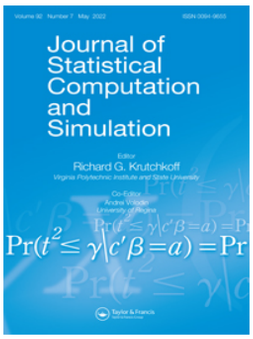

Taylor & Francis
Taylor & Francis Group

Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: www.tandfonline.com/journals/gscs20

Data driven robust estimation methods for fixed effects panel data models

Beste Hamiye Beyaztas & Soutir Bandyopadhyay

To cite this article: Beste Hamiye Beyaztas & Soutir Bandyopadhyay (2022) Data driven robust estimation methods for fixed effects panel data models, Journal of Statistical Computation and Simulation, 92:7, 1401-1425, DOI: 10.1080/00949655.2021.1996576

To link to this article: https://doi.org/10.1080/00949655.2021.1996576

Published online: 27 Oct 2021.

Submit your article to this journal

Article views: 328

View related articles

View Crossmark data

Citing articles: 7 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=gscs20

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION
2022, VOL. 92, NO. 7, 1401–1425
https://doi.org/10.1080/00949655.2021.1996576

Taylor & Francis
Taylor & Francis Group

Check for updates

Data driven robust estimation methods for fixed effects panel data models

Beste Hamiye Beyaztas [a] and Soutir Bandyopadhyay[b]

[a]Department of Statistics, Istanbul Medeniyet University, Istanbul, Turkey; [b]Department of Applied Mathematics and Statistics, Colorado School of Mines, Golden, CO, USA

ABSTRACT

The panel data regression models have gained increasing attention in different areas of research including econometrics, environmental sciences, epidemiology, behavioural and social sciences. However, the presence of outlying observations in panel data may often lead to biased and inefficient estimates of the model parameters resulting in unreliable inferences when the least squares method is applied. We propose extensions of the M-estimation and Exponential squared loss function-based approaches with a data-driven selection of tuning parameters to achieve desirable level of robustness against outliers without loss of estimation efficiency. The consistency and asymptotic normality of the proposed estimators have also been proved under some mild regularity conditions. The finite-sample performance of our proposed methods have been examined via several Monte Carlo experiments and their results are compared with the ones from existing methods. In addition, a macroeconomic dataset is analysed using the proposed methods to demonstrate their superiorities.

ARTICLE HISTORY
Received 27 November 2020
Accepted 18 October 2021

KEYWORDS
Panel data; fixed effects; robust estimation; M-estimation; least squares

1. Introduction

Panel data refers to the two-dimensional data in which cross-sectional units are observed over time [1]. Its grouping structure allows to reflect the nested phenomena so that the characteristics of cross-sectional units are entrenched over time and vice-versa (see, [2]). Over last several years, its increasing availability, demanding methodology and better ability to model the complexity of human behaviour than a pure cross-section or time series data are the primary reasons behind the excessive growth in its study (see, [3]). Consult [4–12] for a comprehensive account of the technical details on linear panel data models. In recent years, the panel data regression models have received increasing attention in the research of different fields such as econometrics and biostatistics, since these models allow the unobserved individual-specific heterogeneity to be taken into account (cf. [4,13] for more details). The statistical appeal of panel data models relies on the fact that these focus particularly on explaining within variations over time and provide controls over individual

CONTACT Beste Hamiye Beyaztas beste.sertdemir@medeniyet.edu.tr

© 2021 Informa UK Limited, trading as Taylor & Francis Group

heterogeneity [1]. The most widely used regression techniques for making statistical inferences regarding the parameters of linear panel data regression models are typically based on the least squares (LS) approaches. However, traditional estimation techniques require a number of quite restrictive and unrealistic assumptions such as the normality of error distribution, strict exogeneity with respect to the error terms and homoscedasticity of the error terms (cf. [14,15]). Also, the panel data may include various types of outliers such as either vertical, horizontal or leverage as noted in [16,17]. Furthermore, the outlying observations may be concentrated in some blocks such that the fraction of contaminated values per one cross-sectional unit constitutes at least a half of the observations over time periods (cf. [18,19]). Hence, the classical ordinary least squares (OLS)-based methods may considerably be affected in the presence of data contamination and outliers caused by the measurement error, typing error, transmission/copying error and naturally unusual observations as noted in [16,20,21]. As a result of this, the well-known estimators, such as within group LS estimators used in panel data models with fixed effects, often lead to unreliable estimates of the model parameters. In addressing the problem, more robust alternatives to LS methods having a high breakdown point (BP), such as Least Trimmed Squares (LTS) and S-estimators have been introduced by Rousseeuw [22] and Rousseeuw and Yohai [23] for linear regression models. One of the various measures characterizing the robustness of the estimator is the breakdown point which measures the minimum proportion of the data that can significantly change the estimates (cf. [18,24,25]). As pointed out by Rousseeuw and Leroy [26], the asymptotic breakdown point of the LS estimator is zero in the presence of contaminated data sets. Hence, for contaminated data sets, the importance using robust and positive breakdown point methods in estimating model parameters has been emphasized by many authors; see, for instance, [20,27–32]. This is more crucial for panel data since the outlying data points may be masked and not be directly detectable using standard outlier diagnostics due to the complex structure of the data. In spite of the fact that building up the robust methods has been well-studied in estimation of the linear regression parameters, the number of available approaches on the robust methods for static panel data models is fairly limited (cf. [18,19]). A few different approaches within the robust estimation framework for the fixed effects panel data models have recently been proposed by Beyaztas and Bandyopadhyay [1], Bakar and Midi [16], Aquaro and Cizek [18], Bramati and Croux [19], Midi and Muhammad [33], Namur and Luneburg [34] and Visek [35]. Bramati and Croux [19] have proposed the robust alternatives to the within group LS estimator with positive breakdown point by extending well-known robust regression estimators, such as LTS estimator of Rousseeuw [22] and a combination of M and S estimates, namely, MS estimates of Maronna and Yohai [36]. Another robust estimation approach has been proposed in [18] based on two different data transformations (i.e. first-difference and pairwise-difference transformation) by applying the efficient weighted least squares estimator of Gervini and Yohai [37] and the reweighted LTS estimator of Cizek [38] in the fixed effects linear panel data framework. Visek [35] has developed a robust algorithm by weighting down the large-order statistics of squared residuals. Moreover, Bakar and Midi [16] have considered a robust centring method by employing the MM-centring procedure to the data. Then, the authors use the within group Generalized M Estimator (WGM) of Bramati and Croux [19] to estimate the parameters of fixed effect panel data model. A Weighted Least Square (WLS) procedure based on MM-centring method, which is highly resistant in the presence of leverage points and vertical outliers, has been proposed by Midi and Muhammad [33].

More recently, the weighted likelihood-based robust versions of the existing traditional panel data estimation methods have been proposed by Beyaztas and Bandyopadhyay [1]. In this study, we concentrate on a class of robust estimators, i.e. M-estimators, and also, an Exponential squared loss function-based method, instead of investigating different data transformations for fixed effects panel data model. A dispersion function (also called the *loss function*), that varies at large values more slowly in comparison to the squared function of the residuals, is attempted to be minimized in M-estimation approaches. However, the robustness against outliers is achieved by some efficiency loss when unnecessarily resistance occurs. (e.g. [29,39–41]). Hence, it is critical to choose a dispersion function with an appropriate resistance level in obtaining efficient estimates of the parameters as noted in [41]. To determine the necessary level of robustness, a tuning parameter (also called the regularization parameter, tuning constant) in the dispersion function requires to be selected appropriately based on the possible proportion of outliers (please, see [41,42] for more details). For some dispersion functions, such as Tukey's bisquare and Huber's functions, a data-driven procedure that automatically chooses the value of the tuning parameter has been introduced by Wang et al. [41,42] in the context of regression models. The main idea behind this method is to achieve desirable level of robustness against outliers without sacrificing estimation efficiency. Also, a data-dependent procedure to choose the value of tuning parameter in Exponential squared loss function has been proposed by Wang et al. [43] in obtaining penalized robust estimators with maximum efficiency and asymptotic breakdown point 1/2. In this paper, we adapt the M-estimation and Exponential squared loss function-based techniques with automatic selection of tuning parameter introduced by Wang et al. [41,43] to fixed effects panel data models to achieve resistant estimation against outliers as well as improving estimation efficiency. To construct robust alternative procedures based on Huber's, Tukey's bisquare and Exponential loss functions to the within group LS estimator, we use within group transformation on mean centred data to eliminate the individual effects. Thus, we do not provide the detailed explanations of the robust methods based on the different data transformations proposed by Aquaro and Cizek [18]. The asymptotic properties of the proposed M-estimation and the Exponential loss function-based methods are investigated by establishing the consistency and asymptotic normality in the context of fixed effects linear panel data models. Also, to investigate the finite sample performances of the proposed robust estimators, Monte Carlo experiments are carried out under various contamination schemes and two levels of contamination. Furthermore, the effects of different types of outliers including vertical outliers and leverage points on the proposed estimation procedures have been examined. The numerical results demonstrate that the proposed robust estimators based on data-driven procedures yield more accurate and precise estimates of the model parameters compared to within group LS estimator, within group MS estimator (WMS) of Bramati and Croux [19], the weighted fixed effects estimator (WFE) of Beyaztas and Bandyopadhyay [1], and the Forward Search (FS) approach of Atkinson et al. [27] for the increasing level of contamination, in general. Moreover, the predictive ability of the models constructed by estimating coefficients using our proposed methods is better than those of WMS, WFE, and FS methods. The remainder of the paper is organized as follows. In Section 2, we present a detailed information on the fixed effects linear panel data models and within group LS estimator, followed by a discussion on the existing robust estimation procedures in Section 3. In Section 4, we describe our method to obtain the proposed estimators based on Tukey's bisquare,

Huber's and Exponential loss functions with a data-dependent tuning in the context of linear fixed effects panel data models. In Section 5, we present the large sample properties of the proposed estimators. The finite sample performances of the proposed estimators are provided by an extensive simulation study and the results are compared with the robust WMS, WFE, FS, and traditional LS methods in Section 6. In Section 7, we apply proposed robust methods to real macroeconomic data. Finally, Section 8 concludes the work with a few remarks.

