STATISTICS IN TRANSITION new series, December 2016
Vol. 17, No. 4, pp. 749–762

M-ESTIMATORS IN BUSINESS STATISTICS[1]

Grażyna Dehnel[2]

ABSTRACT

Recent years have seen a dynamic development in statistical methods for analysing data contaminated with outliers. One of the more important techniques that can deal with outlying observations is robust regression, which represents four decades of research. Until recently the implementation of robust regression methods, such as M-estimation or MM-estimation, was limited owing to their iterative nature. With advances in computing power and the growing availability of statistical packages, such as R and SAS, Stata, the applicability of robust regression methods has increased considerably.The aim of the study is to evaluate one of these methods, namely M-estimation, using data from a survey of small and medium-sized businesses. The comparison involves nine *M-estimators*, each based on a different weighting function. The results and conclusions are formulated on the basis of empirical data from the DG-1 business survey.

Key words: robust regression, M-estimation, business statistics, outliers

1. Introduction

Robust regression provides estimators which eliminate the influence of outliers. In the literature it is sometimes presented as a method designed to ignore outlying observations. It is often contrasted with methods aimed at detecting outliers. The fact is, however, that both detection and robust regression pursue the same objectives – the only difference is how they are achieved (Rousseeuw, Leroy, 1987). In the case of detection, the first step involves identifying outliers; only then are data corrected. In the case of robust regression, first a regression model is fitted to most of the data, then outliers can be detected, based on residual values.

While each of the two approaches has its benefits and drawbacks, it is the techniques of robust estimation that have been attracting growing interest recently. A number of approaches to robust regression have been proposed in an attempt to improve its performance. The starting point for the work was Ordinary Least Squares (OLS). One of its first modifications was *M*-estimation, which is

[1] The project is financed by the Polish National Science Centre, decision DEC- 2015/17/B/HS4/00905.
[2] Poznan University of Economics, Department of Statistics. E-mail: g.dehnel@ue.poznan.pl.

characterized by a low breakdown point by high efficiency. The next development was *S-estimation* and *LTS-estimation*, with a high breakdown point. The group of the latest methods includes *MM-estimation*. The *MM-estimator* was the first estimator with a high breakdown point and high efficiency under normal error (Stromberg, 1993). Some of these methods were developed in the 1970s and 80s of the 20th century, but because they rely on computationally demanding iterative procedures, they had limited applications. Nowadays a number of statistical packages are available, such as R or SAS, Stata, which facilitate the implementation of robust regression methods (Verardi, Croux, 2009). The growing interest in techniques of robust regression is also due to the fact that their application, unlike other methods, does not require earlier detection of outliers.

When choosing robust regression methods, one should keep in mind the properties of particular methods, which have been identified in the literature (Holland, Welsch 1977), (Huber, 1981), (Hampel *et al*., 1986), (Chen, Yin, 2002). One of the latest methods is *MM-estimation*. It is a combination of two different methods: efficient estimation of an *M-estimator* and *S-estimation* or a *LTS-estimation* with a high breakdown point. Hence, the ultimate quality of estimation depends on the quality of each of the two approaches. In each approach it is necessary to make additional decisions about the choice of parameters and functions. The present article is limited to an analysis of the properties of one of these approaches – *M-estimation*.

The study was aimed at evaluating properties of *M-estimators*, where different weighting functions were used. The evaluation was based on an empirical study using data on small and medium-sized enterprises in the transport section of the classification of economic activities (NACE Rev.2).

2. M-estimation

The class of *M-estimators* is a generalization of Maximum likelihood type estimators (MLE). *M-estimators* are classified as part of robust regression estimators of the so-called 1st generation. It is a group which is characterized by a low breakdown point in the case of *x-outliers*. The M-estimator was introduced by Huber in 1964 (Huber, 1964). It is a robust equivalent of the approach represented by the least squares method (Chen, 2007). The loss function of the least squares method is replaced by another loss function $\rho(\cdot)$, which is less sensitive to extreme residual values

$$\hat{\theta}_M = \arg\min_{\theta} \sum_{i=1}^{n} \rho\left(\frac{r_i}{s}(\theta)\right) \tag{1}$$

where: ρ - loss function

s - scale parameter

$r_i = y_i - X\theta$.

To ensure observed response values have comparable variation with respect to the values of the dependent variable, residuals are standardized using a dispersion measure *s*. The use of a classic measure of dispersion for the purposes of standardization, in the presence of outliers, results in overestimated values. For this reason, standard deviation is replaced with other measures of dispersion, such as median absolute deviation (MAD) or interquartile range (IQR).

