www.degruyter.com/view/j/remav

DE GRUYTER OPEN

THE SIGNIFICANCE OF DATABASE SIZE IN MODELLING THE MARKET OF NON-RESIDENTIAL PREMISES [1]

Anna Barańska, D.Sc., Ph.D.
Faculty of Mining Surveying and Environmental Engineering
AGH University of Science and Technology
e-mail: abaran@agh.edu.pl

Abstract

The market of non-residential premises is the subject of analyses less frequently than the housing market. There are two main reasons which probably contribute thereto. First of all, commercial premises are relatively less frequently objects of trade than dwelling units; secondly, they are more diverse due to their various uses. The category includes garages, office premises, commercial premises, as well as warehouses. Such differences in their uses result in significantly different characteristics, such as surface area.

The article attempts to analyse a selected non-residential segment of the commercial property market in Krakow based on a large set of data (280 objects), referring to the transactions concluded in the last five years. The size of the data enabled the use of multidimensional modelling of the selected market in different size variants. This made it possible to draw reliable conclusions which undermine the widespread belief regarding very limited possibilities of using the method of market statistical analysis in the comparative approach, especially in this segment of the real estate market, as well as in others, where transactions are concluded less frequently than on the housing market.

Keywords: real estate valuation, commercial premises, multidimensional functional modelling, database size.

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

Citation: Barańska A., 2016, *The Significance of Database Size in Modelling the Market of Non-Residential Premises*, Real Estate Management and Valuation, Vol. 24, No. 2, pp. 47-56.

DOI: 10.1515/remav-2016-0013

1. Introduction

In practice, the main problem of using statistical methods for the widely understood analyses of the real estate market, also including the valuation of properties on this market, is the size of the available market data, which is often limited. This problem especially concerns those segments of the real estate market which are less typical than, for example, the residential real estate market of a large city. The study attempts to analyze the problem of the size of the data on the example of modelling the commercial property market, and its influence on predicting the model value of premises.

2. Database of commercial premises

The set of dataused as the basis for the performed analyses, consisted of information regarding transactions involving the following types of premises: commercial, office, garages, warehouses and cellars/basements in the Krakow area. Such premises are not marketed as often as residential units. In

[1] The task was carried out within the scope of statutory research in Department of Geomatics, Faculty of Mining Surveying and Environmental Engineering, AGH Krakow, Poland.

the analyzed period of five years (2010 – 2015), the author managed to acquire information about 280 transactions made in this period, which means an average of 4-5 transactions per month. However, due to the fundamentally different character of the premises intended for various uses, it was not possible to treat this set as a whole. Therefore, the entire set of data was divided into four smaller databases according to their use, which resulted in: 173 commercial premises, 28 office premises, 16 warehouses and 63 cellars or basements. Each of the units was described by 16 attributes, most of which were potentially price-determining qualitative attributes, calibrated in 4- and 5-point scales on average, and the remaining five were of a quantitative character. The qualitative attributes were as follows: the surroundings of the building, technical condition of the building, access to the building, access to public transport, availability of other public facilities, location of the premises in the building, exposure of the premises, parking availability, building technology , type of premises and cadastral unit; with quantitative attributes being: the number of floors, number of rooms, surface area of the premises, distance from the town centre and updated unit price. In the first stage, the property prices were deprived of the influence of the *time* factor, by updating them to a common date immediately after the last transaction.

All of the analyzed attributes of the premises (except for the price) could be used to make up the set of independent variables in the process of model estimation, as they successfully passed the verification of independence when considered in pairs, i.e. *each attribute with each attribute*, at the initial stage of the analysis, by means of correlation analysis. Given the different nature of the variables, both parametric and nonparametric correlations were used at this stage. In the end, all of the above attributes taken into account except for the *type of premises*, because the studies focused exclusively on commercial premises.

