

DISSIMILARITY AS A COMPONENT OF THE PROPERTY PRICE MODEL

Jacek Zyga

Faculty of Engineering and Architecture

Lublin University of Technology

e-mail: j.zyga@pollub.pl

Abstract

In the course of discussion on an econometric model of property value and its place in property appraisal, the argument of the main goal of the process (property market value prediction itself) was raised in this article. The need for the consideration of an ontologically perceived, particular element of the real estate market with its distinctive characteristics indicates the specific nature of the interpretation of the data which may be used in the appraisal process.

Therefore, a new shape of the property value model, based on LSM, was presented. It takes into account a specific description of the appraised property. Thus, the factor of dissimilarity between sold properties used in creating the value model and the appraised property was used in its coefficient matrix. The new model clearly shows the advantages and disadvantages of the dissimilarities between sold properties used in creating the coefficient matrix of the value model.

Key words: *Dissimilarity, Property price model, comparables selection, property valuation.*

JEL Classification: *R15, C18, C38, C51.*

Citation Zyga J., 2019, *Dissimilarity as a Component of the Property Price Model*, *Real Estate Management and Valuation*, vol. 27, no. 3, pp. 124-132.

DOI: 10.2478/remav-2019-0030

1. Introduction

Real estate (the land and all improvements affixed permanently to it, as well as proper appurtenances associated with it) is, in economy, a crucial part of real property (understood as a set of rights, interests, benefits and encumbrances). Real estate valuation is therefore very important to all participants of the market (buyers and sellers, owners and investors, creditors, supervisors and tax offices).

The question of valuation seems to have been fully discussed, because it has been subject to deep analysis since the early 1970s. The complete reference list is too long to quote in total here, and therefore only a few milestone positions can be mentioned (SHENKEL, EIDSON 1971; SKAFF 1975; TCHIRA 1979; ALBRITTON 1982; SHLAES 1984; ISAKSON 1986; THOMPSON, GORDON 1987; PETO et al. 1996; MCCLUSKEY, BORST 1997; PAGOURTZI et al. 2003; FRENCH 2003; FRENCH 2004; D'AMATO, KAUKO 2017). Nevertheless, new points of view on valuation are still emerging. AVM is not the only solution for the estimation of value. However, "some institutions consider AVM assisted valuation more reliable than valuation in person"; the valuers' community distinguishes between two notions: "estimation of value" and "opinion of value" (D'AMATO, KAUKO 2017), and reconsiders the question of similarity in valuation tasks (MCCLUSKEY, BORST 2017; JANSEN VAN VUUREN 2017).

This paper aims to show some benefits of the identification of similarities and dissimilarities between relevant properties and sold properties considered for the purposes of a market price modelling process.

2. Multiple regression in property price prediction

A typical general multiple linear regression model of the market price (unitary price), based on LSM, involves equations combining unitary price observations y_i ($i \in \langle 1, n \rangle$) for sold properties, observations $a_{i,j}$ ($j \in \langle 1, k \rangle$) of market features of sold properties (constituting evaluation of several attributes which correspond with factors recognized as affecting the local property market) and fixed (but unknown) regression parameters $x_0, x_1 \dots x_k$ representing the factors mentioned above.

$$y_i = x_0 + a_1 \cdot x_1 + \dots + a_k \cdot x_k + \varepsilon_i \quad (1)$$

Variable ε_i is a stochastic component of each equation. It is normally distributed with mean zero and has a variance of σ_ε^2 .

When market unitary prices y_i of sold properties are collected in the vector Y and evaluation marks $a_{i,j}$ (given to those properties) are sorted for each y_i respectively in rows of the matrix A , they can then together create an information system which can be successfully developed, yielding acceptable results. The information system, in the sense of Pawlak's Rough Sets theory (PAWLAK 1983), consists of sets of observations Y , as well as affecting factors X and a set of elements that describe the relationship $a(y,x)$ between Y and X , $a: Y \times X \rightarrow A$, with the range of possible values Va ascribed to a .

$$S = \langle Y, X, Va, a(y,x) \rangle \quad (2)$$

The system built with these equations (1) represents a common general model of price/value in the considered market (3).

