International Journal of Innovative Mathematics, Statistics & Energy Policies
12(2):41-59, April-June, 2024

 www.seahipaj.org ISSN: 2467-852X

Comparison Of Robust Linear Regression Estimators For Finite Sample Fixed Effect Panel Model With Influential Values

[1]E. M. Eke, [2]I. D. Essi, [3]G. L. Tuaneh

[1]Department of Statistics, Covenant Polytechnic, Owerrinta, Abia State, Nigeria
Email: ekeemenikemong@gmail.com

[2]Department of Mathematics, Rivers State University, Port-Harcourt, Nigeria
Essi.isaac@ust.edu.ng

[3]Department of Agric and Applied Economics, Rivers State University, Port-Harcourt, Nigeria
lebarituaneh@gmail@gmail.com

ABSTRACT
In this paper we compare the robustness of linear regression estimators for finite sample fixed effect panel model having influential values using five centering estimators (Mean, Median, MoptM, MM and DCML) and four regression estimators (OLS, WLS, GM-FIGMT and DCML). The results from Monte Carlo simulation show that OLS outperformed the other estimators when the panel data is uncontaminated, GM-FIGMT outperformed the other estimators in estimating the beta in panel data with 5% and 10% random vertical outliers and bad leverage points, but the DCML yield the smallest standard error of beta-hat; but for 5% and 10% block concentrated vertical outliers and bad leverage points contaminated panel data, DCML regression estimator outperformed the other regression estimators in estimating beta and standard error of beta-hat. The results of GDP growth and FDI dataset show that DCML estimator (centering and regression) recovered the positive effect of FDI on GDP growth which the Mean, Median, MoptM and MM centering estimators of OLS lost. From the findings of this paper, we conclude and recommend thus: when panel data contamination is bi-directional (X and Y), and panel contamination is block concentrated vertical outlier and leverage point, the DCML location and regression estimators yield the best robustness results for all finite samples panel data (DCML recommended); when panel data contamination is bi-directional (X and Y), and panel contamination is having random vertical outlier, GM-FIGMT yields the best robustness parameter estimates of beta but fails in the standard error estimation which the DCML outperformed it.
Keywords: Panel data, Influential values, Finite sample Fixed effect, GM-FIGMT, DCML.

I. INTRODUCTION

Panel data is one of the types of empirical data for econometric and statistical data analysis; the other two are cross-sectional data and time-series. This classification has made it easier to identify suitable statistical models and the estimators of the model parameters for each of the three types of econometric data. The assumptions made about the estimators used for cross-sectional data analysis differ in some magnitudes from those of time series analysis and also from those of panel data analysis. Though there are differences in the estimators of the models used in the three types of econometric data, the estimators sometimes share certain similarities in some specific areas. Advances in modern statistical modeling in general and econometric analysis in particular have shown that panel data which is the newest among these three types of empirical econometric data have extended much of the cross-sectional and time series estimators to adapt to their peculiarities, Hsiao (2022). These account for why Panel data takes different names as longitudinal data, cross-sectional time series data or multi-level data. All these variant names of panel data comes because panel data combines cross-sectional data and time series elements and this involves observing multiple individuals, entities or units over certain period of time; allowing researchers to analyze changes over time within individual units and capture both individual-specific characteristics and time-varying effects, (Pesaran, 2015; Hsiao, 2022 Armstrong et al (2023) and Chesher et al, (2024).

Panel data analysis has found wide applications in several fields of human endeavour including and not limited to Econometrics, Engineering, Epidemiology, Finance, Social and Behavioural Sciences to mention but a few (Gil-García and Puron-Cid, 2015; Beyaztas and Bandyopadhyay, 2020; Ibrahim and Arundina, 2022). In Econometrics, it is known by most researchers as Panel data analysis, to the Physical Sciences and Engineering, it is often called longitudinal data analysis while the Social and Behavioural Scientists refer to it as multi-level analysis (Feng *et al.*, 2022). This multi-disciplinary application of panel data analysis contributes to the increase in the rise of Micro-panel and Macro-panel studies and databases across firms, countries and international organisations; and its application in various academic disciplines, (Jaba *et al*, (2017; Yaffee, 2003 and Gil-García and Puron-Cid, 2015).

The common regression analysis techniques used in panel data analysis are pooled ordinary least squares, fixed effects, random effects and dynamic panel regression. All these techniques are aimed at controlling unobserved time invariant heterogeneity in cross-sectional model, disentangling component of variance in the panel data and estimating transition probability among states so as to study the dynamics of cross-sectional population, Arellano (2003). These have led to the classification of panel data model into fixed effect models and random effects models which are further grouped into static models and dynamic models. To eliminate the unobserved time-invariant heterogeneity in panel data model, three models have been in use: Least Squares Dummy Variable (LSDV), First Differencing (FD) and Fixed Effects (FE). These three methods are also referred to as the fixed effect regression. LSDV model accounts for the time-invariant individual components by explicitly introducing them into the model specification in the form of individual intercepts and thereby the model suffers from large loss of degrees of freedom as many parameters are estimated which aggravates the problem of multicollinearity, Baltagi (2021). FD method which removes the time-invariant individual components by lagging the model and subtracting, has the problem of not being suitable when the number of panels and the time component is small and finite. FE method entails the transformation of the data by subtracting the average over time (individual) to every variable, usually referred to as time-demeaning method. The FE model is preferred to the FD and the LSDV model as it requires only minimal assumption on the nature of heterogeneity, thereby making it the simplest, most efficient and robust specification in panel data analysis, Croissant and Millo (2019); Millimet and Bellemare (2023). The classical regression method used in static fixed effects linear panel data model is Ordinary Least Squares (OLS) which makes us enjoy the best linear unbiased estimator (BLUE) properties of the OLS after demeaning the data, Ebrima and Yahaya (2022).

Data contamination in regression analysis remain the major challenge to the use of OLS in static fixed effects panel data models as many panel data suffer from several forms of data contaminations: outliers (unusual observations in the dependent variable), leverage points (unusual observations in the independent variables), influential values (unusual observations in both the dependent variable and independent variables), heteroscedasticity (non-constant) errors, multicollinearity and wrong data generating processes (Rousseeuw and Van-Zomeren, 1990; Bramati and Croux, 2007; Aquaro and Čížek, 2010). Classical regression models are highly sensitive to the various data contamination especially influential values and it compromises the validity of the OLS regression assumption if a non-robust technique is used ("Robust regression", 2021). Popular among the robust alternatives to the classical OLS regression estimators are Least Absolute Deviations (LAD) estimators, M-estimators, Least Trimmed Squares (LTS) estimator, Least Median of Squares (LMS) estimator, S-estimator, Theil-Sen estimator, MM-estimator, $\tau-$estimator and the distance constrained maximum likelihood (DCML) estimator, Varin (2021).

The OLS estimator as the typical classical regression estimator has been found to fail in robustness and efficiency when the data contains outliers or leverage points as they are seen to have low breakdown point of 0%, Rousseeuw and Van-Zomeren, (1990) and Zaman et al (2001). Though several variants of M-estimators have been proposed and applied in the literature to overcome the problem outliers or leverage points but they are still built on the asymptotic property of large samples (Huber (1981); Bramati and Croux, 2007; Aquaro and Čížek, 2010; Midi and Mohammad, 2018). A shot at the finite sample estimation is the work of Maronna and Yohai (2014), but it is based on cross-sectional data. The task of getting finite sample robustness in panel data can only be achieved by using estimators tailored towards finite-sample estimators' properties. In this paper, we compare the performance of the existing linear regression estimators to access how they adapt to fixed effect panel datasets with influential values.

The rest of this paper are organised as follows: section II reviews the literature of econometrics of panel data model by considering the estimators used in the static linear fixed effects panel data model with additive specific effects with the problems that bedevil them. Section III considers the methodologies applied in the within group linear static fixed effects panel model with additive effects. In section IV, the results of the various analyses are presented in tables. The conclusion drawn from the discussions of the results and the recommendations on the areas for future research are presented in the section V.

II. LITERATURE REVIEW

The first attempt to develop a robust regression estimator on panel data is the work of Wagenvoort and Waldman (2002) caption "On B-robust instrumental variable estimation of the linear model with panel data". They proposed two estimators: Two Stage Generalized M (2SGM) estimator and Robust Generalized Method of Moments (RGMM). The two estimators have bounded influence functions and are consistent as well as asymptotic normally distributed. The estimators are in the linear static panel data but being based on the instrumental variable which suffer from the choice of strong, weak or valid instrument as instrumental variables estimator rely on a strong instrument, Druten (2020). And instrumental variables estimates are biased and inconsistent in finite samples, Bound *et al.* (1995).

