APPLICATION OF HUBER AND HAMPEL M-ESTIMATION IN ANALYSING OF REAL ESTATE PRICE VOLATILITY OVER TIME*

Tomasz Adamczyk, Ph.D.
AGH University of Science and Technology
Faculty of Mining Surveying and Environmental Engineering
Department of Geomatics
Krakow, Poland
e-mail: tomasz.adamczyk@agh.edu.pl

Abstract

Determining market value of the property requires a thorough and complete analysis of the real estate market. Analysis of the market comes down, inter alia, to determine the state of the real estate market at the time. The law doesn't limit the period of market analysis, provided that property prices will be adjusted on the valuation date or the date of analysis. Obligation to update prices follows from the Act on real estate management.

Different methods of determining of the influence of time on real estate prices can be found in the literature. For small data sets can be used interval method, while for large data sets statistical methods are applicable. Among the statistical methods most frequently cited are regression models. Estimation of linear regression model parameters is performed frequently by using least square method, which is not resistant to outlier cases. Even a single outlier can have a negative effect on the results of the estimation. Alternatively, the model parameters estimation can be made of robust estimation methods.

This work presents some robust estimation methods in relation to determining of the influence of time on real estate prices. Parameters estimation results for linear models were compared. Summarizes the results of the estimation of the least squares method and some robust estimation methods. variance analysis being also taken as a basis for conclusion. Analysis and calculations have been carried out on the sample database of properties.

Key words: *robust estimation, Huber m-estimator, Hampel m-estimator, outliers, update of prices*

Introduction

Determining market value of the property is preceded by an analysis of the real estate market. Before performing the market analysis and valuation of real estate appraiser should collect a database of properties similar to the property being valued. One of the elements of market analysis is to determine the state of the market at the time. Analysis of price volatility at the time is the basis for subsequent correction of property prices on the valuation date or the date of the analysis. Analysis of price volatility over a time is the basis for the adjustment of property prices on the valuation date or the date of analysis.

Different methods of determining of the influence of time on real estate prices can be found in the literature. For small data sets can be used interval method. In the case of real estate valuation as well as market analysis, statistical methods are also used. Among the statistical methods most frequently cited are regression models. Property attributes, such as area, location, depreciation, are often considered as parameters of real estate valuation models (CZAJA, PARZYCH, 2007; DĄBROWSKI, ADAMCZYK, 2010; PARZYCH, CZAJA, 2015; BIEDA et al., 2016). Time is usually the only parameter in price volatility models over time (BUDZYŃSKI, 2010; PARZYCH, CZAJA, 2015). Linear models are primarily estimated. Estimation of linear regression model parameters is performed frequently by using least square method, which is not resistant to outlier cases. An alternative is a robust estimation used in property valuation (ADAMCZYK, 2017). Robust estimation methods can also be used as an option in price volatility analysis over time.

* This work is financed from funds science realized at AGH University of Science and Technology, allocated for the year 2017, no 11.11.150.006.

Regression model of price volatility over time

The basis for trend estimation should be homogeneous or very similar properties. Estimation of the trend line requires the compilation of data in databases for real estate from the analyzed local market. The price trend over time can be expressed as a linear function due to the short period of transactions selected for analysis (usually 2 years) (PARZYCH, CZAJA, 2015). The linear trend function for a single transaction can be written as:

$$c = A + B \cdot t \tag{1}$$

where:

c - real property unit price,
t - transaction time that is assessed by months starting from the date of the first transaction,
A, B - regression coefficients.

Equations can be written for several properties. Thus the model takes the form:

$$c_i = A + B \cdot t_i + \delta_i \tag{2}$$

where:

c_i - transaction unit price of the i-th property,
t_i - transaction time of the i-th property,
A, B - regression coefficients,
δ_i - random deviation to the unit price of the i-th property.

