Journal of Engineering and Applied Sciences 13 (17): 7156-7161, 2018
ISSN: 1816-949X

Robust Estimation for Fixed and Random Effects Panel Data Models with Different Centering Methods

[1]Habshah Midi and [2]Sani Muhammad
[1]Faculty of Science and Institute for Mathematical Research
Universiti Putra Malaysia, 43400, UPM Serdang, Selangor, Malaysia
[2]Department of Mathematical Sciences, Federal University Dutsinma, Katsina State, Nigeria

Abstract: In the presence of outlying observations in panel data set, the traditional ordinary least square estimator can be strongly biased, lead to erroneous estimation and misleading inferential statement. However, Weighted Least Squares (WLS) are usually used to remedy the effect of outliers. Visek used Least Weighted Squares (LWS) based on mean-centering technique for data transformation. The mean-centering was found to be very sensitive to outliers. Furthermore, robust method for data transformation is needed in order to down weight the effect of outliers. We employed a new method of transformation based on MM-estimate of location termed MM-Centering method. A simulation study was used to evaluate the performance the proposed method. The Weighted Least Square based on the proposed MM-centering Method (WLSMM) was found to be the best method for both the high leverage points and vertical outliers.

Key words: Centering method, fixed and random effect model, outlier, ordinary least square, weighted least square, panel data

INTRODUCTION

The outlying observation in panel data set cause a misleading estimation prediction, these outliers occur due to typing error, measurement error, unusual values, transmission or copying error (Rousseeuw and Leroy, 2003a, b; Maronna *et al.,* 2006; Bakar and Midi, 2015). Recently, the used of panel data in finance and economics research has been increased due to its importance and advantage of possessing two dimensionalities of cross-section and time series.

Generally, outliers are of different types such as vertical outlier (outlier in Y-direction, horizontal outlier (outlier in X-direction) or leverage (Rousseeuw and Zomeren, 1990; Bramati and Croux, 2007). Moreover, Bramati and Croux (2007) shows that outliers are usually found randomly in a data set and sometimes are found to be concentrated in some few time series (block concentrated outliers). The robustness of an estimator means an estimator is resistance to a small change or modification cause by outlier in the dataset, it can be measure based on breakdown point (Hampel *et al.*, 1986; Bramati and Croux, 2007).

In recent years, many researchers developed robust estimators in many areas, especially the linear regression. Unfortunately, very few researches are available regarding the robust estimation method in panel data.

Robust estimators have some highly desirable properties of high breakdown point, high efficiency and bounded influence (Bakar and Midi, 2015).

The basic models used for the analysis of panel data are Fixed Effects (FE) and the Random Effects (RE) models. In literature some researchers proposed robust method for parameter estimation in panel data such as Bramati and Croux (2007), Verardi and Wagner (2011) and Bakar and Midi (2015).

Recently, Visek (2015) make use of Least Weighted Squares (LWS) to estimate the model with fixed and random effects using residual order statistic but he employed classical centering method (mean centering) which suffer much setback in the presence of outliers.

This motivated us to propose a new estimation technique based on MM-centering Method, so as to reduce the effect of outliers in the data. Two existing centering methods were compared with the proposed method in LWS for both fixed and random effect panel data models.

MATERIALS AND METHODS

A panel data model consists of measurements taken from many individuals over time, cross sectional and time series dimension (Crowder and Hand, 1990). Panel data sets allow to analyze a number of questions that could not

Corresponding Author: Midi Habshah, Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, Serdang, Malaysia

be possible to address using only cross-section or only time-series data sets. The panel data model is given by:

$$y_{it} = u_i + x'_{it}\beta + e_{it} \quad (1)$$
$$i = 1, 2, ..., n \text{ and } t = 1, 2, ..., T$$

Where:

- y_{it} = The dependent (response) variables
- x_{it} = The kth independent (explanatory) variables
- u_i = The unobserved time-invariant effects and
- e_{it} = The error term (disturbances) that is assumed to be normal, uncorrelated across individual units and time, i.e., assumed strict no endogeneity is applied (Wagenvoort and Waldmann, 2002)

