Pak. J. Statist.
2015 Vol. 31(1), 33-48

ROBUST CENTERING IN THE FIXED EFFECT PANEL DATA MODEL

Nor Mazlina Abu Bakar[1,3§] and **Habshah Midi**[1,2]

[1] Institute of Mathematical Research, Universiti Putra Malaysia
Serdang, Malaysia.

[2] Faculty of Science, Universiti Putra Malaysia, Serdang, Malaysia

[3] Universiti Sultan Zainal Abidin, Terengganu, Malaysia

[§] Corresponding author: normazlina@gmail.com

ABSTRACT

In recent years, robust estimators for fixed effect panel data model have been developed to provide alternatives to the least square estimates in the presence of outliers. The robust adaptation involves transformation of data by the median instead of the mean in order to eliminate any unobserved effect. Median centering is chosen due to its robustness, simple derivation and possesses min max property. However, the procedure introduces non-linearity in the data and causes some robust estimators to lose their regression equivariance property. This study proposed MM-centering to provide robust solutions to the Within Group parameter estimates. The numerical results indicate that the proposed methods are more efficient than the existing method.

KEY WORDS

Panel data; Outliers; Fixed effect model; Robust; Linear regression.

MSC2010 Code: **62F35, 62J05, 62P30**

1. INTRODUCTION

Panel data refers to the pooling of observations on a cross-section of households, countries, firms, etc. over multiple time series (Baltagi, 2005). For the past decade, there has been an increasing trend on the use of panel data in the research of economics and finance.

It is important to point out that researchers must be aware that this type of data settings is susceptible to the occurrence of outliers. They occur due to various reasons such as measurement error, typing error, transmission or copying error (Rousseeuw and Leroy, 2003) or naturally unusual values (Maronna et al., 2006). In panel data, the outliers can be found either vertically, horizontally or leverage (Rousseeuw and Zomeren, 1990). More often, they are found to be concentrated in a few time series or known as block concentrated outliers (Bramati and Croux, 2007).

Robust estimation involves systematic procedures to investigate model deviations caused by outliers (Hampel, 2001). The robust procedures are designed to be less sensitive mainly because it considers the majority proportion of data (Rousseeuw and Leroy, 2003). For the last two decades, we have witnessed a remarkable development in the search for robust estimators of the linear regression. As a result, robust techniques are very well developed in this area. Some of the robust estimators even have the desirable properties of high efficiency, high breakdown point and also bounded influence (Simpson and Montgomery, 1998). On the other hand, only scarce literatures are available regarding the robust estimation for panel data.

For the fixed effect panel data model, Bramati and Croux (2007) elegantly applied the robust Generalized M estimator and also a combination of M and S-estimates to provide alternatives to the classical Within Group estimator. Both within GM-estimator and Within MS estimator achieved moderate breakdown points. In a more recent development, Verardi and Wagner (2011) successfully applied S-estimator for another robust Within Group estimator. In both studies, the same type of robust data transformation is used in which data are centered within the time series by the median instead of the mean in order to eliminate the fixed effect in a robust way. Centering by the median is opted since median have a maximum breakdown point, very easy to calculate and at the same time median is min max. Unfortunately, the procedure is found to produce nonlinearity to the resulting data and caused the robust estimators to lose their equivariance properties (Bramati and Croux, 2007). Their work has motivated us to propose a different centering approach whereby data are centered by the MM-estimate of location. This robust procedure is adopted not only to introduce linearity back to the transformed data but also to increase their performances.

The next section gives an overview to the classical fixed effect model. Section 3 discussed on Robust Within Group GM-estimator (RWGM) and the current robust centering procedure. The proposed robust centering procedures and Robust Within Group MM-estimator (RWMM) are explained in details in Section 4. Simulation techniques that have been carried out in this research are discussed in Section 5. Section 6 provides illustrations on two numerical examples and Section 7 concludes.

2. CLASSICAL WITHIN GROUP ESTIMATOR

The fixed effect linear panel data model can be formulated as below:

$$y_{it} = \alpha_i + x'_{it}\beta + \varepsilon_{it} \tag{1}$$

where $i = 1, \dots, n$ are individual units observed at time series $t = 1, \dots, T$. y_{it} is the dependent variable, α_i are the unobservable time-invariant individual effects, β is $K \times 1$ and x_{it} is the i-th observation on K explanatory variables. The ε_{it} denote the error terms which are assumed to be uncorrelated across time and individual units. The assumption of strict no endogeneity is applied. The classical Within Groups estimator is obtained by firstly transformed the data within each time series by the mean:

$$\tilde{y}_{it} = y_{it} - \frac{1}{n}\sum_{t=1}^{n} y_{it} \tag{2}$$

and

$$\tilde{x}_{it} = x_{it} - \frac{1}{n}\sum_{t=1}^{n} x_{it} \tag{3}$$

The procedure is known as data centering and became an essential part by which the unobserved individual effects are eliminated. It follows from (1) that:

$$\tilde{y}_{it} = \tilde{x}_{it}\beta + u_{it} \tag{4}$$

where u_{it} is the new error term. Thus, the classical Within Estimator, $\hat{\beta}_W$ can be determined by the Ordinary Least Square (OLS) method which minimizes the function:

$$\sum_{t=1}^{n}(\tilde{y}_{it} - \tilde{x}_{it}'\beta)^2 = \sum_{t=1}^{n}(r_{it})^2 \tag{5}$$