Such narrowing of the scope of the research resulted in a set of objects, similar to each other to a large extent, representing the same segment of the commercial property market with the same uses and from the same local market. Thus, it was possible to omit the aspect of similarity assessment, which would have otherwise been very important in the comparative approach to real estate valuation. This aspect was discussed in depth in some of the author's other publications, e.g. BARAŃSKA (2010a, 2011); it was not the subject of the present analysis. Additionally, omitting the issue of similarity is supported by test results contained in the work (BARAŃSKA, NOWAK 2015), which shows the independence of the procedures of functional market modelling and similarity assessment of real estate in the valuation process.

3. Modelling the market of non-residential premises

The main objective of this analysis was to assess the influence of the limited availability of data in the selected segment of the real estate market on the possibility of using market statistical analysis in the comparative approach to property valuation. The studies were limited to a variant of the method involving the functional modelling of the market. Modelling of the analyzed market was performed in the basic variant, that is assuming its approximate homogeneity, estimating the parameters of the multiple linear regression model in the following form:

$$C = a_0 + \sum_{i=1}^{m} a_i \cdot X_i \tag{1}$$

where:

C – unit price of commercial premises,
a_0 – intercept of the model,
X_i – attributes of premises,
m – number of analyzed, potentially price-determining attributes of premises.

The case of definitely heterogeneous markets, in which multidimensional non-linear models can apply, was omitted, since the study aimed to analyze not the nature of the market, but the influence of a varied size of data on point estimation of the model value of the real property under valuation.

The relatively large set of collected data allowed for carrying out analyses for the defined purpose in many variants. These variants involve the estimation of the parameters of the same general form of the model based on different database size from the full set of data by gradual and random, but steady reduction. Due to the above-mentioned sizes of individual subsets relating to various uses of premises, the modelling process was based on the database of commercial premises. Each time, a full accuracy analysis was also performed throughout the entire estimation process.

3. 1. Modelling the commercial property market

In the first stage of modelling the commercial property market, having rejected the outliers, the estimation of the multiple regression model parameters was based on 136 premises. The basic criterion for the acceptance of the estimation results was fitting the model to the market data, which was satisfactory and confirmed by an appropriate statistical test. This formed the basis for predicting the model value of a selected real property. The model value of the same property was estimated each time as an element of the database, which made it possible to compare the estimated result of the model value with its actual market price, equal to 11,420.43 [PLN/m²].

In the next stages, 10 randomly selected properties were removed from the database, accepting only those variants of removal which did not decrease the fitting of the model to the data, expressed as the coefficient of determination R^2:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{C}_i - C_{śr})^2}{\sum_{i=1}^{n}(C_i - C_{śr})^2} = 1 - \frac{\sum_{i=1}^{n}(C_i - \hat{C}_i)^2}{\sum_{i=1}^{n}(C_i - C_{śr})^2} \quad (2)$$

where:

C_i – actual unit market prices of commercial premises,

$\hat{C}_i$ – unit model prices of commercial premises,

$C_{śr}$ – average unit market price,

n – number of premises in the database which is the basis of estimation.

Each time, only those independent variables (attributes of premises) whose parameters turned out to be statistically significant remained in the final form of the model. No new independent variables were added either, which could have artificially inflated the degree of fitting the model to market data. Thanks to such conditioning, the only factor differentiating the individual cases of modelling was the varied size of the market database forming its basis.

Table 1

Results of multi-stage modelling of the commercial property market

n	m	w_m [PLNm²]	$\sigma(w_m)$ [PLN/m²]	$\lambda = \frac{\sigma(w_m)}{w_m}$	R^2	R^2_{sk}	$\hat{\sigma}$	$\delta = c - w_m$ [PLN/m²]
136	5	11481.81	504.38	0.044	0.57	0.55	1789.6	-61.38
126	5	11700.98	515.50	0.044	0.58	0.56	1769.0	-280.55
116	4	11246.08	465.83	0.041	0.57	0.56	1790.6	174.35
106	4	11195.10	464.95	0.042	0.57	0.55	1722.4	225.33
96	4	11211.88	480.32	0.043	0.58	0.57	1717.0	208.55
86	3	10941.80	505.15	0.046	0.53	0.52	1743.7	478.63
76	3	11220.48	505.15	0.045	0.57	0.55	1705.6	199.95
66	3	11095.22	575.97	0.052	0.56	0.53	1749.9	325.21
56	3	11755.07	660.01	0.056	0.58	0.56	1672.9	-334.64
46	3	12071.75	588.25	0.049	0.71	0.69	1317.1	-651.32
36	4	12516.56	621.49	0.050	0.78	0.75	1196.2	-1096.13
26	5	12076.56	614.41	0.051	0.85	0.81	1111.1	-656.13
16	3	11994.01	677.80	0.057	0.85	0.81	1166.6	-573.58

Source: own study.

