Pak. J. Statist.
2018 Vol. 34(4), 297-310

ROBUST WITHIN GROUP ESTIMATOR FOR FIXED EFFECT PANEL DATA

Shelan Saied Ismaeel[1,2] and **Habshah Midi**[1§]

[1] Faculty of Science and Institute for Mathematical Research,
Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

[2] Department of Mathematics, Faculty of Science
University of Zakho, Iraq

[§] Corresponding author Email: habshah@upm.edu.my

ABSTRACT

In the presence of outliers, panel data estimators can be extremely biased. In this work we used mm-centering to provide robust solutions to the Within Group parameter estimates. The main contribution of this article is to propose a new version of the Generalized M-estimator (GM) that provides good resistance against bad leverage points. The advantage of this method over the existing methods is that it only minimizes the bad leverage points and outliers. The good leverage points are not down weighted, and this increases the efficiency of this estimator. The effectiveness of the proposed estimator is investigated using real and simulated data sets.

KEYWORDS

Panel data, Fixed effect model, Generalized M-estimator, Outliers.

1. INTRODUCTION

Panel data model combined cross-sectional and time-series data. In panel data, the same cross-sectional unit (industry, firm and country) is surveyed over time, so we have data which is pooled over space as well as time. By combining data in two dimensions, panel data gives more data variation, more degrees of freedom and less collinearity (Gujarati et al., 2009).

Over the past years, there has been a growing trend for using panel data in finance and economics research. Panel data contains multiple observations over multiple time periods for the same firms or individuals. Consider the fixed effect panel data model:

$$y_{it} = \alpha_i + x_{it}^{'}\beta + \varepsilon_{it} i = 1,2,\dots,n\,; t = 1,2,\dots,T \tag{1}$$

where y_{it} is the response 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. At the outset, panel data needs to be transformed within each time series, before any method is applied to the transformed data to estimate the parameters of the model. The classical way of transformation is usually done by computing the mean within each block (time series) in order to eliminate the fixed effect. Such transformation or centering the data using mean centering is done as follows:

and
$$\tilde{y}_{it} = y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it}$$
$$\tilde{x}_{it}^{(k)} = x_{it}^{(k)} - \frac{1}{T}\sum_{t=1}^{T} x_{it}^{(k)} \tag{2}$$

By performing this centering operation, it is observed that the fixed effects α_i have been eliminated and Equation (1) becomes

$$\tilde{y}_{it} = \beta'\tilde{x}_{it} + \varepsilon_{it} \tag{3}$$

The classical Within Group estimator $\hat{\beta}_{WG}$ is obtained when regressing $\tilde{y}_{it}$ on $\tilde{x}_{it}$ by Ordinary Least Squares (OLS) method. Hereafter, we will refer to this classical Within Group estimator as WOLS estimator. By assuming no endogeneity problem, the WOLS is the commonly used method to estimate the parameters of the panel data model on the transformed variables. Similar to classical parameter estimation (OLS) in multiple linear regression for cross sectional data, classical parameter estimation (WOLS) in panel data model also easily affected by outliers.

The presence of outliers may lead to erroneous estimation in panel data model, when employing the WOLS method because it is known to be very sensitive to outliers. As such, a robust method should be used to remedy this problem in panel data model. Unfortunately, not much work is devoted to robust estimation of panel data model.

Wagenvoort and Waldmann (2002) and Bramati and Croux (2007) applied robust Generalized M (GM) estimator using Least Trimmed of Squares (LTS) as an initial estimates and Robust Mahalanobis Distance (RMD) based on multivariate S-estimator to down weight high leverage points (HLPs), respectively. The approach of Bramati and Croux (2007) estimator is more efficient and faster than Wagenvoort and Waldmann (2002) approach. Another robust method to remedy problem of HLPs is by using the MS estimator, the combination of S and M estimator (Yohai et al., 2000). Bramati and Croux (2007) reported that the performances of GM based on Robust Mahalanobis Distance and MS estimator are fairly closed to each other. Recently, Verardi and Wagner (2011) used S-estimator for another robust Within Group estimator. They used the same method of robust data transformation in their studies, in which the data are centered within the time series by using the median instead of the mean in order to eliminate the fixed effect. Median centering is chosen since median have a high breakdown point, very easy to calculate and at the same time median is min max. Unfortunately, this method is found to produce nonlinearity to the resulting data and make the equivariance properties of the robust estimators redundant (Bakar & Midi, 2015).