The pooled OLS Model specifies constant coefficients in Eq. 1, i.e., the usual assumption for cross sectional analysis, fixed effect model is when there is correlation between x_{it} and u_i, i.e., cov(x_{it}, u_i) ≠ 0 and u_i is included in the intercept while for the random effects model the cov (x_{it}, u_i) = 0 and u_i is included in the error term for all i = 1, 2, ..., n, t = 1, 2, ..., T, i.e, $\varepsilon_{it} = u_i + e_{it}$ $E(\varepsilon_{it}) = 0$, $E(\varepsilon_{it}^2) = \sigma^2_u + \sigma^2_e$ and $E(\varepsilon_{it}, \varepsilon_{is}) = var(u_i) = \sigma^2_u$ for all t ≠ s (Baltagi, 2001).

Data transformation within the time series base on mean (mean centering) is not efficient in the presence of outlying observations. The MM-estimate of location is proposed which is more efficient and provide a high breakdown point. The median centering method proposed by Bramati and Croux (2007) was found to produce nonlinearity to the resulting transformed data which caused the estimators to lose their equivariance properties. Mean centering involve transforming Eq. 1 as follows:

$$\tilde{y}_{it} = \tilde{x}_{it}\beta + \varepsilon_{it} \quad (2)$$

where: $\tilde{y}_{it} = y_{it} - \bar{y}_i$ and $\tilde{x}_{it} = x_{it} - \bar{x}_i$ with:

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it},\ \bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$$

and the unobserved time-invariant effect are eliminated. The median centering employed median instead of mean to each within time series as $\tilde{y}_{it} = y_{it}$ median$_t$ $\{y_{it}\}$ and $\tilde{x}_{it} = x_{it}$ median$_t$ $\{x_{it}\}$. The MM-centering uses MM-estimate of location which can be obtain by first computing S-estimates of covariance and location as the initial scale estimate $\hat{\sigma}_n$:

$$\hat{\sigma}_n = \min_t s_n(t) \quad (3)$$

And the corresponding location S-estimate $\hat{\mu}_n$ is given as:

$$\hat{\mu}_n = \arg_t \min s_n(t) \quad (4)$$

Using the loss function of Tukey's bisquare weight with ρ function given by:

$$\rho(x) = \begin{cases} \frac{1}{6c^4}x^6 - \frac{1}{2c^2}x^4 + \frac{1}{2}x^2, \text{ if } |x| \le c \\ \frac{c^2}{6}, |x| > c \end{cases} \quad (5)$$

By considering the location and shape estimates we re-estimate M-estimate which gives 95% efficiency at central model and c is the tuning constant (Ruppert, 1992). The transformed data was given in Eq. 2 where $\tilde{y}_{it} = y_{it} - \hat{\mu}_n\{y_{it}\}$ and $\tilde{x}_{it} = x_{it} - \hat{\mu}_n\{x_{it}\}$ for i = 1, 2, ..., n and t = 1, 2, ..., T.

The estimation based on Visek (2015) technique involves classical mean centering. The residual of the (i, t)-th observations from Eq. 1 is given by:

$$r_{it}(\beta) = y_{it} - x'_{it}.\beta \quad (6)$$

Let denote the q-th squared residuals order statistic by $r^2_{(q)}(\beta)$ where q = 1, 2, …, n. T, So that:

$$r^2_{(1)}(\beta) \le r^2_{(2)}(\beta) \le, ..., \le r^2_{(n.T)}(\beta) \quad (7)$$

Now, instead of minimizing the residual sum of squares, we minimize the weighted sum of squares residuals:

$$WSS(\beta, \vec{w}) = \sum_{q=1}^{nT} w_q \left(y_q - x_q\beta\right)^2 \quad (8)$$

where the weight (wq) is defined as $w_q \in [0, 1]$ for q = 1, 2, …, n.T. The Weighted Least Square (WLS) estimator is given by:

$$\hat{\beta}(WLS) = \arg\min_\beta \sum_{q=1}^{nT} w_q . r_q^2(\beta) \quad (9)$$

