ORIGINAL ARTICLE

The influence analysis of criteria of comparables' selection on the accuracy of the property value estimation

Jacek Zyga [iD] [1*]

[1]Department of Geotechnical Engineering, Lublin University of Technology, 38D Nadbystrzycka Street, 20-618, Lublin, Poland

*j.zyga@pollub.pl

Abstract

The subject of the considerations presented in the article is the question of the criteria according to which the comparables, in the market value estimating process, should be selected. As the most important in the selection of comparative properties, the factor of similarity in relation to the subject property (measured by the Euclidean distance) was considered. As the key issue, the similarity assessment criterion and the influence of the adopted critical value of this criterion on the accuracy of estimates were chosen. The analysis of the above was carried out with taking into account the role of the significance of independent variables (measured by their correlation with the vector of the dependent variable). The results of the simulation tests carried out in the variants set by the criteria adopted were presented. On this basis, it has been shown that there is a potential most relevant solution in the collection of obtainable estimation results. This solution corresponds with the smallest sum of the differences of model prices (accepted as known) and corresponding estimates. The found minimum occurs for a specific layout of the above criterion values only.

Key words: Similarity assessment criterion, property valuation, LS estimation

1 Introduction

The problem of determining the regression model parameters is already resolved and described in numerous literary positions: from basic, purely theoretical (Rao et al., 2008; Sen and Srivastava, 1994) through application descriptions (e.g., Vecchia (1988); Weisberg (1980)) up to summarising positions (Glumac and Des Rosiers, 2018; Manly and Alberto, 2016; McCluskey and Borst, 2017). Against the background of this literature, the question of applying a regression analysis to the real property valuation is also seen as a well-explored field. Because of the big number of the literature reference items, only a few of them are listed: Albritton (1982); d'Amato and Kauko (2017); Bruce and Sundell (1977); French (2003, 2004); Isakson (1986); McCluskey and Borst (1997, 2017); Pagourtzi et al. (2003); Peto et al. (1996); Schlaes (1984); Shenkel and Eidson (1971); Skaff (1975); Tchira (1979); Thompson and Gordon (1987). Despite the extensive writing base in this area, there is very little attention paid to the question of the sold properties similarity. In addition to the very local voices of valuation practitioners, only in a few and rather old scientific publications, for example, Shenkel and Eidson (1971); Skaff (1975); Tchira (1979) or Czaja (1997), there were comments on the comparative objects' selection based on their mutual likeness. At the heart of these concepts was the collection of objects perhaps few, but in a certain way similar to the valued property. Today's trends in the development of real estate valuation methodologies indicate the advantage of the method of developing data over the original information and its usefulness for specific valuations. The issue of similarity is the subject of less and less attention. In some areas of the economy (mainly banking, taxes), econometric techniques as overarching techni-

ques are recommended. This is contrary to the regulations of many professional associations (International Valuation Standards Council (IVSC), 2010; Royal Institution of Chartered Surveyors (RICS), 2017; Standards' Commision of Polish Federation of Valuer's Associations (SCPFVA), 2009) but also to direct legal directives (Bundesministerium für Umwelt Naturschutz Bau und Reaktorsicherheit (BMUB), 2006; ImmoWertV, 2009; Act, 1997; Regulation, 2004). Aims of this paper are: to direct attention again to the role of similarity seen as a factor, that can define the scope of real estate (or land) property local market and to demonstrate relationship between a similarity degree of collected comparables and undertaken estimations accuracy.

2 Property value model

The linear price model consists of several structural parameters x_j and corresponding observations of sale subject features a_{ij}. Combinations of those components juxtaposed with the sale prices y_i create a price model equation (1).

$$y_i = a_{i0} + a_{i1}x_1 + \ldots + a_{ij}x_{ij} + \ldots + a_{ik}x_{ik} + \varepsilon_i, \quad (1)$$

where:
$i = 1, 2, \ldots, n\ i \in N$, where n is a number of observations y_A,
$j = 1, 2, \ldots, k\ j \in N$, where k is a number of structural parameters x_A,
$a_{i0} = 1$.

Same number of such equations can be expressed as the system in matrix form (2). The solution for the system (2) of observations $\mathbf{Y_A}$ and $\mathbf{A}$ is given (according to the Gauss–Markov Theorem) as the best linear unbiased estimator of structural parameters $\widehat{X_A}$ (3) with minimum condition (4):

$$\mathbf{Y_A} = \mathbf{AX_A} + \varepsilon_A, \quad (2)$$

$$\widehat{X_A} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Y_A}, \quad (3)$$

$$\left(\mathbf{Y_A} - \mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{Y_A}\right)^T \left(\mathbf{Y_A} - \mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{Y_A}\right) = \min_{x_0\ldots x_k}. \quad (4)$$

To emphasize the relationship between specific $\mathbf{Y}$ and $\mathbf{X}$ vectors, expressed in the dimensions of matrix $\mathbf{A}$: n i k, complementary markings in indexes were added in the following equations. This issue is important when, by selecting elements y_i and the consistent need of elimination of some independent variables, a_{ij} too poorly correlated with the vector y_i, the shape of matrix $\mathbf{A}$ itself also changes.