The set of equations (2) could be solved using the least squares method. The estimators of regression coefficients can be expressed by the following formulas:

$$B = \frac{\sum_{i=1}^{n}(c_i - \hat{c})(t_i - \hat{t})}{\sum_{i=1}^{n}(t_i - \hat{t})^2} \tag{3}$$

$$A = \hat{c} - B \cdot \hat{t} \tag{4}$$

where:

$\hat{c}$ - average unit price calculated on the basis of the real estate database,
t - average transaction time calculated on the basis of the real estate database,
n - the number of properties from database.

The coefficient of determination R^2 is a measure of goodness of fit for the estimated regression equation:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}{\delta_i}^2}{\sum_{i=1}^{n}(c_i - \hat{c})^2} = 1 - \frac{\sum_{i=1}^{n}(c_i - A - B \cdot t_i)^2}{\sum_{i=1}^{n}(c_i - \hat{c})^2} \tag{5}$$

The price adjustment on the valuation date or the date of analysis is done on the basis of the formula:

$$c_{Ki} = c_i + B \cdot (t_W - t_i) \tag{6}$$

where:

c_{Ki} - unit price of i-th property adjusted on the valuation or the date of analysise,
t_W - time of valuation that is assessed by months starting from the date of the first transaction.

M-estimation

M-estimation idea is based on model deviations function minimizing. It is possible to find a modified smallest squares method among m-estimation class methods (WIŚNIEWSKI, 2009). M-estimators are defined by three functions: objective function, influence function and weighting function. The least squares method is a special case of m-estimation, where the objective function $\rho(v) = v^2$ is replaced by another $\rho(v)$. For modified smallest squares method a function of the objective $\rho(v)$ is represented by:

$$\rho(v) = w(v)v^2 \tag{7}$$

where:

$w(v)$ - weighting function.

Influence function is the first derivative of function of objectives due to v:

$$\psi(v) = \frac{\partial \rho(v)}{\partial v} \tag{8}$$

Weighting function could be determined on the basis of influence function as follows:

$$w(\nu) = \frac{\psi(\nu)}{\nu} = \frac{\partial\rho(\nu)}{\partial(\nu^2)} \tag{9}$$

Functions must meet the listed below criteria:

- objective function:
  - non-negative, $\rho(\nu) \geq 0$,
  - takes zero when its argument is zero, $\rho(0) = 0$,
  - symmetry, $\rho(\nu_i) = \rho(-\nu_i)$,
  - monotonicity in $|\nu_i|$, $\rho(\nu_i) \geq \rho(\nu_j)$ for $|\nu_i| > |\nu_j|$.
- weighting function
  - continuous and even (symmetry),
  - $w(\nu)$ =1 for $\nu = 0$,
  - $w(\nu)$ function values decrese when $|\nu|$ increases,
  - $\lim_{\nu\to\infty} w(\nu) = 0$.

The optimization problem of m-estimation is solved by modified least squares method. This method consists in carrying out subsequent estimations preceded by a modification of observation weights by weighting function (WIŚNIEWSKI, 2009).

Among m-estimators that are recognised in the professional literature, Huber and Hampel m-estimators (HUBER, 1964; HAMPEL, 1974; HUBER, 1981) were considered in this study. It is also possible to create new m-estimators motivated by different weighting functions (BANAŚ, LIGAS, 2014).

The following figures show the functions of m-estimators by Huber and Hampel.

