CORRELATION ANALYSIS IN THE PROCESS OF WEIGHTING REAL PROPERTY ATTRIBUTES[1]

Anna Marta Barańska
Department of Geomatics
AGH University of Science and Technology
e-mail: abaran@agh.edu.pl

Abstract

The key goal of the market analysis carried out for property valuation purposes is to select pricing attributes of real properties and to assign so-called weights to them, which would illustrate their influence on the prices in a given market (Regulation 2004). Correlation analysis is a very useful tool in this respect. However, the fact that it is limited to the use of classical Pearson's linear correlation coefficients is too much of a simplification due to the frequently occurring heterogeneity of the real estate market.

The research paper proposes to use the possibilities offered by the broadly understood correlation analysis which, among linear correlations, takes into account not only Pearson's correlation but also rank correlations. At the same time, analyses of various non-linear correlations are being carried out where linear relationships are not reliable enough. The aim of this research is the simultaneous verification of the existence of various types of correlative relationships, taking into account the nature of random variables that the analyses relate to, and the nature of relationships between them. This approach makes it possible to adjust the weighting of market attributes, the values of which are improved in a few subsequent steps, in order to eventually approach the optimal result as close as possible.

Keywords: *real estate market analysis, property attribute weights, linear and non-linear correlation.*

JEL Classification: *C1, C3, C4, C5, C8, R3.*

Citation Barańska A.M., 2019, *Correlation Analysis in the Process Of Weighting Real Property Attributes*, Real Estate Management and Valuation, vol. 27, no. 4, pp. 74-84.

DOI: 10.2478/remav-2019-0037

1. Introduction

Numerous authors have already tackled the issue of the use of correlation analysis as a possible tool for determining market weights of real estate attributes in explaining their prices (Doszyń 2017; Gaca 2018, and others). However, in a certain generalization, these studies are limited to the use of linear relationships. This issue, being a key element of a thoroughly conducted analysis of the real estate market, is still worth researching and refining the applied methodology, not only due to the market analysis carried out for real estate valuation purposes, but also for the analysis itself, contained in various studies, reports, etc. The professional practice of real estate appraisers in this respect is usually limited to the basic methods of determining weights, such as *ceteris paribus* or a survey of preferences of property buyers. The complexity of the surrounding reality forces us to look for more sophisticated tools that describe market relations better, in order to maximize the accuracy of estimating the weights of market attributes.

The use of correlation analysis for this purpose is a justified alternative in this case, as suggested by

[1] The task was carried out within the scope of research in the Faculty of Mining Surveying and Environmental Engineering AGH UST

various authors (BARAŃSKA 2018). The rank correlation analysis has recently become quite popular (DOSZYŃ 2017; GACA, SAWIŁOW 2014). This method, however, has its limitations despite the seemingly greater relevance to market analyses resulting from the mostly qualitative nature of the variables under consideration. The next step is the analysis of curvilinear correlations, which takes into account a specific observed type of non-linear dependence. The key here is the detection of regularity in the tested dependencies of various types. It is therefore not about the "forced" fitting of a complex model function into real data that does not exhibit apparent regular variability. The model cannot be overfitted, which, in this specific case, means that it cannot be too complicated. A consequence of this would be the lack of universality. Another obvious condition is working on sufficiently large data sets so that the modelling is not based on interpolation, but on approximation. With a small amount of data, interpolation leads to the overfitting of the model. Consequently, it will not work on other data derived from the same market, even if we avoid extrapolation risks.

In this study, the author proposes a combination and a parallel analysis of different types of correlation coefficients as the optimal solution to the studied problem using the statistical tool of correlation analysis.

2. Rank correlation

Rank correlations are essentially a tool for ordinal data analysis. They are also useful in assessing correlation relationships between the attributes, one of which is quantitative and the other – ordinal. A circumstance that undoubtedly facilitates the analysis is the fact that, in this case, in contrast to the ordinary correlation analysis, the normality of distribution is not required of quantitative attributes.

Different rank correlation options are an interesting alternative to Pearson's classic correlation. As the name suggests, they are calculated for variables that have already been subject to ranking, that is, the ranks assigned in the form of consecutive natural numbers, in ascending order, to the original values of random variables. The obtained set may contain so-called tied ranks, which is very common for market characteristics and results directly from the same values of a variable for many cases. The rank in this case is the arithmetic mean of the ranks without taking equal values into account (KRAWCZYK, SŁOMKA 1982).

The following types of rank correlations can be distinguished:

1) Spearman's Rho correlation

$$\rho_S = \frac{\frac{n^3-n}{6} - T_x - T_y - \sum_{i=1}^{n}(x_i - y_i)^2}{\sqrt{\left(\frac{n^3-n}{6} - 2T_x\right)\cdot\left(\frac{n^3-n}{6} - 2T_y\right)}} \tag{1}$$

where:

x_i, y_i – the ranks of the analyzed random variables,
n – the size of a random sample,
T_x, T_y – adjustments due to the occurrence of tied ranks, calculated according to Formula:

$$T = \frac{\sum_{j=1}^{k}\left(t_j^3 - t_j\right)}{12} \tag{2}$$

where:

t_j – the number of tied ranks (with the same variable values) in the j-th tie,
k – the number of ties.