Elements ε_{A_i}, collected in vector ε_A, are stochastic components of each i-th equation.

On the field of property valuation the variable $\mathbf{Y_A}$, can be understood as the vector of sold property prices y_{A_i} (usually recalculated into proper surface unit prices), while the matrix $\mathbf{A}$ is the set of market feature evaluations, recorded while the survey of comparable properties.

The vector of price observations $\mathbf{Y_A}$, market feature evaluations set $\mathbf{A}$ with a set of listed attributes Ψ_A, create together an information system (5) defined by Pawlak (1981, 1983) as:

$$\langle \mathbf{Y_A},\ \Psi_A,\ Va,\ a(y,x)\rangle, \quad (5)$$

where:
$\mathbf{Y_A}$ is the set of price observations,
Ψ_A is the list of attributes (market features), in (2) each element of Ψ_A has its own value shown in $\mathbf{X_A}$,
Va is the set of all possible values of attributes (market features),
$a(y,x)$ are values of attributes, relationship between $\mathbf{Y_A}$ and Ψ_A,

such $\mathbf{Y_A} \times \Psi_A \to \mathbf{A}$.

In this paper, for short, it will be presented as:

$$\langle \mathbf{Y_A},\ \Psi_A,\ \mathbf{A}\rangle \quad (6)$$

The system (5) also represents a common price/value model (2) used to obtain econometric models of the property value (Pagourtzi et al., 2003). In a practice of econometrics, the main goal of the use of regression models is to find the model itself or to confirm the significance of its several listed parameters. In an appraisal practice (based on the econometrics anyway), the prediction of values of some properties is the most important.

The prediction of the value $\widehat{Y_{VA}}$ of the property from the outside of the collected set of comparables ($\mathbf{Y_A}$, $\mathbf{A}$), but based on it, needs to define the vector $\mathbf{F_V}$ comprising evaluation marks given to the appraised property (respectively, to the list of parameters covered by $\widehat{X_A}$):

$$\widehat{Y_{VA}} = \mathbf{F_V}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Y_A}. \quad (7)$$

As well as the value prediction itself, an accuracy of the obtained estimation results and being sure of appropriation of them are essential in appraising activity.

Statistical techniques of LSM gives some solutions regarding how to test the obtained model. The useful parameters are: the residual variance S^2 of the model, presented in equation (8) as S_A^2 in order to emphasise it's reference to matrix $\mathbf{A}$, the coefficient of variation V (9), a variance/covariance matrix of structural parameters (10) or the determination coefficient R^2 (11),

$$S_A^2 = \frac{1}{n-k-1}(\mathbf{Y_A} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Y_A})^T(\mathbf{Y_A} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Y_A}) \quad (8)$$

$$V = \frac{S_A}{\overline{Y}_A}, \quad (9)$$

$$\text{cov}\left(\widehat{X_A}\right) = D^2(\widehat{X_A}) = S_A^2(\mathbf{A}^T\mathbf{A})^{-1}, \quad (10)$$

$$R^2 = 1 - \frac{\mathbf{Y_A}^T\mathbf{Y_A} - \mathbf{Y_A}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Y_A}}{\mathbf{Y_A}^T\mathbf{Y_A} - n\overline{Y}_A}, \quad (11)$$

where:
n is number of observations y_A and rows number of $\mathbf{Y_A}$,
k is number of structural parameters x_A and $k+1$ is a columns number of $\mathbf{A}$,
$\overline{Y}_A$ is mean sale price.

Above indicators are not enough for the valuation purpose, because they give information about the statistical evaluation of a used model only, without any judgement, if this model is good enough for the valuation of the exact object (even if the object is within the scope of the limitations of the considered local market). This question can be solved with indicators connected with the valuation object. For example, with *ex ante* indicators: like the residual variance S_{VA}^2 of $\widehat{Y_{VA}}$:

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

or with the relative prediction error:

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

But the strongest verification of each value prediction is the *ex post* juxtaposition of the obtained result with the independent price eventually created by the market. Therefore, 'the accuracy', mentioned in the title, will be understood in further considerations as an *ex post* prediction error Q, defined as

the difference between the 'true' or 'real' value Y_{VR} and the predicted value $\widehat{Y_{VA}}$ (14) or as its relative form V_{FR} (15):

$$Q = Y_{VR} - \widehat{Y_{VA}}, \qquad (14)$$

$$V_{FR} = \left| \frac{Y_{VR} - \widehat{Y_{VA}}}{Y_{VR}} \right|. \qquad (15)$$

3 The research problem outline

The goal of much statistical modelling is to investigate the relationship between a criterion (dependent) variable and a set predictor (independent) variables. But for a number of studies related to the modelling of property prices on the selected local market, the primary objective remains to predict the price for the next element of the market. The more properties defining the local market are similar to the valued element (the appraisal subject), the more undertaken prediction is convincing. This evaluation might be weak from a statistical point of view. But when the valued element is close to the nearest sold properties (via their feature evaluations), their sales prices are the best data to predict anything among them.