The absolute residuals ($r_{(it)}(\beta)$'s) value distribution function denoted by $F^n_\beta(r)$, its derivation shows that $\hat{\beta}(WLS)$ is among the solutions of the normal equations (Visek, 2011):

$$\sum_{i=1}^{n}\sum_{t=1}^{T} w\left(F_\beta^{nT}\left(\left|r_{it}(\beta)\right|\right)\right)x_{it}\left(y_{it} - x'_{it}\beta\right) = 0 \quad (10)$$

It has been prove that is consistent under the following assumptions.

Assumption 1: The sequence of (x_{it}, e_{it}) for i = 1, 2, …, ∞ t = 1, 2, …, T is independent and identically distributed (p+1) dimensional random variables.

Assumption 2: Weight is generated as a monotone function w i.e., $w_q = w(q\text{-}1/n.T)$, $w_1 \geq w_2 \geq \cdots \geq w_{nT}$ with $w(0) = 1$.

Assumption 3: Taking $F^t_\beta(r) = P(|r_{it}\ (\beta)| \leq r)$. The proof is the direct reformulation of the Visek (2011) for the cross-sectional data framework.

For the Fixed Effect (FE) estimation, the above WLS procedure was applied to the transformed data in Eq. 2 to obtain the fixed weighted effect estimator $(\hat{\beta}(\text{FWE}))$. For the Random Effect (RE) estimation, the transformed data from Eq. 2 will be slightly modified by introducing known as partially demeaned transformation given as:

$$y_{it}\text{-}\theta\bar{y}_{it} = (x_{it}\text{-}\theta\bar{x}_{it})\beta+\varepsilon_{it} \tag{11}$$

where $\theta = 1\text{-}[\sigma^2_{s\cdot}\ (\sigma^2_s+T.\sigma^2_u)^{-1}]$ and the variances σ^2_s and σ^2_s can be estimated by employing the residuals of the OLS estimator $r_{it}(\hat{\beta}(\text{OLS}))$ to obtained $\hat{\sigma}^2_\varepsilon$ and $\hat{\sigma}^2_u$, so that, $\hat{\sigma}^2_s = \hat{\sigma}^2_\varepsilon\text{-}\hat{\sigma}^2_u$. Replace θ with $\hat{\theta}$ and apply WLS method to the partially demeaned transformed data to obtain random weighted effect estimator $(\hat{\beta}(\text{RWE}))$. In general applying OLS to the transformed data gives the classical FE and RE estimates, respectively.

RESULTS AND DISCUSSION

This study focuses on WLS based on three different centering methods (mean, median and MM-centering) for fixed and random effect panel data models as well as pooled OLS Model. We employed, Visek (2015) simulation technique with R = 1000 number of replications to a specified model:

$$\begin{gathered} y_{it} = 2+1.x_{it}+2.x_{it}\text{-}3.x_{it}+4.x_{it}+u_i+e_{it} \\ \text{for } i = 1, 2, \ldots, 30\ \ t = 1, 2, \ldots, 20 \text{ and } k = 1, 2, \ldots, 5 \end{gathered} \tag{12}$$

The explanatory variables x_{it} and idiosyncratic error e_{it} are generated from standard normal distribution where the unobserved time-invariant variables u_i are generated from uniform $u_i \sim U(0, 15)$ based on the assumption of fixed and random effect model mentioned above for fixed effect model the unobserved time invariant effect were added to the explanatory variables.

Two types of data contamination were employed, firstly, random contamination over all observation (random contamination) secondly, concentrating the contamination in a few times series (block concentrated contamination). The contaminations are done at both Y-direction (vertical outlier) and X-direction (leverage point). These make four different types of contaminations, namely: random vertical outliers, random leverage, block concentrated vertical outliers and block concentrated leverage at two different contamination levels 15 and 25%.