It is well known that the significance of the coefficient of determination depends on the size of the data on which the estimation of model parameters is based and on the number of variables that make up the model. Therefore, in each case, the calculated absolute value of the coefficient of determination

R^2 was corrected to the reliable value R^2_{sk} by the relation (3). This eliminates cases of an artificially inflated value of the coefficient of determination resulting from an unfavourable relation between the size of the data and the number of unknowns in the estimation process.

$$R^2_{sk} = 1 - \frac{n-1}{n-m-1} \cdot R^2 \tag{3}$$

where: m, n – denotations as in Formulas (1) and (2).

Table 1 contains only those estimation results that were assessed positively in the following steps of reducing the size of the market data. The columns, following the size of the data n and the number of significant explanatory variables m, include the following: the model value of the selected same real property W_m, its standard deviation $\sigma(W_m)$, the coefficient of variation λ, the coefficient of determination of the model R^2 and its adjusted value R^2_{sk}, the standard error of estimation (square root of the residual variance of the model) $\hat{\sigma}$, and the so-called rest of the model, which is the difference between the actual value and the model value of the valued property.

3. 2. Graphical analysis of the results of modelling the commercial property market

In order to illustrate the dependence of the results of modelling the commercial property market on the factor differentiating its various cases, scatter plots of the selected endpoint parameters were prepared, compiled in Table 1, depending on the size of the set of market data. For each of the scatter plots, the line of the trends of changes was estimated by the method of least squares.

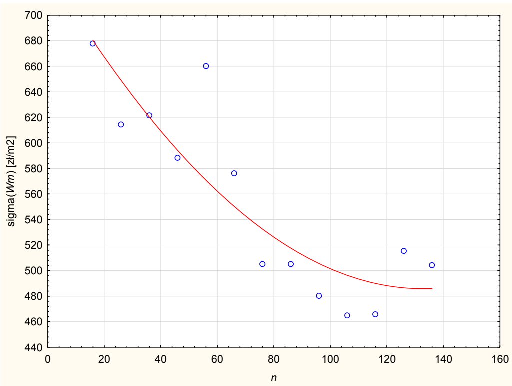

Figure 1. Dependence of the accuracy of the model value of a given premise on the size of the database

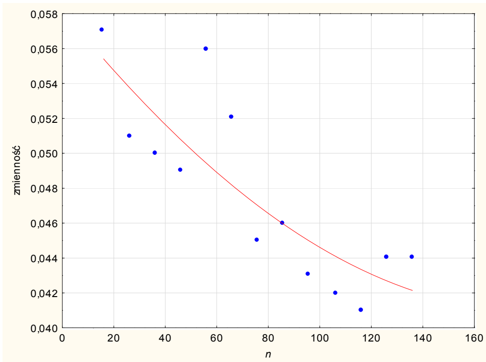

Figure 2. Dependence of the coefficient of variation of the model value of a given premise on the size of the database

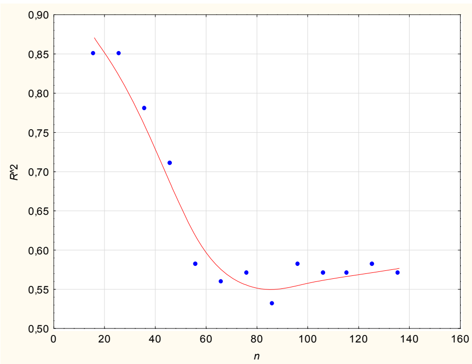

Figure 3. Dependence of the absolute value of the coefficient of fitting the model to the market data on the size of the database