The objective function meets the following conditions (Banaś, Ligas, 2014):

- non-negativity
- equals zero when its argument equals zero ($\rho(0)=0$)
- is symmetric (even function) ($\rho(r_i)=\rho(-r_i)$),
- monotonicity in $|r_i|$ ($\rho(r_i)\geq\rho(r_j)$ for $|r_i|>|r_j|$).

Assuming the scale parameter *s* is known, an estimate of the estimator θ_M is obtained by solving a system of *p* equations with respect to vector θ expressed as a product of independent variables and partial derivatives of the ρ function:

$$\sum_{i=1}^{n}\Psi\left(\frac{y_i-\sum_{k=1}^{p}x_{ik}\theta_k}{s}\right)x_i=0 \tag{2}$$

where: Ψ - influence function, a derivative of ρ function

p – the number of variables *x*.

In addition to the loss function, the influence function is another one that characterizes *M-estimators*. It helps to assess how a single observation affects the value of the estimator. Equation (2) is typically solved by means of *iteratively reweighted least squares* (IRLS) with weights given by the following formula (Trzpiot, 2013):

$$w_i=\Psi\left(\frac{y_i-\sum_{k=1}^{p}x_{ik}\theta_k}{s}\right)\Big/\left(\frac{y_i-\sum_{k=1}^{p}x_{ik}\theta_k}{s}\right) \tag{3}$$

where: w_i - weighting function.

The weighting function w_i, which is a ratio of the influence function and the residual, is the third function which characterizes *M-estimation*. The weighting function meets the following conditions (Banaś, Ligas, 2014):

- is continuous, symmetric,
- decreases when the residual increases,
- is equal to one when its argument is zero,
- decreases to zero for the argument increasing to +/- infinity.

The values of weights depend on what function Ψ is chosen to correspond with function ρ. In the literature of the subject many variations of *M-estimators*

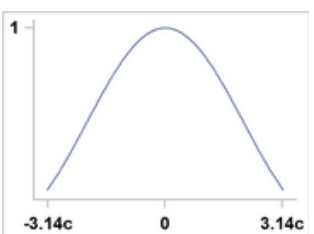
are suggested, with different variants of function Ψ (Fair, 1974), (Holland, Welsch 1977), (Huber, 1981), (Hampel *et al*., 1986), (Chen, Yin, 2002), (Banaś, Ligas, 2014). Curves of the weighting functions are different, but in *M-estimation* one always attempts to minimize or eliminate the influence of outliers. For this reason, all the proposed weighting functions cut off or reduce the influence of large residuals on the estimation of function parameters and/or scale parameter.

The selection of function Ψ is made depending on what weight we want to assign to outliers, among other things. In order to describe some of their properties, evaluation and usefulness, and their influence on estimation results, nine different weighting functions are analysed in this article: Andrews', Tukey's (bisquare), Cauchy's, Fair's, Hampel's, Huber's, logistic, Talworth's and Welsch's weighting functions (see Fig. 1).

Weighting functions have tuning factors, which can be modified. Using tuning factors, it is possible to reduce the impact of outliers with large residuals, but this is achieved at the cost of reducing the estimator efficiency. In the study described in this article, the tuning factors of the weighting function were set in such a way as to ensure 95% efficiency of estimates of M-estimators, see Table 1.

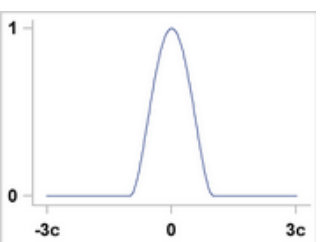
Andrews

$$W(x,c)=\begin{cases}\frac{\sin(\frac{x}{c})}{\frac{x}{c}} & \text{if } |x|\le \pi c\\ 0 & \text{otherwise}\end{cases}$$



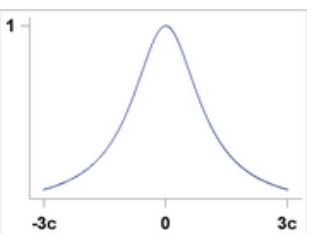
Tukey

$$W(x,c)=\begin{cases}(1-(\frac{x}{c})^2)^2 & \text{if } |x|<c\\ 0 & \text{otherwise}\end{cases}$$