Therefore, in this type of real estate modelling applications, it is important to emphasize the mutual similarity factor. Its presence in the price model was proved in Zyga (2016, 2019). On the other hand, this study focuses on the importance of the similarity criterion for the price model construction. The valued object was considered a benchmark for assessing this similarity.

To make the problem easier to describe, a dissimilarity factor is used in the next steps. The difference (dissimilarity) $d_{i,j}$,

between the $a_{i,j}$ element of the set $\mathbf{A}$ and the proper $f_{V_{1,j}}$ element of set F_V describing the subject of the valuation is defined as:

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

where:
$a_{i,j}$ is the i, j element of matrix $\mathbf{A}$,
$i = 1, 2, \ldots, n, i \in \mathbb{N}$, where n is a number of observations y_A,
$j = 0, 1, 2 \ldots, k, j \in \mathbb{N}$, where k is a number of structural parameters x_A,
$a_{i0} = 1$,
$f_{V_{1,j}}$ is the j element of the single-line matrix of the $\mathbf{F_V}$ pattern.

For the whole set of data, we obtain respectively:

$$\mathbf{D_A} = \mathbf{A} - [\mathbf{1}]\mathbf{F_V}, \qquad (17)$$

where:
$\mathbf{D_A}$ is the matrix of dissimilarity (the differences matrix) between the sold properties with evaluations collected in matrix $\mathbf{A}$ and the appraised property; dimensions in rows and columns $(n \times k)$;
$\mathbf{F_V}$ is the vector comprising evaluation marks given to the appraised property; it is the single-line matrix with dimensions $(1 \times k)$;
$[\mathbf{1}]$ is the vector of elements equal to 1; dimensions in rows x columns $(n \times 1)$.

The small example of $\mathbf{D_A}$ matrix creating is shown in equation (18) (the juxtaposition of the assumed matrix $\mathbf{A}$, the multiplicated pattern single-line matrix $\mathbf{F_V}$ ($\mathbf{F_V} = [4 \ 1 \ 5 \ 1]$), and the dissimilarity matrix $\mathbf{D_A}$). Matrices are set in the same order as in (17).

dissimilarity matrix $\mathbf{D_A}$ / assumed matrix $\mathbf{A}$ / multiplicated $\mathbf{F_V}$ pattern

1	2	3	4		1	2	3	4		1	2	3	4	
...	...	...	...		...	...	...	...		...	...	...	...	...
4	−1	4	1		1	5	1	6		4	1	5	1	6
−2	0	−4	0		2	1	1	1		4	1	5	1	7
−2	0	0	4		2	1	5	5		4	1	5	1	8
−2	4	0	0	=	2	5	5	1		4	1	5	1	9
0	0	0	0		4	1	5	1		4	1	5	1	10
0	4	−4	4		4	5	1	5		4	1	5	1	11
1	0	−4	3		5	1	1	4		4	1	5	1	12
1	0	−3	4		5	1	2	5		4	1	5	1	13
...	...	...	...		...	...	...	...		...	...	...	...	...

$$(18)$$

Putting (17) in (7) gives us a solution for $\widehat{Y_{VA}}$ revealing an involvement of dissimilarity factor on predicting process (Zyga, 2019):

$$\widehat{Y_{VA}} = \mathbf{F_V} \left[([\mathbf{1}]\mathbf{F_V} + \mathbf{D_A})^T ([\mathbf{1}]\mathbf{F_V} + \mathbf{D_A}) \right]^{-1} ([\mathbf{1}]\mathbf{F_V} + \mathbf{D_A})^T \mathbf{Y_A}, \qquad (19)$$

The equation (19) can be easily modified back to the equation (7) because modification (19) does not bias the solution in any way. But it shows how the dissimilarity factor works within this structure and proves that the connections between the dissimilarity and the final effect of estimation in the LS method really exist. The matrix $\mathbf{D_A}$ gives the possibility of easy selection from the whole collected set of comparables $(\mathbf{Y_A}, \mathbf{A})$, such as sales prices y_{A_i} with proper subsets of market feature evaluations a_{ij} (recorded in proper rows of matrix $\mathbf{A}$) that their dissimilarity indicators (16) are the lowest. This step creates a smaller set of selected comparables $(\mathbf{Y_A}, \mathbf{A})$. By reducing the row number of matrix $\mathbf{A}$, the variability of variables represented by individual columns of $\mathbf{A}$ is inevitably limited. This, in turn, forces the rejection of these variables from $\mathbf{X_A}$ as too

poorly correlated with the new vector $\mathbf{Y_A}$. This reduction is the last step in modifying of selected comparables and gives finally new shape of data set still named $(\mathbf{Y_A}, \mathbf{A})$ but with reduced n as a number of observations and new k as a number of properly significant variables.

