ADVANCES IN GEODESY AND GEOINFORMATION
(formerly GEODESY AND CARTOGRAPHY)
Vol. 71, no. 1, article no. e20, 2022

https://doi.org/10.24425/gac.2022.141173

Original paper

Real estate valuation algorithm for large local markets

Agnieszka Bitner*, Malgorzata Frosik

University of Agriculture in Krakow, Krakow, Poland
e-mail: rmbitner@cyf-kr.edu.pl; ORCID: http://orcid.org/0000-0002-5733-3987
e-mail: malgorzata.frosik@gmail.com; ORCID: http://orcid.org/0000-0002-8140-0692

*Corresponding author: Agnieszka Bitner, e-mail: rmbitner@cyf-kr.edu.pl

Received: 2021-01-30 / Accepted: 2022-04-12

Abstract: The article deals with issues related to the application of statistical methods used in the valuation process. The proposed algorithm for real estate valuation can be used in the statistical market analysis method in the process of mass appraisal. The algorithm uses a multiple linear regression model. Legal considerations indicate the need for such an algorithm for the determination of the value of representative properties. Due to the large size of the database of comparables, the proposed algorithm can be used only to appraise typical properties. A good statistical model is parsimonious, that is, it uses as few mathematical concepts as possible in a given situation. A model should extract what is systematic in the results observed, allowing for the presence of purely random deviations. The article discusses the basic principles of building a good statistical model. Attention is drawn to the number of market attributes that are entered into the model and the range of their values. As few explanatory variables as possible should be entered into the model to explain the phenomenon under study. Explanatory variables are only those characteristics of the property that differentiate prices in a given market defined and adopted by the appraiser as the basis for valuation. The article highlights the importance of taking into account market changes during the period under study.

Keywords: mass appraisal, statistical model, multiple regression, real estate valuation

1. Introduction

The development of the free real estate market and spatial data infrastructure (Bielecka et al., 2018) in Poland resulted in an increased demand for more and more efficient real estate valuation tools. Thanks to the tools it is possible to obtain information contained in databases including several thousand items each. Due to the large amount of data being analyzed, it is justifiable to use statistical methods to analyze the real estate market

(Adamczewski, 2011; Bitner et al., 2014; Dydenko, 2020). The choice of a statistical model depends on the type of analysis we want to perform (Waltl, 2016; Nesticò and La Marca, 2020). When valuing real estate property based on a set of comparable properties, linear models appear to be the most appropriate.

The purpose of this paper is to propose an algorithm that can be used in the statistical market analysis method within the comparison approach in mass appraisal (Jahanshiri et al. 2011; Sawilow and Akincza, 2011). The essential element of the algorithm is to solve a system of regression equations. Solving the system involves the use of simple matrix calculus.

In the comparison approach any real estate valuation begins with the analysis of a local market. The purpose of market analysis is to determine the market area, the state of the market, to examine its changes over time, and to determine the appropriate market attributes. Market attributes of property are only those characteristics that differentiate prices in a given market defined and adopted by the appraiser as the basis for valuation (Prystupa, 2014). In the literature, several attributes that significantly affect the variability of real estate prices are analyzed. Maleta and Moscicka (2018) analyze seven attributes of agricultural parcels: area (parcel size), shape, cropland type, soil fertility, distance of paved roads, distance of the municipality centre and distance of the homestead buildings. Rimvydas (2020) uses geodata to select the best real estates for an investment.

The quality of the market analysis conducted has a major influence on the accuracy of the assessed value of property. Even the best statistical model after entering erroneous data will not reflect reality. Therefore, in the further part of the article, general characteristics of the statistical model and the problem of explanatory variables selection are briefly presented. Then, the proposed algorithm for property valuation and its use in the statistical market analysis method within the comparison approach are presented.

2. Principles of good model building

A model is a simplified representation of reality. It is an abstract concept, linking abstract thinking with the real world. A well-constructed model is a compromise between the two extremes (Ostasiewicz, 1998):

– oversimplification;
– excessive accumulation of details.

A statistical model is a system of interdependencies describing a certain process or phenomenon, the parameters of which are estimated on the basis of real data, using statistical methods. The model building process consists of two interrelated steps: determining the general form of the model relationships and the variables present in them, and the statistical estimation of the parameter values based on observed data. A good statistical model is parsimonious, that is, it uses as few mathematical concepts as possible in a given situation. A model should extract what is systematic in the results observed, allowing for the presence of purely random deviations. In a sense, a statistical model decomposes the results of observations into a systematic component that can be described by a specific formula and a random component. Among statistical models, econometric models play

a special role, presenting a system of regression equations linking response variables with explanatory variables (Kukula, 1996).

