Wiadomości Statystyczne. The Polish Statistician, 2023, *68*(10), 24–48 https://doi.org/10.59139/ws.2023.10.2
Statystyka w praktyce / Statistics in practice
Zgłoszony/submitted: 20.09.2022, poprawiony/revised: 16.05.2023, zaakceptowany/accepted: 4.07.2023

Mass appraisal: a statistical approach to determining the impact of a property's attributes on its value

Anna Gdakowicz,[a] Ewa Putek-Szeląg[b]

Abstract. The mass valuation of real estate refers to the simultaneous estimation of the values of a large number of properties using the same method. This method should involve automation that would reduce the human element in the process. The algorithm that meets these requirements is the Szczecin Mass Valuation Algorithm for Real Estate (SAMWN), which was used to determine the values of selected land properties in Szczecin. The article presents a modification of the SAMWN which consists in an objective calculation of the influence of the attributes of a property on its value using dependency coefficients. Various approaches have been proposed to assigning weights to the attributes included in the model. The aim of the study was twofold. Firstly, to identify the coefficients of dependencies that can be used in property valuation based on SAMWN and in the analysis of a property using attributes measured on an ordinal scale. The second aim was to select such a combination of methods for taking into account the influence and weights of attributes which under the SAMWN procedure would produce results closest to property values determined by real estate appraisers. The study used data on 405 land properties located in Szczecin intended for residential purposes, valued individually by real estate appraisers for the purpose of the research.

The study proposes four types of dependency coefficients and their partial values and four ways of including these coefficients in the SAMWN procedure. Additionally, the study assesses six methods of weighing the proposed measures. As a result, 168 ways of measuring the influence of individual attributes on property value were obtained. In order to determine which variants produced values closest to the real values (estimated by real estate appraisers), appraisal error measures were calculated and linear ranking procedures were then adopted to identify the best combination of the applied variants.

The presented mass valuation algorithm may be applied to the estimation of values of various types of properties. However, it requires the procedure to be adapted to the specific characteristics of the appraised property, i.e. the attractiveness zones should be determined as well as the attributes that are relevant to the specific type of property.

Keywords: mass real estate valuation, statistical methods, mass valuation algorithm, dependency coefficients, linear ordering

JEL: C10, R30

[a] Uniwersytet Szczeciński, Instytut Ekonomii i Finansów, Polska / University of Szczecin, Institute of Economics and Finance, Poland. ORCID: https://orcid.org/0000-0002-4360-3755.
E-mail: anna.gdakowicz@usz.edu.pl.

[b] Uniwersytet Szczeciński, Instytut Ekonomii i Finansów, Polska / University of Szczecin, Institute of Economics and Finance, Poland. ORCID: https://orcid.org/0000-0003-0364-615X. Autor korespondencyjny / Corresponding author: ewa.putek-szelag@usz.edu.pl.

Masowa wycena nieruchomości. Statystyczny sposób określenia wpływu cech nieruchomości na jej wartość

Streszczenie. O masowej wycenie nieruchomości mówi się, gdy wartości dużej liczby nieruchomości są szacowane w tym samym czasie przy użyciu tej samej metody. Zastosowana metoda powinna charakteryzować się automatyzacją ograniczającą udział człowieka. Założenia te spełnia Szczeciński Algorytm Masowej Wyceny Nieruchomości (SAMWN), wykorzystywany do wyceny wartości wybranych nieruchomości gruntowych w Szczecinie. W artykule przedstawiono modyfikację SAMWN, polegającą na obiektywnym obliczaniu wpływu cech nieruchomości na jej wartość z wykorzystaniem współczynników zależności. Ponadto zaproponowano różne sposoby nadawania wag atrybutom użytym w modelu. Cel badania omawianego w artykule jest dwojaki. Po pierwsze polega na wskazaniu tych współczynników zależności, które mogą być użyte do wyceny nieruchomości opartej na SAMWN oraz do analizy nieruchomości wykorzystującej atrybuty mierzone na skali porządkowej. Po drugie celem jest wyłonienie takiej kombinacji metod uwzględniania wpływu i wag atrybutów, której zastosowanie w procedurze SAMWN dałoby wyniki najbliższe wartościom nieruchomości oszacowanym przez rzeczoznawców majątkowych. W badaniu wykorzystano dane dotyczące 405 nieruchomości gruntowych w Szczecinie przeznaczonych na cele mieszkaniowe, które na potrzeby badania zostały indywidualnie wycenione przez rzeczoznawców.

Zaproponowano cztery rodzaje współczynników zależności i ich wartości cząstkowych oraz cztery sposoby uwzględniania tych współczynników w procedurze SAMWN. Ponadto przeprowadzono ocenę zastosowania sześciu metod ważenia proponowanych miar. W efekcie uzyskano 168 sposobów pomiaru wpływu poszczególnych atrybutów na wartość nieruchomości. W celu określenia, w którym przypadku uzyskane wartości wyceny są najbardziej zbliżone do wartości rzeczywistych (oszacowanych przez rzeczoznawców majątkowych), obliczono miary błędów wyceny, a następnie wdrożono procedury porządkowania liniowego, aby wskazać najlepszą kombinację rozpatrywanych wariantów.

Przedstawiony algorytm wyceny masowej może być zastosowany do szacowania wartości różnego rodzaju nieruchomości. Wymaga to jednak dostosowania procedury do specyfiki wycenianej nieruchomości, tj. określenia strefy atrakcyjności i atrybutów istotnych dla danego typu nieruchomości.

Słowa kluczowe: masowa wycena nieruchomości, metody statystyczne, algorytm wyceny masowej, współczynniki zależności, porządkowanie liniowe

1. Introduction

One of the problems occurring when estimating the value of a property is determining how the individual attributes of a property affect its value. Until now, this was estimated either on the basis of analyses of prices and market features of similar properties, or by analogy to other properties in terms of the type and area of markets, by examining or observing the preferences of potential buyers, or in other ways (Polska Federacja Stowarzyszeń Rzeczoznawców Majątkowych, 2008). This article refers to the last option and presents an alternative way of dealing with the aforementioned issue. An approach based on statistical methods, using dependency coefficients is introduced as an objective proposal to determine the impact of individual attributes on the value of real estate. The idea of including statistical methods (dependency coefficients) in the procedure of mass real estate valuation is the basis of the statistical approach of the Szczecin Mass Real Estate Valuation Algorithm (Pol. *Szczeciński Algorytm Masowej Wyceny Nieruchomości* – SAMWN).

Considering the specificity of the features (attributes) describing real estate, which in most cases may be presented on an ordinal scale, four coefficients measuring the strength of the relationship between the features and the value of the property were proposed. Since previous studies, e.g. Dmytrów et al. (2020), indicated that the results obtained for partial coefficients reflect the relationships found on the market more efficiently, the analysis was extended to include partial variants of selected coefficients. Then, six procedures for calculating weights in SAMWN were proposed. The obtained results were subsequently ordered on the basis of linear ordering procedures, as described in the work by Strahl (1978) and Pawlukowicz (2010).

The aim of the article was twofold. Firstly, was to identify the coefficients of dependencies that can be used in property valuation based on SAMWN, and also in the analysis of those properties whose attributes are often described on an ordinal scale. The second aim was to identify a combination of methods describing the influence and weights of the attributes which under SAMWN procedure produces results closest to the property values determined by real estate appraisers.

2. Mass appraisal methods

2.1. Literature review

According to the literature, mass property valuation (Jahanshiri et al., 2011) requires the following conditions to occur simultaneously (Hozer et al., 2002; Kuryj, 2007; Telega et al., 2002):

- a large number of properties are subject to valuation;
- the properties are valued using the same method (making the results comparable);
- all properties must be valued at the same time, using the same data and price levels.