3. ROBUST WITHIN GROUP GENERALIZED M-ESTIMATOR (RWGM) – MEDIAN BASED

It is known that OLS produces the best linear unbiased estimator (BLUE) under the usual assumptions of normally distributed, independent and identically distributed errors. However, outliers can immediately alter the normal setting of the data and lead to unreliable estimates of the model. The damaging effect of outliers can be more crucial for the Within Group estimator. Data transformations in (2) and (3) will introduce a lot more outliers into the transformed data due to the non-robust property of the mean. Data in the contaminated time series will be affected in which the values will be greatly inflated or deflated. Thus, a robust data transformation is required to rectify this problem. Bramati and Croux (2007) and Verardi and Wagner (2011) replaced the centering by the mean with the median centering:

$$\tilde{y}_{it} = y_{it} - median\{y_{it}\} \tag{6}$$

and

$$\tilde{x}_{it} = x_{it} - median\{x_{it}\} \tag{7}$$

for $1 \leq i \leq n$ and $1 \leq t \leq T$. Median is chosen simply because it is the simplest robust measure to be derived and also due to its min max property. Median also has the highest breakdown point, in which data can be contaminated up to 50% before the estimate becomes useless. Once data has been robustly transformed by the median, Bramati and Croux (2007) employed the Robust Within Group GM-estimator (RWGM) to estimate the parameters of the panel data model. Generally, the GM-estimators are solutions to normal equations:

$$\sum_{i=1}^{n}\pi_i\psi\left(\frac{y_i - x_i'\beta}{s\pi_i}\right) = 0 \tag{8}$$

This function is known as Schweppe function where π-weights are introduced merely to down weigh any high leverage points with large residuals. Given the right combinations of initial estimator, leverage measures and iteration function, GM-estimator can efficiently identify the real outliers (Bagheri and Habshah, 2009). It has then become standard practices to use the Least Trimmed Square (LTS) or Least Median of Squares (LMS) as initial high breakdown point estimates. They both have 25% and 50% breakdown point respectively. Simpson and Montgomery (1998) indicated that the breakdown property of GM-estimate is usually inherited from the initial estimate which explains why LTS and LMS are more often favoured.

Two types of weighting procedures are considered in order to improve efficiency to the estimates. Initially, the LTS is applied to the robustly centered data to obtain the initial estimates:

$$\hat{\beta}_{LTS}. = \arg\min_{\beta}\sum_{i=1}^{h}[(\tilde{y}_{it} - \tilde{x}_{it}'\beta)^2]_{i:nT} \tag{9}$$

where $[(\tilde{y}_{it} - \tilde{x}_{it}'\beta)^2]_{1:nT} \leq [(\tilde{y}_{it} - \tilde{x}_{it}'\beta)^2]_{2:nT} \leq \ ... \ \leq [(\tilde{y}_{it} - \tilde{x}_{it}'\beta)^2]_{nT:nT}$ are the ranked residuals and h is the number of residuals of the truncated data. According to Rousseeuw and Leroy (2003), the highest breakdown point can be achieved at maximum level of 50% when $h = \left(\frac{nT}{2}\right) + \left(\frac{p+1}{2}\right)$. The LTS is readily installed in statistical softwares such as S-Plus in which h is set at $h = \left(\frac{3nT}{4}\right)$ and 25% breakdown point is achieved. Once the initial estimates have been obtained, the residuals, $r_{it} = \tilde{y}_{it} - \tilde{x}_{it}'\hat{\beta}_{LTS}$ and $\hat{\sigma}_{LTS}^2$ can then be calculated. Tukey's Biweight function is selected as the first weighting scheme:

$$\rho(x) = \begin{cases} \frac{x^2}{2} - \frac{x^4}{2c^2} - \frac{x^6}{6c^4} & if \quad |x| \leq c \\ \frac{c^2}{6} & if \quad |x| > c \end{cases} \tag{10}$$

Under this weighting scheme, any observation with large residual is down weighted. The diagonal elements, W_r in the form of $\rho'(r_{it}/\hat{\sigma}_{LTS})/(r_{it}/\hat{\sigma}_{LTS})$ can be simplified as:

$$W_r = \begin{cases} 0 & if \quad \left|\frac{r_{it}}{\hat{\sigma}_{LTS}}\right| \leq c \\ \left(1 - \left(\frac{r_{it}}{c\hat{\sigma}_{LTS}}\right)^2\right)^2 & if \quad \left|\frac{r_{it}}{\hat{\sigma}_{LTS}}\right| > c \end{cases} \tag{11}$$

c is chosen to be 4.685 to provide a balance between efficiency and robustness (Wagenvoort and Waldmann, 2002). However, it is also important for the GM-estimator to be able to down weigh high leverage points correctly. Bagheri and Habshah (2009) pointed out that the choice of diagnostic measures and weights computation can greatly enhance this ability. Therefore, a second weighting scheme is considered by using the Robust Distance, RD_{it}:

$$RD_{it} = \sqrt{(x_{it} - \hat{\mu})\hat{V}^{-1}(\tilde{x}_{it} - \hat{\mu})'} \tag{12}$$

