Folia Oeconomica Stetinensia
Volume 20 (2020) Issue 2
DOI: 10.2478/foli-2020-0041

ANALYSIS OF THE IMPACT OF THE TYPE OF SAMPLING OF REPRESENTATIVE PROPERTIES ON THE RESULTS OF MASS APPRAISAL*

Sebastian Gnat, Ph.D.

University of Szczecin
Faculty of Economics, Finance and Management
Department Econometrics and Statistics
64 Mickiewicza Street, 71-101 Szczecin, Poland
E-mail: sebastian.gnat@usz.edu.pl
ORCID: 0000-0003-0310-4254

Received 23 March 2020, Accepted 18 September 2020

Abstract

Research background: Mass valuation is a process in which many properties are valued simultaneously with a uniform approach. An example of a procedure used for mass real estate valuation is the Szczecin Algorithm of Real Estate Mass Appraisal (*SAREMA*), which can be developed into a multiple regression model. The algorithm is based on a set of drawn representative properties. This set determines, *inter alia*, the quality of obtained valuations.
Purpose: The objective of the study is to verify the hypothesis whether changing the method of sampling representative properties from the originally used simple random sampling to stratified sampling improves the results of the *SAREMA* econometric variant.
Research methodology: The article presents a study that uses two methods of representative properties sampling – simple random sampling and stratified sampling. Errors of the models of valuation created taking into account both methods of sampling and different number of representative properties are compared. A key aspect of the survey is the choice of a better sampling method.
Results: The study has shown that stratified sampling improves valuation results and, more specifically, allows for lower root mean square errors. Stratified sampling yielded better results in the initial phase of the study with more observations, but reducing the percentage of strata participating in the draws, despite the increase in *RMSE,* guaranteed lower errors than the corresponding results based on simple sampling in all variants of the study.

* The research was conducted within the project financed by the National Science Centre, Project No. 2017/25/B/HS4/01813.

Novelty: The article confirms the possibility of improving the results of mass property valuation by changing the scheme of representative properties sampling. The results allowed for the conclusion that stratified sampling is a better way of creating a set of representative properties.

Keywords: property mass valuation, property market analysis, sampling methods.

JEL classification: C10, R30

Introduction

In the practice of real estate valuation, two main trends can be distinguished: individual and mass valuation. In the process of an individual appraisal the individual assessing the value of real estate focuses on a single property or on a small number of properties. Whereas, in the case of a mass appraisal the subject of an appraisal involves a large number of real properties of one type (e.g. Hozer, Kokot, Kuźmiński, 2002). One of the models used more frequently for property valuation are multiple regression models. Their popularity in property valuation, but also in other areas, is mainly due to their simplicity and ease of interpretation. However, these advantages come at a price. In order to build a good model, a number of conditions must be met (e.g. Doszyń, 2012). One of the disadvantages of this type of model is that they do not take into account the relationship between the objects being modelled in space and the fact that the force and direction of the relationship between the target and explanatory variables may not be constant over the entire modelled area. One of the ways of dealing with the latter situation is to divide the analyzed area into smaller – so-called sub-markets. In such a situation, instead of a single model, there are as many models estimated as a number of sub-markets specified. This solution allows for better, more accurate results of property mass valuation. The process of mass valuation can be carried out using various types of models, algorithms and procedures. Some of these procedures, e.g. the Szczecin's Algorithm of Property Mass Appraisal (*SAREMA*), rely on the preliminary valuation of a specified group of representative properties. The values of the other properties are determined by means of an estimated model. Such an approach requires defining a way of specifying the so-called representatives. As in the case of econometric models, there is more than one possibility or sampling of representative properties. Each sampling method has its advantages and disadvantages. The choice may determine the accuracy of valuation results. It is precisely the problem of the influence of the method of real estate sampling on the results of mass valuation that is the main research problem in this study. The aim of the study is to verify the hypothesis whether the method of selecting representatives within the *SAREMA* significantly affects the accuracy of valuations. If so, which of the approaches considered in this study is superior. Among the known sampling methods two ways of creating a set of representatives were used in the study – simple random sampling and stratified sampling. The latter is selected for the study since the area under evaluation has already been divided into sub-markets (see Hozer, Gnat, Kokot, Kuźmiński, 2019), the so-called location attractiveness zones (*LAZs*). The quality of the valuations obtained using both sampling methods will be assessed using the root mean square errors.