Since the first definition of mass appraisal was formulated in 1984 (International Association of Assessing Officers, 2019), there have been many theoretical ideas for the simultaneous valuation of multiple properties, followed by attempts to apply these ideas in practice. Scientists unanimously agreed that it was necessary to develop a method or model whose application would show the relationship between the property value and the factors affecting it. While statistical, geographical or computer science-based models were used for mass valuations, the proposed solutions can be classified as the 3I-trend: the AI-based model, the GIS-based model and the MIX-based model (Wang & Li, 2019).

The AI-based model includes:

- multiple regression analysis (MRA) – a relatively simple and easy-to-use method, although problematic in terms of choosing the analytical form of the function that best describes the relationship between the value and the selected attributes; additionally, the estimation error is often unsatisfactory (Lin & Mohan, 2011);
- an expert system and a decision support system – systems based on the knowledge of property appraisers. These systems do not learn by themselves but use expert knowledge to develop adjustment factors (Amidu & Boyd, 2018; Kilpatrick, 2011; Lam et al., 2009);
- artificial neural networks (ANNs) – the use of neural networks in the mass valuation process does not require the development of an initial model or validation. Based on a learning sample, the network learns by itself and provides the final results. Despite the fact that the results are at a satisfactory level of matching, researchers pay attention to a part of the ANN procedure called the 'black box' (Abidoye & Chan, 2018; Yacim & Boshoff, 2018; Zhou et al., 2018);
- tree-based model – a group of models based on a decision tree, random forest and boosted tree. These methods perform well in both classification and regression. The results obtained by using these methods are more accurate compared to other models (Antipov & Pokryshevskaya, 2012) and the computational process is fast. However, interpreting the results is problematic (McCluskey et al., 2014);
- hierarchical model – a model which takes into account the hierarchical structure of data. The hierarchical Bayesian approach (Hui et al., 2010), the analytical Bayesian approach (Cervelló-Royo et al., 2016) and the analytical hierarchical process have been used for property valuation;
- cluster analysis – the use of data grouping methods is also applied in mass real estate valuation. The similarity of real estate allows for the isolation of clusters of similar properties; however, it is important to explain the resulting subgroups in practice. Emphasis should also be placed on adopting correct assumptions according to which the cluster analysis process runs, as changing the assumptions changes the clusters. There are several types of classifications: model-based clustering, partitioning clustering, density-based clustering, hierarchical clustering, grid-based clustering, and fuzzy-based clustering (Calka, 2019; Gabrielli et al., 2017; Napoli et al., 2017);
- rough set theory (RST) and fuzzy set theory – methods used for underdeveloped, emerging markets with a low level of computerisation. The use of RST in the real estate market enables mass valuation even when the relevant factors influencing the property value are unknown, and can be used to adjust data weights even if the similarity is at a low level (Alcantud et al., 2017; Guan et al., 2014; Lasota et al., 2011; Ma et al., 2018);
- other models – a group of various other classic models, e.g. the genetic algorithm (GA), the support vector machine (SVM), data envelopment analysis (DEA), and

conformal predictors (CP). The use of these models in the real estate market has only just begun, but their application has already yielded promising results (Ahn et al., 2012; Bellotti, 2017; Chen, Z. et al., 2017; Morano et al., 2018; Zurada et al., 2011).

The GIS-based models involve the use of the geoinformation system/science (GIS). Since an attribute related to its spatial position can be created for each property, scientists take advantage of this possibility and combine the information with other attributes that can affect the property's value. Among these models, the following stand out:

- geographically weighted regression (GWR) – the most commonly used GIS model; it is an extension of the MRA model with spatial analysis. It enables regression analysis for each location. Models of this type are easy to apply and the obtained results allow for an interpretation (Dimopoulos & Moulas, 2016; Lockwood & Rossini, 2011; McCluskey & Borst, 2011);
- geographically weighted principal component analysis – models that refer to principal component analysis (PCA) methods, extended to include spatial elements, allowing the analysis to take into account e.g. the presence of submarkets (Wu et al., 2018);
- spatial error model (SEM) and spatial lag model (SLM) – the most frequently used extension of MRA models with spatial dependence. The SEM assumes that the occurrence of an error in the property valuation depends on the error of its surroundings (Zhang et al., 2015), while the SLM uses the spatially lagged dependent variable of the regression model: the price of a property depends on the prices of the neighbouring properties (Anselin, 2002).

The group of MIX-based models includes a wide range of models (constantly developed) that use (mix) simultaneously different approaches. Researchers propose e.g. forecasting according to different methods and, based on the obtained results, determining the weighted average of the component forecasts (Glennon et al., 2018), combining traditional estimation models with AI and GIS methods (Calka & Bielecka, 2016), or using information from other innovative sources in known models (Chen, J.-H. et al., 2017).

2.2. The SAMWN

As mentioned above, mass valuation involves appraising multiple properties at the same time using the same tool (algorithm). To ensure the comparability of the results and that the generalisation of the calculations is possible, the tool which is used should be automated and the impact of the human factor should be limited to a minimum. SAMWN is such a tool. The algorithm is classified as a MIX-based model, as it represents a non-classical approach. On the one hand, the knowledge of real estate appraisers is used (expert system and decision support system), on the other hand, however, the structure of the model refers to hierarchical solutions, and the entire valuation process is automated. SAMWN can be presented as follows:

$$w_{ji} = wwr_j \cdot pow_i \cdot w_{\mathrm{baz}} \cdot \prod_{k=1}^{K} \prod_{p=1}^{k_p} \left(1 + A_{kpi}\right), \tag{1}$$

where:
w_{ji} is the i-th market value (or cadastral value) of the property in the j-th zone of location attractiveness ($j = 1, 2, \ldots, J$),
wwr_j is the market value coefficient in the j-th zone of location attractiveness,
J is the number of the location attractiveness zone,
pow_i is the area of the i-th real estate,
w_{baz} is the estimated value of 1 sqm of the property with the worst attribute states in the worst location attractiveness zone,
A_{kpi} is the impact of the p-th category of the k-th attribute for the i-th real estate ($k = 1, 2, \ldots, K;\ p = 1, 2, \ldots, k_p$),
K is the number of attributes,
k_p is the number of categories of the k-th attribute,
N is the number of valued properties ($i = 1, 2, \ldots, N$).

The algorithm determines the unit market or cadastral value of the property, not the price. Since the algorithm design (formula 1) requires the multiplication of individual factors, the reference point when determining the value of the property is the base value, i.e. the value of 1 sqm of the property with the worst attribute states, in the worst location attractiveness zone, theoretically of the lowest value. The base value is multiplied by the area of the property, the market value coefficient and the effect of the attribute states of the valued property.

The impact of attribute states (A_{kpi}) can be determined e.g. on the basis of quantitative methods – using statistical methods (Gdakowicz & Putek-Szeląg, 2020a, 2020b). Thus, the objectivity of the procedure is retained and the human factor is limited to a minimum.

The value of a property depends not only on its attributes but also on external factors presented by the demand side. Two properties, very similar in terms of attributes, can have different values if they are located in different zones of location attractiveness (Pol. *strefa atrakcyjności lokalizacji* – SAL). The algorithm influences this type of factor through market value coefficients (wwr_j). These coefficients are determined for each location's attractiveness zone and they show the impact of the location. The market value ratio for the j-th zone of location attractiveness can be determined as

$$wwr_j = \sqrt[n_j]{\prod_{i=1}^{n_j} \frac{w_{ji}^{\mathrm{rz}}}{w_{ji}^{\mathrm{h}}}}, \tag{2}$$

where:
w_{ji}^{rz} is the value of the i-th property in the j-th zone of location attractiveness determined by a property appraiser,
w_{ji}^{h} is the hypothetical value of the i-th property in the j-th zone of location attractiveness,
n_j is the number of representative properties in the j-th zone of location attractiveness.

