

Izabela Basista*

Verification and Efficiency Analysis of an Algorithm for Similar Properties**

1. Introduction

In the process of real estate valuation, the next important step after a real estate market analysis is to select such properties from the database of a given local real estate market that will be similar to the valued property. As far as domestic legislation is concerned, the notion of “similar properties” appears in the *Land Administration Act* and in the *Ordinance of the Council of Ministers from 21 September 2004 on real estate valuation and preparation of appraisal reports*. The act stipulates that “similar property shall be construed as a property comparable with the valued property in terms of location, legal status, purpose, usage and other features that influence its value.”

But, the existing provisions do not define an obligatory manner of selecting similar properties, so real estate appraisers have certain freedom with regard to selecting similar properties in the valuation process. Properties are usually deemed similar when most of the attributes which describe them are the same [1, 2, 5, 7].

In practice it is very difficult to gather the database with several properties that have identical attributes to the valued property. In the valuation process, a real estate appraiser is often forced to select properties with values of attributes different from the values of attributes of the valued property. The choice is not simple, because the differences should be as little as possible. It should also be remembered that every attribute may have a different influence on changes of transaction prices of a property. What is more, usually numerous properties need to be analysed from this angle, and they additionally have several more price-determining attributes [6, 8, 12, 14].

* AGH University of Science and Technology, Faculty of Mining Surveying and Environmental Engineering, Department of Geomatics, Krakow, Poland

** This work was supported by Dean Grant no. 15.11.150.249/13

2. Algorithm for Selection of Similar Properties

Presented algorithm has been developed in view of its implementation in Geographic Information System software (GIS) [4, 13]. Thanks to this combination the user will receive a tool for the selection of similar properties and in addition will be able to quickly verify the location of the property on the map.

2.1. Price-Determining Indexes and Attributes for Real Estate

Market features may be divided into two arbitrary groups: descriptive attributes and price-determining attributes. Descriptive attributes serve for the identification of the property in documents and the space and for description of the property in words. Price-determining attributes should have a definite numeric value and serve for the appraisal of the property value.

The algorithm, to work properly, requires the scales of all the price-determining attributes to correspond to one another in the interpretation of their values.

The following general scale of attributes has been adopted:

- 1 – unfavourable,
- 2 – satisfactory,
- 3 – average,
- 4 – favourable,
- 5 – very favourable.

Thus, the specific value of the price-determining attribute describes the same positive or negative bias of each of the features, as it has been assumed that the quality spectrum of the property has been divided into five equal intervals. The "0" value is reserved for the lack of information about the attribute, and it will not be taken into account in the calculations.

The value assigned to each of the attributes should be determined on the basis of information provided by the user and five-point scale developed for each attribute. As a result, the database will contain a description of these attributes and their value. A detailed description of the process of entering data into the database has been presented in previous works of the author [3, 4].

After each of the price-determining attributes has been assigned a value on the basis of the five-degree scale, a price-determining index can be calculated for each of the properties (d_i). Every property in the database, apart from attributes (a_{ij}), has also a transaction unit price assigned (c_i). As part of the statistical analysis of the market, the correlation coefficient can be calculated (r_j) for each of the attributes in relation to the adjusted transaction unit prices. Values r_j^2 will constitute the weighting factors of each of the attributes. The proposed price-determining index (d_i) will be calculated with the following equation:

$$d_i = \sqrt{\frac{\sum_{j=1}^k a_{ij}^2 r_j^2}{\sum_{j=1}^k r_j^2}} \quad (1)$$

where:

- j – labels of consecutive attributes,
- k – number of attributes in question,
- i – labels (indexes) of consecutive properties in the database.

The standard deviation of the index d_i can be calculated with the equation:

$$\sigma(d_i) = \sqrt{\sum_{j=1}^k \left(\frac{a_{ij} r_j^2}{d_i \sum_{j=1}^k r_j^2} \right)^2 \sigma^2(a_j)} \quad (2)$$

whereas the adopted standard deviation for attributes ($\sigma(a_j)$) is on the level of 0.2, which follows from the assumption of error tolerance, that while determining the value of each of the attributes, the value different by 1 degree may be given, which constitutes 20% of a five-degree scale.

The value of the price-determining index for a given property will reflect its attractiveness on the market. Properties with an index around 5 will have high values of price-determining attributes, so these properties will be very attractive in a given market. Properties with the index of around 1 are not very attractive.

Calculation of this index for each of the properties in the database as well as for the valued (model) property is the essential part of the proposed algorithm [4].

