

A MEAN-VARIANCE OPTIMIZATION APPROACH FOR RESIDENTIAL REAL ESTATE VALUATION

Francisco Guijarro

Universidad Politécnica de Valencia

e-mail: fraguima@upvnet.upv.es

Abstract

This paper introduces a new approach to the sales comparison model for the valuation of real estate that can objectively estimate the coefficients associated with the explanatory price variables. The coefficients of the price adjustment process are estimated from the formulation of a quadratic programming model similar to the mean-variance model in the portfolio selection problem and are shown to be independent of the property to be valued. It is also shown that the sales comparison model should minimize the variance of the adjusted prices, and not their coefficient of variation as indicated by some national and international valuation regulations. The paper concludes with a case study on the city of Medellín, Colombia.

Key words: *pricing, economic valuation, multiple linear regression model, quadratic programming, objective weights.*

JEL Classification: *C44, R30.*

Citation: Guijarro, F. (2021). A mean-variance optimization approach for residential real estate valuation. *Real Estate Management and Valuation*, 29(3), 13-28.

DOI: <https://doi.org/10.2478/remav-2021-0018>

1. Introduction

Price prediction is a matter of great importance in the real estate industry (Ahn et al., 2012). In fact, interest in property valuation has increased over time due to the appearance of new investment organizations such as real estate investment trusts (REITs), as well as for reasons of property taxes, investment decisions, sales transaction, price formation, etc. (Arribas et al., 2016; Raslanas et al., 2010). Since real estate is an important asset for most people, undervaluing or overvaluing it can cause a series of problems to both owners and buyers. Professional valuers estimate a property's equity in such a way that overvaluation implies underestimating the default risk, which can be passed on to buyers or secondary mortgage providers (Guo et al., 2013). Valuers must supplement their skills with valuation methods that can systematically analyze larger datasets with an output that is readily applicable to single-property appraisal (Kane et al., 2004).

Different methodologies have been proposed to estimate the value of real estate. Aznar et al. (2011) classify the different valuation methodologies into (1) Economic methods, (2) Non-economic methods, and (3) Mixed methods. The Economic methods provide a monetary value for the asset being considered, whereas Non-economic methods assign a value measured by a non-monetary scale to each asset. Mixed methods include a set of procedures that combine both methods with the aim of deriving more realistic and effective monetary valuations. Most of the studies in the literature focus on economic methods, since they are simpler and more objective when the necessary information is available (Dmytrow & Gnat, 2019). In this sense, we must point out that the application of one method or another is conditioned precisely by the quantity and quality of the information available. Researchers have found that many factors can affect prices and that the amount of sales evidence varies widely, but generally there are few sales of properties similar enough to be considered *comparable* and none that can be considered *identical*.

Within the Economic methods we can highlight two that are among the most frequently referenced by practitioners and researchers: the sales comparison model (also known as direct comparison) and the multiple linear regression model.

According to Appraisal Institute (1996), the sales comparison model estimates the market value by comparing the subject property to similar properties that have recently been sold, are listed for sale, or are under contract. A major premise of the sales comparison model is that the market value of a property is directly related to the prices of comparable and competitive properties, based on their similarities and differences. Once we know the characteristics of the property to be valued and comparable properties have been identified, the valuer adjusts the known price of the comparable properties according to their similarity with the property to be valued. The prices are *adjusted* so that if a comparable property is very different from the property which is to be valued, its market value must be drastically adjusted, and its adjusted price may differ significantly from its original price. If both properties are very similar, the market value of the comparable property only needs slight adjustment, so that its adjusted price may be similar to its original price. In the unlikely case of both properties being identical, no adjustment is needed and the adjusted price equals the original price.

Finally, the price of the property to be valued is calculated as the average of the adjusted prices and the quality of the result is measured by their dispersion. If the variability of the adjusted prices is small, it is concluded that the adjustment process has been good. On the other hand, if the variance of the adjusted prices is high, it is concluded that the properties cannot really be considered comparable to the property to be valued, and hence necessary to look for other properties with more similar characteristics. To eliminate the effects of scale, some regulations¹ propose measuring the success of price adjustment through the coefficient of variation, measured as the standard deviation of the adjusted prices divided by the average value of those prices, instead of measuring it directly through the variance or standard deviation of the adjusted prices (Antipov & Pokryshevskaya, 2012; Eckert et al., 1990; IAAO, 2003).

The sales comparison method is often questioned for relying too much on subjective judgments to obtain reliable and verifiable data (Wiltshaw, 1995). The estimated value may vary from one valuer to another based on their personal perceptions. The same valuer may even estimate different values for the same property at different times simply because of a change in the perception of the characteristics of the property, its location, its surroundings, etc. Subjectivity associated with the sales comparison model, as well as the increasing availability of information, have meant that the regression method has gained popularity. However, a growing trend is being shown in the application of machine learning methodologies, where non-linear relationships between variables can be captured more accurately.

In the multiple linear regression model, property prices are explained through several predictors and the model is adjusted by minimizing the sum of squared residuals (ordinary least squares regression). The regression model has the important advantage that the valuation function is objectively fitted and is therefore not affected by the bias of the valuer. Nevertheless, the valuation model also faces criticism, such as the need to obtain a large number of properties and predictor variables to explain price differences, nonlinearity within the data, multicollinearity issues in the predictor variables, and the inclusion of outliers in the sample (Ahn et al., 2012; Guijarro, 2019; Cupal et al., 2019).

Both methodologies explain prices from the characteristics that define the property, although the regression model usually requires a large number of properties, comparable to those demanded in the sales comparison model by valuation regulations. In the Spanish case, The Ministry Order EHA/564/2008 on property valuation standards stipulates that, for the sales comparison method, the expert must collect information from at least 6 comparable properties. A wide range of other methods have been proposed by the researchers in the field of real estate valuation that, in general, also estimate the valuation function or the value of the property by comparing a set of comparable properties: e.g. Quantile Regression (Ebru & Eban, 2011; Narula et al., 2012), Geographically Weighted Regression (Wu et al., 2016), Hierarchical Linear Model (Arribas et al., 2016; Brown & Uyar, 2004), Panel Data Analysis (Liu et al., 2017), Data Envelopment Analysis (Lins et al., 2005), Goal

¹ In the Colombian case, the Instituto Geográfico Agustín Codazzi through the Resolution N. 620, establishes that when the coefficient of variation is less than 7.5%, the mean obtained may be adopted as the most probable value assignable to the property.

Programming (Aznar & Guijarro, 2007), Analytic Hierarchy Process (Aznar et al., 2011), Analytic Network Process (Aznar et al., 2010). Artificial Neural Networks (García et al., 2008; Selim, 2009), Support Vector Machine (Kontrimas & Verikas, 2011; Plakandaras et al., 2015), Genetic Algorithm (Gu et al., 2011), Decision Tree (Fan et al., 2006; Cupal et al., 2019), Random Forest (Antipov & Pokryshevskaya, 2012), and Rough Set Theory (d'Amato, 2002, 2004, 2007). Recent papers are focused on the application of machine learning methods in the valuation of real estate, which has currently resulted in a significant increase in research focused on this area (Baldominos et al., 2015; Park & Bae, 2015; Hausler et al., 2018; Hu et al., 2019; Liu & Liu, 2019; Pérez-Rave et al., 2019; Liu & Wu, 2020). Valier (2020) states that "machine learning models are more accurate than traditional regression analysis in their ability to predict value. Nevertheless, many authors point out as their limit their black box nature and their poor inferential abilities".