A step further in the development of robust regression estimator in panel data is the work of Bramati and Croux (2007) similar to the approach of Wagenvoort and Waldmann (2002). They developed a robust within group (RWG) estimator in the linear static fixed effect panel data and compared it to the classical Least squares within group (WG) estimator and MS-estimator of Gervini and Yohai (2002). RGW estimators has a 25% breakdown point for the regression estimator of the panel data model (for both N and T tending to infinity). These three estimators (WG, RWG and MS) where applied to panel data contaminated with five contamination conditions: no outlier, vertical outliers, block concentrated vertical outliers, leverage point and concentrated leverage point. They used the Mean and Median centering to transform their panel data. The RWG and MS estimators studied the problem of outliers and bad leverage points in panel data focusing on fixed effects models with assumption of homoscedasticity and no autocorrelation of error and they are consistent only if the number of time periods increases to infinity which makes them unsuitable for small panels. Baramati and Croux (2007) argued that the presence of the outlying observations affects the classical within group estimators so their estimators RWG (robust within groups) is less sensitive to the presence of unusual observations. They used the LTS estimators (Rousseeuw 1984) as initial regression estimators and a multivariate S estimator to downweigh leverage as the LTS and S estimators are more efficient and faster to compute than the one used by Wagenvoort and Waldman (2002). To increase the statistical efficiency and robustness of their RWG estimators they obtain a residual robust scale estimate based on the LTS estimation and built a weighting diagonal matrix in line with Bisquare function. The simulation results as well as their empirical data results show that RWG outperformed MS and WG in the presence of the data contaminated conditions except for no outlier of which the classical OLS outperformed RWG. The shortcomings of RWG estimator in addition of being not equivariant to rescaling of data, is that it is bias due to the nonlinearity of the procedures if the number of the panels is fixed, Aquaro and Cizek (2010).

Aquaro and Cizek (2010) proposed on alternative robust estimation approach for linear fixed effects panel data models that is equivariant with respect to standard data transformations. They employed two more different data centering/transformations (first-difference and pairwise-difference) to the existing Mean and Median and show that it is possible to apply standard robust estimators of linear regression to the transformed data. Because of the data transformation, the equivariance, robust and asymptotic properties of the proposed estimators were established. The estimators have positive breakdown point equals to 25% and its asymptotically normally distributed. The finite sample performance of the proposed estimators matches the standard between group least squares estimator and the robust properties; thus, does not adversely affect the precision of estimation. They compared six estimators: least squares (LS), least trimmed squares (LTS), within-group generalized M-estimator (WGM), robust and efficient weighted least squares (REWLS) and reweighted least trimmed squares (RLTS) on panel data transformed by Mean, Median, first difference and pairwise difference centering. The efficiency and robustness of the model were compared using uncontaminated panel data and panel data contaminated with vertical outliers and leverage point for clustered outliers and non-clustered outliers with 5% and 20% contamination. They also checked the performance of the estimators given three distributions of the error term (normal, t, and double-exponential). The RLTS outperformed the other estimators for the various data transformation (mean, median, first difference and pairwise difference). The best estimation results were obtained from the pairwise data transformation.

Bramati and Croux (2007) and Aquaro and Cizek (2010) pointed out the problem of centering the dependent and independent variables as the first challenge in fixed effects panel data estimation which the classical static fixed effect panel data estimators use the non-robust mean as the location estimator. And to circumvent this, they use the median as a robust location estimator, though it suffers from low efficiency. In addition to the low efficiency shortcoming of the median centering, the problems of the within group median being non-linearity and also losing their regression equivariance property. The pairwise-differencing centering of Aquaro and Cizek (2010) could not be used in the finite panel data estimation because it requires up to 12-order differencing. To overcoming these problems of the Median centering of the within groups estimators, Bakar and Midi (2015) proposed the use of MM-centering after the MM-estimator of Yohai (1987) to provide robust solution to the classical within group Mean location parameter estimate when the data are contaminated with outlying observations. And to lend more empirical support on the superiority of the MM-centering over the Median and Mean-centering, Midi and Mohammed (2018) compared Mean, Median and MM centering methods. MM centering outperformed the Mean and Median centering. The Mean centering was found to be very sensitive to outliers, while the Median centering was non-linear with low efficiency.

Lee *et al.* (2016) compares the LTS estimator, LMS estimator, M-estimator, S-estimator, and MM-estimator for the growth of 61 countries using GDP as the response variable and four explanatory variables (labor force growth, GDP per worker gap, equipment share and non-equipment share. Their result shows that the MM estimator which reduces the effect of the influential data outperform the other estimators in regression analysis. So, the LTS, LMS, M, S-estimators will not be considered in the simulation and real-life dataset in this study. Abu Bakar and Habshah (2018) compare the robust within group M-estimator, RWGM of Bramati and Croux (2007) and the robust within Group MM (RWGMM) to two panel data under robust MM centering in the presence of high leverage points and the RGMM estimator outperformed the RWGMM. So, we will not consider the RWGM estimator in this study.

Maronna and Yohai (2014) proposed the DCML estimator, a robust linear regression estimator that attains the properties of finite-sample efficiency which is close to one and has its maximum Mean Squared Error (MSE) smaller than that of the MM-estimator in the presence of various data contaminations. The DCML estimator is developed for the linear model with normal errors using the MM-estimator and the distance induced by the Kullback- Leibler divergence. DCML estimator also address the problem of predictor distributions as their study shows that the asymptotic efficiency of the DCML estimator does not depend on the predictor's distribution. A simulation study by Maronna *et al.* (2019) on the finite sample efficiency and maximum mean square errors (MSEs) of MM-estimators and DCML estimator showed that DCML outperformed the MM-estimator. Varin (2021) conducted an extensive and updated review of the predictive performance of OLS and seven robust linear regression estimators on a real epidemiological data and simulated data set contaminated with influential observations and discovered that among the seven robust estimators (Huber M-estimator, Tukey's Bisquare M-estimator, Least Absolute Deviations (LAD), Fast MM-estimator, Fast τ-estimator, high breakdown rank-based (HBR) estimator and distance-constrained maximum likelihood (DCML) estimator, DCML estimator outperformed the other estimators in small samples of $n = 50$ with MSEs lower than that of the other estimators at 5%, 10% and 30% data contamination of outliers and leverage points. Midi and Mohammed (2018) propose a robust fixed effects within group weighted least squares (WLS) estimator which uses the least weighted squares for the within group parameter estimation. Ismael and Midi (2018) propose the robust within group estimator for fixed effects of panel data based on GM-FIMGT an acronym for Generalized-M Fast Improved Generalized Studentized Residuals-Based on Index Set Equity; which they called robust Within Group Generalized-M Fast Improved Generalized Studentized Residuals-Based on Index Set Equity (WGM-FIMGT) estimator. They compare the within group OLS, MM, GM6 and GM-FIMGT estimators. The within group GM-FIMGT estimator (WGM-FIGMT) outperformed the other estimators in simulated and real-life datasets with outliers and bad leverage points. So, we will not consider the MM-estimator in this study.

Research Gap

An in-depth search in the Statistics and Econometrics literature shows that DCML estimator has been applied in some areas of direct applications of cross-sectional linear regression but has not been applied in the linear static fixed effects panel data models. Therefore, this paper seeks to compare the existing classical OLS, two robust location estimators (WLS and GM-FIMGT) and the DCML regression estimator of Maronna and Yohai (2014) to access their finite sample robustness in the linear static fixed effect panel data with influential values with is not existing in the Statistics and Econometrics literature.

III. METHODOLOGY

The linear static fixed effect panel data regression model with additive effects that incorporates the unobserved effects α_i and the regressors is

$$\boldsymbol{y}_{it} = \boldsymbol{\alpha}_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{it} \quad ; i = 1, \dots, N; \quad t = 1, \dots, T \qquad (1)$$

where $\boldsymbol{y}_{it}$ is the dependent variable; $\boldsymbol{x}'_{it}$contains the observable covariates that change across t but not i, variables that change across i but not t and variables that changes across i and t. $\boldsymbol{x}'_{it}$ is assumed to be strictly exogenous with respect to ε_{it}; $\boldsymbol{\beta}$ is a vector of constants; $\boldsymbol{\alpha}_i$ is the effects of the variable peculiar to the i^{th} cross-sectional individual unit in more or less the fashion over time and $\boldsymbol{\varepsilon}_{it}$ is the error term assumed to be an independently identically distributed random variable with mean zero and variance σ_ε^2, i.e., $\boldsymbol{\varepsilon}_{it} \sim N(\boldsymbol{0}, \sigma_\varepsilon^2 \boldsymbol{I})$.

Procedures of Within Group Panel Data Estimation

The within group panel data estimation procedure is divided into two: Centering and regression parameter estimation.

Centering Estimation

The mean centered data also called time demeaned data, $\ddot{y}_{it}$ is given as

$$\ddot{y}_{it} = y_{it} - \bar{y}_{i.}$$
$$\ddot{x}_{it} = x_{it} - \bar{x}_{i.} \qquad (2)$$

$where\ \bar{y}_{i.} = \frac{1}{T}\sum_{t=1}^{T} y_{it}$ is group mean of the dependent variable for the panel data and

$\bar{x}_{i.} = \frac{1}{T}\sum_{t=1}^{T} x_{it}$ is group mean of the dependent variable for the panel data.

