sciendo

DATA SELECTION AS THE BASIS FOR BETTER VALUE MODELLING

Jacek Zyga, PhD
Faculty of Engineering and Architecture
Lublin University of Technology
e-mail: j.zyga@pollub.pl

Abstract

The article is a voice in the debate on the scope of the application of statistical methods in real estate appraisal, written from the comparative perspective. It presents the results of an illustrative valuation of housing units with the use of databases of various sizes, constructed on the basis of publicly available data from the register of property prices and values. Against this background, the article presents an analysis of differences between the objectives and published results of valuations, which exemplify broadly understood property price modelling or property value modelling, as well as of activities focused around appraising a specific object. The conducted experiments demonstrated that, for the purposes of real estate appraisal itself, the selection of data is more useful than searching for a price model.

Key words: *similarity, models of data, selection of data, real estate appraisal*.

JEL Classification: *C13, C51, C52*.

Citation: Zyga J., 2019, *Data Selection as the Basis for Better Value Modelling*, Real Estate Management and Valuation, vol. 27, no. 1, pp. 25-34.

DOI: 10.2478/remav-2019-0003

1. The objectives and results of real estate appraisal

Numerous researchers have been communicating successes achieved in the research area of real estate appraisal. However, when the reader studies the details of the presented analyses, it often turns out (OSLAND 2010; BEAMONTE et al. 2013; KRYVOBOKOV, WILHELMSSON 2007; BERACHA, WINTOKI 2013; DĄBROWSKI 2010) that calculations were based on bidding data, secondary data, data concerning dissimilar objects or only on collections of prices which might even be transaction prices (as if it were about stock exchange shares rather than about the individualized objects of these transactions). Moreover, cases have been brought to light in which the authors opt to use data from markets other than the real estate market, or different data in general (for example, data regarding car sales). In spite of the highest regard for the technical side of many of the presented analyses, which sometimes are a manifestation of statistical analysis virtuosity, the study of many of these cases leads to doubts as to whether the presented analyses are in fact aimed at valuing any real-life object (as they declare to be)? Perhaps they merely serve the purpose of recognizing the mechanisms determining changes on the market or its dynamics, and, at most, can be a starting point for detailed considerations about valuing particular elements of this market? Obviously, one can also find examples in which the application of the methods of mathematical statistics in real estate appraisal is the objective of the activity, where, apart from the proper, strict calculation regime, the basis of the analyses is information about transactions concerning relevantly similar properties, or at least sold properties, and not arbitrary data correlated with the real estate market (CZAJA 1997; HOZER et al. 1999; ADAMCZEWSKI 2011; HRAMADA 2016; BARAŃSKA 2004, 2005, 2010; BOURASSA et al. 2010; KLEIBER 2005; KONTRIMAS, VERIKAS 2011; DEL GIUDICE et al. 2017; MACH 2017).

Inspired by the results of analyses of local markets of housing units in Campania Region, Italy (DEL

GIUDICE et al. 2017) or Krakow (BARAŃSKA 2004), an initiative was launched to clarify whether an analysis of the market in the possibly broadest perspective is, in fact, the most advantageous option from the point of view of valuing a specific property. Or is the precise selection of objects relevantly similar to the object of target valuation possibly a better choice? The issue raised here therefore fits in with the ongoing debate on the validity of employing automated methods of appraisal, based on algorithms which use the tools of mathematical statistics and are in opposition to the simplified methods of appraisal, recommended in a number of regulations and in professional standards of real estate appraisal.

2. Data and methods of data analysis

In order to illustrate the analyzed problem, a study was conducted concerning the structure of a certain collection of data describing housing units which had been sold in Lublin in the years 2004-2014. The complete collection consisted of 8,088 transactions (Fig. 1), out of which a set of 1,949 transactions (from the years 2012-2014 (Fig. 3)) was selected for further analyses, whose target results were to be confronted with the results of simplified valuations. From this set, adequately defined subsets were isolated at the subsequent stages of research in order to make further attempts to value selected control objects (properties).

