ACTA Acta Sci. Pol., Administratio Locorum 23(4) 2024, 349–360.
https://czasopisma.uwm.edu.pl/index.php/aspal plISSN 1644-0749 eISSN 2450-0771 DOI: 10.31648/aspal.10726

ORIGINAL PAPER
Received: 25.10.2024 Accepted: 26.11.2024

APPLICATION OF ROBUST ESTIMATION METHODS IN REAL ESTATE VALUATION

Tomasz Adamczyk✉

ORCID: 0000-0002-4302-6169
AGH University of Krakow
30 Mickiewicza Av., 30-059 Krakow, **Poland**

ABSTRACT

Motives: Conducting real estate appraisals in well-developed markets presents a multitude of data analysis challenges. Some property price data may contain outliers that can significantly affect the valuation process and, as a result, the estimated value. In the case of real estate valuation regression models, estimation is most often based on the least squares method, where outliers are taken into account just like the rest of the data. Eliminating or minimizing the influence of outliers can lead to more reliable estimation results. Such problems can be solved by implementing robust regression methods.
Aim: The main goal is to determine whether robust estimation methods using M-estimators with multiple regression models can provide more accurate estimates of property values.
Results: Research has shown that multiple regression models using robust regression methods can be applied to estimate property values. The use of different types of M-estimators allows for the objective elimination of outliers through algorithms that operate on the entire data set. The calculations are carried out iteratively, and at each iteration step the residuals are verified and the observations are re-weighted. The following M-estimators were considered: Huber, Hampel, Tukey, Faire, Cauchy and Welsch. The reference point was the estimation results from the ordinary least squares method (OLS). All analysed M-estimators led to an increase in the coefficient of determination value and a decrease in standard estimation errors. Each algorithm detected the outliers. The valuation results for the selected properties were also more reliable. The results obtained depend on the characteristics of the data, and the choice of the best estimator may vary across different property markets. The selection of the best estimator may even vary within the same local property market, where the valuer makes subjective assessments of the location or other attributes.

Keywords: robust regression, M-estimator, property valuation, multiple regression, robust estimators

INTRODUCTION

Many current engineering problems can be formulated as multiple linear regression (MLR) problems for data forecasting. Mathematical models based on multiple linear regression are also used to enhance the reliability of price forecasting and analysis in property markets. Price forecasting is defined as the process of using past and present data, along with statistical methods and other analytical

✉tomasz.adamczyk@agh.edu.pl

techniques, to predict future prices. Multiple linear regression is a well-established statistical technique, and MLR models are relatively simple, providing easily interpretable mathematical formulas.

If a data set is characterized by the presence of extreme values, known as outliers, classical estimators may fail to perform efficiently in practical scenarios. An outlier is a data point that differs significantly from other observations. In several cases, such as those involving short-term statistics, the impact of outliers can be considerable and significantly affect the assessed value of a property.

In practical applications of statistical methods, it is not uncommon to encounter situations where the requisite assumptions are not fully satisfied. For instance, an assumption may be made that errors follow a normal distribution, but in practice, this may not hold true. In such circumstances, the efficacy of classical methods can be limited (Ruckstuhl, 2016). Robust estimation strategies, which are not sensitive to outliers, can be employed in these cases.

Several alternative estimation techniques that are less sensitive to outliers have been proposed in the literature. The purpose of this paper is to examine one of the most commonly used robust regression techniques, namely M-estimation, and to conduct an empirical comparison of several well-known M-estimators in the context of property value estimation.

LITERATURE REVIEW

Regression analysis is a significant statistical tool employed across numerous scientific disciplines, and the application of robust M-estimators has been successfully implemented in a wide range of fields, including electrical, electronic, and telecommunications engineering; image processing; econometrics and finance; civil engineering; meteorology and geodesy; chemical engineering; astronomy; mechanical engineering; petrochemical engineering; nuclear engineering; industrial pharmaceutics; and various medical, biomedical, and biotechnological applications (de Menezes et al., 2021).

In electrical engineering, regression analysis is used for power system state estimation (PSSE). One of the earliest estimation methods employed to "clean up" potentially contaminated raw data was Weighted Least Squares (WLS). The non-quadratic estimation criteria for PSSE were first proposed by Merrill H.M and Schweppe F.C in 1971 (Pires et al., 1999).

For image processing, a mask coefficient based on the M-estimator reduces the influence of noisy pixels and pixels in occluded regions. For this purpose, Huber's and Tukey's M-estimators are used (Arya et al., 2007).

