

Tomasz Adamczyk\*

## **Application of the Huber and Hampel M-estimation in Real Estate Value Modeling\*\***

### **1. Introduction**

Methodology of property value estimation based on its comparison with other market real properties is included in a comparative appraisal approach. Knowledge and recognition of transaction prices, market conditions/ characteristics as well as terms of transaction for similar to the one evaluated real properties on the market, are set as fundamentals in the comparative approach. Differences in evaluated and benchmarked properties significant characteristics are taken into account, therefore price adjustments are being implemented for the compared properties based on such differences. In addition to the direct comparison of the real property methods of evaluation, mathematical statistics based methodology is also to be found in the professional literature. Frequently quoted are two and multidimensional variables linear regression methods, as well as non-linear models [1, 4, 6, 10].

Used in valuation and real estate market analysis multidimensional regression models are often based on the assumption of linearity relation between explanatory and dependent variables. Model parameters estimation is based mainly on the smallest squares method, that in its default form takes into account equally all of the observations. It is possible to modify the smallest squares method by applying weights to the observations, however the weighting could often come as a result of comparing the valuated and similar properties characteristics. In such cases, outliers, another words howler or congruent to the model observations, are not being eliminated.

Applicability of multidimensional property evaluation regression in parameter estimation was being considered in the study, the chosen methods of resistant estimation also known as the alignment of geodetic observations. Such methods enable one to get lower levels of standard deviation for model parameters being estimated. Chosen two M-estimation methods were analyzed: Huber and Hampel.

---

\* AGH University of Science and Technology, Faculty of Mining Surveying and Environmental Engineering, Department of Geomatics, Krakow, Poland

\*\* This work was financed from funds science realized at AGH University of Science and Technology, allocated for the year 2015

## 2. Multidimensional Regression Model in Real Property Valuation

All market characteristics and the real estate unit prices are being considered together as a random multidimensional variable in the multidimensional regression. Typically parameters of random multidimensional variable are: average values and standard deviation of market characteristics and prices in the marginal distribution [6, 10].

One of the problems when modelling the studied sizes in the multiple regression model is to determine the mathematical function form to represent analysed phenomenon:

$$c = f(a_i, X_i) + \delta \quad (1)$$

where:

- $c$  – real property unit price,
- $a_i$  – function parameters,
- $X_i$  – real property attributes,
- $\delta$  – random component to the real property unit price.

Assuming linear relation between real properties market characteristics and their unit prices, for many of the independent variables (real property attributes)  $X_1, X_2, \dots, X_k$  multiple regression model is defined by  $(k + 1)$  dimensional space hyper-surfaces, that is:

$$c = a_0 + a_1X_1 + a_2X_2 + \dots + a_kX_k + \delta \quad (2)$$

where:

- $c$  – dependent random variable (forecasted),
- $X_1, X_2, \dots, X_k$  – independent random variables (explanatory),
- $a_0, a_1, \dots, a_k$  – regression coefficients of  $c$  variable in respect to variables  $X_i$ .

If the number of properties representing market turnover database is set as  $n$  and the number of considered attributes would be  $u$  ( $n > u$ ) then the set of equations (2) could be solved using the smallest squares method [6, 10]. The linear set of equations is a stochastic model that could be represented as a matrix:

$$c = XA + \delta \quad (3)$$

where:

- $c$  – real properties unit prices ( $n \times 1$ ),
- $X$  – real properties unit price attributes matrix ( $n \times u$ ),
- $A$  – model parameters matrix ( $u \times 1$ ),
- $\delta$  – random components matrix ( $n \times 1$ ).

According to the smallest squares method the result of the set of equations (3), leads to the shown below dependence [5]:

$$\hat{A} = (X^T X)^{-1} X^T c \tag{4}$$

Statistic model’s estimated parameters accuracy analysis boils down to regression coefficients covariance matrix calculation:

$$\text{cov}[\hat{A}] = \hat{\sigma}_0^2 \cdot (X^T X)^{-1} \tag{5}$$

where  $\hat{\sigma}_0^2$  – residual variance estimator.

