DOI 10.1515/gein-2015-0004

GEOINFORMATICA POLONICA
14: 2015

Anna Barańska*

TWO-STAGE MODELLING OF RANDOM PHENOMENA

AGH University of Science and Technology in Krakow,
Faculty of Mining Surveying and Environmental Engineering, Department of Geomatics
abaran@agh.edu.pl

Keywords: modelling of random phenomena, specifying model value

Abstract

The main objective of this publication was to present a two-stage algorithm of modelling random phenomena, based on multidimensional function modelling, on the example of modelling the real estate market for the purpose of real estate valuation and estimation of model parameters of foundations vertical displacements. The first stage of the presented algorithm includes a selection of a suitable form of the function model. In the classical algorithms, based on function modelling, prediction of the dependent variable is its value obtained directly from the model. The better the model reflects a relationship between the independent variables and their effect on the dependent variable, the more reliable is the model value. In this paper, an algorithm has been proposed which comprises adjustment of the value obtained from the model with a random correction determined from the residuals of the model for these cases which, in a separate analysis, were considered to be the most similar to the object for which we want to model the dependent variable. The effect of applying the developed quantitative procedures for calculating the corrections and qualitative methods to assess the similarity on the final outcome of the prediction and its accuracy, was examined by statistical methods, mainly using appropriate parametric tests of significance. The idea of the presented algorithm has been designed so as to approximate the value of the dependent variable of the studied phenomenon to its value in reality and, at the same time, to have it "smoothed out" by a well fitted modelling function.

DWUETAPOWE MODELOWANIE ZJAWISK LOSOWYCH

Słowa kluczowe: modelowanie zjawisk losowych, doprecyzowanie wartości modelowej

Abstrakt

Głównym celem niniejszej publikacji była prezentacja dwuetapowego algorytmu modelowania zjawisk losowych, opartego na wielowymiarowym modelowaniu funkcyjnym, na przykładzie modelowania rynku nieruchomości na potrzeby szacowania wartości nieruchomości oraz estymacji parametrów modelu przemieszczeń pionowych fundamentów. Dobór odpowiedniej postaci modelu funkcyjnego to pierwszy etap prezentowanego algorytmu. W klasycznych algorytmach, bazujących na modelowaniu funkcyjnym, prognozą zmiennej zależnej jest jej wartość uzyskana wprost z modelu. Im model lepiej odzwierciedla relacje między zmiennymi niezależnymi i ich wpływ na zmienną zależną, tym wartość modelowa jest bardziej wiarygodna. W niniejszej pracy zaproponowano algorytm postępowania polegający na skorygowaniu wartości uzyskanej z modelu, poprawką losową, wyznaczoną z odchyłek modelu dla tych przypadków, które w osobnej analizie uznano za najbardziej zbliżone do obiektu, dla którego chcemy zamodelować zmienną zależną. Wpływ zastosowania opracowanych procedur ilościowych obliczania poprawki oraz metod jakościowych oceny podobieństwa na ostateczny wynik prognozy oraz jej dokładność, został zbadany metodami statystycznymi, głównie za pomocą stosownych parametrycznych testów istotności. Idea zaprezentowanego algorytmu została tak opracowana, by zbliżyć wartość zmiennej zależnej badanego zjawiska do jej wartości występującej w rzeczywistości, a jednocześnie uzyskać pewne jej „wygładzenie" poprzez dobrze dopasowaną funkcję modelującą.

1. INTRODUCTION

The main objective of this paper is to present a two-stage algorithm of modelling random phenomena, based on multidimensional nonlinear function modelling. The effect of applying the algorithm has been illustrated on the example of real estate valuation as a consequence of modelling the local real estate market and for example, determining the parameters of vertical displacement model for foundations (the issue in the field of engineering geodesy).

Additional research issues to be examined include: the effect of linearization technique of a function of several variables on the estimation results of the function model parameters, the effect of the method of assessing similarity between objects which form a base for modelling on the results of point estimation of random correction of the dependent variable model value, as well as the effect of the method of determining the correction on the result of its point estimation and on the result of the point estimation of the dependent variable final prediction.