The choice of a model, the number and the type of attributes (explanatory variables) for a given market is determined by the appraiser. It should be taken into account that each local real estate market has its own specificity and the influence of the same attribute on the prices of the properties may be completely different in different local markets. For this reason, attribute values should be determined based on the analysis of the local market in which the property is being valued. When choosing the form of the model, the focus should be on its simplest formulations. As few explanatory variables as possible should be entered into the model to explain the phenomenon under study. In the literature it is proposed to limit it even to a few variables (Prystupa 2014). When reducing the set of initially selected explanatory variables, one should follow the rule that the model should include variables that are highly correlated with the response variable (that have a significant influence on it) and at the same time very weakly correlated among themselves. Correlated explanatory variables deprive each other of their explanatory power. Failure to follow this rule leads to undesirable econometric modeling results. In this case, we cannot trust the obtained coefficients in the adopted model. If we use too many attributes X_i (e.g., a dozen), some of them may be, in the set of observations, correlated with each other. This correlation is due to the very nature of the attributes adopted. For example, there is an obvious relationship between the intensity of development and the land use designation. Removing one of the correlated variables from the model will improve it significantly, while losing little information about the response variable (property value W). Much of the information about the variable W that one of the variables carries is also contained in the other variable. The ideal situation would be if the attributes were not correlated (no linear relationship between them – collinearity). Then each factor that affects price would convey its share of information about the value of the appraised property – information that no other factor would convey. Unfortunately, this situation does not actually happen. When building a model, we should therefore strive to minimize the undesirable phenomenon of collinearity between variables. One measure of collinearity between two attributes is the correlation coefficient. If we do not eliminate collinear variables from the model, then:

- the values of the obtained model coefficients may be very different from their expected values;
- the signs of the model coefficients are different than expected;
- inclusion or exclusion of one explanatory variable from the model results in great changes in the model coefficient values or changes of their signs;
- removal of certain observations results in great changes in the model coefficient values or changes of their signs.

The following is a proposed property valuation algorithm that can be used to determine property values in large local markets. By “large market” is meant a market with many transactions, and thus market size is measured by the dynamics of transactions rather than by the size of the area. If you want to use this algorithm, you should have a database of at least several dozen properties. This condition can be met when typical

properties are being valued, such as undeveloped land intended for single-family housing or residential premises.

3. Property valuation algorithm

Selecting properties similar to the property we want to valuate. Examining changes in price levels over time in a given local market and possibly adjusting the stated sale prices to account for the effects of changes in the local market. Determining the number of market characteristics (attributes) in a given local market. Determining attribute values of similar properties and the property being appraised.

Let $C^{(1)}, C^{(2)}, \ldots, C^{(n)}$ denote the unit adjusted prices of n similar properties to be included in the analysis.

Assumption: prices $C^{(1)}, C^{(2)}, \ldots, C^{(n)}$ are a linear function of attributes $X_1, \ldots, X_m$.

$$\begin{aligned} C^{(1)} &= w_0^{(1)} + w_1^{(1)} X_1^{(1)} + w_2^{(1)} X_2^{(1)} + \ldots + w_m^{(1)} X_m^{(1)} \\ C^{(2)} &= w_0^{(2)} + w_1^{(2)} X_1^{(2)} + w_2^{(2)} X_2^{(2)} + \ldots + w_m^{(2)} X_m^{(2)} \\ &\ldots \\ C^{(n)} &= w_0^{(n)} + w_1^{(n)} X_1^{(n)} + w_2^{(n)} X_2^{(n)} + \ldots + w_m^{(n)} X_m^{(n)} \end{aligned}$$

– Calculating the unit value of the representative (statistical) property, W_s, and presenting it as a linear function of the arithmetic means of the n values of the individual property attributes accepted for analysis, according to the formula:

$$W_s = a_0 + a_1[X_1] + a_2[X_2] + \ldots + a_m[X_m],$$

where:
a_i – model coefficients calculated according to the algorithm presented in the next point.
The symbol $[\cdot]$ indicates the arithmetic mean.

$$[X_i] = \left(X_i^{(1)} + X_i^{(2)} + \ldots + X_i^{(n)}\right)/n, \qquad i = 1, \ldots, m;$$

$X_i^{(j)}$ – the value of the i-th attribute of the j-th property.