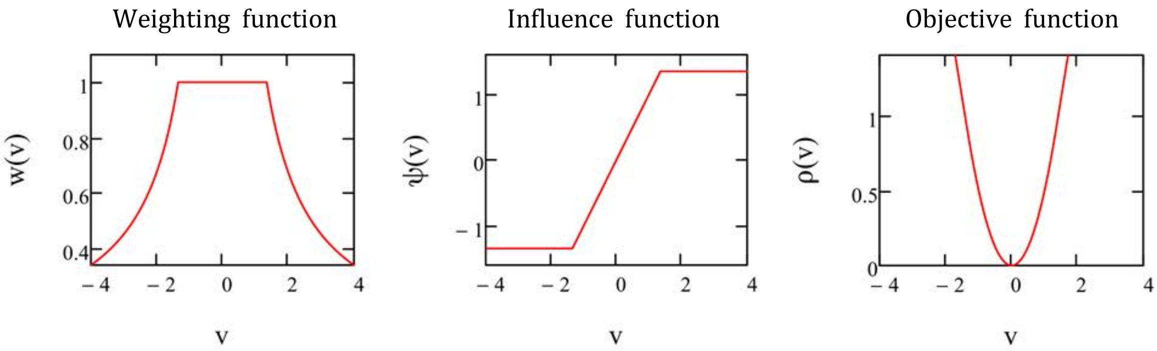

Fig. 1. Weighting, influence and objective function for Huber m-estimator.

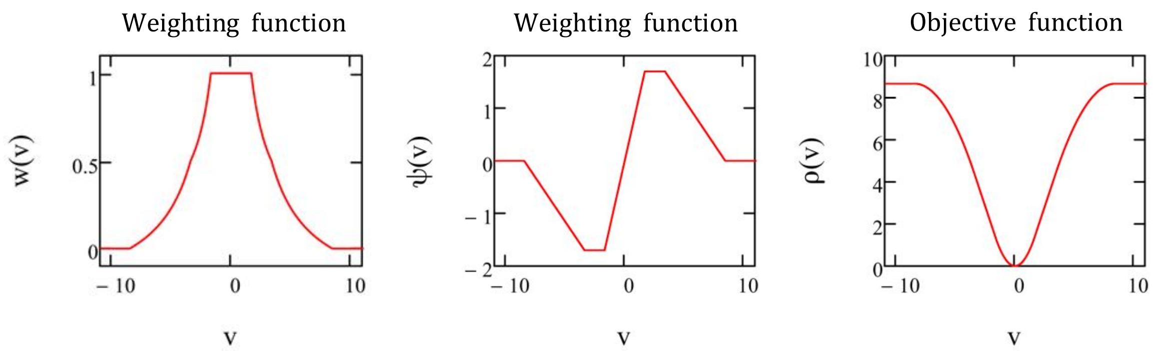

Fig. 2. Weighting, influence and objective function for Hampel m-estimator.

The modifed least square method of estimation consists in carrying out subsequent calculations preceded by a weight modification using the weighting function. The basis for modifying weights are residuals. The influence of observations that have been higher residuals values in the earlier iteration step is dampened.

Verification on numerical examples

A comparison of the price trend estimation was performed on three variants. The first variant is without outliers. The second and third variants include outliers at different levels. The data for testing is shown in the table 1.

Table 1. Data for comparison of estimation methods.

No	Time (in months)	Unit prices - variant 1	Unit prices - variant 2	Unit prices - variant 3
1	0	420	600	550
2	2	415	550	550
3	5	470	600	500
4	7	490	420	420
5	8	390	390	430
6	10	400	400	400
7	11	440	440	440
8	11	480	480	480
9	12	520	500	500
10	15	750	550	550
11	16	610	560	560
12	18	720	600	600
13	19	670	600	600
14	20	790	620	620
15	20	810	600	600
16	22	580	580	580
17	23	540	590	590
18	25	680	450	550
19	26	780	825	250
20	27	790	820	300
21	28	610	840	280
22	29	820	820	620
23	29	850	850	620