The main objective of this work is thus to propose a new version of the sales comparison model for the valuation of real estate from a quadratic programming model, overcoming the inconvenience of subjectivity in the judgments issued by the valuers. The proposed model is compared with the multiple regression model and its similarity with the celebrated mean-variance model proposed by Markowitz (1952) for the portfolio selection problem is shown. It is also shown that optimizing the process of price adjustment through the coefficient of variation generates suboptimal solutions, since the adjusted prices are more dispersed than in the optimization of the variance of the adjusted prices. In addition, a positive bias (overvaluation) is introduced in the estimated price of the property to be valued, so price variance is optimized to obtain more precise and unbiased solutions.

The paper unfolds as follows: Section 2 introduces our modified version of the sales comparison model and highlights the similarities with the regression model and the mean-variance model. In Section 3, we apply the proposed model to a database of apartments in the city of Medellín (Colombia), analyzing the results according to the different variables used as predictors. Lastly, the concluding remarks are given in Section 4.

2. Methodology

Let us assume a set of n properties in addition to the property to be valued, so that $n + 1$, are comparable to each other. To explain the price p_i ($i = 1, \dots, n$), we have a total of k explanatory variables or predictors a_{ij} ($i = 1, \dots, n, j = 1, \dots, k$). The value of these variables is also known in the property to be valued a_j^* ($j = 1, \dots, k$), but not the price, which is precisely what has to be estimated. Since the characteristics of the properties are different from those of this property, the price of the comparable properties must be adjusted appropriately according to their characteristics and in relation to the same characteristics in the problem property. Finally, the price of the property is estimated as the average of the adjusted prices.

Instead of performing price adjustment by means of an expert's subjective judgment, we propose the quadratic mathematical program Eq. (1a) - (1f), a solution by which can estimate the price of the property \bar{h} without the need to consult the opinion of an expert valuer.

The objective function (Eq. 1a) minimizes the variance of the adjusted prices (Eq. 1b) of the properties considered in the comparable set. In the case study described in Section 3, we explain why it is preferable to optimize the variance instead of the coefficient of variation.

The adjusted price of the i -th property, h_i , is obtained by multiplying its original price p_i by the factor F_i . This factor is the product of other factors F_{ij} , so that each of these factors is obtained by comparing the j -th characteristic between the i -th comparable property and the property to be valued. The comparison for the j -th variable is performed by dividing a_j^* by a_{ij} . For example, if the j -th comparable property has an area of 100 m² and the property to be valued has an area of 90 m², the result of this comparison is $a_j^*/a_{ij} = 90/100 = 0.9$. This difference would justify the property to be valued having a price 10% lower than the comparable; i.e., the price of the comparable must be adjusted by a reduction of 10%. The other variables would also adjust the price of the i -th property, so that the different adjustments made have a multiplying effect - they could also be considered to have an additive effect. Logically, not all the variables need to have the same importance in determining the price, thus in constraint (Eq. (1f)), we introduce an adjustment factor f_j that graduates the importance of the k variables. In the example above, a 10% lower surface area would not necessarily imply a 10% lower price. This factor is precisely the variable that model (Eq. (1a) - (1f)) must solve, since the rest of the variables are either known or derived from the said adjustment factor. Once the value of f_j is

solved for the different explanatory variables, the estimated price of the property to be valued f_j (Eq. (1c)) is obtained as the average of the adjusted prices h_i (Eq. (1d)):

$$\text{Min } \sigma_h^2 \quad (1a)$$

$$\text{s.t. } \sigma_h^2 = \frac{1}{n} \sum_{i=1}^n (h_i - \bar{h})^2 \quad (1b)$$

$$\bar{h} = \frac{1}{n} \sum_{i=1}^n h_i \quad (1c)$$

$$h_i = p_i F_i \quad i = 1, \dots, n \quad (1d)$$

$$F_i = \prod_{j=1}^k F_{ij} \quad i = 1, \dots, n \quad (1e)$$

$$F_{ij} = (a_j^*/a_{ij})^{f_j} \quad i = 1, \dots, n \quad j = 1, \dots, k \quad (1f)$$

where:

p_i : i -th property price

a_{ij} : value for the j -th variable in the i -th property

a_j^* : value for the j -th variable in the property to be valued

f_j : factor of adjustment for the j -th variable

F_{ij} : adjusted factor for the j -th variable in the i -th property

F_i : compounded adjusted factor for the i -th property

h_i : adjusted price for the i -th property

σ_h : standard deviation of the adjusted prices

\bar{h} : mean of the adjusted prices

The way in which prices are adjusted explains why the objective function minimizes its variance; the price adjustment can make equivalent properties whose characteristics are not equivalent. When a property has characteristics superior to those of the property to be valued, its original price will be adjusted downwards to make it comparable to the property to be valued. If another property has inferior characteristics to the property to be valued, then its price will be corrected upwards. Finally, both adjusted prices try to estimate the price that the property would have, given the prices and characteristics of other similar properties, but not exactly equal to it, as regards their characteristics. In this way, each adjusted price constitutes an estimate of the value of the problem property. If the n estimates are very similar to each other, the vector of adjusted prices will have little variance and we can conclude that the consensus is high. However, if the adjusted price vector has a high variance, this will be a symptom that the adjustment process has not made consistent comparisons, since the estimated price of the property will vary according to the property used for its valuation.

Although model (Eq. (1a) – (1f)) can be solved using any commercial optimization software, below we show how a slight modification allows an analytical solution to be obtained.

The value of h_i can be expressed as follows:

$$h_i = p_i F_i = p_i \prod_{j=1}^k (a_j^*/a_{ij})^{f_j} \quad (2)$$

If we take logarithms, we reach the next equality:

$$\log(h_i) = \log(p_i \prod_{j=1}^k (a_j^*/a_{ij})^{f_j}) = \log(p_i) + \sum_{j=1}^k (f_j \log(a_j^*/a_{ij})) \quad (3)$$

Eq. (3) can be expressed in vector form:

$$\log(\mathbf{h}) = \log(\mathbf{p}) + \sum_{j=1}^k (f_j \log(a_j^*/\mathbf{a}_j)) \quad (4)$$

So that \mathbf{h} is an n -vector with the adjusted prices of the set of comparable properties, \mathbf{p} is an n -vector with the original prices of such properties, and \mathbf{a}_j is an n -vector with the values of the comparable properties in the j -th variable. The a_j^*/\mathbf{a}_j ratio is the result of the division of a_j^* by each of the n elements of \mathbf{a}_j .

Next we make the following variable changes:

$$\mathbf{x}_0 = \log(\mathbf{p}) \quad (5)$$

$$\mathbf{x}_j = \log(a_j^*/\mathbf{a}_j) \quad (6)$$

$$w_j = f_j \quad (7)$$

with $j = 1, \dots, k$. In this way we can rewrite Eq. (4) as a function of the new variables:

$$\log(\mathbf{h}) = w_0 \mathbf{x}_0 + w_1 \mathbf{x}_1 + \dots + w_k \mathbf{x}_k = \mathbf{w}^t \mathbf{X} \quad (8)$$

where \mathbf{X} is a matrix with n rows and $k + 1$ columns, and $w_0 = 1$. This expression points to greater similarities between the regression model and the sales comparison model than those reported in the literature and by professional valuation associations. Traditionally, both models have been considered within the same family of economic comparison methods. But just as the regression model infers a general valuation function that is then applied to a particular property to estimate its price, the sales comparison model adjusts the prices of each comparable property to finally estimate the price of the property to be valued as the simple average of the adjusted prices. Therefore, the sales comparison model does not estimate a valuation function, but directly estimates the price of the property.