In economic surveys, the flexibility of weighted M-estimation makes it useful for a wide variety of data models (Mulry et al., 2018). M-estimation-based models are also utilized for evaluating economic performance, using data from surveys of small and medium-sized businesses. Functions such as the following can be used for this purpose: Andrews', Tukey's (bisquare), Cauchy's, Fair's, Hampel's, Huber's, logistic, Talworth's, and Welsch's weighting functions (Dehnel & Gołata, 2016).

M-estimation has been continuously evolving in civil engineering, metrology, and geodesy. Wiśniewski proposed some generalizations of M-estimation, known as Msplit estimation (Wiśniewski, 2009) and Msplit(q) estimation (Wiśniewski, 2010). A new approach, inspired by the theory of squared Msplit(q) estimation, was proposed by Nowel in 2019 for the identification of stable potential reference points (Nowel, 2019).

For chemical processes, data reconciliation is an important step in the real-time online optimization of a plant. In addition to Weighted Least Squares (WLS), several other objective functions can be used for data reconciliation. These functions, along with different weighting functions (e.g., for Hampel's redescending M-estimator), are implemented to adapt data for the process's distributed control systems (DCS) (Özyurt & Pike, 2004).

M-estimation methods are useful for a variety of common epidemiological analyses (e.g., estimating the marginal risk difference of anemia on preterm birth), including cases when data are not independently

✉tomasz.adamczyk@agh.edu.pl

and identically distributed (i.e., IID) in data fusion analyses (Ross et al., 2024).

Outliers in property datasets can be identified using surveying methods. Baarda's method and residual analysis can be used to eliminate outliers and improve the accuracy of functional model parameters. A comparative analysis of the two methods based on sample data leads to the conclusion that they are equally effective for property market data (Śpiewak & Barańska, 2020). The elimination of outliers (outlying transactions) in the property market can be achieved using passive robust estimation methods. It provides a consistent dataset free from observations that could be affected by significant human error (Śpiewak, 2018). Active methods of robust estimation have also been applied to real estate market data. In particular, Huber and Hampel functions were considered for real estate market data. Robust estimation methods were analysed in regression models of property price changes over time and in property valuation models (Adamczyk, 2017).

Due to the significant impact of outliers on least squares estimates, the use of robust regression methods is recommended in such cases. The studies highlight the importance of applying robust methods for this purpose and suggest that such approaches may be broadly beneficial for property datasets (Janssen et al., 2001).

The examples presented demonstrate that robust estimation methods based on M-estimators and other statistical methods are applicable for analyzing various types of data.

Multiple regression models are a commonly used analytical method in many practical cases. When modeling the regression, it is first necessary to understand the characteristics of the data to be used. Understanding these characteristics makes it easier to determine the most appropriate method. For data with outliers, a robust regression model is required, for example M-estimation, MM-estimation, or S-estimation.

In the paper "A Robust Method for Multiple Linear Regression" Andrews introduced the theory of robust estimation in multiple linear regression models. The method, based on M-estimators, requires a reliable initial fit, which is then refined to yield a procedure that is relatively efficient for nearly Gaussian data (Andrews, 1974).

A comprehensive view of robust estimation in multiple regression models was presented by Padrul Jana, Dedi Rosadi, and Epha Diana Supandi. The authors compared and presented results using two data sets: one with outliers and one without, applying both the OLS (Ordinary Least Squares) method and robust regression models with M-estimation, MM-estimation, and S-estimation. Their research indicates that understanding the characteristics of the data is crucial for effective analysis, as described in the conclusions (Jana et al., 2023).

Multiple linear regression is an effective method for accurately predicting house prices from a large dataset with a significant number of both categorical and numerical predictors (Abdulhafedh, 2022). The use of multiple regression models for estimating the market value of commercial real estate has been presented in other publications by Banaś, Czaja, and Dąbrowski (Banaś et al., 2022).

Statistical tools are also used in mass appraisal. Valuation models often rely on multiple regression analysis, but literature also includes models using spatial relationship. Property value estimation methods based on geographically weighted regression, spatial autoregressive models, and regression-kriging have also been verified as useful for mass valuation (Walacik et al., 2013). However, beyond the potential application of mathematical models, it is also essential to consider the quality of property data, as well as the legal, economic, and social aspects of conducting mass valuation for property value taxation (Grover & Walacik, 2019).

The promising results of studies predicting property sale prices based on MLS (Multiple Listing Service) data for Ottawa, Canada, in comparison with machine learning methods, have led to the conclusion that multiple regression methods warrant further development and researching on other data sets. The MLS is one of several databases created by collaborating real estate brokers to share information

✉tomasz.adamczyk@agh.edu.pl

on properties available for sale. The results achieved with Random Forest regression were as favorable as those obtained with multiple regression models (Lemeš & Akagic, 2022).