2. THE IDEA OF THE ALGORITHM

The idea of the developed algorithm is based on the relationship which is well-known in modelling random phenomena, which defines the so-called rsidual as a difference between the observed value (e.g. the measured one) of a given variable and its model value (e.g. the most probable one, when it is the value of the estimator of the expected value):

$$\delta = P - M \Leftrightarrow P = M + \delta \tag{1}$$

where:
M – model value of the dependent variable,
P – observed value of the dependent variable,
δ – residual resulting from the model form.

While modelling a specific random phenomenon, we rely on a certain set of information containing the actual data, which are the result of a measurement or observation of a given phenomenon. In the case of modelling the real estate market, these are the prices obtained in the actual market turnover. In the case of vertical displacement modelling of foundation, they are observed vertical displacements of benchmarks. In order to ensure the stability of the model parameter estimators and the reliability of its fitting to the actual data, we usually collect much more data than the number of the estimated parameters. Therefore, residuals occur, as the number of the data is greater than the number of the unknowns, and besides, the model will never perfectly reflect the reality. However, it must be remembered that the unlimited increase in the number of the so-called random sample, which is the basis for estimation model parameters, does not endlessly entail the improvement in the model quality. The researches carried out in this field, regarding modelling the real estate market, proves that, in order to retain stability of model reliability coefficients, it is sufficient when the number of data is equal to about four times the number of estimated parameters. Increasing the number of data above this value does not improve significantly the quality of the function model, especially its accuracy parameters (Barańska 2004).

During the next stage of the developed algorithm, residuals obtained at the modelling stage are used to “correct” the value obtained directly from the model (the model value). Basing on the second variant of the formula (1), it is noticeable that in this way we will move closer to the empirical value of the modelled variable. Let us assume the following designations in this relationship, appropriately to the tested scope of data:

$$W = M + L \tag{2}$$

where:
$L = f(\delta)$,
W – final prediction of the dependent variable,
M – model prediction of the dependent variable,
L – random correction which is the residuals function of the model.

In the above formula, the value of L, which serves as a random correction to the value obtained directly from the model (the model value) is the residuals function corresponding to the original database elements which are most similar to the object being modelled.

Another issue which appears at this point, is the method for assessing similarity between objects, relative to the type of the analysed issue. Below, there are examples of methods for the assessing similarity, which have been tested on the real estate market. In the problem of modelling of vertical displacements of foundations, as the most similar can be allowed the measuring points (benchmarks) located closest to the centre of gravity of considered foundation.

The effect of using the developed quantitative procedures for calculating the correction and qualitative methods for assessing similarity, on the final outcome of the prediction and its accuracy, can be examined by statistical methods, mainly using appropriate parametric tests of significance.

Therefore, to sum up, the proposed algorithm of modelling random phenomena can be divided into the following two stages:

Stage I: estimation of the parameters of multidimensional function model, best suited to the actual data (e.g. market values, measured values etc.), and the point estimation of the dependent variable model value.

Stage II: selection of objects which are most similar to the modelled ones and correcting the value obtained from the model – with a random correction determined from residuals, corresponding to the selected objects.

3. FUNCTION MODELLING – STAGE I OF THE ALGORITHM

Stage I of the developed algorithm includes selection of the function model form, suitable to the variability characterizing the analysed issue. In the classical algorithms, based on function modelling, the prediction of the dependent variable is most frequently its value obtained directly from the model. The better the model reflects a relationship between independent variables and their effect on the dependent variable, the more reliable is the model value.

The forms of the tested function models on the example of modelling real estate market include:

a) linear additive model

$$c = a_0 + \sum_{j=1}^{m} a_i \cdot x_i \tag{3}$$

b) non-linear additive model

$$c = a_0 + \sum_{i=1}^{m} f_i(x_i) \tag{4}$$

c) exponential multiplicative model

$$c = a_0 \cdot \prod_{i=1}^{m} a_i^{x_i} \tag{5}$$

d) multiplicative power model

$$c = a_0 \cdot \prod_{i=1}^{m} x_i^{a_i} \tag{6}$$

or on the example of the model of the vertical component of the foundation movements:

$$w = p_0 + p_1 \cdot x + p_2 \cdot y \tag{7}$$

where:

- c, W – dependent variables of the model (e.g. real estate unit price or the vertical component of the displacement field),
- x_i – independent variables of the model (e.g. real estate attribute values),
- a_i, p_i – parameters of the model,
- m – number of independent variables (e.g. price-determining features of real estate),
- f_i – function of dependency of the dependent variable on the selected independent variable (e.g. real estate price on its attribute i),
- (x, y) – benchmarks horizontal coordinates.