However, Eq. (8) shows that the sales comparison model can also be understood from a holistic perspective in which the different variables contribute individually to the formation of the price, i.e. as a valuation function. Eq. (8) allows for the following quadratic programming model to be proposed:

$$\begin{array}{ll} \text{Min} & \mathbf{w}^t \mathbf{V} \mathbf{w} \quad (9a) \\ \text{s.t.} & \mathbf{0}_1^t \mathbf{w} = 1 \quad (9b) \end{array}$$

where \mathbf{V} represents the $(k + 1) \times (k + 1)$ covariance matrix of \mathbf{X} . The other two new elements are $(k + 1)$ -vectors: $\mathbf{w}^t = [w_0 \ w_1 \ \dots \ w_k]$ and $\mathbf{0}_1^t = [1 \ 0 \ \dots \ 0]$. The objective function in Eq. (9a) minimizes the variance of the logarithm of the adjusted prices \mathbf{h} , while the constraint Eq. (9b) forces the coefficient w_0 to take the value of 1. We must note that Eq. (9b) is actually not a constraint we impose on the model. It comes from Eq. (5) and is closely related to the intercept in the classical regression model.

The model Eq. (9a) – (9b) is similar to the model developed by Markowitz (1952) for the portfolio selection problem. The mean-variance approach is used due to the fact that this model proposes the minimization of the variance of a portfolio (risk) for a given return (average portfolio return). In our case, we do not impose any constraint on the average of prices because this is precisely the variable that we want to solve. But the objective function also consists of minimizing the variance and, as we will see in the next section, we can also impose a constraint on the average adjusted price to obtain a solution frontier in the mean-variance space. As in the model by Markowitz (1952), we also assume that \mathbf{V} is a definite positive. If this assumption is not guaranteed, the variable (or variables) causing the problem must be excluded. This does not influence the achieved solution because this variable is considered redundant (perfectly correlated with some of the remaining variables).

From the quadratic program of Eq. (9a) – (9b) we can state the KKT conditions:

$$L = \mathbf{w}^t \mathbf{V} \mathbf{w} - \lambda (\mathbf{0}_1^t \mathbf{w} - 1) \quad (10)$$

$$\partial L / \partial \mathbf{w} = 2 \mathbf{V} \mathbf{w} - \lambda \mathbf{0}_1 = 0 \quad (11)$$

$$\partial L / \partial \lambda = -(\mathbf{0}_1^t \mathbf{w} - 1) = 0 \quad (12)$$

In Eq. (11) we can clear the value of \mathbf{w} as a function of:

$$\mathbf{w} = \frac{1}{2} \lambda \mathbf{V}^{-1} \mathbf{0}_1 \quad (13)$$

If we premultiply on both sides of the equality by $\mathbf{0}_1^t$, and bearing in mind that $\mathbf{0}_1^t \mathbf{w} = 1$, we get:

$$\mathbf{0}_1^t \mathbf{w} = \frac{1}{2} \lambda \mathbf{0}_1^t \mathbf{V}^{-1} \mathbf{0}_1 = 1 \quad (14)$$

For convenience, we will identify to the matrix product $\mathbf{0}_1^t \mathbf{V}^{-1} \mathbf{0}_1$ as c . We clear the value of λ in Eq. (14):

$$\frac{1}{2} \lambda \mathbf{0}_1^t \mathbf{V}^{-1} \mathbf{0}_1 = \frac{1}{2} \lambda c = 1 \quad (15)$$

$$\lambda = 2/c \quad (16)$$

In this way we can already obtain the vector of coefficients \mathbf{w} :

$$\mathbf{w} = 1/c \mathbf{V}^{-1} \mathbf{0}_1 \quad (17)$$

The product $\mathbf{V}^{-1} \mathbf{0}_1$ returns a vector with the first row of the matrix \mathbf{V}^{-1} , so that Eq. (17) obtains the coefficients associated with each variable by dividing the elements of this vector by the scalar c .

The constant c is the (1,1) element of the matrix \mathbf{V}^{-1} . This scalar represents the inverse of the residual variance of the logarithm of the price versus the set of explanatory variables. In fact, the

elements of the diagonal of \mathbf{V}^{-1} contain the residual variance of a regression between the variable defined by the diagonal and the rest of the variables. If we call s_{ij} to the (i, j) -element of the matrix \mathbf{V}^{-1} , then we get:

$$w_j = s_{1j}/s_{11} \quad (18)$$

where $j = 1, \dots, (k + 1)$. Therefore, the solution to model Eq. (1a) – (1f) is obtained directly from the simple ratio of Eq. (18) between two elements of the matrix \mathbf{V}^{-1} . Here too we find a similarity with the traditional regression model, where the coefficient of the explanatory variable is obtained by dividing the covariance of this variable and the price by the price variance. In our model, the same elements are used in the calculation of the coefficients w_j , but of the matrix \mathbf{V}^{-1} instead of the matrix \mathbf{V} .

From Eq. (18) it is easy to show the relation between the coefficients w_j and the partial correlation between the logarithm of the price and the explanatory price variables. The partial correlation coefficient measures the linear dependence between two variables by controlling the effect of the remaining variables. The partial correlation $r_{12,3..k}$ between the variables $(\mathbf{x}_1, \mathbf{x}_2)$ controlling for the variables $(\mathbf{x}_3, \dots, \mathbf{x}_k)$ is calculated in two steps. First, the regression residual between \mathbf{x}_1 and the set of variables $(\mathbf{x}_3, \dots, \mathbf{x}_k)$ is obtained, along with the residual of the regression between \mathbf{x}_2 and the set of variables $(\mathbf{x}_3, \dots, \mathbf{x}_k)$; the partial correlation coefficient is then calculated as the simple correlation between the two residuals calculated in the first step. The partial correlation can be calculated from the elements of the matrix \mathbf{V}^{-1} :

$$r_{12,3..k} = -s_{12}/\sqrt{s_{11}s_{22}} \quad (19)$$

If we develop Eq. (19), we can express the partial correlation between the logarithm of the price and the j -th explanatory variable as a function of the coefficient w_j :

$$\begin{aligned} r_{1j, \text{rest of variables}} &= -s_{1j}/\sqrt{s_{11}s_{jj}} = -s_{11}s_{1j}/(s_{11}\sqrt{s_{11}s_{jj}}) = \\ &= -s_{11}w_j/\sqrt{s_{11}s_{jj}} = -w_j\sqrt{s_{11}/s_{jj}} \end{aligned} \quad (20)$$

And clearing w_j , we verify that the coefficients calculated in Eq. (18) have opposite signs to the partial correlation between the price logarithm and the explanatory variable analyzed:

$$w_j = -r_{1j, \text{rest of variables}}\sqrt{s_{jj}/s_{11}} \quad (21)$$

The sign of these coefficients is thus the opposite of what we would expect in a classical multiple regression model between the price and explanatory variables. However, as we have argued before, the process of price adjustment involves lowering the price of properties with better characteristics than the property to be valued, and adjusting the price of those with worse characteristics than that property. Hence the apparently counter-intuitive result obtained by estimating the coefficients w_j (f_j).