Residual variance is expressed by the below equation:

$$\hat{\sigma}_0^2 = \frac{(X\hat{A} - c)^T (X\hat{A} - c)}{n - k - 1} \tag{6}$$

### 3. M-estimation

Possibility of using m-estimators for structural parameter estimation of a model was considered in this study. M-estimators as a certain class of estimators are characterized by high comfort of use, flexibility and performance. 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 [11].

M-estimators are defined by three functions: purpose, impact, and weight. For the modified smallest squares method a function of the objective  $\rho(v)$  is represented by:

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

where  $w(v)$  – weight function.

Impact function is the first derivative of function of objectives due to  $v$ :

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

Weight function could be determined on the basis of impact function as follows:

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

Weighting function must meet the listed below criteria:

- $w(v)$  must be a continuous function and even (symmetry),
- $w(v) = 1$  for  $v = 0$ ,
- $w(v)$  function values decrease when  $|v|$  increases,
- $\lim_{v \rightarrow \infty} w(v) = 0$ .

Weighting functions  $w(v)$  usually include parameters that control outliers' impact on the estimation of results. Weighting function  $w(v)$  values are the weights that for the outliers are suppressed in the estimation process. Assuming amendment standardisation, it is possible to express the relations between weighting function  $w(\bar{v})$  and suppression function  $t(\bar{v})$  as:

$$w(\bar{v}) = t(\bar{v})p \quad (10)$$

Among M-estimators that are recognised in the professional literature, Huber and Hampel M-estimators [7–9] were considered in this study. It is also possible to create new M-estimators motivated by different weighting functions [1]. Huber's method of weighting function is represented by:

$$w(v) = \begin{cases} 1 & \text{for } |v| \leq k \\ \frac{k}{|v|} & \text{for } |v| > k \end{cases} \quad (11)$$

Constant  $k$  depends of the number of outliers in the set. Its most commonly set values range between 1.5 and 2.0. It could be assumed, that if the outliers represent approximately 4% of the set than the value is set as 1.5, and in the case of the outliers being less than 1% than it is set as 2.0.

Huber M-estimator suppress function is represented by:

$$t(\bar{v}) = \begin{cases} 1 & \text{for } |\bar{v}| \leq k \\ \frac{k}{|\bar{v}|} & \text{for } |\bar{v}| > k \end{cases} \quad (12)$$

Hampel method weighting function is represented by:

$$w(v) = \begin{cases} 1 & \text{for } |v| \leq a \\ \frac{a}{|v|} & \text{for } a < |v| \leq b \\ \frac{a(c-|v|)}{(c-b)|v|} & \text{for } b < |v| \leq c \\ 0 & \text{for } |v| > c \end{cases} \quad (13)$$

Parameters  $a, b, c$  set intervals boundaries, where correction value could be found, wherein Hampel presents weighting function with  $a = 2, b = 4, c = 8$  parameters for amendments scaled by the median absolute deviation.

Hampel method suppression function is represented by:

$$t(\bar{v}) = \begin{cases} 1 & \text{for } |\bar{v}| \leq a \\ \frac{a}{|\bar{v}|} & \text{for } a < |\bar{v}| \leq b \\ \frac{a(c-|\bar{v}|)}{(c-b)|\bar{v}|} & \text{for } b < |\bar{v}| \leq c \\ 0 & \text{for } |\bar{v}| > c \end{cases} \quad (14)$$

#### 4. Use of Resistance Estimation Methods in Real Estate Valuation

A robust estimation algorithm could be implemented using a modified smallest squares method. Estimation is done iteratively when using either Huber or Hampel method. Taking into account the damping function, in the next iteration step "re-weighting" is conducted. In the next iteration step the weighting matrix is calculated based on the below formula:

$$\hat{P} = T(\bar{V})P \quad (15)$$

where:

$\hat{P}$  – equivalent weight matrix,  
 $T(\bar{V})$  – diagonal suppression matrix,  
 $P$  – weighting matrix.