Cauchy

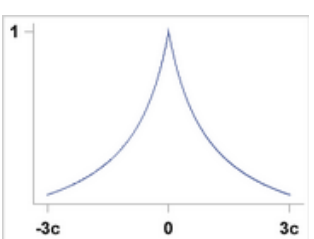
$$W(x,c)=\frac{1}{1+(\frac{|x|}{c})^2}$$



for c=2

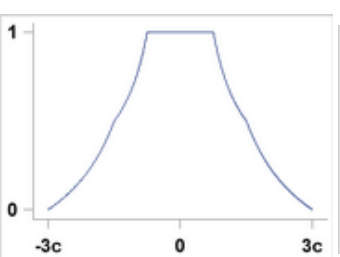
Fair

$$W(x,c)=\frac{1}{(1+\frac{|x|}{c})}$$



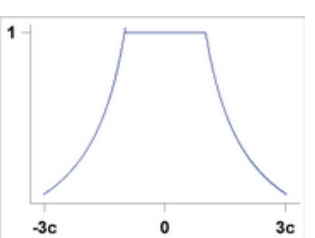
Hampel

$$W(x,a,b,c)=\begin{cases}1 & |x|<a\\ \frac{a}{|x|} & a<|x|\le b\\ \frac{a}{|x|}\frac{c-|x|}{c-b} & b<|x|\le c\\ 0 & \text{otherwise}\end{cases}$$



Huber

$$W(x,c)=\begin{cases}1 & \text{if } |x|<c\\ \frac{c}{|x|} & \text{otherwise}\end{cases}$$



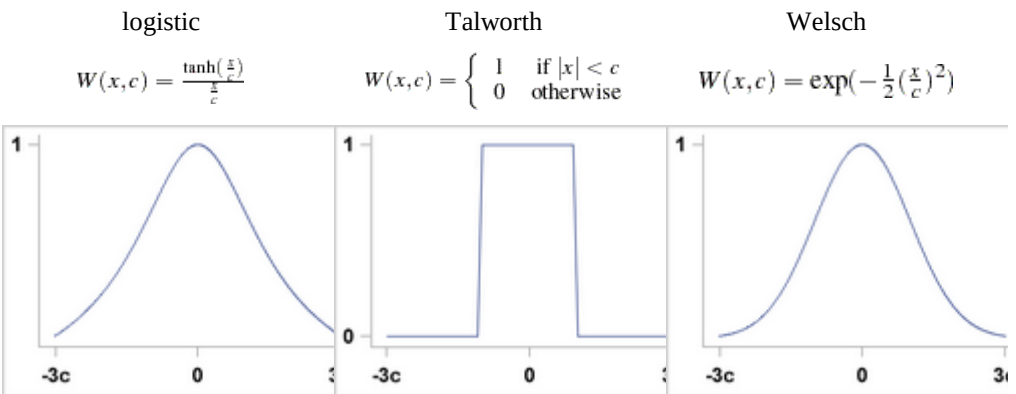


Figure 1. Weighting functions of *M-estimator*

Source: Based on SAS INSTITUTE INC. (2014).

The initial value of $\hat{\theta}_0$ is estimated based on the OLS method. In each iteration *t*, one uses values of residual and weights obtained in iteration *t-1* until convergence is achieved (Alma, 2011). After each iteration, it is also necessary to conduct standardization.

In practice the scale parameter *s* is unknown. One simple and very resistant possibility in these cases is to use the median absolute deviation estimator (Huber, 1964, Ripley, 2004; Trzpiot, 2013) Another possibility is to estimate scale *s* in an MLE-like way (Venables, Ripley, 2002).

Table 1. Tuning factors of the weighting function

Weighting function	Tuning factors a, b, c
Andrews'	1.339
Bisquare	4.685
Cauchy's	2.385
Fair's	1.4
Hampel's	4, 2, 8
Huber's	1.345
Logistic	1,205
Talworth's	2.795
Welsch's	2,985

Source: Based on SAS INSTITUTE INC. (2014).

The *M-estimator* is only resistant to outliers in the *y-direction*; it is not resistant to leverage points. This affects the range of potential applications. Hence, the estimator is frequently used but only in situations where leverage points are not a problem. Its breakdown point is not high and is equal to 1/n. The M-estimator is conditional bias – conditional on the proportion of the outlier in the sample (Cox *et al*., 1995).

3. Evaluation of estimates obtained in the empirical study

A preliminary evaluation of estimates obtained using *M-estimators* was conducted in terms of the goodness of fit of the model, represented by the coefficient of determination. The robust version of the coefficient of determination is defined as:

$$R^2 = \frac{\sum \rho\left(\frac{y_i - \hat{\mu}}{\hat{s}}\right) - \sum \rho\left(\frac{y_i - x_i^T \hat{\theta}}{\hat{s}}\right)}{\sum \rho\left(\frac{y_i - \hat{\mu}}{\hat{s}}\right)} \tag{4}$$

where ρ is the loss function for the robust estimate, $\hat{\mu}$ is the robust location estimator, and $\hat{s}$ is the robust scale estimator in the full model.