for $t = 1, \dots, T$ and $i = 1, \dots, n$. The RD_{it} is known to be the robust counterpart of Mahalanobis Distance, MD_{it} (Rousseeuw and Zomeren, 1990). The classical measures of location and scatter in MD_{it} is simply replaced by the robust estimates. Bramati and Croux (2007) employed S-multivariate of location and scale in their proposed RWGM and achieved a positive breakdown point. Other location and scatter estimators such as the Minimum Volume Ellipsoid (MVE) or Minimum Covariance Distance (MCD) can also be considered. Both MVE and MCD have maximum breakdown point at 50%. This study adopts MCD for the location and scatter estimates with fast computation provided by FastMCD (Salibian-Barrera and Yohai, 2006). When RD_{it} has been calculated, the diagonal elements of the second weighting matrix W_x can then be written as:

$$(W_x)_{it} = \min\left(1, \sqrt{\chi^2_{K,0.975}}/RD_{it}\right) \tag{13}$$

Under the two weighting schemes, the weighted least square estimation is applied to obtain an estimate of high breakdown GM-estimate, $\hat{\beta}_{RWGM}$ and be simplified as:

$$\hat{\beta}_{RWGM} = \left(\tilde{X}'W_xW_r\tilde{X}\right)\left(\tilde{X}'W_xW_r\tilde{Y}\right) \tag{14}$$

4. THE PROPOSED ROBUST WITHIN GROUP ESTIMATORS

In this paper, Robust Within Group MM-estimator (RWMM) is proposed. It is important to note that prior to utilizing the proposed estimators, the data centering procedures need to be employed. As already mentioned, the commonly used mean centering procedure is very sensitive to outliers. As an alternative, the median centering is put forward. However, centering by the median produces nonlinearity to the resulting data and affects the equivariance properties of the robust estimators (Bramati and Croux, 2007). Moreover, in an uncontaminated data, median is known to be less efficient than the mean (Maronna et al., 2006). This will certainly affect the efficiency of robust estimators in the absence of outliers. Thus, different type of robust centering is proposed in order to bring back linearity into the transformed data and at the same time provide more efficiency. We propose centering to be done by MM-estimate of location called MM-centering. The proposed centering procedure is incorporated in the establishment of the proposed robust Within Group Estimator. The proposed Robust Within Group estimator consists of the following steps:

Step 1: Employ the proposed MM-centering procedure to the data.

- *MM-centering* considers the MM-estimate of location which can provide not only efficiency and robustness, but also the high breakdown property. The multistage estimation is defined by first computing S-estimates of location and covariance to obtain the preliminary scale estimate, $\hat{\sigma}_n$:

$$\hat{\sigma}_n = \min_t s_n(t) \tag{15}$$

The scale estimate is derived by minimizing an M-estimate of scale. Thus, the preliminary estimate has a high breakdown of up to 50% and also equivariance. The corresponding location S-estimate $\tilde{\mu}_n$ is defined by:

$$\tilde{\mu}_n = \arg_t \ \min \quad s_n(t) \tag{16}$$

Tukey's Biweight is used for the loss function where ρ is taken as:

$$\rho(x)\begin{cases} \frac{1}{6c^4}x^6 - \frac{1}{2c^2}x^4 + \frac{1}{2}x^2 & if\ |x| \le c \\ \frac{c^2}{6} & if\ |x| > c \end{cases} \tag{17}$$

At Gaussian errors, the preliminary estimates are known to be only 28.7% efficient (Ruppert, 1992). Next, by fixing the scale estimate, the location and shape are re-estimated by a more efficient M-estimator. The step provides 95% efficiency at central model (see Ruppert, 1992). The centering procedure by the MM-estimate of location in the study is referred as MM-centering.

$$\tilde{y}_{it} = \ y_{it} - \hat{\mu}\{\, y_{it}\} \tag{18}$$

and

$$\tilde{x}_{it}^{(k)} = x_{it}^{(k)} - \hat{\mu}\{x_{it}^{(k)}\} \tag{19}$$

for $1 \le i \le n$, $1 \le t \le T$ and $1 \le k \le K$, where $x^{(k)}$ is the k-th explanatory variable.

Step 2: Estimate the parameter of the panel data by using the RWGM proposed by Bramati and Croux (2007) or using our proposed RWMM.

M-estimate is known to have absolutely nice statistical properties with a drawback of having a low breakdown point. Yohai (1987) has then proposed MM-estimator to remedy this problem in which the M-estimate is combined with another high breakdown robust method. As a result, MM-estimator provides high breakdown point and at the same time exhibits only small loss of efficiency in the absence of outliers. Following Yohai (1987) the estimator can be defined in 3 stages and applied to the robustly transformed panel data:

i. In the first step, the initial high breakdown coefficient, $\hat{\beta}_0$ and scale estimates, $\hat{\sigma}_S$ are computed by using S-estimates. Again, fast algorithm by Salibian-Barrera and Yohai (2006) can be used to provide increased computational speed.

ii. Next, the second step involved the computation of the residuals, r_{it} based on the initial estimates where $r_{it} = \tilde{y}_{it} - \tilde{x}'_{it}\hat{\beta}_0$. The M-estimate of scale, $\hat{\sigma}$ with high break down point can now be obtained as a solution of:

$$\frac{1}{n}\sum \rho_0\left(\frac{r_{it}}{\hat{\sigma}}\right) = b \tag{20}$$

where $\frac{b}{a} = 0.5$ and $a = \max \rho_0$.

iii. In the last stage, the M-estimate is computed such that $\hat{\beta}_{RWMM}$ is the solution of

$$\sum \psi_1\left(\frac{r_{it}}{\hat{\sigma}}\right) x_{it} = 0 \tag{21}$$

where $\psi_1 = \rho'_1$ to achieve high efficiency. A weight of 0 is assigned to any unusually large residual. As a result, the efficiency of MM-estimator almost coincides with OLS in a clean dataset (Maronna et al., 2006).