The reduction of initially collected set of comparables $(\mathbf{Y_A}, \mathbf{A})$ must be supervised under assumed criterion indicators. Dissimilarity $d_{i,j}$ of j-th component of i-th comparable object description and the proper elements of benchmark evaluation vector $\mathbf{F_V}$ can be accepted when

$$d_{i,j} \leq dmax_j, \qquad (20)$$

$$d_{max_j} = K_{max}(\max(a_{ij}) - \min(a_{ij})), \qquad (21)$$

where:
$i = 1, 2, \ldots, n_{initial}, i \in \mathbb{N}$, where $n_{initial}$ is a number of initially collected observations y_A,
$j = 0, 1, 2, \ldots, k_{initial}, j \in \mathbb{N}$, where $k_{initial}$ is a number of initially defined structural parameters x_A,
$K_{max} \in (0, 1\rangle$ and is a number assumed by the tester for sub-

sequent tests.

To calculate d_{max_j}, a certain criterion value K_{max} assumed by the tester for each tests is needed. It can take values from 0 to 1 and creates the upper limit of accepted dissimilarities $d_{i,j}$.

In a similar way, a correlation between modified vectors $\mathbf{A}^{\langle j \rangle}$ and modified $\mathbf{Y_A}$ must be controlled afterwards. For this purpose, a next criterion indicator is used. This indicator $K_{corrmin}$ is a declared minimum limit of accepted correlation r_{Y_A, A_j}, so:

$$\left| r_{Y_A, A_j} \right| = \left| \frac{\text{cov}\left(\mathbf{Y_A}, \mathbf{A}^{\langle j \rangle}\right)}{\sigma_{Y_A}\, \sigma_{A^{\langle j \rangle}}} \right| \geq K_{corrmin}, \qquad (22)$$

where:
$\text{cov}\left(\mathbf{Y_A}, \mathbf{A}^{\langle j \rangle}\right)$ is a covariance between indicated variables,
σ_- is standard deviation operator of selected variable,
$K_{corrmin} \in (0, 1)$ and is a number assumed by the tester for subsequent tests.

Technically, $K_{corrmin}$ can take values greater than 0 up to 1, but for practical use, it should be greater than 0.2 or even 0.3. On other side, it should not be greater than 0.5. Under real estate circumstances, the forcing of solutions based on the strong correlations ($K_{corrmin}$ greater than 0.5) can create empty sets of the reduced matrix $\mathbf{A}$ with no real answer about significance of any predictor variables. In the tests carried out and described below, $K_{corrmin}$ values were assumed in the range from 0.05 up to 0.6 in order to test the margin results also.

4   Simulation experiment

The effect of the new parameter, which is K_{max} criterion, on the selection of the set of initially collected records ($\mathbf{Y_A}$, $\mathbf{A}$) and on the accuracy of estimates made with the use of such (limited) subsets, have been tested in the experiments described below. In several experiments, reduced sets of comparable properties, collected separately for each final appraisal, included properties described by the condition (20) with respect to (21).

The accuracy of the obtained estimations was evaluated by the difference Q (14) between the response value $\widehat{Y_{VA}}$ and it's the reference ('true') value $\mathbf{Y_{VR}}$, that was modelled directly on the base of model parameters' values X_A, assumed on the start of simulation:

$$\mathbf{Y_{VR}} = \mathbf{AX_A}. \qquad (23)$$

To illustrate a possible variability of the price estimator results on the performed property market, the artificial (simulated) picture of some local market was taken into account. Simulated market (it was demonstrated earlier in (Zyga, 2019)) was represented by 104 real properties randomly drawn from $625 = 5^4$ possible records (because of 4 features described by marks from 1 to 5). Each simulated object had its own set of feature marks (without any detailed description). Although real estate attributes are usually qualitative variables (measured on ordinal scale), in the experiment, simulated marks can be understood as measurements of several attributes made of interval scale.

Each simulated property was priced as $\mathbf{Y_A} = \mathbf{AX_A}$ and biased with ε_A, accordingly to (2). Errors ε_A were specified randomly with an assumed margin of standard deviation: $E(ER) = 0.000$, $\max(ER) = 0.593$, $\min(ER) = -0.620$, $\text{std}(ER) = 0.200$, $\varepsilon_{A_i} = (1 + ER_i)\mathbf{AX_A}$.