The market value ratios are calculated based on representative real estate valuations carried out by property appraisers on an individual basis, providing real property values (w_{ji}^{rz}). On the other hand, hypothetical values (w_{ji}^{h}) are calculated based on formula 1, but excluding market value coefficients:

$$w_{ji}^{\mathrm{h}} = pow_i \cdot w_{\mathrm{baz}} \cdot \prod_{k=1}^{K} \prod_{p=1}^{k_p} \left(1 + A_{kpi}\right). \tag{3}$$

When the values of representative real estate (w_{ji}^{rz}) are known, as are their attribute states and their impact, their base value (w_{baz}) and the surfaces, the market value coefficients are estimated for each zone of location attractiveness. This involves calculating the geometric mean of the quotients of the real and hypothetical values of the real estate. As the market value coefficients for individual zones of location attractiveness become known, it is possible to estimate the market (cadastral) value of each property within the SAL taking into account the attribute states of the valued property.

The impact of the attributes on the value of a real estate can be calculated according to a statistical approach using dependency coefficients. It should, however, be noted that the use of statistical methods is always limited by certain formal requirements. Therefore, when examining the correlations between attributes in the real estate market, the following problems may occur (Dmytrów et al., 2020):

- the attributes describing real estate subject to valuation are very often qualitative, i.e. measured on an ordinal scale;

- the differences between the successive values of the attributes measured on the ordinal scale are not necessarily constant (changes are not linear);
- when examining the correlations between the value of 1 sqm and the individual attributes, it is necessary to eliminate the influence of other attributes which can significantly disrupt the examined interdependence.

Real estate attributes are most often measured on an ordinal scale, which is why rank-based ratios are the most natural to use, i.e. Spearman's rank correlation coefficient (Kendall, 1948), the Kendall coefficient (Han & Zhu, 2008; Parker et al., 2011) or generalised correlation coefficient Gamma (Siegel & Castellan, 1988).

In practice, however, property appraisers most often use Pearson's linear correlation coefficients[1] in the property valuation process, although they should not be used at all for such features. Partial coefficients have also been calculated for the above-mentioned coefficients to eliminate the influence of other variables on the attribute–property value relationship. If the value of the coefficient is lower than 0, then it is assumed that this attribute has an insignificant influence on the value of the property and, in the further calculations, 0 is assumed.

In addition, four variants were introduced to take into account the impact of the factors on the value of the property. Only statistically significant coefficients or the square of these coefficients and all coefficients (greater than 0), or their square have been considered in the further calculations. Detailed variants of the use of dependence coefficients are presented in Table 1.

Table 1. Dependence coefficients – calculation variants

Dependence coefficients	Significant coefficients	Square of significant coefficients	All coefficients	Square of all coefficients
Spearman's	Si.	Si2.	Sw.	Sw2.
Partial Spearman's	Sci.	Sci2.	Scw.	Scw2.
Kendall's	Ti.	Ti2.	Tw.	Tw2.
Partial Kendall's	Tci.	Tci2.	Tcw.	Tcw2.
Gamma	Gi.	Gi2.	Gw.	Gw2.
Pearson's	ri.	ri2.	rw.	rw2.
Partial Pearson's	rci.	rci2.	rcw.	rcw2.

Source: authors' work.

[1] It should be emphasised that the use of this coefficient is methodologically incorrect (property features / attributes are presented on an ordinal scale, while changes between feature variants are not linear); however, due to the widespread use of this measure by property appraisers, it has been included in the further calculations.

For each of the 28 methods of calculating the coefficients (Table 1), six methods of calculating the weights necessary in the $1 + A_{kpi}$ (formula 1) determination process were proposed (the number of the used method of calculating the weights was added to each coefficient, after the dot):

1. the average value ratio of the real estate with the best attribute states to the average value of the real estate with the weakest attribute states:

$$1 + a_{kp} = \mathrm{e}^{\frac{w_{\mathrm{Amax}k}}{w_{\mathrm{Amin}k}} u_{kp}}, \tag{4}$$

where:

$$u_{kp} = \frac{\rho}{\sum_{k=1}^{K} \rho} \cdot \frac{(p-1)}{(k_p - 1)}, \tag{5}$$

where:

u_{kp} is the weight of categories p of the k-th attribute (used also in methods described in items 2–6),

ρ is the dependence coefficient,

p is the number of states of the k-th attribute ($p = 1, \ldots, k_p$);

2. the median value ratio of the property with the best attribute states to the median value of the property with the weakest attribute states:

$$1 + \alpha_{kp} = \mathrm{e}^{\frac{w_{\mathrm{Mmax}k}}{w_{\mathrm{Mmin}k}} u_{kp}}; \tag{6}$$

3. the average value ratio of the property with the best attribute states to the average value of the property with the weakest attribute states, corrected by factor $\ln\left(\frac{w_{\max}}{w_{\min}}\right)$:

$$1 + \alpha_{kp} = \mathrm{e}^{\ln\left(\frac{w_{\max}}{w_{\min}}\right) \cdot \frac{w_{\mathrm{Amax}k}}{w_{\mathrm{Amin}k}} u_{kp}}; \tag{7}$$

4. the median value ratio of the property with the best attribute states to the median value of the weakest attribute states, corrected by factor $\ln\left(\frac{w_{\max}}{w_{\min}}\right)$:

$$1 + \alpha_{kp} = \mathrm{e}^{\ln\left(\frac{w_{\max}}{w_{\min}}\right) \cdot \frac{w_{\mathrm{Mmax}k}}{w_{\mathrm{Mmin}k}} u_{kp}}; \tag{8}$$

5. the coefficient ratio of the property value variation with the best attribute states to the coefficient of the property value variation with the weakest attribute states:

$$1 + \alpha_{kp} = \mathrm{e}^{\frac{w_{\mathrm{Vmax}k}}{w_{\mathrm{Vmin}k}} u_{kp}}; \tag{9}$$

6. the coefficient ratio of the property value variation with the best attribute states to the coefficient of the property value variation with the weakest attribute states, corrected by coefficient $\ln\left(\frac{w_{\max}}{w_{\min}}\right)$:

$$1+\alpha_{kp}=\mathrm{e}^{\ln\left(\frac{w_{\max}}{w_{\min}}\right)\cdot\frac{w_{\mathrm{Vmax}k}}{w_{\mathrm{Vmin}k}}u_{kp}}, \tag{10}$$

where:

$w_{\mathrm{Amax}k}$, $w_{\mathrm{Mmax}k}$, $w_{\mathrm{Vmax}k}$ are: the average, the median, the variation coefficient of the property value for the highest category of the k-th attribute, respectively,

$w_{\mathrm{Amin}k}$, $w_{\mathrm{Mmin}k}$, $w_{\mathrm{Vmin}k}$ are: the average, the median, the variation coefficient of the property value for the lowest category of the k-th attribute, respectively,

$w_{\max}$ is the maximum value of the property in the data set,

$w_{\min}$ is the minimum value of the property in the data set.

Values $1+\alpha_{kp}$ are determined for all attribute states. The value of $1+A_{kpi}$ corresponds to a specific attribute state of the appraised property.