For both sets of data, the broad one and the one from the reduced period of time, analyses of variation in unit prices of properties over time were conducted; they were also recalculated to the price level from the last day indicated in the set. There were differences between the estimations of unit prices of particular properties for the last day of the analyzed periods. The valuations which were deemed as more accurate were those carried out by means of the model constructed on the basis of data from the period between January 2012 and January 2014 (as in Figure 4), for which the control estimation of the trend of price changes was close to zero ($-5 \cdot 10^{-12}$). The case was opposite for the trend of price changes from the years 2004-2014 nominally reduced on the basis of the polynomial of the fifth degree (coefficient 0.019 in Figure 2).

The data presented above was retrieved from the Register of Property Prices and Values (Rejestr cen i wartości nieruchomości, RCWN) of Lublin City Office, kept as a supplementary collection to the Register of Lands and Buildings (Rejestr gruntów i budynków, RGB), which is the current equivalent of a cadastre in the Polish structures of the real estate management system. In spite of its name, the Register of Property Prices and Values is not only for registering prices and values of properties shown in the Register of Lands and Buildings. The prices are accompanied by descriptive information referring to the sold properties (including the analyzed housing units). Thanks to this, the contents of the Register of Property Prices and Values can be a good example of a database on the basis of which information systems which constitute a starting point (consistent with the definition by Z. Pawlak (PAWLAK 1983, p.16) for analyses of the figures can be constructed.

As research has shown (ZYGA 2016b, 2017), the contents of the Register of Property Prices and Values are far from perfect and, in many cases, do not justify its use as the only source of data regarding the real estate market for the purposes of property appraisal. The reason for this lies in the shortcomings in fulfilling the obligations imposed by the Polish legislation (REGULATION 2001), related to filling the database with adequate contents. In reference to housing units, however, it seems to constitute a database which is extensive enough and contains data with enough details to be used for the above-mentioned purposes. In the current study, the following features of housing units demonstrated in the Register of Property Prices and Values were selected for analysis:

The features presented in the table were supplemented with the types of building construction assigned to each of the properties sold, according to the list in Table 2. This data was only partially provided in the Register of Property Prices and Values, an thus additional inventory in field was necessary.

The validity of employing particular attributes in the statistical analysis was tested by studying the correlations between them, as well as the information capacity of databases, configured according to Hellwig's concept (HELLWIG 1968). As the results from Tables 1 and 2 demonstrate, the majority of variables are weakly correlated with the unit prices of the studied properties. Therefore, the factor of "the age of the building" (AoB) in which the properties sold were located, was chosen as the basic available factor determining the prices. Year of construction (YoC) was presented as raw data found in the RCWN data set but, in further computing, was used for recalculation into Age_of _Building (AoB) only. The range of the AoB variable was numerically better (0-165) so it was used instead of YoC.

Subsequently, the complementary variable "hbmm" (binary variable) was used, which distinguishes buildings constructed using the newest technologies. An analogically characteristic feature distinguishing buildings with a weak structure ("hbmwc") was ultimately ignored, as it minimizes the integral indicator H of Hellwig information capacity basically in any combination of attributes.

Table 1

The features of housing units demonstrated in the Register of Property Prices and Values used in the study

Feature (attribute)	Symbol	Range of variable values	Correlation coefficient with unit prices
District	DIS	---	---
Address	ADR	---	---
Number of storeys in the building	NST	1 - 12	0.012
Year of construction	YoC	1847-2013	0.424
Storey	STO	1-12	-0.015
Premises area [m^2]	PA	13 - 220	-0.003
Number of rooms	NoR	1 - 9	-0.017
Age of the building [yr]*	AoB	0-165	-0.428

*/calculated as a difference between the year of the sale and the year of construction

Source: author's own study on the basis of data from the Register of Property Prices and Values of Lublin City Office.