MATERIALS AND METHODS

In statistics, linear regression is used to model the relationship between a scalar response (or dependent variable) and one or more explanatory (or independent) variables (Susanti et al., 2014). Multiple linear regression is a statistical technique used to investigate the relationship between two or more variables: the dependent variable (also known as the response variable, outcome, or target) and the independent variables (the predictors). Multiple linear regression is also referred to as multiple regression. The multiple regression model is widely used in real estate market analysis and property value modeling.

In 1964, Huber proposed a new type of robust estimator as a generalization of maximum likelihood estimation (MLE), termed M-estimators. From a practical standpoint, an M-estimator can be analysed as a weighted mean, where the weights are designed to minimize the influence of outliers on the estimator. The fundamental concept of M-estimation is based on minimizing a model deviation function (Huber, 1964).

M-estimation is not sufficiently robust with respect to leverage points. Not all leverage points are influential unless they have large residuals. Even though it lacks robustness against leverage points, it is still widely used for analysing data where it is assumed that contamination primarily occurs in the response direction (Chen, 2002). A leverage point is defined as an observation that exhibits an anomalous predictor value, significantly diverging from the bulk of observations. Therefore, it is recommended to employ this approach only in situations where leverage points are absent. This limitation has implications for the range of potential applications. Robust regression methods can have a significant impact on the accuracy of estimates, but they should not automatically be used instead of classical methods (Dehnel, 2016).

The characterization of M-estimators is based on three fundamental functions: the objective (loss) function, the influence function, and the weighting function (Banaś & Ligas, 2014). Compared to the least squares method, the loss function is less sensitive to extreme residual values and increases more slowly than a square (Banaś & Ligas, 2014; Dehnel, 2016).

In its general form, the loss function $\rho(v)$ may be represented by the following equation:

$$\rho(v) = w(v)v^2 \tag{1}$$

where:

$w(v)$– weighting function.

The influence function is defined as the first derivative of the objective function:

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

Influence functions are a valuable analytical tool for assessing the impact of each variable on an estimator's value.

Weighting function is represented by $w(v)$:

$$w(v) = \frac{\psi(v)}{v} = \frac{\partial\rho(v)}{\partial(v^2)} \tag{3}$$

The shapes of the weighting functions are markedly different. The choice of function is contingent upon the desired weighting of outliers, among other factors. The weighting functions [$w(v)$] are controlled by parameters that regulate the effect of outliers on the estimation of results.

Table 1 shows the influence function and weighting function for selected M-estimators used in this work.

M-estimators can be divided into three categories: monotone (e.g., Huber, Fair), soft-redescending (e.g., Cauchy), and hard-redescending (e.g., Andrews, Hampel, Smith).

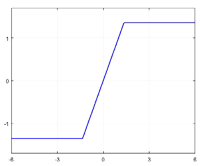
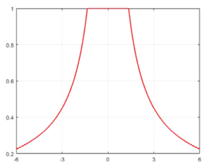
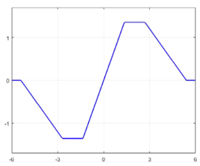
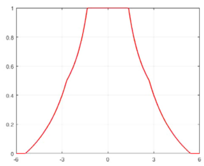
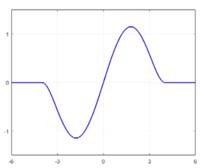
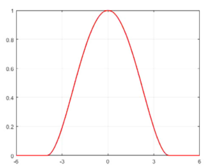
Using M-estimators in data regression problems is often associated with the trade-off between efficiency and robustness. Efficiency refers to the quality of the estimator's fit, while robustness refers to the estimator's performance (Albuquerque & Biegler, 1996). An additional measure of robustness is provided by the concept of Breakdown Points (BP). The Breakdown Point is the proportion of incorrect

✉tomasz.adamczyk@agh.edu.pl

observations in a dataset that a robust regression technique can tolerate. The value of BP cannot exceed 0.5 (50%) (Hampel, 1968; de Menezes, 2021).

