

# Data Driven Robust Estimation Methods for Fixed Effects Panel Data Models

Beste Hamiye Beyaztas<sup>a,\*</sup>, Soutir Bandyopadhyay<sup>b</sup>

<sup>a</sup>Department of Statistics, Istanbul Medeniyet University, Istanbul, Turkey

<sup>b</sup>Department of Applied Mathematics, Statistics Colorado School of Mines, Golden, Colorado, USA

---

## Abstract

The panel data regression models have gained increasing attention in different areas of research including but not limited to econometrics, environmental sciences, epidemiology, behavioral 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 (LS) method is applied. We propose extensions of the M-estimation approach 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 properties of the existing and proposed robust estimators have been examined through an extensive simulation study and an application to macroeconomic data. Our findings reveal that the proposed methods often exhibits improved estimation and prediction performances in the presence of outliers and are consistent with the traditional LS method when there is no contamination.

*Keywords:* Panel data, Fixed effects, Robust estimation, M-estimation, Least squares

*2010 MSC:* 62M10, 62F35

---

## 1. Introduction

Panel data refers to the two-dimensional data in which cross-sectional units are observed over time. 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, Bickel (2007)). Over last several years, its increasing availability, demanding methodology and better ability to model the complexity of human behavior than a pure cross-section or time series data are the primary reasons behind the excessive growth in its study (see, Hsiao (2007)). For a comprehensive account of the technical details on linear panel data models, please refer to Baltagi (2005), Fitzmaurice et al. (2004), Greene (2003), Maddala and Mount (1973), Mundlak (1978),

---

\*Corresponding author

Diggle et al. (2002), Hill et al. (2007), Wallace and Hussain (1969), Wooldridge (2002), and the references therein.

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. Baltagi (2005) and Hsiao (1985) 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 heterogeneity. The most widely used regression techniques for making statistical inferences regarding the parameters of linear panel data regression models are typically based on the 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. Kutner et al. (2004), Greene (2017)). Also, the panel data may include various types of outliers such as either vertical, horizontal or leverage as noted in Rousseeuw and van Zomeren (1990) and Bakar and Midi (2015). 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. Bramati and Croux (2007) and Aquaro and Cizek (2013)). 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 Rousseeuw and Leroy (2003), Maronna et al. (2006) and Bakar and Midi (2015). 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 (1984) and Rousseeuw and Yohai (1984) 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. Genton and Lucas (2003), Davies and Gather (2005) and Aquaro and Cizek (2013)). As pointed out by Rousseeuw and Leroy (1987), 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, Hampel et al. (1986), Simpson et al. (1992), Wagenvoort and Waldmann (2002), Maronna et al. (2006), Cizek (2008). 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 have 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. Bramati and Croux (2007), Aquaro and Cizek (2013)). A few different approaches within the robust estimation framework for the fixed effects panel data models have recently been proposed by Bramati and Croux (2007), Namur and Luneburg (2011), Aquaro and Cizek (2013), Bakar and Midi (2015), Visek (2015) and Midi and Muhammad (2018). Bramati and Croux (2007) 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 (1984) and a combination of M and S estimates, namely, MS estimates of Maronna and Yohai (2000). Another robust estimation approach has been proposed in Aquaro and Cizek (2013) 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 (2002) and the reweighted LTS estimator of Cizek (2010) in the fixed effects linear panel data framework. Visek (2015) has developed a robust algorithm by weighting down the large order statistics of squared residuals. Moreover, Bakar and Midi (2015) have considered a robust centering method by employing the MM-centering procedure to the data. Then, the authors use the within group Generalized M Estimator (WGM) of Bramati and Croux (2007) to estimate the parameters of fixed effect panel data model. More recently, a Weighted Least Square (WLS) procedure based on MM-centering method, which is highly resistant in the presence of leverage points and vertical outliers, has been proposed by Midi and Muhammad (2018).