True (modelled) values of the above properties were calculated simultaneously as (23).

With this data, each from 104 simulated properties was appraised, but not once. The performed market was analysed

each time, whether it was similar enough for each property that was to be appraised. Each appraisal was performed several times, with each assumed values of K_{max} and $K_{corrmin}$. At least 16014 single experiments were performed.

K_{max} was a criterion factor affecting the accepted maximum dissimilarity represented by the distance d_{max_j} (21) between the compared properties (each comparable and appraised one). In each appraisal process, K_{max} limited a selection set of accepted (enough similar) comparables taken to the next steps of the process, accordingly to the rule (20). K_{max} varied from 0.25 to 1.00. For technical reasons and the expected number of tests, the K_{max}-values were assumed at the fixed interval of 0.05. Afterwards, the results themselves have been arranged in three ranges of values, that one can see on Figure 1. This provides an indirect evidence that, with limited variability in the value of the a_{ij}, there was no need to take a smaller jump of the tested K_{max}. Values $K_{max} \in \langle 0.25, 0.45 \rangle$ created a sharp criterion of dissimilarity, that forced the selection of very similar properties only (almost the same as an appraised one). On other end of the scope, values $K_{max} \in \langle 0.75, 0.95 \rangle$ let to collect for calculations almost all properties from the prepared set. The middle interval with $K_{max} \in \langle 0.50, 0.70 \rangle$ let to take into account typically similar properties. The edge of the scope were $K_{max} = 1$ gave the opportunity to take into account all prepared set of 104 'sold' properties. The only cleaning of the start set that was made in this case was the rejection of the subjects with outstanding prices.

Next, the selection of significant variables, in each case of calculation, was performed. Each case of K_{max} included a separate calculation with $K_{corrmin}$ parameter varying from 0.05 to 0.6. $K_{corrmin}$ was a criterion factor affecting the accepted minimum significance. $K_{corrmin} \in \langle 0.05, 0.25 \rangle$ gave a very weak criterion for the rejection of variables (a case formally unacceptable from statistical point of view) and allowed the algorithm to accept all or almost all initially prepared variables. A weak correlation between vectors $\mathbf{A}^{\langle j \rangle}$ and $\mathbf{Y_A}$ ($K_{corrmin} \in \langle 0.30, 0.35t \rangle$) sometimes caused the rejections of some variables as not significant, while in case of $K_{corrmin} \in \langle 0.40, 0.60 \rangle$, most of variable were rejected. Therefore, calculations with $K_{corrmin} > 0.60$ (most wanted from a technological point of view) gave no advantages in the performed research.

Due to the huge number of single calculations, as well as final results, it was decided to show them in the aggregated way. Each single experiment gave as a result the estimation $\widehat{Y_{VA}}$ made with the use of (7) with corresponding reference ('true') value $\mathbf{Y_{VR}}$ (23). This led to the calculation of the difference $Q = \mathbf{Y_{VR}} - \widehat{Y_{VA}}$ (14), that shows how accurate the result $\widehat{Y_{VA}}$ is. Because it was difficult to show all 16014 results, they were summarised in the matrix with reference to values of K_{max} and $K_{corrmin}$. For easier understanding, each collected sum was performed as the quotient with reference to minimum value from the collected sums (24):

$$\frac{\sum\limits_{K_{max}=0.25}^{1} \sum\limits_{K_{corrmin}=0.05}^{0.6} \left(Q_{K_{max}, K_{corrmin}}\right)}{\min\left[\sum\limits_{K_{max}=0.25}^{1} \sum\limits_{K_{corrmin}=0.05}^{0.6} \left(Q_{K_{max}, K_{corrmin}}\right)\right]} \qquad (24)$$

The matrix of such proportional and aggregated results is shown in Table 1 and Figure 1 respectively.