In the case of a valuation model to estimate model parameters using the gross error-tolerant method one should take into account real property unit prices determined by transactions. The model should include attributes of the real property as well as transaction time that is assessed by months starting from the date of the market analysis or valuation:

$$c = a_0 + a_T \cdot t + a_i \cdot X_i + \delta \quad (16)$$

where:

$c$  – real property unit price,  
 $a_0, a_T, a_i$  – function parameters,  
 $X_i$  – real property attributes,  
 $t$  – transaction time,  
 $\delta$  – random component of real property unit price.

It is possible to take into account the cost of creating the components as attributes. It has been shown that the replacement cost is a strong determinant of the market value [2].

To estimate the parameters of the above model credibility, weights should be formulated for the individual equations (transaction prices). Such weights should result from the value accuracy of the chosen facultative attribute values selection for individual properties, because their unit prices should be treated with equal reliability. Thus, the formula of credibility importance may be expressed by using a number of attributes with approximate values ( $n_{AP}$ ):

$$p = \frac{1}{1 + n_{AP}} \quad (17)$$

In the matrix equation the modified smallest squares method could be expressed as:

$$X^T P X \cdot A = X^T P C \quad (18)$$

where:

- $X$  – matrix of a property examined attributes,
- $P$  – credibility weights diagonal matrix,
- $A$  – model parameters matrix,
- $C$  – examined property unit price matrix.

Equation (18) is solved by:

$$\hat{A} = (X^T P X)^{-1} X^T P C \quad (19)$$

The determined parameter vector (19) is used to determine the amendments:

$$\hat{V} = X \hat{A} - C \quad (20)$$

The following step is to calculate standard amendments  $\hat{V}$ , to which in the next iteration step modified weighting matrix  $\hat{P}$  is calculated taking into account the suppression function of Huber (12) and Hampel (14).

To standardize the amendments, amendment variance-covariance matrix should be calculated:

$$\text{cov}(\hat{V}) = \sigma_0^2 (P^{-1} - X(X^T P X)^{-1} X^T) \quad (21)$$

The steps described before are repeated for the next iteration step, this however is done with modified after each iteration step weighting matrix. In the case of modified weighting matrix outliers weights are being suppressed. The solution results

are considered final when the weights are no longer being modified, another words outliers' impact is limited.

The estimated model could be used for market analysis ( $\beta$  coefficients of regression model) and for similar properties valuation for the given time.

## 5. M-estimation Method Case Study

For practical use verification of selected m-estimation methods in the real estate property value modelling, sample-based real property analysis were performed. As for the case study sample base, we have selected 61 local real properties, all located in the central part of cadastral unit Podgórze of Krakow. Our database real properties were all subject to market trade between January 2015 and February 2016.

All the real properties were described by market characteristics as: *location, finishes standard, space/ area, technical condition, floor/ tier*. *Space* is described as number of square meters of the building area and for the remaining market characteristics three-point scale was agreed. Additionally time was included as one of the attributes and it is described as a number of months counting from the oldest transaction. Table 1 shows a part of 61 local real properties information database. As the last item a real property was added and it will be measured based on model estimated parameters.

**Table 1.** Model verification real property database

No.	Time	Location	Surroundings	Standard	Space/area	Technical condition	Floor/tier	Price for 1 m <sup>2</sup> [PLN]
1	0	3	2	2	35.56	3	3	7142.86
2	0	1	2	2	40.19	3	1	5165.96
3	0	2	2	2	39.57	2	3	6772.81
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
59	12	3	2	2	36.34	3	3	5938.66
60	12	2	2	3	39.90	3	2	7142.86
61	13	2	1	2	35.75	2	2	6013.99
NW	15	1	2	3	26.81	2	3	–

Based on our real property database, an estimation of the multiple regression model parameters in three variants was conducted:

- the standard smallest squares method,
- the reweighted smallest squares method with the use of the Huber function,
- the reweighted smallest squares method with the use of the Hampel function.

The testes model is based on the functional relation described by the equation (16).