Another peculiarity of the coefficients w_j is that their values are independent of the property to be valued, since they only depend on the properties included in the comparable set. In the model Eq. (1a) – (1f), we have defined a_j^* as the value of the j -th variable in the property to be valued, so that the original explanatory variables are transformed as $\mathbf{x}_j = \log(a_j^*/\mathbf{a}_j)$. Then the solution of model Eq. (1a) – (1f) may be thought to be dependent on the characteristics of the property to be valued. However, this is not so.

If instead of wanting to value a property whose k -vector of characteristics is \mathbf{a}^* , the purpose is to value property b with characteristics \mathbf{b}^* , the solution $w_j = s_{1j}/s_{11}$ would remain unchanged. In this case, we would have $\mathbf{x}_j = \log(b_j^*/\mathbf{a}_j)$. This change would not affect the value of s_{11} , since the variance of the regression residuals between the logarithm of the price and the rest of variables does not vary by a change in the scale of the regressors. In the same way, s_{1j} would also remain unchanged. While the values of the coefficients w_j would be maintained, the estimated price for the property to be valued, \bar{h} , would change as a result of the change in adjusted prices h_i for the comparable set.

Finally, it is interesting to know to what extent the process of price adjustment has obtained very homogeneous adjusted prices. The ideal situation would be for all adjusted prices to have the same value, i.e., the variance of these prices would be zero. On the other hand, the worst possible situation would be that the adjusted prices coincide with the original prices. Parallel with the regression models, we can calculate the coefficient of determination of the adjustment made by the model with Eq. (22):

$$R^2 = 1 - \frac{\mathbf{w}^t \mathbf{v} \mathbf{w}}{\text{var}(x_0)} = 1 - \frac{(1/c^2) \mathbf{0}_1^t \mathbf{V}^{-1} \mathbf{V} \mathbf{V}^{-1} \mathbf{0}_1}{\text{var}(x_0)} = 1 - \frac{1}{c \times \text{var}(x_0)} \quad (22)$$

If we call v_{ij} to the element in the i -th row and the j -th column of the covariance matrix \mathbf{V} , Eq. (22) can be rewritten as $1 - 1/(s_{11}v_{11})$.

In the following section, we present a case study that serves as an illustrative example of the proposed model.

3. Empirical results

Table 1 shows information on 28 properties in the city of Medellín, Colombia, plus two properties to be valued in the last row. We want to estimate the price from 4 explanatory variables: surface area, Stratum, administration expenses and age. In Colombia, cities are usually segmented into strata, from 1 to 6, so that the poorer neighborhoods with the lowest real estate prices are grouped in Stratum 1, whereas the most expensive ones are classified as Stratum 6. For the properties to be considered comparable, most of the properties considered in the sample were from Strata 4 and 5. Age was another one of the variables considered in the research, though not as a continuous variable but as a categorical variable. The information was gathered from the main real estate website in Colombia, metrocuadrado.com, where age is not expressed in years but as a range of values in the form of a categorical variable: between 0 and 5 years, between 6 and 10 years, between 11 and 20 years, and more than 20 years.

Table 1

Characteristics of $n = 28$ properties used as the comparable set and two properties to be valued

# Property	Price / m ²	Area (m ²)	Stratum	Admin. Expenses	Age
1	1 808	130	4	1.808	0 - 5 years
2	1 622	148	5	0.743	+20 years
3	2 040	125	5	1.840	6 - 10 years
4	1 731	156	5	0.962	11 - 20 years
5	1 188	80	3	1.000	+20 years
6	2 054	112	5	1.830	0 - 5 years
7	1 688	80	4	1.288	11 - 20 years
8	1 566	166	5	0.602	+20 years
9	1 667	69	4	1.087	11 - 20 years
10	1 933	150	5	1.313	6 - 10 years
11	1 786	70	4	1.500	0 - 5 years
12	2 000	80	5	1.438	11 - 20 years
13	1 875	160	5	1.069	6 - 10 years
14	1 827	208	5	1.010	11 - 20 years
15	1 938	160	5	1.125	6 - 10 years
16	2 042	120	5	1.300	6 - 10 years
17	2 042	142	5	1.310	6 - 10 years
18	1 759	108	4	1.111	11 - 20 years
19	1 797	64	4	1.016	0 - 5 years
20	1 797	64	4	0.953	0 - 5 years
21	2 273	110	5	1.809	0 - 5 years
22	1 818	66	4	1.000	6 - 10 years
23	2 083	120	5	1.233	6 - 10 years
24	1 951	82	4	1.463	6 - 10 years
25	1 857	140	4	1.214	11 - 20 years
26	2 368	114	5	1.868	0 - 5 years
27	1 792	173	5	0.636	+20 years
28	1 941	152	4	1.289	6 - 10 years

29 (Example 1)	-	120	5	1.800	0 - 5 years
30 (Example 2)	-	70	4	0.925	11 - 20 years

Source: own study from www.metrocuadrado.com.

The last two rows contain the properties that illustrate the valuation proposal: properties 29 and 30. The first one has a surface area of 120 square meters, is classified as Stratum 5, has annual administration expenses of 1,800,000 Colombian pesos, and was built between 0 and 5 years ago. The second property has a surface area of 70 square meters, the Stratum is 4, administration expenses are 925,000 Colombian pesos, and the age is between 11 and 20 years.

The Stratum needs to be transformed before the application of the valuation model. If we consider this variable without being previously transformed, the result would be assuming a price behavior that does not necessarily meet the market principles. For example, if no transformation was applied on the Stratum, the model would assume that the (price) distance between Stratum 3 and 4 is the same as distance between Stratum 4 and 5, which might be correct, but does not necessarily have to be true in the real estate market. Our proposal is to overcome this limitation by following the same procedure used in the least-squares regression models: transforming n -level variables into $n - 1$ dichotomous variables.

Let's assume that $a_{Stratum}^*$ is the Stratum of the property to be valued, where the Stratum takes values between 3 and 5 in our case. This would translate into two dichotomous variables Stratum_4 and Stratum_5 that follow (23):

$$Stratum_k (k \text{ in } \{4,5\}) = \begin{cases} a_{i,Stratum}/10 & a_{i,Stratum} = a_{Stratum}^* \\ a_{i,Stratum} & a_{i,Stratum} \neq a_{Stratum}^* \end{cases} \quad (23)$$

Hence, values of Stratum_4 and Stratum_5 depend on the values of the Stratum in the property to be valued (Table 2). The same approach was used to transform the Age variable with 4 original levels into 3 dichotomous variables: age_6_10, age_11_20, and age_20.

Table 2

Transformation of Stratum (3 different levels) into dichotomous variables (2 levels) for properties 29 and 30

Original Stratum	Variables to be computed			
	Property 29 to be valued (Stratum 5)		Property 30 to be valued (Stratum 4)	
	Stratum_4	Stratum_5	Stratum_4	Stratum_5
4	0.5	5	0.4	4
5	5	0.5	4	0.4
5	5	0.5	4	0.4
5	5	0.5	4	0.4
3	5	5	4	4
5	5	0.5	4	0.4
4	0.5	5	0.4	4
5	5	0.5	4	0.4
4	0.5	5	0.4	4
5	5	0.5	4	0.4
4	0.5	5	0.4	4
5	5	0.5	4	0.4
5	5	0.5	4	0.4
5	5	0.5	4	0.4
5	5	0.5	4	0.4
5	5	0.5	4	0.4
5	5	0.5	4	0.4
4	0.5	5	0.4	4
4	0.5	5	0.4	4

4	0.5	5	0.4	4
5	5	0.5	4	0.4
4	0.5	5	0.4	4
5	5	0.5	4	0.4
4	0.5	5	0.4	4
4	0.5	5	0.4	4
5	5	0.5	4	0.4
5	5	0.5	4	0.4
4	0.5	5	0.4	4
5	0.5	5	0.4	4
4	5	0.5	4	0.4

Source: own study.