Properties of the estimators analysed in the study were evaluated using the bootstrap method. 1000 iterations of drawing samples were made, which were then used to calculate:

- Relative estimation error (REE)

$$CV\left(\hat{Y}_d\right) = \frac{\sqrt{Var\left(\hat{Y}_d\right)}}{E\left(\hat{Y}_d\right)} = \frac{\sqrt{\frac{1}{999}\sum_{b=1}^{1000}\left(\hat{Y}_{b,d} - \hat{Y}_d\right)^2}}{E\left(\hat{Y}_d\right)} \tag{5}$$

- Mean absolute relative bias (ARB)

$$ARB\left(\hat{Y}_d\right) = \frac{1}{1000}\left|\sum_{b=1}^{1000}\frac{\hat{Y}_{b,d} - Y_d}{Y_d}\right| \tag{6}$$

- Relative root mean square error (RMSE)

$$RMSE\left(\hat{Y}_d\right) = \frac{\sqrt{\frac{1}{1000}\sum_{b=1}^{1000}\left(\hat{Y}_{b,d} - Y_d\right)^2}}{Y_d} \tag{7}$$

4. The description of the study

The empirical study was based on information from official statistics collected in a business survey known as DG-1. It is the largest survey in Polish short-term

business statistics. It collects data from businesses employing over 9 people. The survey collects data from all medium-sized and large enterprises and a 10% sample of small businesses. It is conducted on a monthly basis. Its objective is to collect up-to-date information about basic indicators of economic activity of enterprises. In the empirical study only data about small and medium-sized companies were used (with the number of employees ranging from 10 to 250), which conducted their business activities in December 2011. In the model considered in the study revenue was the dependent variable. Independent variables came from an administrative register. Three independent variables were used in the model: profit, cost and the number of employees. A 10% sample of small and medium-sized companies from the DG-1 survey was treated as the general population. The domain of study was created by cross-classifying the administrative division into provinces with the NACE category of business activity. Results of the study were limited to domains included in one NACE section: *transport*. The section was chosen on the basis of the assessment of the goodness of fit of the regression model to the empirical data. The main motivation for the choice of the section was to ensure that domains it contained were characterized by the presence of outliers, which considerably reduced the quality of the classical model of regression (see. Table 2).

The first stage of the analysis involved assessing the distribution of businesses in terms of variables included in the model. Values of the basic descriptive statistics for all the variables were characterized by high variability and strong asymmetry. In the case of the variable 'Revenue', the coefficient of variation amounted to as much as 405%, while skewness was as high as 5.63.

Table 2. Statistical characteristics of the distribution of the *Revenue* variable (in thousand PLN by province and section 'Transport', 2011

Province	**CV**(%)	**Skewness**	$\mathbf{R^2}$	**Percentage of outliers** (%)	**N**
Dolnośląskie	90	1.24	0.537	7.1	28
Kujawsko-Pomorskie	100	1.45	0.945	16.7	24
Lubelskie	231	4.22	0.995	12.0	25
Lubuskie	303	3,99	0.139	15.0	20
Łódzkie	106	2.47	0.991	6.7	30
Małopolskie	327	5.62	0.999	11.8	34
Mazowieckie	405	5.63	0.999	6.8	73
Opolskie	89	0.91	0.984	28.6	14
Podkarpackie	72	0.27	0.982	18.8	16
Podlaskie	286	3.45	0.999	8.3	12
Pomorskie	138	2.38	0.980	3.0	33
Śląskie	261	5.55	0.992	9.4	64
Świętokrzyskie	143	2.52	0.998	13.0	23
Warmińsko-Mazurskie	65	0.76	0.892	15.4	13
Wielkopolskie	98	1.87	0.993	8.2	49
Zachodniopomorskie	138	2.21	0.970	13.3	30

Source: Own calculations based on DG1 survey.

In addition, Student's t-test and Cook's D confirmed the presence of outliers. These properties indicated the need for the use of robust regression method.

The percentage share of outliers, as well as the value of the coefficient of determination R^2 are presented in Table 2. The assessment of these two parameters indicates that the percentage share of outliers is not correlated with the value of the coefficient of determination. In the case of both more and less numerous sections, even a relatively large number of outliers does not necessarily have a negative impact on the model fit. On the other hand, individual outliers may have a large influence on the quality of the model, for the impact of outliers depends not only on their number but also on their type **(**outliers in the *x-direction*, outliers in the *y-direction*) and their distance from typical observations. The graphic presentation showing the relationship between the type of outliers and the model quality only shows domains with the lowest values of the coefficient of determination, that is for provinces of *Dolnośląskie, Lubuskie, Warmińsko-Mazurskie, Zachodniopomorskie* (see. Fig. 2)