The subject of the study entails 318 properties located in Szczecin – one of the largest Polish cities. These properties were subject to valuation due to the revaluation of perpetual usufruct fees.

1. Literature review

The presented study is related to two research streams. The first concerns the application of various methods in the mass valuation of properties. The second is related to sampling from analyzed populations.

A large number of studies on the application of methods from the groups distinguished by Kauko and d'Amato (2008) can be found in the literature: model-driven methods, data-driven methods, methods based on machine learning and expert methods. Research based on model-driven methods mainly concerns econometric models used for valuation purposes (Benjamin, Randall, Guttery, Sirmans, 2004; Isakson, 1998; Dell, 2017), but also purposes other than property value determination can be found. One such application is the identification of atypical observations on the real estate market (Doszyń, Gnat, Bas 2017). Research related to the use of spatial econometric models constitutes a separate branch of studies. Spatial methods are divided into three categories: *SAR* (Spatial AutoRegressive) models, local models (e.g. *GWR* – Geographically Weighted Regression) and geostatistical models (e.g. kriging). An extensive review of property mass valuation models can be found e.g. in (Jahanshiri, Buyong, Shariff, 2011). Data-driven methods include non-parametric models such as *CART* (classification and regression trees) (McCluskey, Daud, Kamarudin, 2014) or random forests (Antipov, Pokryshevskaya, 2012). Methods based on machine learning are nowadays applied more and more frequently. This class of tools accounts for artificial neural networks, rough set theory, fuzzy logic, and genetic algorithms, etc. An interesting application of machine learning methods could be found e.g. in (Zurada, Levitan, Guan, 2011; Wang, Wen, Zhang, Wang, 2014; Ćetković et al., 2018). Those papers present comparative studies, where several regression and artificial intelligence methods are used in real estate valuation. In the case of the low quality of databases, expert-based methods, such as *AHP* (Analytic Hierarchy Process), could support mass appraisal process. The application of these kinds of methods is presented in e.g. (Kilpatrick, 2011).

Most mass property valuation models use data on the transactional prices of properties that are located within the area being valued. However, in the case of underdeveloped markets one can encounter a problem of insufficient market data for building a good model. The mass valuation model used in the study is a model in which this problem can be circumvented. *SAREMA* is based on the market values of representative properties drawn instead of transaction prices. Moreover, the values of these properties are determined by certified appraisers. Those values were determined in individual (not mass) appraisals. Such an approach requires an appropriate procedure for selecting properties to be assessed by experts. Since its introduction *SAREMA* as a procedure of mass valuation has been modified. In its original form, the so-called "expert approach" was based on the knowledge of valuers, who indicated the appropriate values of the parameters of the model. In this approach, a simple random sampling method was used to create a set of representative properties. Such a method of sampling carries the risk of the drawn properties being similar to each other due to the features they are described with. This lack of attributes variability leads to problems with estimating econometric models, and so the *SAREMA* variant has been used in this study. If simple random sampling may not be an effective way of selecting representative properties, it is worth examining whether other methods of sampling will not prove to be more appropriate.

The issue of appropriate sampling is addressed in various studies. The paper by Burgard, Münnich and Zimmermann (2014) addresses the impact of sampling designs on small area estimates for business data. The authors of the study state that "the sampling designs applied, in general, are non-ignorable and may have a major impact on model-based estimates". They also state that "estimators ignoring the sampling design cannot be recommended since they may yield considerably biased estimates". There are a number of sampling methods. Each of them offers different advantages and disadvantages and can be used in different circumstances. In the work by Schreuder, Gregoire and Weyer (2001) authors focus on indicating "what type or combination of types of sampling can be used in various situations, ranging from a sample designed to establish cause-effect or legal challenge to one involving a simple subjective judgment". Several methods of non-probability sampling are discussed in the paper by Etikan, Musa and Alkassim (2016). The authors indicate for which type of data and research the sampling schemes they analyze can be effectively applied. Issues concerning sampling are also addressed in the context of frequentist and bayesian methods (Rao, 2011). Sampling is also inevitably associated with bootstrap methods. The effects of such methods on small-area estimation are also investigated (Lahiri, 2003). There is no doubt that sampling is an important research problem to be considered in different areas, including the real estate market.