If there are no tied ranks in the analyzed data, Formula (1) takes a simplified form (3). In the case of market attributes, this usually happens only for continuous random variables, such as surface area, distance, price.

$$\rho_S = 1 - \frac{6\cdot\sum_{i=1}^{n}(x_i - y_i)^2}{n\cdot(n^2-1)} \tag{3}$$

Spearman's rank correlation coefficient is, in fact, the classic Pearson measure, but it is not calculated for the values of the "raw" variables being compared, but for their rank values. Thus, the same coefficient determined for ranks may demonstrate a significant relationship between variables, despite the statistical irrelevance of Pearson's correlation for primary variables. When ranking, the primary monotone but non-linear dependence is transformed into a linear one. Therefore, it is sometimes believed that Pearson's linear correlation coefficient applied to ranks measures the strength of non-linear dependence. However, this is a debatable opinion, or at least an incomplete one, as on the one hand it is determined that the studied relationship is not linear, but at the same time the detailed form of its non-linear nature is not investigated. Therefore, it is still not known (not stated numerically) what kind of dependence this is, and hence the specific form of this dependence is not taken into account in the calculated weights of market attributes.

2) Kendall's Tau correlation

$$\tau_K = \frac{P-Q}{P+Q+T} = 2 \cdot \frac{P-Q}{n(n-1)} \tag{4}$$

where:

P, Q, T – respectively: the number of compliant pairs (values in the compared pairs change directly proportionally), the number of non-compliant pairs (the values in the compared pairs change inversely), the number of tied pairs (at least one variable has equal values in the compared pair) - after the observations have been arranged in the sample into all possible pairs,

n – random sample size.

Kendall's tau correlation coefficient is the difference between the probability that the compared variables will be set in the same order and the probability that they will be arranged in the opposite order. It takes values from -1 to 1 inclusive, where +1 means that each of the variables increases with the increase of the second one, and -1 means that each of them decreases with the increase of the second one. Thus, τ-Kendall's, as well as ρ-Spearman's, are measures of the monotone dependence of random variables.

3) Gamma correlation

This coefficient requires assumptions similar to τ-Kendall's and ρ-Spearman's coefficients, and it is used in cases where data contains many tied observations. It is more similar to the τ-Kendall's coefficient, because it is calculated as the difference between the probability that the ordering of two variables is consistent and the probability that it is inconsistent, divided by 1 minus the probability of the occurrence of tied observations.

3. Curvilinear correlation

Due to the limitations of rank correlation described above, in the case of the occurrence of non-monotone dependencies between variables, the curvilinear correlation coefficient q is a very useful tool in calculating the weights of market attributes. It is a generalization of Pearson's correlation coefficient r, and is calculated according to the following Formula:

$$q = \sqrt{\frac{\sum_{i=1}^{n}(f(x_i)-\hat{y})^2}{\sum_{i=1}^{n}(y_i-\hat{y})^2}} = \sqrt{1 - \frac{\sum_{i=1}^{n}\left(y_i-f(x_i)\right)^2}{\sum_{i=1}^{n}(y_i-\hat{y})^2}} = \sqrt{\frac{WSK}{CSK}} = \sqrt{1 - \frac{NSK}{CSK}} \tag{5}$$

where:

x_i – empirical value of the independent random variable (property attribute) for the i observation,

y_i – empirical value of the dependent random variable (property price) for the i observation,

$f(x_i)$ – model value of the dependent random variable for the i observation,

$\hat{y}$ – mean value from the empirical values of the dependent random variable,

n – random sample size,

$CSK = \sum_{i=1}^{n}(y_i - \hat{y})^2$ – total sum of squares, i.e. total scatter of a dependent variable in relation to its average value,

$WSK = \sum_{i=1}^{n}(f(x_i) - \hat{y})^2 = \sum_{i=1}^{n}\left((y_i - \hat{y}) - (y_i - f(x_i))\right)^2$ – sum of squares explained by the regression model,

$NSK = \sum_{i=1}^{n}\left(y_i - f(x_i)\right)^2 = \sum_{i=1}^{n}\delta_i^2$ – unexplained sum of squared deviations of the regression model, illustrating the part unexplained by this model.

The value of q is equal to the Pearson correlation coefficient r, when f in Formula (5) is a linear function. Its square q^2 is the coefficient of determination (fitting), corresponding to the estimated model, describing the dependence of the price on the selected property attribute.

4. Statistical significance of correlation coefficients

Each of the calculated correlation coefficients should be subject to significance analysis before it forms the basis for calculating the weights of market attributes. This task comes down to verification using a parametric test of significance of the hypothesis $H_0\colon r = 0$ against the hypothesis $H_1\colon r \neq 0$. Due to the fact that a thoroughly performed market analysis requires a sufficiently large set of data, the verification of the significance of correlation can be performed using test functions with a normal distribution, which, for each type of correlation, take the following form:

– for Pearson's correlation $$T = \frac{\hat{r}_P}{\sqrt{\frac{1-\hat{r}_P^2}{n-2}}} \quad (6)$$

– for ρ-Spearman's correlation $$Z = \frac{\hat{\rho}_S}{\sqrt{\frac{1}{n-1}}} \quad (7)$$

– for τ-Kendall's correlation $$Z = \frac{3 \cdot \hat{\tau}_K \cdot \sqrt{n \cdot (n-1)}}{\sqrt{2 \cdot (2n+5)}} \quad (8)$$

where:

$\hat{r}_P$, $\hat{\rho}_S$, $\hat{\tau}_K$ – estimators, respectively for: Pearson's, Spearman's and Kendall's correlation coefficients.