More recently, Bakar & Midi (2015) used different centering approach whereby data are centered by MM-estimate of location and the GM and MM within group estimators were then applied to the transformed panel data setting. When the MM and GM estimators are applied to the transformed data within each time series, we call this method as the MM Within Group (WMM) and GM6 Within Group (WGM6) estimators. This robust approach is maintained not only to bring linearity back to the transformed data but also improved their performances. Even though they have shown that the GM and MM estimator using MM centering have improved the performance of the estimates, the efficiency of the estimates is decreasing when good leverage points are present in the data. This is due to the fact that the initial weight of WGM6 estimator of Bakar & Midi (2015) utilized RMD which is based on Minimum Volume Ellipsoid (MVE). The weakness of MVE is that it is not only tends to swamp some low leverage points as high leverage, it also considers good leverage points as bad leverage points and its efficiency tends to decrease as the number of good leverage points increases. Hence, in this paper we propose using our newly developed Fast Improvised Generalized MT (GM-FIMGT) estimator and applied this method to the transformed panel data setting. This paper is organized as follows: Section 2 discusses the GM6 estimator. Section 3 presents the algorithm of the GM-FIGMT Estimator. The proposed Robust Within Group estimator based on GM-FIMGT (WGM-FIGMT) for fixed effect panel data model is presented in Section 4. Section 5 and Section 6 discuss the Monte Carlo simulation study and numerical example, respectively to assess the performance of our proposed method. Finally, some concluding remarks are given in Section 7.

2. ALGORITHM OF THE GM6 ESTIMATOR

The GM6 of Coakley and Hettmansperger (1993) is very popular and commonly used method because it has high breakdown point roughly equals to 50%, bounded influence and high efficiency at normal model. The main limitation of this method is that it is based on initial weight function that considers the good leverage points to be bad leverage points. Subsequently its efficiency tends to decrease as the number of good leverage points increases. The GM estimator is defined as the solution of normal equations which is given by:

$$\sum_{i=1}^{n} \pi_i \psi\left(\frac{y_i - x_i^t\hat{\beta}}{\hat{\sigma}\pi_i}\right) x_i = 0 \tag{4}$$

where $\psi = \rho'$ is a derivative of redescending function (weight function) and $\pi_i, i = 1,2,\dots,n$ is the ith initial weight element of the diagonal matrix W, $\hat{\sigma}$ is the scale estimate, and $\hat{\beta}$ is the vector of parameters estimates.

The main aim of GM estimator is to down weight high leverage points which have large residuals. As such, high leverage points and outliers have to be detected at the outset using HLPs diagnostic methods and once they are detected, their effects should be minimized to increase the efficiency of the GM estimator. Coakley and Hettmansperger (1993) proposed GM6 estimator which has high breakdown, roughly equals to 50%, bounded influence and high efficiency at normal distribution. It is noted that most GM estimators rely on the high leverage points diagnostic methods of identification of HLPs. The GM6 employs the Robust Mahalanobis Distance (RMD) based on Minimum Volume Ellipsoid (MVE) or Minimum Covariance Determinant (MCD) for the detection of high leverage points. The RMD is defined as follows:

$$RMD_i = \sqrt{(X - \bar{X}_R)V_R^{-1}(X - \bar{X}_R)^t}, i = 1,2,\dots,n \tag{5}$$

where $\bar{X}_R$ and V_R are robust locations and scatter estimates based on the MVE, respectively.

The value of the cut-off point for RMD_i as proposed by Rousseeuw and Leroy (1987) is $\sqrt{\chi^2_{p,0.5}}$. The ith observation corresponds to RMD_i that exceeds the cut-off point will be identified as high leverage points. Subsequently initial weight of this GM6 estimator (Coakley and Hettmansperger (1993)) is formulated based on RMD which is given by:

$$\pi_i = \min\left[1, \left(\frac{\chi^2_{(0.95,p)}}{RMD^2}\right)\right], i = 1,2,\dots,n \quad (6)$$

The weakness of this initial weight function is that it tends to swamp some low leverage points (Bagheri and Habshah, 2015). Additionally, the detected HLPs are not classified whether they are good or bad HLPs. Subsequently some of good leverages (GLPs) will be given low weights. Hence, the efficiency of the GM6 estimator tends to decrease with the presence of good leverage points. GLPs have no effect or have very little effect on parameter estimates and may contribute to the precision of parameter estimation (Rousseeuw and Van Zomeren, 1990). On the other hand, bad leverage points (BLPs) have high impact on the regression estimates. This is the reason why the GM6 -estimate is less efficient. Another disadvantage of the GM6 estimator is that, it takes too much computing time due to using MVE or MCD.