These values serve to calculate columns in Tables 3 and 4, which record the logarithmic transformation of the explanatory variables corrected by the values of the property to be valued (x_1 to x_4). Another column shows the logarithm of the price per square metre (x_0) for the 28 properties that form the set of comparable properties.

From these tables, the covariance matrix V and its inverse V^{-1} are calculated. The first 4 columns of Table 5 show the values of matrix V , while the next 4 columns include those of V^{-1} . Note how the V matrix remains constant for properties 29 and 30; that is, it does not depend on the specific values taken by the properties to be valued. The last column of the table contains the coefficients w_j , which have been calculated from Eq. (18). Like the V matrix, w weights are also independent of the property to be valued.

Table 5 includes the weights computed by the proposed model. In other words, these weights compute the factors of adjustment f_j for those variables involved in the valuation process, according to equation (7).

Equations (5)-(8) serve to compute the factors of adjustment by using the ratio between the specific values of the property to be valued and those values corresponding to the comparable set. The higher the value of a_{ij} , the lower the value of $\log(a_j^*/a_{ij})$. This term is multiplied by the factor of adjustment, $f_j \log(a_j^*/a_{ij})$; hence a negative value of f_j (w_j) translates into a positive relation between the estimated price and feature j , while a positive coefficient for w_j translates into a negative relation between the estimated price and feature j .

According to the last column of Table 5, weights of Area, Stratum_4 and Stratum_5 are negative. This would result in a positive impact of area on prices, and higher estimated prices for properties in Stratum 4 and Stratum 5 (when compared with Stratum 3). The coefficient associated with age_6_10 is close to zero (-0.001), hence we can conclude that the estimated prices for properties between 6 and 10 years of age are similar to those of properties between 0 and 5 years old. However, age_11_20 and age_20 obtain positive coefficients: 0.023 and 0.035, respectively. On average, the estimated price for a 11-20 year-old property is lower than the estimated price for a nearly new property (0-5 years old), and even lower for a property built 20 years ago or more, *ceteris paribus*.

Table 3

Logarithmic transformation of the explanatory variables corrected by the values of the property to be valued: property 29

# Pro- perty	x_0 log(Price / m ²)	x_1 log(120/ Area)	x_{21} log(5/Stra- tum_4)	x_{22} log(5/Stra- tum_5)	x_3 log(1.8/Ad min. Expenses)	x_{41} log(1/ag e_6_10)	x_{42} log(1/age _11_20)	x_{43} log(1/ag e_20)
1	3.257	-0.035	1	0	-0.002	0	0	0
2	3.210	-0.091	0	1	0.384	0	0	1
3	3.310	-0.018	0	1	-0.010	1	0	0
4	3.238	-0.114	0	1	0.272	0	1	0

5	3.075	0.176	0	0	0.255	0	0	1
6	3.313	0.030	0	1	-0.007	0	0	0
7	3.227	0.176	1	0	0.145	0	1	0
8	3.195	-0.141	0	1	0.476	0	0	1
9	3.222	0.240	1	0	0.219	0	1	0
10	3.286	-0.097	0	1	0.137	1	0	0
11	3.252	0.234	1	0	0.079	0	0	0
12	3.301	0.176	0	1	0.098	0	1	0
13	3.273	-0.125	0	1	0.226	1	0	0
14	3.262	-0.239	0	1	0.251	0	1	0
15	3.287	-0.125	0	1	0.204	1	0	0
16	3.310	0.000	0	1	0.141	1	0	0
17	3.310	-0.073	0	1	0.138	1	0	0
18	3.245	0.046	1	0	0.210	0	1	0
19	3.255	0.273	1	0	0.248	0	0	0
20	3.255	0.273	1	0	0.276	0	0	0
21	3.357	0.038	0	1	-0.002	0	0	0
22	3.260	0.260	1	0	0.255	1	0	0
23	3.319	0.000	0	1	0.164	1	0	0
24	3.290	0.165	1	0	0.090	1	0	0
25	3.269	-0.067	1	0	0.171	0	1	0
26	3.374	0.022	0	1	-0.016	0	0	0
27	3.253	-0.159	0	1	0.452	0	0	1
28	3.288	-0.103	1	0	0.145	1	0	0

Source: own study.

Table 4

Logarithmic transformation of the explanatory variables corrected by the values of the property to be valued: property 30

#	x_0	x_1	x_{21}	x_{22}	x_3	x_{41}	x_{42}	x_{43}
Pro- perty	log(Price / m ²)	log(70/ Area)	log(4/ Stratum_4)	log(4/Stra- tum_5)	log(0.925/ Admin. Expenses)	log(3/age_6 _10)	log(3/age _11_20)	log(3/age _20)
1	3.257	-0.269	1	0	-0.291	0	0	0
2	3.210	-0.325	0	1	0.095	0	0	1
3	3.310	-0.252	0	1	-0.299	1	0	0
4	3.238	-0.348	0	1	-0.017	0	1	0
5	3.075	-0.058	0	0	-0.034	0	0	1
6	3.313	-0.204	0	1	-0.296	0	0	0
7	3.227	-0.058	1	0	-0.144	0	1	0
8	3.195	-0.375	0	1	0.187	0	0	1
9	3.222	0.006	1	0	-0.070	0	1	0
10	3.286	-0.331	0	1	-0.152	1	0	0
11	3.252	0.000	1	0	-0.210	0	0	0

12	3.301	-0.058	0	1	-0.192	0	1	0
13	3.273	-0.359	0	1	-0.063	1	0	0
14	3.262	-0.473	0	1	-0.038	0	1	0
15	3.287	-0.359	0	1	-0.085	1	0	0
16	3.310	-0.234	0	1	-0.148	1	0	0
17	3.310	-0.307	0	1	-0.151	1	0	0
18	3.245	-0.188	1	0	-0.080	0	1	0
19	3.255	0.039	1	0	-0.041	0	0	0
20	3.255	0.039	1	0	-0.013	0	0	0
21	3.357	-0.196	0	1	-0.291	0	0	0
22	3.260	0.026	1	0	-0.034	1	0	0
23	3.319	-0.234	0	1	-0.125	1	0	0
24	3.290	-0.069	1	0	-0.199	1	0	0
25	3.269	-0.301	1	0	-0.118	0	1	0
26	3.374	-0.212	0	1	-0.305	0	0	0
27	3.253	-0.393	0	1	0.163	0	0	1
28	3.288	-0.337	1	0	-0.144	1	0	0

Source: own study.