2. Fixed effects panel data models

A linear fixed-effects panel data model with a random sample $\{(y_{it}, x_{it}, \alpha_i),\ i = 1, \ldots, N,\ t = 1, \ldots, T\}$ can be represented as

$$y_{it} = x_{it}^T \beta + \alpha_i + \varepsilon_{it},$$

where y_{it} denotes the response variable, the K-dimensional random explanatory variables are denoted by x_{it}'s and $\beta \in \mathbb{R}^K$ is the vector of regression parameters. The subscript i denotes the individuals observed over time periods t. Finally, the α_i's and ε_{it}'s, respectively, represent the unobservable individual specific effects and the independent and identically distributed (i.i.d.) error terms with $E(\varepsilon_{it}|x_{i1}, \ldots, x_{iT}, \alpha_i) = 0, E(\varepsilon_{it}^2|x_i, \alpha_i) = \sigma_\varepsilon^2 I_T$ where I_T is the identity matrix and $E(\varepsilon_{it}\varepsilon_{is}|x_i, \alpha_i) = 0$ for $t \neq s$. For simplifying notation, the panel data regression model can be expressed in more compact form, by stacking observations over time and individuals, as in Equation (1) given below.

$$y = \alpha \otimes e_T + \mathbf{X}\beta + \varepsilon, \tag{1}$$

where $y = (y_1, \ldots, y_N)^T$ and $\mathbf{X} = (x_1, \ldots, x_N)^T$, respectively, are an $NT \times 1$ vector with $y_i = (y_{i1}, \ldots, y_{iT})$ and an $NT \times K$ matrix of regressors with $\mathbf{X}_i = (x_{i1}^T, \ldots, x_{iT}^T)^T$, α is an $N \times 1$ vector consisting of the individual effects α_i for $i = 1 \cdots N$, e_T is a $T \times 1$ vector of ones and $\otimes$ denotes the Kronecker product. An important advantage of the fixed effects model is the elimination of the individual effects from the model equation when estimating the parameter vector, β. The within group transformation removes the fixed effects by using the time averages of y_{it}, x_{it} and ε_{it} for each-cross sectional unit: $\bar{y}_{i.} = T^{-1}\sum_{t=1}^T y_{it}$, $\bar{x}_{i.} = T^{-1}\sum_{t=1}^T x_{it}$ and $\bar{\varepsilon}_{i.} = T^{-1}\sum_{t=1}^T \varepsilon_{it}$. Then, the within group transformed model for the mean-centred data is obtained as follows.

$$\ddot{y}_{it} = \ddot{x}_{it}^T \beta + \ddot{\varepsilon}_{it}$$

where $\ddot{y}_{it} = y_{it} - \bar{y}_{i.}$, $\ddot{x}_{it} = x_{it} - \bar{x}_{i.}$ and $\ddot{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_{i.}$. Under the assumptions of fixed effects models, the within group LS estimator of β, $\hat{\beta}_{(LS)}$ is obtained by running regression of $\ddot{y}_{it}$ on $\ddot{x}_{it}$ using OLS method, as follows.

$$\hat{\beta}_{(LS)} = \left(\sum_{i=1}^N \sum_{t=1}^T \ddot{x}_{it}^T \ddot{x}_{it}\right)^{-1} \left(\sum_{i=1}^N \sum_{t=1}^T \ddot{x}_{it}^T \ddot{y}_{it}\right)$$

The within group estimator fulfills three equivariance properties with respect to scale, regression and affine transformations since it is linear. Suppose $R(\{x_{it}, y_{it}\})$ denotes a panel regression estimator as a function of data.

Definition 2.1 ([19]): If, for any constant $c \in \mathbb{R}$,

$$R\left(\{x_{it}, cy_{it}\}\right) = cR\left(\{x_{it}, y_{it}\}\right)$$

the estimator R is scale equivariant.

Definition 2.2 ([19]): If, for any $K \times 1$ vector of constants ν,

$$R\left(\{x_{it}, y_{it} + x_{it}^T \nu\}\right) = R\left(\{x_{it}, y_{it}\}\right) + \nu$$

it is regression equivariant.

Definition 2.3 ([19]): If it follows for non-singular $A \in \mathbb{R}^{K\times K}$

$$R\left(\{A^T x_{it}, y_{it}\}\right) = A^{-1} R\left(\{x_{it}, y_{it}\}\right)$$

then, this estimator satisfies the affine equivariance property.

The within group LS estimator defined above is known to be highly sensitive to the presence of outliers and aberrant observations. The motivation of this work is to construct robust estimation procedures which are less sensitive in the presence of outliers and erroneous observations. In this paper, we propose extensions of the data-dependent approaches proposed by Wang et al. [41–43] for obtaining robust and more efficient estimates in fixed effects panel data models. This study considers three approaches: the first one is based on exponential squared loss (ESL) function suggested by Wang et al. [43] to improve the robustness of variable selection procedures in context of penalized regression methods. Also, they propose a class of penalized robust regression estimators based on ESL and a data-driven procedure, which allows to select a tuning parameter depending on the proportion of outliers in the data. These procedures yield highly robust and also, efficient estimates by achieving the highest asymptotic breakdown point of 1/2. In other approaches, we propose to use Huber's function and Tukey's bisquare function with data-dependent regularization parameters to obtain more efficient estimates within the fixed effects models framework. Wang et al. [41,42] have proposed a data-driven method for the automatic selection of a tuning constant that can be applied to the various dispersion functions such as the Huber, Tukey's bisquare and ESL functions for regression models. They obtain considerably better results by improving the efficiency in estimating the regression parameters. Next, we briefly discuss the existing robust estimators and describe our proposed method to estimate the fixed effects panel data models.

3. Robust estimation for fixed effects panel data models

The existing literature regarding the robustness of estimators for static fixed effect panel data models is fairly limited. For the purpose of constructing highly robust procedures, Bramati and Croux [19] propose the WGM estimator and the WMS estimator, which have asymptotically breakdown point of 1/4. In the first approach, Bramati and Croux [19] suggest robustly centring the variables by using the within group medians (as described in Equation (2)).

$$\tilde{y}_{it} = y_{it} - \underset{t}{\text{median}}\left(y_{it}\right)$$

$$\tilde{x}_{it} = x_{it} - \underset{t}{\text{median}}\,(x_{it}) \tag{2}$$

Next, LTS regression is suggested to employ on the centred data to obtain initial estimates. Finally, a weighted LS estimation, which uses Tukey's bisquare function with the fixed tuning parameter $c = 4.685$ and a multivariate S-estimator for down-weighting of leverage points, is performed to construct the WGM estimator. Although the WGM estimator can achieve the breakdown point up to 1/4, it has crucial limitations of not being regression and affine equivariant because of the non-linearity of median transformation. To construct the WMS estimator, the fixed effects are first estimated by considering them as regression coefficients. Then, the MS regression estimator of Maronna and Yohai [36] can be implemented to the panel data, which uses M-estimators and S-estimators for the discrete explanatory variables and the continuous ones, respectively. It has been noted that the WMS, which has the advantages of being regression and affine equivariant, and the WGM procedure yield quite similar efficient estimation asymptotically. Thus, we compare our proposed procedures with the WMS estimator by considering its superiority over the WGM estimator. Furthermore, Aquaro and Cizek [18] propose robust estimation procedures based on the pairwise difference transformations by applying the efficient weighted LS estimator (REWLS) of Gervini and Yohai [37] and the reweighted LTS (RLTS) estimator of Cizek [38]. Since the novelty of their approaches is to use different data transformations, the authors demonstrate the equivariance, robust, and asymptotic properties of the proposed estimators. The finite sample performances of the WGM and WMS estimators of Bramati and Croux [19] and the REWLS and RLTS estimators of Aquaro and Cizek [18] are similar and they have equal breakdown points according to a given data transformation, see [18]. Hence, the numerical results discussed in Section 5 do not include the results related to the performances of the methods proposed by Aquaro and Cizek [18]. Recently, Beyaztas and Bandyopadhyay [1] have proposed a weighted likehood-based robust estimation procedure for the fixed effects and random effects panel data models. The authors use the weighted likelihood estimating equations in which the weights are obtained from the Hellinger Residual Adjustment Function of Pearson residuals to estimate model parameters. The WFE estimator of Beyaztas and Bandyopadhyay [1], one of the proposed estimators for the fixed effects panel data model, can attain a 1/2 breakdown point by choosing a root providing minimum disparity measure between kernel density estimate of the model and smoothed model density. Hence, the performance of the proposed estimators are compared with that of WFE estimator in numerical results.

4. New robust estimators using well-known loss functions with automatic selection of tuning parameter

A core task in presence of the contaminated dataset is to figure out a way to achieve robustness while keeping efficiency high in estimation for making valid and reliable inferences (cf. [42]). A broad class of robust estimation methods is the M-estimation approach which is based on minimizing dispersion function that more slowly varies at large values compared to the squared function of the residuals (see [41] for details). In the M-estimation approach, the robustness is achieved by sacrificing some of efficiency in presence of unnecessarily excess resistance (e.g. [29,39–41]). To obtain necessary degree of robustness, a tuning parameter in the dispersion function requires to be chosen appropriately depending

on the possible proportion of outliers, as noted in [41,42]. To this end, for a specific family of dispersion functions, a data-driven approach that automatically selects the value of the tuning constant has been introduced by Wang et al. [41] in the context of regression models. The main idea underlying this method is to achieve desirable robustness level in presence of outliers without loss of estimation efficiency. Moreover, Wang et al. [42] extend this procedure for linear regression models with autoregressive errors in analysing water quality data. In this study, we focus on the robust estimation approaches based on loss functions with automatic selection of tuning parameter proposed by Wang et al. [41,43] to achieve resistant estimation against outliers and improve estimation efficiency in fixed effects panel data models. To construct alternative procedures to the methods discussed in Section 3, we use within group transformation for mean centred data to eliminate the individual effects. Thus, we do not provide the numerical results and the detailed explanations of the robust methods based on the different data transformations proposed by Aquaro and Cizek [18]. The M-estimation approach relies on minimization of a loss function, $\rho(\cdot)$, instead of minimizing sum of squared errors since the LS method is highly sensitive to the distortions caused by outliers [21]. The Huber and Tukey's bisquare functions, which are the most commonly used loss functions in robust regression, are defined as in Equations (3)–(4), respectively. (see [42])

- Huber's function

$$\rho(u) = \begin{cases} \dfrac{u^2}{2} & \text{if } |u| \leqslant c \\ c|u| - \dfrac{c^2}{2} & \text{if } |u| > c \end{cases} \tag{3}$$

- Tukey's bisquare function

$$\rho(u) = \begin{cases} 1 - \left\{1 - \left(\dfrac{u}{c}\right)^2\right\}^3 & \text{if } |u| \leqslant c \\ 1 & \text{if } |u| > c \end{cases} \tag{4}$$

Also, Wang et al. [43] propose to use a different type of loss function, i.e. ESL function to obtain a class of penalized robust regression estimators with asymptotic breakdown point of 1/2 for variable selection, and it can be expressed as follows.