3 ALGORITHM OF THE GM-FIMGT ESTIMATOR

The Generalized M- Support Vector Regression (GM-SVR) of Dhhan et al. (2016) has been proven to achieve a high breakdown point, bounded influence and high efficiency by employing Fixed Parameter-Support Vector Regression (FP-SVR) of Dhhan et al. (2015) which is very successful in identifying outliers and bad leverage points. The only shortcoming of this method is that it is quite complex and required knowledge of support vector regression. Based on our informal survey, the ordinary statistics practitioners difficult to understand the algorithm of FP-SVR for identifying outliers and bad high leverage points. As such, we propose a relatively simple and fast algorithm of identifying outliers and bad leverage points and use it in the computation of the initial weight of our propose GM estimator. The algorithm of GM estimator is similar to that of Dhhan et al. (2016). The only different is the calculation of the initial weight function. The main aim of our propose GM is to first identify outliers and bad leverage points. Subsequently, their effects are reduced by giving smaller weights to those observations. The correct identification of bad leverage points and outliers is very important in order to get efficient GM estimates. The use of RMD to obtain initial weight for GM6 poses certain shortcomings. The RMD fails to accurately identify high leverage points. According to Bagheri and Midi (2015), the RMD has the tendency to declare more observations as high leverage points due to swamping effects. Moreover, it only detects leverage points without classifying whether it is good or bad high leverage points. As such, the efficiency of the GM6 estimates tends to decrease as the number of good leverage points increases. Hence we propose a relatively easy and fast method to detect bad leverage points and outliers. Then only minimize the weights of bad leverage points and outliers. The propose method is based on the procedure of constructing diagnostic plot similar to Alguraibawi and Midi (2015)for classifying observations into outliers, good and bad leverage points, with slight modification to make it fast by employing RMD based on Index Set Inequality (ISE).

3.1 The algorithm of the classification of observations into outliers and bad high leverage points is summarized as follows:

Classification Step I:

Identify the suspected vertical outliers by using the robust Reweighted Least Squares (RLS) based on Least Median of Squares (LMS). Denote these suspected outliers by *L* set.

Classification Step II:

Identify the suspected high leverage points (HLP) by using Diagnostic Robust Generalized Potential based on Index Set Inequality (DRGP (ISE)) proposed by Lim and Midi (2016) whereby the Robust Mahalanobis Distance that they employed is based on ISE. Denote this set of suspected HLPs by *H* set. It has been shown by Lim and Midi (2016) that ISE is much faster than the commonly used method, namely MVE or fast MCD.

Classification Step III:

From steps 1 and 2, observations that correspond to the union of *L* set and *H* set will be considered as deletion group/set *D*, and the remaining data are labeled *as R set.*

Classification Step IV:

Fit the remaining R set using OLS method to estimate the regression coefficients $(\hat{\beta}_R)$, residuals $(\hat{\varepsilon}_{i.R})$, hat values $(w^*_{ii.R})$, standard deviation $(\hat{\sigma}_R)$ and standard deviation with the ith case deleted $(\hat{\sigma}_{R-i})$. Subsequently, the Fast Improvised Generalized Studentized Residuals (FIMGT) is defined as the following by adapting the Generalized Studentized Residuals (GT_i) of Imon (2005);

$$\mathrm{FIMGT_i} = \begin{cases} \dfrac{\hat{\varepsilon}_{i.R}}{\hat{\sigma}_{R-i}\sqrt{1 - w^*_{ii.R}}} & \text{for } i \in R \\ \dfrac{\hat{\varepsilon}_{i.R}}{\hat{\sigma}_{R}\sqrt{1 + w^*_{ii.R}}} & \text{for } i \notin R \end{cases} \tag{7}$$

The observations are declared as vertical outliers if they have values of FIMGT greater than its cutoff point $(\mathrm{CP_{FIMGT}})$. The $\mathrm{CP_{FIMGT}}$ is defined as follows:

$$\mathrm{CP_{FIMGT}} = \mathrm{Median}(\mathrm{FIMGT_i}) + c\,\mathrm{MAD}(\mathrm{FIMGT_i}) \tag{8}$$

where c is equals to 2 or 3. Following Alguraibawi and Midi (2015), observation are classified according to the following rule and presented in Figure 1:

i. Regular Observation (RO): An Observation is declared as a "RO" if
$|FIMGT_i| \le CP_{FIMGT}$ and $p_{ii} \le Median(p_{ii}) + c\,MAD(p_{ii})$

ii. Vertical Outlier (VO): An Observation is declared as a "VO" if
$|FIMGT_i| > CP_{FIMGT}$ and $p_{ii} \le Median(p_{ii}) + c\,MAD(p_{ii})$

iii. GLPs: An Observation is declared as a GLP if
$|FIMGT_i| \le CP_{FIMGT}$ and $p_{ii} > Median(p_{ii}) + c\,MAD(p_{ii})$

iv. BLPs: An Observation is declared as a BLP if
$|FIMGT_i| > CP_{FIMGT}$ and $p_{ii} > Median(p_{ii}) + c\,MAD(p_{ii})$