2. Methodology of the research

As was previously mentioned, there are a number of methods of mass appraisal. One example of such a method is the Szczecin Algorithm of Real Estate Mass Appraisal (*SAREMA*). In the survey, an econometric form of this algorithm is used:

$$\ln\left(w_{ji}\right) = \alpha_0 + \sum_{k=1}^{K}\sum_{p=2}^{k_p}\alpha_{kp}x_{kpi} + \sum_{j=2}^{J}\alpha_j laz_{ji} + u_i \tag{1}$$

where:

w_{ji} – market value of 1 square meter of i–th real estate in j–th location attractiveness zone,
N – number of real estates ($i = 1, 2, ..., N$),
J – number of location attractiveness zones ($j = 2, 3, ..., J$),
$surf_i$ – surface of i–th real estate (in m^2),
α_0 – constant term,
K – number of real estate attributes,
k_p – number of states of k-th attribute,
α_{kp} – impact of p-th state of attribute k,
x_{kpi} – zero-one variable for p-th state of attribute k,
α_j – market value coefficient for j-th location attractiveness zone,
laz_{ji} – dummy variable equal one for j-th location attractiveness zone,
u_i – random component.

The explained variable is a natural logarithm of a real estate unit value. Real estate values are determined by certified appraisers in individual appraisals. Real estate attributes are qualitative characteristics measured on an ordinal scale, so they are introduced into the model (1) through dummy variables for each state of an attribute.

In the model (1) there is a constant term. In order to avoid any strict multi-collinearity of the explanatory variables, each dummy variable for the worst attribute states is skipped.

Hence the summation of $p = 2, ..., k_p$ in the formula (1). In the interpretation, the ignored state of an attribute serves as a point of reference for the remaining states.

There are also market value coefficients (α_j) in the model (1). They could be treated as a proxy for location. They are estimated by introducing dummy variables for each location attractiveness zone. Location attractiveness zones are constructed by experts. They are defined as areas with a similar impact of location. Therefore, location attractiveness zones are constructed in such a way that the impact of location in a given area is homogenous.

Because of the strict collinearity of the explanatory variables the worst (cheapest) location attractiveness zone is skipped. Thus, the omitted location attractiveness zone creates a point of reference.

There are many ways to draw sample elements from the population. The basic classification indicates two basic categories: probability sampling and non-probability sampling.
In the case of probability sampling one uses randomization to ensure that each element is equally likely to be included in the drawn sample. There are several ways of such sampling:

- simple random sampling,
- stratified sampling,
- systematic sampling,
- cluster sampling,
- multi-stage sampling.

In turn, non-probability sampling means that individual objects in the population do not have the same chance of being selected for the sample, and the selection is based on the subjective judgment of the researcher. There are several such ways of drawing:

- convenience sampling,
- quota sampling,
- judgment or purposive sampling,
- snowball sampling.

Chun and Griffith (2013) describes various sampling techniques in the context of spatial research. The authors indicate, among other things, the significant impact of the way the population is sampled on the results of spatial statistical analyses.

This study will use two methods – simple random sampling and stratified sampling. In the case of simple sampling, each property has the same probability of being drawn. In the case of stratified sampling, properties are drawn from strata, which in this case means combinations of states of properties' attributes.

The selection of representative properties is an important stage of valuation carried out using *SAREMA* for two reasons:

1. In the expert variant of *SAREMA,* the values of the representatives are used to determine market value coefficients, which are responsible for the assessment of the locations of the properties subject to valuation. Location is a key feature determining market value.
2. The *SAREMA* variant used in this study, the so-called econometric approach, requires an appropriate number of representative properties and appropriate variability of attributes describing valuation objects.

The accuracy of valuations will be assessed on the basis of a root mean square error (*RMSE*):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(w_i - \hat{w}_i\right)^2}{N}} \quad (2)$$

where:

w_i – the actual unit value of the property determined by the property valuer,

$\hat{w}_i$ – theoretical, unit value of the property determined from the model.

The study of the impact of the sampling scheme of representative properties on the results of the mass valuation was carried out in the following way:

1. Errors of valuations obtained in the original studies (Doszyń, 2020) are presented,
2. Analogously to the set of representatives used in point 1, the representatives were drawn using the scheme of stratified sampling, so that in each of the location attractiveness zones they come from all specified strata of the analyzed real estate. On the basis of such a set, model (1) was estimated and the *RMSE* was calculated for it.
3. Multiple simple random samplings with an increasing number of representatives were carried out (1,000 times for each number of representative properties) and the obtained valuation errors were compared with those of the model estimated in point 2.
4. A multiple stratified sampling of representative properties was carried out, taking into account the decreasing fraction of strata (1,000 times for each considered fraction of strata), and the obtained valuation errors were compared with those of the original model presented in point 1.