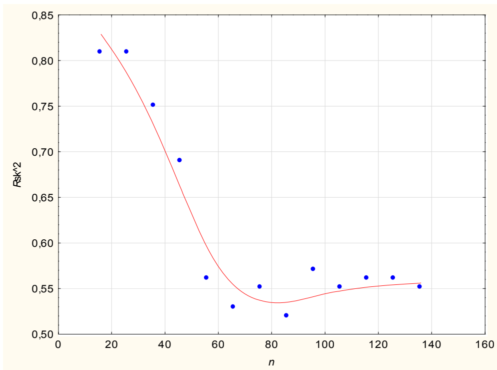

Figure 4. Dependence of the reliable value of the coefficient of fitting the model to the market data on the size of the database

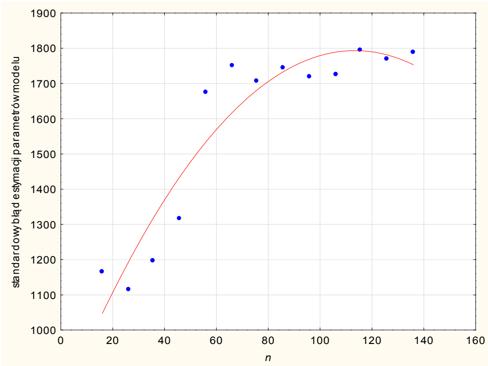

Figure 5. Dependence of the value of the standard error of estimation on the size of the database

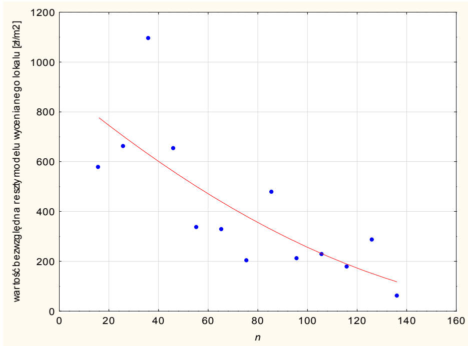

Figure 6. Dependence of the residual of the model value of a premises on the size of the database

Source: own study.

The charts presented above illustrate many interesting relationships. With the increase in the size of the market data on which we can model a selected real estate market, the standard error of the estimation of the model parameters increases as well (Figure 5). At first glance, this seems to be a negative trend, but easily explainable at the same time: more data usually means greater diversity. Therefore, that what depends directly on the diversity of the data, increases. However, at the same time, it is worth noting that the increasing standard error of the estimation does not necessarily entail a decline in the accuracy of estimating the model value of a selected premise (Figure 1). On the contrary, accuracy increases with the increase in the data size. There is also an increase in the precision of determining the model value, evaluated by its compliance with the actual market price, and so the residual of the model value decreases (Figure 6). However, its value in each case remains at a good level, not exceeding 10% of the premise price of 11,420.43 [PLN/m²].

The coefficient of variation for the result of the model value point estimation clearly decreases (Figure 2). This is a direct consequence of the improving accuracy of this estimation. However, more importantly, observing the values of λ and δ (Table 1), it appears that, for any value of n, we are able to model the market value of a real property with a very good coefficient of variation, oscillating in the range of 4-5%, and reflect the local market very well (small residual). Therefore, the large size of market data is not an absolute requirement to obtain reliable valuation results using statistical methods, although, undoubtedly, it is a great convenience.

An interesting regularity can be observed in Figures 3 and 4: the dependence of the coefficient of determination (its absolute and adjusted values) of the market data size; a greater diversity of large data sets generally results in its decrease. However, it is clear that, from a certain size, fitting the model to the data remains more or less constant.

3.3. Statistical analysis of the results of modelling the commercial property market

Figures 1 to 6 distinctly depict a monotonic character of changes in the selected parameters characterizing the quality of modelling the market of non-residential premises. In the previous section, on the example of the values of the coefficient of variation and model residuals, it was pointed out that visual perception needs to be clarified by a closer look at the specific values so as not to draw too hasty conclusions. In the context of the objective of this study defined at the beginning, hasty inference regarding the absolute requirement of a large size of data for the application of statistical methods is mentioned.