FIMGT		
	Vertical Outliers	ad Leverage Points
	Regular Observations	Good Leverage Points
	Vertical Outliers	Bad Leverage Points

DRGP

Figure 1: DRGP against Fast Improvised Generalized Studentized Residuals

It is clearly seen from Figure 1, that the vertical outliers and bad leverage points are detected based on our proposed FIGMT method. Hence, we only down weight the detected bad leverage points and outliers to increase its efficiency. The regular observation and GLPs are given weight equals 1. Therefore, the initial estimate of our propose GM-FIGMT is defined as;

$$\pi_i = \min\left[1, \left(\frac{\mathrm{CP}_{\mathrm{FIMGT}}}{FIMGT}\right)\right], i = 1,2, \dots, n \tag{9}$$

where $\mathrm{CP}_{\mathrm{FIMGT}}$ is defined as in Equation (8).

3.2 Algorithm of the Proposed GM-FIGMT Estimator

The GM6 of Coakley and Hettmansperger (1993) is very popular and commonly used method because it has high breakdown point roughly equals to 50%, bounded influence and high efficiency at normal model. The main limitation of this method is that it is based on initial weight function that considers the good leverage points to be bad leverage points. Subsequently, its efficiency tends to decrease as the number of good leverage points increases. As such, we propose a relatively simple and fast algorithm of identifying outliers and bad leverage points and use it in the computation of the initial weight of our propose GM estimator. We anticipate that this estimator is more efficient than the GM6 estimator because it only minimizes the bad leverage points and outliers, but not the good leverage points. The algorithm of our proposed GM estimator is summarized as follows:

Step 1: Use the S estimator (Rousseuw, 1984) as an initial estimator to achieve a high breakdown of 50% with a $n^{-1/2}$ rate of convergence, and calculate the residuals (r_i).

Step 2: Based on the residuals in Step 1, compute the estimated scale of the residuals, denoted as s and it is defined as $s = (1.4826)$ the median of the largest $(n - p)$ of the $|r_i|$), (see Coakley and Hettmansperger, 1993 and Dhnn et al., (2016) 1.4826 is used to achieve consistency at the Gaussian distribution.

Step 3: Using the estimated residuals (r_i) and the estimated scale (s), find the standardized residuals (e_i), where, $e_i = r_i/s$

Step 4: Compute the initial weight based on FIMGT (Equation 9), where $\pi_i = min[1, \frac{\mathrm{CP}_{\mathrm{FIMGT}}}{FIMGT}]$.

Step 5: Employ the initial weight (step 4) and the standardized residuals (step 3) to achieve a bounded influence function for bad leverage points, $t_i = e_i/\pi_i$.

Step 6: Use the weighted residuals (t_i) in the first iteration of the weighted least squares (WLS) to estimate the parameters of the regression, $\hat{\beta} = (X^T W\, X)^{-1} X^T W\, Y$, where the weight w_i is reduced for large residuals to get good efficiency (Tukey weight function is used in this paper).

Step 7: Calculate the new residuals (r_i) from WLS and repeat steps (2-6) until convergence.

4. ROBUST WITHIN GROUP ESTIMATOR BASED ON GM-FIMGT FOR FIXED EFFECT PANEL DATA MODEL

The robust Within Group estimator based on FIMGT (WGM-FIMGT) is proposed to be applied to the transformed data based on MM-centering.

MM-estimator was originally developed by Yohai (1987). The procedure is to provide an estimator which has high efficiency and high breakdown point.

In the transformed algorithm, each of the variables is considered as univariate data. Let us first consider univariate data for each of the variable $x_1, x_2, ..., x_n$. The location – scale model of this univariate data can be written as follows:

$$x_i = \mu + \sigma \varepsilon_i$$

4.1 The MM Location and Scale of x_i is Computed in Three Steps

Step 1: The S-estimator is computed to obtain the initial consistent estimate μ_0 and scale estimate, $\hat{\sigma}_0$.

