resulting residuals.

When Z is stationary, $\mu(s) = \mu$ is constant-mean and unknown, the ordinary kriging (co-kriging) method is applied to the data. When we know μ then the version of kriging (co-kriging) is called simple kriging (co-kriging).

Unfortunately, in practice, it can be difficult to decide between the models (variable-mean and constant-mean) based on the data itself. The methods that require the drift need to estimate a bigger number of parameters. For example, when the drift is quadratic we need to estimate six parameters. In general, the parsimony principle will take us to select simple models. Furthermore, it is possible to make hypotheses of quasi-stationarity when we define sliding neighborhoods that hold the stationary hypothesis (Journel, A.G. and Huijbregts Ch.J. 1978). Kitanidis (1997) suggests adopting the variable-mean model only if it fits the data significantly better than the constant-model. Besides, if the drift is mild, as occurs in the following application, the variogram for short distances is often enough for the requisites of kriging (Chilès and Delfiner 1999).
Cross-validation allows us to select between different models or methods. Cross-validation removes, one at a time, each data location and predicts data value in this location. This procedure is repeated for all experimental points. In this way we compare the predicted value to the observed value. Using the simple regression between predicted and observed values and, subsequently, their scatter plot, we are given a measure to compare.

In the application described in section 3 we observed (Figure 2) that the price of housing, Z, can show second-grade drift. We applied the detrend method and ordinary kriging to estimate Z and below we compare the results using the cross-validation method. Figures 7a and 7b show the maps of estimated price using ordinary kriging and the detrend method, respectively. In general both figures are similar in form, except in the north and south zones, where Figure 7b has a more homogeneous aspect due to the quadratic drift's effect.

The simple regression between predicted and observed values and their scatter plot are included in Figure 3. This figure shows that ordinary kriging has a better fit than the detrend method because the coefficient of determination is larger and, in addition, the regression line is closer to the 1:1 line. For all these aforementioned reasons we have selected the ordinary kriging (co-kriging) method.

3. APPLICATION

The study area is the city of Granada, Spain. Granada is an ancient city located in the south of the country (Figure 1). Figure 1 shows the city divided into 14 districts. The Central Business District (CBD) is located in the district 10. District 9 is an historic quarter from where the city has grown out towards districts 8, 10 and 11. Districts 6 and 14 are working-class areas. The more conflictive districts, with the highest crime rate are in the northern part of the city, districts 1, 2 and 3 (the most conflictive being district 2). The application has been carried out on a sample of 287 apartments, Figure 2.a. The data comes from market research carried out by the Centro de Gestión Catastral (Official Cadastre Agency) of Granada and completed mainly during the fourth quarter[3] of 1995.

[3] Significant changes in the house prices were not observed in these months.

ISBN 978-84-8439-519-5

Id	District
1	Polígono de Almanjáyar
2	Barrio de la Paz
3	Polígono de Cartuja
4	Barrio de Cartuja
5	Colonia San Francisco
6	Chana
7	Camino de Ronda
8	Constitución-Doctores
9	Albayzín
10	Centro
11	San Matías-Realejo
12	San Antón-Ciudad Jardín
13	Cervantes-Vergeles
14	Zaidín

Id	Item
H1	University
H2	Cathedral
H3	Hospital
H4	University
H5	Town council

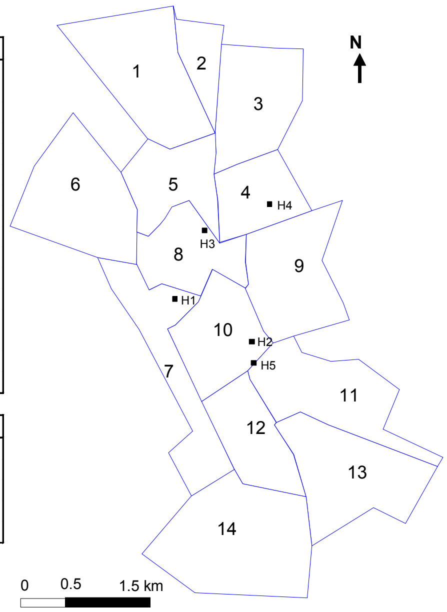

Figure 1. Study area, district and item.