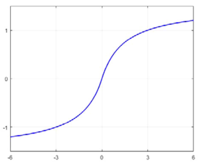
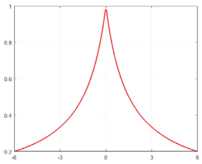
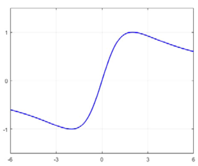
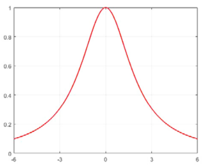
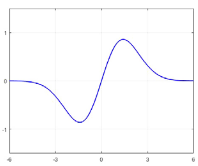
Huber devised the M-estimation method, which has become the most prevalent general robust regression approach (Huber, 1964). This work was further developed by other researchers who introduced various types of estimators, such as Yohai (MM-estimators), Rousseeuw (S-estimators, GM-estimators), Siegel's Repeated Median Estimators, Least Trimmed Squares (LTS) estimators, Minimum Volume Ellipsoid (MVE estimators), Baraud & Chen

Table 1. Characteristic functions of selected M-estimators

Huber M-estimator	
Influence function	Weighting function
1	2
$\Psi(v) = \begin{cases} v & if \quad \lvert v\rvert \le k \\ k \cdot \frac{v}{\lvert v\rvert} & if \quad \lvert v\rvert > k \end{cases}$	$w(v) = \begin{cases} 1 & if \quad \lvert v\rvert \le k \\ \frac{k}{\lvert v\rvert} & if \quad \lvert v\rvert > k \end{cases}$
	
Hampel M-estimator	
Influence function	Weighting function
$\Psi(v) = \begin{cases} v & if \quad \lvert v\rvert \le a \\ a \cdot sign(v) & if \quad a < \lvert v\rvert \le b \\ \frac{a \cdot (c \cdot sign(v) - v)}{c - b} & if \quad b < \lvert v\rvert \le c \\ 0 & if \quad \lvert v\rvert > c \end{cases}$	$w(v) = \begin{cases} 1 & if \quad \lvert v\rvert \le a \\ \frac{a}{\lvert v\rvert} & if \quad a < \lvert v\rvert \le b \\ \frac{a(c - \lvert v\rvert)}{(c - b)\lvert v\rvert} & if \quad b < \lvert v\rvert \le c \\ 0 & if \quad \lvert v\rvert > c \end{cases}$
	
Tukey M-estimator	
Influence function	Weighting function
$\Psi(v) = \begin{cases} v \cdot \left(1 - \left(\frac{v}{c}\right)^2\right)^2 & if \quad \lvert v\rvert \le c \\ 0 & if \quad \lvert v\rvert > c \end{cases}$	$w(v) = \begin{cases} \left(1 - \left(\frac{v}{c}\right)^2\right)^2 & if \quad \lvert v\rvert \le c \\ 0 & if \quad \lvert v\rvert > c \end{cases}$
	

✉tomasz.adamczyk@agh.edu.pl

cont. **Table 1**

Influence function	Weighting function
Fair M-estimator	
$\Psi(v) = \frac{v}{1 + \frac{\lvert v \rvert}{c}}$	$w(v) = \frac{1}{1 + \frac{\lvert v \rvert}{c}}$
	
Cauchy M-estimator	
$\Psi(v) = \frac{v}{1 + (\frac{v}{c})^2}$	$w(v) = \frac{1}{1 + (\frac{v}{c})^2}$
	
Welsh M-estimator	
$\Psi(v) = \frac{v}{1 + (\frac{v}{c})^2}$	$w(v) = \frac{1}{1 + (\frac{v}{c})^2}$
	

Source: own elaboration based on de Menezes et al. (2021).

(robust estimators for exponential families of distributions), Hampel (three-part M-estimators), and Tukey (biweight or bisquare M-estimators), as well as Andrews, Cauchy, and Fair (Raza et al., 2024).

RESULTS AND DISCUSSION

Verification of the selected M-estimation methods was carried out on the basis of properties from Tarnów. The database included 103 properties, which were selected from the central part of the city. The properties were sold over a 12-month period. The properties were described by typical characteristics: location (3-degree scale), surroundings (3-degree scale), transport accessibility (3-degree scale), standard (3-degree scale), floor location (3-degree scale), belonging premises (3-degree scale). The location of the sold properties is shown on the signature map (Fig. 1). The unit price level of the properties is shown in terms of colour intensity.