3. Methodology for selecting the best variant

Based on the applied variants of determining the impact of attributes in real estate valuation using SAMWN, 168 cases of real estate value estimations were received. The obtained real estate values were compared with the real values, valued individually by property appraisers. The measures used were as follows:

1. root means square error (RMSE):

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{n}(w_{\mathrm{r}i}-w_{\mathrm{t}i})^2}{n}}, \tag{11}$$

where:

$w_{\mathrm{r}i}$ is the real unit value of the real estate determined by a real estate appraiser,

$w_{\mathrm{t}i}$ is the theoretical unit value of the property determined by the SAMWN,

n is the number of observations;

2. RMSE variation coefficient:

$$V_{\mathrm{RMSE}}=\frac{\mathrm{RMSE}}{\bar{w}_{\mathrm{r}i}}\cdot 100\%, \tag{12}$$

where:

$\bar{w}_{\mathrm{r}i}$ is the average value of the real unit values of the real estate designated by a property appraiser;

3. the mean absolute percentage error (MAPE):

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|w_{\mathrm{r}i} - w_{\mathrm{t}i}|}{w_{\mathrm{r}i}}; \tag{13}$$

4. based on the percentage error of PE:

$$\text{PE}_i = \frac{w_{\mathrm{r}i} - w_{\mathrm{t}i}}{w_{\mathrm{r}i}} \cdot 100\%; \tag{14}$$

5. share of B+ valuations for which $\text{PE}_i > 0$:

$$\text{B}^+ = \frac{\sum_{i=1}^{n^+} \text{PE}_i^+}{n^+} \cdot 100\%; \tag{15}$$

6. share of B- valuations for which $\text{PE}_i < 0$:

$$\text{B}^- = \frac{\sum_{i=1}^{n^-} \text{PE}_i^-}{n^-} \cdot 100\%, \tag{16}$$

where:
n^+ is the number of observations for which $\text{PE}_i > 0$,
n^- is the number of observations for which $\text{PE}_i < 0$.

The linear ordering of the proposed variants for determining the impact of the attributes on the value of the real estate estimated using the SAMWN was carried out according to the valuation error measures (formulae 11–13 and 15–16). The next steps of the linear ordering procedure included:

1. preparing a data matrix based on the collected data (valuation errors calculated for the proposed variants);
2. the standardisation transformation of the variables (Gatnar & Walesiak, 2011; Zeliaś, 2002);
3. defining the variables as stimulants or destimulants (Hellwig, 1968);
4. applying the upper development pattern (ordering the elements of the set of objects according to the increasing values of the distance measure);
5. calculating the distances between objects by means of the GDM1 measure for variables measured on an interval scale (Walesiak, 2016):

$$d_{ik} = \frac{1}{2} - \frac{\sum_{j=1}^{m}(x_{ij} - x_{kj})(x_{kj} - x_{ij}) + \sum_{j=1}^{m}\sum_{\substack{l=1 \\ l \neq i,k}}^{n}(x_{ij} - x_{lj})(x_{kj} - x_{lj})}{2\left[\sum_{j=1}^{m}\sum_{l=1}^{n}(x_{ij} - x_{lj})^2 \cdot \sum_{j=1}^{m}\sum_{l=1}^{n}(x_{kj} - x_{lj})^2\right]^{\frac{1}{2}}}, \tag{17}$$

where: x_{ij}, x_{kj}, x_{lj} are the i-th, k-th and l-th observation of the j-th variable;

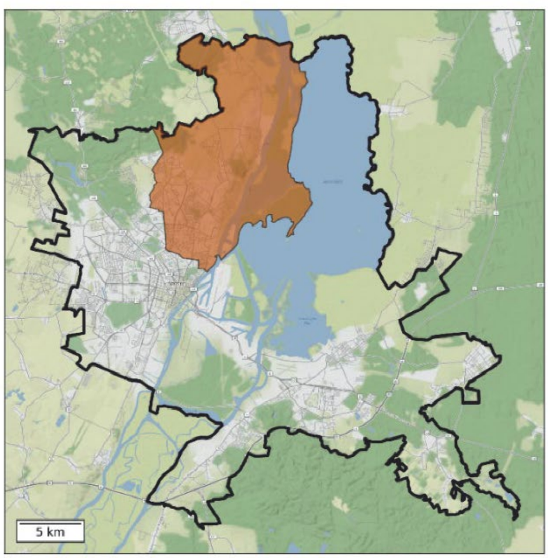
6. a graphical presentation of the results obtained.

All calculations related to the linear ordering were made using the cluster SIM package in the R programme as developed and presented in the work by Walesiak and Dudek (2020).

4. Empirical data

The study used data relating to 405 plots of land in Szczecin, intended for housing purposes, which were located in the northern part of the city (Figure 1). The real estate has been grouped into 17 location attractiveness zones – SALs (Figure 2). The work by Hozer et al. (2019) describes the methodology applied for setting SALs. All properties have been individually valued by property appraisers. The base which consisted of all 405 plots of land was divided into two groups. The first group included 117 properties, which was the learning base. Its composition included real estate representatives from all SALs. The second group, with 288 properties, formed the testing base. The properties on which the algorithm was tested were located in the 13th, 14th and 15th SAL.

Figure 1. The northern region of Szczecin



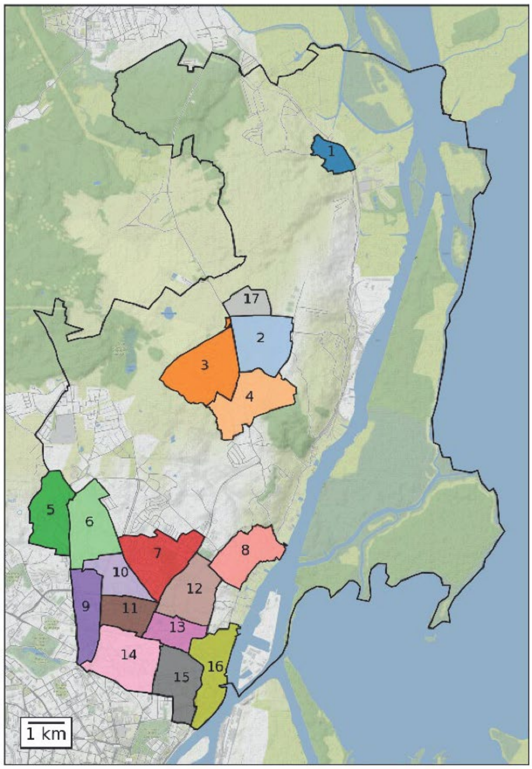
Source: authors' work.

Each real estate was described using the following set of attributes:

- $x1$ is the area: large, average, small;
- $x2$ are the utilities: none, incomplete, full;
- $x3$ is the neighbourhood: troublesome, unfavourable, average, favourable;

- $x4$ is the communication accessibility: unfavourable, average, good;
- $x5$ are the physical features: unfavourable, average, favourable.

Figure 2. The northern region of Szczecin with location attractiveness zones marked