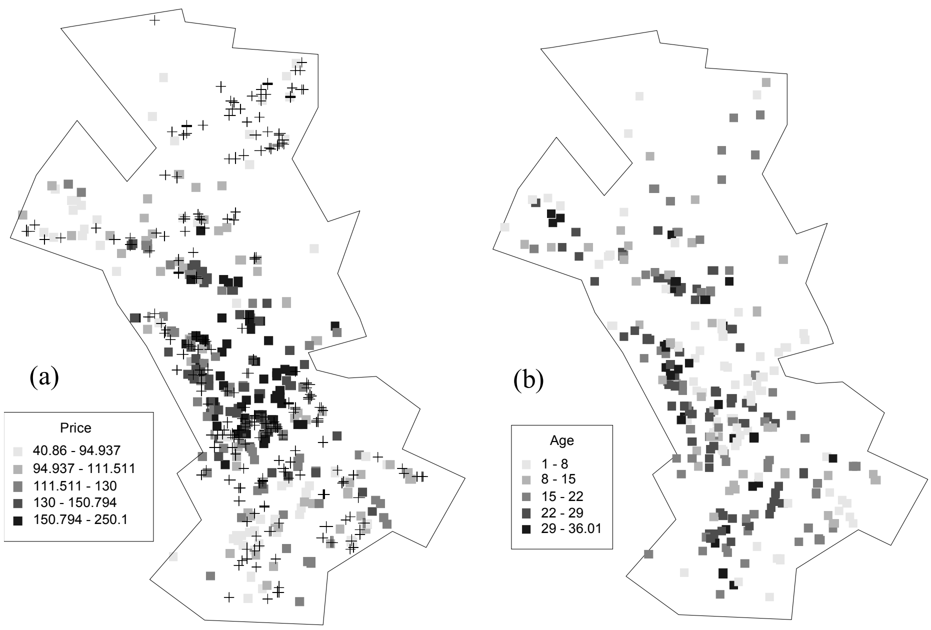

Figure. 2. Location of sample apartments. (a) represents with symbol square the price of the sample housing of 1995. The cross symbol represents sample of 1991. (b) represents the age of the sample housing of 1995.

The variable of interest PRICE represents the apartment price per square meter. The auxiliary variables are:

- AGE: apartment age in years.
- AMP: Amplitude, quotient between the number of constructed square meters of the apartment and the number of rooms.
- HEAT: Central Heating, binary variable that takes the value 1 if the apartment has central heating and 0 if not.

The Table 1 shows statistics of variables. The average price is 124.654 thousand pesetas (749.19 euros) and the mean age is 15.84 years. The Table 2 shows the simple correlation coefficients of the variables examined. All coefficients of linear correlation are significant with p-value equal to 0.01. Negative correlation between Age and other variables is as expected, as is the positive correlation between Price, Amplitude and Central Heating.

To apply kriging with heterotopic data, we have used another sample with 259 houses (the crosses in Figure 2.a represent their locations). For these houses we know the variables Age and Amplitude.

Table 1. Statistics of variables.

	PRICE	AGE	AMP	HEAT
Average	124.654	15.840	37.047	0.560
Standard deviation	33.597	9.986	8.935	0.497
Skewness	0.439	0.017	1.106	-0.261
Kurtosis	0.231	-1.109	1.349	-1.945
Minimum	40.860	1.000	23.750	–
25%tile	100.000	6.000	30.000	–
Median	122.951	16.000	35.750	–
75%tile	145.312	23.000	42.000	–
Maximum	250.000	36.000	73.000	–

Table 2. Correlation coefficients between variables.

	PRICE	AGE	AMP	HEAT
PRICE	1	-0.389	0.439	0.504
AGE	-0.389	1	-.370	-0.321
AMP	0.439	-0370	1	0.283
HEAT	0.504	-0.321	0.283	1

3.1. Analysis of spatial structure of variability. Variogram.

The variogram is the tool most commonly used in geostatistics to analyze the spatial continuity (or variability). The empirical direct-variograms and cross-variograms for the four variables are shown in Figure 3. Four direct-variograms and six cross-variograms are displayed. The calculated variograms are omnidirectional and they have been calculated for a maximum distance equal to 1000 meters (100 lag spacing and 10 number of lags). The examined direct-variograms of the variables show an increasing behavior, which suggests that variables are correlated spatially. The variogram of the price variable presents a more continuous behavior in the space than the rest of the variables. The variables Age and Amplitude behave very similarly.