Statistical inference runs normally with a determined significance level.

5. Research problem

The research material used in this study comprised a collection of data derived from the residential market in the city of Rawa Mazowiecka. The collected database consists of 121 housing units, described using 16 attributes that were identifiable on the studied market. The list of those attributes is contained in Table 1.

Table 1

List of attributes of real properties analyzed on the residential market

No.	Property attribute	No.	Property attribute
1	Transaction date	9	Availability of public facilities
2	Number of chambers	10	Positive elements of the surroundings
3	Usable floor space	11	Accessibility of recreational areas
4	Location	12	Heating and gas
5	Storey	13	Noise and other nuisances
6	Transport accessibility	14	Lift
7	Parking	15	Window exposure
8	Auxiliary rooms	16	Unit price

Source: own study.

Prior to the analyses resulting from the aim of this study, the volatility of prices at the time the transactions were concluded (about 5 years, with an average of 2 transactions per month) was examined. Due to the price stability in the analyzed period, which can be observed in Figure 1, correlation analysis was performed on the transaction prices.

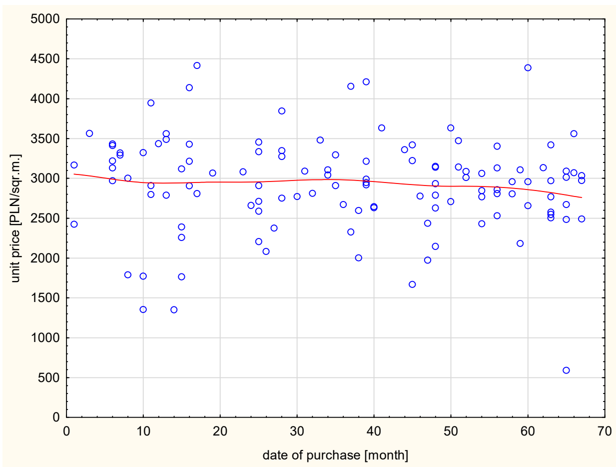
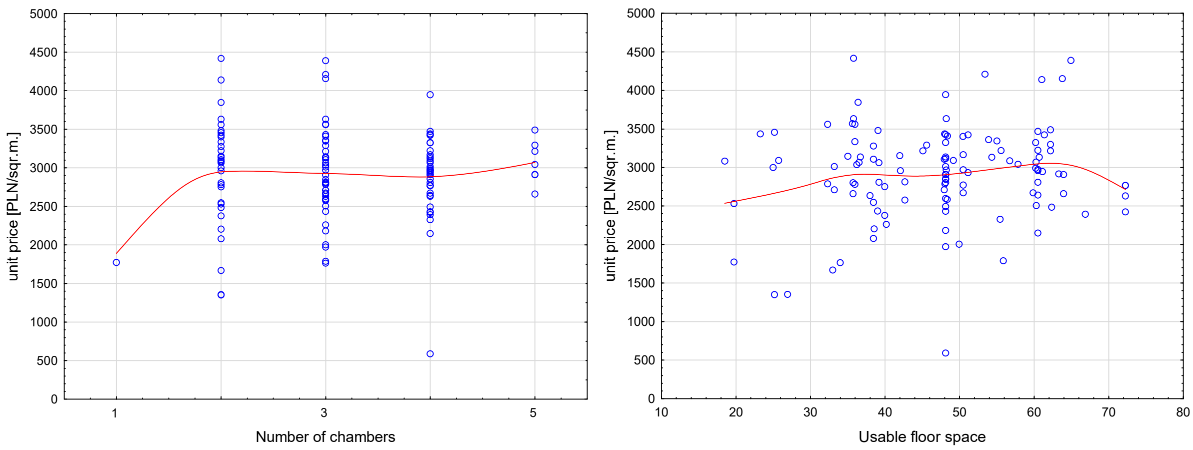


Fig. 1 A scatter plot of the unit price in relation to the transaction date. *Source*: own study.

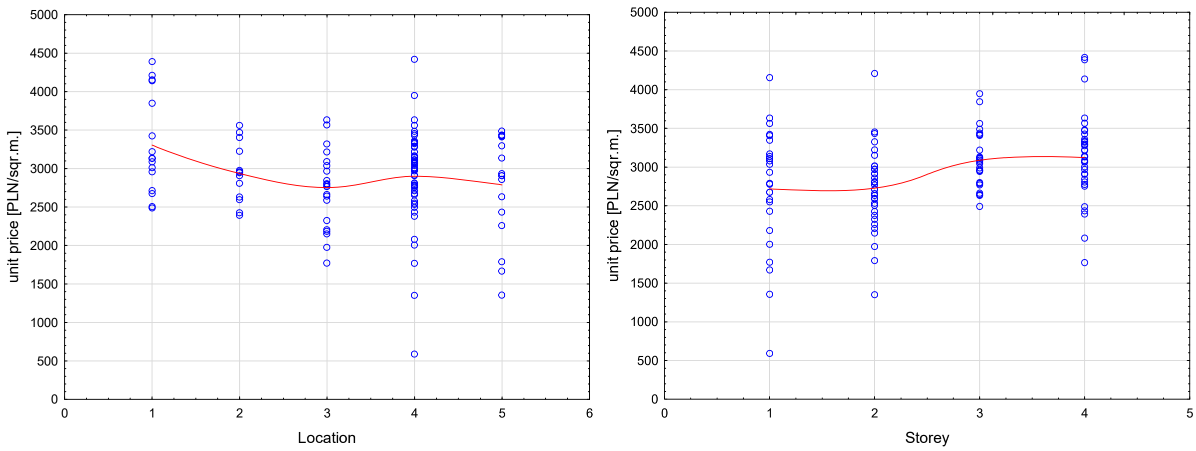
5.1. Graphic representation of the dependence *price*(*attribute* of real estate)

The first stage of the proper analyses was a graphic representation of the dependence of the unit price of the property on each of the attributes in the form of scatter plots with the regression line superimposed on them, estimated by the method of least squares (LSM).