Since the aim of the study is not to analyze the models themselves, but only the scale of valuation errors, each model will be described only by the means of the root mean square error.

3. Results

The described *SAREMA* procedure will be used for the appraisal of 318 land plots located in the northern part of Szczecin (known as region 3 within the existing geodetic division), which is the capital of the West Pomeranian voivodeship, one of 16 Polish voivodeships. The real properties constitute a set for which an update of annual perpetual usufruct charges was conducted by municipal authorities. The real properties were located in three clusters (referred as *LAZ*s) of various numbers of real properties. The area within which the appraised real properties are located is shown in Figure 1.

Figure 1. Location of the valued properties

Source: own work.

Attributes and their states are presented in Table 1. It could be noted that all attributes, except surface area, are qualitative variables. They are introduced into the econometric model (1) as a dummy variable for each state of an attribute (with the exclusion of the first, the worst state). Properties' area is a quantitative variable, but it is treated as a qualitative one. This is because market participants often treat this variable in such a way. This conclusion stems from appraisers. With respect to real estate unit value, it is assumed that a small area is better than an average one, and an average area is better than a large one.

Table 1. Real estate attributes and their states

No.	Attribute	Attribute category/symbol
1	Utilities	None Incomplete Complete
2	Neighbourhood	Onerous Unfavourable Average Favourable
3	Transport availability	Unfavourable Average Favourable
4	Physical plot properties	Unfavourable Average Favourable
5	Plot area	Large (>1200 m^2) Average (500–1200 m^2) Small (<500 m^2)

Source: own work.

Unit values of real properties were within the range of 502.11 PLN/m^2 – 701.43 PLN/m^2, with a median equal to 592.28 PLN/m^2. The values of all of the analyzed properties were estimated by property appraisers. The results of their work will be used to assess the quality of valuations obtained by means of *SAREMA*.

In the original study (Doszyń, 2020), out of 318 properties, 30 representative properties were selected by means of simple random sampling, from which the *SAREMA* model was estimated according to formula (1). The model was estimated based on 30 observations. For all 318, theoretical and actual values were compared to determine the element of a mean square error according to formula (2). Actual values were estimated by valuers on a case-by-case basis. The element of average square error for the initial model was 57.29 PLN/m^2.

In order to carry out stratified sampling, the concept of stratum in the test shall be defined. In order to ensure that each type of real estate in each *LAZ* will be present in the set of representative real estate, a group of real estate with the same variants of all of the analyzed attributes was defined as a stratum. Such a definition of strata results in the following number of them in *LAZs*[1]*:*

– zone 13–25 strata,
– zone 14–20 strata,
– zone 15–20 strata.

For the model for which the representative properties were set included a total of 65 properties, *RMSE* was obtained at the level of PLN 32.10/m^2.

The lower error of valuations obtained by using the stratified sampling of representatives is the first indication that this is a better method of sampling. However, this conclusion is distorted by two factors. First of all, both models were built with a different number of observations. In the case of simple random sampling it was 30 representatives, and in the case of stratified sampling – 65. Secondly, sampled properties were drawn in single draws and it can, by chance, affect the results obtained. Therefore, in order to generalize the conclusions, successive stages of the study were carried out. In the case of simple random sampling, the set of representatives was drawn 1,000 times and the number of representatives was gradually increased until the number of representatives was similar to the number of representatives in the stratified sampling (which was 65). 7 set sizes were determined for representative properties – from 30 to 66. As was mentioned before, each time the representatives were drawn a thousand times. This resulted in 7,000 sets and the same number of models were estimated according to formula (1). Valuation errors determined using formula (2) are presented in Figures 2 and 3. The first problem that can be observed concerns single draws, in which very high errors occurred, even exceeding the average value of the properties being valued. This is due to the possible strong collinearity of the explanatory variables or their trace variation for some draws. Valuation errors, including errors in extremely unfavorable draws, decrease with the increase of the number of representatives. The error of 57 PLN/m^2 for the original sample was therefore accidental. The number of 30 representatives for a simple draw should be considered too small.