Table 2

Breakdown of buildings by construction type

Type of construction	Symbol	Correlation coefficient with unit prices
High buildings (4-5 storeys), masonry construction	hbm	-0.017
High buildings (3-8 storeys), modern masonry frame construction	hbmm	0.409
High buildings (5-6 storeys), frame construction	hbfc	-0.055
High buildings, masonry construction with wooden ceilings	hbmwc	-0.290
High buildings, masonry construction with brick ceilings	hbmbc	-0.162
Small masonry buildings with ceilings from materials other than wood	smb	-0.002
Small masonry buildings with wooden ceilings	smbwc	-0.052
Low buildings (2-3 storeys), masonry construction with ceilings from materials other than wood	lmb	-0.003
High buildings (5-12 storeys) made of prefabricated elements	hbpe	-0.081
High buildings (5-12 storeys), reinforced concrete construction	hbrcc	0.014

Source: the author's own study.

In certain analyses of the correlation coefficient of particular variables in reference sets of varying sizes, different partial conclusions were obtained. The variables rejected at the stage of the analysis of the biggest set (1,948 observations), e.g. Storey, Number of Rooms, Premises area, were relevant variables in smaller sets. Table 3 illustrates the evolution of the estimations of the correlation coefficient between the values of selected variables and unit prices, as well as the changes in the sizes of reference sets. Some of the variables included in Table 3, but left out of Table 5 (e.g.: hbmwc, hbmbc, Storey, Number of Rooms) were rejected as their correlation with other variables was too strong.

Table 3

Correlation coefficient between selected variables and unit prices
in sets with different numbers of price observations

Number of price observations (n)	AoB	PA	STO	NoR	hbmwc	hbmbc	hbmm
1,948	-0.428	-0.003	-0.015	-0.017	-0.290	-0.162	0.409
22	-0.457	-0.354	0.103	-0.261			
7	-0.327	-0.277	-0.057				

Source: the author's own study.

Table 4

Summary of parameters of appraised control properties (benchmarks of appraisal)

	Control property A	Control property B	Control property Z
District	27	21	34
Address	Agatowa 11/xx	Balladyny 12/xx	Żmigród 12/xx
Sold on	2013-07-05	2013-01-08	2012-05-18
Unit transaction price	5,921	4,850	2,945
Unit price corrected for 2014-01-22	5,139	3,713	1,666
Feature symbol			
NST	4	5	4
STO	1	0	0
PA	56.59	54.5	63.4
NoR	3	4	4
AoB	7	47	107

Source: the author's own study.

As the sets of properties on the basis of which subsequent appraisals of control properties were performed were being reduced (or, in the process of the selection of increasingly similar properties), other variables also identified as relevant were "Premises area" ("PA") and "Storey" ("STO"). Their correlation coefficients with unit prices are presented in Table 3, in a summary of parameters describing the subsequent variants of the conducted analyses. The two above variables were in fact measured on interval scales, although putting them into ordinal scales might have resulted in better assessment of the mentioned features (FORYŚ, GACA 2016). In the present experiment, the decision not to do so was made because transformation of variables into an ordinal scale would require additional information derived from the outside the main, rough data source, i.e. RCWN. Each attempt of putting these variables in any kind of orderliness would have to rely on the decision which variable mark or variable value is better/worse than others.

The variants in which calculation experiments were conducted were listed beginning with those involving the most extensive sets of transactions and ending with experiments based on a few transactions referring to the properties which were the most similar to the three selected control (benchmark) properties.

It should be noted that the aim of research was to determine, from the point of view of the purposes of valuing a specific property, whether the more advantageous option is to analyse the market in the broadest perspective possible, or to base the valuation on a low-abundant set of objects relevantly similar to the object of target appraisal. Therefore, the guiding principle adopted in the research was to systematically reduce, in subsequent steps of the experiment, the size of reference sets, at the same time selecting objects increasingly more similar to the benchmark property. Due to the limited possibilities of identifying the features of similarity, only two degrees of reduction were applied: to the subgroup of "properties located in the same street," and subsequently, the subgroup of "properties located in the same building." A certain exception to this rule was the case of the series of valuations of the "Z" property, in which the reduction of the basic set (of 1,948 properties) to the group of properties located on the same street led to creating a subset of only five elements already in the first step.