$$Y = AX + \varepsilon \quad (3)$$

In order to emphasize the fact that the vector of market unit prices Y , values of market attributes X as well as the vector of stochastic component ε are strongly connected with a part of the information collected in Table A, notation (4) of the considered system is preferred:

$$Y_A = AX_A + \varepsilon_A \quad (4)$$

A solution for X_A is given as the best linear unbiased estimator (5):

$$\widehat{X}_A = (A^T A)^{-1} A^T Y_A \quad (5)$$

$$(Y_A - A(A^T A)^{-1} A^T Y_A)^T (Y_A - A(A^T A)^{-1} A^T Y_A) = \min_{x_0, \dots, x_k} \quad (6)$$

The above-described regression method is the one used most widely to obtain econometric models (PAGOURTZI et al. 2003) which help to track the correlation between defined variables, and to identify the relevant factors functioning on the given market (parameters included in vector X_A). Although prediction is formally seen by statistical theory (Sen, Srivastava 1994) as the principal purpose of regression, the prediction of values of market elements outside the system (Y, X, A) rather seems to be merely an additional benefit in the practice of econometrics as well as of statistical property valuation. This impression may come from the fact that the estimation of the value (predicted price) Y_{VA} of any subsequent property which is, in general, external to the above-described system (4) is mathematically simple. Thus, in the literature on the subject, there are no further divagations regarding it.

The prediction of the value of any "new" property is just a combination (7) of the examined vector \widehat{X}_A and the matrix F_V comprising evaluation marks given to the appraised property (respectively, to parameters covered by X_A):

$$\widehat{Y}_{VA} = F_V \widehat{X}_A \quad (7)$$

By putting equation (6) into (7), we obtain the formula (8). It indicates that, in order to predict the price/value of a property described by a set of characteristics F_V , we use the elements of system (4) taking into account F_V itself.

$$\widehat{Y}_{VA} = F_V (A^T A)^{-1} A^T Y_A \quad (8)$$

This clearly proves that the predicted price/value \widehat{Y}_{VA} is the product of an information system consisting of the vector of price observations Y_A supplemented with the associated matrix A .

To verify how good the developed model (4) is and how well the regression line fits the observed reality, one can use a set of parameters (SEN, SRIVASTAVA 1994; KEENER 2006); for example: the residual

variance S^2 of the model (9) or its square root (standard deviations S), the coefficient of residual variability V (10), a variance/covariance matrix of structural parameters (11) or the determination coefficient R^2 ,

$$S_A^2 = \frac{1}{n-k} (Y_A - A(A^T A)^{-1} A^T Y_A)^T (Y_A - A(A^T A)^{-1} A^T Y_A) \quad (9)$$

$$V = \frac{S_A}{Y_A} \quad (10)$$

$$\text{cov}(\widehat{X}_A) = D^2(\widehat{X}_A) = S_A^2 (A^T A)^{-1} \quad (11)$$

As the *ex ante* verification of the performed prediction, the following are usually used: residual variance S_{VA}^2 of \widehat{Y}_{VA}

$$S_{VA}^2 = S_A^2 (F_V (A^T A)^{-1} F_V^T) + S_A^2 \quad (12)$$

standard deviation S_{VA} or its possible relative prediction error:

$$V_{FP} = \left| \frac{S_{VA}}{\widehat{Y}_{VA}} \right| \quad (13)$$

However, the last and the most important way of verifying each prediction of value is always its juxtaposition with the price eventually created by market participants. In statistics, it is called *ex post* prediction error, meant as the difference Q between the true value of Y_{VR} (actual price) and the predicted value \widehat{Y}_{VA} or its relative form.

$$Q = Y_{VR} - \widehat{Y}_{VA} \quad (14)$$

$$V_{FR} = \left| \frac{Y_{VR} - \widehat{Y}_{VA}}{Y_{VR}} \right| \quad (15)$$

3. Introduction of dissimilarity factor

If one is not only interested in general research on some market phenomena, but wants to obtain a very real answer in a specific case of real estate value, attention is particularly drawn to the end of the procedure described above. Considering the statement that the predicted price/value \widehat{Y}_{VA} is the product of calculation based on the vector of price observations Y_A and matrix A only, one can ask whether the developed information system is really good enough for estimating the value \widehat{Y}_{VA} of a specific property? Then, even if the system $\langle Y, X, Va, a(y,x) \rangle$, represented by observations in the vector Y_A , matrix A and solutions X_A , was tested (as far as the collected data coherence and the optimal choice of predictors, e.g. proposed by (HELLWIG 1968), are concerned), that entire system should be suitable for research focused on the prediction of the price/value of a particular property, as there is no universal model for predicting anything anywhere. Of course, we have to assume that all the records collected in Y_A and A demonstrate a minimum degree of similarity to each other. However, it should be proved that the above-mentioned records are also similar to the specified property.