The cross-variogram between Age and other variables is negative because these variables have a negative correlation. The city of Granada has an historic center with more old housing located downtown. In general, the distribution spatial of the Age variable would be expected to be convex (older housing downtown and more modern houses on the outskirts). However, in the housing for sale sample this does not occur because the houses located downtown have been

 ISBN 978-84-8439-519-5

rehabilitated, causing a contrary effect to that expected, Figure 2.b. Hence the cross-variogram among Age and other variables has a decreasing tendency as the distance becomes larger.

The rest of the cross-variogram shows a positive correlation. The rising trend suggests that variables are less correlated as the distance increases.

We have also examined the presence of anisotropy in the studied variables. A variable is anisotropic if the spatial correlation structure depends on the direction. All variables were assumed to be isotropic because we have not detected any significant anisotropy for a maximum distance of 1000 m. We have used exploratory tools to examine anisotropy: the variogram surface and the directional variograms (directions 0º, 45º, 90º and 135º). Variogram surface represents the variogram values on a map using a gray color scale. Figure 4 shows the variogram surface and directional variograms of the Price variable, other variables not shown. In these variograms we do not observe a specific behavior in any of the directions on the map.

We have used the exponential model to fit the empirical direct-variograms and cross-variograms. The method used to carry out the fitting of the direct-variograms and cross-variograms is by non-linear least squares (Cressie 1991). The values of the estimated parameters out of every direct-variogram and cross-variogram are shown in table 3. The column %Nugget of Table 3 shows spatial discontinuity (Nugget) out of every variable or variables in terms of the total variability (Sill). Thus, the variable with the main relative discontinuity is Age (53.3%) while the variable with the least discontinuity is price (20.235%). The estimated range[4] for spatial correlation is 992.47 meters.

Table 3. Parameters of model direct-variograms and cross-variograms.

Variable	Nugget	Partial Sill	%Nugget	Range
PRICE	170.300	671.280	20.235	992.470
AGE	58.827	51.935	53.300	992.470
AMP	43.175	40.743	51.400	992.470
HEAT	0.123	0.135	47.700	992.470
PRICE-AGE	-76.197	-75.117	50.400	992.470
PRICE-AMP	8.400	88.844	8.600	992.470
PRICE-HEAT	0.000	7.898	0.000	992.470
AGE-AMP	-13.940	-21.993	38.800	992.470
AGE-HEAT	-0.720	-0,801	47.300	992.470
AMP-HEAT	0.000	1.300	0.000	992.470

[4] To resolve the co-kriging system of equations the condition that the matrix Γ be positive-definite must be met. This condition comes true if the direct and cross-variograms satisfy the inequality of Cauchy-Schwarz:

$$\left|\gamma_{ij}(h)\right| \leq \sqrt{\gamma_{ii}(h)\gamma_{jj}(h)}$$

for two variables i and j.
For said inequality to be met, one criterion to be considered is that all the variograms have the same range.