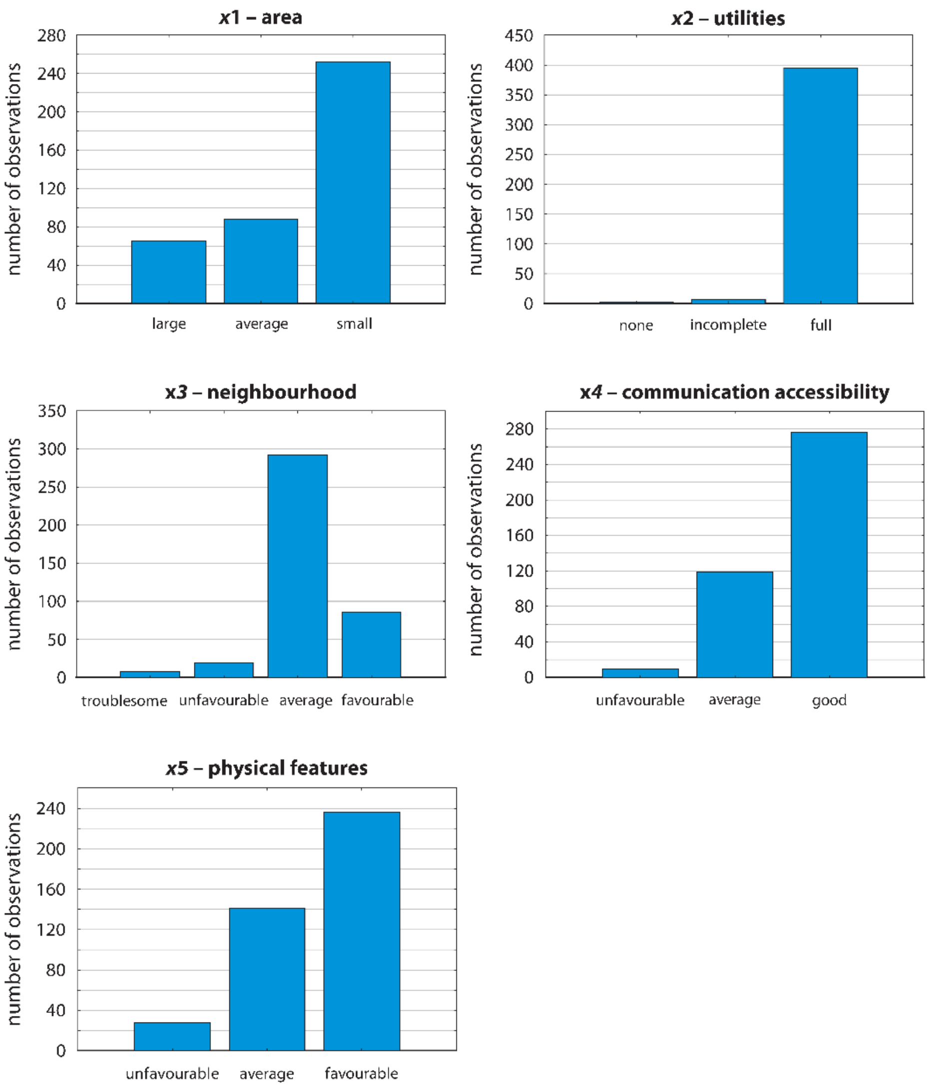


Source: authors' work.

The attributes were presented on ordinal scales, with the value 1 designating the weakest attribute level and subsequent numbers more favourable levels (Figure 3). This method of data encoding should result in a positive correlation between the consecutive attributes and the property value.

Among the properties analysed, those with attributes at the highest levels dominated (Figure 3), except for the neighbourhood attribute, as most properties were situated in average surroundings. Since the analysis concerned land properties located in the city, only a few of them were characterised by none or incomplete infrastructure.

Figure 3. Distributions of individual attribute variants



Source: authors' work.

5. Research results

For the real estate designated as the learning base, seven dependency coefficients between the real estate attributes and their values were calculated. Then, the impact of individual variables on the value of the property $(1 + A_{kpi})$ was determined according to the assumed variants (Table 1) and methods of determining weights (formulae 4 and 6–10). 168 options were received. For each variant based on formula 3, the hypothetical values of the analysed properties were calculated and the market coefficient values were determined (formula 2). Then, the values calculated in wwr_j were used and based on formula 1, the values of properties located within SAL 13, 14, and 15 (the test base) were estimated. In each case, valuation matching measures were calculated (formulae 11–13 and 15–16). A list of fit measures for the first method of determining weights is presented in Table 2.

Table 2. Measurement errors of real estate valuations according to dependency coefficients for the first method of calculating weights (formula 4)

Dependency coefficients	RMSE	V_{RMSE}	MAPE	B^+	B^-
		in %			
Si.1	495.35	82.53	54.87	3.37	–51.50
Si2.1	495.35	82.53	54.87	3.37	–51.50
Sw.1	303.47	50.56	34.97	3.64	–31.33
Sw2.1	303.47	50.56	34.97	3.64	–31.33
Sci.1	106.30	17.71	14.12	13.62	–0.51
Sci2.1	106.30	17.71	14.12	13.62	–0.51
Scw.1	106.30	17.71	14.12	13.62	–0.51
Scw2.1	106.30	17.71	14.12	13.62	–0.51
Ti.1	495.35	82.53	54.87	3.37	–51.50
Ti2.1	495.35	82.53	54.87	3.37	–51.50
Tw.1	311.42	51.89	35.82	3.62	–32.20
Tw2.1	311.42	51.89	35.82	3.62	–32.20
Tci.1	106.33	17.72	14.12	13.61	–0.51
Tci2.1	106.33	17.72	14.12	13.61	–0.51
Tcw.1	106.33	17.72	14.12	13.61	–0.51
Tcw2.1	106.33	17.72	14.12	13.61	–0.51
Gi.1	495.35	82.53	54.87	3.37	–51.50
Gi2.1	495.35	82.53	54.87	3.37	–51.50
Gw.1	313.49	52.23	36.05	3.62	–32.43
Gw2.1	313.49	52.23	36.05	3.62	–32.43
ri.1	193.63	32.26	21.89	3.47	–18.42
ri2.1	193.63	32.26	21.89	3.47	–18.42
rw.1	174.77	29.12	20.33	3.36	–16.97
rw2.1	174.77	29.12	20.33	3.36	–16.97
rci.1	495.35	82.53	54.87	3.37	–51.50
rci2.1	495.35	82.53	54.87	3.37	–51.50
rcw.1	98.09	16.34	13.90	6.42	–7.48
rcw2.1	98.09	16.34	13.90	6.42	–7.48

Source: authors' work.

In many cases, the variants used have produced the same results, which is a direct result of the number of dependency factors taken into account. The adopted assumption that dependencies on the negative direction are ignored narrowed the number of coefficients used in the subsequent steps. Only one relationship was statistically significant (for all coefficients): between communication accessibility and the value of the property, and in the case of partial coefficients (for Spearman's, Kendall's and the Gamma coefficient) two relationships: between the utilities and surroundings and the value of the property. According to the Pearson coefficient, the relationship between utilities, transport accessibility, the environment and the value of the property was statistically significant (and for the partial coefficient, all relationships were statistically significant).

From the point of view of the results obtained, it was irrelevant to include the squares of the relationship between the attributes and the value of the property in the further calculations, as the results obtained for both variants were the same. This was the case for Kendall and Spearman's partial coefficients: the fit measures were the same for all four options. The further calculations omitted 54 variants that generated the same results.

The proposed measures of errors between the real values of the real estate and the values estimated based on SAMWN showed a wide variation resulting from the approach used. The basic descriptive statistics of the received errors are summarised in Table 3.

Table 3. Descriptive statistics of the valuation errors

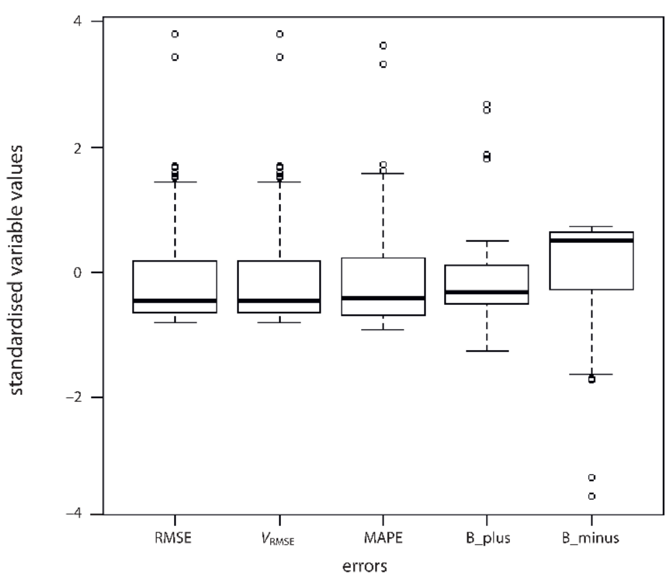
Selected descriptive statistics	RMSE	V_{RMSE}	MAPE	B^+	B^-
		in %			
Minimum	46.56	7.76	6.23	1.99	−161.07
Maximum	1753.73	292.19	165.32	17.50	−0.51
Arithmetic mean	313.35	52.21	33.79	5.70	−28.09
Standard deviation	378.77	63.11	36.16	4.37	37.08
Variation coefficient	120.88	120.88	107.01	76.64	132.00
Median	133.53	22.25	18.07	4.20	−9.07
Semi-interquartile range	187.38	31.22	18.69	1.42	18.92
Positional variation coefficient	140.33	140.33	103.43	33.70	208.72

Source: authors' work.