Contaminations were created by adding to the Y's originally computed a term ~N(20, 1) for both random and block concentrated Vertical outliers andsimilarly, replacing the original generated value by a term ~N(10, 1) for both random and block concentrated leverage points. For each of the Replication (R) the coefficient β was estimated for all the three methods WLS, FEW and RWE for the three centering methods mentions. The empirical means and variances of the estimates in this simulation with R = 1000 number of replications over the kth index was computed using R package as:

$$\hat{\beta}_k^{(\text{index})} = \frac{1}{R}\sum_k^R \hat{\beta}_k^{(\text{index},k)}\ \text{var}\left(\hat{\beta}_k^{(\text{index})}\right) = \frac{1}{R}\sum_k^R \left[\hat{\beta}_k^{(\text{index},k)}\right]^2\text{-}\left[\hat{\beta}_k^{(\text{index})}\right]^2 \tag{13}$$

The performance of the methods was assessed based on the bias and variance of the estimate. The estimate with same value of the true coefficient and smallest variance is the best and the most efficient. Table 1-5 present the performance of the three methods at different contamination level.

Table 1 shows that robust methods employed have equivalent performance with classical method in an uncontaminated data (clean data). Although, classical estimator performed poorly in the presence of outliers in a data set for all the different contamination level considered. Table 2-5 shows that the MM-centering Method yield good and efficient result for both WLS, FWE and RWE by providing an estimate very close to the true parameter (smaller bias) and the smallest variance. For the FE Model (when there is correlation between unobserved effect and explanatory variables) FWE estimator produces the best result but as in the transformation method the unobserved effect is eliminated this clearly shows that the intercept of FEW we will be zero. Although, for the RE model (when there is no correlation between unobserved effect and explanatory variables) RWE estimator is the best. Moreover, for the random and block contamination almost all the methods produce the same result which indicates that employing the order squared residual in obtaining weight has played a significant roll. In this study only result for random contamination was presented.

Table 1: Parameter estimate of simulated panel data without contamination for WLS, FWE and RWE

Uncontaminated (clean) data

	Mean centering			Median centering			MM centering		
True coefficient	β (WLS) (var(β))	β (FWE) (var(β))	β (RWE) (var(β))	β (WLS) (var(β))	β (FWE) (var(β))	β (RWE) (var(β))	β (WLS) (var(β))	β (FWE) (var(β))	β (RWE) (var(β))
$\beta_0 = 2$	2.00	0.00	1.99	2.00	0.00	1.99	2.00	0.00	1.99
	(0.002)	(0.000)	(0.033)	(0.003)	(0.000)	(0.031)	(0.002)	(0.000)	(0.042)
$\beta_1 = 1$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$\beta_2 = 2$	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$\beta_3 = -3$	-3.00	-3.00	-3.00	-3.00	-3.00	-3.00	-3.00	-3.00	-3.00
	(0.001)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
$\beta_4 = 4$	4.00	4.00	4.00	4.00	3.99	4.00	4.00	4.00	4.00
	(0.001)	(0.002)	(0.002)	(0.002)	(0.005)	(0.002)	(0.002)	(0.003)	(0.002)

Table 2: Parameter estimate of simulated panel data when there is no correlation between unobserved effect and explanatory variables with random vertical contamination for WLS, FWE and RWE