- ESL function

$$\rho(u) = 1 - \exp\left\{-\left(\frac{u^2}{c}\right)\right\}$$

In the ESL function, the data points, producing large residuals, lead to large losses of $\rho(u_i)$ and thus, the influence of those observations on the estimators reduces when the tuning parameter c is small (cf. [43]). Hence, the tuning parameter regulate the impact of outliers on the estimates. Then, the robust estimator of β is obtained as the solution of the following estimating functions by minimizing the loss functions $\sum_{i=1}^{N}\sum_{t=1}^{T}\rho\{\frac{(y_{it}-x_{it}^T\beta-\alpha_i)}{\hat{\sigma}}\}$

for a given estimate of the scale parameter, $\hat{\sigma}$

$$U(\beta) = \sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}\psi\left\{\frac{\left(y_{it} - x_{it}^T\beta - \alpha_i\right)}{\hat{\sigma}}\right\} = \underset{K\times 1}{(\mathbf{0})}$$

where ψ denotes the sub-gradient function of $\rho()$. It can be expressed as in Equations (5)–(7), for Huber's, Tukey's bisquare and ESL functions, respectively.

- Huber's function

$$\psi(u) = \begin{cases} u & \text{if } |u| \leqslant c \\ \text{sign}\,(u)\,c & \text{if } |u| > c \end{cases} \tag{5}$$

- Tukey's bisquare

$$\psi(u) = \begin{cases} u\left\{1 - \left(\frac{u}{c}\right)^2\right\}^2 & \text{if } |u| \leqslant c \\ 0 & \text{if } |u| > c \end{cases} \tag{6}$$

- The ESL function

$$\psi(u) = u\exp\left\{-\left(\frac{u^2}{c}\right)\right\} \tag{7}$$

Let us consider the residuals $e_{it} = \ddot{y}_{it} - \ddot{x}_{it}^T\hat{\beta}_{(LS)}$ when OLS regression is performed on the within group transformed data. Then, the estimating equation can be rearranged as follows

$$\begin{aligned} U(\beta) &= \sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{x}_{it}\psi\left\{\frac{\left(\ddot{y}_{it} - \ddot{x}_{it}^T\beta\right)}{\hat{\sigma}}\right\} \\ &= \sum_{i=1}^{N}\sum_{t=1}^{T} \omega_{it}\ddot{x}_{it}\bar{e}_{it} \end{aligned} \tag{8}$$

where $\omega_{it} = \psi(\bar{e}_{it})/\bar{e}_{it}$ denote the weights obtained by the weighting function $W(e) = \rho'(e)/e$ and $\bar{e}_{it} = (\ddot{y}_{it} - \ddot{x}_{it}^T\beta)/\hat{\sigma}$. The U defined in Equation (8) has a form of score functions with the weights, ω_{it}. Thus, our proposed robust estimators can be obtained as follows

$$\hat{\beta}_M = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{x}_{it}^T\omega_{it}\ddot{x}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{x}_{it}^T\omega_{it}\ddot{y}_{it}\right).$$

Since W is a function of β and σ, an iterative procedure has been followed at the kth iteration based on the previous $\hat{\beta}^{(k-1)}$ to determine the weights, $\hat{\omega}_{it} = \omega_{it}|_{\beta=\hat{\beta}^{(k-1)}}$. The M-estimation method assumes that $E\psi(\varepsilon_i) = 0$ and $E\psi^2(\varepsilon_i) = \sigma_\psi^2$. Let's also suppose that

$$b = \left.\frac{\partial E\psi\,(\varepsilon_{it} + \delta)}{\partial\delta}\right|_{\delta=0}.$$

Under the finite variance assumption with some regularity conditions defined in detail in Section 5 and $E|\psi(\varepsilon_i + \delta) - \psi(\varepsilon_i)|^2 = o(1)$ as $\delta \to 0$, the consistent estimator $\hat{\beta}_M$ is

obtained by solving $U(\beta) = 0$ and it follows that

$$\sqrt{N}\left(\hat{\beta}_M - \beta\right) \sim N\left(0, \tau^{-1}\sigma_\varepsilon^2\left[E\left(\ddot{\mathbf{X}}_i^T\ddot{\mathbf{X}}_i\right)\right]^{-1}\right),$$

where $E(\ddot{\mathbf{X}}_i^T\ddot{\mathbf{X}}_i)$ is estimated by $N^{-1}\sum_{i=1}^N \ddot{\mathbf{X}}_i^T\ddot{\mathbf{X}}_i$ (or $N^{-1}\sum_{i=1}^N\sum_{t=1}^T \ddot{x}_{it}^T\ddot{x}_{it}$) and $\tau = b^2/\sigma_\psi^2$ denotes the scalar efficiency factor. The subgradient function ψ, which has large values of the efficiency factor τ, lead to some gain in efficiency of the estimator β as noted in [41]. Thus, the loss function $\rho(\cdot)$ with the largest value of τ should be chosen to obtain more efficient estimates of the parameters. As noted in [19,41], the choice of constant, c may have a great impact to achieve a good trade-off between efficiency and robustness degree. For example, in constructing the WGM estimator proposed by Bramati and Croux [19], the value of c is chosen as 4.685 where Tukey's bisquare function is used. For Huber's function, its default value in the `R` package `rlm` equals to 1.345 to obtain about 90% efficiency when the data follow Normal distribution, see [41]. In traditional robust procedures, the value of c for any loss function must be predetermined by taking into account the desirable level of robustness. However, the efficiency of the estimators varies significantly with the different choices of c. Thus, selecting an appropriate tuning parameter c is a crucial issue in M-estimation approaches. The data driven approaches have improved performances over the the traditional approaches with fixed tuning parameter c in the presence of contamination as shown in the study of Wang et al. [41]. In fact, many procedures are available for choosing proper tuning parameter such as Median-of-Means (MoM), cross-validation (CV) and its generalized version (GCV), Information Criterion-based approaches (Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC)) etc. (cf. [39,44]). However, some of those methods such as MoM may produce unstable and unsatisfactory numerical results. Also, CV methods may be computationally intensive, intractable and time-consuming especially as the number of parameters increases.

For the selection of tuning parameter, we use data driven procedures which have less computational burden than the methods mentioned above. Also, the value of tuning parameter c requires to be chosen depending on the proportion of the outlying points in the data or distribution of the data to maximize asymptotic efficiency of the estimators since the main purpose is to improve the efficiency of the estimators besides robustness. Thus, in this study, best value of the tuning parameter c is chosen by maximizing the value of $\tau = b^2/\sigma_\psi^2$, and the nonparametric estimate of τ, proposed by Wang et al. [42], is calculated by

$$\hat{\tau}(c) = \frac{\left\{\sum_{i=1}^N\sum_{t=1}^T \psi'(e_{it})\right\}^2}{NT\sum_{i=1}^N\sum_{t=1}^T \psi^2(e_{it})} \tag{9}$$

where $e_{it} = \frac{(\ddot{y}_{it} - \ddot{x}_{it}^T\hat{\beta})}{\hat{\sigma}}$, $\hat{\beta}$ and $\hat{\sigma}$ denote the current estimates of β and σ, respectively. In this study, the value of efficiency factor $\hat{\tau}(c)$ is assessed within a finite set of possible values of c ranging from 0 to 3 for the Huber's loss function and a range of values of 1 to 10 for the Tukey's bisquare function since the efficiency factor does not indicate a smooth function of the tuning parameter c (cf. [42]). For the ESL function, it is estimated from the data using an iterative method described in the algorithm below and the convergence is achieved very quickly. Also, some smoothing techniques such as the nearest neighbours

and the expectation-maximization (EM) algorithm can be used in obtaining efficiency factor when choosing the optimal tuning parameter. However, a direct application of the EM algorithm is not possible without determining an initial value and the smoothing methods may increase the computational burden till the convergence is obtained. In fact, when a good balance between efficiency and robustness is investigated, a limiting behaviour of the estimators should be addressed since the ill effects of contamination becomes more apparent in large sample sizes as noted in [45]. When the limiting distribution of an estimator is normal, better approximation can be achieved for the substantial central part of its distribution based on the asymptotic variance than the case of actual variance as emphasized by Huber [45]. Thus, a proper measure of robustness for the estimators, which follow asymptotic normal distribution, can be the supremum of asymptotic variance (cf. [45,46]). Following the definition of Huber [45], let F denote common distribution function within some convex set C of distribution functions. The most robust M-estimator is the maximum likelihood estimator for the unique $F_0 \in C$ with the smallest Fisher Information number $I(F) = \int (f'/f)^2 f \, \mathrm{d}t$ among $F \in C$ (cf. [45]). This implies that by minimizing the loss function of the standardized residuals $\sum \rho(e_i/\sigma)$, a maximum likelihood estimate of β is obtained if the loss function ρ is convex (cf. [46]). Then, Huber [45] demonstrated that the limiting distribution of an M-estimator $\hat{\beta}_M$ is multivariate normal with mean β and covariance matrix proportional to $\mathbf{X}^T\mathbf{X}$. The only effect of ρ on the asymptotic distribution is expressed by the proportionality given below

$$\sigma^2 E\left[\psi^2\left(e/\sigma\right)\right]/E^2\left[\psi'\left(e/\sigma\right)\right]$$

where $\psi(u) = \mathrm{d}\rho(u)/\mathrm{d}u$ and $\psi'(u) = \mathrm{d}^2\rho(u)/\mathrm{d}u^2$ (cf. [46]). As pointed out by Wang et al. [41], the function ψ with a larger value of efficiency factor yields more efficient estimators for β, and $\tau\sigma^2$ indicates a measure of relative efficiency of the estimators considered with respect to the LS estimator. Thus, the efficiency factor τ is calculated using Equation (9) to attain high efficiency as well as high robustness for the proposed estimators. Based on the above, we present the data-driven procedure introduced by Wang et al. [41] to obtain the proposed robust estimators of β in panel data regression models when Huber's and Tukey's bisquare functions are used as loss functions.