The proposed way of focusing the research on a certain object (property), but also on the final prediction as the objective of the undertaken analysis, consists in taking into consideration the issue of similarity.

Each description of a sold property, represented as a record of marks a_i , related to the unitary price y_i of that property and confronted with an analogical record of marks given to the appraised property and presented in the pattern F_V , provides an opportunity to recognize the differences between each sold and appraised property.

The found differences allow dissimilarity to be defined as:

$$d_{i,j} = a_{i,j} - f_{V1,j} \quad (16)$$

where:

$a_{i,j}$ – is the i,j element of matrix A

$f_{V1,j}$ – is the j element of the single-line matrix of the F_V pattern

Taking into the consideration the whole set of data, we obtain:

$$D_A = A - [1]F_V \quad (17)$$

where:

D_A - is the matrix of dissimilarity (the differences matrix) between the collected sold properties and the model – the appraised property; dimensions in rows and columns ($n \times k$),

F_V - is the single-line matrix of characteristics describing the appraised property – the pattern for dissimilarity analysis; dimensions in rows \times columns ($1 \times k$),

[1] - is the vector of elements equal to “1”; dimensions in rows \times columns ($n \times 1$).

The content of matrix D_A clearly describes whether the collected sold properties (with their descriptions in matrix A) are really similar to the relevant property to be evaluated. For easier evaluation of all single similarities/dissimilarities collected in matrix D_A , cumulative evaluation is suggested (such as mean cumulative dissimilarity $CDiss$):

$$CDiss = \frac{\sum_{i=1}^n \sum_{j=1}^k (D_{i,j})^2}{n} \quad (18)$$

Using the substitution (17), we can transform Formula (4) into:

$$Y_A = ([1]F_V + D_A) X_A + \varepsilon_A \quad (19)$$

and consequently transform Formula (8) into:

$$\widehat{Y}_{VA} = F_V([1]F_V + D_A)^T([1]F_V + D_A)^{-1}([1]F_V + D_A)^T Y_A \quad (20)$$

Through the equation presented above, a new point of view on data is demonstrated. This point of view, taking into account the issue of similarity between properties (measured in practical terms through dissimilarity D_A between them) is completely ignored in constructing some strictly econometric models of market price and in testing their adequacy. However, overlooking the actual property similarity issue can be accepted in solving intrinsic problems of the market; however when a model of the property price market is to be used for valuation of any piece of real estate, it cannot abstract from the essence and natural peculiarity of each real estate. Equation (20) shows that, after redefining the collected records into two groups: those more and less similar to the model property, matrix F_V has a deeper role in estimating the sought value Y_{VA} . It does not only transform factors (recognized as the ones affecting the local property market - regression parameters $x_0, x_1 \dots x_k$) into price prediction, but it also actively interacts with dissimilarity factors (indicators). The matrix F_V , being the model description for all collected comparable properties, makes it possible to see which comparable is closer (more similar).

Most of the available information concerning the advancements in valuation theory shows that the biggest emphasis is laid on research (OSLAND 2010; BEAMONTE et al. 2013; KRYVOBOKOV, WILHELMSSON 2007; BERACHA, WINTOKI 2013), on the methodology of data analysis (multiple regression is one of the most frequently used), on the quantity of data, on the size of the data set or on the ingenuity of the employed solutions. Almost all researchers declare that they employ the comparability approach, but consistently overlook the problem of similarity or dissimilarity. Obviously, there are also examples in which the application of any statistical methods in real estate valuation is related to the proper data collecting process, which takes into account the principles of similarity (CZAJA 1997; HOZER et al. 1999; BARAŃSKA 2004, 2005, 2010; KLEIBER 2005; KRAUSE, KUMMEROV 2009; BOURASSA et al. 2010; ADAMCZEWSKI 2011; KONTRIMAS, VERIKAS 2011; HROMADA 2016; DEL GIUDICE et al. 2017A; DEL GIUDICE et al. 2017B).