As similarity measures *p*, the author adopted estimation based on Euclidean distance averaged over a given sample, calculated on the basis of the values of variables used in a given variant according to the formula (ZYGA 2009):

$$p = 1 - \frac{1}{n}\sum_{i=1}^{n}\sqrt{\sum_{k=1}^{s+2}(x_{Bk} - x_{ik})^2} \tag{1}$$

where: *p* stands for the averaged similarity coefficient, n - the number of compared properties in a given set, *s* - the number of explanatory variables (increased by the variables taken into consideration in the addresses of particular properties, e.g. the location of the comparative property on the same street as the benchmark property (0/1) and the location of the comparative property in the same building as the benchmark property (0/1)), x_B, x_i stand for the standardized values of explanatory variables of the benchmark property (B) and the subsequent property in the set (i), respectively.

There are many other ways of measuring similarity. Some of them, based on different models of the distance e.g.: Spearman, Bray-Curtis, Canberra, Minkowski or Chebyshev/Maximum distances, were discussed in (WALESIAK 2002; ZYGA 2009). Other techniques of assessing similarity, based on different weighting schemes, variation or rank transformation analysis were presented in (MCCLUSKEY, BORST 2017). The selected option of similarity computation sufficiently includes the characteristics of each property. Due to the manner of identifying property location, it strongly underlines the metric of nearness as well as the hedonic distance between the benchmark property and other properties in the sample. Therefore, it is similar to the ideas of (ISAKSON (1986) and (KRAUSE, KUMMEROW 2009), who used Mahalanobis distance with or without an additional distance penalty.

Moreover, particular sets of prices were characterized by the coefficient of relative scatter Sc quoted after (ADAMCZEWSKI 2011):

$$Sc = 2\frac{P_{max} - P_{min}}{P_{max} + P_{min}} \tag{2}$$

where: Pmax and Pmin stand for the maximum and minimum unit price in a given set, respectively.

The aim of each of the experiments was the best possible valuation of the benchmark property (one of the three selected from an extensive collection of 1,949 sold properties), and, at the same time, the selection of these features of the properties which best describe the variation of prices in the adopted reference set. The required value was understood as market value, in accordance with both the provisions of law (ImmoWertV 2009, RPM Act 1997) and the standards of valuation (ASB 2017, IVSC, 2010, RICS 2017, SCPFVA 2009). As a consequence, it was necessary to take into consideration the characteristics of each of the appraised properties separately, and not to view them as representatives of groups of properties (properties typical for the whole city, a selected quarter, neighborhood, etc.).

In each of the variants, the market value of appropriate benchmark properties was estimated. Estimation was performed using the least-squares procedure in the multiple regression model. In the subsequent steps of the experiment, when the number of reference properties was being reduced, the accuracy of the calculated values of appraised benchmark properties was subject to control. The control was performed by determining the differences between particular results, i.e. estimated values of each of the benchmark properties and the their transactional prices, adjusted for the passage of time. Also subject to control were measures of the average estimation error of subsequently estimated values, and the quality of reference sets expressed by standard deviation from particular samples of prices, as well as standard deviation from the adopted models.

The aggregated results of calculations are presented in Table 7.

3. Conclusions of the conducted analyses

As it can be concluded from the analysis of the contents of Tables 5 and 6, the results of estimations of the values of the appraised properties are more accurate in relation to their relevant model prices every time after reducing the set of data which is the basis for calculations to the most similar properties (in the analyzed examples, similarity was expressed in the location on the same street and in the same building). As reference properties are selected in a well-considered manner, the similarity measure p in reference sets is steadily increasing. Deteriorating, on the other hand, are the values of indicators describing the informational power of the subsequently created subsets (integral indicator of Hellwig information capacity H) and sample standard deviations, as well as those calculated on the basis of corrections to model equations. Correlating particular indicators with the results of the accuracy of the subsequent estimations of values of the appraised properties (measured as differences between the estimated values and model prices) led to demonstrating which of the used measures of the quality of estimation can be helpful in identifying a truly accurate case of calculating value. The results of the above-mentioned analysis are presented in Table 7.