Table 5

Inverse of the covariance matrix and coefficients solutions for the proposed model

Covariance matrix							
0.0030	-0.0013	-0.0044	0.0113	-0.0042	0.0092	-0.0039	-0.0120
-0.0013	0.0222	0.0421	-0.0475	-0.0038	-0.0133	0.0013	-0.0114
-0.0044	0.0421	0.2385	-0.2245	-0.0046	-0.0332	0.0446	-0.0561
0.0113	-0.0475	-0.2245	0.2449	0.0018	0.0459	-0.0357	0.0255
-0.0042	-0.0038	-0.0046	0.0018	0.0163	-0.0105	0.0041	0.0305
0.0092	-0.0133	-0.0332	0.0459	-0.0105	0.2296	-0.0893	-0.0510
-0.0039	0.0013	0.0446	-0.0357	0.0041	-0.0893	0.1875	-0.0357
-0.0120	-0.0114	-0.0561	0.0255	0.0305	-0.0510	-0.0357	0.1224

Inverse of the covariance matrix								w
2,526.64	-28.08	-353.96	-450.11	418.79	-2.42	57.62	89.16	1.000
-28.08	92.18	18.63	34.26	-10.77	8.50	8.14	15.79	-0.011
-353.96	18.63	103.84	112.51	-100.53	6.36	-2.06	18.11	-0.140
-450.11	34.26	112.51	132.66	-108.54	5.71	-3.86	11.07	-0.178
418.79	-10.77	-100.53	-108.54	245.57	-15.13	-10.85	-53.80	0.166
-2.42	8.50	6.36	5.71	-15.13	9.10	6.35	11.69	-0.001
57.62	8.14	-2.06	-3.86	-10.85	6.35	12.39	15.24	0.023
89.16	15.79	18.11	11.07	-53.80	11.69	15.24	47.09	0.035

Source: own study.

The adjusted prices are obtained from the coefficients estimated by the model, whose average is 3,150 for property 29 and 3,105 for property 30 (Table 6). If we undo the logarithmic transformation, we obtain the estimated price per square meter for the property to be valued: $10^{3.150} = 1,414.03$

Colombian pesos per square meter for property 29 and $10^{3.105} = 1,273.90$ Colombian pesos per square meter for property 30. The variance of the adjusted log prices is obtained by the expression $\mathbf{w}^t \mathbf{V} \mathbf{w} = 0.00039578$.

Table 6

Log adjusted prices for the comparable set

# Property	Market Price / m ²	Property to be valued:	
		29	30
		Log adjusted Price	Log adjusted Price
1	1 808	3.117	3.072
2	1 622	3.132	3.087
3	2 040	3.129	3.084
4	1 731	3.129	3.084
5	1 188	3.150	3.105
6	2 054	3.133	3.088
7	1 688	3.132	3.087
8	1 566	3.132	3.087
9	1 667	3.138	3.093
10	1 933	3.131	3.086
11	1 786	3.122	3.077
12	2 000	3.160	3.115
13	1 875	3.133	3.087
14	1 827	3.151	3.105
15	1 938	3.143	3.098
16	2 042	3.154	3.109
17	2 042	3.155	3.109
18	1 759	3.162	3.117
19	1 797	3.153	3.107
20	1 797	3.157	3.112
21	2 273	3.178	3.132
22	1 818	3.158	3.113
23	2 083	3.167	3.121
24	1 951	3.162	3.117
25	1 857	3.181	3.135
26	2 368	3.193	3.148
27	1 792	3.187	3.142
28	1 941	3.172	3.127
Mean log price for property 29		3.150	
Mean log price for property 30			3.105

Source: own study.

The value of the determination coefficient is $R^2 = 1 - 1/(s_{11}v_{11}) = 1 - 1/(2,526.64 \times 0.0030) = 0.8680 \cong 86.8\%$, which shows that a relatively good price adjustment has been obtained, albeit far from ideal and unrealistic, which would imply a determination coefficient of 100%.

The coefficient of variation was also calculated as the standard deviation of the adjusted prices divided by their average, since, as mentioned in the introductory section, this is the measure used by

some national and international associations to measure the quality of valuation processes. The result obtained was a coefficient of variation of $0.00039578^{0.5}/3.150 = 0.63\%$ for property 29 and $0.00039578^{0.5}/3.105 = 0.64\%$ for property 30, much lower than the maximum threshold of 7.5% marked by the Colombian regulations.

The solution found in model Eq. (9) minimizes the variance of the logarithm of the adjusted prices of the 28 comparable properties. Continuing with the similarity between the proposed model and that proposed by Markowitz (1952), we can obtain a solution frontier for different values of the logarithm of the estimated price. For this, we propose the model Eq. (24a) – (24c):

$$\text{Min} \quad \mathbf{w}^t \mathbf{V} \mathbf{w} \quad (24a)$$

$$\text{s.t.} \quad \mathbf{0}_1^t \mathbf{w} = 1 \quad (24b)$$

$$\frac{1}{n} \mathbf{w}^t \mathbf{X} \mathbf{1} = h^* \quad (24c)$$

where $\mathbf{1}^t = [1 \ \dots \ 1]$ is a $(k + 1)$ - vector and h^* is the particular value of the logarithm of the price we want to obtain.

Figure 1 represents the frontier obtained by the model Eq. (24a) – (24c) for different values of h^* . The leftmost point of the curve labeled *Minimal variance* represents the solution obtained with model Eq. (9). In the same figure, we also represent the solution of the problem in which, instead of minimizing the variance of prices, we had minimized the coefficient of variation, as proposed by some valuation regulations. The straight line in Figure 1 represents the inverse of the coefficient of variation, and its steepest slope is at the tangent point labeled *Minimal coef. of variation*, which is precisely the solution we would get by minimizing the coefficient of variation instead of the variance.

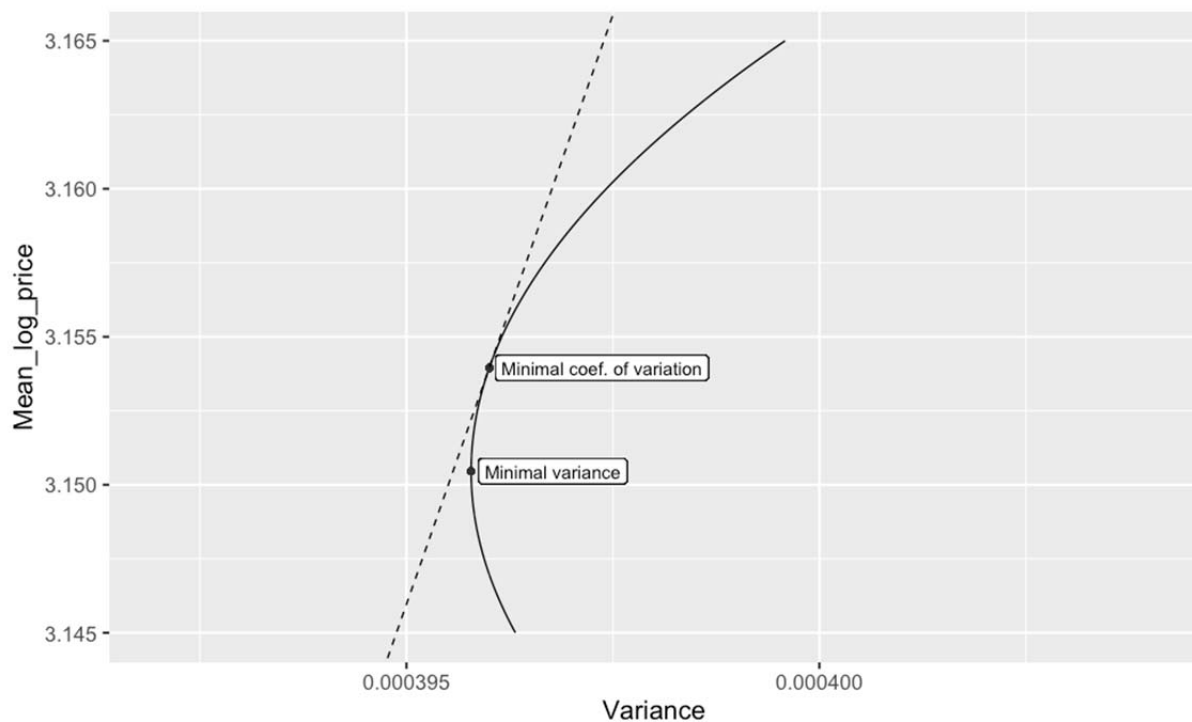


Fig. 1. Mean-variance frontier. Source: own study.