4. Dissimilarity-based property price model

The control of dissimilarity indicators $D_{A,i,j}$ can demonstrate that the set of comparables (described by unitary price observations in the vector Y_A and a set of market feature evaluations collected in several lines of matrix A) is not homogenous and includes properties being more and less similar to the model property. This allows a distinction in matrices Y_A and A to be introduced, such as:

$$Y_A = \begin{bmatrix} Y_H \\ Y_B \end{bmatrix} \quad A = \begin{bmatrix} H \\ B \end{bmatrix} \quad (21)$$

where:

H - is a matrix of factors describing sold items less similar to F_V ,

B - is a matrix of factors describing sold items more similar to F_V .

Taking into consideration the distinctions from (21), one can modify equations (8) and (17) as:

$$\widehat{Y}_{VA} = F_V \left(\begin{bmatrix} H \\ B \end{bmatrix}^T \begin{bmatrix} H \\ B \end{bmatrix} \right)^{-1} \begin{bmatrix} H \\ B \end{bmatrix}^T \begin{bmatrix} Y_H \\ Y_B \end{bmatrix} \quad (22)$$

$$\begin{bmatrix} H \\ B \end{bmatrix} = \begin{bmatrix} 1_H \\ 1_B \end{bmatrix} F_V + \begin{bmatrix} D_H \\ D_B \end{bmatrix} \quad (23)$$

where:

D_H, D_B - are matrices of dissimilarity (in analogy to D_A) constructed with reference to matrices H and B

F_V - is the single-line matrix of characteristics describing the appraised property - the model for dissimilarity analysis; dimensions in rows x columns ($1 \times k$)

$[1_H], [1_B]$ - are vectors of elements equal to "1"; their dimensions in rows x columns are respectively ($n_H \times 1$) and ($n_B \times 1$), where $n_H + n_B = n$.

Matrices D_A (17) and D_H, D_B from (23) and their norms and traces remain in the following relations:

$$\|D_A^T D_A\| > \|D_H^T D_H\| > \|D_B^T D_B\| \quad (24)$$

$$tr(D_A^T D_A) > tr(D_H^T D_H) > tr(D_B^T D_B) \quad (25)$$

And after putting (23) into (22), we obtain:

$$\widehat{Y}_{VA} = F_V \left(\begin{bmatrix} [1_H] \\ [1_B] \end{bmatrix} F_V + \begin{bmatrix} D_H \\ D_B \end{bmatrix} \right)^T \begin{bmatrix} [1_H] \\ [1_B] \end{bmatrix} F_V + \begin{bmatrix} D_H \\ D_B \end{bmatrix} \right)^{-1} \begin{bmatrix} [1_H] \\ [1_B] \end{bmatrix} F_V + \begin{bmatrix} D_H \\ D_B \end{bmatrix} \right)^T \begin{bmatrix} Y_H \\ Y_B \end{bmatrix} \quad (26)$$

or after developing (26):

$$\widehat{Y}_{VA} = F_V (n_H F_V^T F_V + D_H^T ([1_H] F_V) + ([1_H] F_V)^T D_H + D_H^T D_H + n_B F_V^T F_V + D_B^T ([1_B] F_V) + ([1_B] F_V)^T D_B + D_B^T D_B)^{-1} \cdot (([1_H] F_V + D_H)^T Y_H + ([1_B] F_V + D_B)^T Y_B) \quad (27)$$

After dividing the collected records into two groups: more and less similar to the model property, equation (27) shows that the final estimation \widehat{Y}_{VA} is the result of a sort of cooperation of dissimilarity indicators D_H and D_B and the model property matrix F_V . Through an analysis of equation (27), it is easy to demonstrate that dissimilarity is the factor protecting the entire equation (27) (but also its basic shape in (19) as well as in (8)) against singularity. Therefore, it should be admitted that dissimilarity is really the basis of any predictions made using regression methods. However, reducing dissimilarity to some extent can lead to the improvement of the quality of these predictions.