Table 5

Summary of parameters describing the subsequent variants of the conducted analyses

Symbol of the variant	Number of price observations (n)	Number of explanatory variables (s)	Symbols of the variables used		Coefficient of relative variability Sc	Average price in the sample [PLN]	Averaged similarity coefficient *p*	Integral indicator of Hellwig information capacity H
Unit price of the benchmark property A 5,139 PLN								
A0	1,948	1	AoB		1.84	4,305	0.52	0.2304
A1	1,948	2	AoB	hbmm	1.84	4,305	0.59	0.2675
A2	22	1	AoB		0.32	5,191	0.76	0.2084
A2-2	22	1	PA		0.32	5,191	0.75	0.1254
A3	22	2	AoB	PA	0.32	5,191	0.80	0.2736
A4	7	2	STO	PA	0.21	5,371	0.82	0.0588
Unit price of the benchmark property B 3,713 PLN								
B0	1,948	1	AoB		1.84	4,304	0.53	0.2305
B1	1,948	2	AoB	hbmm	1.84	4,304	0.62	0.2677
B2	8	1	AoB		0.32	4,235	0.81	0.0806
B2-2	8	1	PA		0.32	4,235	0.75	0.3157
B3	8	2	AoB	PA	0.32	4,235	0.76	0.3804
B4	5	2	STO	PA	0.19	4,305	0.78	0.0245
Unit price of the benchmark property Z 1,666 PLN								
Z0	1,948	1	AoB		1.84	4,305	0.51	0.2289
Z1	1,948	2	AoB	hbmm	1.84	4,305	0.61	0.2667
Z2	5	1	AoB		1.37	2,998	0.87	0.3191
Z3	5	2	AoB	STO	1.37	2,998	0.81	0.5289
Z4	5	2	STO	PA	1.37	2,998	0.79	0.2655

Source: the author's own study.

Table 6

Results of calculations (highlighted positions indicate the most accurate estimations)

Symbol of the variant	Estimated value [PLN]	Mean estimation error of the estimated value MSE [PLN]	Difference between the estimated value and the price of the pattern property [PLN]	Sample standard deviation *s* [PLN]	Standard deviation from the adopted model σ [PLN]	Coefficient of determination R2
Valuation of benchmark property A						
A0	4,823	30	-316	1,046	918	0.23
A1	5,037	36	-102	1,045	894	0.26
A2	5,118	96	-20	478	425	0.21
A2-2	5,158	97	20	478	447	0.13

A3	5,056	150	-82	478	371	0.43
A4	5,207	260	68	368	381	0.10
Valuation of benchmark property B						
B0	4,113	22	400	1,046	918	0.23
B1	3,996	24	283	1,045	894	0.27
B2	4,180	114	467	544	301	0.08
B2-2	4,191	93	478	544	260	0.32
B3	4,141	176	428	544	461	0.38
B4	4,534	825	820	323	362	0.06
Valuation of benchmark property Z						
Z0	3,036	57	1,370	1,044	917	0.23
Z1	3,311	61	1,645	1,045	894	0.27
Z2	2,448	277	783	1,332	480	0.32
Z3	1,693	892	28	1,332	1,055	0.06
Z4	2,046	1,062	380	1,332	1,170	0.42

Source: the author's own study.

Table 7

Analysis of the relationship between selected precision characteristics
and the accuracy of estimating the value of benchmark properties

Benchmark property	A	B	Z
Controlled parameters	Correlation coefficients of indicated parameters with differences between estimated values and reference prices of benchmark properties		
s	0.74	-0.79	-0.91
σ	0.75	-0.57	-0.31
MSE	-0.44	0.95	-0.90
H	0.33	-0.67	-0.73
R2	0.17	-0.61	0.21
p	-0.78	0.53	-0.78
Sc	0.75	-0.65	0.91

Source: the author's own study.