The median centered data, $\tilde{y}_{it}$ is given as

$$\tilde{y}_{it} = y_{it} - \underset{t}{median}(y_{it})$$
$$\tilde{x}_{it} = x_{it} - \underset{t}{median}(x_{it}) \qquad (3)$$

The Modified Optimal M location estimator, $\hat{\mu}_{MoptM}$ is the solution of the maximum likelihood estimator (4) below

$$\sum_{i=1}^{n} \psi\left(\frac{y_i - \hat{\mu}}{\hat{\sigma}}\right) = 0 \qquad (4)$$

where $\hat{\sigma} =$ median absolute deviation normalized to be unbiased and $\psi(x)$ is the modified optimal psi function given as

$$\psi_a^{(mod)}(\varepsilon) = \left(\frac{\phi(1)}{\phi(1) - a}\right)\psi_a^{(0)}(\varepsilon) = \begin{cases} \varepsilon & -1 \le \varepsilon \le 1 \\ \left(\frac{\phi(1)}{\phi(1) - a}\right) y - sign(y)\frac{a}{\phi(y)} & if\ |y| < upper \\ 0 & otherwise \end{cases} \qquad (5)$$

The Modified Optimal M-estimator of location has 99% efficiency and is implemented in the locScaleM function of RobStatTM package of R software.

The Modified Optimal M-estimator of location estimate $\hat{\mu}_{MoptM}$ centered data becomes

$$\ddot{y}_{it} = y_{it} - \hat{\mu}_{MoptM}(y_{it}); \qquad (6)$$
$$\ddot{x}_{it} = x_{it} - \hat{\mu}_{MoptM}(x_{it})$$

The MM location estimate denoted in this paper as $\hat{\mu}_{MM}$ is the solution of $\sum \psi\left(\frac{\varepsilon_i}{\hat{\sigma}}\right) x_i = 0$. For the detailed explanation of estimation procedure of MM-location estimator for fixed effect panel data see Bakar and Midi (2015); Eke, Essi and Nse (2024).

The MM- location estimate centered data becomes

$$\overset{\frown}{y}_{it} = y_{it} - \hat{\mu}_{MM}(y_{it}) \qquad (7)$$
$$\overset{\frown}{x}_{it} = x_{it} - \hat{\mu}_{MM}(x_{it})$$

The DCML estimator of location $\hat{\mu}_{DCML}$ see Eke, Essi and Nse (2024) is given as

$$\hat{\mu}_{DCML} = arg\min_{\mu} \sum_{it=1}^{NT} r_{it}^2(\mu) \quad Subject\ to\ d_{KL}(\mu_0, \mu) = \frac{1}{\hat{\sigma}^2}(\hat{\mu}_0 - \hat{\mu})'\boldsymbol{C}(\hat{\mu}_0 - \hat{\mu}) \qquad (8)$$

with $\boldsymbol{C} = \boldsymbol{E}(\alpha_i \alpha_i')$ and MM $-$ estimator optimizer $= \sum_{i=1}^{n} \psi_B\left(\frac{y_{it} - \hat{\mu}_0}{c_0 \hat{\sigma}_0}\right)$

The DCML Centered data, $\breve{y}_{it}$will be given as

$$\breve{y}_{it} = y_{it} - \hat{\mu}_{DCML}(y_{it}) \qquad (9)$$
$$\breve{x}_{it} = x_{it} - \hat{\mu}_{DCML}(x_{it})$$

Regression Parameter Estimation.

The OLS estimation is applied to the various centered data to obtain $\hat{\beta}$ estimate referred to in this paper as estimate of within group OLS ($\hat{\beta}_{WGOLS}$) given as

$$\hat{\beta}_{WGOLS} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{\boldsymbol{x}}_{it}'\ddot{\boldsymbol{x}}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{\boldsymbol{x}}_{it}'\ddot{\boldsymbol{y}}_{it}\right) \qquad (10)$$

The lm function of the R statistical package is the implementation (10) above.

The parameter β estimation of WGWLS for Midi and Mohammed (2018) is obtained as

$$\hat{\beta}_{WGWLS} = \arg\min_{\beta} \sum_{q=1}^{NT} W_q\, r_q^2(\beta) \qquad (11)$$

where $W_q \in [0,1]\ for\ q = 1, 2, \dots, NT$ is the weight and is generated as a monotone function W i.e., $W_q = W(q - 1/NT), w_1 \geq w_2 \geq \dots \geq w_{NT}$ with $w(0) = 1$. And $r_q^2(\beta)$ is the squared residuals of the OLS estimator.

The robust within group estimator for fixed effects of panel data proposed by Ismael and Midi (2018) is based on Generalized-M Fast Improved Generalized Studentized Residuals-Based on Index Set Equity (GM-FIMGT). The algorithm of GM-FIGMT estimator given by:

- Step 1: Calculate the residuals (r_i) based on S estimator developed by Rousseeuw (1984).
- Step 2: Calculate the estimated scale (σ) of the residuals s = (1.4826) (the median of the largest (n-p) of the $|r_i|$), where r_i is obtained from Step 1.
- Step 3: Compute the standardized residuals (e_i), where $e_i = r_i/s$.
- Step 4: Calculate the initial weight, denoted as $\pi_i = min\left[1, \left(\frac{CP_{FIMGT}}{FIMGT}\right)\right]$.
- Step 5: Compute the bounded influence function for bad leverage points, $t_i = e_i/\pi_i$.
- Step 6: Employ the weighted least squares (WLS) to estimate the parameters of the regression, $\hat{\beta} = (X'WX)^{-1}X'WY$, where the weight w_i is reduced for large residuals to get good efficiency (they employed the Tukey weight function).
- Step 7: Calculate the new residuals (r_i) from WLS and repeat steps (2-6) until convergence.

The centered dependent variable is regress on the centered independent variable(s) using GM-FIMGT estimator to obtain the robust within group estimator (WGM-FIMGT). The algorithm of GM-FIGMT estimator above is implemented in R-Software.

Maronna and Yohai (2014) proposed the DCML regression estimator of linear model with normal error which has a high finite sample efficiency close to one and a breakdown point of 50%, thereby being asymptotically fully efficient and has a good balance between efficiency and robustness and it is given as

$$\sum_{i=1}^{n} r_i(\beta)^2 \qquad Subject\ to\ \hat{d}_{KL,C}(\hat{\beta}_0, \beta) \leq \delta \qquad (12)$$

So, $\hat{\beta}$ is defined as the minimizer of $\sum_{i=1}^{n} r_i(\beta)^2$ subject to $Subject\ to\ \hat{d}_{KL,C}(\hat{\beta}_0, \beta) \leq \delta$. The DCML regression of Maronna and Yohai (2014) is implemented in RobStatTM package of R.

Simulation Study and Statistical Software

This simulation study investigates the behaviour of the estimators when the sample sizes vary and when different kinds of outlying observations are present by comparing the classical within-group estimator $\hat{\beta}_{WGOLS}$ and the robust within group estimators. The robust within group estimators considered are WGWLS estimator of Midi and Mohammed (2018), WGM-FIMGT estimator of Ishmael and Midi (2018) and DCML estimator of Maronna and Yohai (2014). The statistical software used for this simulation study is R version 4.1.3 version 2022.

The data generating process of a static fixed effect panel data with additive effect model of (1) given below $y_{it} = \alpha_i + \beta_1 x_{1it} + \varepsilon_{it}$, where $\varepsilon_{it} \sim N(\mu = 0, \sigma^2 = 1)$. The parameter $\beta_1 = 3$ and the $\alpha_i \sim U(0,12)$. In generating x_{it}, the format of Aquaro and Cizek (2013) and Beyaztas and Bandyopadhyay (2020) is that $x_{it} = \chi_2^2 - 2$. But Bramati and Croux (2007), Bakar and Midi (2015) and Ismaeel and Midi (2018) toyed the line that each of the explanatory variables be generated from a multivariate standard normal distribution, and that we applied such that $x_{it} \sim N(0,1)$. The sample sizes are the combination and product of the three cross-sectional dimensions, N = 5, 10, 40 and the time period T = 10, 30, 50 which yield the sample sizes of 50, 100, 150, 250, 300, 400, 500, 1200 and 2000.

To ascertain the robustness performances of the estimation procedures, the random contamination and concentrated (or clustered) contamination as given in Bramati and Croux (2007) and Beyaztas and Bandyopadhyay (2020) are considered. The proportions of contamination chosen are 5% and 10% in line with the recommendation of Shu (1978, p. 6). The combination of the proportions of outlying contaminations, the types of outlying contaminations and the description of outlying contaminations yield the following panel data contamination considered in this simulation experiment: Uncontaminated data i.e., 0%RVO+BLP, 5% Random Vertical Outliers and Bad Leverage Point (5%RVO+BLP), 5% Block Concentrated Vertical Outliers and Bad Leverage Point (5%BCVO+BLP), 10% Random Vertical Outlier and Bad Leverage Point (10%RVO+BLP) and 10% Block Concentrated Vertical Outlier and Bad Leverage Point (10%BCVO+BLP).