Let us assume that there is no possibility to extract the matrix D_B (there is no item really "more similar" to the model property). In that case, the estimation of the sought value Y_V (of the item described by the matrix of features F_V) must be based on matrices H and Y_H only. Hence, the prediction result for Y_V is:

$$\widehat{Y}_{VH} = F_V (n_H F_V^T F_V + D_H^T ([1_H] F_V) + ([1_H] F_V)^T D_H + D_H^T D_H)^{-1} \cdot ([1_H] F_V + D_H)^T Y_H \quad (28)$$

Alternatively, a simultaneous estimation \widehat{Y}_{VB} of Y_V , performed on the basis of subsystem (Y_B, B) only, can provide another solution:

$$\widehat{Y}_{VB} = F_V (n_B F_V^T F_V + D_B^T ([1_B] F_V) + ([1_B] F_V)^T D_B + D_B^T D_B)^{-1} \cdot ([1_B] F_V + D_B)^T Y_B \quad (29)$$

The two variants of prediction of Y_V presented above provide an opportunity to compare how the quality of similarity (or rather dissimilarity) affects the quality of the prediction of the price of the appraised object.

5. Simulation experiment

To examine the presented theory, certain estimations were performed. In the face of the fact that real data cannot provide an answer regarding the true value of any real estate (in the sense of a true and unbiased reflection of some indicated market factors), only simulation tests were performed.

An artificial (simulated) picture of a local market was created using 104 records of sales described by a set of also artificial attributes, which were marked on a scale from 1 to 5 (collected in the described matrices A, B, H). Also, three simulated model properties were prepared. Due to the known

values of market parameters X (because they had been assumed), all true and unbiased price observations were estimated (according to (1)), as well as the so-called "true values" of model properties. Next, noise was added to the data concerning prices of 104 "sales" Y_A . Errors ε were specified randomly with an assumed margin of standard deviation.

Next, the matrix A , including simulated "features of 'sold properties,'" was transformed (separately in each variant described beneath) into the matrix of dissimilarity indicators D_A (17). Dissimilarity indicators then allowed the matrices D_A as well as A and Y_A (21) to be divided into matrices D_H, D_B, H, B, Y_H, Y_B respectively in each variant. Two-step predictions of the value of each model property were performed. In the first step, prediction was performed with an assumption that the price model includes all the adopted attributes. In the second one, in particular variants, prediction only involved these attributes which were confirmed as significant variables in particular price models based on properly modified H, B, A matrices. Collected data were controlled with such coefficients as: mean cumulative dissimilarity $CDiss$, the norm of dissimilarity matrix $\|D^T D\|$, the trace of dissimilarity matrix $tr(D^T D)$, determination coefficient R^2 . In each case, the predicted value \hat{Y}_V (26-28) was calculated, as well as the coefficients describing the quality of the performed predictions: *ex ante* standard deviations S_V and *ex post* relative true prediction error V_{FR} . All numerical assumptions were presented in Tables 1 and 2. Results were shown in Tables 3, 4 and 5.

Assumptions:

matrix A was established as a group of 104 records,
 randomly selected from the 5^4 feature records (4 features described by marks from 1 to 5)
 A dimensions $104 \times (1+4)$
 X_A dimensions $1 \times (1+4)$
 Assumptive values of $X^T = (100 \ 80 \ 30 \ 20 \ 10)$
 Simulated observations $Y_A: Y_A = AX_A + \varepsilon_A$ include randomly generated errors ε_A
 $E(ER) = 0.000$, $\max(ER) = 0.593$, $\min(ER) = -0.620$, $\text{stdev}(ER) = 0.200$
 $\varepsilon_{Ai} = (1 + ER_i) AX_A$

From matrices Y_A and A , matrices Y_B, B were extracted, as well as Y_H, H , in the subsequent variants respectively; matrices H and B were complemented with columns of [1] respectively.