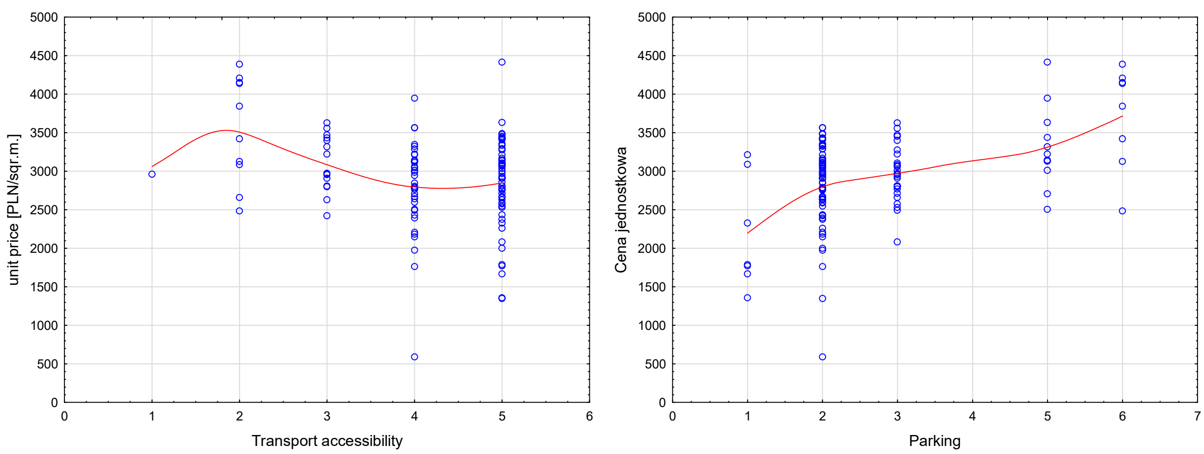
Functions *f* were selected based on the LSM line, from among the elementary functions. On the graphs below, the fitted functions were: a quadratic function (Figures 3, 8 and 12), a third degree polynomial (Figures 4, 6, 9, 10 and 11) and a cyclometric function (Fig. 5).



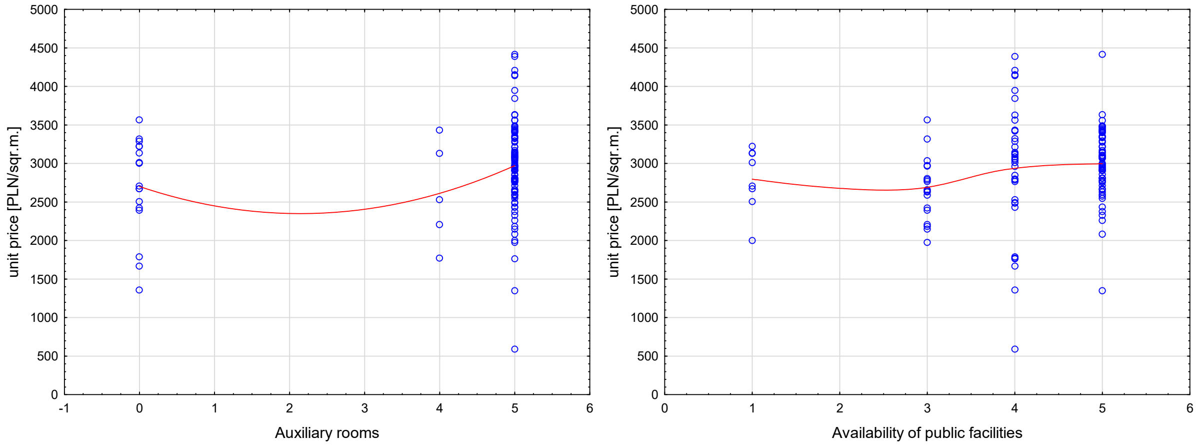
Figs. 2 and 3 Graphs of the dependence of residential unit prices on the number of chambers and their usable floor space. *Source*: own study.