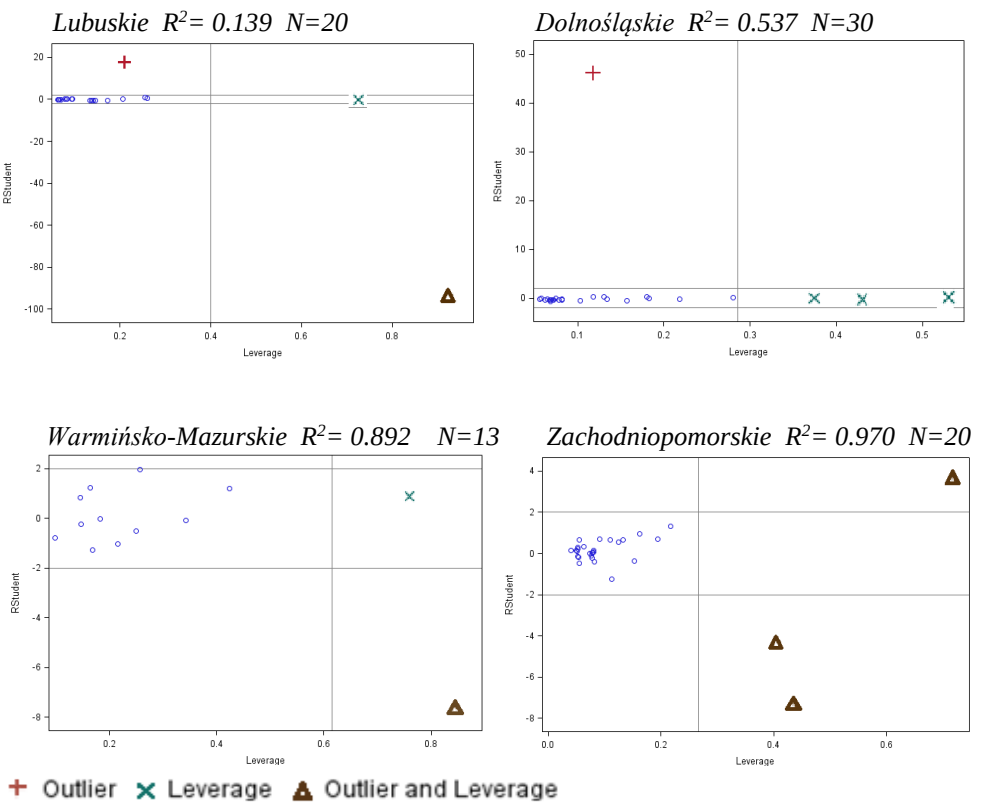


Figure 2. Outlier and Leverage diagnostic for *transport* in selected provinces

Source: Calculations based on the DG1 survey and the tax register of December 2011.

5. Empirical results of the study

The aim of the study was to compare the properties of *M-estimators*, in which different weighting functions were used. 9 types of *M-estimators* were analysed. The analysis was divided into two parts. The first part involved assessing the quality of the model's goodness of fit based on the robust version of the coefficient of determination and estimation errors of the equation parameters.

Values of the coefficient of determination are shown in Fig. 3. Differences in values obtained for each type of *M-estimator* reflect their sensitivity to the presence of different kinds of outliers (in the *x-direction* or in the *y-direction)* and their distance from the bulk of the data. The analysis of the results suggests that the use of *M-estimation* improves the goodness of fit of the model only when *y-outliers* are present. In the case of *x-outliers*, the application of *M-estimation* resulted in lower values of the coefficient of determination (compared to OLS), see Table 2 and Fig. 3.

In the domains analysed in the study, the highest values of the coefficient of determination were recorded for Fair's and Huber's functions, while the lowest ones for Cauchy's and Hampel's functions. Also, Talworth's and Tukey's functions are noteworthy. As can be seen from the results, the application of these functions in domains where the influence of outliers on the model quality is large results in a considerable improvement in efficiency and the robustness of *M-estimators*. This is due to the fact that they completely ignore observations for which large residuals were recorded.