The MM-estimator is very well developed and has been modified by quite a number of authors. Its regression function is also readily available in S-Plus as lmRobMM and in R as lmrob. By default, the function gives a highly robust and highly efficient estimate.

5. SIMULATION STUDY

The performances of the MM-centering will be compared to the median centering for the two robust estimators by the Monte Carlo simulation. Following Bramati and Croux (2007), the dependent variable is set to accord the model in Equation (1) by generating $\varepsilon_{it} \sim N(0,1)$, $\alpha \sim U(0,20)$ and the vector of the slope coefficients β set equal to a vector of ones. The explanatory variables are generated from a multivariate standard normal distribution $N(10 \times 1,1)$ where 1 is a $K \times 1$ vector of ones.

Data is contaminated either randomly over all observations (random contamination) or concentrating the contamination in a few times series (block concentrated contamination). Both types of contaminations are done at two different locations; in y-direction and x-direction or leverage. All together, four different types of contamination cases are studied; vertical outliers, leverage, block concentrated vertical outliers and block concentrated leverage, at 5% and 10% level of contamination. The non-contaminated case is also studied for comparisons. For the block concentrated contamination, a few time series are randomly selected from the panel data set and be contaminated only up to 50% as suggested by Bramati and Croux (2007).

Vertical outliers or outliers in the y-direction are generated by inflating the randomly chosen y's from a few time series with $\sim N(20,1)$. Further, to generate block concentrated x-outliers or the leverage points, we inflate x's of the contaminated y's with data points from K-variate normal distribution $N(10 \times 1,1)$. This is done in order to create influential leverage points. In the experiments, we considered panel datasets of $T = 5, 10, 15$ and 20; representing small, medium, and large time series, each with $n = 25, 50, 100$ and 200 units for small, medium and large samples. Univariate regression is considered where $K = 1$ with $M = 1000$ Monte Carlo replications.

Once panel datasets are generated, data are immediately transformed by applying the classical mean centering and two other types of robust centering procedures - the median centering and MM-centering. The classical β coefficients are estimated by the OLS and the robust β coefficients are estimated by the RWGM and RWMM estimators. The average mean square error (MSE) for each case is calculated by comparing the robust estimator's parameter estimates to the true parameter values using the formula:

$$MSE(\hat{\beta}) = \frac{1}{M}\sum_{j=1}^{M}(\hat{\beta}^{(j)} - \beta)^2 \tag{22}$$

where $\hat{\beta}^{(j)}$ is the estimated slope in jth-replication. The root mean square error (RMSE) is given by $[MSE(\hat{\beta})]^{1/2}$. Following Riazoshams et al. (2010) the performance of each technique is evaluated based on the percentage of robustness measures using the ratio of the RMSEs of the estimators compared with the WG-mean centering based estimator for the good data.

The robustness measures for different types of contaminations are presented in Tables 1 to 5. High percentage indicates the improved performance of the robust estimators.

Table 1 shows the robustness measures of the simulated panel data set in the uncontaminated data. The results indicate that RWGM and RWMM are rather efficient under the median centering but their efficiencies quickly deteriorate with the decrease in number of time series, T. The efficiencies for both robust estimators improved vastly under MM-centering with more efficient and stable results. The overall results in Table 1 show that the RWMM provide better estimates than the RWGM in the uncontaminated panel data. Hence, RWMM under MM-centering provides the most efficient and consistent results for the uncontaminated data.

Results for block concentrated contaminations are produced in Table 2 and 3 for vertical outliers and leverage, respectively. It is observed that the classical WG estimations are largely affected in both types of outliers; only less severe when contaminated vertically. On the other hand, both RWGM and RWMM estimators are able to provide improved estimations under the two robust centering. Their performances are seen to increase with the increase of number of time series, T but rather low under the median centering. More stable and greater performances are observed for the robust estimations under MM-centering compared to robust estimations under median centering.

Similar results are obtained when blocks or time series are contaminated in the x-direction. Leverage points are known to cause severe effects to the classical estimates, resulting in low percentage on the robustness measures in all cases. Under the median centering, RWGM and RWMM are able to provide good results with increasing trend as T increases. Once again, the better results are found under the MM-centering regardless of the size of time series. It is also observed that RWMM performs more superior than RWGM under different types of robust centering and contamination levels.

Table 4 and 5 give the results of random contamination for vertical outliers and leverage, respectively. The results are similar to the results of block contamination cases where RWMM performs better than RWGM and their performances under MM-centering outperform other types of centering.