Prediction variants (various patterns of records containing the evaluations of appraised/sought properties)

Variant 1: matrix $F = (1, 1, 1, 1, 1)$
 Variant 2: matrix $F = (1, 3, 3, 3, 3)$
 Variant 3: matrix $F = (1, 5, 5, 5, 5)$

6. Conclusions

Viewing the contents of Tables 3-5, one can observe that the value prediction for the object whose features were given extreme evaluations (in variants 1 and 3) is better when a set of objects with similar characteristics is used (comparable properties are most similar with respect to each model object). The combination of sets of objects with characteristics different from the model (matrix H) and objects with more similar characteristics (matrix B) improves the quality of prediction, measured by the relative prediction error V_{FR} .

Figure 1 shows the values of the correlation coefficient extracted for the pairs: *ex post* relative true prediction error $V_{FR}(15)$ vs *ex ante* standard deviations S_V , and *ex post* relative true prediction error $V_{FR}(15)$ vs mean cumulative dissimilarity $CDiss$ (18). It can be observed that the first type of correlations, in the first step of estimations as well as in the second one, takes negative values, while the correlations between *ex post* relative true prediction error $V_{FR}(15)$ and mean cumulative dissimilarity factor $CDiss$ (18) are positive. Furthermore, in the second-step estimation (after extracting significant variables), the values of the relevant coefficient were higher each time (in each variant). This shows that the factor of dissimilarity is a better indicator of estimation accuracy than the variance S_V^2 (or standard deviations S_V) of the obtained result.

Variance (or *ex ante* standard deviation S_V) is a good indicator of estimation accuracy only when the appraised object, in terms of its features, is similar to the average values of these features in the set.

The same role (of a good indicator of estimation accuracy) and under the same circumstances can be ascribed to the determination coefficient R^2 (Figure 2). In two other variants, where the appraised objects are significantly different from the main group of comparables, the true accuracy is unfortunately correlated negatively with coefficient R^2 , with its significant increase in the second step of estimations (after the correction of the variable set). This negative correlation can constitute a positive observation after all. It shows that, in these boundary conditions, a better (higher) coefficient R^2 goes with better (smaller) true accuracy.

The conclusions presented above are confirmed by parallel research ZYGA (2016), as well as ZYGA (2019). The presented research confirms that the deliberate selection of comparable sold properties helps to achieve a better (more accurate) final valuation result (among other choices, offered by possibilities of different, alternative sets of data). It is important to observe that the described proposal of data selection is based on the awareness of dissimilarities between properties, and it is an attempt at using this information in order to come closer to the essence of the value of a specific and real piece of real estate, without the need of modelling the entire local market. It is not "the selection made accordingly to pre-empt outcome of a valuation," as was stated in (KAHNEMAN 2011). It is just proper selection "from computation and not intuition," as it was written in (JANSEN VAN VUUREN 2017).

We must also keep in mind that the search for comparable properties, those most similar to the relevant property, must be properly restrained. An unlimited selection process leads to a drastic reduction of Y_B price sets. This reduction affects calculations, for example, by causing singularity of the model covariance matrix, or by deteriorating the parameters of the formal evaluation of the model.

Unfortunately, there is no table showing the effects of the decomposition of matrix A , H , B according to formulas (26-28). This kind of demonstration would be too large to present here. The decomposition itself does not affect the final results of several experiments. However, observing parts of decomposed matrices of "features of 'sold properties'" and their numbers game is very illuminating; even a thorough analysis of the arrangement of the component matrices in each of the formulas (26-28) can explain how important the issue of dissimilarity is. On the one hand, there is a need to find market items as similar to the appraised property (model property) as possible. This forces the researcher to choose only "the best" part of market information. As it was demonstrated above, this can really help to obtain "the best" prediction of value (however, it can be demonstrated in simulation experiments only). On the other hand, the selection of data, even well-motivated by sets of market features (and their evaluations), can lead to the formation of unsolvable systems, due to the singularity of the covariance matrix from (28)

$$(A^T A)^{-1} = \begin{pmatrix} n_H F_V^T F_V + D_H^T ([1_H] F_V) + ([1_H] F_V)^T D_H + D_H^T D_H & \\ n_B F_V^T F_V + D_B^T ([1_B] F_V) + ([1_B] F_V)^T D_B + D_B^T D_B \end{pmatrix}^{-1} \quad (30)$$

As it can be noticed, the components of $(A^T A)^{-1}$, such as $n_H F_V^T F_V$ or $n_B F_V^T F_V$, are singular by definition.