The conducted study has demonstrated that typical measures of the quality of information in a set of data and of the quality of estimation itself, used in the least-squares procedure, i.e.: sample standard deviation - *s*, standard deviation calculated on the basis of corrections to model equations – σ, integral indicator of Hellwig information capacity H, or coefficient of determination R^2, are either weakly or negatively correlated with the differences between the estimated values and the prices of benchmark properties. These results indicate their low diagnostic value in determining which of the possible results of estimation (from Table 7) is the best. Only the mean error of estimation of the estimated value MSE (Table 7), in the case of estimating the value of property B, provides an example of a clear and correct correlation which can identify the best result. In the remaining variants of calculations, on the other hand, a clear diagnostic signal is given by: coefficient *p* of mean similarity between reference and benchmark properties, and the coefficient of the relative scatter of prices in reference sets - *Sc*. A negative correlation of *p* coefficient with the differences between the estimated values and the prices of benchmark properties provides yet another empirical proof of the significance of the factor of similarity in analyses of value of this kind (the more similar the reference properties, the smaller the error of estimation of the value of the appraised property). The simultaneously observed correlation of the coefficient of relative scatter *Sc* is a hint that what is very advantageous for the quality of estimation is the coherence of the set of prices of reference properties.

The observations above contain an element of an old truth, well known to every active appraiser. This, however, is a warning for other authors of publications concerning the topic of real estate appraisal. Where conclusions aim to express opinions concerning the value of objects on the real estate market, which are inevitably individualized in terms of their characteristics, the success of particular calculation processes, which are aimed at determining the value of subsequent properties, is not decided by the amount of data (prices), but by the type and characteristics of reference objects included in the analysis.

Looking from a certain distance at the presented results, it is worth stressing that the above-mentioned conclusions do not contradict the usefulness of the published results of valuations in which the power of quantitative methods was used. This is demonstrated by the example B1 (Table 6) of the appraisal of property B. This property, in contrast to properties A and Z, was, in many respects, average. It was not located on the outskirts of the city nor in its center. It was neither very old nor new. Its surface area classified it as a medium-sized property. It was under such conditions that quantitative methods proved effective. Very similar conclusions were drawn by (Bitner 2010) as well as (McCluskey, Borst 2017). The most accurate result (in the sense specified above) was obtained with the basic, extensive reference set, though it might not have been indicated unambiguously enough by s, σ, H, MSE or R^2.

The conducted calculation experiments confirm earlier results of similar tests (Zyga 2016a) and demonstrate that, in real estate appraisal, a major role is played by the factor of similarity, whose analysis, as well as the adequate selection of data based on it, leads to more accurate results than those achieved as a result of price modelling with a considerable amount of input data. In relation to the real estate market, the selection of this data is usually very limited, and the lack of correlation between most of the available factors (Number of storeys in the building, Storey, Premise area, Number of rooms) and prices considerably limits the possibilities offered by statistics. In such cases, the large amounts of data gathered become of little use in reality. Therefore, the technique of adequate selection of data is more effective from the point of view of the economy of undertaken actions than the selection of variables. It is only in the case of problems with the identification of factors of similarity that the very manner of analyzing the relationships between the variables in increasingly bigger sets becomes significant, and thus becomes justified.

It is also worth indicating that, in particular analyses of significance (measured with the absolute value of the correlation coefficient) of different variables in reference sets of varying sizes, different partial conclusions were obtained whether variables (and which of them) can be suspected to affect the variability of prices of the analyzed properties. The variables rejected at the stage of the analysis of the set of 1,948 observations (e.g. Storey, Number of Rooms, Premises area) were relevant variables in smaller sets. They were rejected due to the fact that correlation between them was too high.

The case studies provide empirical evidence that models obtained from large data sets are less efficient. This lower efficiency is not related to their size itself, but to the less factor of similarity. It must be underlined that the selection described above cannot be understood as merely a way of simplifying market data analysis. It has to be the result of this analysis. Simple reduction of the number of data, as well as reducing the amount of information, does not automatically increase the quality of the final result. This result can be increased only if the selected data provides a more coherent and adequate set of properties, much more similar to the subject (benchmark) property. One must remember that numerousness of these sets cannot decline too much because, in such a case, the role of the random component rises too much.

Against this background, an essential feature of the technique of selecting data based on similarity is performing it as a one-stage procedure. Further reduction of the number of objects more and more similar to one another can, at most, lead to obtaining a set of formally indistinguishable properties with dissimilar prices (Zyga 2015). To avoid this, the RCWN database should be developed by adding a number of characteristics of sold properties. In fact, the structure of RCWN registry is ready to collect such information, but the system delivering this information does not work properly. RCWN appears to be a good starting point for property valuation but, in the context of presented analysis, turns out to be incomplete. In Polish appraisal practice, appraisers must replenish basic RCWN databases with missing (but very relevant) additional information.