The poor performances of robust estimators in the median-centred data may due to the non-linearity in the median transformed data. Under the robust MM-centering, linearity is brought back into the data and provided improved performances for both RWGM and RWMM. In both newly proposed robust centering, data are required to be centered close to the value of the mean in the non-contaminated data. This also explains the increased performances for the uncontaminated data of both robust estimators, and hence the ability to provide efficient estimates under normality. MM-centering is found to provide more efficiency, stable and consistent results to both RWGM and RWMM.

Table 1
Robustness Measures (%) of Simulated Panel Data Sets for Uncontaminated Data

Contamination Level	N	T	WG Mean Centering	RWGM Median Centering	RWMM MM-Centering	RWGM	RWMM
0%	25	5	-	50.9	63.5	82.6	91.9
		10	-	64.4	79.9	87.9	95.3
		15	-	68.1	79.0	90.3	97.7
		20	-	76.9	87.1	90.8	97.4
	50	5	-	38.9	50.4	81.8	89.2
		10	-	56.6	70.7	91.5	97.0
		15	-	57.2	69.1	90.9	97.4
		20	-	64.1	76.3	90.5	96.9
	100	5	-	29.5	38.1	83.1	90.7
		10	-	43.6	56.5	90.3	96.7
		15	-	47.2	58.4	91.9	98.1
		20	-	52.4	64.5	90.0	97.1
	200	5	-	21.3	27.9	85.3	94.0
		10	-	32.4	43.5	90.7	97.4
		15	-	32.4	41.0	89.7	96.6
		20	-	39.7	50.7	88.7	96.6

Table 2
Robustness Measures (%) of Simulated Panel Data Sets for Block Concentrated Vertical Contamination

Contamination Level	N	T	WG	RWGM	RWMM	RWGM	RWMM
			Mean Centering	Median Centering		MM-Centering	
Block Vertical 5%	25	5	9.2	50.3	60.6	83.7	91.6
		10	9.6	64.5	74.7	87.6	94.1
		15	10.9	67.4	75.4	89.7	94.9
		20	9.9	72.2	81.1	91.2	96.6
	50	5	8.8	37.1	45.5	82.3	89.9
		10	9.5	55.6	67.3	90.8	96.1
		15	10.2	55.8	65.7	88.0	94.3
		20	10.7	63.7	73.8	91.8	97.1
	100	5	8.9	28.7	36.4	83.9	90.4
		10	9.2	43.0	53.9	87.6	94.9
		15	10.1	43.9	53.5	87.6	94.4
		20	9.9	50.9	60.7	89.6	96.4
	200	5	8.9	20.6	26.5	82.5	90.2
		10	9.4	32.9	42.1	90.0	95.2
		15	10.2	32.7	39.9	88.6	94.3
		20	10.5	41.4	50.8	90.9	96.1
Block Vertical 10%	25	5	6.6	49.5	57.8	81.0	87.5
		10	6.5	63.9	73.0	88.9	94.1
		15	7.4	65.8	72.8	89.8	94.0
		20	7.4	72.7	79.2	91.4	95.1
	50	5	6.2	37.4	43.6	84.6	90.0
		10	6.7	53.1	61.7	86.8	90.6
		15	7.7	53.2	61.2	88.7	93.2
		20	7.6	61.6	69.6	91.1	94.9
	100	5	6.4	27.9	33.1	81.0	85.9
		10	6.6	41.3	49.6	87.4	91.6
		15	7.0	41.6	48.9	89.5	93.4
		20	7.2	48.5	56.2	88.8	92.5
	200	5	6.2	20.5	25.0	83.4	87.7
		10	7.0	33.1	40.6	91.4	95.1
		15	7.2	31.3	37.2	88.2	92.1
		20	7.5	38.9	46.2	90.3	93.9

Table 3
Robustness Measures (%) of Simulated Panel Data Sets for Block Concentrated Leverage Contamination

Contamination Level	N	T	WG Mean Centering	RWGM Median Centering	RWMM Median Centering	RWGM MM-Centering	RWMM MM-Centering
Block Leverage 5%	25	5	3.0	53.2	64.8	82.7	90.8
		10	2.1	72.5	84.4	90.9	96.7
		15	1.7	72.1	82.3	89.8	95.0
		20	1.4	80.7	89.6	90.6	95.8
	50	5	2.2	42.9	53.6	83.1	89.4
		10	1.5	61.2	74.5	87.2	93.6
		15	1.2	63.6	75.5	86.9	93.3
		20	1.0	76.5	87.3	90.1	95.5
	100	5	1.5	31.5	40.1	81.5	89.5
		10	1.0	50.3	63.8	90.6	95.6
		15	0.9	53.3	65.7	89.1	94.0
		20	0.7	66.7	81.7	90.7	95.9
	200	5	1.1	24.4	31.4	84.8	92.1
		10	0.7	39.5	52.5	89.7	95.6
		15	0.6	41.8	53.6	88.4	95.0
		20	0.5	53.4	68.3	90.0	95.2
Block Leverage 10%	25	5	2.7	55.5	64.3	82.9	88.9
		10	1.9	74.6	83.5	88.3	92.5
		15	1.6	79.4	86.4	88.8	93.1
		20	1.3	82.5	89.3	89.2	92.6
	50	5	2.0	47.2	56.7	83.2	89.7
		10	1.4	68.7	79.6	89.2	92.7
		15	1.1	71.6	79.8	88.4	92.7
		20	0.9	83.9	91.8	91.3	95.8
	100	5	1.4	35.7	43.7	82.0	87.5
		10	1.0	56.7	68.7	89.2	92.6
		15	0.8	60.2	71.8	87.9	92.6
		20	0.7	75.2	86.8	89.4	93.4
	200	5	1.0	26.5	33.7	83.0	87.8
		10	0.7	44.9	57.2	88.7	93.1
		15	0.6	52.5	65.8	91.7	95.3
		20	0.5	67.1	80.7	88.3	92.4