	Mean centering			Median centering			MM centering		
True coefficient	β(WLS) (var(β))	β(FWE) (var(β))	β(RWE) (var(β))	β(WLS) (var(β))	β(FWE) (var(β))	β(RWE) (var(β))	β(WLS) (var(β))	β(FWE) (var(β))	β(RWE) (var (β))
15% contamination level									
$\beta_0 = 2$	1.96	0.00	1.98	1.97	0.00	1.97	1.99	0.00	1.99
	(0.021)	(0.000)	(0.018)	(0.023)	(0.000)	(0.016)	(0.007)	(0.000)	(0.005)
$\beta_1 = 1$	0.96	0.95	0.97	0.96	0.97	0.98	1.00	1.00	1.00
	(0.023)	(0.026)	(0.015)	(0.017)	(0.014)	(0.015)	(0.008)	(0.007)	(0.004)
$\beta_2 = 2$	1.95	1.97	1.99	1.95	2.00	1.95	1.98	1.96	2.00
	(0.025)	(0.019)	(0.014)	(0.018)	(0.014)	(0.016)	(0.007)	(0.007)	(0.004)
$\beta_3 = -3$	-2.96	-3.00	-3.00	-2.96	-2.96	-2.97	-2.97	-2.98	-3.00
	(0.027)	(0.012)	(0.009)	(0.014)	(0.015)	(0.014)	(0.009)	(0.006)	(0.003)
$\beta_4 = 4$	3.94	3.97	3.98	3.95	3.97	3.96	3.99	4.00	4.00
	(0.027)	(0.018)	(0.012)	(0.006)	(0.015)	(0.013)	(0.007)	(0.006)	(0.004)
25% contamination level									
$\beta_0 = 2$	1.85	0.00	2.03	0.00	1.87	1.88	1.96	0.00	1.98
	(0.165)	(0.000)	(0.121)	(0.123)	(0.000)	(0.135)	(0.015)	(0.000)	(0.012)
$\beta_1 = 1$	0.79	0.67	0.85	0.86	0.79	0.86	0.97	0.95	1.01
	(0.141)	(0.172)	(0.093)	(0.107)	(0.121)	(0.124)	(0.012)	(0.016)	(0.008)
$\beta_2 = 2$	1.81	1.75	1.79	1.85	1.83	1.88	1.98	1.84	1.99
	(0.153)	(0.164)	(0.081)	(0.113)	(0.116)	(0.124)	(0.012)	(0.016)	(0.007)
$\beta_3 = -3$	-2.63	-2.78	-2.68	-2.79	-2.70	-2.74	-2.97	-2.93	-3.00
	(0.144)	(0.171)	(0.079)	(0.122)	(0.109)	(0.123)	(0.012)	(0.016)	(0.007)
$\beta_4 = 4$	3.72	3.69	3.72	3.67	3.67	3.67	3.97	3.91	4.00
	(0.141)	(0.166)	(0.093)	(0.152)	(0.109)	(0.124)	(0.013)	(0.018)	(0.006)

Table 3: Parameter estimate of simulated panel data when there is correlation between unobserved effect and explanatory variables with random vertical contamination for WLS, FWE and RWE

	Mean centering			Median centering			MM centering		
True coefficient	β(WLS) (var(β))	β(FWE) (var(β))	β(RWE) (var(β))	β(WLS) (var(β))	β(FWE) (var(β))	β(RWE) (var(β))	β(WLS) (var(β))	β(FWE) (var(β))	β RWE) (var(β))
15% contamination level									
$\beta_0 = 2$	1.92	0.00	1.93	1.95	0.00	1.94	1.97	0.00	1.95
	(0.017)	(0.000)	(0.012)	(0.015)	(0.000)	(0.014)	(0.006)	(0.000)	(0.009)
$\beta_1 = 1$	0.94	0.96	0.93	0.94	0.99	0.95	0.98	1.00	0.98
	(0.012)	(0.008)	(0.009)	(0.0012)	(0.007)	(0.007)	(0.005)	(0.003)	(0.006)
$\beta_2 = 2$	1.95	1.97	1.95	1.96	1.98	1.96	1.98	2.00	2.00
	(0.012)	(0.008)	(0.009)	(0.012)	(0.008)	(0.009)	(0.006)	(0.002)	(0.006)
$\beta_3 = -3$	-2.93	-2.96	-2.96	-2.95	-2.97	-2.97	-2.99	-3.00	-2.99
	(0.011)	(0.009)	(0.010)	(0.020)	(0.009)	(0.009)	(0.005)	(0.003)	(0.005)
$\beta_4 = 4$	3.95	3.97	3.95	3.96	3.98	3.96	3.98	3.99	4.02
	(0.010)	(0.009)	(0.013)	(0.016)	(0.009)	(0.010)	(0.006)	(0.003)	(0.009)
25% cntamination level									
$\beta_0 = 2$	1.83	0.00	2.04	1.86	0.00	2.03	1.92	0.00	2.02
	(0.103)	(0.000)	(0.142)	(0.104)	(0.000)	(0.014)	(0.009)	(0.000)	(0.012)
$\beta_1 = 1$	0.79	0.82	0.83	0.90	0.92	0.90	0.91	1.00	0.98
	(0.131)	(0.098)	(0.097)	(0.103)	(0.100)	(0.086)	(0.008)	(0.004)	(0.007)