In order to estimate the parameters of the models presented above by the least squares method, their non-linear forms had to be subjected to linearization. In the cases where it was possible (multiplicative models), it was carried out in two ways:

• injective function transformation:

$$\ln c = \ln a_0 + \sum_{i=1}^{m} x_i \cdot \ln a_i \tag{8}$$

$$\ln c = \ln a_0 + \sum_{i=1}^{m} a_i \cdot \ln x_i \tag{9}$$

• expansion of the function into a Taylor series:

$$c = \dot{c} + \sum_{i=1}^{m} \frac{\partial c}{\partial a_i} \cdot da_i \tag{10}$$

where:

- $\dot{c}$ – approximate value of the model function, defined for the approximate values of the parameters,
- $\frac{\partial c}{\partial a_i}$ – partial derivative of the model function with respect to the parameter i,
- da_i – correction to the approximate value of the parameter a_i

The least squares method (OLS), used to estimate the parameters of the model, leads to the development of the so-called system of normal equations, solved using iterative numerical procedures. Estimation was each time accompanied by a full variance analysis of unknown parameters of the model, model values of the dependent variable and residuals of the model.

Table 1 contains examples of the results of point estimation of exponential model parameters in the form (5) for the 9 parameters – the column $\hat{a}_i$ and $\sigma(\hat{a}_i)$, having applied two types of linearization – logarithmisation and expansion into a series. The second part of the table are the results of parametric tests of significance comparing the corresponding parameters estimators and their accuracies (standard deviations). The values of the test functions (Z_{obl} or F_{obl}) exhibiting statistical significance of differences (exceeding the tests critical value Z_{tab} or F_{tab}) have been emphasized in bold. It can be seen that each time the linearization method does not significantly affect the resulting parameter value, while it significantly affects the accuracy of its estimation. Each time it is in favour of the linearization by injective transformation.

Table 2 contains the results of point estimation of the model value (real estate price) for the two models, for which it was possible to use two methods of their linearization in the estimation process. Also this time, a series of statistical tests comparing both the models and the methods of their linearization were performed. As above, we do not observe any significant differences between the model values of the dependent variable themselves. However, there are differences in accuracies of their estimation (test function values marked in bold) in favour of the exponential model and function linearization.

Table 1. Comparison of parameters estimation results of a exemplary exponential model after linearization carried out in a different manner
Tabela 1. Porównanie wyników estymacji parametrów przykładowego modelu wykładniczego po różnie wykonanej linearyzacji

	Logarithmisation		Taylor series		tests results			
	$\hat{a}_i$	$\sigma(\hat{a}_i)$	$\hat{a}_i$	$\sigma(\hat{a}_i)$	Z_{obl}	Z_{tab}	F_{obl}	F_{tab}
$\hat{a}_0$	1061,817	140,804	1066,973	261,312	0,02		**3,44**	
$\hat{a}_1$	1,057	0,012	1,074	0,017	0,82		**2,11**	
$\hat{a}_2$	1,052	0,016	1,075	0,023	0,82		**2,06**	
$\hat{a}_3$	0,966	0,014	0,942	0,023	0,90		**2,50**	
$\hat{a}_4$	1,058	0,015	1,085	0,022	1,01	1,96	**2,08**	1,35
$\hat{a}_5$	1,149	0,046	1,160	0,080	0,12		**3,11**	
$\hat{a}_6$	1,056	0,018	1,066	0,027	0,30		**2,21**	
$\hat{a}_7$	0,993	0,001	0,991	0,002	0,71		**2,39**	
$\hat{a}_8$	1,101	0,029	1,077	0,045	0,45		**2,32**	