**Table 2.** Selection of parameters values and their standard deviations using different estimation methods

		$a_0$	$a_T$	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$
Smallest squares method	parameter value	4717.74	-18.12	263.31	226.29	585.86	-36.20	194.04	112.81
	standard deviation	463.28	20.09	100.64	115.99	117.55	10.62	106.15	78.72
Huber method	parameter value	4772.62	-22.62	250.62	220.19	663.54	-35.48	168.97	76.89
	standard deviation	400.13	17.79	89.72	102.82	108.42	9.16	95.21	69.29
Hampel method	parameter value	4667.77	-17.01	253.96	231.65	643.52	-34.71	172.35	91.39
	standard deviation	421.47	18.62	93.55	107.75	112.90	9.68	100.09	73.15

In accordance to the thesis stated in the beginning of the study, the use of Huber and Hampel M-estimators gave more precise estimation results than the smallest squared method. This is evidenced by lower standard deviation results for given parameters obtained when using M-estimators (Tab. 2).

Based on the computed model parameters for different estimation methods, the value of the appraised property for which the market characteristics have been recorded in the last row of Table 1 was set. Standard deviations of the estimated property value with the use of the variance-covariance matrix for the model parameters were set. The last iteration matrices were used for the Huber and Hampel methods. The results of value the sample property estimation are shown and standard deviations of the estimated values are presented in Table 3.

**Table 3.** A statement of the calculated property value along with the standard deviations

	Unit value [PLN/m <sup>2</sup> ]	Standard deviation of unit value [PLN/m <sup>2</sup> ]	The confidence interval width for a unit value of $p = 0.95$ [PLN/m <sup>2</sup> ]	Real property value [PLN]	Standard deviation of real property value [PLN]	The confidence interval width for a real property value of $p = 0.95$ [PLN]
Smallest squares method	6675.47	260.13	±507.25	178 969	6974	±13 599
Huber method	6732.45	220.99	±430.93	180 497	5925	±11 553
Hampel method	6748.89	232.07	±452.54	180 938	6222	±12 133



## 6. Conclusions

This article presents the application of robust estimation methods in the property value modelling. For the purpose of the study the smallest squares estimation result and two different robust estimation methods, Huber and Hampel methods, comparative analysis was conducted.

M-estimators use in estimation results in gross-errors observation and observations that are not aligned with the model impact minimizing. As a result of robust estimation, lower than for the smallest squares method the values of the standard deviation model parameters were obtained. The model parameters variance minimizing, that is secured by the robust estimation method, results in estimated reliability increase based on the property value model.

## References

- [1] Adamczyk T., Dąbrowski J.: *Application of GAM additive non-linear models to estimate real estate market value*. Geomatics and Environmental Engineering, vol. 4, no. 2, 2010, pp. 55–62.
- [2] Adamczyk T., Ruchel J.: *Analysis of comparative approach and cost approach in the aspect of real property market value*. Geomatics and Environmental Engineering, vol. 9, no. 4, 2015, pp. 15–24.
- [3] Banaś M., Ligas M.: *Empirical tests of performance of some M-estimators*. Geodesy and Cartography, vol. 63, no. 2, 2014, pp. 127–146.
- [4] Bieda A.: *Wycena nieruchomości w aspekcie standardów międzynarodowych*. [in:] *Materiały Krakowskiej Konferencji Młodych Uczonych 2008: Kraków, 25–27 września 2008*, AGH GN Pro Futuro, Kraków 2008, pp. 387–393.
- [5] Czaja J.: *Modele statystyczne w informacji o terenie*. Wydawnictwa AGH Kraków 1997.
- [6] Czaja J., Parzych P.: *Szacowanie rynkowej wartości nieruchomości w aspekcie międzynarodowych standardów wyceny*. Stowarzyszenie Naukowe im. St. Staszica, Kraków 2007.
- [7] Hampel F.: *The Influence Curve and Its Role in Robust Estimation*. Journal of American Statistical Association, vol. 69, no. 346, 1974, pp. 383–393.
- [8] Huber P.J.: *Robust Estimation of a Location Parameter*. The Annals of Mathematical Statistics, vol. 35, no. 1, 1964, pp. 73–101.
- [9] Huber P.J.: *Robust Statistics*. John Wiley & Sons, New York 1981.
- [10] Parzych P., Czaja J.: *Szacowanie rynkowej wartości nieruchomości*. Wydawnictwa AGH, Kraków 2015.
- [11] Wiśniewski Z.: *Rachunek wyrównawczy w geodezji*. Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego, Olsztyn 2009.