On the basis of the *RMSE* calculated for each of the models, average errors were calculated for each size of the representative property sets. The results presented in Figure 3 clearly demonstrate two facts. Firstly, the more representative properties there are, the lower the average *RMSE* is. The highest average *RMSE* was PLN 82.76 per square meter. The lowest

[1] The presented study concerns a part of a larger real estate database. The entire database features 17 attractiveness zones of the location. For the three indicated ones, an individual valuation was carried out regarding all the properties located in them, which are designated for an update of the annual perpetual usufruct fee.

equalled 53.99 PLN/m^2. This range indicates a second conclusion. The root mean square error for the model with 65 representatives coming from the stratified sampling was lower than the average *RMSE* in each analyzed case. It can therefore be concluded that stratified sampling is a better way to create a set of representative properties.

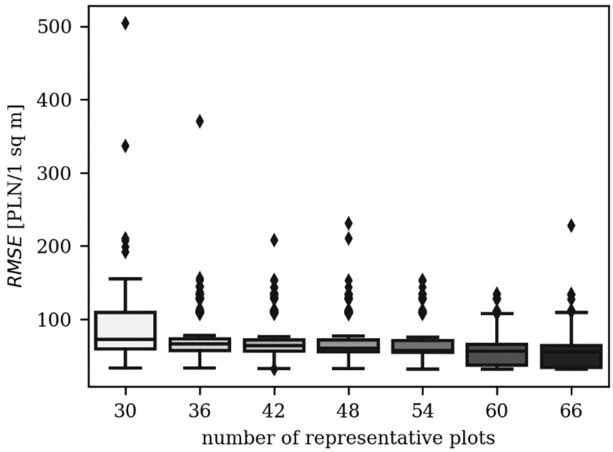

Figure 2. *RMSE* for considered numbers of representative plots drawn with simple random sampling

Source: own work.

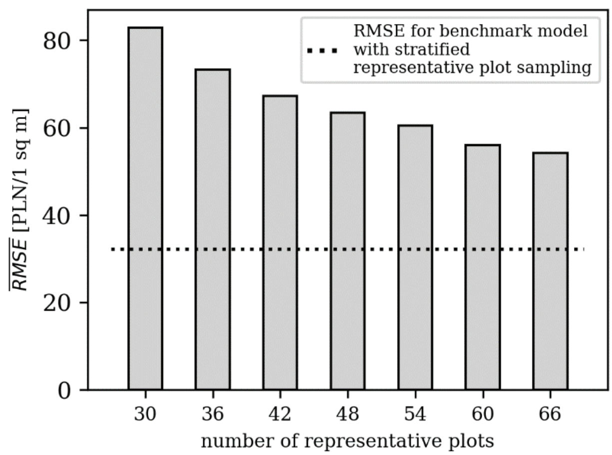

Figure 3. Average *RMSE* for considered numbers of representative plots drawn with simple random sampling

Source: own work.

The process of the mass valuation of real estate, especially in the context of possible general property taxation in Poland, is characterized by an important economic factor, which is the cost of its implementation. One of the elements of this cost is, in the case of *SAREMA*, the cost of the valuation of representative properties carried out by valuers. One faces a trade-off between valuation errors and a number of representative properties. Since stratified sampling is more advantageous, it is also necessary to check how few representatives can be drawn using this scheme so that valuation errors are acceptably high. At this stage of the survey, representative properties were drawn from a decreasing percentage of strata in individual *LAZs*. The results obtained were compared with the *RMSE* of the initial model, in which the simple random sampling of 30 representative properties was specified. This percentage was reduced from 90 to 40% of the number of strata in each zone. This translated into the number of representative properties ranging from 58 to 26. This means that in an extreme case the number of representative properties is even slightly lower than the base number of 30. Figures 4 and 5 present the errors of valuations obtained in a thousand draws for each of the selected percentages of representative property layers. The box diagrams presented in Figure 5 show, first of all, that stratified sampling is much less sensitive to extremely unfavorable draws. The *RMSE* in such cases was much lower than in the case of simple sampling. There were no situations when errors were higher than the average actual value of the analyzed properties.