5   Experiment results and conclusions

It was shown that nominally, the best aggregated results of accuracy (proportional cumulated accuracy factor equals 1) are obtained with configuration $\{K_{max} = 1 | K_{corrmin} = 0.05\}$. The
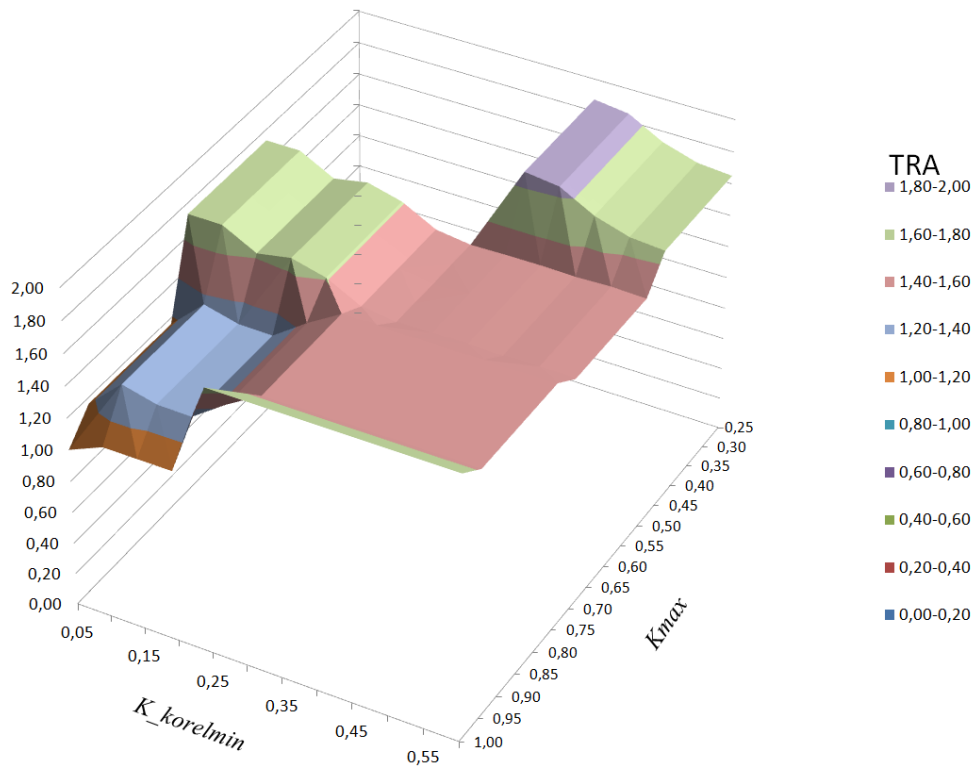
Figure 1. Chart of the total, relative assessment (TRA) of the estimation accuracy according to the imposed criteria parameters K_{max} and $K_{corrmin}$ (22)
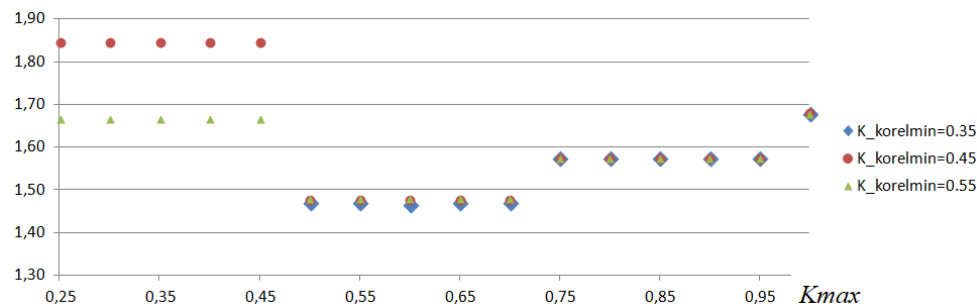

Figure 2. The cross-sections of the surface chart of TRA at selected lines of $K_{corrmin}$. Notice that for $K_{corrmin} < 0.40$ no numerical results were obtained.

next minimal aggregated results have occurred for the configuration $\{K_{max} = 1 | K_{corrmin} \in \langle 0.0.10, 0.20 \rangle\}$ when the result is 1.0913 and $\{K_{max} \in \langle 0.75, 0.95 \rangle | K_{corrmin} = 0.05\}$ with result equal 1.1612. Within the range of coefficients discussed above, one can also find another slight, local minimum (1.2985). All of that means that all (or almost all) data initially collected in the set (Y_A, A) can be used in each discussed calculation. Moreover, the numbers in Table 1 as well as the chart, show that while K_{max} is greater than 0.7 – the stronger a correlation remand is, the worst value estimations can be obtained. In other words, this conclusion could mean that there is no need to respect any similarity issues as well as a signification problem of independent variables. But it is not true. Skipping this unacceptable proposal, one can also find that within the rest of possible configurations of K_{max} and $K_{corrmin}$ different conclusions are also to be drawn. For ranges of $K_{max} \in \langle 0.25, 0.45 \rangle$ as well as $K_{max} \in \langle 0.50, 0.70 \rangle$, the relationship between proportional, aggregated results of estimation accuracy and the criterion factor $K_{corrmin}$ indicates the opposite tendency: better result of accuracy can be obtained when the correlation demand arises.