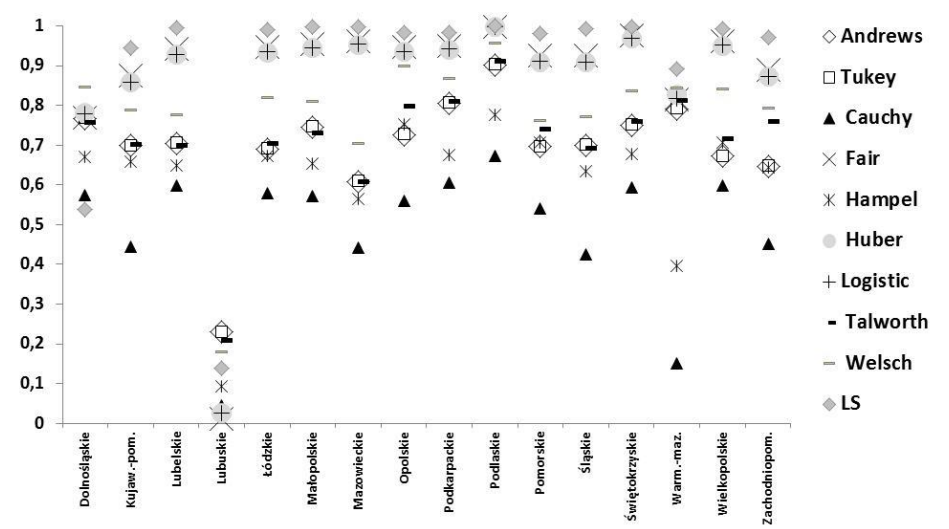


Figure 3. The coefficient of determination for the regression models of *Transport* in provinces

Source: Calculations based on the DG1 survey and the tax register of December 2011.

The study also investigated the values of model parameters and their estimation errors (standard errors) for the following variables: *profit*, *cost* and *the number of employees*, see Fig. 4. Both the estimated parameters and standard errors indicate high similarity of estimates in the domains of interest for all the weighting functions considered. The only case in which there were significant differences in estimates of the equation parameters (the slope) as well as the standard error was the number of employees in the province of Lubuskie. It is noteworthy that this situation applies to domains with the lowest values of the coefficient of determination for all kinds of the *M-estimator* ($R^2 \in (0{,}03;0{,}23)$). Additionally, in this domain there is a leverage point far removed from the bulk of the data, see Fig. 2.

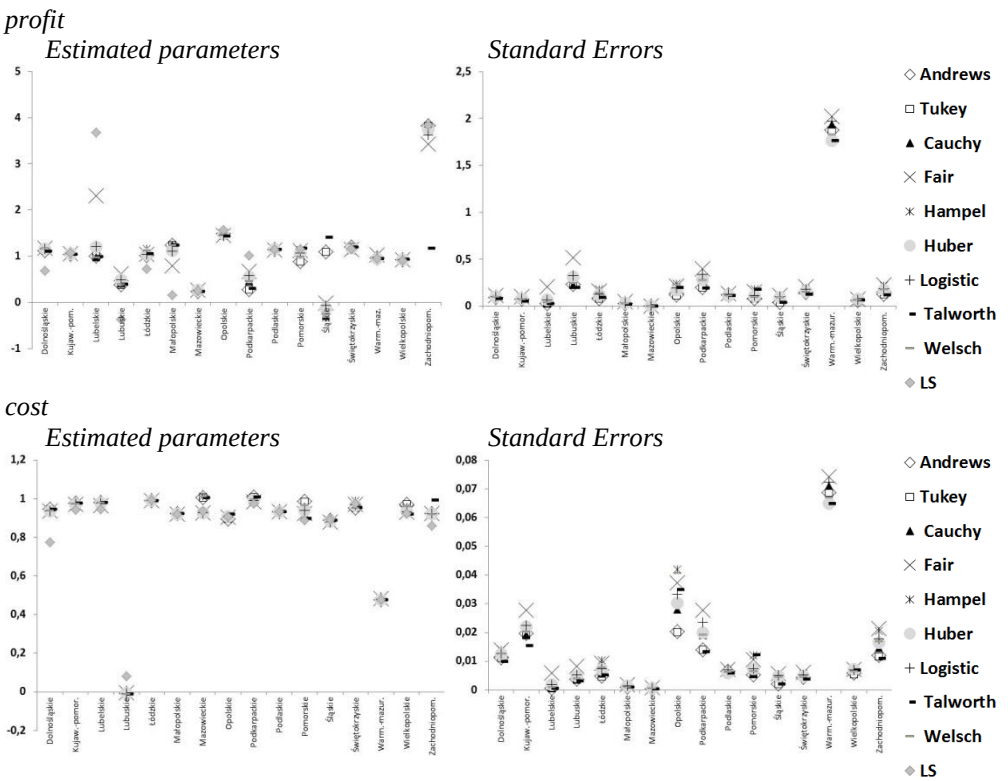


Figure 4. Estimated parameters and Standard Errors of the weighting functions for profit, cost and the number of employees