The most decent level of the analysed errors was as close as possible to 0, which would indicate no differences between the value estimated by property appraisers and the value estimated by SAMWN. The range of the obtained results was very large (Table 3). The smallest observed MAPE value was 6.23%, while the highest was 165.32%. The coefficient of variation for this measure was over 100% (also in a narrowed area of variation). Moreover, RMSE and V_{RMSE} errors were characterised by very high variability (over 120% and 140% in the narrowed area of variability). The B^+ error indicates in what percentage the values estimated by the model are higher than the actual values. Similarly, the B^- error provides information about the values estimated by the model below the actual values. It turns out that the used variants caused a significant shift in the obtained results in plus or in minus (the range between the errors for individual variants extended from 6% to even 165%). This is not a desirable occurrence, because after all, the use of SAMWN in mass valuation aims to achieve results close to the real values, calculated in individual appraisals. Therefore, when selecting the best variant of the algorithm, not only the valuation errors related to the differences between the actual and estimated value (formulae 11–13) should be taken into account, but also the errors indicating how much the model underestimates or overestimates the value of the property (formulae 15–16).

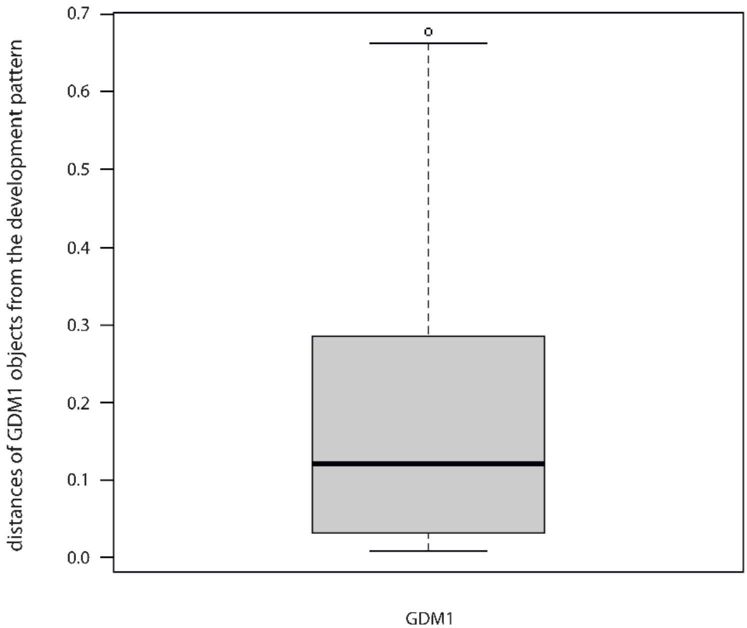
The procedure of linear ordering of the used dependency coefficients according to the received valuation errors was carried out. All variables were characterised by very high variability (in most cases the variation coefficient exceeded 100%). To achieve comparable variables, standardised variables were derived. The first four variables were designated as destimulants whose smaller value is desirable, and B^{-} as a stimulant. The reference level for all variables was 0, i.e. the most desirable level of the selected variables. Then distance measure GDM1 (formula 17) and the upper development pattern were used. Figure 4 presents the evolution of the variables after standardisation, while Figure 5 shows the diversity of the synthetic variable calculated using the GDM1 distance measure.

Figure 4. Error differentiation after standardisation



Source: authors' calculations.

Figure 5. Differentiation of a synthetic (aggregate) variable covering GDM1 distances from the development pattern for all SAMWN estimation variants

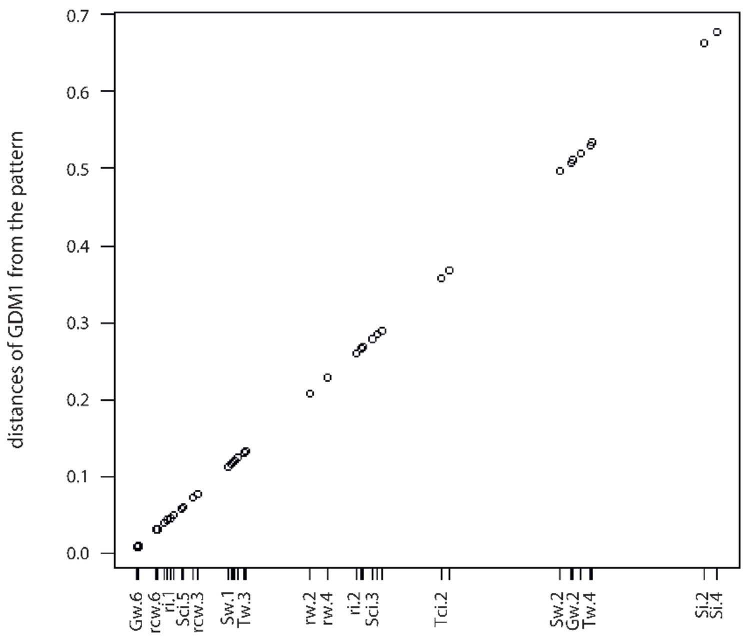


Source: authors' calculations.

Distributions of valuation errors indicated strong right-hand asymmetrical distributions for RMSE, V_{RMSE}, and the MAPE variables (Figure 4). Atypical values occurred, which indicated that in some of the proposed variants the obtained valuations significantly differed from the actual values. The valuation error deviating in plus was also distinguished by right-hand asymmetry, but proved weaker than in the case of the previous errors. Unusual values were also observed in this distribution. The left-hand asymmetry was characterised by the B^{-} error distribution, which is related to the nature of this variable being a stimulant. The aggregate variable calculated based on GDM1 distances showed right-hand asymmetrical distribution, but with no atypical values (Figure 5).

Figure 6 presents the results of linear ordering and Table 4 lists the three best and three worst (according to GDM1) valuation options using appropriate relationship factors.

Figure 6. Ordering of the dependence coefficients from the best to the worst due to error valuations obtained using SAMWN



Source: authors' calculations.

Table 4. Ordering of the dependence coefficients from the best to the worst according to the GDM1 measure value

Dependence coefficients	GDM1 measure	Dependence coefficients	GDM1 measure
Best		**Worst**	
Gw.6	0.008	Gw.4	0.534
Tw.6	0.008	Si.2	0.662
Sw.6	0.008	Si.4	0.677

Source: authors' calculations.

The least distant from the standard (representing the best results) were cases calculated based on the Gamma, Kendall and Spearman coefficients and taking into account all measures of dependence at the 6th weighing method (results differed from each other at the 4th and 5th decimal place). The furthest from the standard (representing the worst results) were obtained for cases where the weighing was based on the median, for the Spearman coefficients and taking into account only significant dependencies (only one relationship), or the Gamma coefficient (all dependencies, weighing based on the median).

In the presented study, the best results were obtained for weighing methods based on a coefficient of variation (formulae 9–10), and then for balances constructed with an arithmetic mean (formula 4 and 7). The worst results concerned the 3rd and 4th weighing method (formula 6 and 8), which were associated with the use of the median.

6. Discussion

The use of automated algorithms in the process of property valuation still raises controversy in Poland. According to the Polish law, only real estate appraisers are authorised to issue opinions on property values. However, there are attempts to apply advanced computational procedures in the valuation of real estate, especially in the context of the held discussions on changing the basis for calculating property tax from surface area (as it currently is) to its value. If such a change were to occur, proper tools (computational algorithms) would be required to allow for the mass valuation of various types of properties. The Szczecin algorithm for mass valuation is such a tool.