The most interesting result of the above investigation is that when the criterion of the demand of correlation level between vectors $\mathbf{A}^{\langle j \rangle}$ and $\mathbf{Y_A}$ is medium or strong ($K_{corrmin} \geq 0.30$), then the similarity issue starts to be significant. It is to be noticed that in each column in Table 1, when $K_{corrmin} \geq 0.30$, or in each corresponding cross-section (Figure 2) of the surface chart on Figure 1, the total, relative assessment (TRA) of accuracy is the lowest for medium range of K_{max} ($K_{max} \in \langle 0.50, 0.70 \rangle$). Moreover, within this range, it has its local minimum equal 1.4672 for $K_{max} = 0.60$ and $K_{corrmin} = 0.354$, telling that the most accurate estimations (in the sense of (14)) were obtained when the criterion of dissimilarity (21) was strong enough ($K_{max} = 0.60$) and the demand of correlation level of independent variable was in the medium range. Each extreme demand on similarity of comparables or signification of descriptive variables acted against the effectiveness of estimations process as well as against the estimation accuracy.

These considerations and the results of the experiment show that, in the estimation process based on a linear price model and being resolved by n the LS method, there are some conditions not considered yet which may affect the prediction

Table 1. Total, relative assessment (TRA) of the estimation accuracy according to the imposed criteria parameters K_{max} and $K_{corrmin}$ (22)

K_{max}	$K_{corrmin}$											
	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60
0.25								1.8740	1.8476	1.7331	1.6677	1.6501
0.30								1.8740	1.8476	1.7331	1.6677	1.6501
0.35								1.8740	1.8476	1.7331	1.6677	1.6501
0.40								1.8740	1.8476	1.7331	1.6677	1.6501
0.45								1.8740	1.8476	1.7331	1.6677	1.6501
0.50	1.7324	1.7320	1.6219	1.6612	1.6077	1.4979	1.4698	1.4746	1.4764	1.4721	1.4797	1.4730
0.55	1.7324	1.7320	1.6219	1.6612	1.6077	1.4979	1.4698	1.4746	1.4764	1.4721	1.4797	1.4730
0.60	1.7324	1.7320	1.6219	1.6612	1.6077	1.4979	1.4672	1.4746	1.4764	1.4721	1.4797	1.4730
0.65	1.7324	1.7320	1.6219	1.6612	1.6077	1.4979	1.4698	1.4746	1.4764	1.4721	1.4797	1.4730
0.70	1.7324	1.7320	1.6219	1.6612	1.6077	1.4979	1.4698	1.4746	1.4764	1.4721	1.4797	1.4730
0.75	1.1612	1.3553	1.3024	1.2985	1.4473	1.5744	1.5744	1.5744	1.5744	1.5744	1.5744	1.5744
0.80	1.1612	1.3553	1.3024	1.2985	1.4473	1.5744	1.5744	1.5744	1.5744	1.5744	1.5744	1.5744
0.85	1.1612	1.3553	1.3024	1.2985	1.4473	1.5744	1.5744	1.5744	1.5744	1.5744	1.5744	1.5744
0.90	1.1612	1.3553	1.3024	1.2985	1.4473	1.5744	1.5744	1.5744	1.5744	1.5744	1.5744	1.5744
0.95	1.1612	1.3553	1.3024	1.2985	1.4473	1.5744	1.5744	1.5744	1.5744	1.5744	1.5744	1.5744
1.00	1.0000	1.0913	1.0913	1.0913	1.6802	1.6802	1.6802	1.6802	1.6802	1.6802	1.6802	1.6802

relevance of endogenous variable. The essence of these conditions is contained in the relationship between the characteristics of the subjects taken into the analysis. In the conducted experiment, the role of the endogenous variable has been assigned as the unit price of a hypothetical property. The starting point for the conducted considerations was the real estate market, since the prediction quality and its reliability are specifically conditioned by the question of similarity on that market. The carried out studies demonstrate that the link between the similarity (or dissimilarity) of the sold properties, used as comparables and the valued property, affects the undertaken estimations' accuracy. The observations made show that there is a niche that is worthy of further research.

References

Act (1997). The Act of 21 August 1997 on real estate management. Official Journal 2018, No. 0, item 121. consolidated text, as amended.

Albritton, H. D. (1982). *Controversies in real property valuation: a commentary*. Amer Inst of Real Estate appraisers, Chicago.

Bruce, R. W. and Sundell, D. J. (1977). Multiple regression analysis: history and applications in the appraisal profession. *Real Estate Appraiser*, 43(1):37–44.

Bundesministerium für Umwelt Naturschutz Bau und Reaktorsicherheit (BMUB) (2006). Wertermittlungsrichtlinien – WertR 2006. https://www.bmi.bund.de/SharedDocs/downloads/DE/gesetztestexte/richtlinie/wertr.pdf. Accesed: 26 March 2020.

Czaja, J. (1997). *Modele statystyczne w informacji o terenie*. Wydawnictwo AGH, Kraków.

d'Amato, M. and Kauko, T., editors (2017). *Advances in automated valuation modeling*. Springer International Publishing AG, doi:10.1007/978-3-319-49746-4.

