

VARIANTS OF MODELING DWELLING MARKET VALUE¹

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*

Sylwia Michalik, M.Sc., Eng.

*Faculty of Mining Surveying and Environmental Engineering
AGH University of Science and Technology
e-mail: michaliksylwi@gmail.com*

Abstract

The object of this paper is to determine real estate market value on the basis of a multidimensional function model in different variants: A - directly from the model estimated on the basis of a big database, B - from the same model form, but estimated on the basis of a reduced database consisting of dwellings most similar to the estimated one, and C - based on modeled prices corrected by random correction, calculated from random deviations for dwellings most similar to the assessed one. In the framework of statistical inference procedures, the resulting comparison was carried out by parametric significance tests. They were applied to draw conclusions on the analyzed variants.

Keywords: *real estate valuation, multidimensional functional modeling, reduction of database.*

JEL Classification: C1, C3, C4, C5, C8, R00.

Citation: Barańska A., Michalik S., 2014, *Variants of modeling of dwelling market value*, Real Estate Management and Valuation, vol. 22, no. 3, s. 28-35.

DOI: 10.2478/remav-2014-0024

1. Introduction

Real estate market modeling is extremely difficult due to its fluctuations. Hardly ever is the market homogenous; therefore, the classic multidimensional multiple regression model rarely suffices to represent market trends. The second difficulty, occurring at the beginning of the modeling process, is obtaining an adequately large dataset which is to constitute the basis for estimating model parameters. The dataset should be sufficiently large, both with respect to the number of analyzed variables, i.e. price-related features of dwellings in this case, as well as the number of records, i.e. described cases, which includes an adequate number of dwellings that one is able to collect all necessary information about.

The condition concerning the size of the market modeling database is sometimes extremely difficult to meet (variant A). With reference to the above, one should try to provide an answer to the question of whether this is always necessary, or perhaps an equally reliable real estate value projection can be obtained on the basis of a model estimated from a reduced dataset (variant B). However, the reduced dataset cannot be random. The authors will assess whether limiting the number of dwellings to those most closely resembling the priced object will lead to similar projections.

The third of the investigated real estate market value modeling variants is also based on a set of

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

objects most similar to the priced dwelling according to the determined similarity criteria. However, on the basis of this dataset one can only estimate random correction for the modeled value (variant C) as opposed to the final value of the market projection, as in the second scenario.

2. Database

A database – the starting point of the modeling process – contains market information on transactions carried out in the given period of time and taken from notarized deeds or possibly the Price and Value Register, which are considered to be the most reliable sources of transaction prices. Nevertheless, the description of data by means of scalable, potentially price-determining real estate features is the effect of the tedious work of a researcher studying market modeling. All records should be ascribed with “normalized” variable values, describing records in such a way that makes the database, as the modeling foundation, a coherent dataset.

The data for the calculations presented in the present paper come from two cities in the Malopolska region of Poland: Cracow and Tarnow. To collect an adequately large dataset, the authors concentrated on residential units. Transactions were carried out in 2011-2013. Residential units from Tarnow (118 objects) and Cracow (226 objects) were described by means of 11 and 14 attributes respectively, presented in the 3-, 4- or 5-tier scales if the attributes were not numerical. Dwelling transaction prices collected from various databases were corrected to a common date constituting the date of appraisal.

3. Multi-dimensional model

The general form of the multi-dimensional function model tested in various particular forms is as follows:

$$c = a_0 + \sum_{i=1}^m f_i(X_i) \quad (1)$$

where:

c – dwelling price (corrected),

a_0 – absolute term,

X_i – dwelling attributes,

f_i – price-attribute i - dependence function,

m – number of investigated, potentially price-determining dwelling features.

On the basis of price distribution diagrams with respect to given dwelling features, a preliminary selection of the forms of the f_i function was carried out to create a multi-dimensional model of the form (1) for each local real estate market separately. Multiple model parameter estimation was conducted to obtain the best adjustment to market data, measured firstly by means of the coefficient of determination - R^2 . The preliminary value of the coefficient for Tarnow amounted to 0.46, which meant that 46% of the price fluctuations on the selected market were explained by the estimated model. For the Podgorze district of the Cracow market, this value was 58%, probably due to almost twice as much data. To increase R^2 , detached cases and statistically non-significant variables were gradually eliminated, observing the effect of these actions on the level of model adjustment to the data. Eventually, the highest values of the coefficient of determination were obtained, amounting to 0.60 in Tarnow after removing 20 cases (17%) and 4 statistically non-significant variables (36%) for the model in the following form:

$$c = a_0 + a \cdot \log(\text{area}) + b \cdot \exp(\text{loc.}) + c \cdot \exp(\text{surr.}) + d \cdot \exp(\text{trans.}) + g \cdot \exp(\text{tech. state}) + h \cdot \exp(\text{parking}) + i \cdot \log(\text{windows}) \quad (2a)$$

and 0.74 for the Cracow data after eliminating 38 cases (also 17% of the original amount of data) and 8 statistically non-significant variables (57%) for the model in the following form:

$$c = a_0 + a \cdot \text{area} + b \cdot \exp(\text{loc.}) + c_2 \cdot \text{aux. prem.}^2 + c_1 \cdot \text{aux. prem.} + d \cdot \exp(\text{object}) + g \cdot \exp(\text{surr.}) + h \cdot \text{tech. state} \quad (2b)$$

where:

area – dwelling area,

loc. – location,

surr. – dwelling's neighborhood and surroundings,

trans. – public transport accessibility,

tech.state – building’s technical condition,
parking – parking facilities,
windows – window exposure,
aux.prem. – auxiliary premises,
object – public facility.

To confirm that both models were adequate for further calculations of market value of the dwelling, i.e. that both models properly reflected market trends, a parametric test verifying the statistical significance of the models was performed, assuming the following form of the investigated hypotheses:

$$\begin{aligned} H_0: R^2 &= 1 - R^2 \\ H_0: R^2 &> 1 - R^2 \end{aligned} \quad (3)$$

At the same time, it was assumed that the coefficient of determination reflects the variance of a portion of the explained dependable variable (price) by means of the regression model in the general distribution of the variable.

The statistics for the test were determined for both models: $F_T = 19.29$ for (2a) and $F_K = 73.19$ for (2b), according to the following formula:

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-u}{u-1} \quad (4)$$

where:

R^2 – coefficient of determination ($R^2=0.60$ for Tarnow, $R^2=0.74$ for Cracow),

n – number of dwellings in the database ($n=98$ for Tarnow, $n=188$ for Cracow),

u – number of estimated model parameters ($u=8$ in both cases)

and compared with critical values read from statistical tables, and amounting to: $F_{\alpha; k1, k2} = F_{0,05; 7, 90} = 2.11$ for Tarnow, and $F_{\alpha; k1, k2} = F_{0,05; 7, 180} = 2.06$ for Cracow. Therefore, in both cases the critical value was significantly exceeded, which meant the rejection of the zero hypothesis H_0 for the alternative hypothesis H_1 , and was tantamount to a high reliability of both coefficients of determination and related models.

4. Selection of the most similar dwellings

In the second phase of the applied algorithm, a set of dwellings most closely resembling the priced object, one of the elements of the initially aggregated data taken from the complete dataset, was selected. To assess similarity, two qualitative methods were applied: comparative analysis and sequencing analysis, described in detail by CZAJA and PARZYCH (2007).

Both qualitative methods were applied in two variants: taking into account all (respectively 11 and 13) investigated dwelling features, and covering only those dwelling features which proved to be significant in the modeling process (respectively 7 and 6). In doing so, the authors introduced an additional aspect of research, where the independence of quantitative methods (market modeling phase) and qualitative methods (dwelling similarity assessment phase) was verified in the process of dwelling market value projection on the basis of the most similar objects.

Table 1 presents the results of selecting the most similar objects from the complete object dataset. The variants in the table have the following meaning:

- I – comparative analysis covering all attributes,
- II – sequencing analysis covering all attributes,
- III – comparative analysis covering relevant attributes,
- IV – sequencing analysis covering relevant attributes.

It should be noted that in both datasets – regardless of the significant difference in size – the level of the most similar objects to the priced dwelling amounted to 16% on average, which seemed to indicate the homogenous operation of similarity assessment methods irrespective of the size of the original dataset. In table 1, the numbers of dwellings selected in at least three out of four variants have been indicated in bold (on average 25% of selected objects in Tarnow and 38% in Cracow) and objects selected in two out of four variants have been underlined (on average 44% of selected objects in Tarnow and 36% in Cracow). To sum up, there were almost 70% and around 74% similar objects respectively recognized as the most similar in at least two out of four applied variants.

The dwellings recognized as the most similar to the priced objects in all four variants constituted the basis of estimating parameters in models (2a) and (2b), and next, projecting dwelling market value

in variant B, described in the introduction.