This section focuses on the precise, statistical assessment of the significance of differences between the final estimation results of the model value of premises. A series of parametric tests of significance were performed, which were at a standard level of significance equal to 0.05, assessing both the differences between the accuracies of the obtained model values $\sigma(w_m)$ and the differences between the values themselves w_m. Therefore, the following hypotheses were subject to verification, in relation to the accuracy of predictions:

$$H_0: \sigma^2(w_{m_i}) = \sigma^2(w_{m_j})$$
$$H_1: \sigma^2(w_{m_i}) > \sigma^2(w_{m_j})$$
(4)

and in relation to the predictions of model values themselves:

$$H_0: w_{m_i} = w_{m_j}$$
$$H_1: w_{m_i} \neq w_{m_j} \tag{5}$$

Therefore, comparing the accuracies, we are interested in their actual increase or decrease, and by comparing the predictions of the model value, we are interested in their constancy only. Information on the increase or decrease does not contribute anything here, because it does not apply to the variability in time.

As part of the verification of the hypotheses (4) in pairs, each variant of modelling with each one - the values of the *F*-statistic of the Fisher-Snedecor distribution were calculated:

$$F = \left(\frac{\sigma(w_{m_i})}{\sigma(w_{m_j})}\right)^2 \geq 1 \tag{6}$$

and compared with the critical values of this distribution, read from the statistical tables. The calculation results of the test function (6) are contained in Table 2. The values pointing to the statistical significance of the analysed difference between the variances of the model values have been marked in red.

Table 2

The values of the *F*-statistic in the test comparing the accuracy of the model values of a selected premises

n		136	126	116	106	96	86	76	66	56	46	36	26
	m	5	5	4	4	4	3	3	3	3	3	4	5
136	5	-											
126	5	1.04	-										
116	4	1.17	1.22	-									
106	4	1.18	1.23	1.00	-								
96	4	1.10	1.15	1.06	1.07	-							
86	3	1.00	1.04	1.18	1.18	1.11	-						
76	3	1.00	1.04	1.18	1.18	1.11	1.00	-					
66	3	1.30	1.25	**1.53**	**1.53**	1.44	1.30	1.30	-				
56	3	**1.71**	**1.64**	**2.01**	**2.02**	**1.89**	**1.71**	**1.71**	1.31	-			
46	3	1.36	1.30	**1.59**	**1.60**	1.50	1.36	1.36	1.04	1.26	-		
36	4	1.52	1.45	**1.78**	**1.79**	**1.67**	1.51	1.51	1.16	1.13	1.12	-	
26	5	1.48	1.42	**1.74**	**1.75**	1.64	1.48	1.48	1.14	1.15	1.09	1.02	-
16	3	1.81	1.73	**2.12**	**2.13**	**1.99**	1.80	1.80	1.38	1.05	1.33	1.19	1.22

Source: own study.

The results presented in Table 2 show that significant differences in the results of the accuracy analysis of the estimated model value are observed for data sets of fewer than 70 elements. By definition, these are disadvantageous differences, which increase the standard deviation of the model value along with a decrease in the size of the data available in the process of market modelling (c.f. values in Table 1). But do we also observe a negative effect of decreasing database size on the actual model values?

In order to verify the hypotheses (5), one of the following three statistics (6a), (6b), (6c) was used, depending on the size of the data from which a given model value was calculated, and depending on the test results comparing the variances. Therefore, when compared values are derived from a random sampling of over 30 elements, we use the statistics of the normal distribution:

$$Z = \frac{w_{m_i} - w_{m_j}}{\sqrt{\sigma^2(w_{m_i}) + \sigma^2(w_{m_j})}} \tag{6a}$$

When at least one of the random samples consists of up to 30 elements, and in addition the standard deviations of the comparable values can be considered to be statistically the same, that is, there is no reason to reject the null hypothesis H_0 in the hypotheses (4), we use the statistics of the Student's T-distribution:

$$T = \frac{w_{m_i} - w_{m_j}}{\sqrt{\frac{n_i + n_j}{n_i \cdot n_j} \cdot \frac{k_i \cdot \hat{\sigma}_i^2 + k_j \cdot \hat{\sigma}_j^2}{k_i + k_j}}} \tag{6b}$$