Source: based on LIGAS, 2010

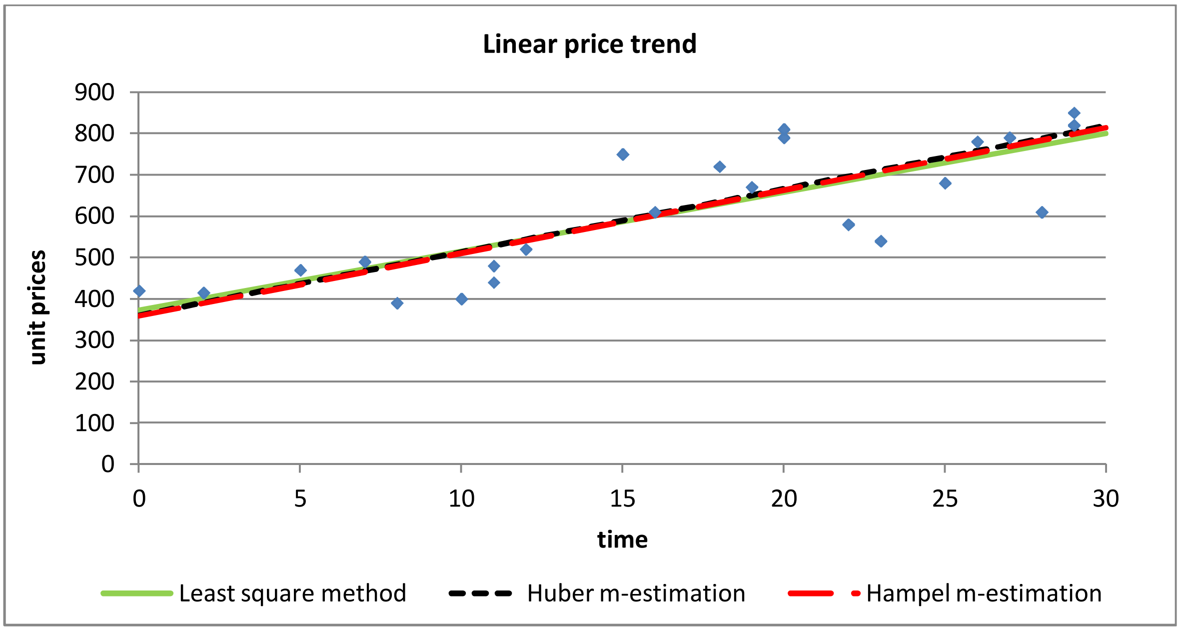

Fig. 3. Comparison of estimation methods - variant 1. *Source: own study.*

Table 2. Regression coefficients and their standard deviations - variant 1.

	Least square method		Huber m-estimation		Hampel m-estimation	
	Parameter value	Standard deviation	Parameter value	Standard deviation	Parameter value	Standard deviation
A	372,89	42,62	360,27	30,86	358,12	34,96
B	14,23	2,27	15,29	1,69	15,16	1,91