The considered contamination schemes depending on the types of outliers are described as:

1. To generate random vertical outliers (RVO) for the dependent variable (y_{it}^{r}), randomly selected original values of the dependent variable are replaced by the observations from a normal distribution $y_{it}^{r} \sim \mathrm{N}(0,20)$.
2. To generate random vertical outliers for the independent variable (x_{it}^{r}), randomly selected original values of the independent variable are replaced by the observations from a normal distribution $x_{it}^{r} \sim \mathrm{N}(0,10)$.
3. Block concentrated vertical outliers (BCVO) are inserted into the data by substituting the random observations from a normal distribution $y_{it}^{c} \sim$ N(30, 2) for the randomly selected blocks of the original values of dependent variable.
4. The concentrated bad leverage points (BLP) are inserted into the original blocks of the independent variables corresponding to the blocks already contaminated in dependent variable by replacing them by the random values from a normal distribution $N(\mu = 0, \sigma^2 = 25)$.

The scatter diagrams of the generated panel data confirm that they have the features of the expected panel data. The plots of y_{it} and x_{it} for several panel data show a typical panel data with cross-sectional heterogeneous effect which cannot be model by OLS if no transformation is done to the data and is good for the within group panel data models. The boxplots of the various panel data, y_{it} indicate that the panel data is without outlier even with the classical outlier dictation used by the boxplot. This uncontaminated panel data illustrates the panel data most suitable for the OLS estimator.

The performance of each estimator is evaluated using R = 1000 simulated samples, measured by the mean squared error (MSE) of $\hat{\beta}_1$ and the standard error of the $\hat{\beta}_1$. The MSE of $\hat{\beta}_1$ is

$$MSE = \frac{1}{R}\sum_{r=1}^{R}\left\|\hat{\beta}^{r} - \beta\right\|^{2} \qquad (13)$$

where $\hat{\beta}^{r}$, $r = 1,\ldots,R$, are the estimates for the R simulated samples. Martin (2018) gave expansion of the finite-sample MSE of (13) as $MSE_F(\hat{\theta}_n, \gamma) = Var_F(\hat{\theta}_n, \gamma) + bias_F^2(\hat{\theta}_n, \gamma)$.

Economic Datasets for Illustrations of Finite Sample Linear Panel Data Regression

Dataset 1: The first dataset used in this economic variable empirical illustration has lngdp as the explanatory variable and the imports as the dependent variable obtained from the World Development Index (WDI) 2022 edition. The panel dataset is formed by taking seven countries (cross-sectional units) and twenty-six years (1995-2020). Each of the seven countries are the economic giant of each economic continental block. The seven countries alphabetically arranged are Argentina, Australia, China, France, Israel, Nigeria and United States.

Dataset2: The second economic dataset is GDP Growth (dependent variable) and foreign direct investment (FDI), the independent variable. The panel dataset2 has five countries (cross-sectional units) and twenty years (2001-2020). The five countries are all the Anglophone countries in West

Africa. The five countries alphabetically arranged are Gambia, Ghana, Liberia, Nigeria and Sierra Leone. Dataset2 has all the attributes of influential values. Dataset2 is also sourced from WDI 2022 edition. And because all the five Anglophone countries of West Africa is taken, the fixed effect panel data is most suitable to analyse the dataset2.

IV. RESULTS AND DISCUSSIONS

Table 1: MSE($\hat{\beta}$) and standard error of $\hat{\beta}$ with Mean centering for various for linear static fixed effect panel-data model contaminations

			0%RVO+BLP				5%RVO+BLP				10%RVO+BLP				5%BCVO+BLP				10%BCVO+BLP			
N	T	Parameter	WGOLS	WGWLS	WGM-FIMGT	WGDCML	WGOLS	WGWLS	WGM-FIMGT	WGDCML	WGOLS	WGWLS	WGM-FIMGT	WGDCML	WGOLS	WGWLS	WGM-FIMGT	WGDCML	WGOLS	WGWLS	WGM-FIMGT	WGDCML
5	10	MSE	**0.34**	0.38	1.40	0.34	4.79	4.66	**2.02**	3.70	7.18	7.14	**4.20**	**6.70**	0.69	0.87	24.52	**0.22**	0.90	1.11	1.15	**0.43**
5	10	$s_{\hat{\beta}}$	**0.51**	0.61	0.96	0.59	0.44	3.96	**2.20**	0.43	0.37	3.07	**2.32**	**0.30**	0.76	0.69	8.41	**0.40**	0.81	0.86	0.88	**0.52**
5	30	MSE	**0.11**	0.13	0.42	0.11	6.13	6.14	**1.93**	5.21	7.44	7.40	**3.93**	**7.32**	0.23	0.30	8.68	**0.04**	0.23	0.30	0.34	**0.09**
5	30	$s_{\hat{\beta}}$	**0.29**	0.45	0.62	0.33	0.22	4.90	**2.54**	0.20	0.20	3.17	**2.37**	**0.15**	0.43	0.44	5.89	**0.17**	0.44	0.58	0.59	**0.29**
5	50	MSE	**0.07**	0.08	0.27	0.07	5.89	5.83	**1.61**	4.95	7.41	7.43	**3.85**	**7.35**	0.13	0.15	4.84	**0.02**	0.13	0.17	0.18	**0.05**
5	50	$s_{\hat{\beta}}$	**0.23**	0.42	0.55	0.26	0.17	4.79	**2.37**	0.16	0.15	3.18	**2.38**	**0.12**	0.32	0.32	3.50	**0.13**	0.33	0.51	0.52	**0.22**
10	10	MSE	**0.14**	0.18	0.74	0.14	5.79	5.65	**2.39**	4.05	7.35	7.26	**3.82**	**6.81**	0.32	0.39	9.59	**0.06**	0.36	0.43	0.42	**0.10**
10	10	$s_{\hat{\beta}}$	**0.36**	0.51	0.77	0.42	0.28	4.63	**2.80**	0.28	0.14	3.13	**2.32**	**0.19**	0.53	0.48	5.67	**0.20**	0.55	0.63	0.63	**0.28**
10	30	MSE	**0.05**	0.06	0.24	0.05	6.07	6.03	**2.14**	4.65	7.35	7.32	**3.96**	**7.32**	0.09	0.12	2.16	**0.01**	0.11	0.13	0.13	**0.02**
10	30	$s_{\hat{\beta}}$	**0.21**	0.40	0.52	0.24	0.15	4.88	**2.81**	0.15	0.14	3.16	**2.41**	**0.10**	0.30	0.25	2.94	**0.11**	0.31	0.47	0.47	**0.16**
10	50	MSE	**0.03**	0.04	0.13	0.03	6.14	6.12	**2.04**	4.87	7.41	7.39	**3.92**	**7.38**	0.06	0.07	1.26	**0.01**	0.07	0.08	0.07	**0.01**
10	50	$s_{\hat{\beta}}$	**0.16**	0.35	0.46	0.19	0.12	4.93	**2.78**	0.11	0.11	3.18	**2.41**	**0.08**	0.23	0.21	2.15	**0.08**	0.23	0.42	0.41	**0.12**
40	10	MSE	**0.04**	0.05	0.18	0.04	6.07	5.98	**1.40**	2.73	7.44	7.42	**2.33**	**7.33**	0.06	0.08	0.87	**0.01**	0.07	0.09	0.07	**0.01**
40	10	$s_{\hat{\beta}}$	**0.18**	0.37	0.50	0.22	0.13	4.87	**2.27**	0.15	0.12	3.18	**1.93**	**0.08**	0.25	0.21	1.87	**0.08**	0.26	0.44	0.41	**0.11**
40	30	MSE	**0.01**	0.01	0.06	0.01	6.22	6.20	**1.41**	3.38	7.41	7.41	**2.34**	**7.39**	0.02	0.03	0.36	**0.002**	0.02	0.03	0.02	**0.00**
40	30	$s_{\hat{\beta}}$	**0.10**	0.28	0.40	0.12	0.08	4.97	**2.35**	0.08	0.07	3.18	**1.96**	**0.05**	0.15	0.13	1.12	**0.05**	0.15	0.34	0.32	**0.06**
40	50	MSE	**0.01**	0.01	0.04	0.01	6.23	6.21	**1.38**	3.61	7.45	7.44	**2.36**	**7.44**	0.01	0.02	0.25	**0.001**	0.02	0.02	0.02	**0.002**

40	50	$s_{\hat{\beta}}$	**0.08**	0.25	0.35	0.10	0.06	4.98	**2.33**	0.06	0.05	3.19	**1.97**	**0.04**	0.11	0.10	0.94	**0.04**	0.12	0.31	0.30	**0.05**

Table 2: MSE($\hat{\beta}$) and standard error of $\hat{\beta}$ with Median centering for various for linear static fixed effect panel-data model contaminations