Step 1. Compute the within group LS estimate, $\hat{\beta}_{(LS)}$ of the parameter vector, β for the panel data regression model.

Step 2. Calculate the residuals $e_{it} = \ddot{y}_{it} - \ddot{x}_{it}^T\hat{\beta}_{(LS)}$ and obtain the initial estimate of σ with the following equation.

$$\hat{\sigma} = \text{Median}\left\{\left|\ddot{y}_{it} - \ddot{x}_{it}^T\hat{\beta}_{(LS)}\right|\right\}/0.6745$$

Step 3. Obtain the c value (within a range of values of 0 to 3 for Huber's function and 1 to 10 for Tukey's bisquare function as suggested by Wang et al. [42]) which has the largest efficiency factor $\hat{\tau}(c)$.

Step 4. Finally calculate the robust estimator of β by using the Huber's function and Tukey's bisquare function with $c = \hat{c}$ obtained in the previous step.

The importance of selecting optimal tuning parameter has been emphasized in many empirical and methodological researches within the fields of optimization, robotics, robust

and penalized regressions for different purposes such as estimation, making statistical inferences, variable selection, model selection etc. (cf. [47–51]). The choice of tuning parameters allows to adjust the degree of compromise between robustness and efficiency properties of the estimators and makes substantial differences on the behaviour of the estimators (cf. [48,50,52,53]). For example, different tuning parameter selection approaches for the optimization of the penalized least squares estimators have been investigated by Xiao and Sun [53], and the authors note that the performances of penalized estimators significantly change depending on the tuning parameter selection procedure used. Different robust loss functions associated with the different robust estimators have one or more tuning parameters which regularize the effect of the data. However, there is a source of uncertainty which practitioners encountered in implementing the method to a new problem: manually choosing the tuning constants that control the strength of the regularization is not intuitive (cf. [47]). In fact, no universal way exists to choose a proper tuning parameter for different type of estimators (cf. [48,49]) since selecting this parameter represents a kind of optimization problem. When the value of tuning parameter is determined large, then the robustness level of the method increases by sacrificing efficiency (cf. [54]). On the other hand, if the tuning parameter is small, then the efficiency is gained at the price of some loss of robustness. At this stage, a critical consideration in a particular case is how to select a proper tuning parameter depending on the choice of robust method. Note that the loss functions considered in this study have some weaknesses and strengths relative to each other. For example, the M-estimator based on Huber's loss function may not be resistant when high leverage points, representing the contamination in the explanatory variables, are present in the data. Also, the existence and uniqueness of the redescending M-estimator is not ensured since the Tukey's biweight function allows for multiple solutions regarding the multiple local minima. Thus, it may not be feasible to converge to a global solution by using any algorithm. Hence, to select optimal tuning parameter achieving best compromise between robustness and efficiency in the presence of non-normal errors and outliers, we consider two different data-driven algorithms proposed by Wang et al. [41,43] for the loss functions in the class of M-estimators and the Exponential squared loss function. Next, the complete algorithm to determine the value of tuning parameter c in the ESL function is provided by extending the data-driven procedure for penalized regression of Wang et al. [43] to the linear panel data models with fixed effects. For more detailed information, please see Wang et al. [43].

Step 1. Let $\mathbf{Z} = \{\ddot{x}_{it}, \ddot{y}_{it}\}_{i=1,t=1}^{N,T}$ be a random sample that contains m bad points and $NT-m$ good points denoted by $\mathbf{Z}_m = \{Z_1, \ldots, Z_m\}$ and $\mathbf{Z}_{NT-m} = \{Z_{m+1}, \ldots, Z_{NT}\}$, respectively. Then, obtain the residuals $e_{it}(\hat{\beta}_0) = \ddot{y}_{it} - \ddot{x}_{it}^T\hat{\beta}_0$ using the initial high breakdown coefficient $\hat{\beta}_0$ for $i = 1, \ldots, N$, $t = 1, \ldots, T$ and calculate median absolute deviation estimator (MAD), $S_N = 1.4826 \times median|e_{it}(\hat{\beta}_0) - \underset{t}{median}(e_{it}(\hat{\beta}_0))|$. Then, generate pseudo outlier set $\mathbf{Z}_m = \{(\ddot{x}_{it}, \ddot{y}_{it}) : |e_{it}(\hat{\beta}_0)| \geq 2.5S_N\}$ where $m = \#\{1 \leq i \leq N, 1 \leq t \leq T : |e_{it}(\hat{\beta}_0)| \geq 2.5S_N\}$ and $\mathbf{Z}_{NT-m} = \mathbf{Z}/\mathbf{Z}_m$.

Step 2. Let c_N denote the minimizer of $\det(\hat{V}(c))$ in the set $G = \{c : \xi(c) \in (0, 1]\}$ where

$$\xi(c_N) = \frac{2m}{NT} + \frac{2}{NT}\sum_{t=m+1}^{T} \rho_{c_N}\left\{e_{it}\left(\hat{\beta}_0\right)\right\} \tag{10}$$

for $i = 1, \ldots, N$ for a contaminated sample $\mathbf{Z}$ and $\det(\cdot)$ denotes the determinant operator, $\hat{V}(c) = \{\hat{I}(\hat{\beta}_0)\}^{-1}\tilde{\Sigma}\{\hat{I}(\hat{\beta}_0)\}^{-1}$, and

$$\hat{I}\left(\hat{\beta}_0\right) = \frac{2}{c}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\exp\left(-\frac{e_{it}^2\left(\hat{\beta}_0\right)}{c}\right)\left(\frac{2e_{it}^2\left(\hat{\beta}_0\right)}{c}-1\right)\right\} \times\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{x}_{it}\ddot{x}_{it}^T\right),$$

$$\tilde{\Sigma} = \operatorname{cov}\left\{\exp\left(-\frac{e_{it}^2\left(\hat{\beta}_0\right)}{c}\right)\frac{2e_{it}\left(\hat{\beta}_0\right)}{c}\ddot{x}_{1t},\ldots, \exp\left(-\frac{e_{it}^2\left(\hat{\beta}_0\right)}{c}\right)\frac{2e_{it}\left(\hat{\beta}_0\right)}{c}\ddot{x}_{Nt}\right\}_{t=1}^{T}.$$

Step 3. Update $\hat{\beta}_0$ based on the selected value of c_N in Step 2.

It should be noted that we use the MM-estimator of Yohai [55] to obtain the inital estimate $\hat{\beta}_0$ when detecting the outliers in Step 1 in the algorithm for ESL function. This provides to compute $\xi(c_N)$ by using Equation (10). Then, the Steps 1-3 are repeated until $\hat{\beta}_0$ and c_N converge. To achieve high efficiency, the tuning parameter c_N is selected by minimizing the determinant of asymptotic covariance matrix as in Step 2 as noted in [43]. In the last Step, to obtain $\hat{\beta}_M$, $\hat{\beta}_0$ is updated in Step 3 and the algorithm is repeated until convergence. Wang et al. [43] have emphasized that for the convergence of their computing algorithm, it is enough to repeat Steps 1-3 once. If the initial estimator $\hat{\beta}_0$ is robust having asymptotic breakdown point of 1/2 and c_N is chosen such that $\xi(c_N) \in (0, 1]$, then the breakdown point of $\hat{\beta}_M$, $BP(\hat{\beta}_M; \mathbf{Z}_{NT-m}, c_N)$ is asymptotically 1/2, see [43], Theorem 2. This leads us to choose c_N to achieve the highest efficiency.

5. Asymptotic properties

In this section, we established the asymptotic properties of the regression estimators of under fixed effects linear panel data models. Before introducing the asymptotic results, we begin with listing some assumptions. To obtain the consistent estimates of the parameters of fixed effects linear panel data models, the within group LS method requires to satisfy some assumptions (see [12]) as listed below:

Assumption 5.1: (A1) *$E(\varepsilon_{it}|\mathbf{x}_i, \alpha_i)$ for $t = 1, \ldots, T$ implies that the strict exogeneity of $\{x_{it} : t = 1, \ldots, T\}$ conditional on α_i's.*

(A2) $\operatorname{rank}(\sum_{t=1}^{T} E(\ddot{x}_{it}^T\ddot{x}_{it})) = \operatorname{rank}[E(\ddot{\mathbf{X}}_i^T\ddot{\mathbf{X}}_i)] = K$ *where $\ddot{\mathbf{X}}_i$ is a $T \times K$ matrix of predictors.*

(A3) $E(\varepsilon_i\varepsilon_i^T|\mathbf{x}_i, \alpha_i) = \sigma_\varepsilon^2 I_T$.

The following objective function is required to be minimized in obtaining our proposed estimators $\hat{\beta}_M$s based on M-estimation techniques,

$$L(\beta) = \sum_{i=1}^{N} \rho_\tau \left(\frac{\ddot{y}_i - \ddot{\mathbf{X}}_i \beta}{S_N} \right) \tag{11}$$

where the dimension of $\ddot{y}_i$, $\ddot{\mathbf{X}}_i$ and β, respectively, are $T \times 1$, $T \times K$ and $K \times 1$, S_N denotes a scale parameter, and is a normalized MAD estimator obtained using the residuals $e_{it} = \ddot{y}_{it} - \ddot{x}_{it}^T \hat{\beta}_0$ and τ is the efficiency factor (cf. [39]). It should be noted that the proposed estimators $\hat{\beta}_M$s are regression and scale equivariant since S_N has the advantages of being scale equivariant and invariant to the regression.

In fixed effects linear panel data regression models, the regularity conditions required for the consistency and asymptotic normality of the proposed estimators are as follows:

Assumption 5.2: (A1) $E(\ddot{\mathbf{X}}_i^T \ddot{\mathbf{X}}_i)$ *is positive definite, and* $E\|\mathbf{X}^3\| < \infty$.
(A2) $E\psi_\tau(\varepsilon_i) = 0$ *and* $E\psi_\tau^2(\varepsilon_i) = \sigma_{\psi_\tau}^2$.
(A3) $E[\psi_\tau'(\varepsilon/\sigma)] > 0$ *for* τ *where prime denotes the derivative.*

The assumption $E\psi_\tau(\varepsilon_i) = 0$ is needed to provide Fisher consistent estimating functions $U(\beta)$ as noted in [41].