Step 2: Compute residuals e_i from Step 1 and then compute M estimate of scale, $\hat{\sigma}$ where $\hat{\sigma}$ is the solution to

$$\frac{1}{n}\sum \rho_0\left(\frac{e_i}{\hat{\sigma}}\right) = 0.5$$

$\psi = \rho'$ has to be re-descending ρ function such as Hampel, Tukey's Bisquare and Tanh. In this paper, we employed Tukey's Bisquare function

Step 3: Compute M estimate of $\hat{\mu}$ using ρ_1. Yohai (1987) highlighted that for Tukey's Bisquare weight function, employing $c_1 = 4.68$, result in high efficiency. $\hat{\mu}$ is a solution to $\sum \psi\left(\frac{e_i}{\hat{\sigma}}\right) x_i = 0$ where $\psi = \rho'\left(\frac{t}{c_1}\right)$, and $t = \frac{e_i}{\hat{\sigma}}$. Upon convergence, $\hat{\mu}$ and $\hat{\sigma}$ is the mm location and mm scale estimates.

4.2 The WGM-FIMGT Estimator is summarized as follows:

Step 1: Compute the location MM estimate ($\hat{u}_{mm}$), for each dependent and independent variable of panel data where MM-centering considers the MM-estimate of location which can provide not only efficiency and robustness, but also high breakdown property.

Step 2: Employ the MM–centering procedure to the data by using Equation.

$$\begin{aligned} \tilde{y}_{it} &= y_{it} - \hat{\mu}_{mm}\{y_{it}\} \\ \tilde{x}_{it}^{(j)} &= x_{it}^{(j)} - \hat{\mu}_{mm}\{x_{it}^{(j)}\} \end{aligned} \tag{10}$$

where j is the number of predictor variables.

Step 3: Regress $\tilde{y}_{it}$ on $\tilde{x}_{it}$ using Fast Improvised Generalized Studentized Residuals based on Index Set Equality (FIMGT) already explained in Section 2 to estimate the parameters of fixed effect panel data model. The resultant estimator is call Robust Within Group Estimator based on FIMGT and is denoted as WGM-FIMGT.

5. MONTE CARLO SIMULATION STUDIES

In this section, we report a Monte Carlo Simulation study that is designed to investigate the performance of our new proposed WGM-FIMGT method under a variety of situations. We consider the fixed effect linear panel data as in Equation (11):

$$y_{it} = \alpha_i + x_{it}^{'}\beta + \varepsilon_{it} \tag{11}$$

where, ε_{it} is normally distributed as $N(0,1)$. Following Bakar and Midi (2015), each of the explanatory variables is generated from a multivariate standard normal distribution $N(0,1)$. Here, we consider panel datasets of (T=10, 20 and 30), each with (n=5, 10) and p=3.

The data is contaminated randomly over all observations (random contamination). Different types of contamination cases are studied; vertical outliers and leverage points at 5% and 10% level of contamination. The contamination is done by replacing randomly some good observations with certain percentage of outliers by a fixed number equals to 100. The contaminated data are done as follows:

1) For vertical outliers, we contaminated randomly by replacing a certain percentage of good observations y with a fixed value equals 100.
2) For bad leverage point, we contaminated randomly by replacing certain percentage of good observations in x_1, x_2 and x_3 with a fixed value equals 100.
3) For combination of bad leverage points and vertical outliers scenario, the bad HLPs are contaminated randomly by replacing certain percentage of good observations for x_1, x_2 and x_3 with a fixed value equals 100. For vertical outliers, the contaminated is done by randomly replacing certain percentage of good observations in y with a fixed value equals 100.

For each of the R = 1000 replications, we estimate the coefficient β of model by applying the classical within group estimator WOLS, WGM6, WMM and our proposed WGM-FIMGT methods. For each scenario, some statistical measures are calculated such as Root Mean Square Error (MSE) and bias (Croux et al., 2003). Bias and Mean-Squared Error (MSE) measures were computed as follows:

$$bias = ||\frac{1}{R}\sum_{J=1}^{R}\hat{B}_j - B|| \tag{12}$$

and

$$MSE = \frac{1}{R}\left\|Y - X\hat{\beta}\right\|^2 \tag{13}$$

where ‖.‖ indicates the Euclidean norm. As per Dhnn et al. (2016), the efficiency (*eff*) is defined as follows:

$$eff \ = \frac{\text{MSE}(\hat{\beta}_{\text{OLSc}})}{\text{MSE}(\hat{\beta})\text{of robust}} \times 100\% \tag{14}$$