In this study, we concentrate on a class of robust estimators, i.e. M-estimators, 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., Hampel et al. (1986), Lindsay (1994), Wang et al. (2007) and Jiang et al. (2019)). Hence, it is critical to choose a dispersion function with an appropriate resistance level in obtaining efficient estimates of the parameters as noted in Wang et al. (2007). To determine the necessary level of robustness, a tuning parameter (also called the regularization parameter) in the dispersion function requires to be selected appropriately based on the possible proportion of outliers (please, see Wang et al. (2007) and Wang et al. (2018) 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. (2007) and Wang et al. (2018) 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 loss function has been proposed by Wang et al. (2013) in obtaining

penalized robust estimators with maximum efficiency and asymptotic breakdown point  $1/2$ . In this paper, we adapt the M-estimation techniques with automatic selection of tuning parameter introduced by Wang et al. (2007) and Wang et al. (2013) 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 centered 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 (2013). The asymptotic properties of the proposed M-estimation 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 M-estimators based on data-driven procedures yield more accurate and precise estimates of the model parameters compared to within group LS estimator and within group MS estimator (WMS) of Bramati and Croux (2007) 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 that of WMS method.

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 M-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 M-estimators are provided by an extensive simulation study and the results are compared with the robust WMS 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  $\{(y_{it}, x_{it}, \alpha_i), i = 1, \dots, N, t = 1, \dots, T\}$  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  $\beta \in \mathbb{R}^K$  are the vector of regression parameters. The subscript  $i$  denotes the individuals observed over time periods  $t$ . Finally, the  $\alpha_i$ 's and  $\varepsilon_{it}$ 's respectively represent the unobservable individual specific effects and the independent and identically distributed (i.i.d.) error terms with  $E(\varepsilon_{it}|x_{i1}, \dots, x_{iT}, \alpha_i) = 0$ ,  $E(\varepsilon_{it}^2|x_i, \alpha_i) = \sigma_\varepsilon^2 I_T$  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.

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

where  $y = (y_1, \dots, y_N)^T$  and  $\mathbf{X} = (x_1, \dots, x_N)^T$ , respectively, are an  $NT \times 1$  vector with  $y_i = (y_{i1}, \dots, y_{iT})$  and an  $NT \times K$  matrix of regressors with  $\mathbf{X}_i = (x_{i1}^T, \dots, x_{iT}^T)^T$ ,  $\alpha$  is an  $N \times 1$  vector consisting of the individual effects  $\alpha_i$  for  $i = 1 \dots N$ ,  $e_T$  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,  $\beta$ . The within group transformation removes the fixed effects by using the time averages of  $y_{it}$ ,  $x_{it}$  and  $\varepsilon_{it}$  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-centered data is obtained as follows.

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

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

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

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.** If, for any constant  $c \in \mathbb{R}$ ,

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

the estimator  $R$  is scale equivariant.

**Definition 2.2.** *If, for any  $K \times 1$  vector of constants  $v$ ,*

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

*it is regression equivariant.*

**Definition 2.3.** *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. (2007), Wang et al. (2013) and Wang et al. (2018) 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. (2013) 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. (2007) and Wang et al. (2018) 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 (2007) propose the WGM estimator and the WMS estimator, which have asymptotically breakdown point of  $1/4$ . In the

first approach, Bramati and Croux (2007) suggest robustly centring the variables by using the within group medians (as described in Equation (2)).

$$\begin{aligned}\tilde{y}_{it} &= y_{it} - \text{median}_t(y_{it}) \\ \tilde{x}_{it} &= x_{it} - \text{median}_t(x_{it})\end{aligned}\tag{2}$$

Next, LTS regression is suggested to employ on the centered 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 (2000) 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.

Recently, Aquaro and Cizek (2013) propose robust estimation procedures based on the pairwise difference transformations by applying the efficient weighted LS estimator (REWLS) of Gervini and Yohai (2002) and the reweighted LTS (RLTS) estimator of Cizek (2010). 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 (2007) and the REWLS and RLTS estimators of Aquaro and Cizek (2013) are similar and they have equal breakdown points according to a given data transformation, see Aquaro and Cizek (2013). 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 (2013).

#### **4. New Robust Estimators Using Well-Known Loss Functions with Automatic Selection of Tuning Parameter**

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 Wang et al. (2007) for details). In the M-estimation approach, the robustness is achieved by sacrificing

some of efficiency in presence of unnecessarily excess resistance (e.g., Hampel et al. (1986), Lindsay (1994), Wang et al. (2007) and Jiang et al. (2019)). 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 Wang et al. (2007) and Wang et al. (2018). 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. (2007) 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. (2018) 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. (2007) and Wang et al. (2013) 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 centered 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 (2013).