Theorem 5.1: *Let Assumption* 5.2 *hold. If* $S_N \xrightarrow{\text{p}} \sigma$ *as* $N \to \infty$, *then a local minimum of* $L(\beta)$ *occurs at* $\hat{\beta}_M$ *such that* $\|\hat{\beta}_M - \beta\| = \mathcal{O}_p(N^{-1/2})$ *where* '$\xrightarrow{\text{p}}$' *represents the convergence in probability.*

Proof: For any $\nu > 0$, a large constant C exists satisfying

$$P\left(\inf_{\|u\|=C} L\left(\beta + N^{-1/2}u\right) > L(\beta) \right) > 1 - \nu$$

To prove that the existence of a local minimum of $L(\beta)$ in the ball $\{\beta + N^{-1/2}u : \|u\| \leq C\}$ with probability at least $1 - \nu$, under A1 and A3 of Assumption 5.2,

$$\begin{aligned}
&L\left(\beta + N^{-1/2}u\right) - L(\beta) \\
&= \sum_{i=1}^{N} \rho_\tau \left(\frac{\ddot{\mathbf{y}}_i - \ddot{\mathbf{X}}_i\left(\beta + N^{-1/2}u\right)}{S_N} \right) - \sum_{i=1}^{N} \rho_\tau \left(\frac{\ddot{\mathbf{y}}_i - \ddot{\mathbf{X}}_i\beta}{S_N} \right) \\
&= -\frac{1}{\sqrt{N}S_N} \sum_{i=1}^{N} \psi_\tau \left(\frac{\ddot{\mathbf{y}}_i - \ddot{\mathbf{X}}_i\beta}{S_N} \right) \ddot{\mathbf{X}}_i u \\
&\quad + \frac{1}{2S_N^2} u^T \left[\frac{1}{N} \sum_{i=1}^{N} \psi_\tau' \left(\frac{\ddot{\mathbf{y}}_i - \ddot{\mathbf{X}}_i\beta}{S_N} \right) \ddot{\mathbf{X}}_i^T \ddot{\mathbf{X}}_i + o_p(1) \right] u \\
&\triangleq I_1 + I_2
\end{aligned}$$

where $\ddot{\mathbf{y}}_i$ is a $T \times 1$ vector. Then, by using the fact that $S_N \xrightarrow{\text{p}} \sigma$, the law of large numbers and central limit theorem, the proof of Theorem 5.1 direclty follows from Theorem 1 of [39]. ■

Remark 5.1: Theorem 5.1 guarantees the existence of a consistent estimator under some regularity conditions (cf. [39]). It should be noted that the existence of redescending M-estimator is not ensured for the unbounded loss function. Also, Maronna and Yohai [56] provide the results for the existence of the redescending M-estimator when some conditions hold.

Theorem 5.2: *Under Assumption* 5.2 *and* $E|\psi_\tau(\varepsilon_i + \delta) - \psi_\tau(\varepsilon_i)|^2 = o(1)$ *as* $\delta \to 0$, *if* $S_N \xrightarrow{\text{p}} \sigma$ *as* $N \to \infty$, *then, we obtain*

$$\sqrt{N}\left(\hat{\beta}_M - \beta\right) \xrightarrow{\text{d}} N\left(0, \frac{E\left[\psi_\tau^2\left(\varepsilon/\sigma\right)\right]}{\left(E\left[\psi_\tau'\left(\varepsilon/\sigma\right)\right]\right)^2}\sigma^2\left[E\left(\ddot{\mathbf{X}}_i^T\ddot{\mathbf{X}}_i\right)\right]^{-1}\right)$$

where '$\xrightarrow{\text{d}}$' *denotes the convergence in distribution.*

Proof: Let assume

$$\phi_N(\theta) = \frac{1}{NS_N}\sum_{i=1}^{N}\psi_\tau\left(\frac{\ddot{\mathbf{y}}_i - \ddot{\mathbf{X}}_i\theta}{S_N}\right)\ddot{\mathbf{X}}_i$$

Then, we have the following by Taylor's expansion

$$\phi_N\left(\hat{\beta}_M\right) = \phi_N(\beta) + \phi_N'(\beta)\left(\hat{\beta}_M - \beta\right) + \frac{\left(\hat{\beta}_M - \beta\right)^T\phi_N''(\beta^*)\left(\hat{\beta}_M - \beta\right)}{2}$$

where β^* is a vector that lies between $\hat{\beta}_M$ and β, and also, $\phi_N'(\cdot)$ and $\phi_N''(\cdot)$ represent the first-order and second-order derivatives of $\phi_N(\cdot)$, respectively. Then, $\phi_N(\hat{\beta}_M) = 0$ from Equation (11) (cf. [39]). Under (A1)–(A2) of Assumption 5.2, we obtain

$$= \frac{S_N}{\sqrt{N}}\sum_{i=1}^{N}\psi_\tau\left(\frac{\ddot{\mathbf{y}}_i - \ddot{\mathbf{X}}_i\beta}{S_N}\right)\ddot{\mathbf{X}}_i = \sqrt{N}\left(\hat{\beta}_M - \beta\right)\frac{1}{N}\sum_{i=1}^{N}\psi_\tau'\left(\frac{\ddot{\mathbf{y}}_i - \ddot{\mathbf{X}}_i\beta}{S_N}\right)\ddot{\mathbf{X}}_i^T\ddot{\mathbf{X}}_i + o_p(1)$$

by Theorem 5.1. The proof of Theorem 5.2 is completed since $S_N \xrightarrow{\text{p}} \sigma$ as $N \to \infty$. ■

6. Numerical results

In this section, we conduct an extensive simulation study to examine the finite-sample properties of proposed and some existing panel data estimators. To investigate the performances of the proposed procedures, three scenarios, namely (i) different sample sizes, (ii) different error distributions, and (iii) various types of outliers, have been considered. The following simulations and all calculations have been carried out using R 3.6.0. on an Intel-Core i7 6700HQ 2.6 GHz PC. (The codes can be obtained from the author upon request.)

For the data generation process, we consider the following static fixed-effect panel-data model

$$y_{it} = x_{it}^T \beta + \alpha_i + \varepsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{12}$$

where ε_{it}'s are assumed to be iid $\mathrm{N}(\mu = 0, \sigma^2 = 1)$ and the vector of parameters is chosen as $\beta^T = (\beta_1, \beta_2) = (2.4, -1.2)$. Similar to experimental design of Aquaro and Cizek [18], the unobservable individual effects are generated as follows

$$\alpha_i = \sum_{t=1}^{T} x_{it}^T \gamma / \sqrt{T} + \eta_i$$

where $\gamma = (2, 4)^T$ and $\eta_i \sim U(0, 12)$ to guarantee homogeneity of the variances of deterministic γ and random parts η_i of α_i. To avoid completely symmetric design as in [18], the independent variables x_{itK} for $K = 1, 2$ are generated according to

$$x_{itK} \sim \begin{cases} \chi_2^2 - 2 & \text{if } K = 1 \\ N(0, 1) & \text{if } K = 2, \end{cases}$$

where χ_2^2 denote the chi-squared distribution with 2 degrees of freedom. The performances of the proposed estimators are evaluated based on $S = 10.000$ simulations. In case of the changing sample sizes and presence of outliers, we compute the mean squared errors (MSE): $MSE = \frac{1}{S} \sum_{s=1}^{S} \|\widehat{\beta}^s - \beta\|^2$, where $\widehat{\beta}^s$, $s = 1, \ldots, S$, represents the estimates obtained from S simulated samples. Also, we calculated the squared errors (SE): $SE = \|\widehat{\beta}^s - \beta\|^2$ to examine the effect of the error distributions on the estimators. As noted above, the robust estimators under consideration are the WMS estimator of Bramati and Croux [19] and WFE estimator of Beyaztas and Bandyopadhyay [1] whereas the LS estimator is considered as a reference method. Furthermore, we consider another robust alternative by applying the FS method of Atkinson et al. [27] to the within group transformed model for the performance evaluation of proposed methods (see e.g. [30,57–60], for more information about the FS method). To this end, we use `fwdlm` function in the `R` package `forward`. Throughout the experiments, all calculations are performed for the finite sample breakdown points $BP = 0.1$ and $BP = 0.5$ for WMS estimator. However, we present only the results obtained for the choice of $BP = 0.1$ since WMS procedure with this breakdown point produced better results.

6.1. Sample sizes

In this subsection, different values of cross-sectional dimension, $N = 50, 100, 150, 200, 250$ (by keeping time period fixed at $T = 3$) and time periods, $T = 4, 6, 9, 12, 24$ (for a fixed number of cross-sectional units $N = 50$) are considered to investigate the influence of panel sizes on our proposed estimators. Figure 1 illustrates calculated MSE values of the estimators when the errors follow standart normal distribution, $\varepsilon_{it} \sim N(\mu = 0, \sigma^2 = 1)$. In this figure, the lines representing the performances of the LS estimator, WFE estimator, FS-based estimator, and proposed loss functions (Huber, Tukey, Exponential) based estimators using data-dependent regularization parameters overlap as $N \to \infty$ and $T \to \infty$ since they have almost same performances under normal errors. However, the WMS procedure are not consistent for fixed time dimension while it is consistent for increasing

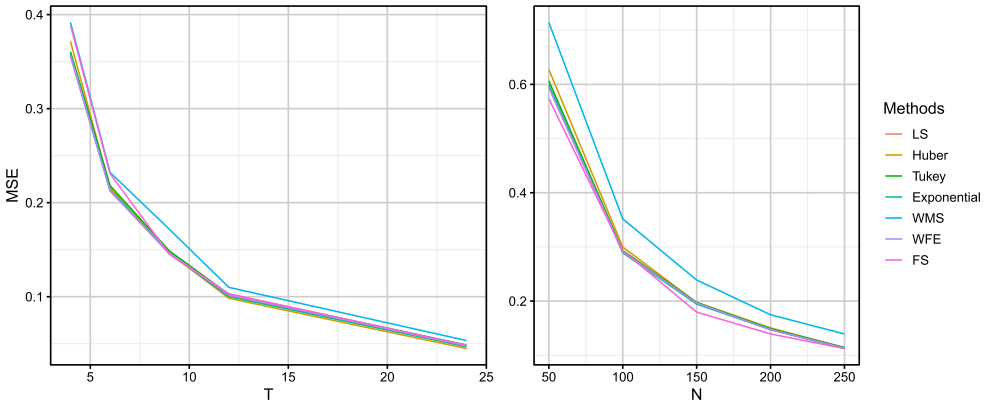

Figure 1. The MSEs of all estimators when $T = 4, 6, 9, 12, 24$ for fixed cross-sectional dimension $N = 50$ and $N = 50, 100, 150, 200, 250$ for fixed time dimension $T = 3$ and $\varepsilon_{it} \sim N(\mu = 0, \sigma^2 = 1)$.

values of T as noted in [18]. These results also confirm that the proposed methods are asymptotically equivalent to LS estimator as N and/or T increases.