✉tomasz.adamczyk@agh.edu.pl

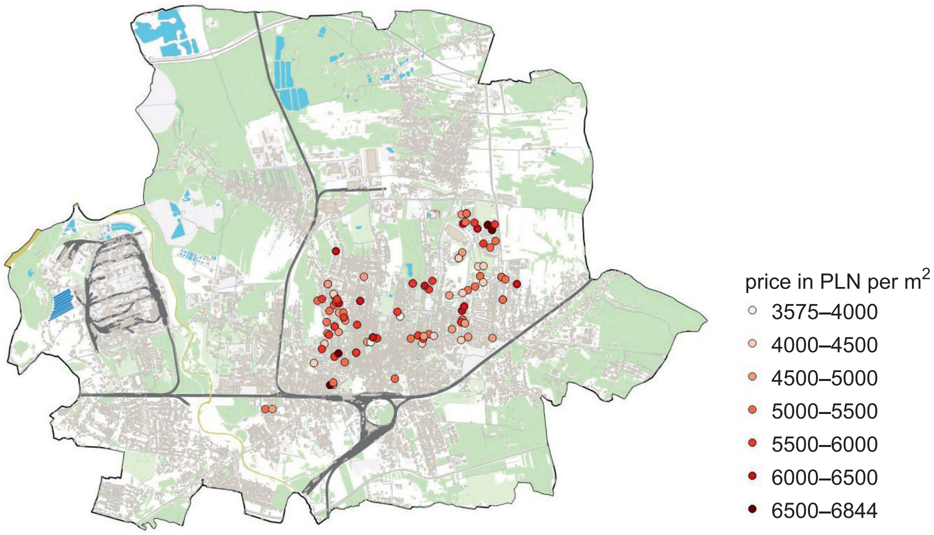


Fig. 1. Transactions in the form of signatures on the map of Tarnów
Source: own elaboration.

The calculations were made taking into account the following models and functions:
– ordinary least squares (OLS) method,
– iteratively reweighted least squares (IRLS) method with Huber weighting function,
– IRLS method with Hampel weighting function,
– IRLS method with Tukey weighting function,
– IRLS method with Fair weighting function,
– IRLS method with Cauchy weighting function,
– IRLS with Welsch weighting function.

Parameters of weighting function are usually tuned to 95% efficiency in respect to the normal distribution. In this work, the parameters were assumed according to (de Menezes, 2021).

The MAD (absolute median deviation), which is commonly used as a robust scale estimator, was used to scale the residuals. MAD uses the median twice, the first time to obtain the median value in a set of observations, and then to determine the median of a set composed of the absolute differences between the median value and the individual observation values.

The model tested is based on a linear relationship between the variables, the estimated parameters for the different estimation variants are shown in Table 2:

When M-estimation was applied, more reliable model parameter values were obtained in each case compared to the ordinary least squares estimation method. Additionally, one of the model's parameters was found to be insignificant when using the least squares method. In contrast, with M-estimation, each model parameter was statistically significant. This was due to the suppression of the influence of outliers. It is worth noting that no observations were excluded during the M-estimation process. Even methods that assign a weight of zero to large residuals, did not result in such changes. This outcome was attributed to the relatively good selection of variables for the model and the consistency of the data.

Based on statistical tests for the coefficient of determination, it can be concluded that the variance of the part of the dependent variable explained by the model is significantly greater than the unexplained part (Tab. 4). The coefficient of determination for models estimated using M-estimation increased by 0.03 to 0.07, depending on the applied estimation method.

The property values presented in Table 5 indicate an increase in reliability for the results obtained from models estimated using M-estimation compared to the model based on the least squares method.

✉tomasz.adamczyk@agh.edu.pl

Table 2. Estimated model parameters and their standard deviations

	Model parameter							
	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7
	Standard deviation							
	σa_0	σa_1	σa_2	σa_3	σa_4	σa_5	σa_6	σa_7
Ordinary Least Square	3,091.71	593.69	332.20	245.00	169.28	359.78	95.01	166.08
	164.88	75.84	81.47	80.22	61.91	54.78	55.96	53.72
Huber M-estimator	2,998.64	622.07	340.02	291.55	161.97	351.10	116.17	172.46
	140.76	66.81	69.77	69.73	53.28	46.81	48.24	46.26
Hampel M-estimator	2,987.33	629.16	343.73	293.01	160.27	349.11	119.59	171.78
	139.98	66.56	69.36	69.37	52.92	46.59	48.00	45.97
Tukey M-estimator	2,981.69	631.75	345.90	295.65	159.86	353.36	112.22	168.32
	139.67	66.28	69.31	68.91	52.65	46.48	47.70	46.00
Fair M-estimator	3,029.89	614.73	336.04	283.76	163.57	355.68	100.24	169.43
	132.61	64.17	66.31	64.59	50.47	44.34	45.74	44.69
Cauchy M-estimator	3,005.51	621.91	340.39	289.18	162.86	353.93	107.76	170.56
	137.77	65.68	68.42	67.68	52.14	45.73	47.18	45.65
Welsch M-estimator	3,041.67	609.15	337.58	267.52	166.31	355.93	104.93	168.17
	151.53	70.89	75.02	74.26	57.04	50.36	51.63	49.67

Source: own elaboration.