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 (Rousseeuw and Leroy (2003)). 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, Wang et al. (2018))

- Huber's function

$$\rho(u) = \begin{cases} \frac{u^2}{2} & \text{if } |u| \leq c \\ c|u| - \frac{c^2}{2} & \text{if } |u| > c \end{cases} \quad (3)$$

- Tukey's bisquare function

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

Also, Wang et al. (2013) propose to use 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

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



Then, the robust estimator of  $\beta$  is obtained as the solution of the following estimating functions by minimizing the loss functions  $\sum_{i=1}^N \sum_{t=1}^T \rho \left\{ \frac{(y_{it} - x_{it}^T \beta - \alpha_i)}{\hat{\sigma}} \right\}$  for a given estimate of the scale parameter,  $\hat{\sigma}$

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

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

- Huber's function

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

- Tukey's bisquare

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

- The ESL function

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

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

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

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

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

Since  $W$  is a function of  $\beta$  and  $\sigma$ , an iterative procedure has been followed at the  $k$ -th 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  $\delta \rightarrow 0$ , the consistent estimator  $\hat{\beta}_M$  is obtained by solving  $U(\beta) = 0$  and it follows that

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

where  $E\left(\ddot{\mathbf{X}}_i^T \ddot{\mathbf{X}}_i\right)$  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  $\psi$ , which has large values of the efficiency factor  $\tau$ , lead to some gain in efficiency of the estimator  $\beta$  as noted in Wang et al. (2007). Thus, the loss function  $\rho(\cdot)$  with the largest value of  $\tau$  should be chosen to obtain more efficient estimates of the parameters.

As noted in Bramati and Croux (2007) and Wang et al. (2007), 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 (2007), 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 `r1m` equals to 1.345 to obtain about 90% efficiency when the data follow Normal distribution, see Wang et al. (2007). 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  $c$ . Therefore, 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 the estimation efficiency. Thus, in this study, best value of the tuning parameter  $c$  is chosen by maximizing the value of  $\tau$ , and the nonparametric estimate of  $\tau$ , proposed by Wang et al. (2018), is calculated by

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

where  $\hat{e}_{it} = \frac{(y_{it} - \mathbf{x}_{it}^T \hat{\beta})}{\hat{\sigma}}$ ,  $\hat{\beta}$  and  $\hat{\sigma}$  denote the current estimates of  $\beta$  and  $\sigma$ , respectively.

Based on the above, we present the data-driven procedure introduced by Wang et al. (2007) to obtain the proposed robust estimators of  $\beta$  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,  $\beta$  for the panel data regression model.

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

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

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. (2018)) which has the largest efficiency factor  $\hat{\tau}(c)$ .

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

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. (2013) to the linear panel data models with fixed effects. For more detailed information, please see Wang et al. (2013).

Step 1. Let  $\mathbf{Z} = \{\ddot{x}_{it}, \dot{y}_{it}\}_{i=1, t=1}^{N, T}$  be a random sample that contains  $m$  bad points and  $NT - m$  good points denoted by  $\mathbf{Z}_m = \{Z_1, \dots, Z_m\}$  and  $\mathbf{Z}_{NT-m} = \{Z_{m+1}, \dots, Z_{NT}\}$ , respectively. Then, obtain the residuals  $e_{it}(\hat{\beta}_0) = \dot{y}_{it} - \dot{x}_{it}^T \hat{\beta}_0$  using the initial high breakdown coefficient  $\hat{\beta}_0$  for  $i = 1, \dots, N, t = 1, \dots, T$  and calculate median absolute deviation estimator (MAD),  $\hat{\sigma}_{MAD} = 1.4826 \times \text{median}_t |e_{it}(\hat{\beta}_0) - \text{median}_t(e_{it}(\hat{\beta}_0))|$ . Then, generate pseudo outlier set  $\mathbf{Z}_m = \{(\ddot{x}_{it}, \dot{y}_{it}) : |e_{it}(\hat{\beta}_0)| \geq 2.5(\hat{\sigma}_{MAD})\}$  where  $m = \#\{1 \leq i \leq N, 1 \leq t \leq T : |e_{it}(\hat{\beta}_0)| \geq 2.5(\hat{\sigma}_{MAD})\}$  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  $G = \{c : \zeta(c) \in (0, 1]\}$  where

$$\zeta(c_N) = \frac{2m}{NT} + \frac{2}{NT} \sum_{t=m+1}^T \rho_{c_N} \{e_{it}(\hat{\beta}_0)\} \quad (9)$$