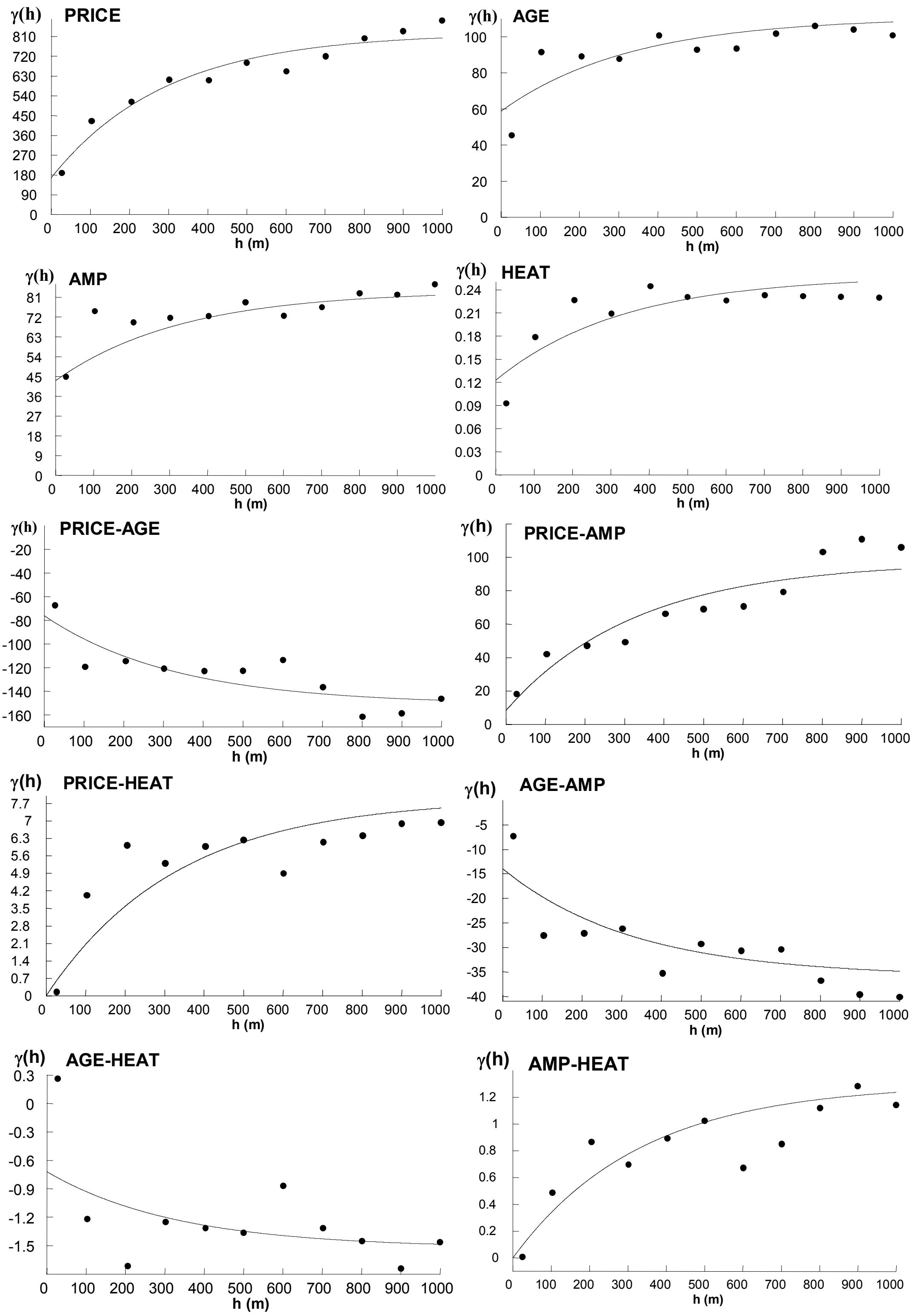

Figure 3. Variograms and cross-variograms of variables (empirical, dots, and model, solid line).

ISBN 978-84-8439-519-5

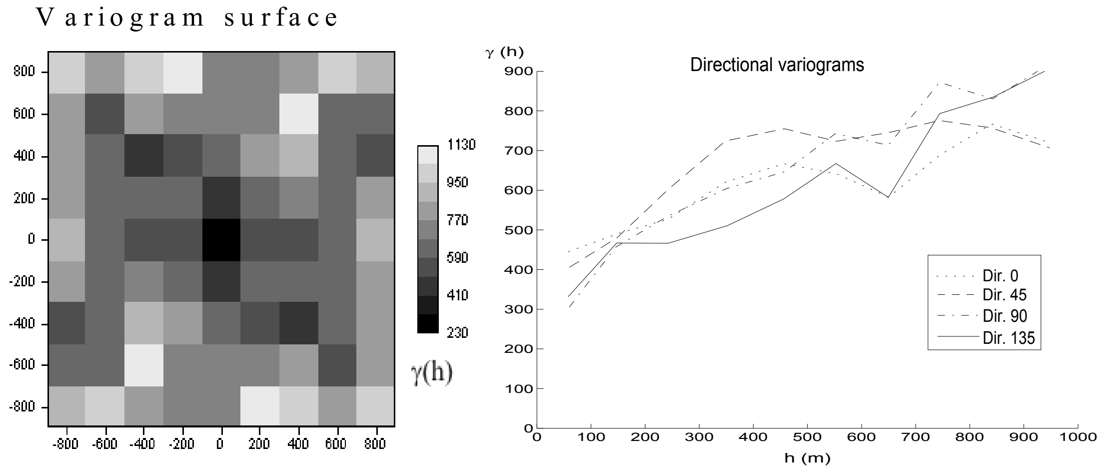

Figure 4. Variogram surface and directional variograms of price variable.

