MUNICH, 12 october 2006

Estimation of parameters of multiplicative exponential function model for real estate market value prediction

PhD Anna Barańska
Terrain Information Department
Faculty of Mining Surveying and Environmental Engineering
University of Science and Technology
Krakow, POLAND

INTRODUCTION

The previous methods of estimating real estates assumed "a priori" that a market value was created as a sum of the attribute shares. These shares could be presented by different elementary functions, also non-linear. This assumption is satisfied especially by multiple linear regression model:

$$F(X_i, a) = a_0 + \sum_{k=1}^{u} X_{ik} * a_k$$

However some question arise:

• is the additive form of the market value variability model for all real estates markets optimum?

Among different non-linear functions of many variables, in modelling of a real estate unit price or value, a multiplicative combination of exponential variability's of the particular attributes was analysed.

The advantages of an exponential form of the function:

• always positive value of real estates,
• describing variability as a monotone function,
• possibility to present the market or cadastral value variability in percentages in relation to its value at determined attributes (zero for example).

Estimation of model parameters may be done in many ways. In the present paper, the run system for Markow method (in form of least squares weighing method) will be submitted. And verification of estimated model will be presented.

model of real estates unit values in form of multiplicative exponential function

$$c = B_0 \cdot B_1^{x_1} \cdot B_2^{x_2} \cdot \ldots \cdot B_m^{x_m} \qquad (1)$$

where:

c – *real estate unit price or unit value,*
$x_1, x_2, \ldots, x_m$ – *real estate attributes, including the attribute of transaction time,*
B_j – *estimated model factors,*
B_0 – *real estate unit value, for zero of all attributes.*

ESTIMATION OF MODEL PARAMETERS USING THE LEAST SQUARES WEIGHING METHOD

For the estimation of the coefficients B_j, the function (1) has to be brought to a linear form. For this purpose, we take the logarithms of both sides, using the natural logarithm, and we receive:

$$\ln c = \ln B_0 + x_1 \cdot \ln B_1 + x_2 \cdot \ln B_2 + \ldots + x_m \cdot \ln B_m \qquad (2)$$

The system of equations (2) has the features of a probabilistic model, taking on in matrix notation the following form:

$$\{\ln \delta\} + \{A\} \cdot \{\ln B\} = \{\ln c\} \qquad (3)$$

where:

$\{A\}$ - matrix $[n \times (m+1)]$, containing ones and attributes values of real estates in database,
$\{\ln c\}$ - vector $[n \times 1]$, containing natural logarithms of database real estates prices,
$\{\ln B\}$ - vector $[(m+1) \times 1]$, containing estimated values of logarithms of the model natural coefficients B_j,
$\{\ln \delta\}$ - vector $[n \times 1]$, containing random deviations for natural logarithms of database real estates prices,

where ***n*** is the number of real estates in the database and ***m*** is the number of attributes.

By applying the least squares weighing method, we receive the following formulas for estimated parameters of a non-linear model:

$$\{\ln \hat{B}\} = \left(\{A^T\}\{P\}\{A\}\right)^{-1} \times \{A\}^T \{P\}\{\ln c\} \qquad (4)$$

To determine a covariance matrix for estimated regression coefficients, we use the formula:

$$Cov\{\ln \hat{B}\} = \hat{\sigma}_0^2 \left(\{A\}^T \{P\}\{A\}\right)^{-1} \qquad (5)$$

where σ_0^2 is the variance of the multiplicative non-linear model estimation. Its estimator is given by the following expression:

$$\hat{\sigma}_0^2 = \frac{\{\ln c\}^T \{P\}\{\ln c\} - \{\ln \hat{B}\}^T \{A\}^T \{P\}\{\ln c\}}{n-m-1} \qquad (6)$$

PROCEDURE OF VERIFICATION OF AN ESTIMATED MODEL

a) investigation of model acceptability regarding the values of **the factors variability** and **the convergence**:

- coefficient of variation V:

$$V = \frac{\sigma_\delta}{\hat{c}} \qquad (7)$$

where: σ_δ - standard error of remainders,
$\hat{c}$ - mean value of real estates prices in database,

- coefficient of convergence φ^2:

$$\varphi^2 = \frac{\sum_{i=1}^{n} \delta_i^2}{\sum_{i=1}^{n} (c_i - \hat{c})^2} \qquad (8)$$

where: $\delta_i = c_i - c_{i(p)}$ - remainder for i^{th} real estate (difference between a price observed and a price forecast according to the model).