2.2. Algorithm for Similar Properties Selection – How It Works

The algorithm serves for the selection of a set of properties as similar as possible to the valued (model) one. The algorithm works on the basis of price-determining indexes calculated for each of the properties in the database and for the valued property in question [4].

The algorithm for selecting set of properties similar to the valued one may be divided into the following stages:

1. Determining the value of attributes of the valued property on the basis of the input information and attributes scales.
2. Calculating price-determining index d_w and its standard deviation $\sigma(d_w)$ for the valued property.
3. Defining intervals for price-determining indexes d_x for properties similar to the valued property.
4. Selection of properties assigned to a given interval.

Interval limits are determined depending on the level of confidence ($p = 1 - \alpha$) for the selection of similar properties. If the index d_w for the valued property and its standard deviation has been determined $\sigma(d_w)$, then the width of the confidence interval containing the indexes d_x of similar properties should be defined with the following inequality (for quantiles of normal distribution):

$$p = 1 - \alpha \quad d_w - z(1 - \alpha/2) \cdot \sigma(d_w) < d_x < d_w + z(1 - \alpha/2) \cdot \sigma(d_w) \quad (3)$$

3. Verification of the Algorithm on an Exemplary Database

Below, an example with numbers has been presented to verify the algorithm for selecting a set of properties similar to the valued property. Table 1 contains information about undeveloped land properties for residential development in the area of the Skawina commune, with transaction unit prices adjusted as of March 2009. The values of attributes have been determined on the basis of five-degree scales of price-determining attributes. Choice of price-determining attributes has been made based on the literature and questionnaires on the real estate market [9–11].

Table 1. Database of undeveloped land properties together with attributes, prices and price-determining indexes

Property	Part of the city	Access to transportation	Standard of the access road	Shape	Utility infrastructure	Additional information	Area [m ²]	Unit price - adjusted [PLN/m ²]	Price-determining index d_x
1	4	3	3	4	4	3	1583	90	3.44
2	4	3	3	4	4	3	1579	90	3.44
3	3	1	1	3	2	3	973	67	2.29
4	2	2	2	2	1	2	2169	17	1.85
5	3	3	1	3	3	1	979	50	2.35
6	3	3	1	3	3	1	809	50	2.35
7	3	3	1	3	3	1	790	50	2.35
8	3	1	3	3	3	3	1000	55	2.78
9	3	3	3	1	4	2	900	69	2.89
10	4	3	2	4	2	4	3014	100	3.21
11	3	1	2	3	2	2	1255	41	2.15
12	3	2	2	4	2	4	690	81	2.94
13	5	4	5	4	5	4	1395	141	4.53
14	2	3	1	3	2	2	2186	41	2.17
15	3	2	3	3	2	2	2200	54	2.46
16	4	3	3	3	3	2	749	82	2.94

Table 1 cont.

17	4	3	3	3	2	3	645	82	2.99
18	3	3	1	3	2	2	5500	33	2.31
19	3	2	2	4	3	4	1152	87	3.10
20	3	2	4	4	4	4	2000	122	3.64
21	4	3	2	3	1	1	1600	59	2.34
22	4	2	3	3	2	4	851	78	3.13
23	3	2	3	2	4	2	926	79	2.80
24	3	1	1	3	2	1	1900	29	1.82
25	3	4	2	4	3	4	1000	103	3.39
26	3	4	2	2	3	4	1300	93	3.21
27	3	3	2	3	3	4	1200	98	3.11
28	3	1	2	3	2	3	700	56	2.41
29	2	3	3	2	3	1	900	67	2.45
30	4	3	2	3	2	2	1484	84	2.60
31	3	1	1	3	2	2	2000	36	2.01
32	3	2	2	3	3	1	2003	61	2.31
33	3	3	2	2	3	2	1361	74	2.52
34	4	1	3	3	1	3	777	56	2.66
35	4	4	4	5	4	5	1501	115	4.36
36	2	1	2	3	2	1	2665	61	1.81
37	1	2	3	3	3	3	3200	52	2.69
38	4	1	2	3	1	3	925	54	2.47
39	4	4	3	3	3	2	2234	90	3.13
40	3	1	1	3	2	2	5000	31	2.01
41	3	3	3	3	4	4	885	101	3.46
42	1	2	3	3	3	2	800	56	2.45
43	4	5	5	4	3	4	96	110	4.23
44	3	3	1	3	1	3	2000	66	2.44
45	3	2	4	2	3	3	612	73	3.01
46	2	1	2	3	2	2	126	32	2.00
47	3	2	2	3	4	4	814	98	3.20
48	3	2	4	3	3	4	1263	104	3.35
49	2	1	3	2	3	2	3500	54	2.33
average	3.10	2.39	2.41	3.04	2.67	2.65	-	-	-
standard deviation	0.2	0.2	0.2	0.2	0.2	0.2	-	-	-
correlation r	0.51	0.58	0.64	0.46	0.63	0.71	-	-	-
correlation squared r^2	0.263	0.341	0.414	0.208	0.399	0.508	-	-	-
							3500	-	2.7787
								$\sigma(d) = 0.085$	