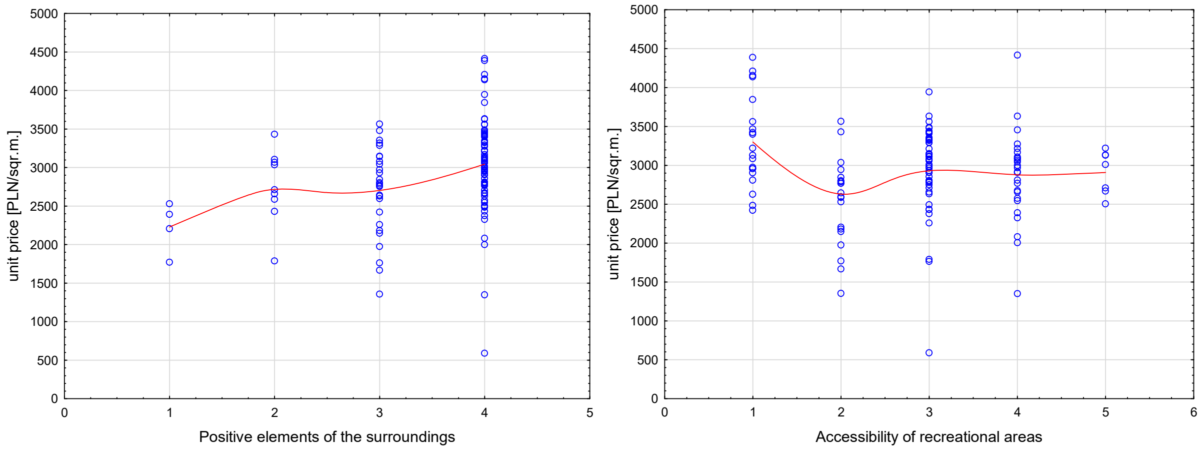
Figs. 4 and 5 Graphs of the dependence of house prices on the location relative to the city center and the storey. *Source*: own study.



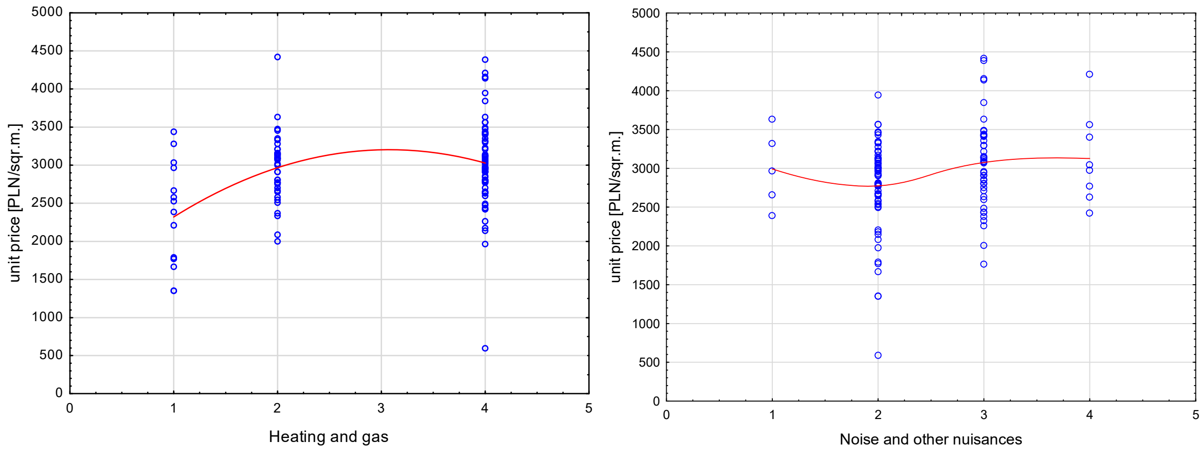
Figs. 6 and 7 Graphs of the dependence of residential unit prices on the accessibility of public transport and availability of parking spaces next to the building. *Source*: own study.



Figs. 8 and 9 Graphs of the dependence of residential unit prices on the surface areas of auxiliary rooms and on the availability of public facilities. *Source*: own study.

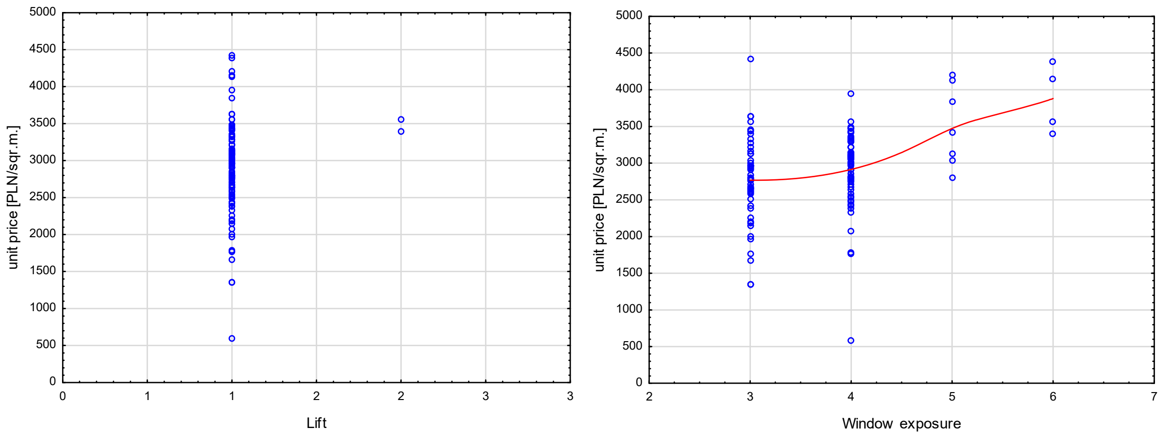


Figs. 10 and 11 Graphs of the dependence of residential unit prices on the occurrence of positive elements of the surroundings and on the accessibility of recreational areas. Source: own study.



Figs. 12 and 13 Graphs of the dependence of residential unit prices on the type of heating and availability of gas network, and on the occurrence of nuisances in the building's surroundings. *Source*: own study.