It can be observed from Tables 1-6 that the WOLS is very poor, having the smallest efficiency and largest bias in all contamination scenarios. The performance of WGM6 is fairly close to the WGM-FIMGT when the data is contaminated with bad leverage points (BLP). This can be expected because WGM6 is based on MVE estimator which down weight HLPs irrespective whether it is good or bad HLPs. The WGM-FIMGT consistently gives the best result because it is based on FIMGT which is capable of down weighting bad leverage points and vertical outliers. It can be seen that in all contamination scenarios, the WGM-FIMGT has the highest efficiency and smallest bias irrespective of sample size and time series. The WMM always inferior than WGM-FIMGT because MM estimator does not have bounded influence property.

Table 1
Efficiency (%) and Bias (Parenthesis), 5% and 10% of Bad Leverage Points (BLP) and Vertical Outliers (VO)

Methods n	T	WOLS	WMM	WGM6	WGM-FIMGT
5% (BLP &VO)					
10	10	6.065893 (3.187289)	90.30579 (0.0755529)	87.02098 (0.0754993)	**92.99068 (0.074045)**
	20	5.46989 (2.993133)	94.385 (0.02218)	91.7338 (0.02472)	**96.10319 (0.0179597)**
	30	4.973048 (2.997027)	94.77246 (0.006972)	92.9972 (0.0059721)	**95.05038 (0.0012391)**
10%(BLP &VO)					
30	10	5.4234 (2.999272)	91.4125 (0.052525)	84.3327 (0.052059)	**92.6523 (0.050522)**
	20	5.539628 (3.001207)	92.177 (0.01292)	86.3134 (0.013052)	**93.12738 (0.012993)**
	30	4.701945 (2.999415)	92.7259 (0.0206759)	85.11074 (0.024035)	**92.80076 (0.012017)**

Table 2
Efficiency (%) and Bias (Parenthesis), 5% and 10% of Bad Leverage Points (BLP)

Method n	T	WOLS	WMM	WGM6	WGM-FIMGT
5% (BLP)					
10	10	11.10527 (2.998455)	92.5195 (0.05062)	92.61981 (0.055036)	**94.35073 (0.0504751)**
	20	7.561168 (2.997855)	94.2065 (0.016642)	94.44503 (0.019458)	**94.99992 (0.017471)**
	30	6.139189 (2.998442)	92.95739 (0.01374)	92.70966 (0.018914)	**94.50858 (0.007169)**
10% (BLP)					
30	10	11.22463 (2.999846)	92.2246 (0.055589)	92.53778 (0.0681835)	**92.88457 (0.048701)**
	20	7.83835 (2.998907)	92.3825 (0.012618)	90.87349 (0.0271314)	**92.61651 (0.012381)**
	30	6.18918 (2.998829)	91.9573 (0.00493)	90.95283 (0.0154768)	**92.5201 (0.005619)**

Table 3
Efficiency (%) and Bias (Parenthesis), 5% and 10% of Vertical Outliers (VO)

Method n	t	WOLS	WMM	WGM6	WGM-FIMGT
5% (VO)					
10	10	15.46326 (2.997475)	93.79605 (0.04769)	93.0791 (0.051202)	**94.67848 (0.051094)**
	20	12.1171 (2.998671)	93.4652 (0.0140259)	92.33277 (0.02291)	**95.44363 (0.008259)**
	30	8.365529 (2.997941)	93.03661 (0.021731)	94.26835 (0.025281)	**94.4318 (0.012321)**
10%(VO)					
30	10	14.82604 (2.999751)	81.52144 (0.052384)	89.27276 (0.052385)	**90.25635 (0.051754)**
	20	12.28001 (2.998983)	93.54864 (0.0144)	92.21883 (0.028409)	**93.57283 (0.009665)**
	30	8.323474 (2.998988)	92.82697 (0.006399)	92.07799 (0.012099)	**92.20927 (0.01023)**

Table 4
Efficiency (%) and Bias (Parenthesis), 5% and 10% of Bad Leverage Points (BLP) and Vertical Outliers (VO)