Table 2. Estimation results of the premises model value
Tabela 2. Wyniki estymacji wartości modelowej lokalu

	Logarithmisation		Taylor series		comparison of linearization methods			
Model	w_M [zł/m^2]	$\sigma(w_M)$	w_M [zł/m^2]	$\sigma(w_M)$	Z_{obl}	Z_{tab}	F_{obl}	F_{tab}
exponential	2153,98	54,43	2181,14	68,79	0,31	1,96	**1,60**	1,35
power	2113,77	79,58	2140,95	86,42	0,23		1,18	1,37

comparison of multiplicative models				
	Logarithmisation w_M [zł/m^2]	Logarithmisation $\sigma(w_M)$	Taylor series w_M [zł/m^2]	Taylor series $\sigma(w_M)$
Z_{obl} , F_{obl}	0,42	**2,39**	0,36	**1,76**
Z_{tab} , F_{tab}	1,96	1,35	1,96	1,35

Each model, prior to its practical use, must be statistically verified. The process of verifying model quality is a separate issue, which is not the subject of this publication. Therefore, the aspects considered in the process of verification will only be mentioned briefly:

1) assessing the degree of compliance of the model with the empirical data,
2) determining a set of independent variables considering a strong influence on the dependent variable,
3) verifying the assumptions regarding stochastic structure of the model.

and the tools which can be used to examine them:

Ad.1) coefficients of: determination, convergence, variability,

Ad. 2) Fisher-Snedecor test of significance of the variables, Student's T-test of significance of each variable separately,

Ad.3) the series test (Stevens) to test the randomness of the model residuals, residual significance test (test for symmetry of residuals distibution), nonparametric test for normal distribution of residuals, parametric test for mean value of residuals, test for homoscedasticity of residuals (e.g the Goldfeld-Quandt test), test for autocorrelation of residuals (e.g. the Durbin-Watson test).

4. CORRECTION OF THE MODEL VALUE – STAGE II OF THE ALGORITHM

The initial part of the stage II of the proposed algorithm involves the selection of objects which are most similar to the one for which we want to accurately estimate the value, which served as the dependent variable in stage I. The selection should be performed from the full database, which was the basis for the estimation of the model parameters in stage I. The issue of assessing similarity may be treated completely independently in the modelling process, that is, taking into account all the identifiable attributes of the objects. Another approach is based on using at this stage only those attributes which in the modelling process proved important in the shaping of the dependent variable of the model. The research studies, whose sample results have been presented in this publication, use the first approach, however, the author does not rule out a limitation to the statistically significant independent variables. It was the subject of other studies.

In the example of modelling the market of premises, the following methods for selecting similar objects (real estate most similar to the valued premises) were employed:

1. Number of consistent attributes (objects, which had at least half of the attributes identical with those of the valued object, were considered to be the most similar; depending on the number of the identical attributes, the weights were assigned determining the degree of similarity).
2. Analysis of relative comparison.
3. Analysis of sequencing (ranking).

Methods 2) and 3) have been described in detail in (Czaja, Parzych 2007) and (Barańska 2010). Basically, the difference between them lies in the relative comparison analysis, where only the occurrence of the difference in the value of a given attribute between the two compared objects is considered, and the analysis of sequencing also takes into account the value of this difference, and on this basis, the objects are sequenced in terms of the degree of similarity to an established reference point.

In the case of displacement model , as the most " similar" to the point for which we determine the vertical translation movement – can be considered the points located closest to the centre of gravity of the foundation.

Having selected the objects which are most similar to the one which is subject to valuation, the proper part of the stage II of the algorithm follows, which is the prediction of the final value of the dependent variable for this object. In the presented example, it is the prediction of the final market value of premises. It was made according to the two variants presented below, while maintaining the idea of the algorithm expressed by the formula (2):

1. Model value corrected by the random correction, determined as the arithmetic mean of residuals for the selected objects (e.g. for real estate),
2. Model value corrected by the random correction, determined as the weighted mean of residuals, corresponding to the objects (real estate) which are most similar (weight matrix = the inverse of the covariance matrix of residuals).