– Determining the a_i coefficients. It consists in solving the linear matrix equation:

$$\mathbf{M} \cdot \mathbf{a} = \mathbf{b},$$

in which the unknown is the matrix $\mathbf{a}$. The individual matrices of the equation have the following form:
$\mathbf{M}$ – the symmetric square matrix of dimension $(m{+}1) \times (m{+}1)$

$$\mathbf{M} = \begin{bmatrix} 1 & [X_1] & \cdots & [X_m] \\ [X_1] & [X_1X_1] & \cdots & [X_1X_m] \\ \cdots & \cdots & & \cdots \\ [X_m] & [X_mX_1] & \cdots & [X_mX_m] \end{bmatrix}$$

a – the *coefficient matrix* a_i of dimension $(m+1) \times 1$

$$\mathbf{a} = \begin{bmatrix} a_0 \\ a_1 \\ \dots \\ a_m \end{bmatrix}$$

b – the matrix of dimension $(m+1) \times 1$

$$\mathbf{b} = \begin{bmatrix} [\mathrm{C}] \\ [\mathrm{CX}_1] \\ \dots \\ [\mathrm{CX}_m] \end{bmatrix}$$

The matrix **a** is the product of the inverse matrix **M** and the matrix **b**.

$$\mathbf{a} = \mathbf{M}^{-1} \cdot \mathbf{b}$$

– Calculating the value of the property W being appraised as the sum of the products of the coefficients a_i and the value of the attributes X_i of the property being appraised.

$$W = a_0 + a_1\mathrm{X}_1 + a_2\mathrm{X}_2 + \ldots + a_m\mathrm{X}_m \,,$$

where:
a_i – the calculated model coefficients,
X_i – the value of the i-th attribute of the property being appraised, $i = 1, \ldots, m$.

The proposed algorithm for determining the a_i coefficients in borderline cases behaves as follows:

– in the case of a fully deterministic linear model (in which there is no influence of random factors on the property value) in the form $W = w_0 + w_1\mathrm{X}_1 + \ldots + w_m\mathrm{X}_m$ the values of a_i coefficients calculated according to the algorithm are equal to the values of respective w_i coefficients, i.e. $a_i = w_i$;
– if the i-th attribute has no effect on the value of the property, the a_i coefficient is equal to 0.

The proposed algorithm is used for large local markets to obtain a sufficiently large base of similar properties accepted for analysis. Using this method requires the analysis of the market over a longer period of time (minimum six months) in order to correctly determine the state of the market. Price adjustment according to the changes of the market during the period under study is an essential correction. The influence of market attributes on property prices is determined for adjusted prices. Therefore, the validity of further analysis depends on the reliability of the determined price change index. In this case, the most appropriate technique, is a linear regression method (Bitner, 2017). The directional coefficient of the linear regression line fitted to the data using the least squares method, is the indicator of price changes during the period under study. If the state of the market has changed significantly during the period under study, this period

of time should be divided into shorter periods and prices should be adjusted in several stages (Bitner, 2015).

The adopted form of the property value model $W = a_0 + a_1X_1 + a_2X_2 + \ldots + a_mX_m$ contains the constant term a_0. This is justified on the grounds that, even in theory, the model should rule out any situation in which the value of the property being appraised is not positive. This condition will be satisfied if attribute scales starting at zero are adopted and the constant term a_0 is entered into the model. The constant term determines the unit value of the hypothetical worst property with the lowest values of all attributes equal to zero. The value of the worst property according to the adopted model is then $W = a_0$.

Each property, even the worst one, has value, therefore the constant term a_0 should have a positive value. Otherwise, the model is incorrect and the assumption is not met. In such situation, it is necessary to verify the database of similar properties once again, to check whether we have included all significant variables (attributes) in the model. It should be emphasized again that the proposed model can be applied only to the valuation of typical properties. Of course, there are properties with a negative market value. Examples of such properties include waste disposal sites and former surface mining areas, which require costly reclamation (Al-Ruzouq and Al Rawashdeh, 2014; Lechner et al., 2016; Huber, 2021). These are special types of properties that require individual valuation and the proposed algorithm cannot be applied to them.

The number of attributes $X_1, \ldots, X_m$ included in the model should not exceed several. Arguments to support this thesis can be found, inter alia, in psychological literature (Miller, 1956; Tomaszewski, 1975) as well as in literature on real estate valuation (Prystupa, 2014). Psychological research has shown that humans cannot properly value traits according to their importance when their number exceeds seven. Thus, the number of attributes that a potential real estate participant considers important is several. In the valuation process in the comparison approach when determining the number and type of market attributes one should take into account the behavior of the real estate market participants.