Table 1

Tabulated summary of the most similar dwellings in four variants

Variant	Number of selected dwellings	Dwelling numbers	Average transaction price [zł/m ²]
Tarnow			
I	17	15, 16, 19, 22, 34, 35, 38, 60, 63, 72, 92, 94, 95, 97, 100, 105, 116	2,560.02
II	15	15, 16, 19, 28, 34, 44, 47, 51, 56, 60, 84, 97, 103, 105, 107	2,725.05
III	17	15, 16, 27, 29, 38, 41, 44, 51, 72, 75, 90, 93, 95, 107, 108, 113, 117	2,607.39
IV	15	15, 16, 26, 28, 41, 43, 45, 60, 74, 81, 93, 97, 100, 114, 117	2 525.02
Cracow			
I	29	11, 14, 15, 19, 26, 34, 38, 42, 47, 62, 67, 69, 73, 75, 107, 134, 142, 144, 161, 168, 169, 191, 199, 204, 207, 212, 217, 222, 225	4,720.86
II	32	38, 42, 45, 61, 67, 71, 77, 78, 107, 112, 119, 120, 142, 149, 154, 161, 168, 169, 176, 182, 188, 189, 191, 193, 196, 199, 201, 204, 207, 215, 222, 224	4,239.59
III	32	25, 31, 38, 49, 51, 52, 61, 77, 80, 104, 107, 123, 128, 131, 133, 142, 145, 149, 154, 161, 166, 176, 182, 184, 188, 189, 196, 199, 204, 210, 212, 222	4,172.48
IV	23	25, 38, 43, 49, 61, 104, 107, 123, 133, 142, 145, 154, 161, 166, 184, 189, 193, 196, 199, 204, 207, 222, 225	4,031.04

Source: own elaboration.

5. Model parameter estimation on the basis of a reduced dataset

Among the three options presented in the introduction, dwelling market value projection in variant B is carried out by means of determining the (2a) or (2b) function modeled values, estimated on the basis of the most similar object dataset selected from the original database. Due to the fact that, according to table 1, four dwelling datasets were obtained after the application of various similarity assessment methods, the re-estimation of model parameters (2) was also conducted four times, with the coefficients of determination for each case presented in table 2:

Table 2

Coefficients of determination for models (2a) and (2b) for reduced data from table 1

Variant	R ² in Tarnow	R ² in Cracow
I	0.74	0.82
II	0.74	0.61
III	0.25	0.62
IV	0.27	0.55

Source: own elaboration.

The results summarized in table 2 for Tarnow clearly demonstrate that qualitative similarity assessment methods should be applied in the pricing process regardless of quantitative methods, therefore irrespective of the modeling process itself. The model adjustment to the reduced dataset is three times better when all information about dwellings was taken into account in the similarity assessment. Moreover, in the case of Cracow the model adjustment was also better for the application of the similarity assessment method covering all information about qualitative dwelling features (compare variant I with III, and II with IV).

With respect to the above, further calculations were carried out only for models estimated in variants I and II.

6. Dwelling market value projection

6.1. Market value as a modeled value

The parameters of models (2a) and (2b) were estimated together with the complete variance analysis by means of the least-squares method for the complete dataset (variant A - for 98 and 188 dwellings respectively) and for the reduced datasets (variant B - for around 16 and 30 dwellings) selected on the basis of comparative analysis and sequencing analysis of dwellings.

Table 3

Point estimation of dwelling market value in variants A and B

Variant and database size	Residual variance	Square root of residual variance	Priced dwelling value [zł/m ²]	Standard deviation of the modeled dwelling value [zł/m ²]
Tarnow				
a) 98	75,360.55	274.52	2,519.62	40.36
BI 17	44,367.66	210.64	2,591.75	113.50
BII 15	55,453.71	235.49	2,749.20	117.20
Cracow				
b) 188	128,054.65	357.85	3,933.46	63.58
BI 29	83,598.17	289.13	3,858.86	157.44
BII 32	101,131.32	318.01	4,038.38	126.81

Source: own elaboration.

The dwelling value obtained from the largest database proves to be the best in both cases with respect to accuracy, which is understandable providing for the way of calculating projected standard deviation in the least-squares method. However, it is worth noting that all projection accuracies are within 5% of the projection value, which should be regarded as very accurate estimation. Taking into consideration the real, corrected transaction price of the priced dwelling, which amounted to **2,779.95 zł/m²** in Tarnow and **3,697.08 zł/m²** in Cracow, it should be noted that the obtained results slightly diverge from the value, which can be proven by certain significance tests. However, the value closest to the real price was obtained on the basis of model parameter estimation from the database with the most similar objects, i.e. the dwellings selected by means of sequencing analysis in Tarnow and comparative analysis in Cracow, covering all investigated dwelling features.