Table 3. Verification of statistical hypotheses on the significance of model parameters

	Model parameter							
	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7
Test function OLS	18.75	7.83	4.08	3.05	2.73	6.57	**1.70**	3.09
Test function Huber	21.30	9.31	4.87	4.18	3.04	7.50	2.41	3.73
Test function Hampel	21.34	9.45	4.96	4.22	3.03	7.49	2.49	3.74
Test function Tukey	21.35	9.53	4.99	4.29	3.04	7.60	2.35	3.66
Test function Fair	22.85	9.58	5.07	4.39	3.24	8.02	2.19	3.79
Test function Cauchy	21.82	9.47	4.98	4.27	3.12	7.74	2.28	3.74
Test function Welsch	20.07	8.59	4.50	3.60	2.92	7.07	2.03	3.39
Critical area $\alpha = 0.05$	$(-\infty ; -1.96) \cup (1.96 ; \infty)$							

Source: own elaboration.

✉tomasz.adamczyk@agh.edu.pl

Table 4. Determination coefficients of models with statistical tests

	R^2	R^2-R_0^2	Test function F	Critical area α = 0.05
Ordinary Least Square	0.75	–	40.71	(3.27; ∞)
Huber M-estimator	0.81	0.06	57.86	
Hampel M-estimator	0.81	0.06	57.86	
Tukey M-estimator	0.81	0.06	57.86	
Fair M-estimator	0.82	0.07	61.83	
Cauchy M-estimator	0.81	0.06	57.86	
Welsch M-estimator	0.78	0.03	48.12	

Source: own elaboration.

The property values derived from the various models are similar to each other. For the standard deviations of the estimated values, an improvement of several to several dozen percent was achieved compared to the standard deviation values obtained from the least squares method.

A dataset containing information on properties sold in the local market of Tarnów was used for the calculations. The market value was estimated using various M-estimators from the group of monotone estimators (Huber estimator and Fair estimator), soft-redescending estimators (Cauchy estimator and Welsch estimator), and hard-redescending estimators (Hampel estimator and Tukey estimator). All analysed M-estimators led to an increase in the coefficient of determination value and a decrease in standard estimation errors. Each algorithm detected the outliers. The highest value of the coefficient of determination was observed for the monotonic estimation using Fair's estimator. For Fair estimator, the standard errors were the smallest. Results for most estimators are comparable and clear. The smallest changes in results compared to the least squares method were observed for the Welsch estimator.

The results obtained depend on the characteristics of the data, and the choice of the best estimator may

Table 5. Determined value for several properties with analysis of accuracy

	Value [PLN/m^2]					
	W_1	W_2	W_3	W_4	W_5	W_6
	Standard deviation [PLN/m^2]					
	σW_1	σW_2	σW_3	σW_4	σW_5	σW_6
Ordinary Least Square	5,052.73	6,408.98	3,791.49	5,001.03	5,455.93	4,510.75
	61.63	145.70	151.00	157.65	135.54	182.76
Huber M-estimator	5,053.98	6,466.66	3,757.47	4,947.65	5,486.90	4,538.24
	53.46	124.45	128.38	137.15	115.74	156.84
Hampel M-estimator	5,053.97	6,478.49	3,749.03	4,944.22	5,494.29	4,534.56
	53.15	124.02	127.75	136.64	115.30	155.97
Tukey M-estimator	5,048.73	6,466.76	3,742.91	4,932.80	5,486.98	4,532.90
	53.08	123.75	126.90	135.16	114.66	155.04
Fair M-estimator	5,053.34	6,437.35	3,769.57	4,946.81	5,475.97	4,545.14
	51.01	119.38	119.26	125.91	108.12	146.70
Cauchy M-estimator	5,052.09	6,455.58	3,756.37	4,943.08	5,482.94	4,539.20
	52.54	122.15	124.74	132.54	112.70	152.73
Welsch M-estimator	5,051.25	6,437.39	3,770.05	4,973.02	5,470.78	4,520.34
	57.19	133.95	138.07	145.81	124.27	168.09

Source: own elaboration.

✉tomasz.adamczyk@agh.edu.pl

vary across different property markets. The absence of universal methods for fitting estimators is demonstrated by the continually growing number of available estimators, with more than 50 currently known. The selection of the best estimator may even vary within the same local property market, where the valuer makes subjective assessments of the location or other attributes. These estimators also influence the characteristics of the data analysed. The use of M-estimators makes it possible to objectively reduce the impact of data that may be related to abnormal behavior in the property market, which could otherwise affect the valuation.