From the database, three groups of properties have been selected on three levels of confidence:

1. Price-determining indexes d_x of properties similar to the valued property, on the confidence level $p = 1 - \alpha \approx 0.68$, should fall within interval $d - \sigma(d) < d_x < d + \sigma(d)$, thus $2.69 < d_x < 2.86$. For the same confidence level two properties have been selected (Tab. 2).

Table 2. Set of properties similar to the one in question on the confidence level $1 - \alpha \approx 0.68$

Property	Part of the city	Access to transportation	Standard of the access road	Shape	Utility infrastructure	Additional information	Area [m ²]	Unit price – adjusted [PLN/m ²]	Price-determining index d_x
8	3	1	3	3	3	3	1000	55	2.78
23	3	2	3	2	4	2	926	79	2.80

2. Price-determining indexes d_x of properties similar to the valued property, on the confidence level $p = 1 - \alpha \approx 0.95$, should fall within interval $d - 2\sigma(d) < d_x < d + 2\sigma(d)$, thus $2.61 < d_x < 2.95$. For the same confidence level seven properties have been selected (Tab. 3).

Table 3. Set of properties similar to the one in question on the confidence level $1 - \alpha \approx 0.95$

Property	Part of the city	Access to transportation	Standard of the access road	Shape	Utility infrastructure	Additional information	Area [m ²]	Unit price – adjusted [PLN/m ²]	Price-determining index d_x
8	3	1	3	3	3	3	1000	55	2.78
9	3	3	3	1	4	2	900	69	2.89
12	3	2	2	4	2	4	690	81	2.94
16	4	3	3	3	3	2	749	82	2.94
23	3	2	3	2	4	2	926	79	2.80
34	4	1	3	3	1	3	777	56	2.66
37	1	2	3	3	3	3	3200	52	2.69

3. Price-determining indexes d_x of properties similar to the valued property, on the confidence level $p = 1 - \alpha \approx 0.999$, should fall within interval $d - 3\sigma(d) < d_x < d + 3\sigma(d)$, thus $2.52 < d_x < 3.04$. For the same confidence level ten properties have been selected (Tab. 4).

Table 4. Set of properties similar to the one in question on the confidence level $1 - \alpha \approx 0.999$

Property	Part of the city	Access to transportation	Standard of the access road	Shape	Utility infrastructure	Additional information	Area [m ²]	Unit price – adjusted [PLN/m ²]	Price-determining index d_x
8	3	1	3	3	3	3	1000	55	2.78
9	3	3	3	1	4	2	900	69	2.89
12	3	2	2	4	2	4	690	81	2.94
16	4	3	3	3	3	2	749	82	2.94
17	4	3	3	3	2	3	645	82	2.99
23	3	2	3	2	4	2	926	79	2.80
30	4	3	2	3	2	2	1484	84	2.60
34	4	1	3	3	1	3	777	56	2.66
37	1	2	3	3	3	3	3200	52	2.69
45	3	2	4	2	3	3	612	73	3.01

4. Algorithm Effectiveness

Effectiveness of the proposed algorithm for similar properties selection has been determined with the coefficient defined by Aranowski [1]. This factor defines the level of similarity of selected properties to the valued property. Its value falls within the interval $<0,1>$. The higher the value, the greater the similarity of the selected properties to the valued property. The similarity level of the property is calculated with the following equation:

$$V_w = 1 - \frac{1}{m \cdot n} \sum_{j=1}^m w_j \sum_{i=1}^n \frac{|a_{w,j} - a_{i,j}|}{a_{\max,j} - 1} \quad (4)$$

where:

- V_w – coefficient defining the level of similarity of selected properties similar to the valued one,
- m – number of price-determining attributes,
- n – number of selected similar properties,
- $a_{w,j}$ – value of j -price-determining attribute of the valued property,
- $a_{i,j}$ – value of j -price-determining attribute of i -similar property,
- $a_{\max,j}$ – maximum value of j -price-determining attribute,
- w_j – weight of j -price-determining attribute.

In order to define the level of the algorithm effectiveness, for each of the properties from the above exemplary database, groups of properties similar to them have been selected on three confidence levels. Next, the coefficient value has been calculated for each group, as a result of which 49 coefficient values have been found on each of the three confidence levels.