These variants differ in terms of taking (or not taking) into consideration the results of the variance analysis for residuals while correcting the model value

$W = M + L$. Therefore, in each of these approaches, point estimation of the random correction $L \pm \sigma(L)$ is done in a slightly different way, which can be expressed by the following formulas:

Ad.1)

$$L = \frac{\sum_{i=1}^{k} \delta_{w_i}}{k}, \quad \sigma(L) = \sqrt{\frac{\underline{1} \cdot Cov(\delta_w) \cdot \underline{1}^T}{k^2}}, \tag{11}$$

where:

- w_L – random correction to the model value,
- δ_w – vector of model residuals for the selected similar object,
- k – number of the selected objects,
- $Cov(\delta_w)$ – covariance matrix of residuals, corresponding to the selected objects which are the most similar ones, with the dimensions of $k \times k$,
- 1 – row of ones.

Ad.2)

$$L = [\underline{1} \cdot P \cdot \underline{1}^T]^{-1} \cdot [\underline{1} \cdot P] \cdot [\delta_w],$$

$$\sigma(L) = \sqrt{\hat{\sigma}_{0_w}^2 [\underline{1} \cdot P \cdot \underline{1}^T]^{-1}}, \tag{12}$$

where:

$$\hat{\sigma}_{0_w}^2 = \frac{[\delta_w^T P \cdot \delta_w] - L[\underline{1} \cdot P \cdot \delta_w]}{k-1} \tag{12a}$$

the residual variance determined for the group of k selected most similar objects,

$$P = Cov^{-1}[\delta_w] \tag{12b}$$

the weight matrix for the group of k selected most similar objects, dependent on the results of the variance analysis of residuals.

In the case of the model of vertical displacements as the weights we can also take the inverse of the variances for the heights of benchmarks or inverse of the distance between the benchmarks and the foundation's centre of gravity.

These formulas are applicable for linear models, at least with respect to their parameters, and for nonlinear models, linearized by means of expansion into a series. However, for multiplicative power or exponential models, linearized by logarithmic transformation, the point estimate $L \pm \sigma(L)$ is performed as follows:

$$\ln L = [\underline{1} \cdot P \cdot \underline{1}^T]^{-1} \cdot [\underline{1} \cdot P] \cdot [\ln\delta_w],$$

$$\sigma(\ln L) = \sqrt{\hat{\sigma}_{0_w}^2 [\underline{1} \cdot P \cdot \underline{1}^T]^{-1}}, \tag{13}$$

where the residual variance $\hat{\sigma}_{0_w}^2$ and the weight matrix P are expressed by the formulas:

$$\hat{\sigma}_{0_w}^2 = \frac{[\ln\delta_w]^T P \cdot [\ln\delta_w] - \ln(L) \cdot \underline{1} \cdot P \cdot [\ln\delta_w]}{k-1},$$

$$P = Cov^{-1}[\ln \delta_w]. \tag{13a}$$

Then the final, corrected, prediction of the dependent variable we determine as follows:

$$\ln W = \ln M + \ln L \quad \text{or:} \quad W = M \cdot L \tag{14}$$

The following tables 3 and 4 contain the results of the point estimation of the value w_L, which we use to correct the model value of the dependent variable. It plays the role of a random correction (2) or of a correction factor (14), with the multiplicative form of the function model, linearized using the logarithmic transformation.

As it is apparent from the results of the parametric tests of significance included in these tables, the technique for calculating the value of L, having used the established method for assessing similarity of objects, did not affect the results of its point estimation.

In addition, the statistical significance of the differences between the accuracies of the estimated values of L was examined, separately for both linearization methods, comparing the methods of selecting objects which are the most similar to the modelled one. The values of the test functions contained in Table 5 show a significantly higher accuracy of the random correction (or the correction factor) calculated from residuals corresponding to the objects considered as the most similar to the modelled one, based on at least half of the considered attributes. The methods of relative comparison analysis and sequencing proved to be less accurate.

The final results of the two-stage modelling of a random phenomenon, which in the discussed example was the valuation of the market value of real estate, have been contained in Table 6. The results presented there regard the two linearization methods for multivariate model function, the three variants of selecting the objects most similar to the modelled one and two variants of calculating the correction value w_L.