Source: own study

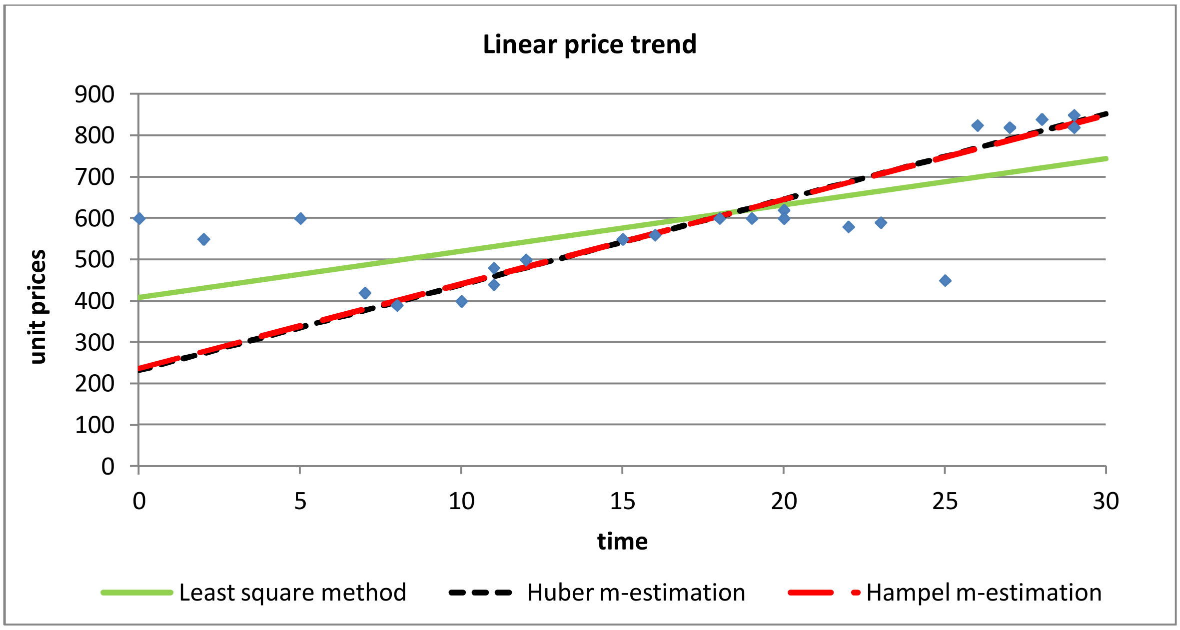

Fig. 4. Comparison of estimation methods - variant 2. *Source: own study.*

Table 3. Regression coefficients and their standard deviations - variant 2.

	Least square method		Huber m-estimation		Hampel m-estimation	
	Parameter value	Standard deviation	Parameter value	Standard deviation	Parameter value	Standard deviation
A	408,47	49,01	231,91	21,82	236,50	24,80
B	11,20	2,61	20,68	1,13	20,43	1,27

Source: own study

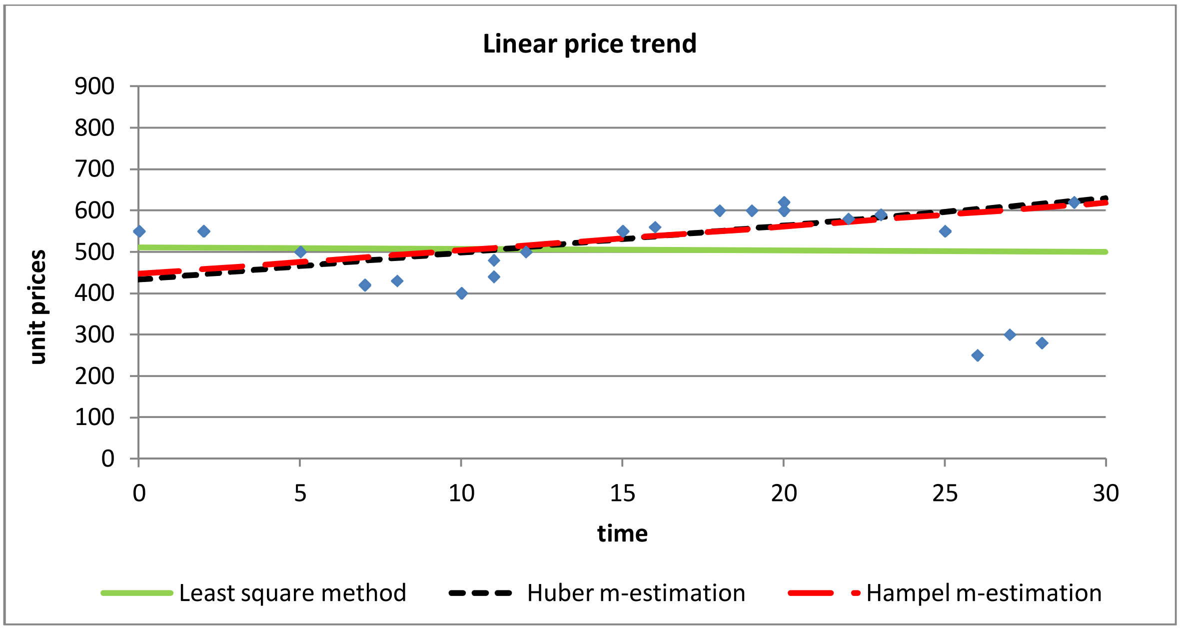

Fig. 5. Comparison of estimation methods - variant 3. *Source: own study.*

Table 4. Regression coefficients and their standard deviations - variant 3.