For $m = 1$ (in the case of one attribute) the formulas for the coefficients a_0 and a_1 are as follows:

$$a_0 = ([X_1][CX_1] - [C][X_1X_1]) / \left([X_1]^2 - [X_1X_1]\right)$$

$$a_1 = ([C][X_1] - [CX_1]) / \left([X_1]^2 - [X_1X_1]\right)$$

The formulas for the coefficients a_i determined according to the algorithm when considering a linear model come down to the formulas for the coefficients in a multiple regression model calculated by the least squares method.

4. The statistical market analysis method

In accordance with Article 4(2) of the Regulation of the Council of Ministers on property valuation and the preparation of a real estate appraisal report (Regulation, 2004), the

method of statistical market analysis is one of three valuation methods to be used within the comparison approach.

Article 4.2. *In the comparison approach, the pairwise comparison method, the average price adjustment method or the statistical market analysis method shall be used.*

In Article 4(2) the method of statistical market analysis was developed in more detail in the following provision:

Article 4.5. *In the method of statistical market analysis the set of transaction prices appropriate for determining the value of representative real estate referred to in Article 161(2) of the Act shall be adopted. The value of real estate shall be determined using methods applied for statistical analyses.*

In accordance with Article 161 of the real estate management act (Act, 1997), which is referred to in Article 4(5) of the ordinance, the values of representative properties are determined in the process of mass appraisal.

Article 161. *1. Mass appraisal shall be aimed at determining the cadastral value of real estate and shall be carried out, in accordance with this Act and separate regulations, by the authorities maintaining the real estate cadastre.*

2. The cadastral value of real estate shall be determined on the basis of assessments of the real estate representative of the different types of real estate on the territory of a municipality. The provisions of chapter 1 of this section shall apply to the determination of the value of these properties. The value of representative real estate shall be determined using transaction prices of real estate on the territory of a municipality, and in case of insufficient number of transactions, on the territory of neighboring municipalities.

The proposed statistical market analysis method is shown below:

1. Identifying the local market in which the property being appraised is located by determining its type, area, and the period of transaction price research.
2. Determining the type and the number of market attributes X_i, $i = 1, \ldots, m$ that differentiate prices in a given local market.
3. Providing the characteristics of the property being appraised, highlighting its market attributes.
4. Selecting for the analysis at least several dozen properties, whose prices are reliable, that are most similar to the property being appraised in terms of market attributes, giving their brief description and general characteristics.
5. Assigning values to the property market attributes on a scale starting at 0.
6. Determining the price change index, which is a measure of changes in the state of the market over time, and possibly adjusting the prices of the properties accepted for analysis.
7. Determining the value W of the representative property for the set n of data accepted for analysis according to the formula:

$$W = a_0 + a_1[X_1] + a_2[X_2] + \ldots + a_m[X_m],$$

where:
a_i – the model coefficients calculated according to the algorithm,
$[X_i]$ – the average of the values of the i-th attribute for the set n of properties,

$[X_i] = (X_i^{(1)} + X_i^{(2)} + \ldots + X_i^{(n)})/n, i = 1, \ldots, m,$
$X_i^{(j)}$ – the value of the i-th attribute of the j-th property.

8. Determining the a_i coefficients according to the algorithm.
9. Calculating the value W of the property being appraised, as the sum of the products of the coefficients a_i (calculated according to the algorithm referred to in sections 7 and 8) and the value of the attributes X_i of the property being appraised, according to the formula:

$$W = a_0 + a_1X_1 + a_2X_2 + \ldots + a_mX_m\,,$$

where:
a_i – the calculated model coefficients,
X_i – the value of the i-th attribute of the property being appraised, $i = 1, \ldots, m$.

5. Concluding remarks

This paper proposes an algorithm that can be used in the statistical market analysis method in the process of mass appraisal. Legal considerations indicating the need for such an algorithm for determination of the value of representative properties are also provided. Because the method makes use of statistical tools, it is necessary to have a sufficiently large database of properties that are subjects of the transaction. Due to the large size of the database of comparable, the proposed algorithm can be used only to appraise typical properties. Only a few market attributes that differentiate property prices on the market under study should be entered into the algorithm. The proposed model includes a constant term, a_0, that, when the model assumptions are met, ensures that the model can be used to estimate the value of properties with extremely low attribute values.