Table 4
Robustness Measures (%) of Simulated Panel Data Sets for Random Vertical Contamination

Contamination Level	N	T	WG	RWGM	RWMM	RWGM	RWMM
			Mean Centering	Median Centering		MM-Centering	
Vertical 5%	25	5	9.0	57.9	79.7	71.1	83.2
		10	8.9	75.2	89.4	81.7	88.0
		15	9.4	78.6	94.6	89.1	90.5
		20	9.3	80.3	93.3	87.0	90.8
	50	5	8.9	46.8	74.6	63.9	84.2
		10	9.5	69.9	90.8	81.9	92.6
		15	8.9	64.8	92.2	85.5	92.5
		20	9.0	72.3	92.1	85.2	90.4
	100	5	9.1	35.8	64.0	52.6	83.4
		10	8.7	52.8	80.0	70.5	88.7
		15	8.9	51.9	86.3	78.2	88.9
		20	9.5	62.5	90.8	83.2	91.1
	200	5	9.0	26.7	52.6	41.9	82.5
		10	9.4	41.5	72.3	61.3	87.5
		15	9.1	40.1	80.4	71.4	90.4
		20	9.1	49.2	83.7	74.8	87.1
Vertical 10%	25	5	6.6	56.9	75.3	69.6	86.7
		10	6.7	72.3	85.7	80.2	89.9
		15	6.8	73.6	89.5	84.0	89.8
		20	6.7	81.2	92.9	88.0	91.1
	50	5	6.5	42.2	64.7	59.2	82.9
		10	6.7	62.4	80.8	74.9	89.8
		15	6.7	60.7	84.7	78.7	88.9
		20	6.8	70.6	88.8	84.0	91.5
	100	5	7.1	35.4	59.6	52.8	85.7
		10	6.9	49.7	72.1	65.1	87.4
		15	6.2	47.1	77.3	70.7	89.2
		20	7.0	59.1	84.0	78.0	87.8
	200	5	7.1	25.5	46.0	39.8	83.4
		10	7.0	40.1	65.0	57.3	89.8
		15	6.3	37.2	69.8	62.7	87.3
		20	6.2	44.8	76.4	68.9	87.1

Table 5
Robustness Measures (%) of Simulated Panel Data Sets
for Random Leverage Contamination

Contamination Level	N	T	WG	RWGM	RWMM	RWGM	RWMM
			Mean Centering	Median Centering		MM-Centering	
Leverage 5%	25	5	3.2	56.0	67.7	83.3	90.6
		10	2.1	70.0	80.8	90.8	95.7
		15	1.6	69.1	78.3	89.4	94.5
		20	1.4	75.0	84.5	90.2	95.7
	50	5	2.1	43.3	54.0	82.4	90.6
		10	1.4	60.8	73.7	88.9	95.2
		15	1.2	60.7	71.5	91.9	95.8
		20	1.0	67.3	78.3	90.4	95.0
	100	5	1.5	33.2	42.4	83.2	90.5
		10	1.0	51.0	64.6	90.3	95.1
		15	0.8	46.7	57.1	90.5	95.9
		20	0.7	55.4	67.4	89.5	94.9
	200	5	1.1	25.2	32.5	83.3	89.6
		10	0.8	39.5	51.9	88.3	93.8
		15	0.6	36.0	45.0	88.4	93.6
		20	0.5	44.3	55.1	89.8	95.8
Leverage 10%	25	5	2.8	68.0	77.8	79.9	80.9
		10	1.9	83.3	85.1	84.6	86.7
		15	1.5	79.9	90.6	88.0	90.2
		20	1.3	82.6	90.9	88.2	89.2
	50	5	1.9	56.1	71.8	76.0	81.8
		10	1.3	77.1	83.4	85.2	87.1
		15	1.1	72.9	90.5	90.0	91.9
		20	0.9	79.1	88.8	86.7	88.5
	100	5	1.4	46.9	75.0	78.9	82.4
		10	1.0	68.9	75.7	81.4	87.7
		15	0.7	57.9	84.3	85.6	89.0
		20	0.6	67.0	87.6	85.6	87.2
	200	5	1.0	38.3	73.2	81.5	85.7
		10	0.7	59.2	69.6	77.2	86.7
		15	0.6	49.3	80.6	85.0	89.9
		20	0.4	58.6	82.7	84.1	88.7

6. NUMERICAL EXAMPLES

In this section, the RWGM and RWMM estimators will be applied to real panel data to evaluate their performances. The MM-centering performances will be compared with the median centering by considering two numerical examples taken from Greene (2007).