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

$$\begin{aligned} \hat{I}(\hat{\beta}_0) &= \frac{2}{c} \left\{ \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \exp \left( -\frac{e_{it}^2(\hat{\beta}_0)}{c} \right) \left( \frac{2e_{it}^2(\hat{\beta}_0)}{c} - 1 \right) \right\} \\ &\times \left( \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \ddot{x}_{it} \ddot{x}_{it}^T \right), \\ \hat{\Sigma} &= \text{cov} \left\{ \exp \left( -\frac{e_{it}^2(\hat{\beta}_0)}{c} \right) \frac{2e_{it}(\hat{\beta}_0)}{c} \ddot{x}_{1t}, \dots, \exp \left( -\frac{e_{it}^2(\hat{\beta}_0)}{c} \right) \frac{2e_{it}(\hat{\beta}_0)}{c} \ddot{x}_{Nt} \right\}_{t=1}^T. \end{aligned}$$

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 (1987) to obtain the initial 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 (9). 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 Wang et al. (2013). 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. (2013) 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 Wang et al. (2013), 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 Wooldridge (2002)) as listed below:

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

*A2.  $\text{rank}\left(\sum_{t=1}^T E(\ddot{x}_{it}^T \ddot{x}_{it})\right) = \text{rank}[E(\ddot{\mathbf{X}}_i^T \ddot{\mathbf{X}}_i)] = K$  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  $\hat{\beta}_{MS}$  based on M-estimation techniques,

$$L(\beta) = \sum_{i=1}^N \rho_\tau \left( \frac{\ddot{y}_i - \ddot{\mathbf{X}}_i^T \beta}{S_N} \right) \quad (10)$$

where  $S_N$  denotes a scale parameter, and is a normalized MAD estimator obtained using the residuals  $e_{it} = \ddot{y}_{it} - \ddot{x}_{it}^T \hat{\beta}_0$  (cf. Jiang et al. (2019)). It should be noted that the proposed estimators  $\hat{\beta}_{MS}$  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  $E\|\ddot{\mathbf{X}}_i^3\| < \infty$ .*

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  $\tau$  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 Wang et al. (2007).

**Theorem 5.1.** *Let Assumption 5.2 hold. If  $S_N \xrightarrow{P} \sigma$  as  $N \rightarrow \infty$ , 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 “ $\xrightarrow{P}$ ” represents the convergence in probability.*

**Proof 5.1.** *For any  $\nu > 0$ , a large constant  $C$  exists satisfying*

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

*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,*

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

*where  $\ddot{\mathbf{y}}_i$  is a  $T \times 1$  vector. Then, by using the fact that  $S_N \xrightarrow{P} \sigma$ , the law of large numbers and central limit theorem, the proof of Theorem 5.1 directly follows from Theorem 1 of Jiang et al. (2019).*

**Remark 1.** *Theorem 5.1 guarantees the existence of a consistent estimator under some regularity conditions (cf. Jiang et al. (2019)). It should be noted that the existence of redescending M-estimator is not ensured for the unbounded loss function. Also, Maronna and Yohai (1981) 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  $\delta \rightarrow 0$ , if  $S_N \xrightarrow{P} \sigma$  as  $N \rightarrow \infty$ , then, we obtain*

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

*where “ $\xrightarrow{d}$ ” denotes the convergence in distribution.*

**Proof 5.2.** *Let assume*

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

*Then, we have the following by Taylor's expansion*

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

*where  $\beta^*$  is a vector that lies between  $\hat{\beta}_M$  and  $\beta$ , 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 10 (cf. Jiang et al. (2019)). Under A1-A2 of Assumption 5.2, we obtain*

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

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

## 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 IntelCore 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

$$y_{it} = x_{it}^T \beta + \alpha_i + \varepsilon_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (11)$$

where  $\varepsilon_{it}$ 's are assumed to be iid  $N(\mu = 0, \sigma^2 = 1)$  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 (2013), 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  $\gamma$  and random parts  $\eta_i$  of  $\alpha_i$ . To avoid completely symmetric design as in Aquaro and Cizek (2013), 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  $\chi_2^2$  denote the chi-squared distribution with 2 degrees of freedom.