	Least square method		Huber m-estimation		Hampel m-estimation	
	Parameter value	Standard deviation	Parameter value	Standard deviation	Parameter value	Standard deviation
A	509,79	51,89	432,88	26,76	446,52	28,29
B	-0,35	2,76	6,56	1,48	5,74	1,57

Source: own study

In accordance to thesis stated in the beginning of the study, the use of Huber and Hampel's m-estimators gave more precise estimation results than least squared method. This is evidenced by lower standard deviation results for given parameters obtained when using m-estimators.

In a variant without observing outliers (variant 1), the results of the estimation are similar. However, the obtained standard deviations of the regression coefficients indicate that the results obtained with the m-estimation are more accurate. The visual assessment of the graphs for the other two variants shows a better fit of the trend line to the data using the m-estimation. In these cases, outlier observations completely changed the trend line estimated by the least squares method.

Summary

This article presents application of robust estimation methods in analysing of real estate price volatility over time. For the purpose of the study the least squares estimation result and two different robust estimation methods, Huber and Hampel methods, comparative analysis was conducted.

As a result of m-estimation, lower than for the least squares method values of standard deviation model parameters were obtained. Robust estimation methods allow to minimize the variance of price trend line parameters. Targeted price adjustment on the valuation date or analysis is more reliable. Consequently, this can lead to increased reliability of property valuation.

References

ADAMCZYK T., Application of the Huber and Hampel M-estimation in Real Estate Value Modeling, Geomatics and Environmental Engineering, Kraków, 2017. vol. 11 no. 1 s.15-23

BANAŚ M., LIGAS M., Empirical tests of performance of some M – estimators. Geodesy and Cartography, Vol. 63, No 2, Warszawa 2014.

BIEDA A., BIEDA A., ADAMCZYK T., PARZYCH P. The procedure for the valuation of real properties development with passive and energy-efficient houses, SGEM 2016 : 16-th international multidisciplinary scientific geoconference: informatics, geoinformatics and remote sensing: 30 June–6 July, 2016, Albena, Bulgaria: conference proceedings. Vol. 2, Geodesy and mine surveying, photogrammetry and remote sensing., s. 633–640.

BUDZYŃSKI T. Metodyka aktualizacji cen na przykładzie nieruchomości lokalowych, w: Studia i Materiały Towarzystwa Naukowego Nieruchomości, vol. 18, nr 1, 2010, ss. 107-118

CZAJA J., PARZYCH P., Szacowanie rynkowej wartości nieruchomości w aspekcie międzynarodowych standardów wyceny, Stowarzyszenie Naukowe im. St. Staszica, Kraków 2007.

DĄBROWSKI J., ADAMCZYK T., Application of GAM additive non-linear models to estimate real estate market value — Zastosowanie addytywnych modeli nieliniowych GAM do szacowania rynkowej wartości nieruchomości Geomatics and Environmental Engineering, Kraków, 2010 vol. 4 no. 2, s. 55–62.

HAMPEL F., The Influence Curve and Its Role in Robust Estimation. Journal of American Statistical Association, vol. 69, no. 346, 1974.

HUBER P. J., Robust Estimation of a Location Parameter. The Annals of Mathematical Statistics, vol. 35, no. 1, 1964.

HUBER P. J., Robust Statistics. New York: John Wiley & Sons, 1981.

LIGAS M., Metody statystyczne w wycenie nieruchomości — Statistical methods in real estate valuation, Studia i Materiały Towarzystwa Naukowego Nieruchomości = Journal of the Polish Real Estate Scientific Society, 2010 vol. 18 no. 1, s. 49–64.

PARZYCH P., CZAJA J., Szacowanie rynkowej wartości nieruchomości — Estimation of real estates market value, Wydawnictwa AGH, Kraków, 2015.

WIŚNIEWSKI Z., Rachunek wyrównawczy w geodezji. Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego. Olsztyn 2009.