The SAMWN has already been successfully applied in business practice. During the process of valuating properties based on SAMWN, certain problems were identified which required improvement. One of them was the automation of the process of determining market value coefficients wwr, which was presented in the article. The use of SAMWN in real estate valuation is, of course, limited by certain conditions.

Obtaining accurate valuation results requires a well-prepared database containing information about the valued properties (including their attributes). An important aspect of creating such a database is to ensure the comparability of the levels of attributes describing individual properties so that e.g. an ‘average’ neighbourhood is understood in the same way for each property. Next, the elementary units (SALs) for which market value coefficients wwr are calculated must be determined. The proper selection of representative properties for the training sample is a very important stage where the representative properties must have all attribute states. In addition, the sample should be sufficiently large to ensure the stability of the calculated dependencies. The representatives should be valued individually by authorised specialists, i.e. property appraisers.

The article systematises the coefficients of dependencies used to determine the impact of individual attributes on the value of a real estate. Additionally, the best method for weighing the impact of individual attribute states was proposed. The advantage of the presented method is that the algorithm does not require additional assumptions regarding the attributes, such as an appropriate type of distribution.

Additionally, SAMWN takes into account various attributes of real estate (which can be adjusted depending on the type of the real estate), including the mode. The proposed algorithm is universal and can be used to value different types of real estate.

7. Conclusions

Estimating the value of real estate in mass can be based on calculation algorithms. In addition, statistical methods are used in the mass valuation process. One of the proposals for mass valuation is the Szczecin Mass Valuation Algorithm. One of the aims of the article was to present the possibility of using various dependency coefficients to determine the impact of the attributes on the value of the real estate. The use of dependency coefficients was proposed, dedicated to ordinal variables (real estate attributes were presented on ordinal scales) as was Pearson's correlation coefficient due to its high popularity among property appraisers.

Various methods were used to consider dependency coefficients in the subsequent property valuation procedure: only significant dependencies or their squares were taken into account, as were all dependencies or their squares. The study showed that the consideration of the coefficients or their squares did not change the scale of the valuation errors.

Finally, 54 variants of property valuation using SAMWN were further analysed. Linear ordering methods were used to receive valuation errors in order to select the best option, which was the second aim of the article. As a result of the used ordering procedure, it turned out that the choice of the coefficient was not as significant (although the use of the Person coefficient did not produce the best results, nor did it give the worst). The valuation results were mainly influenced by the formula used to calculate the impact weights of individual attributes on the value of the real estate in SAMWN.

The analysis shows that the best combination is the use of one of the coefficients: Gamma, Kendall, or Spearman. Additionally, all coefficients greater than 0 between the attributes and the value of the property should be taken into account, as well as the use of the formula no. 6 (designated in formula 10) for calculating the weights: the ratio of the variation coefficient of the property value with the best states of the attributes to the variation coefficient of the property value with the weakest attribute states, corrected by the $\ln\left(\frac{w_{\max}}{w_{\min}}\right)$ coefficient. The weakest results occurred when the median was considered in the method of calculating the weights.

References

Abidoye, R. B., & Chan, A. P. C. (2018). Improving property valuation accuracy: A comparison of hedonic pricing model and artificial neural network. *Pacific Rim Property Research Journal*, *24*(1), 71–83. https://doi.org/10.1080/14445921.2018.1436306.

Ahn, J. J., Byun, H. W., Oh, K. J., & Kim, T. Y. (2012). Using ridge regression with genetic algorithm to enhance real estate appraisal forecasting. *Expert Systems with Applications*, *39*(9), 8369–8379. https://doi.org/10.1016/J.ESWA.2012.01.183.

Alcantud, J. C. R., Rambaud, S. C., & Torrecillas, M. J. M. (2017). Valuation Fuzzy Soft Sets: A Flexible Fuzzy Soft Set Based Decision Making Procedure for the Valuation of Assets. *Symmetry*, *9*(11), 1–19. https://doi.org/10.3390/SYM9110253.

Amidu, A.-R., & Boyd, D. (2018). Expert problem solving practice in commercial property valuation: an exploratory study. *Journal of Property Investment & Finance*, *36*(4), 366–382. https://doi.org/10.1108/JPIF-05-2017-0037.

Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, *27*(3), 247–267. https://doi.org/10.1111/J.1574-0862.2002.TB00120.X.

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. https://doi.org/10.1016/J.ESWA.2011.08.077.

Bellotti, A. (2017). Reliable region predictions for automated valuation models. *Annals of Mathematics and Artificial Intelligence*, *81*(1–2), 71–84. https://doi.org/10.1007/S10472-016-9534-6.

Calka, B. (2019). Estimating Residential Property Values on the Basis of Clustering and Geostatistics. *Geosciences*, *9*(3), 1–14. https://doi.org/10.3390/GEOSCIENCES9030143.

Calka, B., & Bielecka, E. (2016). The Application of Geoinformation Theory in Housing Mass Appraisal. In *2016 Baltic Geodetic Congress (BGC Geomatics)* (pp. 239–243). IEEE. https://doi.org/10.1109/BGC.GEOMATICS.2016.50.

Cervelló-Royo, R., Guijarro, F., Pfahler, T., & Preuss, M. (2016). An Analytic Hierarchy Process (AHP) framework for property valuation to identify the ideal 2050 portfolio mixes in EU-27 countries with shrinking populations. *Quality & Quantity*, *50*(5), 2313–2329. https://doi.org/10.1007/s11135-015-0264-3.

Chen, J.-H., Ong, C. F., Zheng, L., & Hsu, S.-C. (2017). Forecasting spatial dynamics of the housing market using Support Vector Machine. *International Journal of Strategic Property Management*, *21*(3), 273–283. https://doi.org/10.3846/1648715X.2016.1259190.

Chen, Z., Hu, Y., Zhang, C. J., & Liu, Y. (2017). An Optimal Rubrics-Based Approach to Real Estate Appraisal. *Sustainability*, *9*(6), 1–19. https://doi.org/10.3390/SU9060909.

Dimopoulos, T., & Moulas, A. (2016). A Proposal of a Mass Appraisal System in Greece with CAMA System: Evaluating GWR and MRA techniques in Thessaloniki Municipality. *Open Geosciences*, *8*(1), 675–693. https://doi.org/10.1515/GEO-2016-0064.

Dmytrów, K., Gdakowicz, A., & Putek-Szeląg, E. (2020). Methods of analyzing qualitative variable correlation on the real estate market. *Real Estate Management and Valuation*, *28*(1), 80–90. https://doi.org/10.2478/REMAV-2020-0007.

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

Gabrielli, L., Giuffrida, S., & Trovato, M. R. (2017). Gaps and overlaps of urban housing sub-market: Hard clustering and fuzzy clustering approaches. In S. Stanghellini, P. Morano, M. Bottero & A. Oppio (Eds.), *Appraisal: From Theory to Practice* (pp. 203–219). Springer. https://doi.org/10.1007/978-3-319-49676-4_15.

Gatnar, E., & Walesiak, M. (Eds.). (2011). *Analiza danych jakościowych i symbolicznych z wykorzystaniem programu R*. Wydawnictwo C.H. Beck.

Gdakowicz, A., & Putek-Szelag, E. (2020a). Is It Possible to Overfit the Algorithm? Case Study of Mass Valuation of Land Properties in Szczecin. *European Research Studies Journal*, *23*(SI2), 110–122. https://doi.org/10.35808/ERSJ/1812.

Gdakowicz, A., & Putek-Szeląg, E. (2020b). The Use of Statistical Methods for Determining Attribute Weights and the Influence of Attributes on Property Value. *Real Estate Management and Valuation*, *28*(4), 33–47. https://doi.org/10.1515/REMAV-2020-0030.

Glennon, D., Kiefer, H., & Mayock, T. (2018). Measurement error in residential property valuation: An application of forecast combination. *Journal of Housing Economics*, *41*, 1–29. https://doi.org/10.1016/J.JHE.2018.02.002.