French, N. (2003). The rics valuation and appraisal standards. *Journal of Property Investment & Finance*, 21(6):495–501, doi:10.1108/14635780310508649.

French, N. (2004). The valuation of specialised property. *Journal of Property Investment & Finance*, 22(6):533–541, doi:10.1108/14635780410569506.

Glumac, B. and Des Rosiers, F. (2018). Real estate and land property automated valuation systems: A taxonomy and conceptual model. *Luxembourg Institute of Socio-Economic Research (LISER) Working Paper Series*, 9, doi:10.2139/ssrn.3183132.

ImmoWertV (2009). Verordnung über die Grundsätze für die Ermittlung der Verkehrswerte von Grundstücken (Immobilienwertermittlungsverordnung – ImmoWertV). Bundesrat Drucksache 296/09, 3 April 2009, https://www.gesetze-im-internet.de/immowertv/ImmoWertV.pdf. Accesed: 13 September 2017.

International Valuation Standards Council (IVSC) (2010). Exposure Draft of New International Valuation Standards.

Isakson, H. R. (1986). The nearest neighbors appraisal technique: An alternative to the adjustment grid methods. *Journal of the American Real Estate & Urban Economics Association*, 14(2):274–286, doi:10.1111/1540-6229.00387.

Manly, B. F. and Alberto, J. A. N. (2016). *Multivariate statistical methods: a primer*. CRC Press LLC, Boca Raton.

McCluskey, W. and Borst, R. A. (1997). An evaluation of MRA, comparable sales analysis, and artificial neural networks (ANNs) for the mass appraisal of residential properties in northern ireland. *Assessment Journal*, 4(1):47–55.

McCluskey, W. J. and Borst, R. A. (2017). The theory and practice of comparable selection in real estate valuation. In d'Amato M. & Kauko T., editor, *Advances in Automated Valuation Modeling*, pages 307–330. Springer International Publishing, doi:10.1007/978-3-319-49746-4.

Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T., and French, N. (2003). Real estate appraisal: a review of valuation methods. *Journal of Property Investment & Finance*, 21(4):383–401, doi:10.1108/14635780310483656.

Pawlak, Z. (1981). Information systems theoretical foundations. *Information Ssystems*, 6(3):205–218, doi:10.1016/0306-4379(81)90023-5.

Pawlak, Z. (1983). *Systemy Informacyjne. Podstawy Teoretyczne [Information Systems. Theoretical Foundations]*. Wydawnictwa Naukowo-Techniczne, Warsaw.

Peto, R., French, N., and Bowman, G. (1996). Practice paper: Price and worth developments in valuation methodology. *Journal of Property Valuation and Investment*, 14:79–100, doi:10.1108/14635789610153489.

Rao, C. R., Rao, C. R., Statistiker, M., Rao, C. R., and Rao, C. R. (2008). *Linear statistical inference and its applications*, volume 2. Wiley, New York, second edition edition, doi:10.1002/9780470316436.

Regulation (2004). Regulation of the Council of Ministers of 21 September 2004 on the property valuation and appraisal reports. Official Journal 2004, no. 207, item 2109.

Royal Institution of Chartered Surveyors (RICS) (2017). RICS Valuation – Global Standards 2017. https://www.rics.org/. Accessed: 26 March 2020.

Schlaes, J. (1984). The market in market value. *The Appraisal Journal*, pages 494–518.

Sen, A. and Srivastava, M. (1994). *Regression analysis: theory, methods, and applications*. Springer-Verlag, New York, 2nd corrected printing. edition.

Shenkel, W. and Eidson, A. (1971). Comparable sales retrieval systems. *Appraisal Journal*, 39(4):540–556.

Skaff, M. S. (1975). The search for comparable sales: A new approach. *Assessors Journal*, pages 7–16.

Standards' Commision of Polish Federation of Valuer's Associations (SCPFVA) (2009). Standards. `https://pfsrm.pl/aktualnosci/item/13-standardy`. Accessed: 26 March 2020.

Tchira, A. (1979). Comparable sales selection—a computer approach. *The Appraisal Journal*, 47(1):86–98.

Thompson, J. F. and Gordon, J. F. (1987). Constrained regression modeling and the multiple regression analysis-comparable sales approach. *Property Tax Journal*, 6(4):251–262.

Vecchia, A. V. (1988). Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(2):297–312, doi:10.1111/j.2517-6161.1988.tb01729.x.

Weisberg, S. (1980). *Applied Linear Regression*. John Wiley & Sons.

Zyga, J. (2016). Connection between similarity and estimation results of property values obtained by statistical methods. *Real Estate Management and Valuation*, 24(3):5–15, doi:doi:10.1515/remav-2016-0017.

Zyga, J. (2019). Dissimilarity as a component of the property price model. *Real Estate Management and Valuation*, 27(3):124–132, doi:10.2478/remav-2019-0030.