6.2. Market value as a corrected modeled value

In the third pricing variant C, the final market value was the modeled value obtained from the complete dataset corrected by random correction calculated from model (2) random deviations for the most similar dwellings. The following formula was applied:

$$W = C - \delta \quad \text{equal to} \quad C = W + \delta \quad (5)$$

where:

C – real dwelling price (corrected),

W – modeled dwelling price,

δ – modeled price random deviation.

Therefore, to obtain results closer to the real dwelling price in its market value projection, one can use correction calculated from deviations. In this case, deviations for objects selected as the most similar to the priced dwelling were applied, marked as δ_{w_i} , and the correction P_δ was calculated as the mean of the deviations:

$$P_\delta = \frac{1}{k} \cdot \sum_{i=1}^k \delta_{w_i} \quad (6)$$

where:

δ_{w_i} – model random deviations for the selected, most similar dwellings,

k – number of selected dwellings.

To conduct a complete accuracy analysis, standard deviation of the final projection on the basis of standard deviation of the random correction P_δ was also determined in this variant:

$$\sigma(P_\delta) = \sqrt{\frac{\mathbf{1} \cdot \text{Cov}(\delta_w) \cdot \mathbf{1}^T}{k^2}} \quad (7)$$

where:

$\mathbf{1}$ - row vector consisting of ones,

$\text{Cov}(\delta_w)$ - similar dwelling random deviation covariance matrix, selected from the complete covariance matrix of model random deviations (2).

The complete calculation algorithm can be found in Barańska (2010). Table 4 below presents the corrected modeled values and their accuracies.

Table 4

Tabulated summary of corrections and corrected modeled values

Database size	P_δ	$\sigma(P_\delta)$	Corrected modeled value W_0	$\sigma(W_0)$	Corrected value (W_0+P_δ)	$\sigma(W_0+P_\delta)$	Marking of corrected value
[zł/m ²]							
Tarnow							
17	-19.91	59.62	2,519.62	40.36	2,499.71	99.98	CI
15	146.01	63.09			2,665.63	103.45	CII
Cracow							
29	-14.78	57.71	3,933.46	63.58	3,918.68	121.29	CI
32	-68.29	52.03			3,865.17	115.61	CII

Source: own elaboration.

Similarly to variant B for Tarnow, when market value matched the modeled value determined from the reduced dataset (table 3), the projection obtained by the application of dwelling sequencing analysis to select dwellings most similar to the priced object (variant CII in table 4 for both cities) proved to be the closest to the real dwelling market value. This may indicate the advantage of this qualitative method of similarity assessment over comparative analysis.

7. Result comparison by parametrical significance tests

To reach conclusions about differences between dwelling market value projections in three variants, parametrical significance tests were applied to compare both: projections and their accuracy. For the comparison of market values the following tests were applied: the Student's t-test, Cochran's C test, and the z (normal distribution) test (formulas (8a), (8b), and (8c)) comparing mean values with regard to random sample size and variance comparison test results. Additionally, for the projection accuracy comparison, the variance-comparing Fisher-Snedecor test (formula (9)) was applied. The test functions in the aforementioned tests are as follows:

$$T = \frac{w_1 - w_2}{\sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2} \cdot \frac{k_1 \cdot \sigma^2(w_1) + k_2 \cdot \sigma^2(w_2)}{k_1 + k_2}}} \quad (8a)$$

$$C = \frac{w_1 - w_2}{\sqrt{\sigma^2(w_1) + \sigma^2(w_2)}} \quad (8b)$$

$$Z = \frac{w_1 - w_2}{\sqrt{\sigma^2(w_1) + \sigma^2(w_2)}} \quad (8c)$$

$$F = \frac{\sigma^2(w_1)}{\sigma^2(w_2)} \cdot \frac{k_2}{k_1} \quad F = \frac{\sigma^2(w_1)}{\sigma^2(w_2)} \cdot \frac{k_2}{k_1} \quad (9)$$

where:

w_1, w_2 - compared market values,

$\sigma^2(w_1), \sigma^2(w_2)$ - compared market value variances,

k_1, k_2 - number of degrees of freedom for the compared parameters,

n_1, n_2 - random sample size where w_1, w_2 were determined respectively.

All parametrical significance tests were carried out at the standard significance level of 0.05.

7.1. Comparison of market value accuracy

Due to the fact that the application of a given model in the mean value comparison test depends, in some cases, on the result of the variance comparison test, the accuracies of the particular projections were compared first.