A selection of comparable properties identical to the model property generates matrices D_H and D_B with elements equal or almost equal to zero; hence the singularity of the covariance matrix of system (3). This last issue brings about a new question: is there a limit in recognizing dissimilarity which will still allow a value prediction process with the use of the regression method to be successfully performed?

7. Bibliography:

- ADAMCZEWSKI Z., 2011, *Elementy modelowania matematycznego w wycenie nieruchomości Podejście porównawcze (The elements of mathematical modeling in property appraisal Comparative approach)*, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa.
- AHMAD A., DEY L., 2007, *A method to compute distance between two categorical values of same attribute in unsupervised learning for categorical data set. Pattern Recognition Letters*, 28:1, pp. 110-118.
- ALBRITTON H., 1982, *Controversies in real property valuation: a commentary*, American Institute of Real Estate Appraisers, Chicago.
- BARAŃSKA A., 2004, *Wybór cech nieruchomości do modelowania matematycznego wartości rynkowej na przykładzie kilku baz nieruchomości gruntowych, (Selection of real estate features for mathematical modelling of the market value exemplified by several databases of ground properties)*, *Geodezja*, v. 10. nr 1. pp. 31-38

- BARAŃSKA A., 2005, *Estymacja parametrów nieliniowych modeli funkcyjnych dla potrzeb predykcji rynkowej wartości nieruchomości*, Geodezja, t. 11. z. 2, Kraków.
- BARAŃSKA A., 2010, *Modele multiplikatywne w procesie wyceny*, Studia i materiały Towarzystwa Naukowe Nieruchomości, Vol. 18, no 1.
- BERACHA E., BABAJIDE WINTOKI M., 2013, *Forecasting Residential Real Estate Price Changes from Online Search Activity*, Journal of Real Estate Research, Vol. 35, No.3, pp. 283-312
- BORST R. A., 2014, *Improving mass appraisal valuation models using spatio-temporal methods*, Toronto. Ontario, Canada, International Property Tax Institute.
- BORST R. A., MCCLUSKEY W. J., 2007, *Comparative evaluation of the comparable sales method with geostatistical valuation models*, Pacific Rim Property Research Journal, 13:1, pp. 106-129.
- BOURASSA S.C., CANTONI E., HOESLI M., 2010, *Predicting House Prices with Spatial Dependence: A Comparison of Alternative Methods*, JRER, Vol. 32, No. 2, pp. 139-159.
- CZAJA J., 1997, *Modele statystyczne w informacji o terenie*, Wydawnictwa AGH, Kraków.
- D'AMATO M., KAUKO T., 2017, *Advances in Automated valuation methodology*, Springer International Publishing AG, DOI 10.1007/978-3-319-49746-4
- DEL GIUDICE V., MANGANELLI B., DE PAOLA P., 2017A, *Hedonic Analysis of Housing Sales Prices with Semiparametric Methods*, International Journal of Agricultural and Environmental Information Systems, Vol. 8, Issue 2, DOI: 10.4018/IJAEIS.2017040105.
- DEL GIUDICE V., DE PAOLA P., FORTE F., 2017B, *Using Genetic Algorithms for Real Estate Appraisals*, Buildings 7(2), DOI: 10.3390/buildings7020031.
- PAGOURTZI E., ASSIMAKOPOULOS V., HATZICHRISTOS T., FRENCH N., 2003, *Real estate appraisal: a review of valuation methods*, Journal of Property Investment Finance, 21:4, pp.383-40, <https://doi.org/10.1108/14635780310483656>.
- FRENCH N., 2003, *The RICS valuation and appraisal standards*. Journal of Property Investment and Finance, 21:6, pp. 495-501, <https://www.emeraldinsight.com/doi/full/10.1108/14635780310508649>
- FRENCH N., 2004, *The valuation of specialised property: A review of valuation methods*, Journal of Property Investment Finance, 22:6, pp. 533-541, <https://doi.org/10.1108/14635780410569506>.
- HELLWIG Z., 1968, *On the optimal choice of predictors*, Study VI, Paris, UNESCO.
- HOZER J., ED. FORYŚ I., ZWOLANKOWSKA M., KOKOT S., KUŹMIŃSKI W., 1999, *Ekonometryczny algorytm masowej wyceny nieruchomości gruntowych*, Uniwersytet Szczeciński, Stowarzyszenie "Pomoc i Rozwój", Szczecin.
- HROMADA E., 2016, *Real estate valuation using data mining software*, Procedia Engineering, 164, pp. 284-291.
- ISAKSON H., 1986, *The nearest neighbors appraisal technique: An alternative to the adjustment grid methods*, Journal of the American Real Estate and Urban Economics Association, 14, pp. 274-286.
- JANSEN VAN VUUREN D., 2017, *Modified sales comparison method: valuing under uncertainty*, Journal of Property Investment Finance, 35:1, pp.101-110.
- KAHNEMAN D., 2011, *Thinking. Fast and Slow*, Penguin Books, London.
- KEENER R. W., 2006, *Statistical Theory: Notes for a Course in Theoretical Statistics*, Springer.
- KLEIBER W., 2005, *Immobilienbewertung in der Bundesrepublik Deutschland*. in: *Immobilienmärkte und Immobilienbewertung*, München, pp. 173-207.
- KONTRIMAS V., VERIKAS A., 2011, *The mass appraisal of the real estate by computational intelligence*, Applied Soft Computing 11, pp 443-448, <https://ac.els-cdn.com>. accessed 2018-01-23.
- KRAUSE A., KUMMEROW M., 2009, *An iterative approach to minimizing valuation Errors using an automated comparables sales model*, Journal of Property Tax Assessment Administration, 82, pp. 39-52.
- KRYVOBOKOV M., WILHELMSSON M., 2007, *Analysing Location Attributes With A Hedonic Model For Apartment Prices In Donetsk*, International Journal of Strategic Property Management, no 11, pp. 157-178.
- MCCLUSKEY W.J., BORST R.A., 1997, *An evaluation of MRA. comparables sale analysis and ANNs for the mass appraisal of resident properties in Northern Ireland*, Assesment Journal, Vol, 4 No. 1. pp. 47-55.
- MCCLUSKEY W.J., BORST R.A., 2017, *The Theory and Practice of Comparable Selection in Real Estate Valuation* D'AMATO M. AND KAUKO T. ED. *Advances in Automated valuation methodology*, Studies in Systems, Decision and Control, Springer International Publishing, pp. 307-330, DOI 10.1007/978-3-319-49746-4