We can observe that the solution obtained when optimizing the coefficient of variation is worse than that obtained by the proposed model, since its variance is greater. We must remember that the greater the variance of the adjusted prices, the less the similarity of the set of comparable properties to the property studied, taking into account the explanatory variables. In an extreme case, a model that presented zero variance in the adjusted prices would indicate that the comparable properties allow us to perfectly adjust the price of the property to be valued, with a coefficient of determination of 100%. In addition, since prices are always going to take positive values, it is easy to show that the solution found in optimizing the coefficient of variation will always obtain a price higher than that obtained when optimizing variance, i.e. this model will lead to a positive bias over the estimated price. In the

case study, this bias is only 1,000 Colombian pesos, but if we look closely at Figure 1, we can see how the bias increases as the variance of adjusted prices -at equal prices- increases, i.e. the adjustment process was less reliable.

Finally, we can also analyze the relative importance of the explanatory variables that we consider in our model. Figure 2 compares different models with two explanatory variables and the one in the previous figure with all four explanatory price variables. We can see how the model that includes the stratum and administration expenses is close to the general model that includes all variables, with $R^2 = 83.3\%$. The combination of Stratum and age also gets a high $R^2 = 81.6\%$. However, all models that include the area obtain a worse fit.

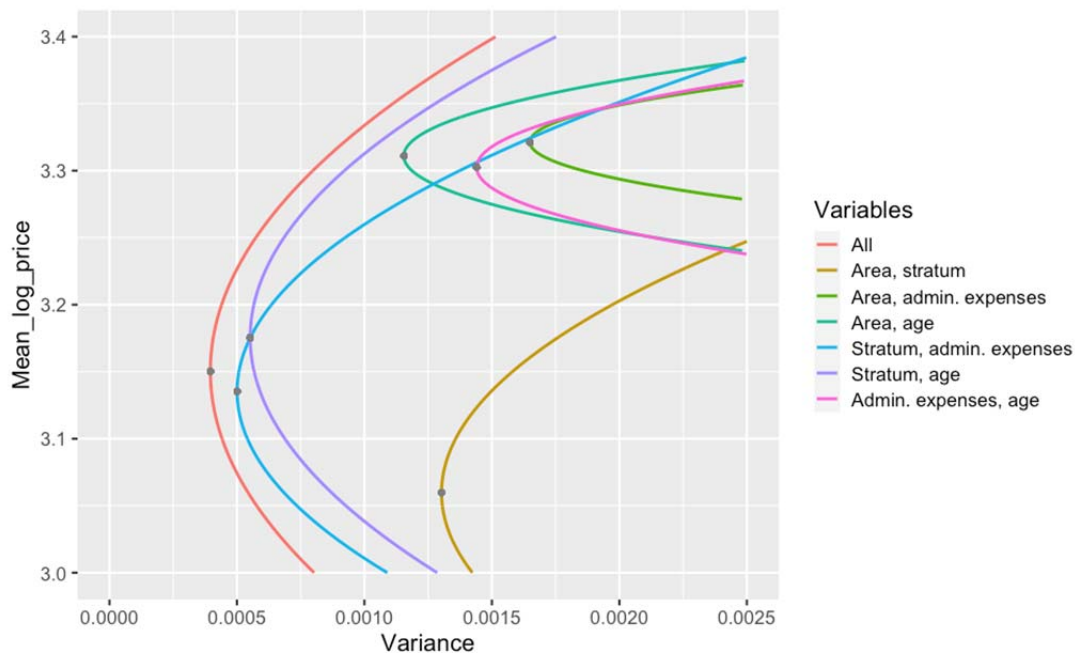


Fig. 2. Frontiers obtained from the subsets of explanatory variables. *Source:* own study.

4. Conclusions

This paper proposes a new approach to the sales comparison model for the valuation of real estate. Within the economic valuation models, the sales comparison and multiple regression models share some similarities and differences that have been highlighted in this study. The main disadvantage of the sales comparison model versus the regression models is that it involves a subjective comparison by an expert valuer between the property to be valued and the properties included in the comparable set, and thus the price adjustment process can be seriously biased. The model proposed in this paper determines such comparisons objectively through a quadratic programming model similar to that used in the portfolio selection problem.

Although it has been considered that the sales comparison model does not obtain a valuation function, our work shows that the proposed model does indeed generate a valuation function that is independent of the property to be valued, so that the coefficients of the variables are obtained as a simple ratio between two elements of the inverse of the covariance matrix.

The study also shows that the process of price adjustment should not be carried out according to the criterion of minimizing the coefficient of variation, contrary to what is suggested by different national and international valuation regulations. This procedure obtains solutions that are less precise and have a positive bias on the estimated price, compared to the proposed alternative of minimizing the variance of the adjusted prices. The present work thus shows that these regulations should be modified to update the valuation procedures along the lines indicated here. The results obtained from the case study verified that a model that barely obtains a coefficient of determination of 86.8% would comply with the quality requirements imposed by valuation regulations, which should also lead to a revision of these standards to increase the validation requirements of their models.