3.2. Spatial estimation. Kriging, Co-kriging and Hedonic Regression

Kriging (co-kriging) can be used to carry out estimations for any point or house on the map. Thus, in order to carry out estimations at a certain point the estimation system has to be geometrically defined, which means that we must establish the neighborhood of the data which will intervene in the estimation. In the estimation we can use all the houses - global neighborhood, or a set of data - local neighborhood. In practice, only the houses that are found inside a circumference or ellipse centered on the point to estimate are used. In this application a minimum of three houses and a maximum of 24 have been used, located in a circumference with a maximum radius of 1000 meters. Obviously, the proximity of the data for each point to be estimated will be different, which makes it necessary to solve a different kriging (co-kriging) equations system for each of the points on the map at which we wish to carry out an estimation.. On the other hand, this method allows the synthetic representation of the estimations of housing prices in the form of isoline maps. These maps are obtained from the estimations carried out on the nodes of a regular mesh. In this piece of work, the housing price has been estimated at each of the nodes of a 50 meter-sided regular mesh.

Figure 5.a shows the prices map obtained applying ordinary kriging. On this map it can be seen how the more highly valued areas coincide with the areas closest to the CBD. The least valued areas are, principally, in districts 2, 3, 6 and 14. The cross-validation results are shown in Figure 8.

Figure 6.a shows the pricings map using ordinary co-kriging with isotopic data (original sample of 287 houses). Figure 6.b shows the pricings map using ordinary co-kriging with heterotopic data (original sample of 287 houses plus a second sample of 259 houses). We observed greater discontinuity in figure 6.b than in figure 6.a. The cross-validation (Figure 7) does not show significant differences between ordinary co-kriging with isotropic and heterotopic data.

To finalize, we have obtained a hedonic regression model using data and the original sample variables. The goal is to compare the results obtained with the different procedures applied. The model parameters have been estimated by means of OLS (Ordinary Least Squares). Table 4 shows that for OLS all variables are significant using a level of confidence equal to 95%. Furthermore, the signs of the coefficients are as expected. For the purpose of making comparisons between the distinct methods applied, we present the graphs of the predicted versus measured values as well as the linear regression between both. As we can see in Figure 8, the results obtained with OLS are worse than with any of the other methods.

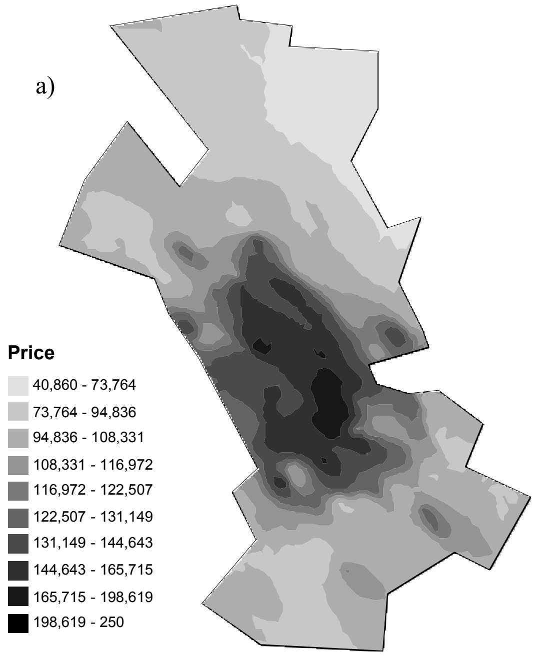

Figure 5. Map of price estimated by ordinary kriging.