- OSLAND L., 2010, *An Application of Spatial Econometrics in Relation to Hedonic House Price Modeling*, JRER, Vol. 32, No.3, pp. 289-320.
- PAWLAK Z., 1983, *Systemy informacyjne. Podstawy teoretyczne. (Information systems. Theoretical basis)*, WNT, Warszawa.
- PETO R., FRENCH N., BOWMAN G., 1996, *Price and worth: developments in valuation methodology*, Journal of Property Valuation and Investment, 14:4, pp. 79-100.
- SEN A., SRIVASTAVA M., 1994, *Regression Analysis Theory Methods and Applications*, Springer-Verlag.
- SHENKEL W.M., EIDSON A.S., 1971, *Comparable sales retrieval systems*, Appraisal Journal, pp. 540-555.
- SHLAES J., 1984, *The Market in Market Value*, Appraisal Journal, 52, pp.494-518.
- SKAFF M.S., 1975, *The search for comparable sales: A new approach*, Assessors Journal, pp. 7-16.
- TCHIRA A., 1979, *Comparable sales selection – A computer approach*, The Appraisal Journal, pp. 86- 98.
- THOMPSON J.F., GORDON J.F., 1987, *Constrained regression modeling and the multiple regression analysis - comparable sales approach*, Property Tax Journal, 64, 251-262.
- ZYGA J., 2016, *Connection Between Similarity and Estimation Results of Property Values Obtained by Statistical Methods*, Real Estate Management and Valuation, Vol. 24, No. 3, pp. 5-15. DOI:10.1515/remav-2016-0017.
- ZYGA J., 2019, *Data Selection As The Basis For Better Value Modeling*, Real Estate Management and Valuation, not published, accepted 2018-08-26.