CONCLUSIONS

The use of robust estimates enables the objective filtering of data through algorithms, reducing the risk of discretion and error in analysis or the failure to detect anomalies. It is important to understand the characteristics of the data being used, as this understanding facilitates the selection of the most appropriate robust estimation method, including the choice and fitting of the best weighting function or estimation technique. The choice largely depends on the nature of the outliers. It is important to note that robust analysis with M-estimation is not resistant to leverage points, and robust regression should not be automatically used as a replacement for classical methods. The higher the Breakdown Point (BP) of an estimator, the more robust it is. However, as stated by Albuquerque and Biegler (1996), "The more robust an estimator is, the less efficient it is".

Developed to study the large-sample properties of robust statistics, M-estimation is a general statistical approach that simplifies and unifies estimation. However, it may not be appropriate for small samples. M-estimation can be easily implemented in statistical analysis software, with examples of solutions available in applications such as Matlab, R, SAS, or Python.

M-estimation is a useful tool when data are not independent and identically distributed, a common occurrence in data fusion analyses. Consequently, data from the real estate market can be subjected to this type of analysis. This assumption may also be relevant for property data, which may contain certain imperfections.

Robust estimation models can be successfully applied to the analysis of real estate markets. Using robust estimation for data "clean-up" allows for the removal of transactions exhibiting price discrepancies from the data set, treating them as outliers. This is a complex process that requires the selection of appropriate estimators, which can vary depending on the specific property market.

Many M-estimators have been proposed in the robust statistics literature, and some of these have been used extensively for regression analysis.

REFERENCES

Abdulhafedh, A. (2022). Incorporating Multiple Linear Regression in Predicting the House Prices Using a Big Real Estate Dataset with 80 Independent Variables. *Open Access Library Journal*, *9*(1). https://doi.org/10.4236/oalib.1108346

Adamczyk, T. (2017). Application of the Huber and Hampel M-estimation in real estate value modeling. *Geomatics and Environmental Engineering*, *11*(1), 15–23. https://dx.doi.org/10.7494/geom.2017.11.1.15

Adamczyk, T. (September, 2017). *Application of Huber and Hampel m-estimation in analysing of real estate price volatility over time* [Conference proceedings]. Geographic Information Systems Conference and Exhibition "GIS ODYSSEY 2017", Trento – Vattaro, Italy.

Albuquerque, J. S, & Biegler, L. T. (1996). Data reconciliation and gross-error detection for dynamic systems. *AIChE Journal*, *42*(10), 2841–2856. https://doi.org/10.1002/aic.690421014

Anderws, D. F. (1974). A Robust Method for Multiple Lenear Regression. *Technometrics*, *16*(4), 523–531. https://doi.org/10.2307/1267603

Arya, K. V., Gupta, P., Kalra, P. K., & Mitra, P. (2007). Image registration using robust M-estimators. *Pattern Recognition Letters*, *28*(15), 1957–1968. https://doi.org/10.1016/j.patrec.2007.05.006

Banaś, M., & Ligas, M. (2014). Empirical test of performance of some M-estimators. *Geodesy and Cartography*, *63*(2), 127–146. https://doi.org/10.2478/geocart-2014-0010

✉tomasz.adamczyk@agh.edu.pl

Banaś, M., Czaja, J., & Dąbrowski, J. (2022). The Application of Regression Analysis for Estimating the Market Value of Commercial Real Estate. *Geomatics and Environmental Engineering*, *16*(1), 17–35. https://doi.org/10.7494/geom.2022.16.1.17

Chen, C. (2002). *Robust Regression and Outlier Detection with the ROBUSTREG Procedure* [Conference Paper]. Proceedings of the Twenty-Seventh Annual SAS Users Group International Conference, SAS Institute Inc., Cary, NC.

Dehnel, G. (2015). Robust regression in Monthly Business Survey, *Statistic in Transition new series*, *16*(1), 137–152. https://doi.org/10.59170/stattrans-2015-008

Dehnel, G., & Gołata, E. (2016). M-estimation in a Small Business Survey. *Krakow Review of Economics and Management Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie*, *1*(949), 5–21. https://doi.org/10.15678/ZNUEK.2016.0949.0101

Grover, R., & Walacik, M. (2019). Property valuation and taxation for fiscal sustainability–lessons for Poland. *Real Estate Management and Valuation*, *27*(1), 35–48. https://doi.org/10.2478/remav-2019-0004

Hampel, F. R. (1968). *Contribusion to the theory of robust estimation*. Univeristy of California.