The second part of Table 6 contains a comparison of the final results of using the two-stage algorithm, both in relation to the very estimates of the corrected value of the dependent variable, as well as in relation to its standard deviation. The differentiating factor was a different method for determining the correction L. In some cases, the accuracy of the estimated values differ significantly – always in favour of the use of the random correction, determined as the arithmetic mean of residuals of the selected objects (the last column of Table 6). This suggests a simpler method for correcting the model value of the dependent variable.

Table 3. Factors correcting the model value of the premises after logarithmic linearization
Tabela 3. Współczynniki korygujące wartość modelową lokalu po linearyzacji logarytmicznej

Method for selecting similar objects	w_L as the weighted mean		w_L as the arithmetic mean		Comparing techniques for calculating the factor w_L – test results			
	w_L	$\sigma(w_L)$	w_L	$\sigma(w_L)$	Z_{obl}	Z_{tab}	F_{obl}	F_{tab}
1	0,980	0,029	0,981	0,033	0,02	1,96	1,25	3,18
2	1,052	0,081	1,025	0,047	0,29		2,99	6,39
3	0,992	0,054	0,989	0,049	0,04		1,19	6,39

Table 4. Random corrections to the model value of the premises after linearization using Taylor series
Tabela 4. Poprawki losowe do wartości modelowej lokalu po linearyzacji za pomocą szeregu Taylora

Method for selecting similar objects	w_L as the weighted mean		w_L as the arithmetic mean		Comparing techniques for calculating the random correction w_L – test results			
	w_L [zł/m^2]	$\sigma(w_L)$	w_L [zł/m^2]	$\sigma(w_L)$	Z_{obl}	Z_{tab}	F_{obl}	F_{tab}
1	–55,23	71,66	–54,26	81,66	0,01	1,96	1,30	3,18
2	84,33	153,16	58,89	116,53	0,13		1,73	6,39
3	–38,12	118,75	–41,46	122,52	0,02		1,06	6,39

Table 5. Comparing the accuracy of L for various methods of assessing similarity
Tabela 5. Porównanie dokładności L dla różnych metod oceny podobieństwa

logarithmic linearization											
L as the weighted mean				Critical values of the test				L as the arithmetic mean			
F_{obl}	1	2	3	F_{tab}	1	2	3	F_{obl}	1	2	3
1	–	**17,44**	**7,62**	1	–	3,63	3,63	1	–	**4,66**	**5,10**
2		–	2,29	2		–	6,39	2		–	1,09
3			–	3			–	3			–
Taylor series linearization											
L as the weighted mean				Critical values of the test				L as the arithmetic mean			
F_{obl}	1	2	3	F_{tab}	1	2	3	F_{obl}	1	2	3
1	–	**10,28**	**6,18**	1	–	3,63	3,63	1	–	**4,58**	**5,07**
2		–	1,66	2		–	6,39	2		–	1,11
3			–	3			–	3			–

Table 6. Final predictions of the market values of the valued premises
Tabela 6. Ostateczne prognozy wartości rynkowych wycenianego lokalu

Method for selecting similar objects	Point estimation results $W \pm \sigma(W)$				Results of parametric tests of significance		Classification of techniques for calculating L in accuracy respect
	L as the weighted mean – a)		w_L as the arithmetic mean – b)		$Z_{tab}=Z_{0,05}=1{,}96$ $F_{tab}=F_{0,05}=1{,}37$		
	W [zł/m²]	$\sigma(W)$	W [zł/m²]	$\sigma(W)$	Z_{obl}	F_{obl}	
	after logarithmic linearization						
1	2071,80	99,55	2073,43	104,36	–0,01	1,10	*a* i *b*
2	2223,59	191,47	2166,39	128,77	0,25	**2,21**	*b*, *a*
3	2096,94	138,54	2090,94	130,62	0,03	1,12	*b* i *a*
	after Taylor series linearization						
1	2085,72	112,27	2086,69	118,90	–0,01	1,12	*a* i *b*
2	2225,28	175,86	2199,85	145,08	0,11	**1,47**	*b*, *a*
3	2102,83	146,87	2099,49	149,93	0,02	1,04	*a* i *b*

Algorithm verification variants – summary

- **2** different methods for pre-linearization of nonlinear functions of several variables.
- **3** different methods for assessing qualitative similarity in order to select these real estates which are the most similar to the valued object, from the database which is the basis for the estimation of the model parameters.
- **2** different methods for determining the final prediction of the market value.