Boundary values for V and φ^2 are determined arbitrarily.

Mostly they are: $V_0 = 0{,}10$ and $\varphi_0^2 = 0{,}40$.

b) investigating the random components symmetry:

Suppose k is the number of positive (or negative) remainders δ_i. We verify the hypothesis concerning the remainder structure indicator with established sign (so-called fraction p) in the following form: H_0: $p=0.5$, against the alternative hypothesis H_1: $p \neq 0.5$. The test statistics takes on the form:

$$Z = \frac{\left|\frac{k}{n} - 0.5\right|}{\sqrt{\frac{\frac{k}{n}\left(1-\frac{k}{n}\right)}{n-1}}} \qquad (9)$$

If the test indicates the necessity to reject the null hypothesis, the analytic form of the model has to be modified.

c) examining the significance of model coefficients:

- verification of a parameters system significance on the basis of Fisher-Snedecor's statistics, using the following hypothesis H_0: $\sum_{j=0}^{m} B_j^2 = 0$ against the alternative hypothesis H_1: $\sum_{j=0}^{m} B_j^2 \neq 0$ statistic form:

$$F = \frac{1-\varphi^2}{\varphi^2} \cdot \frac{n-m-1}{m} \qquad (10)$$

- verification of particular regression coefficients significance on the basis of T-Student statistics, under the null hypothesis H_0: $B_j = 0$ against the alternative hypothesis H_1: $B_j \neq 0$ statistic form:

$$T = \frac{\ln \hat{B}_j}{\sigma\left(\ln \hat{B}_j\right)} \qquad (11)$$

If for any of explaining variables, the statistical test does not demonstrate reasons for rejecting the null hypothesis, we eliminate this variable from the model and we reestimate the parameters.

FORECASTING A MARKET UNIT VALUE OF A REAL ESTATE

The prediction of a real estate unit price c_p (market value) is determined by substituting to the evaluated model, the values of the attributes a_j. By converting this, we receive the forecast value for the analysed real estate:

$$c_p = \exp\left(\ln B_0 + a_1 \cdot \ln B_1 + a_2 \cdot \ln B_2 + \ldots + a_m \cdot \ln B_m\right)$$

The standard deviation, accordingly to the propagation of errors law, is calculated by formula:

$$\sigma(c_p) = c_p \cdot \sigma\left(\ln c_p\right)$$

EXAMPLE

Results of non-linear multiplicative models estimation for two different real estate data bases:

Model 1

multiplicative model with 13+1 variables:

$$c = B_0 \cdot B_1^{x_1} \cdot B_2^{x_2} \cdot B_3^{x_3} \cdot \ldots \cdot B_{13}^{x_{13}}$$

coefficient of convergence: $\varphi^2 = 0.32$

variance of the multiplicative non-linear model estimation: $\sigma_0^2 = 0.02$

standard error of estimation: $\sigma_0 = 0.14$

Table 1. Parameters of model 1

Number of parameter	$\ln \hat{B}_i$	$\sigma(\ln \hat{B}_i)$	T_{cal}	T_{tab}
0	0,5335	0,1705		
1	0,0165	0,0041	**4,05**	
2	0,0068	0,0018	**3,78**	
3	-0,0148	0,0530	-0,28	
4	0,0308	0,0285	1,08	
5	-0,0908	0,0232	**-3,92**	1,99
6	-0,0186	0,0247	-0,76	
7	0,0427	0,0129	**3,32**	
8	-0,0885	0,0302	**-2,93**	
9	-0,0223	0,0067	**-3,31**	
10	0,0520	0,0276	1,89	
11	0,1070	0,0217	**4,93**	
12	-0,0221	0,0333	-0,66	
13	-0,0485	0,0391	-1,24	