The first example is a set of artificial data consisting 30 observations on investment (y) and profit (x) for 3 firms over the period of 10 years as shown in Table 6. In order to study the effect of outliers, we purposely introduced three leverage points into the data by inflating the values of cases number 4, 5 and 7 (written in parentheses in Table 6). In this example, we consider a univariate regression in order to gain a better understanding. Before the investment data are ready to be regressed, the data are transformed by the centering procedures - the mean, median and MM-centering.

The classical and robust estimates are then performed accordingly to the original and modified data. Results for the classical and robust estimates under different types of centering are reported in Table 7. The standard errors of the parameter coefficients are estimated by bootstrapping method and written in parentheses. In this study, we employed fixed-x resampling method to generate the bootstraps of 1000 replications. The method is also known as bootstrapping the residuals of linear regression models and commonly discussed in the literature (Davison and Hinkley, 1997).

Table 6
Panel Data for Investment, y and Profit, x

T	***n=1***		***n=2***		***n=3***	
	Y	***X***	***Y***	***X***	***Y***	***X***
1	13.32	12.85	20.30	22.93	8.85	8.65
2	26.30	25.69	17.47	17.96	19.60	16.55
3	2.62	5.48	9.31	9.16	3.87	1.47
4	14.94	13.79(68.95)	18.01	18.73	24.19	24.91
5	15.80	15.41(77.05)	7.63	11.31	3.99	5.01
6	12.20	12.59	19.84	21.15	5.73	8.34
7	14.93	16.64(83.20)	13.76	16.13	26.68	22.70
8	29.82	26.45	10.00	11.61	11.49	8.36
9	20.32	19.64	19.51	19.55	18.49	15.44
10	4.77	5.43	18.32	17.06	20.84	17.87

From Table 7, both robust estimators; RWGM and RWMM perform very well under median and MM-robust centering for the uncontaminated data. Their standard errors remain low, indicating highly efficient estimates at normal errors. In the modified data, the leverage values introduced have greatly influenced the classical estimate. Most data are swamped, or falsely identified as outliers due to the non-robust mean centering procedure. The two leverage values have induced every other data in the first and third firms to become outliers. Thus, the classical estimates are no longer useful in the presence of the leverage points. However, robust estimations under the three types of robust centering are found to be able to provide robust results for the modified data.

Table 7
Parameter Estimates for the Original and Modified Data of Investment

	OLS	**RWGM**	**RWMM**	**RWGM**	**RWMM**
	Mean Centering	**Median Centering**		**MM-Centering**	
Original data, $\boldsymbol{\beta_1}$	1.1022 (0.0477)	1.1000 (0.0492)	1.1199 (0.0517)	1.1460 (0.0547)	1.1128 (0.0500)
Modified data, $\boldsymbol{\beta_1}$	0.1381 (0.0711)	0.8352 (0.0589)	0.8154 (0.0521)	1.1405 (0.0249)	1.1131 (0.0181)

The second example is a panel data set of six airline firms, also taken from Greene (2007). The data consists of 90 yearly observations (1970 to 1984) and originally used to study efficiency in production of airline services. A straight forward equation can be formed as an illustration where the total cost of production is fitted into a multiple linear regression model:

$$\ln cost_{it} = \alpha_i + \beta_1 \ln output + \beta_2 \ln fuel\ price + \beta_3 load\ factor + \varepsilon_{it}$$

From Greene (2007), this model measured the output in "revenue passenger miles". The load factor is defined as the average rate at which seat's on the airline's planes are filled.

The effects of outliers are again studied in this example by introducing two large values into the dataset. The results obtained for the classical Within Group, RWGM and robust RWMM estimation for the modified data are also reported in Table 8. It is observed that in the existence of outliers, both robust estimators are performing well under the robust MM-centering. Under the procedures, both RWGM and RWMM are able to produce resistant estimates with low standard errors. Similar resistant results are also obtained for the modified data where robust estimations also provide equally low standard errors.

Table 8
Parameter Estimates of the Cost Equation with Fixed Effects

		OLS	**RWGM**	**RWMM**	**RWGM**	**RWMM**
		Mean Centering	**Median Centering**		**MM-Centering**	
Original Data	β_1	0.9193 (0.0229)	0.8688 (0.0312)	0.8604 (0.0266)	0.9161 (0.0309)	0.9165 (0.0277)
	β_2	0.4175 (0.0061)	0.4167 (0.0148)	0.4294 (0.0147)	0.4103 (0.0146)	0.4108 (0.0140)
	β_3	-1.0700 (0.1914)	-0.6726 (0.2004)	-0.9501 (0.1939)	-1.0854 (0.1972)	-1.0766 (0.1836)
Modified Data	β_1	-7.0641 (3.0828)	0.8649 (0.0276)	0.8612 (0.0248)	0.9187 (0.0266)	0.9175 (0.0231)
	β_2	7.4311 (0.7578)	0.4172 (0.0122)	0.4291 (0.0098)	0.4066 (0.0110)	0.4104 (0.0091)
	β_3	-21.5600 (24.0655)	-0.7030 (0.2062)	-0.9413 (0.1958)	-1.0491 (0.1930)	-1.0796 (0.1810)

7. CONCLUSION

This study is aimed to develop two new centering procedures for robust estimation of the fixed panel model. The MM-centering are developed to correct the non-linearity caused by data transformation using the median. We have applied the RWGM and RWMM under three different types of robust centering. Simulation study indicates that data transformation under MM-centering provides more stable and superior results than transformation by the median.