Some of the graphs indicate that there is no justification for the attempts to fit a non-linear dependence, either because the value of a given attribute is not diversified enough (only two values - Figure 14) or due to the relatively explicit rectilinear character of the regression line determined by the least square method (Figures 7, 13, 15), neglecting one outlier that disrupted the apparent linear tendency (Figure 2). A list of fitted basic non-linear functions that serve as a two-dimensional regression is provided in Table 2.



Figs. 14 and 15 Graphs of the dependence of residential unit prices on the presence of a lift in a building and on window exposure. *Source*: own study.

Table 2

List of fitted elementary non-linear functions

No.	Function name	Function form
1	Quadratic function	$y = a + b \cdot x + c \cdot x^2$
2	Third degree polynomial	$y = a + b \cdot x + c \cdot x^2 + d \cdot x^3$
3	Exponential function	$y = a + b \cdot \exp(x)$
4	Logarithmic function	$y = a + b \cdot \ln(x)$
5	Homographic function	$y = a + \frac{b}{x}$
6	Trigonometric functions	$y = a + b \cdot sin(x)$
7		$y = a + b \cdot cos(x)$
8		$y = a + b \cdot tg(x)$
9		$y = a + b \cdot ctg(x)$
10	Cyclometric function	$y = a + b \cdot arctg(x)$

Source: own study.

5.2. Estimation of two-dimensional non-linear regression parameters

Due to the fact that all tested non-linear elementary functions maintain linearity with respect to their parameters, the estimation of these parameters is carried out according to the well-known algorithm for estimating parameters of multidimensional linear regression described in numerous publications (CZAJA 2001, BARAŃSKA 2010, and others). The modification applies only to the structure of matrix A of coefficients of the created system of linear equations AX=C. This matrix will have the size of $(n \times u)$, where n is the random sample size, and u is the number of the searched parameters of the selected non-linear function. Thus, apart from the column of ones, resulting from including a constant term in each function, the remaining columns will result from the transformations of the independent variable (selected attribute) according to the function f. For example, for the cyclometric function, we will have column $arctg(x)$ following the column of ones.

The estimation process proceeds according to the known Formulas (9)-(11):

$$A \cdot X = C \tag{9}$$

$$\hat{X} = (A^T \cdot A)^{-1} \cdot A^T \cdot C \tag{10}$$

$$\hat{C} = f(x) = A \cdot \hat{X} \tag{11}$$

where:

$A = \begin{bmatrix} 1 & arctg(x_1) \\ 1 & arctg(x_2) \\ \vdots & \vdots \\ 1 & arctg(x_n) \end{bmatrix}$ – sample coefficient matrix with the size of $(n \times u)$ of the system of equations

$A \cdot X = C$, linear due to the parameters (for the function *f*, in the form of a cyclometric function, $u = 2$),

$X = \begin{bmatrix} a \\ b \end{bmatrix}$ – sample matrix of unknown function parameters, here: for $f(x) = a + b \cdot arctg(x)$, with the size of $(u \times 1)$,

$C = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$ – vector of the empirical values of a dependent variable (price), with the size of $(n \times 1)$,

$\hat{X}$ – vector of the estimators of the *f* function parameter, with the size of $(u \times 1)$,

$\hat{C} = f(x)$ – vector of the model values of a dependent random variable with the size of $(n \times 1)$.

A detailed description of the accuracy analysis in the process of model parameter estimation can be found in other publications (CZAJA 2001, BARAŃSKA 2010, and others). These analyses come down to determining the coefficient of determination of the q^2 model from Formula (5).

5.3. Correlation analysis

The analysis of various types of *attribute-price* correlations was the main part of the performed calculations. It had been preceded by a test of feature redundancy, carried out using Pearson's correlation and rank correlation. The Gamma correlation coefficient indicated one redundant attribute – *Lift*, strongly correlated with six others. The *attribute-price* analyses confirmed its irrelevance. This attribute has therefore been deleted in Table 3.

Table 3

Calculation of the weights of residential unit attributes based on various types of correlations

i	Attribute	$\hat{r}_P$	$k_i(\hat{r}_P)$	$\hat{\rho}_S$	$k_i(\hat{\rho}_S)$	$\hat{\tau}_K$	$k_i(\hat{\tau}_K)$	$\hat{r}_G$	$k_i(\hat{r}_G)$	*Price*(*attribute*) function	q_i
1	~~Number of chambers~~	0.02	0	-0.01	0	0.00	0	0.00	0	Linear function	
2	Usable floor space	0.13	2	0.05	0	0.03	0	0.03	0	Quadratic function	0.21
3	**Location**	-0.20	5	-0.07	1	-0.05	1	-0.06	1	Third degree polynominal	0.28
4	Storey	0.30	10	0.28	12	0.22	12	0.26	10	Cyclometric function	0.27
5	**Public transport accessibility**	-0.28	8	-0.18	5	-0.14	5	-0.17	4	Third degree polynominal	0.33
6	**Parking**	0.49	27	0.38	22	0.30	22	0.38	21	Linear function	
7	**Auxiliary rooms**	0.16	3	0.14	3	0.11	3	0.21	7	Quadratic function	0.20
8	**Public facilities**	0.15	3	0.19	6	0.13	5	0.17	4	Third degree polynominal	0.25
9	**Positive elements of the surroundings**	0.31	11	0.32	16	0.25	16	0.36	20	Third degree polynominal	0.33
10	Recreational areas	-0.11	1	-0.06	1	-0.04	1	-0.05	0	Third degree polynominal	0.22
11	Heating and gas	0.28	9	0.24	9	0.19	9	0.25	10	Quadratic function	0.38
12	Noise and other nuisances	0.21	5	0.18	5	0.14	5	0.18	5	Linear function	
13	~~Lift~~	0.12	2	0.16	4	0.13	4	0.03	0	Linear function	
14	Window	0.37	16	0.32	16	0.26	17	0.34	17	Linear function	

exposure				
Total:	**102 %**	**100 %**	**100 %**	**99 %**

Source: own study.