the number of employees

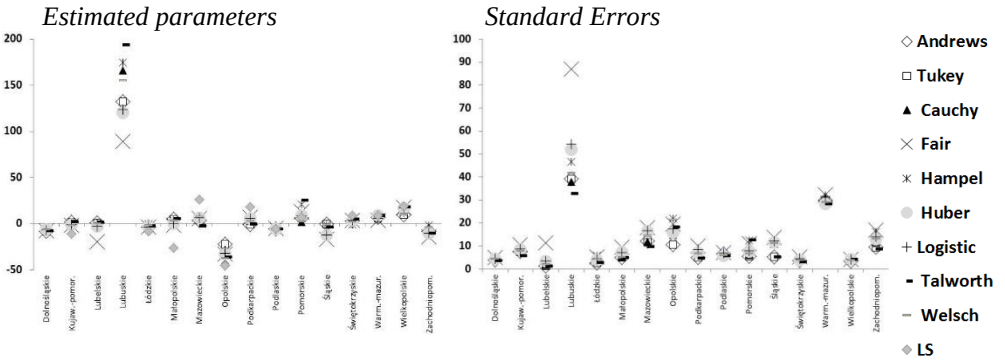


Figure 4. Estimated parameters and Standard Errors of the weighting functions for profit, cost and the number of employees (cont.)

Source: calculations based on the DG1 survey and the tax register of December 2011.

The second stage of the analysis consisted in comparing the properties of parameter estimators for the regression equations derived on the basis of the weighting functions. The bootstrap method was applied to determine measures for the assessment of the efficiency, bias and MSE, see Fig. 5. In the case of the variable *'The number of employees*' mean absolute relative bias (ARB) does not exceed 30%, while the relative estimation error is below 20%. This variable is least correlated with the variable of interest.

Relative estimation error (REE)

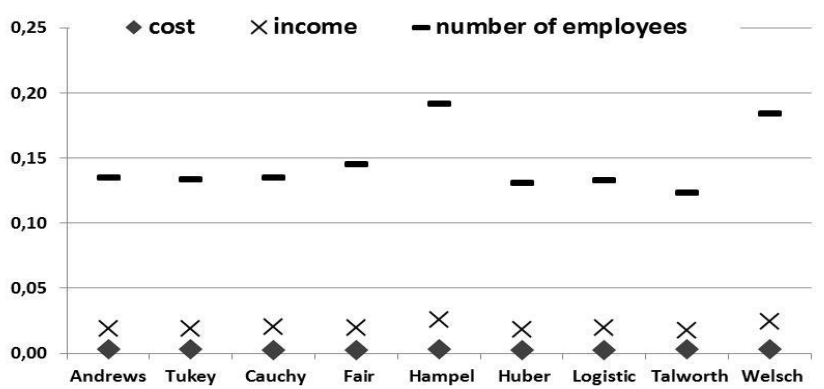
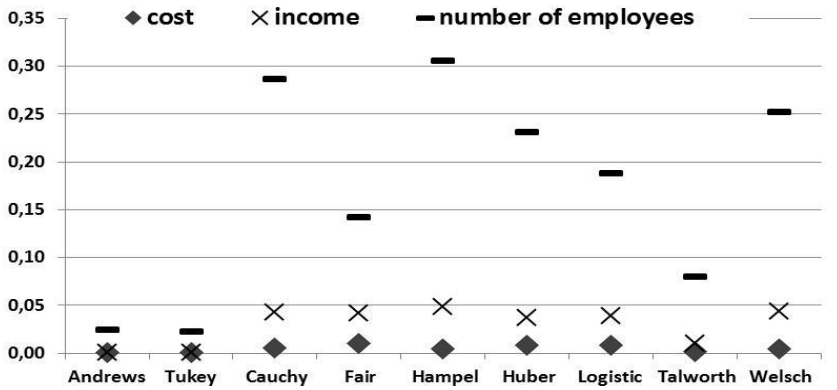


Figure 5. Performance criteria of estimates of equation parameters for profit, cost and the number of employees

Mean absolute relative bias (ARB)



Relative root mean square error (RMSE)

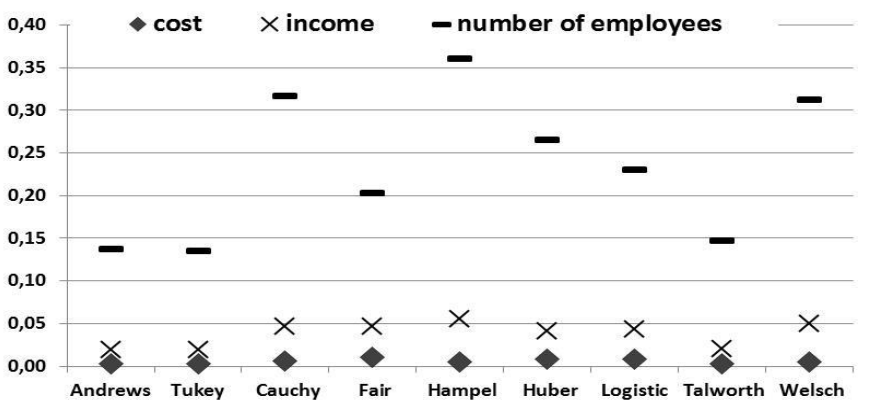


Figure 5. Performance criteria of estimates of equation parameters for profit, cost and the number of employees (cont.)