Below average values of that coefficient have been presented:

- at the confidence level $p \approx 0.68$, the average value of the coefficient is $V_w = 0.9833$;
- at the confidence level $p \approx 0.95$, the average value of the coefficient is $V_w = 0.9804$;
- at the confidence level $p \approx 0.999$, the average value of the coefficient is $V_w = 0.9789$.

The results allow us to state that the properties selected with the proposed algorithm show great similarity to the valued property, so this algorithm works effectively.

5. Conclusions

Real estate market analysis is an important process, which is very difficult at the same time and requires a real estate appraiser to have experience and great knowledge about the given market. Its results determine which properties will be selected for evaluation and thus the appraised value of property. A real estate appraiser must choose from among properties with attributes' values different from attributes of the valued property. Those differences should be as little as possible, and additionally the appraiser must take into consideration the fact that every attribute may have a different influence on changes in the transaction prices of properties.

The presented algorithm serves to facilitate such decisions of appraisers. The algorithm allows to select from the properties in the market defined by the appraiser – the properties similar to the property with the given price-determining attributes. The resulting group is certainly no final solution. The real estate appraiser should verify that choice. Nevertheless he has a group of properties smaller than the initial one, selected taking into account the influence of attributes on price changes in a given real property market. Moreover, a real estate appraiser may determine the confidence level, and thus the width of the interval to include similar properties.

The proposed algorithm simplifies and partly automates the process of market analysis, and that is why it can be used to build systems of real estate valuation.

References

- [1] Aranowski A.: *Analiza efektywności metod statystycznych w badaniu podobieństwa obiektów na rynku nieruchomości*. Szczecin 2010 [Ph.D. thesis, unpublished].
- [2] Barańska A.: *Metody jakościowe i ilościowe na usługach wyceny nieruchomości*. Acta Scientiarum Polonorum – Geodesia et Descriptio Terrarum, vol. 7(2), 2008, pp. 3–13.

-
- [3] Basista I.: *Wstępny projekt podsystemu geoinformatycznego do analizy rynku nieruchomości*. Studia i Materiały Towarzystwa Naukowego Nieruchomości, vol. 18, no. 4, 2010, pp. 45–56.
- [4] Basista I.: *Podsystem geoinformatyczny do analizy rynku nieruchomości*. Kraków, 2012 [Ph.D. thesis, unpublished].
- [5] Czaja J., Parzych P.: *Szacowanie rynkowej wartości nieruchomości w aspekcie Międzynarodowych Standardów Wyceny*. Wydawnictwo Stowarzyszenia Naukowego im. St. Staszica, Kraków 2007.
- [6] Dąbrowski J.: *Real Estate Market Analysis for Market Value Assessment*, Geomatics and Environmental Engineering [AGH UST quarterly], vol. 2, no. 4, 2008, pp. 35–46.
- [7] Drelich P.: *Podobieństwo a wycena nieruchomości*. Polskie Towarzystwo Rzeczoznawców Majątkowych, 2011, [on-line:] <http://www.ptm.pl> [access: December 2013].
- [8] Foryś I.: *Wykorzystanie metod taksonomicznych do wyboru obiektów podobnych w procesie wyceny lokali mieszkalnych*. Studia i Materiały Towarzystwa Naukowego Nieruchomości, vol. 18, no. 1, 2010, pp. 95–106.
- [9] Foryś I., Kokot S.: *Preferencje potencjalnych nabywców w określaniu wag cech rynkowych nieruchomości*. Studia i Materiały Towarzystwa Naukowego Nieruchomości, vol. 16, no. 2, 2008, pp. 45–62.
- [10] Jurczak F.: *Preferencje klientów na rynku nieruchomości w Krakowie*, Kraków 2009, [on-line:] <http://artykuly.krn.pl> [access: December 2014].
- [11] Krajewska M., Gaca R.: *Preferencje nabywców nieruchomości na wybranych rynkach rodzajowych*. Studia i Materiały Towarzystwa Naukowego Nieruchomości, vol. 15, no. 3–4, 2007, pp. 35–42.
- [12] Krajewska M.: *The problem of real estate research area selection for the purpose of appraisal process*. Real Estate Management and Valuation 2013, vol. 21, no. 3, pp. 25–34.
- [13] Rodriguez M., Sirmans C.F., Marks A.P.: *Using Geographic Information Systems to Improve Real Estate Analysis*. Journal of Real Estate Research, vol. 10, no. 2, 1995, pp. 163–174.
- [14] Zyga J.: *Podobieństwo w wycenie nieruchomości*. Budownictwo i Architektura, vol. 5(2), 2009, pp. 61–76.