The performances of the proposed estimators are evaluated based on  $S = 1000$  simulations. In case of the changing sample sizes and presence of outliers, we compute the mean squared errors (MSE):  $MSE = \frac{1}{S} \sum_{s=1}^S \|\hat{\beta}^s - \beta\|^2$ , where  $\hat{\beta}^s, s = 1, \dots, S$ , represents the estimates obtained from  $S$  simulated samples. Also, we calculated the squared errors (SE):  $SE = \|\hat{\beta}^s - \beta\|^2$  to examine the effect of the error distributions on the estimators. Note that throughout the experiments, all calculations are performed for the finite sample breakdown points  $BP = 0.1$  and  $BP = 0.5$  for WMS estimator of Bramati and Croux (2007). 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 and proposed loss functions (Huber, Tukey, Exponential) based estimators using data-dependent regularization parameters overlap as  $N \rightarrow \infty$  and  $T \rightarrow \infty$  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 values of  $T$  as noted in Aquaro and Cizek (2013). 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  $\chi_{(4)}^2$  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 Aquaro and Cizek (2013). 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 and LS estimators have similar SE values under all the error distributions, except for standard Cauchy distribution. Also, the mean and variance of the SE values for LS method increases more than the other estimators in case of the Cauchy distribution of the error terms while the proposed robust estimators and WMS of Bramati and Croux (2007) are 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  $m = 12$  and  $m = 24$ . Two main types of contamination, namely, random contamination and concentrated (or clustered) contamination as in Bramati and Croux (2007) 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 (2007). 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 1-2. 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) and WMS of Bramati and Croux (2007) 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 procedure. 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 for  $(N_1, T_1) = (120, 2)$ . The performances of the robust estimators can be sorted based on minimizing MSE values as follows: Tukey, Exponential, Huber estimators and WMS for  $T_1 = 2$ . 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 Aquaro and Cizek (2013) and Bramati and Croux (2007). Although there are no substantial differences between WMS and Exponential loss function based estimator at 5% contamination for  $(N_2, T_2) = (80, 3)$ , 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 estimator especially for increasing level of contamination.

To confirm the supremacy over the traditional LS and WMS 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

$$RMSE = \frac{1}{S} \sum_{s=1}^S \sqrt{\sum_{i=1}^{N_{test}} \sum_{t=1}^T (y_{it}^s - \hat{y}_{it}^s)^2 / (N_{test} \times T)} = \frac{1}{S} \sum_{s=1}^S \sqrt{\sum_{t=1}^{T_i} (y_{it}^s - \hat{y}_{it}^s)^2 / T_i} \quad (12)$$

where  $i = 1, 2, \dots, N_{test}$ ,  $y_{it}^s$  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  $N_{test} \times T_1$  for  $(N_1, T_1) = (120, 2)$  and  $N_{test} \times T_2$  for  $(N_2, T_2) = (80, 3)$  with the choice of

$N_{test} = 50$ . 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 Croissant and Millo (2008)). For this panel, the gasoline demand model studied by Baltagi and Griffin (1983) 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  $i = 1, 2, \dots, 18$  and  $t = 1, 2, \dots, 19$  denote the OECD countries and years, respectively; see Baltagi and Griffin (1983) 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 WMS estimators produce more precise estimates of individual coefficients among the robust procedures. However, the proposed methods, especially Exponential loss function based estimator, provide competitive results in estimating  $\beta_2$  and  $\beta_3$ .

To investigate the predictive performances of the LS, WMS 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 ( $N - N_{test} = 15$ ) are used to build a predictive model, and the logarithms of the gasoline consumption of remaining 3 countries ( $N_{test} = 3$ ) are predicted by using the estimated model coefficients. This process is performed  $S = 100$  times, and for each time, RMSEs are obtained by using the Equation (12). 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 method.

## 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 (2007) 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 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.