Table 3 lists all the determined correlation coefficients (linear and non-linear ones), together with their weights calculated based on the linear correlations (Pearson's and rank correlations). The correlation coefficients that demonstrated statistical significance were highlighted in red. There are also non-linear functions listed, which were selected based on two-dimensional scatter plots, for which the highest values of the curvilinear correlation coefficient (q_i) were obtained.

The weight contributions k_i, resulting from the degree of fitting of the functional model to a specific dependence *price(attribute)*, were calculated based on known Formula (CZAJA 2001):

$$k_i(r_i) = \frac{r_i^2}{\sum_{i=1}^{n} r_i^2} \tag{12}$$

where:

r_i – the largest of the calculated correlation coefficients between the *i* attribute and the price of the property.

In the performed analyses, a limit of 0.05 (5%) of the weight contribution considered to be significant in explaining the scatter of the prices by individual attributes was determined arbitrarily. Therefore, those attributes for which the weight contribution k_i was smaller were considered irrelevant and crossed out in the first column of Table 3 (*Number of Chambers* and *Lift*).

Where the linear regression demonstrates a better fitting, or where the selection of a non-linear function is impossible due to technical reasons (too small diversification of attributes) - the linear relationship is assumed as the optimal one (5 attributes).

The first difference that is observed between the results of the linear correlation analysis and the curvilinear correlation analysis is, in a few cases, a significant increase in the latter ones (e.g. for *Usable floor space, Location,* and others). The use of curvilinear correlations, even for the simplest functional variants, can therefore significantly improve the reliability of the determined weight contributions.

Table 4

Calculation of the optimized and final weight contributions of residential unit attributes based on various types of correlations

i	Attribute	$\hat{r}_{opt}$	$\hat{r}^2_{0pt}$	$k_i(\hat{r}_{opt})$	$k_{ost}(\hat{r}_{opt})$
1	~~Usable floor space~~	0.21	~~0.0441~~	4	
2	Location	0.28	0.0784	7	8
3	Storey	0.30	0.0900	8	9
4	Public transport accessibility	0.33	0.1089	9	11
5	Parking	0.49	0.2401	20	24
6	~~Auxiliary rooms~~	0.21	~~0.0441~~	4	
7	Public facilities	0.25	0.0625	5	6
8	Positive elements of the surroundings	0.36	0.1296	11	13
9	~~Recreational areas~~	0.22	~~0.0484~~	4	
10	Heating and gas	0.38	0.1444	12	15
11	~~Noise and other nuisances~~	0.21	~~0.0441~~	4	
12	Window exposure	0.37	0.1369	12	14
	Total:		**1.1715**	**100 %**	**100 %**
	The sum limited to the significant ones:		**0.9908**		

Source: own study.

In Table 3, the correlation coefficient which had finally been used to determine the optimized weight contributions (Table 4) was underlined for each of the attributes. Therefore, Table 4 contains the results of the last two stages of weighting calculations: after the first elimination of two non-significant attributes and after the selection of the optimal (the strongest, i.e. with the maximum absolute value) correlation dependence and after the second elimination of the four relatively least significant attributes (in relation to the adopted threshold of 0.05), and then the recalculation of the standardized (adding up to 1) weight contributions. The reasons behind the second elimination results not only from the low relative weight contributions after the optimization, smaller than 0.05 (5%), but also from the high sum of squares of the optimized correlation coefficients equal to 1.1715, indicating the existence of redundant attributes. The removal of four variables reduces this sum to the desired optimal value, i.e. close to 1 (0.9908).

Ultimately, out of 15 attributes analyzed on the housing market, we obtain 8 that are recognized as pricing attributes. As illustrated by the values in the last column of Table 4, compared with weight contributions k_i calculated from different types of linear correlations in Table 3, there was a significant increase in the estimated weight contribution in the case of three attributes, after taking into account the non-linear correlation of a given attribute with the price of the residential units (*Location, Transport accessibility* and *Heating and gas*). This means that more than one-third of the market information would be estimated incorrectly in terms of the specific gravity that it brings if we were limited to basic statistical tools. The risk of information loss cannot be underestimated either, which in the case of the first two attributes would be particularly counter-intuitive, as it is against common belief that *Location* and *Transport accessibility* are among the most important attributes affecting the prices of residential units in every city.

6. Conclusions

The study is a continuation of the author's research on the key issue of the real estate market analysis, i.e. the estimation of the so-called weights of market attributes. In previous studies, the problem of the effects of oversimplification in the analysis of the real estate market has been signalled. Specific examples of the analyses have demonstrated significant discrepancies in the estimation of weights, including non-linear dependencies which are not uncommon in the real estate market, in relation to the weights estimated assuming the homogeneity of the market.