			0% RVO+BLP				5%RVO+BLP				10%RVO+BLP				5%BCVO+BLP				10%BCVO+BLP			
N	T	Parameter	WGOLS	WGWLS	WGM-FIMGT	WGDCML	WGOLS	WGWLS	WGM-FIMGT	WGDCML	WGOLS	WGWLS	WGM-FIMGT	WGDCML	WGOLS	WGWLS	WGM-FIMGT	WGDCML	WGOLS	WGWLS	WGM-FIMGT	WGDCML
5	10	MSE	**0.41**	0.48	2.42	0.46	5.09	4.82	**3.25**	**2.76**	7.37	7.28	**4.02**	**5.43**	0.68	3.07	1.46	**0.34**	0.87	5.39	1.30	**0.36**
5	10	$s_{\hat{\beta}}$	**0.53**	0.78	1.80	0.58	0.44	4.05	**3.07**	**0.43**	0.36	5.29	**3.72**	**0.31**	0.78	2.88	1.51	**0.43**	0.82	4.14	1.40	**0.36**
5	30	MSE	**0.17**	0.20	1.05	0.18	6.48	6.42	**3.46**	**3.67**	7.64	7.58	**4.16**	**6.27**	0.25	3.47	0.75	**0.12**	0.26	5.17	0.63	**0.11**
5	30	$s_{\hat{\beta}}$	**0.30**	0.58	1.42	0.33	0.22	5.01	**3.57**	**0.22**	0.19	5.48	**3.99**	**0.15**	0.43	3.50	1.34	**0.22**	0.44	4.38	1.23	**0.20**
5	50	MSE	**0.14**	0.16	0.76	0.15	6.26	6.16	**3.35**	**3.25**	7.62	7.63	**4.18**	**6.46**	0.16	3.07	0.55	**0.09**	0.16	5.04	0.50	**0.07**
5	50	$s_{\hat{\beta}}$	**0.24**	0.58	1.33	0.25	0.17	4.93	**3.55**	**0.18**	0.15	5.51	**4.02**	**0.11**	0.33	3.35	1.20	**0.17**	0.34	4.40	1.18	**0.15**
10	10	MSE	**0.15**	0.20	0.85	0.15	5.95	5.76	**2.00**	**2.36**	7.40	7.33	**2.51**	**5.15**	0.32	3.04	0.50	**0.11**	0.34	4.84	0.42	**0.09**
10	10	$s_{\hat{\beta}}$	**0.37**	0.43	0.96	0.42	0.28	4.67	**2.54**	**0.32**	0.25	5.37	**2.96**	**0.23**	0.53	3.12	0.75	**0.27**	0.55	4.16	0.75	**0.23**
10	30	MSE	**0.06**	0.07	0.31	0.06	6.24	6.19	**1.89**	**2.14**	7.47	7.43	**2.57**	**6.39**	0.09	3.01	0.19	**0.03**	0.11	4.80	0.16	**0.03**
10	30	$s_{\hat{\beta}}$	**0.21**	0.32	0.65	0.24	0.15	4.95	**2.64**	**0.19**	0.14	5.44	**3.14**	**0.11**	0.30	3.34	0.58	**0.15**	0.31	4.30	0.55	**0.13**
10	50	MSE	**0.04**	0.05	0.20	0.04	6.31	6.28	**1.85**	**2.14**	7.52	7.50	**2.53**	**6.68**	0.06	2.97	0.14	**0.02**	0.07	4.79	0.12	**0.02**
10	50	$s_{\hat{\beta}}$	**0.16**	0.28	0.57	0.18	0.12	4.99	**2.65**	**0.14**	0.11	5.47	**3.14**	**0.08**	0.23	3.36	0.56	**0.11**	0.24	4.33	0.52	**0.10**
40	10	MSE	**0.04**	0.05	0.19	0.04	6.12	6.03	**1.13**	**1.69**	7.46	7.44	**1.66**	**7.00**	0.06	2.77	0.09	**0.02**	0.07	4.65	0.07	**0.01**
40	10	$s_{\hat{\beta}}$	**0.18**	0.19	0.37	0.21	0.13	4.89	**2.02**	**0.17**	0.12	5.45	**2.51**	**0.09**	0.25	3.22	0.29	**0.13**	0.26	4.26	0.26	**0.10**
40	30	MSE	**0.01**	0.02	0.07	0.01	6.26	6.25	**1.14**	**1.92**	7.44	7.44	**1.63**	**7.36**	0.02	2.88	0.04	**0.01**	0.02	4.67	0.03	**0.00**
40	30	$s_{\hat{\beta}}$	**0.11**	0.13	0.27	0.12	0.08	4.99	**2.10**	**0.09**	0.07	5.45	**2.53**	**0.05**	0.15	3.36	0.22	**0.07**	0.15	4.30	0.20	**0.06**

40	50	MSE	**0.01**	0.01	0.04	0.01	6.27	6.25	**1.12**	**1.96**	7.48	7.47	**1.62**	**7.45**	0.01	2.84	0.02	**0.004**	0.02	4.70	0.02	**0.003**
40	50	$s_{\hat{\beta}}$	**0.08**	0.11	0.21	0.10	0.06	5.00	**2.09**	**0.07**	0.05	5.46	**2.53**	**0.04**	0.11	3.35	0.18	**0.06**	0.12	4.32	0.18	**0.05**

Table 3: MSE($\hat{\beta}$) and standard error of $\hat{\beta}$ with MoptM centering for various for linear static fixed effect panel-data model contaminations

			0% RVO+BLP				5%RVO+BLP				10%RVO+BLP				5%BCVO+BLP				10%BCVO+BLP			
N	T	Parameter	WGOLS	WGWLS	WGM-FIMGT	WGDCML	WGOLS	WGWLS	WGM-FIMGT	WGDCML	WGOLS	WGWLS	WGM-FIMGT	WGDCML	WGOLS	WGWLS	WGM-FIMGT	WGDCML	WGOLS	WGWLS	WGM-FIMGT	WGDCML
5	10	MSE	**0.23**	**0.26**	1.15	0.23	**4.71**	4.42	**1.21**	1.09	**7.14**	7.02	**1.67**	2.82	0.54	2.61	0.71	**0.17**	0.71	4.84	0.56	**0.12**
5	10	$s_{\hat{\beta}}$	**0.46**	**0.41**	0.82	0.54	**0.42**	3.84	**1.53**	0.46	**0.36**	5.18	**2.12**	0.41	0.74	2.58	0.61	**0.37**	0.81	3.88	0.56	**0.28**
5	30	MSE	**0.07**	**0.08**	0.34	0.07	**6.12**	6.07	**0.85**	1.08	**7.40**	7.32	**1.26**	3.33	0.18	2.98	0.18	**0.04**	0.19	4.74	0.16	**0.03**
5	30	$s_{\hat{\beta}}$	**0.26**	**0.21**	0.47	0.30	**0.22**	4.87	**1.56**	0.26	**0.19**	5.39	**2.05**	0.21	0.42	3.21	0.34	**0.18**	0.44	4.18	0.29	**0.15**
5	50	MSE	**0.05**	**0.06**	0.21	0.05	**5.89**	5.79	**0.72**	0.84	**7.39**	7.39	**1.15**	3.72	0.10	2.62	0.10	**0.02**	0.11	4.62	0.08	**0.01**
5	50	$s_{\hat{\beta}}$	**0.20**	**0.18**	0.36	0.23	**0.17**	4.77	**1.49**	0.20	**0.15**	5.42	**2.03**	0.15	0.31	3.06	0.23	**0.14**	0.33	4.20	0.22	**0.11**
10	10	MSE	**0.12**	**0.15**	0.62	0.12	**5.78**	5.59	**1.08**	1.28	**7.28**	7.20	**1.44**	3.89	0.29	2.83	0.35	**0.08**	0.32	4.63	0.28	**0.05**
10	10	$s_{\hat{\beta}}$	**0.35**	**0.29**	0.62	0.41	**0.28**	4.60	**1.72**	0.34	**0.25**	5.32	**2.12**	0.27	0.52	2.99	0.43	**0.25**	0.54	4.06	0.41	**0.20**
10	30	MSE	**0.04**	**0.05**	0.20	0.04	**6.07**	6.00	**0.84**	1.09	**7.37**	7.32	**1.29**	5.09	0.08	2.77	0.10	**0.02**	0.10	4.60	0.09	**0.01**
10	30	$s_{\hat{\beta}}$	**0.20**	**0.18**	0.32	0.23	**0.15**	4.87	**1.68**	0.19	**0.14**	5.40	**2.18**	0.13	0.29	3.20	0.24	**0.14**	0.30	4.21	0.21	**0.11**
10	50	MSE	**0.03**	**0.03**	0.11	0.03	**6.14**	6.11	**0.79**	1.15	**7.40**	7.39	**1.25**	5.42	0.05	2.73	0.06	**0.01**	0.06	4.58	0.05	**0.01**
10	50	$s_{\hat{\beta}}$	**0.15**	**0.13**	0.25	0.18	**0.12**	4.92	**1.67**	0.15	**0.11**	5.43	**2.18**	0.09	0.22	3.22	0.19	**0.11**	0.23	4.23	0.16	**0.09**
40	10	MSE	**0.03**	**0.05**	0.18	0.03	**6.07**	5.98	**0.92**	1.37	**7.44**	7.42	**1.38**	6.86	0.06	2.72	0.08	**0.02**	0.07	4.61	0.06	**0.01**
40	10	$s_{\hat{\beta}}$	**0.18**	**0.16**	0.32	0.21	**0.13**	4.87	**1.79**	0.17	**0.12**	5.44	**2.28**	0.09	0.25	3.19	0.23	**0.12**	0.26	4.24	0.20	**0.10**
40	30	MSE	**0.01**	**0.01**	0.06	0.01	**6.22**	6.20	**0.88**	1.57	**7.41**	7.41	**1.32**	7.30	0.02	2.82	0.03	**0.01**	0.02	4.61	0.02	**0.003**