Model 2

multiplicative model with 14+1 variables:

$$c = B_0 \cdot B_1^{x_1} \cdot B_2^{x_2} \cdot B_3^{x_3} \cdot \ldots \cdot B_{14}^{x_{14}}$$

coefficient of convergence: $\varphi^2 = 0.16$

variance of the multiplicative non-linear model estimation: $\sigma_0^2 = 0.01$

standard error of estimation: $\sigma_0 = 0.07$

Table 2. Parameters of model 2

Parameter	$\ln \hat{B}_i$	$\sigma(\ln \hat{B}_i)$	T_{cal}	T_{tab}
0	1,2896	0,1223		
1	-0,0012	0,0033	-0,36	
2	-0,0026	0,0010	**-2,63**	
3	0,0736	0,0229	**3,22**	
4	-0,0357	0,0189	-1,89	
5	-0,0263	0,0136	-1,94	
6	-0,0123	0,0164	-0,75	2,00
7	-0,0110	0,0091	-1,21	
8	0,0515	0,0183	**2,81**	
9	-0,0038	0,0054	-0,72	
10	0,0129	0,0235	0,55	
11	0,1510	0,0182	**8,28**	
12	-0,1033	0,0168	-0,79	
13	-0,0168	0,0155	-1,09	
14	-0,1614	0,0347	**-4,65**	

Table 3. Results of verification of model 1 and 2

Model	1	2
V	0,13	0,08
φ^2	0,32	0,16
k	50	42
Z_{cal}	0,55	0,80
Z_{tab}	1,96	1,96
F_{cal}	12,49	23,63
F_{tab}	1,85	1,86

Model 1a:

$$c = B_0 \cdot B_1^{x_1} \cdot B_2^{x_2} \cdot B_3^{x_3} \cdot \ldots \cdot B_7^{x_7}$$

coefficient of convergence: $\varphi^2 = 0.36$

variance of the multiplicative non-linear model estimation: $\sigma_0^2 = 0.02$

standard error of estimation: $\sigma_0 = 0.15$

Table 4. Parameters of model 1a

Number of parameter	$\ln \hat{B}_i$	$\sigma(\ln \hat{B}_i)$	T_{cal}	T_{tab}
0	0,4223	0,0936		
1	0,0199	0,0036	**5,56**	
2	0,0083	0,0016	**5,32**	
3	-0,0907	0,0213	**-4,26**	1,99
4	0,0395	0,0128	**3,09**	
5	-0,0918	0,0284	**-3,23**	
6	-0,0215	0,0063	**-3,40**	
7	0,1078	0,0190	**5,67**	

Model 2a:

$$c = B_0 \cdot B_1^{x_1} \cdot B_2^{x_2} \cdot B_3^{x_3} \cdot B_4^{x_4} \cdot B_5^{x_5}$$

coefficient of convergence: $\varphi^2 = 0.18$

variance of the multiplicative non-linear model estimation: $\sigma_0^2 = 0.01$

standard error of estimation: $\sigma_0 = 0.07$

Table 5. Parameters of model 2a

Number of parameter	$\ln \hat{B}_i$	$\sigma(\ln \hat{B}_i)$	T_{cal}	T_{tab}
0	1,0533	0,0706		
1	-0,0019	0,0007	**-2,59**	1,99
2	0,0617	0,0182	**3,39**	
3	0,0404	0,0160	**2,52**	
4	0,1628	0,0145	**11,25**	
5	-0,1202	0,0280	**-4,30**	

Comparing coefficients of determination $R^2=1-\varphi^2$ obtained for models 1 and 1a (0,68 and 0,64) as well 2 and 2a (0,84 and 0,82), we can notice that the reduction of number of variables is negligible for quality of model.

Table 6. Results of verification of model 1a and 2a

Model	1a	2a
V	0,14	0,09
φ^2	0,36	0,18
k	48	37
Z_{cal}	0,53	0,34
Z_{tab}	1,96	1,96
F_{cal}	21,53	23,63
F_{tab}	2,13	2,35

The result of verification is positive for both models.

CONCLUSIONS

The example shows, that the model in multiplicative form can be also good in modelling of market value of real estate and can give also high accuracy of prediction results. So, taking account of his advantages, it can be even better than a model in additive form.

An analogical analysis can be done for other methods of model parameters estimation. The equality of parameters, estimated by different methods, may be verified using suitable statistical tests. Within the framework of the comparison of estimation methods, also the covariance matrixes for parameters achieved by different estimation methods should be compared.