This study suggests the optimization of determining the weights of market attributes based on their gradual estimation, with the simultaneous use of all commonly used types of correlations: Pearson's linear correlations, rank correlations and curvilinear correlations. This allows the different character of variables describing the market and the different nature of dependencies between them to be accounted for in one analysis. It turns out that, even if the dependence of price on a given variable can be considered linear, it is not necessarily the standard Pearson correlation coefficient that is the optimal measure of this relationship.

7. Bibliography

ADAMCZEWSKI Z., 2011, *Elementy modelowania matematycznego w wycenie nieruchomości, Podejście porównawcze,* Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa (Elements of mathematical modelling in real estate valuation, comparative approach, Publishing House of the Warsaw University of Technology, Warsaw).

BARAŃSKA A., 2010, *Statystyczne metody analizy i weryfikacji proponowanych algorytmów wyceny nieruchomości,* WYDAWNICTWA AGH, KRAKÓW (Statistical methods for the analysis and verification of the proposed algorithms in real estate appraisal, publications AGH, Krakow).

BARAŃSKA A., 2018, *Linear and nonlinear correlation in calculation of weights of real estate features,* 18th International Multidisciplinary Scientific GeoConference SGEM Albena, Bulgaria 2018.

BITNER A., 2010, *O użyteczności metod statystycznych w wycenie nieruchomości*, Polska Akademia Nauk, Kraków, Komisja Technicznej Infrastruktury Wsi, nr 12 (On the usefulness of statistical methods in property valuation. Polish Academy of Sciences, Krakow, Technical Committee of Rural Infrastructure).

CELLMER R., SZCZEPANKOWSKA K., 2015, *Use of Statistical Models for Simulating Transactions on the Real Estate Market*, Real Estate Management and Valuation, Vol. 23, No. 2, pp. 99-108.

CZAJA J., 2001, *Metody szacowania wartości rynkowej i katastralnej nieruchomości,* KOMP-SYSTEM, Kraków (Valuation methods of market and cadastral value of real estate, KOMP-SYSTEM, Krakow).

DOSZYŃ M., 2017, *Statistical Determination of Impact of Property Attributes for Weak Measurement Scales,* Real Estate Management and Valuation, 25(4), pp. 75-84.

DOSZYŃ M., GNAT S., 2017, *Econometric Identification of the Impact of Real Estate Characteristics Based on*

Predictive and Studentized Residuals, Real Estate Management and Valuation, vol. 25, no. 1, pp. 84-92.
GACA R., 2018, *Parametric and Non-Parametric Statistical Methods in the Assessment of the Effect of Property Attributes on Prices,* Real Estate Management and Valuation, vol. 26, no. 2, pp. 83-91.
GACA R., SAWIŁOW E., 2014, *Zastosowanie współczynników korelacji rang Spearmana do ustalenia wag cech rynkowych nieruchomości*, Rzeczoznawca Majątkowy, nr 82, pp. 24-30 (Application of Spearman's rank correlation coefficient for establishing ranks of real estate characteristics, Real Estate Appraiser, no 82, pp. 24-30).
GREŃ J., 1978, *Statystyka matematyczna, modele i zadania*, Wydawnictwo Naukowe PWN, Warszawa (Mathematical Statistics, models and tasks, Scientific Publishers PWN, Warsaw).
KRAWCZYK A., SŁOMKA T., 1982 - *Podstawowe metody modelowania w geologii. Materiały pomocnicze do ćwiczeń*. AGH Kraków, s. 186.
MALETA M., 2013, *Methods for Determining the Impact of the Trend in the Valuation of Land Property*, Real Estate Management and Valuation, vol. 21, no. 2, pp. 29-36.
MCALLISTER P., 2007, *Valuation accuracy: A contribution to the debate*, Journal of Property Research, Vol. 12, Issue 3, pp. 203-216.
NÚÑEZ TABALES J., REY CARMONA F., CARIDAD Y OCERIN J. Mª., 2013, *Implicit Prices in Urban Real Estate Valuation*, Revista de la construcción, Vol. 12, No 2, pp. 116-126.
Rozporządzenie Rady Ministrów z dnia 21 września 2004 r. *w sprawie wyceny nieruchomości i sporządzania operatu szacunkowego*. Dz. U. nr 207, poz. 2109 z późniejszymi zmianami (Regulation of the Council of Ministers of 21 September 2004 r. On real estate valuation and estimation. Journal of Laws no 207, item 2147 with later changes).
SANCHEZ-RODRIGUEZ M. I., 2014, *Modeling and partial least squares approaches in OODA,* Biometrical Journal, Vol. 56, No 5, pp. 771-773.
SAWIŁOW E., 2010, *Problematyka określania wartości nieruchomości metodą analizy statystycznej rynku* (The problems of qualifying the value of real estate with the method of the statistical analysis of the market), Studia i Materiały Towarzystwa Naukowego Nieruchomości, Vol. 8, No. 1, pp. 21-32.
ZAVADSKAS E. K., TURSKIS Z., USTINOVICHIUS L., SHEVCHENKO G., 2010, *Attributes Weights Determining Peculiarities in Multiple Attribute Decision Making Methods*, Inzinerine Ekonomika-Engineering Economics, vol. 21, nr 1: 32-43.
Ustawa *o gospodarce nieruchomościami* z dnia 21 sierpnia 1997 r. tekst jednolity: Dz.U. 2016, poz. 2147, z późn. zmianami (Act of 21 August 1997 *on the real estate economy*, Journal of Laws, 2016, item 2147, as amended).