4. References

ADAMCZEWSKI Z., 2011, *Elementy modelowania matematycznego w wycenie nieruchomości. Podejście porównawcze (Elements of Mathematical Modelling in Property Appraisal. Comparative Approach)* Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa.

APPRAISAL STANDARD BOARD (ASB), 2017, Uniform Standards of Professional Appraisal Practice 2016-2017 (USPAP), available at: http://www.uspap.org/, accessed 2017-08-31.

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. No. 1. pp. 31-38.

BARAŃSKA A., 2005, *Estymacja parametrów nieliniowych modeli funkcyjnych dla potrzeb predykcji rynkowej wartości nieruchomośc*i *(Estimation of Non-linear Function Models for the Needs of Prediction the Market Value of Real Estate)*. Geodezja, t. 11, z. 2, Kraków.

BARAŃSKA A., 2010, *Modele multiplikatywne w procesie wyceny (Multiplicative Models in the Valuation Process)*, Studia i materiały Towarzystwa Naukowe Nieruchomości , v. 18, nr 1, Olsztyn.

BEAMONTE A., GARGALLO P., SALVADOR M.J., 2013, *Retrospective Analysis of the Dwelling Price by Means of STAR Models with Neighborhood Effects*, Journal of Real Estate Research, Vol. 35, No.2, pp. 173-197

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.

BITNER A., 2010, *O użyteczności metod statystycznych w wycenie nieruchomości (On the Usefulness of Statistical Methods in Real Estate Valuation)*, Infrastruktura i Ekologia Terenów Wiejskich, No. 12/2010, 145-158.

BOURASSA S.C., CANTONI E., HOESLI M., 2010, *Predicting House Prices with Spatial Dependence: A Comparison of Alternative Methods*, JRER, Vol. 32, No. 2– 2010, pp. 139-159.

CZAJA J., 1997, *Modele statystyczne w informacji o terenie (Statistical Models in Information on the Area)* Wydawnictwa AGH, Kraków.

DĄBROWSKI J., 2010, *Zastosowanie Wybranych Metod Statystycznych do analizy rynku nieruchomości (Application of Selected Statistical Methods for Real Estate Market Analysis)*, available at: https://www.statsoft.pl/czytelnia.html, accessed 2016-02-13.

DEL GIUDICE V., MANGANELLI B., DE PAOLA P. 2017, 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.

FORYŚ I., GACA R., 2016, Teoretyczne i praktyczne aspekty opisu cech jakościowych nieruchomości w modelach regresji wielorakiej wyceny wartości mieszkań, (*Theoretical and Practical Aspects of Qualitative Variable Descriptions of Residential Property Valuation Multiple Regression Models*), Zeszyty Naukowe UEK, 9 (957), pp. 89–100.

HELLWIG Z., 1968, *On the Optimal Choice of Predictors.* Study VI in Z. Gostkowski (ed.): Toward a system of quantitative indicators of components of human resources development; Paris: UNESCO,

HOZER J. ED, FORYŚ I., ZWOLANKOWSKA M., KOKOT S., KUŹMIŃSKI W., 1999, Ekonometryczny algorytm masowej wyceny nieruchomości gruntowych (Economic Algorithm of Land Real Estate Valuationn), Uniwersytet Szczeciński, Stowarzyszenie "Pomoc i Rozwój", Szczecin.

INTERNATIONAL VALUATION STANDARDS COUNCIL (IVSC), 2010, Exposure Draft of New International Valuation Standards, IVSC, London.

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.

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 11/2011,pp 443–448. available at: 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 Mode*l. Journal of Property Tax Assessment & Administration, 8(2), pp. 39–52.