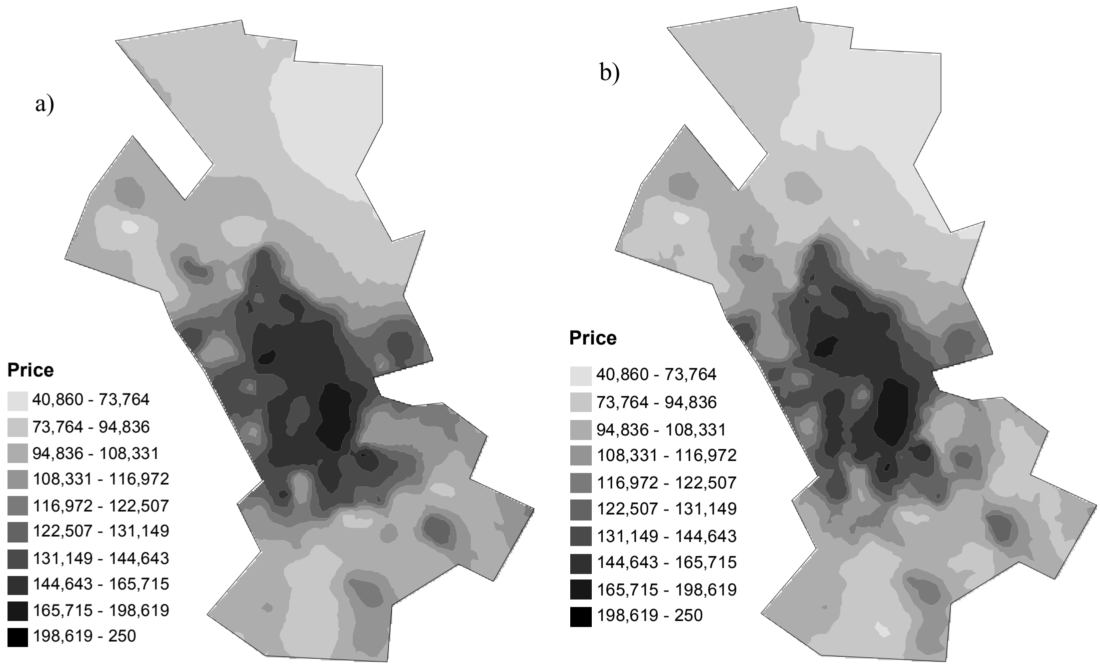

Figure 6. Cokriging with a) isotopic data and b) heterotopic data.

 ISBN 978-84-8439-519-5

Table 4. OLS regression.

	Coefficient	t-value
Constant	81.722	9.409
AGE	-0.568	-3.232
AMP	1.018	5.244
HEAT	25.215	7.364
R-squared: 0.372		

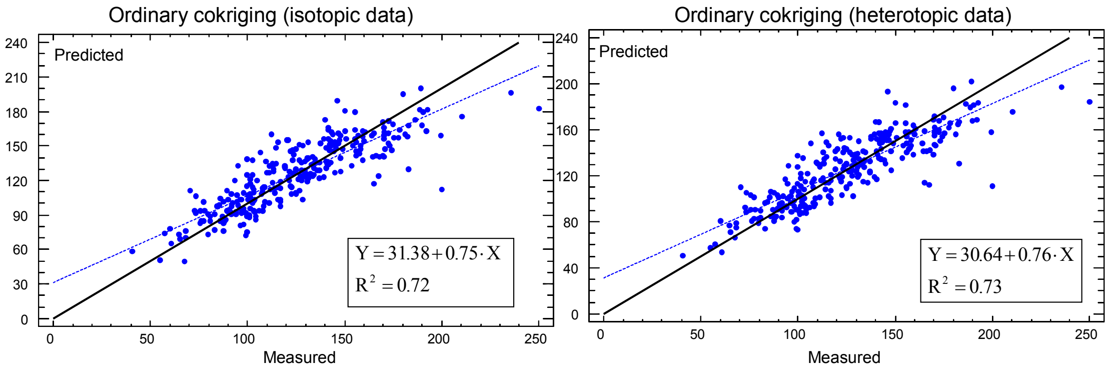

Figure 7. Cross-validation ordinary cokriging with isotopic and heterotopic data.