Table 3: Continue

	Mean centering			Median centering			MM centering		
True coefficient	β(WLS) (var(β))	β(FWE) (var(β))	β(RWE) (var(β))	β(WLS) (var(β))	β(FWE) (var(β))	β(RWE) (var(β))	β(WLS) (var(β))	β(FWE) (var(β))	β(RWE) (var(β))
$\beta_2 = 2$	1.84	1.87	1.86	1.88	1.89	1.89	1.93	1.98	2.00
	(0.108)	(0.105)	(0.107)	(0.102)	(0.093)	(0.086)	(0.009)	(0.004)	(0.008)
$\beta_3 = -3$	-2.86	-2.87	-2.84	-2.91	-2.90	-2.90	-2.92	-3.00	-2.99
	(0.106)	(0.096)	(0.106)	(0.103)	(0.093)	(0.087)	(0.009)	(0.004)	(0.008)
$\beta_4 = 4$	3.89	3.81	3.79	3.89	3.92	3.90	3.93	3.99	4.01
	(0.109)	(0.098)	(0.110)	(0.105)	(0.094)	(0.086)	(0.009)	(0.005)	(0.009)

Table 4: Parameter estimate of simulated panel data when there is no correlation between unobserved effect and explanatory variables with random leverage contamination for WLS, FWE and RWE

	Mean centering			Median centering			MM centering		
True coefficient	β(WLS) (var(β))	β(FWE) (var(β))	β(RWE) (var(β))	β(WLS) (var(β))	β(FWE) (var(β))	β(RWE) (var(β))	β(WLS) (var(β))	β(FWE) (var(β))	β RWE) (var(β))
15% contamination level									
$\beta_0 = 2$	1.78	0.00	1.89	1.76	0.00	1.90	1.86	0.00	1.95
	(0.103)	(0.000)	(0.100)	(0.132)	(0.000)	(0.020)	(0.067)	(0.000)	(0.022)
$\beta_1 = 1$	0.84	0.81	0.87	0.82	0.80	0.87	0.91	0.88	0.97
	(0.126)	(0.120)	(0.105)	(0.122)	(0.121)	(0.105)	(0.067)	(0.083)	(0.020)
$\beta_2 = 2$	1.79	1.77	1.82	1.80	1.75	1.84	1.89	1.87	1.96
	(0.110)	(0.121)	(0.105)	(0.122)	(0.121)	(0.104)	(0.065)	(0.086)	(0.020)
$\beta_3 = -3$	-2.82	-2.80	-2.87	-2.80	-2.79	-2.89	-2.93	-2.88	-2.97
	(0.102)	(0.100)	(0.108)	(0.120)	(0.122)	(0.104)	(0.065)	(0.087)	(0.021)
$\beta_4 = 4$	3.85	3.82	3.90	3.83	3.84	3.91	3.91	4.90	3.98
	(0.110)	(0.109)	(0.019)	(0.121)	(0.121)	(0.105)	(0.067)	(0.083)	(0.020)
25% contamination level									
$\beta_0 = 2$	1.71	0.00	1.75	1.72	0.00	1.78	1.83	0.00	1.93
	(0.682)	(0.000)	(0.744)	(0.656)	(0.000)	(0.545)	(0.102)	(0.000)	(0.042)
$\beta_1 = 1$	0.80	0.76	0.81	0.78	0.74	0.87	0.87	0.84	0.96
	(0.713)	(0.870)	(0.665)	(0.587)	(0.645)	(0.534)	(0.091)	(0.103)	(0.035)
$\beta_2 = 2$	1.72	1.71	1.70	1.74	1.71	1.80	1.85	1.83	1.93
	(0.710)	(0.856)	(0.665)	(0.586)	(0.645)	(0.532)	(0.095)	(0.106)	(0.033)
$\beta_3 = -3$	-2.76	-2.74	-2.80	-2.76	-2.74	-2.83	-2.90	-2.82	-2.94
	(0.713)	(0.876)	(0.666)	(0.587)	(0.586)	(0.534)	(0.093)	(0.103)	(0.034)
$\beta_4 = 4$	3.81	3.71	3.79	3.79	3.77	3.85	3.88	4.85	3.95
	(0.714)	(0.879)	(0.669)	(0.587)	(0.591)	(0.532)	(0.093)	(0.103)	(0.034)