6.2. Error distributions

The SE values of the estimators are calculated under four different error distributions: $N(\mu = 0, \sigma^2 = 1)$, Student's t distribution with 5 degrees of freedom (t_5), the chi-squared distribution with 4 degrees of freedom $\chi^2_{(4)}$ and standard Cauchy distribution Cauchy(0, 1) to demonstrate the effects of error distributions on the estimation methods. Also, three pairs of cross-sectional sizes and time periods: $(N_1, T_1) = (30, 20)$, $(N_2, T_2) = (75, 8)$ and $(N_3, T_3) = (200, 3)$ are considered as in [18]. The simulation results are presented in Figure 2. The skewness of SE values of WMS procedure (with $BP = 0.1$) is greater than the other methods for Normal, t_5 and $\chi^2_{(4)}$ distribution of the errors especially for increasing N or decreasing T. This means that the variability and mean of SE values obtained for WMS estimator are slightly higher than the other estimators. All the proposed, WFE, FS, and LS estimators have similar SE values under all the error distributions, except for standard Normal and standard Cauchy distributions. Under normally distributed errors, FS method produces slightly smaller SE values compared with other methods as N increases. However, the mean and variance of the SE values for LS, and FS methods increase more than the other estimators in case of the Cauchy distribution of the error terms while the proposed robust estimators, WFE of Beyaztas and Bandyopadhyay [1] and WMS of Bramati and Croux [19] are considerably less affected.

6.3. Outliers

This subsection presents the robustness performances of the estimation procedures in case of the various types of outliers. Throughout the simulations, two different levels of cross-sectional sizes and time periods, namely, $N_1 = 120$, $T_1 = 2$ and $N_2 = 80$, $T_2 = 3$, are considered for the fixed panel size consisting of a total of 240 observations. The proportions of contamination are chosen as 5% and 10% by determining the number of outliers as

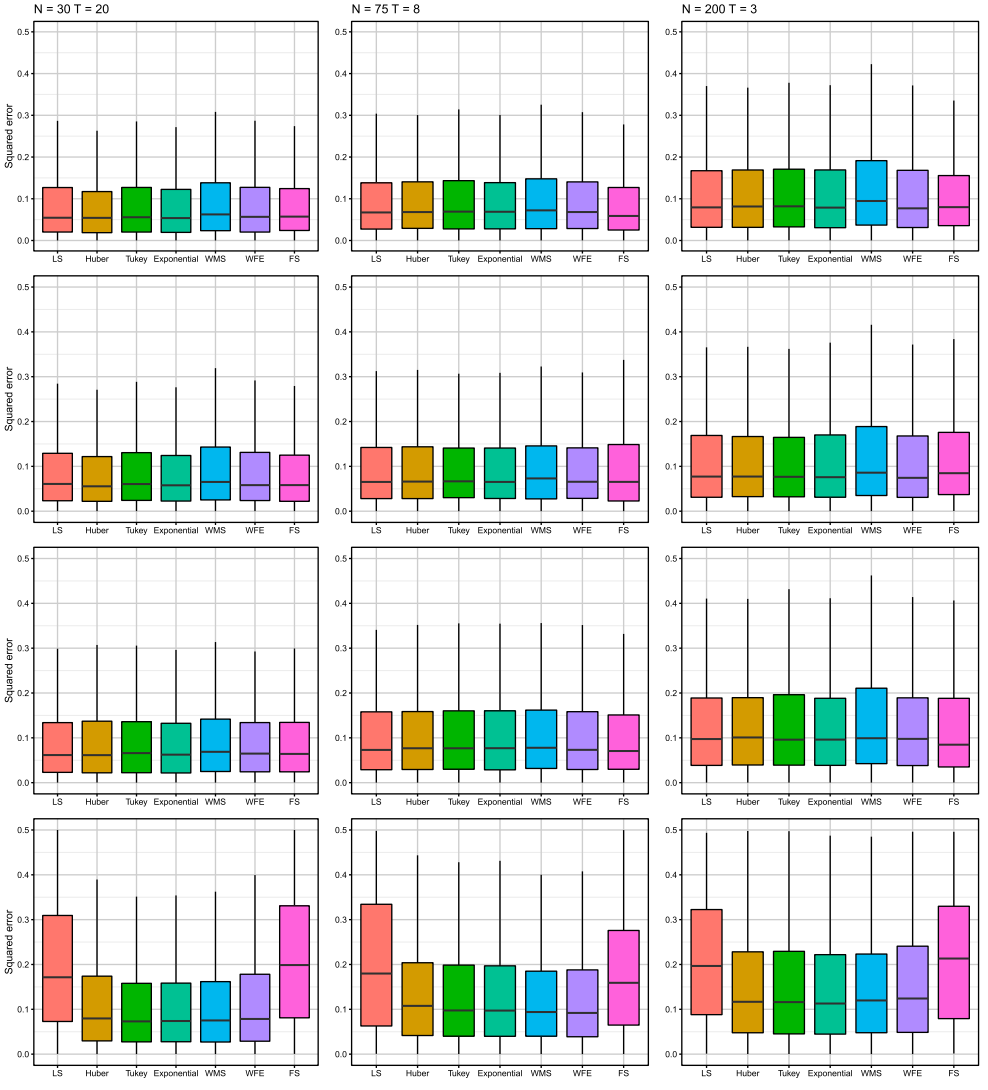

Figure 2. The boxplots of the SEs of all estimators under different error distributions when $(N_1, T_1) = (30, 20)$, $(N_2, T_2) = (75, 8)$ and $(N_3, T_3) = (200, 3)$. First row: $N(0, 1)$ distribution, second row: t_5 distribution, third row: $\chi^2_{(4)}$ distribution and fourth row: $Cauchy(0, 1)$ distribution.

$m = 12$ and $m = 24$. Two main types of contamination, namely, random contamination and concentrated (or clustered) contamination as in [19] are used to generate contaminated datasets. The outlying points are randomly distributed over all observations for random contamination while the outliers concentrating in some blocks constitute at least a half of the observations within cross-sectional units for concentrated contamination. For more detailed information on the types of contamination and graphs of contamination schemes, please see Figure 1 of Bramati and Croux [19]. The considered contamination schemes depending on the types of outliers are described as follows.

(1) To generate random vertical outliers (y_{it}^r), randomly selected original values of the dependent variable are replaced by the observations from an Uniform distribution $y_{it}^r \sim U(20, 80)$.
(2) At first, randomly chosen values of the dependent variable are contaminated following the same rule of first scheme. After that, the values of the independent variables corresponding to the observations contaminated in the dependent variable are replaced by the points coming from a Normal distribution $N(\mu = 8, \sigma^2 = 4)$ to obtain the random leverage points.
(3) Concentrated vertical outliers are inserted into the data by substituting the random observations from an Uniform distribution $y_{it}^c \sim U(79, 80)$ for the randomly selected blocks of the original values of dependent variable.
(4) Firstly, the contaminated blocks of the dependent variable are obtained by applying the same rule used in third scheme. In the next step, the concentrated leverage points 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 $N(\mu = 8, \sigma^2 = 4)$.

The simulation results are summarized in Tables 1 and 2. At first glance, it is conspicuous from Table 1 that the LS method leads to obtain distorted estimates of the parameters of interest in the presence of outliers. It can further be seen that LS estimators can be severely degraded when the data is contaminated by the leverage points, compared to the case of vertical outliers. Additionally, the MSE values obtained for LS increase as the level of contamination increases regardless of the types and levels of contaminations and are severely affected in presence of concentrated contamination. Conversely, the proposed estimators (named according to the corresponding loss functions: Huber, Tukey, Exponential), WFE of Beyaztas and Bandyopadhyay [1] and WMS of Bramati and Croux [19] are highly resistant to outliers. In particular, the proposed robust estimators are quite less affected by the choice of contamination level and/or scheme compared to WMS, WFE, and FS procedures. The MSE values for FS method are generally similar with those of LS method in the presence of vertical outliers. Although FS method yields substantially better results than LS method in case of leverage points, other robust methods produce relatively better results than the FS. One of the remarkable results is that Tukey estimator based on the Tukey's bisquare function produces best results with smallest MSEs among the robust estimators under all contamination schemes except for concentrated contamination with level 5% for $(N_1, T_1) = (120, 2)$. Although WFE estimator produce slightly better MSEs compared to Tukey estimator under 5% concentrated contamination, Tukey estimator has significantly better performance than the WFE estimator as the contamination level increases. The performances of the robust estimators can be sorted based on minimizing MSE values as follows: Tukey, Exponential and Huber estimators for $T_1 = 2$. The proposed estimators have improved performance over the WFE and WMS estimators which produce closer results to each other especially with the increasing level of contamination. We further see that, the performance of WMS estimator is sensitive to the increasing level of contamination and concentrated contamination when $T_1 = 2$ as noted in [18,19]. When $(N_2, T_2) = (80, 3)$ is considered, Exponential loss function-based estimator has the best performance among all estimators under random contamination. Although there are no substantial differences between WMS and Exponential loss function based estimator at 5%

Table 1. The MSEs for $(N_1, T_1) = (120, 2)$, $(N_2, T_2) = (80, 3)$ in the presence of 5% and 10% random and concentrated contamination by setting the number of outliers as $m = 12, 24$.

	Random contamination				Concentrated contamination			
$(N_1, T_1) = (120, 2)$	Vertical outliers		Leverage points		Vertical outliers		Leverage points	
Method	5%	10%	5%	10%	5%	10%	5%	10%
LS	1.593	2.641	7.487	8.950	3.321	5.344	25.075	30.292
Huber	0.741	0.990	0.695	0.982	0.756	0.973	0.690	0.968
Tukey	0.668	0.848	0.615	0.823	0.541	0.663	0.535	0.613
Exponential	0.672	0.865	0.625	0.847	0.562	0.681	0.561	0.645
WMS	0.795	1.599	0.749	1.663	1.105	3.634	1.012	4.022
WFE	0.756	1.739	0.702	1.791	0.529	3.831	0.517	4.439
FS	1.517	2.808	1.637	2.395	3.182	5.430	3.170	5.428
$(N_2, T_2) = (80, 3)$								
LS	1.188	1.927	7.588	8.984	2.345	4.312	25.240	30.213
Huber	0.552	0.815	0.630	0.807	0.612	0.943	0.620	0.981
Tukey	0.497	0.746	0.575	0.729	0.487	0.628	0.513	0.670
Exponential	0.492	0.732	0.573	0.721	0.495	0.674	0.524	0.699
WMS	0.502	0.787	0.581	0.788	0.447	0.940	0.492	1.010
WFE	0.587	1.171	0.674	1.187	0.591	1.811	0.615	1.906
FS	1.214	1.925	1.180	1.938	2.239	4.397	2.438	4.108

Table 2. The RMSEs of predicted values for $(N_1, T_1) = (120, 2)$, $(N_2, T_2) = (80, 3)$ in the presence of 5% and 10% random and concentrated contamination by setting the number of outliers as $m = 12, 24$.