Author contributions

Conceptualization: A.B.; writing the article: A.B.; critical revision of the article: A.B. and M.F.; final approval of the article: A.B. and M.F.

Data availability statement

No datasets were used in this research.

Acknowledgements

This research was funded by the program “Incubator of Innovation 4.0” implemented as part of the non-competitive project entitled “Support for the management of scientific research and commercialization of the results of R&D works in research units and enterprises” under the Intelligent Development Operational Program 2014 – 2020 (Measure 4.4).

References

Act (1997). Act of 21 August 1997, on real estate management. *Journal of Laws of 2021*, item 1899.

Adamczewski, Z. (2011). *Elementy modelowania matematycznego w wycenie nieruchomości: podejście porównawcze*. Warszawa: Oficyna Wydawnicza Politechniki Warszawskiej.

Al-Ruzouq, R., and Al Rawashdeh, S. (2014). Geomatics for rehabilitation of mining area in Mahis, Jordan. *J. Geogr. Inf. Sys.*, 6, 123–134. DOI: 10.4236/jgis.2014.62014.

Bielecka, E., Dukaczewski, D., and Janczar, E. (2018). Spatial data infrastructure in Poland – lessons learnt from so far achievements. *Geod. Cartogr.*, 67(1), 3–20. DOI: 10.24425/118702.

Bitner, A., Nawrocki, P., Litwin U. (2014). Use of methods of cartographic representation and econometrics for defining the correlation between real estate prices and a distance to a city centre. *Geomat Landmanagem. Landscape*, 4, 13–23. DOI: 10.15576/GLL/2014.4.13.

Bitner, A. (2015). Metoda składania trendów dla rynku nieruchomości. *Przegląd Geodezyjny*, 2, 3–5.

Bitner, A. (2017). Wyznaczanie wskaźnika zmian cen nieruchomości z wykorzystaniem modelu regresji. *Wycena nieruchomości i przedsiębiorstw*, 1(1).

Dydenko, J. (2020). *Szacowanie nieruchomości*. Warszawa: Wydawnictwo Wolters Kluwer.

Huber, B.R. (2021). Negative-Value Property. *Washington University Law Review*, 98, 1461–1515.

Jahanshiri, E., Buyong, T., and Shariff A.R.M. (2011). A Review of property mass valuation models. *Pertanika J. Sci. Technol.*, 19(1), 23–30.

Kukula, K. (1996). *Wprowadzenie do ekonometrii w przykładach i zadaniach*. Warszawa: PWN Warszawa.

Lechner, A.M., Kassulke O., and Unger C. (2016). Spatial assessment of open cut coal mining progressive rehabilitation to support the monitoring of rehabilitation liabilities. *Resour. Policy*, 50, 234–243. DOI: 10.1016/j.resourpol.2016.10.009.

Maleta, M., and Moscicka, A. (2018). Selection and significance evaluation of agricultural parcels determinants. *Geod. Cartogr.*, 67(2), 239–253. DOI: 10.24425/gac.2018.125473.

Miller, G.A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.*, 63(2). DOI: 10.1037/h0043158.

Nesticò, A., and La Marca, M. (2020). Urban real estate values and ecosystem disservices: An estimate model based on regression analysis. *Sustain.*, 12, 6304. DOI: 10.3390/su12166304.

Ostasiewicz, W. (1998). *Statystyczne metody analizy danych*. Wrocław: Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.

Prystupa, M. (2014). *Wycena nieruchomości i przedsiębiorstw w podejściu porównawczym*. Poznań: Wydawnictwo Replika.

Regulation (2004). Regulation of the Council of Ministers of 21 September 2004, on property valuation and the preparation of an appraisal report. *Journal of Laws of 2021*, item 555.

Rimvydas, G. (2020). Geodata influencing land market and sustainable development. *Geod. Cartogr.*, 69(1), 3–17. DOI: 10.24425/gac.2020.131078.

Sawilow, E., and Akincza, M., (2011). Zastosowanie teorii modelowania dla potrzeb powszechnej taksacji nieruchomości. *Infrastructure and Ecology of Rural Areas*, 4, 129–140.

Tomaszewski, T. (1975). *Psychologia*. Warszawa: PWN Warszawa.

Waltl, S.R. (2016). Variation across price segments and locations: A comprehensive quantile regression analysis of the Sydney housing market. *Real Estate Econ.*, 47(2), 723–756. DOI: 10.1111/1540-6229.12177.