KRYVOBOKOV M., WILHELMSSON M., 2007, *Analysing Location Attributes with a Hedonic Model for Apartment Prices in Donetsk, Ukraine*, International Journal of Strategic Property Management, No. 11, pp. 157–178

MACH Ł., 2017, *The Application of Classical and Neural Regression Models for the Valuation of Residential Real Estat*e, Folia Oeconomica Stetinensia, vol. 17, is.1. DOI: 10.1515/foli-2017-0004, pp.44-56.

MCCLUSKEY W.J., BORST R.A., 2017, *The Theory and Practice of Comparable Selection in Real Estate Valuation in D'AMATO M. AND KAUKO T.*, Advances in Automated valuation methodology pp. 307-330.

OSLAND L., 2010, A*n Application of Spatial Econometrics in Relation to Hedonic House Price Modeling*, JRER, Vol. 32, No. 3– 2010, pp.289-320.

PAWLAK Z., 1983, *Systemy informacyjne. Podstawy teoretyczne, (Information Systems. Theoretical Basis),* WNT, Warszawa.

RICS, 2017, Royal Institution of Chartered Surveyors (RICS), 2017, RICS Valuation – Global Standards 2017, RICS, London.

SCPFVA, 2009, Standard's Commision of Polish Federation of Valuer's Associations (SCPFVA), 2009, National Valuation Standard Basic No. 1, PFVA, Warszawa (Komisja Standardów Polskiej Federacji Stowarzyszeń Rzeczoznawców Majątkowych (SCPFVA), 2009, Krajowy Standard Wyceny Podstawowy Nr 1, PFVA, Warszawa)

WALESIAK M., 2002, Pomiar podobieństwa obiektów w świetle skal pomiaru i wag zmiennych, (S*imilarity Measures from the Point of View Scales of Measurement and Variables Weights*), Prace Naukowe Akademii Ekonomicznej We Wrocławiu, Ekonometria 10, No 950, pp. 71-85.

ZYGA J. 2009: Podobieństwo w wycenie nieruchomości, Budownictwo i Architektura Vol. 5 (2): pp.61-76

ZYGA J. 2015: *Search for the Dissimilarity Factors for Nominally Indiscernible Facilities,* Real Estate Managment and Valuation, Volume 23, Issue 3, Pages 65–72, ISSN (Online) 2300-5289, DOI: 10.1515/remav-2015-0026,

ZYGA J., 2016A, *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. 2016B: *The Usefulness of Real Estate Price and Value Register in Appraisal by Comparative Methods, on the Bassis of Lublin Voivodeship Registers*, INFRASTRUKTURA I EKOLOGIA TERENÓW WIEJSKICH, Vol. IV/3/2016, pp. 1673-1688, DOI: http://dx.medra.org/10.14597/ infraeco. 2016.4.3.126

ZYGA J. 2017: *Evaluation of Usefulness of Real Estate Data Contained in the Estates' Prices and Values Register*, INFRASTRUKTURA I EKOLOGIA TERENÓW WIEJSKICH, Vol. III/1/2017, pp.1017–1030, DOI: http://dx.medra.org/10.14597/infraeco.2017.3.1.078

Law Acts:

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

REGULATION 2004, Rozporządzenie Rady Ministrów z dnia 21 września 2004 r. W sprawie wyceny nieruchomości i sporządzania operatu szacunkowego (Dz.U. 2004 nr 207 poz. 2109 z późniejszymi zmianami). [Regulation of the Council of Ministers of 21 September 2004. On the valuation of the property and the preparation of the valuation report, Journal of Laws 2004 No. 207, item. 2109, as amended.]

REGULATION 2001, Rozporządzenie Ministra Rozwoju Regionalnego i Budownictwa z dnia 29 marca 2001 r. W sprawie ewidencji gruntów i budynków (Dz. U. z 2015 r. poz. 542 z późniejszymi zmianami) [Regulation of the Minister Regional Development and Building Industry of 29 March 2001. On land and buildings register, Journal of Laws 2015, item. 542, as amended.]

RPM ACT, 1997, Ustawa z dnia 21 sierpnia 1997 o gospodarce nieruchomościami r. (Dz. U. z 2015r. poz. 782) [The Act of 21 August 1997. On Real Property Management ,Journal of Laws 1997 No. 115, item 741 as amended.]