40	30	$s_{\hat{\beta}}$	**0.10**	**0.09**	0.19	0.12	**0.08**	4.97	**1.84**	0.10	**0.07**	5.44	**2.27**	0.05	0.15	3.33	0.14	**0.07**	0.15	4.27	0.11	**0.06**
40	50	MSE	**0.01**	**0.01**	0.03	0.01	**6.22**	6.21	**0.86**	1.57	**7.45**	7.44	**1.32**	7.40	0.01	2.77	0.02	**0.003**	0.01	4.64	0.01	**0.002**
40	50	$s_{\hat{\beta}}$	**0.08**	**0.07**	0.14	0.09	**0.06**	4.98	**1.83**	0.07	**0.05**	5.45	**2.28**	0.04	0.11	3.31	0.10	**0.06**	0.12	4.30	0.10	**0.05**

Table 4: MSE($\hat{\beta}$) and standard error of $\hat{\beta}$ with MM centering for various for linear static fixed effect panel-data model contaminations

			0% RVO+BLP				5%RVO+BLP				10%RVO+BLP				5%BCVO+BLP				10%BCVO+BLP			
N	T	Parameter	WGOLS	WGWLS	WGM-FIMGT	WGDCML	WGOLS	WGWLS	WGM-FIMGT	WGDCML	WGOLS	WGWLS	WGM-FIMGT	WGDCML	WGOLS	WGWLS	WGM-FIMGT	WGDCML	WGOLS	WGWLS	WGM-FIMGT	WGDCML
5	10	MSE	**0.23**	**0.26**	1.15	0.23	4.71	4.42	1.21	**1.09**	0.54	2.61	0.71	**0.12**	7.14	7.02	1.67	**2.82**	0.71	4.84	0.56	**0.12**
5	10	$s_{\hat{\beta}}$	**0.46**	**0.41**	0.82	0.54	0.42	3.84	1.53	**0.46**	0.74	2.58	0.61	**0.28**	0.36	5.18	2.12	**0.41**	0.81	3.88	0.56	**0.28**
5	30	MSE	**0.07**	**0.08**	0.34	0.07	6.12	6.07	0.85	**1.08**	0.18	2.98	0.18	**0.03**	7.40	7.32	1.26	**3.33**	0.19	4.74	0.16	**0.03**
5	30	$s_{\hat{\beta}}$	**0.26**	**0.21**	0.47	0.30	0.22	4.87	1.56	**0.26**	0.42	3.21	0.34	**0.15**	0.19	5.39	2.05	**0.21**	0.44	4.18	0.29	**0.15**
5	50	MSE	**0.05**	**0.06**	0.21	0.05	5.89	5.79	0.72	**0.84**	0.10	2.62	0.10	**0.01**	7.39	7.39	1.15	**3.72**	0.11	4.62	0.08	**0.01**
5	50	$s_{\hat{\beta}}$	**0.20**	**0.18**	0.36	0.23	0.17	4.77	1.49	**0.20**	0.31	3.06	0.23	**0.11**	0.15	5.42	2.03	**0.15**	0.33	4.20	0.22	**0.11**
10	10	MSE	**0.12**	**0.15**	0.62	0.12	5.78	5.59	1.08	**1.28**	0.29	2.83	0.35	**0.05**	7.28	7.20	1.44	**3.89**	0.32	4.63	0.28	**0.05**
10	10	$s_{\hat{\beta}}$	**0.35**	**0.29**	0.62	0.41	0.28	4.60	1.72	**0.34**	0.52	2.99	0.43	**0.20**	0.25	5.32	2.12	**0.27**	0.54	4.06	0.41	**0.20**
10	30	MSE	**0.04**	**0.05**	0.20	0.04	6.07	6.00	0.84	**1.09**	0.08	2.77	0.10	**0.01**	7.37	7.32	1.29	**5.09**	0.10	4.60	0.09	**0.01**
10	30	$s_{\hat{\beta}}$	**0.20**	**0.18**	0.32	0.23	0.15	4.87	1.68	**0.19**	0.29	3.20	0.24	**0.11**	0.14	5.40	2.18	**0.13**	0.30	4.21	0.21	**0.11**
10	50	MSE	**0.03**	**0.03**	0.11	0.03	6.14	6.11	0.79	**1.15**	0.05	2.73	0.06	**0.01**	7.40	7.39	1.25	**5.42**	0.06	4.58	0.05	**0.01**
10	50	$s_{\hat{\beta}}$	**0.15**	**0.13**	0.25	0.18	0.12	4.92	1.67	**0.15**	0.22	3.22	0.19	**0.09**	0.11	5.43	2.18	**0.09**	0.23	4.23	0.16	**0.09**
40	10	MSE	**0.03**	**0.05**	0.18	0.03	6.07	5.98	0.92	**1.37**	0.06	2.72	0.08	**0.01**	7.44	7.42	1.38	**6.86**	0.07	4.61	0.06	**0.01**
40	10	$s_{\hat{\beta}}$	**0.18**	**0.16**	0.32	0.21	0.13	4.87	1.79	**0.17**	0.25	3.19	0.23	**0.10**	0.12	5.44	2.28	**0.09**	0.26	4.24	0.20	**0.10**

40	30	MSE	**0.01**	**0.01**	0.06	0.01	6.22	6.20	0.88	**1.57**	0.02	2.82	0.03	**0.00**	7.41	7.41	1.32	**7.30**	0.02	4.61	0.02	**0.00**
40	30	$s_{\hat{\beta}}$	**0.10**	**0.09**	0.19	0.12	0.08	4.97	1.84	**0.10**	0.15	3.33	0.14	**0.06**	0.07	5.44	2.27	**0.05**	0.15	4.27	0.11	**0.06**
40	50	MSE	**0.01**	**0.01**	0.03	0.01	6.22	6.21	0.86	**1.57**	0.01	2.77	0.02	**0.002**	7.45	7.44	1.32	**7.40**	0.01	4.64	0.01	**0.002**
40	50	$s_{\hat{\beta}}$	**0.08**	**0.07**	0.14	0.09	0.06	4.98	1.83	**0.07**	0.11	3.31	0.10	**0.05**	0.05	5.45	2.28	**0.04**	0.12	4.30	0.10	**0.05**

Table 5: MSE($\hat{\beta}$) and standard error of $\hat{\beta}$ with DCML centering for various for linear static fixed effect panel-data model contaminations

			0% RVO+BLP				5%RVO+BLP				10%RVO+BLP				5%BCVO+BLP				10%BCVO+BLP			
N	T	Parameter	WGOLS	WGWLS	WGM-FIMGT	WGDCML	WGOLS	WGWLS	WGM-FIMGT	WGDCML	WGOLS	WGWLS	WGM-FIMGT	WGDCML	WGOLS	WGWLS	WGM-FIMGT	WGDCML	WGOLS	WGWLS	WGM-FIMGT	WGDCML
5	10	MSE	**0.23**	**0.26**	1.22	0.23	**4.81**	4.58	**1.32**	1.05	**7.10**	6.84	**1.61**	3.12	0.54	2.57	0.75	**0.16**	0.63	4.89	0.58	**0.12**
5	10	$s_{\hat{\beta}}$	**0.46**	**0.41**	0.83	0.54	**0.44**	3.90	**1.55**	0.46	**0.37**	5.11	**2.07**	0.39	0.74	2.48	0.65	**0.36**	0.79	3.94	0.56	**0.28**
5	30	MSE	**0.07**	**0.08**	0.35	0.06	**6.08**	5.92	**0.93**	1.23	**7.31**	7.26	**1.28**	3.37	0.18	2.98	0.18	**0.04**	0.19	4.62	0.15	**0.03**
5	30	$s_{\hat{\beta}}$	**0.26**	**0.21**	0.44	0.30	**0.22**	4.80	**1.67**	0.25	**0.20**	5.36	**2.09**	0.21	0.42	3.21	0.34	**0.18**	0.43	4.15	0.30	**0.15**
5	50	MSE	**0.04**	**0.05**	0.18	0.04	**5.89**	5.80	**0.73**	0.86	**7.34**	7.29	**1.21**	3.75	0.10	2.59	0.11	**0.02**	0.11	4.65	0.08	**0.01**
5	50	$s_{\hat{\beta}}$	**0.20**	**0.16**	0.32	0.23	**0.17**	4.78	**1.50**	0.20	**0.15**	5.39	**2.06**	0.15	0.32	3.05	0.26	**0.14**	0.33	4.22	0.21	**0.11**
10	10	MSE	**0.13**	**0.16**	0.61	0.13	**5.66**	5.45	**1.02**	1.24	**7.24**	7.18	**1.49**	3.90	0.26	2.83	0.34	**0.07**	0.28	4.73	0.27	**0.04**
10	10	$s_{\hat{\beta}}$	**0.35**	**0.30**	0.64	0.41	**0.28**	4.54	**1.65**	0.34	**0.25**	5.32	**2.17**	0.27	0.52	2.98	0.47	**0.25**	0.54	4.10	0.40	**0.20**
10	30	MSE	**0.04**	**0.05**	0.18	0.04	**6.05**	5.98	**0.85**	1.18	**7.39**	7.33	**1.24**	5.05	0.09	2.72	0.10	**0.02**	0.09	4.59	0.08	**0.01**
10	30	$s_{\hat{\beta}}$	**0.20**	**0.18**	0.34	0.23	**0.15**	4.86	**1.70**	0.19	**0.14**	5.40	**2.14**	0.13	0.29	3.18	0.25	**0.14**	0.30	4.20	0.21	**0.11**
10	50	MSE	**0.03**	**0.03**	0.12	0.03	**6.16**	6.11	**0.81**	0.93	**7.41**	7.40	**1.29**	5.58	0.05	2.75	0.06	**0.01**	0.06	4.59	0.04	**0.01**
10	50	$s_{\hat{\beta}}$	**0.15**	**0.15**	0.27	0.18	**0.12**	4.93	**1.70**	0.15	**0.11**	5.43	**2.21**	0.09	0.22	3.23	0.19	**0.10**	0.23	4.24	0.16	**0.09**
40	10	MSE	**0.03**	**0.04**	0.17	0.03	**6.11**	6.03	**0.94**	1.38	**7.38**	7.36	**1.39**	6.87	0.07	2.74	0.09	**0.02**	0.07	4.53	0.07	**0.01**