Source: Calculations based on the DG1 survey and the tax register of December 2011.

For the two remaining variables included in the model (*cost* and *profit*), the relationships between the values of characteristics of different types of *M-estimators* are very similar, both in terms of efficiency, and bias, which has a direct influence of MSE. The values of REE and ARB are close to zero. The amount of bias and efficiency for most cases is almost insignificant.

6. Conclusion

- The use of the *M-estimator* in the presence of outliers can considerably improve the quality of the model's fit compared to the classical method of estimation – it largely depends on the type of outliers. The *M-estimator* is only resistant to *y-outliers* but is not resistant to *leverage points*. It should therefore be used in situations where there are no *leverage points*.
- In practical applications of *M-estimation*, the selection of function Ψ is not a key choice for obtaining good robust estimates. The adoption of each of the nine weighting functions analysed in the study yielded similar results from the viewpoint of the values of estimated parameters and their standard errors. The least adequately fitted models were those based on Cauchy's and Hampel's functions. The best fit was obtained for the models based on Fair's and Huber's functions; one drawback in their case was the relatively high level of standard errors.
- The largest gain in efficiency and robustness of *M-estimators* was obtained when Talworth's and Tukey's functions were used. This result was particularly visible for domains in which the influence of outliers on the quality of the classical LS model was very strong. Owing to the curve shapes of Talworth's and Tukey's functions, observations with large residuals are ignored.

REFERENCES

ALMA, Ö. G., (2011). Comparison of Robust Regression Methods in Linear Regression, [in:] Int. J. Contemp. Math. Sciences, Vol. 6, No. 9, pp. 409–421.

BANAŚ, M., LIGAS, M., (2014). Empirical tests of performance of some M-estimators, Geodesy And Cartography, Vol. 63, No. 2, pp. 127–146.

CHEN, C., (2007). Robust Regression and Outlier Detection with the ROBUSTREG Procedure, SUGI, http://www2.sas.com/proceedings/sugi27/ pp.265–27.pdf.

CHEN, C., YIN, G., (2002). Computing the Efficiency and Tuning Constants for M-Estimation, Proceedings of the 2002 Joint Statistical Meetings, 478–482.

COX, B. G., BINDER, A., CHINNAPPA, N. B., CHRISTIANSON, A., COLLEDGE, M. J., KOTT, P. S., (1995). Business Survey Methods, John Wiley and Sons.

FAIR, R. C. (1974). On the robust estimation of econometric models, Ann. Econ. Social Measurement, 3, pp. 667–678.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. STAHEL, W. A., (1986). Robust Statistics: The Approach Based on Influence Functions, Wiley, New York.

HOLLAND, P., WELSCH, R., (1977). Robust Regression Using Interactively Reweighted Least-Squares, Commun. Statist. Theor. Meth., 6, 813–827.

HUBER, P. H., (1964). Robust estimation of a location parameter, The Annals of Mathematical Statistics, 35, pp. 7–101.

HUBER, P. H., (1981). Robust Statistics, New York: John Wiley and Sons.

RIPLEY, B. D., (2004). Robust Statistics, M.Sc. in Applied Statistics MT2004, 1992-2004, https://www.stats.ox.ac.uk/pub/StatMeth/Robust.pdf.

ROUSSEEUW, P. J., LEROY, A. M., (1987). Robust Regression and Outlier Detection. Wiley-Interscience, New York.

SAS INSTITUTE INC., (2014). SAS/STAT® 13.2 User's Guide. The Robustreg Procedure Cary, NC: SAS Institute Inc.

STROMBERG, A. J., (1993). Computation of high breakdown nonlinear regression parameters, [in:] Journal of the American Statistical Association, 88 (421).

TRZPIOT, G., (2013). Wybrane statystyki odporne [Selected resistant statistics], [in:] Studia Ekonomiczne, No. 152, pp. 162–173, Uniwersytet Ekonomiczny w Katowicach.

VENABLES, W. N., RIPLEY, B. D., (2002). Modern Applied Statistics with S-PLUS. Springer-Verlag.

VERARDI, V., CROUX, C., (2009). Robust regression in Stata, [in:] The Stata Journal, 9, No. 3, pp. 439–453.