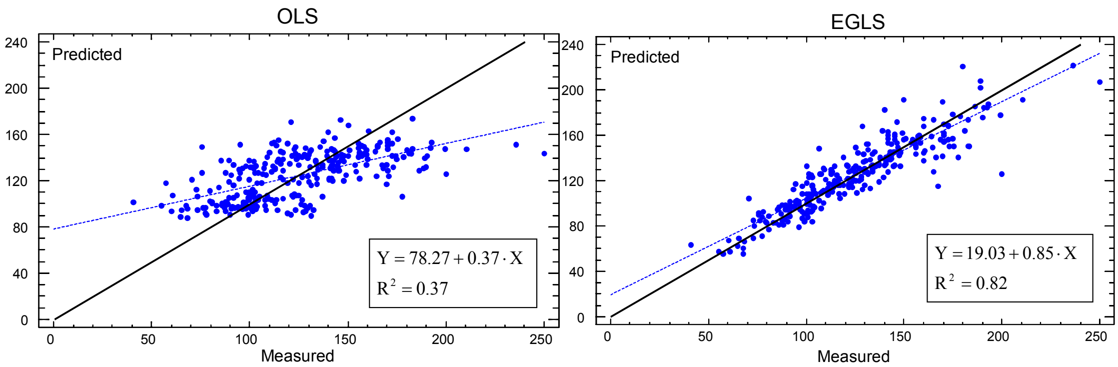

Figure 8. Predicted versus measured plot with OLS and EGLS.

4. CONCLUSIONS

This study suggests that using the kriging and co-kriging methods can be of interest for carrying out mass appraisal. Using these methods, we can obtain continuous maps of housing price that provide appraisers with an overall view of pricing.

The ordinary kriging method, under the hypothesis of quasi-stationarity, has provided results that are somewhat better than the detrend method. But, within the geostatistical methods applied, co-kriging has given the best results in the cross-validation. This multivariate method allows us to use isotopic and heterotopic data and, furthermore, enables house price appraisals to be made when the only available characteristic is location. On the other hand, the regression model can only operate with isotopic data and, for housing price appraisal, the houses' characteristics need to be known. Another interesting property of co-kriging, compared with the regression model, is that the presence of multicolinearity among explanatory variables is desirable in the former but not in the latter.

However, the method that produced the best values in the estimated values-observed values relation is the hedonic regression model applying EGLS. The worst method in this sense was OLS.

5. REFERENCES

Anselin, L.; Florax, R. J. (1995). *New directions in Spatial Econometrics*. Springer-Verlag.

Basu, S.; Thibodeau, T. G. (1998). *Analysis of Spatial Autocorrelation in House Prices*, Journal of Real Estate Finance and Economics 17:1, 61-85.

Can, A. (1990). *The measurement of neighborhood dynamics in urban house prices*. Economic Geography 66, 254-272.

Can, A. (1992). *Specification and estimation of hedonic housing price models*. Regional Science and Urban Economics 22, 453-474.

Cano Guervos, R.; Chica-Olmo, J.; Gutiérrez Hermoso, J. (2003). *A Geo-Statistical Method to Define Districts Within a City*, Journal of Real Estate Finance and Economics 27:1.

Chica-Olmo, J. (1995). *Spatial Estimation of Housing Prices and Locational Rents*. Urban Studies 32: 8, 1331-1344.

Chica-Olmo, J. (1992). *Adjusting of Regionalized Variable Theory to the economic spatial variables. Application to the housing prices in town from Granada.* Doctoral thesis.

Chilès, J.P.; Delfiner, P. (1999). *Geostatistics. Modeling Spatial Uncertainty*. USA: Wiley

Cressie, N. (1991). *Statistics for Spatial Data.* USA: John Wiley & Sons.

Dubin, R.A. (1992). *Spatial autocorrelation and neighborhood quality.* Regional Science and Urban Economics 22, 433-452.

Isaaks, E. H.; Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics*. New York: Oxford University Press.

Journel, A.G. ; Huijbregts, Ch. J. (1978). *Mining Geostatistics*. London: Academic Press.

Kitanidis, P.K. (1997). *Introduction to Geostatistics. Applications in Hydrogeology.* USA: Cambridge University Press.

Matheron, G. (1970). *La Théorie des variables regionalisées et ses applications.* Centre de Géostatistique et de Morphologie Mathematique. Fas. 1.

Neuman, S.P. ; Jacobson, E.A. (1984). *Analysis of non-intrinsic spatial variability by residual kriging with application to regional groundwater levels.* Mathematical Geology 16, 499-521.

Tse, Raymond Y. C. (2002). *Estimating Neighbourhood Effects in House Prices: Towards a New Hedonic Model Approach.* Urban Studies 39:7, 1165-1180.

Wackernagel, H. (1995). *Multivariate Geostatistics.* Germany: Springer-Verlag.

 ISBN 978-84-8439-519-5