40	10	$s_{\hat{\beta}}$	**0.18**	**0.17**	0.33	0.21	**0.13**	4.89	**1.83**	0.17	**0.12**	5.42	**2.29**	0.09	0.25	3.20	0.23	**0.12**	0.26	4.20	0.22	**0.10**
40	30	MSE	**0.01**	**0.01**	0.06	0.01	**6.19**	6.17	**0.87**	1.58	**7.43**	7.43	**1.32**	7.27	0.02	2.75	0.03	**0.01**	0.02	4.57	0.02	**0.003**
40	30	$s_{\hat{\beta}}$	**0.10**	**0.09**	0.19	0.12	**0.08**	4.96	**1.83**	0.10	**0.07**	5.45	**2.27**	0.05	0.14	3.28	0.13	**0.07**	0.15	4.26	0.12	**0.06**
40	50	MSE	**0.01**	**0.01**	0.04	0.01	**6.20**	6.19	**0.87**	1.62	**7.44**	7.42	**1.31**	7.36	0.01	2.77	0.02	**0.003**	0.01	4.59	0.01	**0.002**
40	50	$s_{\hat{\beta}}$	**0.08**	**0.07**	0.14	0.10	**0.06**	4.97	**1.84**	0.07	**0.05**	5.45	**2.28**	0.04	0.11	3.31	0.11	**0.06**	0.12	4.27	0.09	**0.05**

Table 6: $\hat{\beta}_1$ and its Standard Error for the lnimports/lngdp Panel Data having 0%, 5%, 10% BCVO, RVO, BLP and GDPgrowth/FDI Panel Data

Outliers & percentages	Parameter Estimators	Panel Data Within Group Estimator									
		WGOLS					WGDCML				
		Mean	Median	MoptM	MM	DCML	Mean	Median	MoptM	MM	DCML
0% Outliers	$\hat{\beta}_1$	0.509	0.502	**0.585**	**0.585**	**0.585**	0.509	0.502	0.585	0.585	0.585
	$\hat{\sigma}_{\hat{\beta}_1}$	0.074	0.074	**0.073**	**0.073**	**0.073**	0.095	0.095	0.089	0.089	0.089
5% RVO +BLP	$\hat{\beta}_1$	0.008	0.006	-0.005	-0.005	-0.005	0.036	0.368	**0.413**	**0.413**	**0.413**
	$\hat{\sigma}_{\hat{\beta}_1}$	0.049	0.048	0.049	0.049	0.049	0.028	0.074	**0.068**	**0.068**	**0.068**
10% RVO +BLP	$\hat{\beta}_1$	-0.006	-0.009	-0.012	-0.012	-0.012	0.008	0.022	**0.021**	**0.021**	**0.021**
	$\hat{\sigma}_{\hat{\beta}_1}$	0.033	0.035	0.037	0.037	0.037	0.023	0.014	**0.014**	**0.014**	**0.014**
5% BCVO +BLP	$\hat{\beta}_1$	0.012	0.028	0.034	0.034	0.034	0.044	0.426	**0.514**	**0.514**	**0.514**
	$\hat{\sigma}_{\hat{\beta}_1}$	0.049	0.046	0.047	0.047	0.047	0.029	0.079	**0.075**	**0.075**	**0.075**
10% BCVO +BLP	$\hat{\beta}_1$	-0.018	-0.027	-0.018	-0.018	-0.018	0.033	0.049	**0.066**	**0.066**	**0.066**
	$\hat{\sigma}_{\hat{\beta}_1}$	0.070	0.069	0.071	0.071	0.071	0.049	0.030	**0.029**	**0.029**	**0.029**
GDP growth and FDI	$\hat{\beta}_1$	-0.048	-0.042	-0.011	-0.011	0.011	-0.019	0.004	**0.026**	**0.026**	**0.026**
	$\hat{\sigma}_{\hat{\beta}_1}$	0.037	0.037	0.038	0.038	0.038	0.025	0.018	**0.023**	**0.023**	**0.023**

Tables 1-5 give the results of the simulated panel data having Mean, Median, MoptM, MM and DCML centering estimators. The results show the following for the various centering estimators: OLS outperformed the other estimators when the panel data has uncontaminated data in estimating the beta and its standard error; for panel data with 5% RVO+BLP, the GM-FIGMT outperformed the other estimators in the mean square of beta but the DCML (for smaller sample sizes of less than 1000) and OLS (for larger sample sizes of greater than 1000) yield the smallest standard error of beta-hat; for 10% RVO+BLP contamination, the GM-FIGMT yields the lower MSE($\hat{\beta}$) values, but the DCML has the lowest standard error of the beta-hat; for 5% and 10% BCVO+BLP contaminated panel data, DCML regression estimator outperformed the other regression estimators in estimating beta and standard error of beta-hat.

Table 6 gives the results of the real-life economic datasets (lnimports/lngdp panel dataset1 having 0%, 5%, 10% RVO+BLP, BCVO+BLP and also for the GDP growth/FDI panel dataset2). For the panel dataset1 having 0%, 5%, 10% RVO+BLP, BCVO+BLP, the MoptM, MM and DCML centering of OLS and DCML regression estimators yield the largest estimated values of beta-hat and the smallest standard error values for the various panel data contaminations against the Mean and Median centering. For the GDP growth and FDI dataset, the DCML estimator (centering and within group) recovered the positive effect of FDI on GDP growth which the Mean, Median, MoptM and MM centering estimators of OLS lost as their $\hat{\beta}_1$ changed from -0.048, -0.042, -0.011 and -0.011 to 0.011. the DCML regression estimator recovered the positive effect of FDI on GDP growth for the Median, MoptM and MM centering estimators, but couldn't recover it for the Mean centering. Also, the standard errors of $\hat{\beta}_1$ reduced from 0.038 using the OLS to 0.023 using the WGDCML.

V. CONCLUSION AND RECOMMENDATIONS

This paper compares the robustness of the linear regression estimators for finite sample fixed effect panel model having influential values using five centering estimators and four regression estimators. The mean square error of the beta-hat and standard error of beta-hat were the goodness of fit statistics used in the comparison. Monte Carlo simulation and real-life datasets analysis of panel data regression were performed and the following conclusions were drawn from the study: when the panel data contamination is bi-directional (X and Y), and the panel contamination is block concentrated vertical outlier and leverage point, the DCML location and regression estimators yield the best robustness results for all finite samples panel data; when the panel data contamination is bi-directional (X and Y), and the panel contamination is having random vertical outlier, the GM-FIGMT yields the best robustness parameter estimates but fails in the standard error estimation for finite samples panel data which the DCML outperformed it. Hence, for bi-directional finite samples panel contamination having block concentrated vertical outlier and bad leverage point, we recommend the use of DCML regression estimator to obtain the best robust beta estimates and their standard errors; for bi-directional panel contamination having random vertical outlier, we recommend the use of GM-FIGMT estimator to get the best robust parameter estimates for finite samples panel data.

REFERENCES

Abu Bakar N. M. and Habshah M. (2018). The Applications of Robust Estimation in Fixed Effect Panel Data Model. Advances in Social Science, Education and Humanities Research, volume 292

Aquaro, M. & P. Čížek (2010). One-step robust estimation of fixed effects panel data models. *Computational Statistics and Data Analysis,* 57 (1), 536–548.