5. References

- Ahn, J. J., Byun, H. B., Oh, K. Y., & Kim, Y. K. (2012). Using ridge regression with genetic algorithm to enhance real estate appraisal forecasting. *Expert Systems with Applications*, 39, 8369–8379. <https://doi.org/10.1016/j.eswa.2012.01.183>
- Antipov, E. A., & Pokryshevskaya, E. B. (2012). Mass appraisal of residential apartments: An application of random forest for valuation and a cart-based approach for model diagnostics. *Expert Systems with Applications*, 39(2), 1772–1778. <https://doi.org/10.1016/j.eswa.2011.08.077>
- Appraisal Institute. (1996). The appraisal of real estate. Appraisal Institute.
- Arribas, I., García, F., Guijarro, F., Oliver, J., & Tamosiuniene, R. (2016). Mass appraisal of residential real estate using multilevel modelling. *International Journal of Strategic Property Management*, 20(1), 77–87. <https://doi.org/10.3846/1648715X.2015.1134702>
- Aznar, J., Ferris-Oñate, J., & Guijarro, F. (2010). An ANP framework for property pricing combining quantitative and qualitative attributes. *The Journal of the Operational Research Society*, 61(5), 740–755. <https://doi.org/10.1057/jors.2009.31>
- Aznar, J., & Guijarro, F. (2007). Estimating regression parameters with imprecise input data in an appraisal context. *European Journal of Operational Research*, 176(3), 1896–1907. <https://doi.org/10.1016/j.ejor.2005.10.029>
- Aznar, J., Guijarro, F., & Moreno-Jiménez, J. M. (2011). Mixed valuation methods: A combined AHP-GP procedure for individual and group multicriteria agricultural valuation. *Annals of Operations Research*, 190(1), 221–238. <https://doi.org/10.1007/s10479-009-0527-2>
- Baldominos, A., Blanco, I., Moreno, A. J., Iturrarte, R., Bernárdez, Ó., & Afonso, C. (2018). Identifying real estate opportunities using machine learning. *Applied Sciences (Basel, Switzerland)*, 8(11), 2321. <https://doi.org/10.3390/app8112321>
- Brown, K., & Uyar, B. (2004). A hierarchical linear model approach for assessing the effects of house and neighborhood characteristics on housing prices. *Journal of Real Estate Practice and Education*, 7(1), 15–24. <https://doi.org/10.1080/10835547.2004.12091603>
- Cupal, M., Sedlacik, M., & Michalek, J. (2019). The assessment of a building's insurable value using multivariate statistics: The case of the Czech Republic. *Real Estate Management and Valuation*, 27(3), 81–96. <https://doi.org/10.2478/remav-2019-0027>
- d'Amato, M. (2002). Appraising property with rough set theory. *Journal of Property Investment & Finance*, 20(4), 406–418. <https://doi.org/10.1108/14635780210435074>
- d'Amato, M. (2004). A comparison between MRA and rough set theory for mass appraisal. A case in Bari. *International Journal of Strategic Property Management*, 8(4), 205–217. <https://doi.org/10.3846/1648715X.2004.9637518>
- d'Amato, M. (2007). Comparing rough set theory with multiple regression analysis as automated valuation methodologies. *International Real Estate Review*, 10(2), 42–65.
- Dmytrow, K., & Gnat, S. (2019). Application of AHP method in assessment of the influence of attributes on value in the process of real estate valuation. *Real Estate Management and Valuation*, 27(4), 15–26. <https://doi.org/10.2478/remav-2019-0032>
- Ebru, C., & Eban, A. (2011). Determinants of house prices in Istanbul: A quantile regression approach. *Quality & Quantity*, 45(2), 305–317. <https://doi.org/10.1007/s11135-009-9296-x>
- Eckert, J. K., Gloudemans, R. J., & Almy, R. R. (1990). Property appraisal and assessment administration. International Association of Assessing Officers.
- Fan, G. Z., Ong, S. E., & Koh, H. C. (2006). Determinants of house price: A decision tree approach. *Urban Studies (Edinburgh, Scotland)*, 43(12), 2301–2315. <https://doi.org/10.1080/00420980600990928>
- García, N., Gámez, M., & Alfaro, E. (2008). ANN+GIS: An automated system for property valuation. *Neurocomputing*, 71(4–6), 733–742. <https://doi.org/10.1016/j.neucom.2007.07.031>
- Gu, J., Zhu, M., & Jiang, L. (2011). Housing price forecasting based on genetic algorithm and support vector machine. *Expert Systems with Applications*, 38(4), 3383–3386. <https://doi.org/10.1016/j.eswa.2010.08.123>
- Guijarro, F. (2019). Assessing the impact of road traffic externalities on residential price values: A case study in Madrid, Spain. *International Journal of Environmental Research and Public Health*, 16(24), 5149. <https://doi.org/10.3390/ijerph16245149> PMID:31861055

- Hausler, J., Ruscheinsky, J., & Lang, M. (2018). News-based sentiment analysis in real estate: A machine learning approach. *Journal of Property Research*, 35(4), 344–371. <https://doi.org/10.1080/09599916.2018.1551923>
- Hu, L., He, S., Han, Z., Xiao, H., Su, S., Weng, M., & Cai, Z. (2019). Monitoring housing rental prices based on social media: An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies. *Land Use Policy*, 82, 657–673. <https://doi.org/10.1016/j.landusepol.2018.12.030>
- Guo, J., Xu, S., & Bi, Z. (2013). An integrated cost-based approach for real estate appraisals. *Information Technology and Management*, 15(2), 131–139. <https://doi.org/10.1007/s10799-012-0152-7>
- IAAO. (2003). Standard on automated valuation models (AVMs). International Association of Assessing Officers.
- Kane, M. S., Linne, M. R., & Johnson, J. A. (2004). Practical Applications in Appraisal Valuation Modeling: Statistical Methods for Real Estate Practitioners. Appraisal Institute.
- Kontrimas, V., & Verikas, A. (2011). The mass appraisal of the real estate by computational intelligence. *Applied Soft Computing*, 11(1), 443–448. <https://doi.org/10.1016/j.asoc.2009.12.003>
- Lins, M. P. E., Novaes, L. F. L., & Legey, L. F. L. (2005). Real estate appraisal: A double perspective data envelopment analysis approach. *Annals of Operations Research*, 138(1), 79–96. <https://doi.org/10.1007/s10479-005-2446-1>
- Liu, F., Liu, D., Malekian, R., Li, Z., & Wang, D. (2017). A measurement model for real estate bubble size based on the panel data analysis: An empirical case study. *PLoS One*, 12(3), e0173287. <https://doi.org/10.1371/journal.pone.0173287> PMID:28273141
- Liu, L., & Wu, L. (2020). Predicting housing prices in China based on modified Holt's exponential smoothing incorporating whale optimization algorithm. *Socio-Economic Planning Sciences*, 72, 100916. Advance online publication. <https://doi.org/10.1016/j.seps.2020.100916>
- Liu, R., & Liu, L. (2019). Predicting housing price in China based on long short-term memory incorporating modified genetic algorithm. *Soft Computing*, 23(22), 11829–11838. <https://doi.org/10.1007/s00500-018-03739-w>
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77–91. <https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>
- Narula, S. C., Wellington, J. F., & Lewis, S. A. (2012). Valuating residential real estate using parametric programming. *European Journal of Operational Research*, 217(1), 120–128. <https://doi.org/10.1016/j.ejor.2011.08.014>
- Park, B., & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. *Expert Systems with Applications*, 42(6), 2928–2934. <https://doi.org/10.1016/j.eswa.2014.11.040>; <https://doi.org/10.1016/j.eswa.2015.03.005>
- Pérez-Rave, J. I., Correa-Morales, J. C., & González-Echavarría, F. (2019). A machine learning approach to big data regression analysis of real estate prices for inferential and predictive purposes. *Journal of Property Research*, 36(1), 59–96. <https://doi.org/10.1080/09599916.2019.1587489>
- Plakandaras, V., Gupta, R., Gogas, P., & Papadimitriou, T. (2015). Forecasting the US real house price index. *Economic Modelling*, 45, 259–267. <https://doi.org/10.1016/j.econmod.2014.10.050>
- Raslanas, S., Zavadskas, E. K., Kaklauskas, A., & Zabulenias, A. R. (2010). Land value tax in the context of sustainable urban development and assessment. Part II - Analysis of land valuation techniques: The case of Vilnius. *International Journal of Strategic Property Management*, 14(2), 173–190. <https://doi.org/10.3846/ijspm.2010.13>
- Selim, H. (2009). Determinants of house prices in turkey: Hedonic regression versus artificial neural network. *Expert Systems with Applications*, 36(2), 2843–2852. <https://doi.org/10.1016/j.eswa.2008.01.044>
- Valier, A. (2020). Who performs better? AVMs vs hedonic models. *Journal of Property Investment & Finance*, 38(3), 213–225. <https://doi.org/10.1108/JPIF-12-2019-0157>
- Wiltshaw, D. (1995). A comment on methodology and valuation. *Journal of Property Research*, 12(2), 157–161. <https://doi.org/10.1080/09599919508724139>
- Wu, C., Ye, X., Ren, F., Wan, Y., Ning, P., & Du, Q. (2016). Spatial and social media data analytics of housing prices in Shenzhen, China. *PLoS One*, 11(10), e0164553. <https://doi.org/10.1371/journal.pone.0164553> PMID:27783645