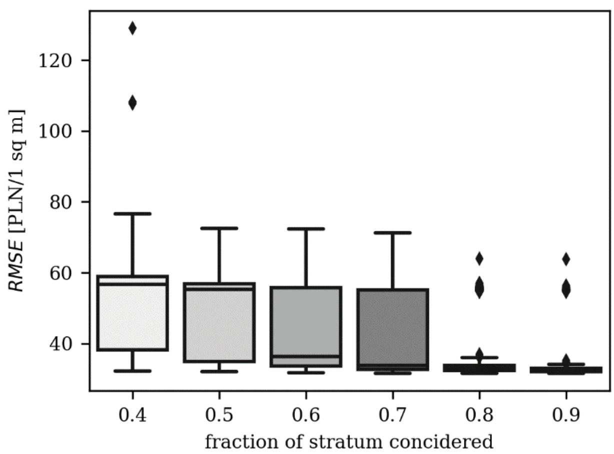

Figure 4. *RMSE* for considered numbers of representative plots drawn with stratified sampling

Source: own work.

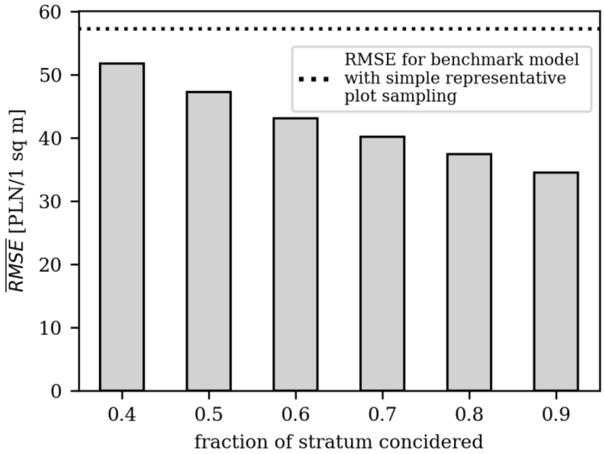

Figure 5. Average *RMSE* for considered numbers of representative plots drawn with stratified sampling

Source: own work.

Valuation errors increase as the number of strata included in the samples decreases. However, as can be seen in Figure 5, there was no case with an average *RMSE* greater than the error for the base model. The stratified sampling allows for better (lower) valuation errors with fewer representative properties. This type of drawing is also more resistant to randomly unfavorable sets of representative properties.

Conclusions

The article presents the problem of the selection of representative real estate drawn for the purposes of the Szczecin's Algorithm of Real Estate Mass Appraisal. The described algorithm can be used in one of the following three approaches – expert, statistical and an econometric one. The stage of selecting representative properties is particularly important for the statistical and econometric approach. In an expert approach, the impact of a given attribute states on the value of a property is determined based on expert knowledge, e.g. using the *AHP* method (e.g. Dmytrów, Gnat 2019). The statistical approach is based on the measures of the correlation between the values of representative plots and their attributes. In the econometric approach, a model is estimated, in which estimations of structural parameters illustrate the relationship between a value of representative properties and individual states of attributes (independent variables). In order to determine the correlation coefficients or to estimate parameters of the econometric model, it is necessary to provide an adequate amount of data from representative properties. Used in the classic *SAREMA* expert variant, simple random sampling can provide a sufficiently large set, but will not provide a sufficient diversity of data. The conducted numerical experiment showed that in the case of simple random sampling, there can be a set of properties drawn, whose attributes will feature collinearity or a lack of appropriate variability. This causes problems with estimating the econometric model and generates high valuation errors. The stratified sampling, on the other hand, allows for drawing a set of representative properties which eliminates these problems. Moreover, it has also been evidenced that even partial representativeness (simulated by the percentage of strata participating in a draw) is a better solution than simple sampling. This is an important economic conclusion, as fewer representatives translate into the lower cost and shorter time of valuation, which has an economic aspect as well. In order to confirm the results obtained in the survey, further research on the impact of the sampling of representative real estate will be conducted on another real estate market.

References

Antipov, E.A., Pokryshevskaya, E.B. (2012). Mass appraisal of residential apartments: An application of Random forest for valuation and a CART-based approach for model diagnostics. *Expert Systems with Applications*, *39* (2), 1772–1778. DOI: 10.1016/j.eswa.2011.08.077.

Benjamin, J.D., Randall, S., Guttery, R.S., Sirmans, C.F. (2004). Mass Appraisal: An Introduction to Multiple Regression Analysis for Real Estate Valuation. *Journal of Real Estate Practice and Education*, *7* (1), 65–77. DOI: 10.1080/10835547.2004.12091602.

Burgard, J.P., Münnich, R., Zimmermann, T. (2014). The Impact of Sampling Designs on Small Area Estimates for Business Data. *Journal of Official Statistics*, *30* (4). 749-771. DOI: 10.2478/jos-2014-0046.