Huber, P. (1964). Robust Estimator of a Location Parameter. *Ann. Math. Statist*, *35*(1), 73–101. https://doi.org/10.1214/aoms/1177703732

Jana, P., Rosadi, D., & Supandi, E. D. (2023). Comparison of robust estimation on multiple regression model. *Journal of Mathematics and Its Applications*, *17*(2). https://doi.org/10.30598/barekengvol17iss2pp0979-0988

Janssen, Ch., Söderberg, B., & Zhou, J. (2001). Robust estimation of hedonic models of price and income for investment property. *Journal of Property Investment & Finance*, *19*(4), 342–360. http://dx.doi.org/10.1108/EUM0000000005789

Lemeš, L., & Akagic, A. (2022). Prediction of Real Estate Market prices with Regression Algorithms. In Ademović, N., Mujčić, E., Mulić, M., Kevrić, J., Akšamija, Z. (Eds.), *Advanced Technologies, Systems, and Applications VII. IAT 2022. Lecture Notes in Networks and Systems* (Vol. 539) (pp. 401–411). Springer. http://dx.doi.org/10.1007/978-3-031-17697-5_32

de Menezes, D. Q. F., Prata, D. M., Secchi, A. R., & Pinto, J. C. (2021). A review on robust M-estimators for regression analysis. *Computer & Chemical Engineering*, *147*, Article 107254. https://doi.org/10.1016/j.compchemeng.2021.107254

Mulry, M. H., Kaputa, S., & Thompson, K. J. (2018). Setting M-Estimation Parameters for Detection and Treatmen of Infuential Values. *Journal of Office Statistics*, *34*(2), 483–501. https://doi.org/10.2478/jos-2018-0022

Nowel, K. (2019). Squared $M_{split(q)}$ S-transformation of control network deformations. *Journal of Geodesy*, *93*, 1025–1044. https://doi.org/10.1007/s00190-018-1221-4

Özyurt, D. B., & Pike, R. W. (2004). Theory and practice of simultaneous data reconciliation and gross error detect for chemical processes. *Computer and Chemical Engineering*, *28*(3), 381–402. https://doi.org/10.1016/j.compchemeng.2003.07.001

Pires, R. C., Simoes Costa, A., & Mili, L. (1999). Iteratively Reweighted Least-Squares State Estimation Through Givens Rotations. *IEEE Transactions on Power System*, *14*(4). https://doi.org/10.1109/59.801941

Raza, A., Talib, M., Noor-ul-Amin, M., Gunaime, N., Boukhris, I., & Nabi, M. (2024). Enhancing performance in the presence of outliers with redescending M-estimators. *Scientific Report*, *14*, Article 13529. https://doi.org/10.1038/s41598-024-64239-6

Ross, R. K., Zivich, P. N., Stringer, J. S. A., & Cole, S. R. (2024). M-estimation for common epidemiliological measures: introduction and applied examples. *International Journal of Epidemiology*, *53*(2). https://doi.org/10.1093/ije/dyae030

Ruckshtul, A. (2016). *Robust Fitting of Parametric Models Based on M-Estimation. IDP Institute of Data Analysis and Process Design; ZHAW Zurich University of Applied Sciences in Winterthur*. https://ethz.ch/content/dam/ethz/special-interest/math/statistics/sfs/Education/Advanced%20Studies%20in%20Applied%20Statistics/course-material-1719/Robuste/robstat18E.pdf

Susanti, Y., Pratiwi, H., Sulistijowati, S. H., & Liana, T. (2014). M estimation, S estimation, and MM estimation in robusty regression. *International Journal of Pure and Applied Mathematics*, *91*(3), 349–360. https://doi.org/10.12732/ijpam.v91i3.7

Śpiewak, B. (2018). Application of Passive Methods of Robust Estimation: Baarda's and Pope's in Real Estate Market Analysis. *Real Estate Management and Valuation*, *26*(1). http://dx.doi.org/10.2478/remav-2018-0001

Śpiewak, B., & Barańska A., (2020). *Chosen statistical methods for the detection of outliers in real estate*

tomasz.adamczyk@agh.edu.pl

market analysis. Acta Scientiarum Polonorum Administratio Locorum, *19*(2), 109–118. https://doi.org/10.31648/aspal.4608

Walacik, M., Cellmer, R., & Źróbek, S. (2013). Mass appraisal–international background, Polish solutions and proposal of new methods application. *Geodetski list*, *67*(4), 255–269.

Wiśniewski, Z. (2009). Estimation of parameters in a split functional model of geodetic observations (M_{split} estimation). *Journal of Geodesy*, *83*(2), 105–120. https://doi.org/10.1007/s00190-008-0241-x

Wiśniewski, Z. (2010). $M_{split(q)}$ estimation: estimation of parameters in a multi split functional model of geodetic observations. *Journal of Geodesy*, *84*, 355–372. https://doi.org/10.1007/s00190-010-0373-7

✉tomasz.adamczyk@agh.edu.pl