Table 5: Parameter estimate of simulated panel data when there is correlation between unobserved effect and explanatory variables with random leverage contamination for WLS, FWE and RWE

	Mean centering			Median centering			MM centering		
True coefficient	β(WLS) (var(β))	β(FWE) (var(β))	β(RWE) (var(β))	β(WLS) (var(β))	β(FWE) (var(β))	β(RWE) (var(β))	β(WLS) (var(β))	β(FWE) (var(β))	β(RWE) (var(β))
15% contamination level									
$\beta_0 = 2$	1.75	0.00	1.87	1.76	0.00	1.84	1.85	0.00	1.90
	(0.112)	(0.000)	(0.107)	(0.132)	(0.000)	(0.107)	(0.064)	(0.000)	(0.062)
$\beta_1 = 1$	0.84	0.88	0.87	0.82	0.86	0.83	0.91	0.96	0.89
	(0.128)	(0.101)	(0.105)	(0.122)	(0.112)	(0.121)	(0.065)	(0.027)	(0.054)
$\beta_2 = 2$	1.79	1.86	1.82	1.80	1.85	1.81	1.89	1.97	1.91
	(0.110)	(0.101)	(0.105)	(0.122)	(0.109)	(0.117)	(0.065)	(0.026)	(0.053)
$\beta_3 = -3$	-2.83	-2.89	-2.87	-2.80	-2.90	-2.82	-2.93	-2.98	-2.92
	(0.108)	(0.102)	(0.108)	(0.120)	(0.109)	(0.115)	(0.065)	(0.025)	(0.053)
$\beta_4 = 4$	3.84	3.90	3.92	3.83	3.92	3.87	3.90	4.97	3.90
	(0.108)	(0.102)	(0.109)	(0.121)	(0.105)	(0.116)	(0.067)	(0.025)	(0.062)
25% contamination level									
$\beta_0 = 2$	1.73	0.00	1.74	1.73	0.00	1.75	1.83	0.00	1.84
	(0.694)	(0.000)	(0.644)	(0.673)	(0.000)	(0.645)	(0.102)	(0.000)	(0.103)
$\beta_1 = 1$	0.81	0.83	0.78	0.77	0.86	0.82	0.87	0.97	0.91
	(0.721)	(0.530)	(0.654)	(0.687)	(0.555)	(0.678)	(0.091)	(0.033)	(0.094)
$\beta_2 = 2$	1.72	1.71	1.70	1.74	1.79	1.75	1.85	1.93	1.83
	(0.710)	(0.532)	(0.655)	(0.586)	(0.555)	(0.672)	(0.095)	(0.034)	(0.103)
$\beta_3 = -3$	-2.75	-2.82	-2.82	-2.75	-2.84	-2.78	-2.90	-2.95	-2.87
	(0.722)	(0.533)	(0.656)	(0.686)	(0.554)	(0.678)	(0.093)	(0.034)	(0.103)
$\beta_4 = 4$	3.80	3.81	3.76	3.78	3.86	3.80	3.88	3.96	3.86
	(0.718)	(0.541)	(0.659)	(0.686)	(0.554)	(0.678)	(0.093)	(0.033)	(0.121)