Arellano, M. (2003). Panel Data Econometrics, Oxford: Oxford University Press.

Armstrong, T. B., Weidner, M. & Zeleneev, A. (2023). Robust estimation and inference in panels with interactive fixed effects. The Institute for Fiscal Studies, Department of Economics. Centre for Microdata Methods and Practice working paper CWP14/23.

Bakar, N. M. A. & Midi, H. (2015). Robust Centering in the fixed effect panel data model. *Pakistan Journal of Statistics,* 31, 33-48.

Baltagi, B. H. (2021). Econometric Analysis of Panel Data, 6th edition. Springer Nature: Switzerland.

Beyaztas, B. H. & Bandyopadhyay, S. (2020). Data Driven Robust Estimation Methods for Fixed Effects Panel Data Models. *Journal of Statistical Computation and Simulation*, 92(7), 1-25.

Bounds, J., Jaeger, D. A. & Baker, R. M. (1995). Problems with instrumental variables when the correlation between the instruments and the endogenous explanatory variables is weak. *Journal of the American Statistical Association,* 90(430).

Bramati, M. C. & Croux, C. (2007). Robust estimators for the fixed effects panel data model. *Econometrics Journal*, 10 (3), 521–540.

Chesher, A., Rosen, A. M. & Zhang, Y. (2024). Robust Analysis of Short panels. The Institute for Fiscal Studies Department of Economics, UCL. Centre for Microdata Methods and Practice working paper CWP01/24.

Chiang, J. T. (2007). The Masking and Swamping Effects Using the Planted Mean-Shift Outliers Models. *International Journal of Contemporary Mathematical Sciences*, 2(7), 297 – 307.

Croissant, Y. & Millo, G. (2019). Panel Data Econometrics with R. John Wiley & Sons, Inc., New York, USA.

Druten, J. (2020). On robust instrumental variables estimator performance in contaminated environments. Unpublished thesis Bachelor Thesis, Erasmus School of Economics, programme Econometrics and Operations Research.

Ebrima, K. C. & Yahaya, M. M. (2022). Pooled ordinary least-square, fixed effects and random effects modelling in a panel data regression analysis: a consideration of international commodity price and economic growth indicators in 35 Sub-Saharan African countries. *International Journal of Technology Transfer and Commercialisation*, 19(1).

Eke, E. M., I. D. Essi, C. A. Nse (2024) Robust Location-Scale Estimator In Static Linear Finite Sample Fixed Effect Panel Model With Outliers International Journal of Innovative Mathematics, Statistics & Energy Policies 12(2):22-40

Feng, G., Gao, J. & Peng, B. (2022). Multi-Level Panel Data Models: Estimation and Empirical Analysis Working Paper No.4/22 Department of Econometrics and Business Statistics Working Paper Series.

Gervini, D. & Yohai, V. J. (2002). A class of robust and fully efficient regression estimators. Technical Report. Unpublished. Available from the first author at request, or via anonymous ftp at ulises.ic.fcen.uba.ar (files: rewlse.dvi, rewlse.ps, rewlse.pdf).

Gil-García, J. R. & Puron-Cid, Y. G. (2015). Using Panel Data Techniques for Social Science Research: An Illustrative Case and Some Guidelines. *Network of Scientific Journals from Latin America, the Caribbean, Spain and Portuga*l, 21(3), 203-216.

Hsiao, C. (2022). Analysis of Panel Data, 4^{th} Edition. Cambridge University Press. United Kingdom.

Huber P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35, 73–101.

Huber, P. J. (1981). Robust Statistics. Wiley, New York.

Ibrahim, M. & Arundina, T. (2022). Practical Panel Modelling. National Committee of Islamic Economy and Finance, Indonesia.

Ismaeel, S. S. & Midi, H. (2018). Robust Within Group Estimator for Fixed Effect Panel Data. *Pakistan Journal of Statistics*, 34(4), 297-310.

Jaba, E., Robu, I. & Balan, C. B. (2017). Panel data analysis applied in financial performance assessment. Romania Statistical review nr.2

Least absolute deviations. (2021). In wikipedia https://en.wikipedia.org/wiki/Least_absolute_deviations

Lee, G. M., Yeon, K. & Kim, H. (2016). A Comparative Study on Robust Regression Methods. *International Advanced Research Journal in Science, Engineering and Technology*, 3(8), 166-170.

Maronna, R. A. & Yohai, V. J. (2014). High finite-sample efficiency and robustness based on distance-constrained maximum likelihood. *Computational Statistics and Data Analysis*, 83, 262–274.

Maronna, R. A., Martin, R. D., Yohai, V. J. & Salibi´an-Barrera, M. (2019). Robust Statistics: Theory and Methods (with R). Wiley, New York.

Martin, R. D. (2018). Robust Statistics for Quantitative Finance. Online Tutorial and Lecture Note. www.rinfinance.com.

Martin, R. D. (2021). Optimal Bias Robust Regression Psi and Rho Revisited. Electronic copy available at: https://ssrn.com/abstract=3902862.

Martin, R. D., Yohai, V. J. & Zamar, R.H. (1989). Min-max bias robust regression. *Annals of Statistics*, 17, 1608-1630.

Midi, H. & Mohammad, S. (2018). Robust estimation for fixed and random effects panel data models with different centering methods. *Journal of Engineering and Applied Sciences* 13(17), 7156-7161.

Millimet, D. L. & Bellemare, M. F. (2023). Fixed Effects and Causal Inference. Southern Methodist University and University of Minnesota. New York, USA.

Pesaran, M. H. (2015). Time Series and Panel Data Econometrics, First Edition. Oxford University Press, New York, United States of America.

Riazoshams, H., Midi, H. & Ghilagaber, G. (2019). Robust Nonlinear Regression: with Applications using R. John Wiley & Sons Ltd, New York USA.

Robust regression. (2021, July 08). In Wikipedia. https://en.wikipedia.org/wiki/Robust_regression

Rousseeuw P. J. & Yohai V. (1984). *Robust Regression by Means of S estimators, in Robust and Nonlinear Time Series Analysis*. Lecture Notes in Statistics 26. New York: Springer Verlag, 256–274.

Rousseeuw, P. J. & Huber, M. (2011). Robust statistics for outlier detection John Wiley & Sons, Inc, New York.

Rousseeuw, P. J. & Van-Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, 85, 633–651.

Rousseeuw, P. J. & Wagner, J. (1994). Robust regression with a distributed intercept using least median of squares. *Computational Statistics and Data Analysis,* 17, 65–76.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of American Statistical Association*, 79, 871-880.

Rousseeuw, P. J. (2019). Robust Statistics Part 1: Introduction and univariate data, LARS-IASC School Online Lecture Note.

Rousseeuw, P. J. and Leroy A. (1987). Robust Regression and Outlier Detection. Wiley, New York.

Shu, V. S. (1978). Robust estimation of a location parameter in the presence of outliers. Retrospective Theses and Dissertations. 6419.

Svarc, M., Yohai, V. J. & Zamar, R. H. (2002). Optimal Bias Robust M-Estimates of Regression. In*: Statistical Data Analysis Based on the L1-Norm and Related Methods*. Springer, 191–200.

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. Pages 444-485 I. Olkin, ed. Contributions to probability and statistics. Stanford University Press, Stanford, California.

Varin, S. (2021). Comparing the Predictive Performance of OLS And Seven Robust Linear Regression Estimators on A Real and Simulated Datasets. *International Journal of Engineering Applied Sciences and Technology*, 5(11), 9-23.

Wagenvoort, R. & Waldmann, R. (2002). On B-robust instrumental variable estimation of the linear model with panel data. *Journal of Econometrics,* 106. 297-324.

Wooldridge, J. M. (2010). Econometric Analysis of Cross Section and Panel Data, 2nd Edition, The MIT Press, London.

Wooldridge, M. (2021). Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators Jeffrey, Department of Economics Michigan State University.

Yadolah, D. (2008). The Concise Encyclopedia of Statistics (1st ed. 2008). New York, Springer.

Yaffee, R. (2003). A Primer for Panel Data Analysis. New York university information services.

Yohai V. J. (1987). High Breakdown Point and High Efficiency Robust Estimates for Regression, *Annals of Statistics*, 15, 642–656.

Yohai, V. & Zamar, R. (1988). High breakdown point estimates of regression by means of the minimization of an efficient scale. *Journal of the American Statistical Association*, 83, 406 – 413.

Yohai, V. J. & Zamar, R. H. (1997). Optimal locally robust M-estimates of regression. *Journal of Statistics, Planning and Inference*, 64, 309-323.

Zaman, A., Rousseeuw, P. J. & Orhan, M. (2001). Econometric applications of high-breakdown robust regression techniques. *Economics Letters,* 71 (1), 1–8.

Zhang, J. & Li, G. (1998). Breakdown Properties of Location M-Estimators. *The Annals of Statistics*, 26(3), 1170–1189.