When at least one of the random samples is small and, at the same time, we reject the null hypothesis from the hypotheses (4) – we use the statistics of the Cochran-Cox distribution:

$$C = \frac{w_{m_i} - w_{m_j}}{\sqrt{\sigma^2(w_{m_i}) + \sigma^2(w_{m_j})}} \tag{6c}$$

where:

w_{m_i}, w_{m_j} – model values of the selected premises from two different modelling variants,

$\sigma^2(w_{m_i}), \sigma^2(w_{m_j})$ – variances of the compared model values,

$\hat{\sigma}_i, \hat{\sigma}_j$ – standard errors of estimation in different modelling variants,

n_i, n_j – database sizes which are the basis for the estimation of model parameters,

k_i, k_j – number of degrees of freedom in various modelling variants.

The values of these statistics have been compared with the critical values for the significance level of 0.05. The statistic (6a) was related to the critical value in the normal distribution $z_{0.05}$, independent of the degrees of freedom, corresponding to the quantile $z(0.975) = 1.96$. The statistic (6b), was compared with the critical values of the Student's T-distribution $t_{0.05;k=k_i+k_j}$, contained in Table 3. The critical values for the statistic (6c) were calculated from the formula:

$$c_{0.05;k_i,k_j} = \frac{\sigma^2(w_{m_i}) \cdot t_{0.05;k_i} + \sigma^2(w_{m_j}) \cdot t_{0.05;k_j}}{\sigma^2(w_{m_i}) + \sigma^2(w_{m_j})} \tag{7}$$

where:

$t_{0.05;k_i}, t_{0.05;k_j}$ – critical values of the Student's T-distribution for the significance level of 0.05 and the number of the degrees of freedom k_i , k_j.

Table 3

Critical values relating to the statistics *T, C*, in the test comparing the model values

n		136	126	116	106	96	86	76	66	56	46	36	26
	m	5	5	4	4	4	3	3	3	3	3	4	5
26	5	1.976	1.977	2,048	2.049	1.982	1.982	1.987	1.990	1.994	1.999	2.008	-
16	3	1.977	1.978	2,116	2.116	2.114	1.987	1.988	1.992	1.998	2.005	2.017	2.037

Source: own study.

The values of the test functions, accordingly selected in each comparison pair from (6a) – (6c), have been compiled in Table 4. The values indicating the statistical significance of the studied difference have been highlighted in red.

Significant differences between the model values appear only if the database size falls below 30 elements. It is, however, worth proving that a significant difference between the model values does not necessarily mean a significant difference between the model value and the actual market price of the premises, which for the analysed premises was 11,420.43 [PLN/m²] and is the main point of reference here.

Table 4

Values of the statistics *Z, T, C* in the test comparing the model values of the selected premises

n		136	126	116	106	96	86	76	66	56	46	36	26
	m	5	5	4	4	4	3	3	3	3	3	4	5
136	5	-											
126	5	0.304	-										
116	4	-0.343	-0.655	-									
106	4	-0.418	-0.729	-0.077	-								
96	4	-0.388	-0.694	-0.051	0.025	-							
86	3	-0.756	-1.052	-0.443	-0.369	-0.387	-						
76	3	-0.366	-0.666	-0.037	0.037	0.012	0.390	-					
66	3	-0.505	-0.784	-0.204	-0.135	-0.156	0.200	-0.164	-				
56	3	0.329	0.065	0.630	0.694	0.665	0.979	0.643	0.753	-			
46	3	0.761	0.474	1.100	1.169	1.132	1.457	1.098	1.186	0.358	-		
36	4	1.293	1.010	1.636	1.703	1.661	**1.966**	1.618	1.677	0.840	0.520	-	
26	5	1.620	1.031	1.077	1.144	**2.407**	**3.093**	**2.362**	**2.620**	0.881	0.016	-1.469	-
16	3	1.110	0.641	0.909	0.972	0.941	**2.274**	1.715	1.932	0.530	-0.208	-1.464	-0.229

Source: own study.