The performances of robust estimations under the newly proposed procedures have also improved vastly in small data sets, with small number of time series. Overall results showed that the performances of RWMM are more superior than RWGM under different types of contamination levels, sample size and number of time series.

8. REFERENCES

1. Arellano, M. (1987). Computing Robust Standard Errors for Within-Groups Estimators. *Oxford Bulletin of Economics and Statistics*, 49(4), 431-434.
2. Bagheri, A. and Habshah, M. (2009). Robust Estimations as a Remedy for Multicollinearity Caused by Multiple High Leverage Points. *J. Math. and Statist.*, 5(4), 311-321.
3. Baltagi, B.H. (2005). *The Econometrics of Panel Data*. John Wiley & Sons, New York.
4. Bramati, M.C. and Croux, C. (2007). Robust Estimators for the Fixed Effects Panel Data Model. *Econometrics Journal*, 10(3), 521-540.
5. Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press.
6. Donoho, D.L. and Huber, P.J. (1983). The Notion of Breakdown Point, In: *A Festschrift for Erich L. Lehmann*, 157-184.
7. Hampel, F.R. (1971). A General Qualitative Definition of Robustness. *The Annals of Mathematical Statistics*, 42(6), 1887-1896.
8. Hampel, F.R. (2001). Robust Statistics: A Brief Introduction and Overview. *Research Report No. 94*, Seminar für Statistik, Eidgenossische Technische Hochschule (ETH), Zürich.
9. He, X. and Portnoy, S. (1992). Reweighted LS Estimators Converge at the Same Rate as the Initial Estimator. *The Annals of Statist.*, 20(4), 2161-2167.
10. Greene, W.H. (2007). *Econometric Analysis*, 6th Edition, Prentice Hall.
11. Kezdi, G. (2004). Robust Standard Error Estimation in Fixed-Effects Panel Models. *Hungarian Statistical Review*, 9, 95-116.
12. Lucas, A., Van Dijk, R. and Kloek, T. (2007). *Outlier Robust GMM Estimation of Leverage Determinants in Linear Dynamic Panel Data Models*. Unpublished manuscript, Vrije Universiteit, Amsterdam, the Netherlands.
13. Maronna, R.A., Martin, R.D. and Yohai, V.J. (2006). *Robust Statistics: Theory and Methods*, John Wiley, New York, ISBN: 10: 0470010924, 436.
14. Petersen, M.A. (2005). Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. *The Review of Financial Studies*, 22(1), 435-480.
15. Preminger, A. and Franck, R. (2007). Forecasting Exchange Rates: A Robust Regression Approach. *International Journal of Forecasting*, 23(1), 71-84.

16. Riazoshams, H., Habshah, M. and Sharipov, O. (2010). The Performance of Robust Two-Stage Estimator in Nonlinear Regression with Autocorrelated Error. *Commun. in Statist. – Simul. and Compu.*, 39(6), 1251-1268.
17. Rousseeuw, P. and Leroy, A.M. (2003). *Robust Regression and Outlier Detection*. John Wiley, New York.
18. Rousseeuw, P. and Zomeren, B.C.V. (1990). Unmasking Multivariate Outliers and Leverage Points. *J. Amer. Statist. Assoc.*, 85(411), 633-639.
19. Rousseeuw, P., Daniels, B. and Leroy, A. (1984). Applying Robust Regression to Insurance. *Insurance: Mathematics and Economics*, 3(1), 67-72.
20. Ruppert, D. (1992). Computing S Estimators for Regression and Multivariate Location/Dispersion. *Journal of Computational and Graphical Statistics*, 1, 253-270.
21. Salibian-Barrera, M. and Yohai, V.J. (2006). A Fast Algorithm for S-regression Estimates. *Journal of Computational and Graphical Statistics*, 15(2), 414-427.
22. Simpson, J.R. and Montgomery, D.C. (1998). The Development and Evaluation of Alternative Generalized M-Estimation Techniques. *Commun. in Statist. – Simul. and Compu.*, 27(4), 999-1018.
23. Stock, J.H. and Watson, M.W. (2006). Heteroskedasticity-Robust Standard Errors for Fixed Effects Panel Data Regression. *NBER Working Paper Series*, Vol. t0323. Available at SSRN: http://ssrn.com/abstract=912424
24. Verardi, V. and Wagner, J. (2011). Robust Estimation of Linear Fixed Effects Panel Data Models with an Application to the Exporter Productivity Premium. *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*, 231(4), 546-557.
25. Wagenvoort, R. and Waldmann, R. (2002). On B-robust Instrumental Variable Estimation of the Linear Model with Panel Data. *Journal of Econometrics*, 106, 297-324.
26. Yohai, V.J. (1987). High Breakdown-Point and High Efficiency Estimates for Regression. *The Annals of Statistics*, 15, 642-65.
27. Zaman, A., Rousseeuw, P.J. and Orhan, M. (2001). Econometric Applications of High-Breakdown Robust Regression Techniques. *Economics Letters*, 71(1), 1-8.