	Random contamination				Concentrated contamination			
$(N_1, T_1) = (120, 2)$	Vertical outliers		Leverage points		Vertical outliers		Leverage points	
Method	5%	10%	5%	10%	5%	10%	5%	10%
LS	4.805	4.867	5.162	5.248	4.967	5.135	6.121	6.337
Huber	4.741	4.738	4.750	4.780	4.768	4.788	4.791	4.783
Tukey	4.734	4.729	4.742	4.768	4.750	4.762	4.775	4.749
Exponential	4.736	4.728	4.743	4.769	4.752	4.763	4.778	4.752
WMS	4.745	4.785	4.752	4.833	4.798	5.005	4.819	5.023
WFE	4.742	4.792	4.751	4.841	4.750	5.007	4.772	5.019
FS	4.784	4.926	4.832	4.899	4.957	5.099	4.944	5.126
$(N_2, T_2) = (80, 3)$								
LS	5.547	5.647	6.030	6.058	5.671	5.825	7.058	7.314
Huber	5.485	5.544	5.547	5.498	5.516	5.529	5.508	5.536
Tukey	5.479	5.540	5.542	5.493	5.503	5.500	5.498	5.506
Exponential	5.479	5.538	5.542	5.492	5.504	5.505	5.499	5.510
WMS	5.480	5.539	5.541	5.501	5.499	5.539	5.495	5.556
WFE	5.488	5.573	5.551	5.534	5.515	5.622	5.510	5.630
FS	5.577	5.652	5.579	5.650	5.657	5.888	5.670	5.790

concentrated contamination, the performance of WMS method is slightly better than the proposed robust estimators. However, Exponential and Tukey's bisquare functions-based estimators produce more accurate and precise estimates of the parameters for random contamination and concentrated contamination, respectively, when the level of contamination increases. Our results clearly demonstrate that the all proposed estimators (Huber, Tukey, Exponential) are reasonable competitors that often exhibit improved performance over the traditional LS method and robust WMS, WFE, and FS estimators especially for increasing level of contamination.

To confirm the supremacy over the traditional LS and robust methods, we also calculate the root mean squared errors (RMSE) of predicted values. To this end, the simulated samples are divided into the following two parts: the predictive model is constructed based on training sample to compute the prediction errors from testing sample with cross-sectional size N_{test}. Then, RMSEs of the predicted values are calculated by

$$RMSE = \frac{1}{S}\sum_{s=1}^{S}\sqrt{\sum_{i=1}^{N_{\text{test}}}\sum_{t=1}^{T}(y_{it}^{s}-\hat{y}_{it}^{s})^2/\left(N_{\text{test}}\times T\right)} = \frac{1}{S}\sum_{s=1}^{S}\sqrt{\sum_{t=1}^{T_i}(y_{it}^{s}-\hat{y}_{it}^{s})^2/T_i} \quad (13)$$

where $i = 1, 2, \ldots, N_{\text{test}}$, y_{it}^{s} and $\hat{y}_{it}^{s}$, respectively, denote the observed values and predictions obtained from the predictive model for each of S simulated samples. Throughout simulations, we determine the sizes of test samples as $N_{\text{test}} \times T_1$ for $(N_1, T_1) = (120, 2)$ and $N_{\text{test}} \times T_2$ for $(N_2, T_2) = (80, 3)$ with the choice of $N_{\text{test}} = 50$. Table 2 displays the RMSEs obtained from the predictive model constructed by the estimated regression coefficients using all the estimators considered in this study. The results provided by the robust estimators are similar and also, better than traditional LS method especially for concentrated leverage points. However, proposed Tukey's bisquare and Exponential loss functions-based estimators have slightly better performance among the robust estimators, in general.

7. Case study

In this section, we compare the performances of all the estimators to check the supremacy of proposed M-estimation procedures based on automatic selection of the tuning parameters via a real macroeconomic application. The data set consists of a total of 342 (N = 18, T = 19) annual observations covering a cross-section of 18 OECD countries over the period 1960-1978 (available in the `R` package `plm`, please see [61]). For this panel, the gasoline demand model studied by Baltagi and Griffin [62] is as follows

$$\ln GC_{it} = \alpha_i + \beta_1 \ln IPC_{it} + \beta_2 \ln GP_{it} + \beta_3 \ln CS_{it} + \varepsilon_{it}$$

where GC represents the gasoline consumption per car as dependent variable, IPC, GP and CS, respectively, denote real income per capita, real gasoline price and car stock per capita as independent variables. The index $i = 1, 2, \ldots, 18$ and $t = 1, 2, \ldots, 19$ denote the

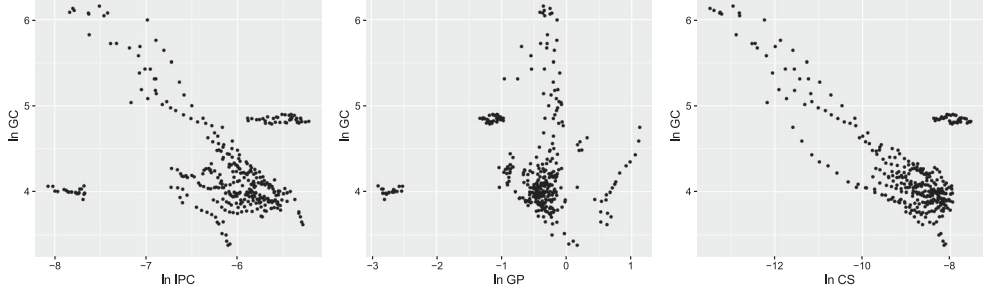

Figure 3. Scatter plots of the logarithm of gasoline consumption (ln *GC*) against the logarithms of real income (ln *IPC*), real gasoline price (ln *GP*) and car stock (ln *CS*) for 18 OECD countries.

Table 3. Estimates of individual coefficients (upper rows) and standard errors of the estimates (lower rows) for gasoline consumption data of 18 OECD countries over the period 1960–1978.

Method	β_1	β_2	β_3
LS	−1.043	−0.264	0.113
	(0.054)	(0.066)	(0.021)
Huber	0.579	−0.317	−0.559
	(0.048)	(0.029)	(0.019)
Tukey	0.479	−0.302	−0.451
	(0.040)	(0.024)	(0.016)
Exponential	0.482	−0.307	−0.452
	(0.027)	(0.018)	(0.012)
WMS	0.902	−0.356	−0.651
	(0.017)	(0.014)	(0.011)
WFE	0.578	−0.300	−0.574
	(0.058)	(0.034)	(0.023)
FS	0.662	−0.322	−0.640
	(0.051)	(0.030)	(0.020)

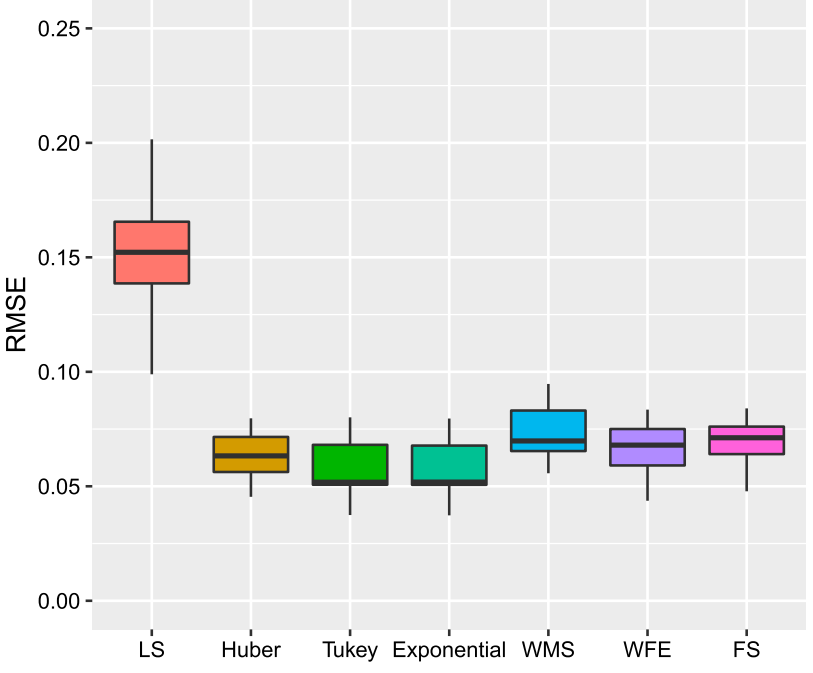

Figure 4. The boxplots of the RMSEs of predicted and observed values of the logarithm of the gasoline consumption for three OECD countries.

OECD countries and years, respectively; see [62] for more detailed description of the data set. Figure 3 shows the scatter plots of the logarithm of response variable against the logarithms of explanatory variables. From Figure 3, we observe that there exist outliers in the response and the independent variables. Table 3 reports the estimates of individual coefficients and standard errors of the estimates. We can see from Table 3 that the robust methods yield more efficient estimates than the traditional LS method. The proposed

Huber estimator and FS method-based estimator exhibit similar performance in terms of efficiency. The WMS estimator produces more precise estimates of individual coefficients among the robust procedures since *BP*, a measure of robustness level, is chosen as 0.1. However, it should be noted that the precision of the estimates may be adversely affected by increasing level of robustness. Also, the proposed methods, especially Exponential loss function-based estimator, provide competitive results in estimating β_2 and β_3.