Let us verify the following hypotheses at the end of the discussion:

$$H_0: w_m = 11420{,}43 \quad [\text{zł}/\text{m}^2]$$

$$H_1: w_m \neq 11420{,}43 \quad [\text{zł}/\text{m}^2] \qquad (8)$$

The statistic of Formula (9) was used for verification, which, depending on the size of the data set, has a normal distribution (more than 30 elements) or Student's T-distribution (up to 30 elements). Table 5 contains values of the statistic (9) for each *n* and the critical values which they were compared with.

$$\frac{\hat{w}_m - 11420{,}43}{\sigma(\hat{w}_m)} \qquad (9)$$

Table 5

Values of the statistic and critical values in the test comparing model values with the market price

n	*m*	Statistics	Critical value
136	5	0.122	1.960
126	5	0.544	1.960
116	4	-0.374	1.960
106	4	-0.485	1.960
96	4	-0.434	1.960

86	3	-0.948	1.960
76	3	-0.396	1.960
66	3	-0.565	1.960
56	3	0.507	1.960
46	3	1.107	1.960
36	4	1.764	1.960
26	5	1.068	2.086
16	3	0.846	2.179

Source: own study.

None of the values of the test function (9) exhibit a statistically significant difference between the model value in a given modelling variant and the actual market price of the selected premises. Therefore, even with a very limited set of data, we are able to achieve a reliable prediction of the market value of a real property using algorithms fitted in the method of market statistical analysis. The absolute condition, however, is sufficient fitting of the functional model to the market database on which we estimate its parameters.

4. Conclusions

The research objective, formulated at the beginning of the article, was an attempt to demonstrate that a sufficiently large size of the market database on transactions involving real estate of a specific type on the local market is not an absolute condition for using algorithms based on statistical methods in the process of property valuation.

In order to be able to unambiguously abstract the influence of the factor of the database size on predicting the market value of real estate, the potential effect of other factors was eliminated, considering the same general form of the multidimensional functional model, the same type of non-residential premises (commercial premises), a fixed set of attributes considered to be essential in this market at the stage of its comprehensive analysis (for the full set of data comprising 173 premises) and predicting the model value of the same premises in each modelling variant. Each time, with a randomly but steadily decreased set of input data (by 10 elements), only those estimated models were accepted, which exhibited good fitting to the market data. In this way, it was possible to obtain 13 variants of the point estimate of the model value of the selected premises.

Studies on the significance of differences performed at the very end, both of the predictions of values and their accuracy, and of predictions regarding the market price of the premises, revealed the greatest influence of the data size on the accuracy of the estimate, a smaller one on the diversity of the model value prediction, and a minimal one on what is most essential, namely the similarity of the model value to the market price.

The performed analyses demonstrate that it is possible to reliably use more advanced algorithms in the process of property valuation, that is, such that fit the method of market statistical analysis, including markets which are less typical, with less frequently concluded transactions. This conclusion may be considered to be a special case of one of the conclusions involving the use of the Monte Carlo simulation method to verify the restrictiveness of assumptions regarding the required minimum number of random samples when using selected statistical tools. At this point, it is worth mentioning the Polish mathematician Stanislaw Ulam from the famous Lviv School of Mathematics, author of the Monte Carlo method. This is, undoubtedly, one of the statistical methods which, as other researchers studying their use in comparative approach note (e.g. Cupal 2014, Sawiłow 2010), creates new prospects for the improvement of algorithms possible to apply within this approach to property valuation.

However, it is important to remember not only the advantages, but also the disadvantages of the Monte Carlo method. The advantages undoubtedly include: the ability to solve difficult problems, a simple form to replace analytical solutions with empirical solutions, the increasing calculating capacity of computers which frees the user from complicated theories and formulas and allows one to focus on the essence of the question to be answered by the statistics. The disadvantage, however, is that experiments take place for a finite number of samples; their results will always be only an approximation, and they depend on random factors providing input data. Therefore, in order to strengthen the reliability of the conclusions formulated above, similar studies should be conducted on other local markets as well.