Method n	t	WOLS	WMM	WGM6	WGM-FIMGT
5%(BLP &VO)					
5	10	8.007664 (2.998968)	92.87965 (0.046458)	90.35659 (0.036623)	**92.18667 (0.032593)**
	15	7.8987 2.993761	93.7274 0.00437704	91.04437 0.031297	**95.24868 (0.011731)**
	30	5.884518 (2.999001)	93.72742 (0.00437)	90.41367 (0.004809)	**95.7928 (0.004729)**
10%(BLP &VO)					
5	10	5.4234 (2.999272)	91.4125 (0.052525)	84.3327 (0.052059)	**92.6523 (0.050522)**
	15	5.539628 (3.001207	92.177 (0.01292)	86.3134 (0.013052)	**93.12738 (0.013993)**
	30	4.701945 (2.999415)	92.7259 (0.0206759)	85.11074 (0.024035)	**92.80076 (0.012017)**

Table 5
Efficiency (%) and Bias (Parenthesis), 5% and 10% of Vertical Outliers (VO)

Method n	t	WOLS	WMM	WGM6	WGM-FIMGT
5% (VO)					
5	10	5.564356 (0.215801)	92.5095 (0.037823)	90.52095 (0.030558)	**93.59219 (0.062925)**
	15	4.916625 (0.145266)	94.1192 (0.02411)	88.80737 (0.029834)	**95.2088 (0.020445)**
	30	4.483283 (0.269339)	(93.98203) (0.082298)	87.71519 (0.01025)	**94.61697 (0.01258)**
10%(VO)					
5	10	3.26458 0.277661	89.99722 (0.081308)	72.28979 (0.071106)	**92.30196 (0.070793)**
	15	3.398262 (0.2168260	91.97976 (0.0343728)	76.90227 (0.038788)	**94.07129 (0.025773)**
	30	3.424634 (0.318043)	91.07577 (0.022244)	78.09512 (0.021811)	**91.18304 (0.017117)**

Table 6
Efficiency (%) and Bias (Parenthesis), 5% and 10% of Bad Leverage Points (BLP)

Method n	t	WOLS	WMM	WGM6	WGM-FIMGT
5%(BLP)					
5	10	15.46326 (2.997475)	93.79605 (0.04769)	93.0791 (0.051202)	**94.67848 (0.051094)**
	15	12.1171 (2.998671)	93.4652 (0.014025)	92.33277 (0.02291)	**95.44363 (0.008259)**
	30	8.365529 (2.997941)	93.03661 (0.02173157)	94.26835 (0.025281)	**94.4318 (0.012321)**
10%(BLP)					
5	10	14.82604 (2.999751	81.52144 (0.052384)	**89.27276 (0.052385)**	80.25635 (0.064754)
	15	12.28001 (2.998983)	93.54864 (0.0144)	92.21883 (0.028409)	**92.57283 (0.009665)**
	30	8.323474 (2.998988)	92.82697 (0.006399)	92.07799 (0.012099)	**92.20927 (0.01023)**

6. NUMERICAL EXAMPLE

In this section, we applied (WGM-FIMGT, WOLS, WMM, WGM6) methods to real panel data to evaluate the performance of our newly developed method.

Airline Data Set

This airline firms data is taken from Greene (2007) used to study the efficiency in production of airline services. It represents the relationship between the response variable (cost) and three predictor variables (output, fuel price, and load factor) over 15 yearly observations (1970-1984). To see the effect of outliers, the good observations of original data for x1 are replaced with a fixed value equals to 50. Also, 4 observations of original data for y are replaced with a fixed value equals to 50 (10%). According to 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. A straight forward equation can be formed as an illustration where the total cost of production is fitted into a multiple linear regression model:

$$\ln \cos t_{it} = \alpha_i + \ln output_{it} + \ln\ fuelprice_{it} + load\ \ factor_{it} + \varepsilon_{it} \tag{13}$$

The WOLS, WGMM, WGM6 and WGM-FIMGT were then applied to the data. Table 7 illustrates the result for clean data (original) and contaminated data (modified). The evaluation is based on the standard error. The standard errors of the parameter coefficients are estimated by bootstrapping method and written in parentheses. We employed fixed $x-$resampling method to generate the bootstraps samples of 1000 replications. The method is also known as bootstrapping the residuals of linear regression models (Davison and Hinkley, 1997). It can be observed that when there is no outlier in the data (original) the WOLS performs the best, having the smallest standard errors of the estimates compared to other estimators. However its performance becomes poor in the presence of outliers. In this situation the WGM-FIMGT outperformed other estimators evident by having the least standard errors compared to other methods.