Table 5 presents values of calculated statistics (9) and critical values, which the former were compared to. Red font marks test function values indicating a statistical significance of the investigated differences, and bold font in the first column shows the calculation variant which proved to be the most accurate in the given pair. The last two columns cover numbers of degrees of freedom in each calculation variant applied to calculate the test statistics.

Table 5

Tabulated summary of results of parametrical significance tests comparing market value accuracy

Compared calculation variants	Calculated statistic values (9)		Critical values read from the Fisher's distribution statistical tables		Calculation variants	Number of degrees of freedom	
	Tarnow	Cracow				Tarnow	Cracow
BI-BII	1.37	1.76	3.29	2.02	BI	9	21
CI-CII	1.04	1.08	1.46	1.32	BII	7	24
BI-CI	10.31	12.04	2.02	1.62	CI	72	150
BI-CII	9.90	12.98	2.02	1.62	CII	74	147
BII-CI	14.13	6.83	2.14	1.59	A	90	180
BII-CII	13.57	7.37	2.14	1.59			
A-BI	79.08	52.56	1.99	1.62			
A-BII	108.42	29.84	2.11	1.58			
A-CI	7.67	4.37	1.45	1.30			
A-CII	7.99	4.05	1.45	1.31			

Source: own elaboration.

Table 6

Tabulated summary of results of parametrical significance tests comparing projected market values

Test and calculation variants for comparison	Tarnow	Cracow	Critical values from statistical tables (for the t-test and z-test, or calculated for the C test)		Auxiliary critical values <i>T</i> for the C test			
					Tarnow		Cracow	
					First compared variant	Second compared variant	First variant	Second variant
<i>T</i> BI-BII	3.86	0.00	2.12	2.01				
<i>Z</i> CI-CII	1.15	0.32	1.96	1.96				
<i>C</i> BI-CI	0.09	0.30	2.15	2.04	2.2622	1.994	2.0796	1.9759
<i>C</i> BI-CII	0.48	0.03	2.14	2.04	2.2622	1.992	2.0796	1.9765
<i>C</i> BII-CI	1.62	0.68	2.21	2.02	2.3646	1.994	2.0639	1.9759
<i>C</i> BII-CII	0.53	1.01	2.20	2.02	2.3646	1.992	2.0639	1.9765
<i>C</i> A-BI	0.60	0.44	2.23	2.06	1.9867	2.2622	1.9730	2.0796
<i>C</i> A-BII	1.85	0.74	2.32	2.05	1.9867	2.3646	1.9730	2.0639
<i>Z</i> A-CI	0.18	0.11	1.96	1.96				
<i>Z</i> A-CII	1.31	0.52	1.96	1.96				

Source: own elaboration.

The results presented in the table demonstrate complete regularity: only projections corresponding to each other within variants B and C show statistically uniform accuracies. The results indicate the

highest accuracy of projections of values identical to the modeled values determined from the complete dataset, and the lowest accuracy of projections of values identical to the modeled values determined from the reduced dataset. Dwelling market values estimated by means of random correction are at a medium level as far as accuracy is concerned.

7.2. Comparison of projected market values

The results of the comparison of market value projections determined in three variants have been presented in table 6. With one exception, all values are statistically significant at the same level. Apart from one case marked in red, statistically significant differences between the projections were not observed. This indicates that all variants lead to market value at the same level, regardless of the qualitative method applied to select dwellings most similar to the priced objects from the complete dataset.

9. References

- 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.
- BARAŃSKA A., 2010, *Statystyczne metody analizy i weryfikacji proponowanych algorytmów wyceny nieruchomości (Statistical methods for the analysis and verification of the proposed algorithms in property valuation)*, WYDAWNICTWA AGH, KRAKÓW.
- CZAJA J., PARZYCH P., 2007, *Szacowanie rynkowej wartości nieruchomości w aspekcie międzynarodowych standardów wyceny (Market value estimation of real estate in aspect of international valuation standards)*, Stowarzyszenie Naukowe im. Stanisława Staszica, Kraków.
- GREŃ J., 1978, *Statystyka matematyczna, modele i zadania (Mathematical statistics, models and tasks)*, Wydawnictwo Naukowe PWN, Warszawa.
- KRYSICKI W., BARTOS J., DYCZKA W., KRÓLIKOWSKA K., WASILEWSKI M., 2013, *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach (Probability and Mathematical Statistics in tasks)*, Wydawnictwo Naukowe PWN, Warszawa.