5. CONCLUSIONS

Based on numerous variants of the performed analyses, a number of detailed conclusions can be drawn. They have been divided thematically, due to different research aspects and the effect of the factors differentiating the results in the subsequent stages of the presented algorithm. It should be noted that the conclusions have been formulated on the basis of a much more extensive research (regarding modelling a variety of local real estate markets) than presented in this publication, which increases their reliability.

The conclusions on linearization of the function valuation model:

- the method of model linearization does not significantly affect the estimators of its parameters, but it can have a significant effect on the accuracy of these estimators,
- in the case of identifying a significant difference between the variances of the corresponding model parameters estimators, obtained after differently performed pre-linearization of the function, in the vast majority of cases, a significantly smaller standard deviation of the estimator is obtained after linearization performed by injective transformation.

The conclusions on the model value of the dependent variable:

- the method of bringing the model to a linear form does not significantly affect the model value of the dependent variable of the model,
- where differently performed linearization affected the accuracy of the model value, it always was in favour of the linearization using injective function transformation,

The conclusions on the random correction to the model value:

- point estimation results of the random correction do not depend on the technique of its calculation; occasionally we observe significant differences in accuracies of the random corrections – each time it is in favour of the correction calculated as a weighted mean of residuals,

- in the group of methods for assessing similarity, generating more accurate random corrections, there is always a selection of objects which are similar on the basis of at least half of identical attributes.

The conclusions on the final (corrected) value prediction of the dependent variable of the model:

- different methods of function model linearization do not differentiate significantly the final value predictions of the dependent variable of the model, however, they may affect significantly standard deviations of these predictions, whereas it is always the case that linearization by injective transformation leads to more accurate results,
- the largest standard deviation is often associated with a small number of the objects recognized as the most similar to the modelled one,
- in the case when the effect of the method of calculating the value correcting the model value is essential for the accuracy of the final prediction of the dependent variable, more frequently it is in favour of the estimation L as the arithmetic mean of the model residuals for the selected objects most similar to the modelled one.

The idea of the presented algorithm has been designed so as to approximate the value of the dependent variable of the studied phenomenon to its value in reality and, at the same time, to have it “smoothed out” by a well fitted modelling function. The algorithm has been tested so far on various types of the real estate markets (lands, premises) and theoretically analyzed on the example of the issue of determining the vertical displacements of benchmarks representing the foundations of buildings. Probably it could have also be applied in many other issues describing random phenomena subject to modelling.

BIBLIOGRAPHY

Adamczewski Z. 2002. *Nieliniowe i nieklasyczne algorytmy w geodezji*. Warszawa: Oficyna Wydawnicza Politechniki Warszawskiej.

Barańska A. 2004. *Kryteria stosowania modeli stochastycznych w predykcji rynkowej wartości nieruchomości*. Rozprawa doktorska. AGH. Kraków

Barańska A. 2010. *Statystyczne metody analizy i weryfikacji proponowanych algorytmów wyceny nieruchomości*. Kraków: Wydawnictwa AGH.

Barańska A., Michalik S. 2014. *Variants of Modeling Dwelling Market Value*. Real Estate Management and Valuation, vol. 22, no. 3, s. 28–35.

Czaja J. 1997. Modele statystyczne w informacji o terenie. Kraków: Wydawnictwa AGH.

Czaja J. 1996. Wybrane zagadnienia z geodezji inżynieryjnej. Kraków: Wydawnictwa AGH.

Czaja J., Parzych P. 2007. *Szacowanie rynkowej wartości nieruchomości w aspekcie międzynarodowych standardów wyceny*. Kraków: Stowarzyszenie Naukowe im. St. Staszica.

Rao R.C. 1978. Modele liniowe statystyki matematycznej. Warszawa: PWN.