Guan, J., Shi, D., Zurada, J. M., & Levitan, A. S. (2014). Analyzing Massive Data Sets: An Adaptive Fuzzy Neural Approach for Prediction, with a Real Estate Illustration. *Journal of Organizational Computing and Electronic Commerce*, *24*(1), 94–112. https://doi.org/10.1080/10919392.2014.866505.

Han, L., & Zhu, J. (2008). Using matrix of thresholding partial correlation coefficients to infer regulatory network. *Biosystems*, *91*(1), 158–165. https://doi.org/10.1016/J.BIOSYSTEMS.2007.08.008.

Hellwig, Z. (1968). Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr. *Przegląd Statystyczny*, *15*(4), 307–327.

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. https://doi.org/10.2478/REMAV-2019-0024.

Hozer, J., Kokot, S., & Kuźmiński, W. (2002). *Metody analizy statystycznej rynku w wycenie nieruchomości.* Polska Federacja Stowarzyszeń Rzeczoznawców Majątkowych.

Hui, S. K., Cheung, A., & Pang, J. (2010). A Hierarchical Bayesian Approach for Residential Property Valuation: Application to Hong Kong Housing Market. *International Real Estate Review*, *13*(1), 1–29. https://ideas.repec.org/a/ire/issued/v13n012010p1-29.html.

International Association of Assessing Officers. (2019). *Standard on Mass Appraisal of Real Property. A criterion for measuring fairness, quality, equity and accuracy*. https://www.iaao.org/media/standards/StandardOnMassAppraisal.pdf.

Jahanshiri, E., Buyong, T., & Shariff, A. R. M. (2011). A Review of Property Mass Valuation Models. *Pertanika Journal of Science & Technology*, *19*(*special issue*), 23–30. https://core.ac.uk/download/pdf/153832642.pdf#page=34.

Kendall, M. G. (1948). *Rank correlation methods* (4th edition). Charles Griffin & Company Limited.

Kilpatrick, J. (2011). Expert systems and mass appraisal. *Journal of Property Investment and Finance*, *29*(4/5), 529–550. https://doi.org/10.1108/14635781111150385.

Kuryj, J. (2007). Metodyka wyceny masowej nieruchomości na bazie aktualnych przepisów prawnych. *Wycena. Wartość, Obrót, Zarządzanie Nieruchomościami*, (4), 50–58.

Lam, K. C., Yu, C. Y., & Lam, C. K. (2009). Support vector machine and entropy based decision support system for property valuation. *Journal of Property Research*, *26*(3), 213–233. https://doi.org/10.1080/09599911003669674.

Lasota, T., Telec, Z., Trawiński, B., & Trawiński, K. (2011). Investigation of the eTS Evolving Fuzzy Systems Applied to Real Estate Appraisal. *Journal of Multiple-Valued Logic and Soft Computing*, *17*(2), 229–253.

Lin, C. C., & Mohan, S. B. (2011). Effectiveness comparison of the residential property mass appraisal methodologies in the USA. *International Journal of Housing Markets and Analysis*, *4*(3), 224–243. https://doi.org/10.1108/17538271111153013.

Lockwood, T., & Rossini, P. (2011). Efficacy in modelling location within the mass appraisal process. *Pacific Rim Property Research Journal*, *17*(3), 418–442. https://doi.org/10.1080/14445921.2011.11104335.

Ma, X., Zeng, D., Wang, R., Gao, J., & Qin, B. (2018). An intelligent price-appraisal algorithm based on grey correlation and fuzzy mathematics. *Journal of Intelligent and Fuzzy Systems*, *35*(3), 2943–2950. https://doi.org/10.3233/JIFS-169650.

McCluskey, W. J., & Borst, R. A. (2011). Detecting and validating residential housing submarkets: A geostatistical approach for use in mass appraisal. *International Journal of Housing Markets and Analysis*, *4*(3), 290–318. https://doi.org/10.1108/17538271111153040.

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. https://doi.org/10.1108/JFMPC-06-2013-0022.

Morano, P., Tajani, F., & Locurcio, M. (2018). Multicriteria analysis and genetic algorithms for mass appraisals in the Italian property market. *International Journal of Housing Markets and Analysis*, *11*(2), 229–262. https://doi.org/10.1108/IJHMA-04-2017-0034.

Napoli, G., Giuffrida, S., & Valenti, A. (2017). Forms and Functions of the Real Estate Market of Palermo (Italy). Science and Knowledge in the Cluster Analysis Approach. In S. Stanghellini, P. Morano, M. Bottero & A. Oppio (Eds.), *Appraisal: From Theory to Practice* (pp. 191–202). Springer. https://doi.org/10.1007/978-3-319-49676-4_14.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining Nonoverlap and Trend for Single-Case Research: Tau-U. *Behavior Therapy*, *42*(2), 284–299. https://doi.org/10.1016/J.BETH.2010.08.006.

Pawlukowicz, R. (2010). Wykorzystanie metodyki porządkowania liniowego do określania wartości rynkowej nieruchomości. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, *17*(107), 377–385.

Polska Federacja Stowarzyszeń Rzeczoznawców Majątkowych. (2008). *Powszechne Krajowe Zasady Wyceny (PKZW) – nota interpretacyjna. Zastosowanie podejścia porównawczego w wycenie nieruchomości*. http://wycena.net.pl/standardy/NI-Zastosowanie.podejscia.porownawczego.pdf.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd edition). McGraw-Hill. http://www.sciepub.com/reference/20400.

Strahl, D. (1978). Propozycja konstrukcji miary syntetycznej. *Przegląd Statystyczny*, *25*(2), 205–215.

Telega, T., Bojar, Z., & Adamczewski, Z. (2002). Wytyczne przeprowadzania powszechnej taksacji nieruchomości (projekt). *Przegląd Geodezyjny*, *74*(6), 6–11.

Walesiak, M. (2016). *Uogólniona miara odległości GDM w statystycznej analizie wielowymiarowej z wykorzystaniem programu R* (2nd edition). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

Walesiak, M., & Dudek, A. (2020). The Choice of Variable Normalization Method in Cluster Analysis. In K. S. Soliman (Ed.), *Education Excellence and Innovation Management: A 2025 Vision to Sustain Economic Development During Global Challenges* (pp. 325–340). International Business Information Management Association.

Wang, D., & Li, V. J. (2019). Mass Appraisal Models of Real Estate in the 21st Century: A Systematic Literature Review. *Sustainability*, *11*(24), 1–14. https://doi.org/10.3390/SU11247006.

Wu, C., Ye, X., Ren, F., & Du, Q. (2018). Modified Data-Driven Framework for Housing Market Segmentation. *Journal of Urban Planning and Development*, *144*(4), 1–15. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000473.

Yacim, J. A., & Boshoff, D. G. B. (2018). Combining BP with PSO algorithms in weights optimisation and ANNs training for mass appraisal of properties. *International Journal of Housing Markets and Analysis*, *11*(2), 290–314. https://doi.org/10.1108/IJHMA-02-2017-0021.

Zeliaś, A. (2002). Some Notes on the Selection of Normalization of Diagnostic Variables. *Statistics in Transition*, *5*(5), 787–802.

Zhang, R., Du, Q., Geng, J., Liu, B., & Huang, Y. (2015). An improved spatial error model for the mass appraisal of commercial real estate based on spatial analysis: Shenzhen as a case study. *Habitat International*, *46*, 196–205. https://doi.org/10.1016/J.HABITATINT.2014.12.001.

Zhou, G., Ji, Y., Chen, X., & Zhang, F. (2018). Artificial Neural Networks and the Mass Appraisal of Real Estate. *International Journal of Online and Biomedical Engineering*, *14*(3), 180–187. https://doi.org/10.3991/IJOE.V14I03.8420.

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. https://doi.org/10.1080/10835547.2011.12091311.