Table 7
Parameter Estimates for the Original and Modified Data in Airline Data

		Mean centering	**mm-centering**		
		WOLS	**WMM**	**WGM6**	**WGM-FIMGT**
Original Data	$\hat{\beta}_1$	**0.919285 (0.0229)**	0.9165 (0.0277)	0.9186127 (0.0313)	0.921756869 (0.02939463)
	$\hat{\beta}_2$	**0.4175 (0.0061)**	0.4108 (0.0140)	0.407768626 (0.014078)	0.417572035 (0.01224403)
	$\hat{\beta}_3$	**-1.0700 (0.1914)**	-1.0766 (0.1836)	-1.023110 (0.185789)	-0.95208886 (0.20724239)
Modify data	$\hat{\beta}_1$	-0.03479 (0.08131)	0.87744 (0.02996)	0.9224339 (0.052789)	**0.907779152 (0.0157360)**
	$\hat{\beta}_2$	-0.30806 (1.22679)	0.42387 (0.03372)	0.410947718 (0.0234966)	**0.42048643 (0.0149187)**
	$\hat{\beta}_3$	8.73823 (22.04315)	-0.86095 (0.38262)	-0.9855288 (0.32329049)	**-0.9880553 (0.257718)**

7. CONCLUSION

In this paper, we proposed new Within Group WGM-FIMGT method to estimate the parameters of fixed effect panel data model. The developed WGM-FIMGT method is compared with WOLS, WGM6 and WMM estimators. The new WGM-FIMGT technique outperformed other methods and has the highest efficiency as it down weight bad leverage points and vertical outliers in the first step of the algorithm. The effectiveness of the proposed approach is tested on real and simulated data sets. The results from both the real and the simulated data show that the performance of the WGM-FIMGT is robust against both vertical outliers and bad leverage points. Hence we suggest using WGM-FIMGT as a robust alternative to the WGM6 method for parameter estimation in fixed effect panel data model.

REFERENCES

1. Gujarati, D.N. (2009). *Basic econometrics*. Tata McGraw-Hill Education.
2. 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.
3. Bramati, M.C. and Croux, C. (2007). Robust estimators for the fixed effects panel data model. *The Econometrics Journal*, 10(3), 521-54
4. Yohai, V.J. (1987). High Breakdown-Point and High Efficiency Estimates for Regression. *The Annals of Statistics*, 15, 642-65.
5. Verardi, V. and Wagner, J. (2011). Robust estimation of linear fixed effects panel data models with an application to the exporter productivity premium. *Jahrbücher für Nationalökonomie und Statistik*, 231(4), 546-557.

6. Bakar, N.M.A. and Midi, H. (2015). Robust centering in the fixed effect panel data model. *Pakistan Journal of Statistics*, 3(1), 33-48.
7. Coakley, C.W. and Hettmansperger, T.P. (1993). A bounded influence, high breakdown, efficient regression estimator. *Journal of the American Statistical Association*, 88(423), 872-880.
8. Bagheri, A. and Midi, H. (2015). Diagnostic plot for the identification of high leverage collinearity-influential observations. *SORT*, 39(1), 51-70.
9. Rousseeuw, P.J. and Van Zomeren, B.C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, 85(411), 633-639.
10. Dhhan, W., Rana, S. and Midi, H. (2015). Non-sparse $\in$-insensitive support vector regression for outlier detection. *Journal of Applied Statistics*, 42(8), 1723-1739.
11. Dhhan, W., Rana, S. and Midi, H. (2016). A high breakdown, high efficiency and bounded influence modified GM estimator based on support vector regression. *Journal of Applied Statistics*, 44(4), 700-714.
12. Alguraibawi, M., Midi, H. and Imon, A. (2015). A new robust diagnostic plot for classifying good and bad high leverage points in a multiple linear regression model. *Mathematical Problems in Engineering*, Volume 2015, Article ID 279472, 1-12.
13. Lim, H.A. and Midi, H. (2016). Diagnostic Robust Generalized Potential Based on Index Set Equality (DRGP (ISE)) for the identification of high leverage points in linear model. *Computational Statistics*, 3(31), 859-877.
14. Leroy, A.M. and Rousseeuw, P.J. (1987). *Robust regression and outlier detection*. Wiley Series in Probability and Mathematical Statistics, New York: Wiley.
15. Croux, C. and Haesbroeck, G. (2003). Implementing the Bianco and Yohai estimator for logistic regression. *Computational Statistics & Data Analysis*, 44(1), 273-295.
16. Rousseeuw, P.J. (1984). Least median of Squares Regression. *Journal of the American Statistical Association*, 79(388), 871-880.
17. Yohai, V.J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 15(2), 642-656.
18. Greene, W.H. (2007). *Econometric Analysis*. 6th Edition, Prentice Hall.
19. Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and their Application*. Cambridge University Press.