Chun & Griffith (2013). *Spatial statistics and geostatistics*, SAGE: Los Angeles.

Ćetković, J., Slobodan, L., Lazarevska, L., Žarković, M., Vujošević, S., Cvijović, J., Gogić, M. (2018). Assessment of the Real Estate Market Value in the European Market by Artificial Neural Networks Application. *Complexity 2018*. DOI: 10.1155/2018/1472957.

Dell, G. (2017). Regression, Critical Thinking, and the Valuation Problem Today. *Appraisal Journal*, *85* (3), 217–230.

Dmytrów, K., Gnat, S. (2019). Application of AHP Method in Assessment of the Influence of Attributes on Value in the Process of Real Estate Valuation, *Real Estate Management and Valuation*, *27* (4), 15–26. DOI: 10.2478/remav-2019-0032.

Doszyń, M., Gnat, S., Bas, M. (2017). The Econometric Procedures of Specific Transactions Identification. *Folia Oeconomica Stetinensia*, *17* (1), 20–30. DOI: 10.1515/foli-2017-0002.

Doszyń, M. (2012). Ekonometryczna wycena nieruchomości. *Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania Uniwersytetu Szczecińskiego*, *26*, 41–52.

Doszyń, M. (ed.) (2020). *System kalibracji macierzy wpływu atrybutów w szczecińskim algorytmie masowej wyceny nieruchomości*. Szczecin: Wydawnictwo Naukowe Uniwersytetu Szczecińskiego.

Hozer, J., Gnat, S., Kokot, S., Kuźmiński ,W. (2019). The Problem of Designating Elementary Terrains for the Purpose of Szczecin Algorithm of Real Estate Mass Appraisal. *Real Estate Management and Valuation*, *27* (3), 42–58. DOI: 10.2478/remav-2019-0024.

Etikan, I., Musa, S.A., Alkassim, R.S. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics*, *5* (1), 1–4. DOI: 10.11648/j.ajtas.20160501.11.

Hozer, J., Kokot, S., Kuźmiński, W. (2002), *Methods of statistical market analysis in real estate valuation.* Warsaw: Polish Federation of Valuers' Associations.

Isakson, H.R. (1998). The Review of Real Estate Appraisals Using Multiple Regression Analysis. *Journal of Real Estate Research*, *15* (2), 177–190.

Jahanshiri, E., Buyong, T., Shariff, A.R.M. (2011). A review of Property Mass Valuation Models. *Pertanika Journal of Science & Technology*, *19*, 23–30.

Kauko, T., d'Amato, M. (2008). *Mass Appraisal Methods: An International Perspective for Property Valuers*. Blackwell Publishing Ltd.

Kilpatrick, J. (2011). Expert Systems and Mass Appraisal. *Journal of Property Investment & Finance*, *29* (4/5), 529–550. DOI: 10.1108/14635781111150385.

Lahiri, P. (2003). On the Impact of Bootstrap in Survey Sampling and Small-Area Estimation. *Statistical Science*, *18* (2), 199–210. DOI: 10.1214/ss/1063994975.

McCluskey, W.J., Daud, D.Z., Kamarudin, N. (2014). Boosted regression trees: An application for the mass appraisal of residential property in Malaysia. *Journal of Financial Management of Property and Construction*, *19* (2), 152–167. DOI: 10.1108/JFMPC-06-2013-0022.

Rao, J.N.K. (2011). Impact of Frequentist and Bayesian Methods on Survey Sampling Practice: A Selective Appraisal. *Statistical Science*, *26* (2), 240–256. DOI: 10.1214/10-STS346.

Schreuder, H.T., Gregoire, T.G., Weyer, J.P. (2001). For what applications can probability and non-probability sampling be used? *Environmental Monitoring and Assessment*, *66*, 281–291. DOI: 10.1023/A:1006316418865.

Wang, X., Wen, J., Zhang, Y., Wang, Y. (2014). Real estate price forecasting based on SVM optimized by PSO. *Optik*, *125* (3), 1439–1443. DOI: 10.1016/j.ijleo.2013.09.017.

Zurada, J., Levitan, A.S., Guan, J. (2011). A Comparison of Regression and Artificial Intelligence Methods in a Mass Appraisal Context. *Journal of Real Estate Research*, *33* (3), 349–387. DOI: 10.1080/10835547.2011.12091311.