The performance of median-centering is not much significant due to the non-linearity causes by its transformation technique (Maronna *et al.,* 2006). Although, mean-centering did not resist to the effect of outliers due its low power and 0% breakdown point.

CONCLUSION

A very low level of contamination by means of outliers or leverage points in panel data set causes bias in ordinary least square estimation. This study focus on robust estimation for both fixed and random effect models using WLS with different centering methods. It is very necessary for econometricians to know the effect of outlying observations and to also know the importance of robust methods. The simulation study in this paper shows that the new proposed $\hat{\beta}(\text{RWE})$ and $\hat{\beta}(\text{FWE})$ based on MM-centering method is the most highly robust and efficient as it provides the smallest bias and smallest variance of the estimate at different contamination levels.

RECOMMENDATIONS

To improve the robust method of estimation based on WLS in panel data, the robust technique used for the identification of outliers is needed to first identify the outlying observations and then apply a robust weighting function such as Huber, Hampel and Turkey bisquare to down weight the effect of outlier.

REFERENCES

Bakar, N.M.A. and H. Midi, 2015. Robust centering in the fixed effect panel data model. Pak. J. Stat., 31: 33-48.

Baltagi, B.H., 2001. Econometric Analysis of Panel Data. 2nd Edn., John Wiley and Sons, New York, ISBN: 0471499374.

Bramati, M.C. and C. Croux, 2007. Robust estimators for the fixed effects panel data model. Econom. J., 10: 521-540.

Crowder, M.J. and D.J. Hand, 1990. Analysis of Repeated Measures. Chapman & Hall, London, England, UK., ISBN:9780412318306, Pages: 259.

Hampel, F.R., P.J. Rousseeuw, E. Ronchetti and W.A. Stahel, 1986. Robust Statistics: The Approach Based on Influence Functions. 1st Edn., Wiley, New York, Pages: 536.

Maronna, R.A., D.R. Martin and V.J. Yohai, 2006. Robust Statistics: Theory and Methods. John Wiley & Sons, Hoboken, New Jersey, USA., ISBN:9780470010921, Pages: 436.

Rousseeuw, P.J. and A.M. Leroy, 2003a. Robust Regression and Outlier Detection. John Wiley & Sons, Hoboken, New Jersey, USA., ISBN:9780471488552, Pages: 335.

Rousseeuw, P.J. and A.M. Leroy, 2003b. Robust Regression and Outlier Detection. 1st Edn., John Wiley, New York, ISBN: 978-0471488552, pp: 360.

Rousseeuw, P.J. and B.C.V. Zomeren, 1990. Unmasking multivariate outliers and leverage points. J. Am. Stat. Assoc., 85: 633-639.

Ruppert, D., 1992. Computing S estimators for regression and multivariate location/dispersion. J. Comput. Gr. Stat., 1: 253-270.

Verardi, V. and J. Wagner, 2011. Robust estimation of linear fixed effects panel data models with an application to the exporter productivity premium. J. Econ. Stat., 231: 546-557.

Visek, J.A., 2011. Consistency of the least weighted squares under heteroscedasticity. Kybernetika, 47: 179-206.

Visek, J.A., 2015. Estimating the model with fixed and random effects by a robust method. Method. Comput. Appl. Probab., 17: 999-1014.

Wagenvoort, R. and R. Waldmann, 2002. On B-robust instrumental variable estimation of the linear model with panel data. J. Econom., 106: 297-324.