To investigate the predictive performances of the LS, WMS, WFE, FS, and proposed robust methods further, the RMSE of predicted and observed values of the logarithms of the gasoline consumption are calculated by dividing the dataset into the two parts as mentioned in Section 6.3. For this purpose, the randomly selected 15 countries ($N - N_{\text{test}} = 15$) are used to build a predictive model, and the logarithms of the gasoline consumption of remaining 3 countries ($N_{\text{test}} = 3$) are predicted by using the estimated model coefficients. This process is performed $S = 100$ times, and for each time, RMSEs are obtained by using Equation (13). The results are presented in Figure 4. This figure demonstrates that all the robust methods have better predictive ability compared to the traditional LS method. Furthermore, our proposed robust methods yield considerably better predictions than WMS, WFE, and FS methods.

8. Conclusions

In this paper, we propose robust and efficient estimators based on minimizing loss functions in estimating the parameters of fixed effects panel data models. The proposed procedures uses a data-driven approach for selection of regularization parameters, which is important to determine the necessary level of robustness without sacrificing estimation efficiency in the presence outlying observations and/or blocks. The asymptotic properties of the proposed estimators are investigated in the context of fixed effects linear panel data models. The finite sample performances of the proposed methods are illustrated via extensive simulation studies and an empirical application, and we compare the results of our methods with those for existing WMS estimator of Bramati and Croux [19], WFE estimator of Beyaztas and Bandyopadhyay [1], FS method of Atkinson et al. [27], and traditional LS method. Our thorough study demonstrates that the proposed estimators produce more accurate and efficient parameter estimates with better predictions compared to WMS, WFE, FS, and traditional LS methods under data contamination, and consistent results with the LS method when no outliers are present in the data and sample size increases.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Beste Hamiye Beyaztas http://orcid.org/0000-0002-6266-6487

References

[1] Beyaztas BH, Bandyopadhyay S. Robust estimation for linear panel data models. Stat Med. 2020;39(29):4421–4438.

[2] Bickel R. Multilevel analysis for applied research: it's just regression. New York: The Guilford Press; 2007.
[3] Hsiao C. Panel data analysis-advantages and challenges. Test. 2007;16(1):1–22.
[4] Baltagi BH. Econometric analysis of panel data. Chichester: John Wiley and Sons; 2005.
[5] Diggle PJ, Heagerty P, Liang K-Y, et al. Analysis of longitudinal data. UK: Oxford University Press; 2002.
[6] Fitzmaurice GM, Ware JH. Applied longitudinal analysis. New York: John Wiley and Sons; 2004.
[7] Greene WH. Econometric analysis. New Jersey: Prentice Hall; 2003.
[8] Hill RC, Griffiths WE, Lim GC. Principles of econometrics. USA: John Wiley and Sons; 2007.
[9] Maddala GS, Mount TD. A comparative study of alternative estimators for variance components models used in econometric applications. J Amer Statist Assoc. 1973;68(342):324–328.
[10] Mundlak Y. On the pooling of time series and cross section data. Econometrica. 1978;46(1):69–85.
[11] Wallace TD, Hussain A. The use of error components models in combining cross section and time-series data. Econometrica. 1969;37(1):55–72.
[12] Wooldridge JM. Econometric analysis of cross section and panel data. Cambridge: The MIT Press; 2002.
[13] Hsiao C. Benefits and limitations of panel data. Econom Rev. 1985;4(1):121–174.
[14] Greene WH. Econometric analysis. New York: Prentice Hall; 2017.
[15] Kutner MH, Nachtsheim CJ, Neter J. Applied linear regression models. New York: McGraw-Hill Education; 2004.
[16] Bakar NMA, Midi H. Robust centering in the fixed effect panel data model. Pakistan J Statist. 2015;31:33–48.
[17] Rousseeuw PJ, van Zomeren BC. Unmasking multivariate outliers and leverage points. J Amer Statist Assoc. 1990;85(41):633–639.
[18] Aquaro M, Cizek P. One-step robust estimation of fixed-effects panel data models. Comput Statist Data Anal. 2013;57(1):536–548.
[19] Bramati MC, Croux CP. Robust estimators for the fixed effects panel data model. Econom J. 2007;10(3):521–540.
[20] Maronna RA, Martin RD, Yohai VJ. Robust statistics, theory and methods. New York: John Wiley and Sons; 2006.
[21] Rousseeuw PJ, Leroy AM. Robust regression and outlier detection. New York: John Wiley and Sons; 2003.
[22] Rousseeuw PJ. Least median of squares regression. J Amer Statist Assoc. 1984;79(388): 871–880.
[23] Rousseeuw PJ, Yohai VJ. Robust regression by means of S estimators. In: Franke J, Hardle W, Martin RD, editors. Robust and nonlinear time series analysis. Lecture Notes in Statistics. New York: Springer; 1984. p. 256–274.
[24] Davies PL, Gather U. Breakdown and groups. Ann Statist. 2005;33(3):977–1035.
[25] Genton MG, Lucas A. Comprehensive definitions of breakdown points for independent and dependent observations. J R Stat Soc Ser B Stat Methodol. 2003;65(1):81–94.
[26] Rousseeuw PJ, Leroy AM. Robust regression and outlier detection. Chichester: John Wiley and Sons; 1987.
[27] Atkinson AC, Riani M, Cerioli A. The forward search: theory and data analysis. J Korean Stat Soc. 2010;39(2):117–134.
[28] Cizek P. Robust and efficient adaptive estimation of binary-choice regression models. J Amer Statist Assoc. 2008;103(4):687–696.
[29] Hampel FR, Ronchetti EM, Rousseeuw PJ, et al. Robust statistics: the approach based on influence functions. New York: John Wiley and Sons; 1986.
[30] Riani M, Atkinson AC, Perrotta D. A parametric framework for the comparison of methods of very robust regression. Stat Sci. 2014;29(1):128–143.
[31] Simpson DG, Ruppert D, Carroll RJ. On one-step gm estimates and stability of inferences in linear regression. J Amer Statist Assoc. 1992;87(41):439–450.

[32] Wagenvoort R, Waldmann R. On b-robust instrumental variable estimation of the linear model with panel data. J Econom. 2002;106(2):297–324.
[33] Midi H, Muhammad S. Robust estimation for fixed and random effects panel data models with different centering methods. J Eng Appl Sci. 2018;13(17):7156–7161.
[34] Namur VV, Luneburg JW. Robust estimation of linear fixed effects panel data models with an application to the exporter productivity premium. Jahrb f Nationalok u Stat. 2011;231(4):546–557.
[35] Visek JA. Estimating the model with fixed and random effects by a robust method. Methodol Comput Appl Probab. 2015;17(4):999–1014.
[36] Maronna RA, Yohai VJ. Robust regression with both continuous and categorical predictors. J Statist Plann Inference. 2000;89(1–2):197–214.
[37] Gervini D, Yohai VJ. A class of robust and fully efficient regression estimators. Ann Statist. 2002;30(2):583–616.
[38] Cizek P. Reweighted least trimmed squares: an alternative to one-step estimators. CentER Discussion Paper Series 91/2010; 2010.
[39] Jiang Y, Wang YG, Fu L, et al. Robust estimation using modified Huber's functions with new tails. Technometrics. 2019;61(1):111–122.
[40] Lindsay B. Efficiency versus robustness: the case for minimum Hellinger distance and related methods. Ann Stat. 1994;22(2):1018–1114.
[41] Wang YG, Lin X, Zhu M, et al. Robust estimation using the huber function with a data-dependent tuning constant. J Comput Graph Stat. 2007;16(2):468–481.
[42] Wang N, Wang YG, Hu S, et al. Robust regression with data-dependent regularization parameters and autoregressive temporal correlations. Environ Model Assess. 2018;23(6):779–786.
[43] Wang X, Jiang Y, Huang M, et al. Robust variable selection with exponential squared loss. J Amer Statist Assoc. 2013;108(5):632–643.
[44] Cai X, Xue L, Lu F. Robust estimation with a modified Huber's loss for partial functional linear models based on splines. J Korean Stat Soc. 2020;49(4):1214–1237.
[45] Huber PT. Robust estimation of a location parameter. Ann Math Statist. 1964;35(1):73–101.
[46] Schrader RM, Hettmansperger TP. Robust analysis of variance based on upon a likelihood ratio criterion. Biometrika. 1980;67(1):93–101.
[47] Agamennoni G, Furgale P, Siegwart R. Self-tuning m-estimators. 2015 IEEE International Conference on Robotics and Automation (ICRA); 2015. p. 4628–4635.
[48] Basu A, Harris IR, Hjort NL, et al. Robust and efficient estimation by minimising a density power divergence. Biometrika. 1998;85(3):549–559.
[49] Jones MC, Hjort NL, Harris IR, et al. A comparison of related density-based minimum divergence estimators. Biometrika. 2001;88(3):865–873.
[50] Kelly GE. Robust regression estimators-the choice of tuning constants. J R Stat Soc Ser D. 1992;41(3):303–314.
[51] Roy S, Chakraborty K, Bhadra S, et al. Density power downweighting and robust inference: some new strategies. J Math Statist. 2019;15(1):333–353.
[52] Fujisawa H. Normalized estimating equation for robust parameter estimation. Electron J Stat. 2013;7:1587–1606.
[53] Xiao H, Sun Y. On tuning parameter selection in model selection and model averaging: a Monte Carlo study. J Risk Financ Manag. 2019;12(3):1–16.
[54] Fujisawa H, Eguchi S. Robust parameter estimation with a small bias against heavy contamination. J Multivar Anal. 2008;99(9):2053–2081.
[55] Yohai VJ. High breakdown-point and high efficiency robust estimates for regression. Ann Statist. 1987;15(2):642–656.
[56] Maronna RA, Yohai VJ. Asymptotic behavior of general m-estimates for regression and scale with random carriers. Probab Theory Relat Fields. 1981;58(1):7–20.
[57] Atkinson AC. Econometric applications of the forward search in regression: robustness, diagnostics and graphics. Econom Rev. 2009;29(1–3):21–39.
[58] Atkinson AC, Riani M, Cerioli A. Exploring multivariate data with the forward search. New York: Springer-Verlag; 2004.

[59] Hadi AS, Simonoff JS. Procedures for the identification of multiple outliers in linear models. J Amer Statist Assoc. 1993;29(424):1264–1272.
[60] Riani M. Extensions of the forward search to time series. Stud Nonlinear Dyn Econom. 2004;8(2):1–23.
[61] Croissant Y, Millo G. Panel data econometrics in R: the PLM package. J Stat Softw. 2008;27(2):1–43.
[62] Baltagi BH, Griffin JM. Gasoline demand in the OECD: an application of pooling and testing procedures. Eur Econ Rev. 1983;22(2):117–137.