5. Bibliography

ADAMCZEWSKI Z., 2011, *Elementy modelowania matematycznego w wycenie nieruchomości, Podejście porównawcze* (*Elements of Mathematical Modelling in Real Estate Valuation, Comparative Approach*), Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa, (Publishing House of the Warsaw University of Technology, Warsaw)

ASMUSSEN S., Glynn P.W., 2007, *Stochastic Simulation, Algorithms and Analysis*, Springer

BARAŃSKA A., 2010, *Different Methods for Assessing the Similarity of Real Estate in Two-Stage Estimation Algorithm Based on Multiplicative Functions*, FIG SYDNEY 2010: FACING THE CHALLENGES – BUILDING THE CAPACITY : THE XXIV FIG INTERNATIONAL CONGRESS 2010 : 11–16 APRIL 2010, SYDNEY, AUSTRALIA : PROCEEDINGS. ISBN: 978-87-90907-87-7

BARAŃSKA A., NOWAK D., 2015, *Function Modelling of the Market and Assessing the Degree of Similarity Between Real Properties – Dependent or Independent Procedures in the Process of Office Property Valuation*, REAL ESTATE MANAGEMENT AND VALUATION ; ISSN 2300-5289. Vol. 23, No. 3, pp. 36–46

BARAŃSKA A., 2011, *Qualitative and Quantitative Methods for Assessing the Similarity of Real Estate*, FIG WORKING WEEK 2011: Bridging the gap between cultures: Marrakech, Morocco, May 18–22, 2011 : proceedings. ISBN: 978-87-90907-92-1. —pp. 1–13

BARAŃSKA A., 2010, *Statystyczne metody analizy i weryfikacji proponowanych algorytmów wyceny nieruchomości* (*Statistical Methods for the Analysis and Verification of the Proposed Algorithms in Real Estate Appraisal*), Wydawnictwa AGH, Kraków (AGH Publishers, Krakow)

BITNER A., 2007, *Konstrukcja modelu regresji wielorakiej przy wycenie nieruchomości* (*The Construction of the Multiple Regression Model for the Valuation of Real Estate*), Acta Scientiarum Polonorum, Administratio Locorum, No. 6(4), 59-66

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

CUPAL M., 2014, *The Comparative Approach theory for real estate valuation*, Procedia – Social and Behavioral Sciences 109 (2014), pp. 19-23

GREŃ J., 1978, *Statystyka matematyczna, modele i zadania* (*Mathematical Statistics, Models and Tasks*), Wydawnictwo Naukowe PWN, Warszawa (Scientific Publishers PWN, Warsaw).

ISAKSON H. R., 1998, *The Review of Real Estate Appraisals Using Multiple Regression Analysis*, Journal of Real Estate Research, Vol. 15, Issue 2, pp. 177-190.

KOKOT S., BAS M., (2013), *Evaluation of the Applicability of Statistical Methods in Studies on Price Dynamics on the Real Estate Market*, Real Estate Management and Valuation, Vol. 21, No. 1: 49-58.

MCALLISTER P., 2007, *Valuation Accuracy: A Contribution to the Debate*, Journal of Property Research, Vol. 12, Issue 3, pp. 203-216.

MCCLUSKEY W. J., MCCORD M., DAVIS P. T., HARAN M. & MCILHATTON D., 2013, *Prediction Accuracy in Mass Appraisal: a Comparison of Modern Approaches,* Journal of Property Research, Vol. 30, Issue 4, pp. 239-265.

SAWIŁOW E., 2010, *Problematyka określania wartości nieruchomości metodą analizy statystycznej rynku* (*Problems of Determining the Value of Real Estate with the Method of Statistical Market Analysis*), Studia i Materiały Towarzystwa Naukowego Nieruchomości (Journal of the Polish Real Estate Scientific Society) Vol. 8, No. 1, pp. 21-32

STATISTICA PL, *Poradnik Użytkownika* (User's Guide), 1997, StatSoft, Kraków.

WILEY J. A., LIU Y., KIM D., SPRINGER T., 2014, *The Commercial Office Market and the Markup for Full Service Leases*, Journal of Real Estate Research, Vol. 36 No. 3, USA.