## References

## References

- Aquaro, M., Cizek, P., 2013. One-step robust estimation of fixed-effects panel data models. *Computational Statistics and Data Analysis* 57, 536–548.
- 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., 2005. *Econometric Analysis of Panel Data*. John Wiley and Sons, Chichester.
- Baltagi, B.H., Griffin, J.M., 1983. Gasoline demand in the oecd: An application of pooling and testing procedures. *European Economic Review* 22, 117–137.
- Bickel, R., 2007. *Multilevel Analysis for Applied Research: It's Just Regression*. The Guilford Press, New York.
- Bramati, M.C., Croux, C.P., 2007. Robust estimators for the fixed effects panel data model. *Econometric Journal* 10, 521–540.
- Cizek, P., 2008. Robust and efficient adaptive estimation of binary-choice regression models. *Journal of the American Statistical Association* 103, 687–696.

- Cizek, P., 2010. Reweighted least trimmed squares: an alternative to one-step estimators. CentER Discussion Paper Series 91/2010 .
- Croissant, Y., Millo, G., 2008. Panel data econometrics in r: The plm package. *Journal of Statistical Software* 27, 1–43.
- Davies, P.L., Gather, U., 2005. Breakdown and groups. *The Annals of Statistics* 33, 977–1035.
- Diggle, P.J., Heagerty, P., Liang, K.Y., Zeger, S.L., 2002. *Analysis of Longitudinal Data*. Oxford University Press, United Kingdom.
- Fitzmaurice, G.M., M.Laird, N., Ware, J.H., 2004. *Applied Longitudinal Analysis*. John Wiley and Sons, New York.
- Genton, M.G., Lucas, A., 2003. Comprehensive definitions of breakdown points for independent and dependent observations. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 65, 81–94.
- Gervini, D., Yohai, V.J., 2002. A class of robust and fully efficient regression estimators. *The Annals of Statistics* 30, 583–616.
- Greene, W.H., 2003. *Econometric Analysis*. Prentice Hall, New Jersey.
- Greene, W.H., 2017. *Econometric Analysis*. Prentice Hall, New York.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 1986. *Robust Statistics: The Approach Based on Influence Functions*. John Wiley and Sons, New York.
- Hill, R.C., Griffiths, W.E., Lim, G.C., 2007. *Principles of Econometrics*. John Wiley and Sons, USA.
- Hsiao, C., 1985. Benefits and limitations of panel data. *Econometric Reviews* 4, 121–174.
- Hsiao, C., 2007. Panel data analysis-advantages and challenges. *Test* 16, 1–22.
- Jiang, Y., Wang, Y.G., Fu, L., Wang, X., 2019. Robust estimation using modified huber’s functions with new tails. *Technometrics* 61, 111–122.
- Kutner, M.H., Nachtsheim, C.J., Neter, J., 2004. *Applied Linear Regression Models*. McGraw-Hill Education, New York.
- Lindsay, B., 1994. Efficiency versus robustness: The case for minimum hellinger distance and related methods. *Annals of Statistics* 22, 1018–1114.

- Maddala, G.S., Mount, T.D., 1973. A comparative study of alternative estimators for variance components models used in econometric applications. *Journal of the American Statistical Association* 68, 324–328.
- Maronna, R.A., Martin, R.D., Yohai, V.J., 2006. *Robust Statistics. Theory and Methods*. John Wiley and Sons, New York.
- Maronna, R.A., Yohai, V.J., 1981. Asymptotic behavior of general m-estimates for regression and scale with random carriers. *Probability Theory and Related Fields* 58, 7–20.
- Maronna, R.A., Yohai, V.J., 2000. Robust regression with both continuous and categorical predictors. *Journal of Statistical Planning and Inference* 89, 197–214.
- Midi, H., Muhammad, S., 2018. Robust estimation for fixed and random effects panel data models with different centering methods. *Journal of Engineering and Applied Sciences* 13, 7156–7161.
- Mundlak, Y., 1978. On the pooling of time series and cross section data. *Econometrica* 46, 69–85.
- Namur, V.V., Luneburg, J.W., 2011. Robust estimation of linear fixed effects panel data models with an application to the exporter productivity premium. *Jahrbucher f. National okonomie u. Statistik* 231, 546–557.
- Rousseeuw, P.J., 1984. Least median of squares regression. *Journal of the American Statistical Association* 79, 871–880.
- Rousseeuw, P.J., Leroy, A.M., 1987. *Robust Regression and Outlier Detection*. John Wiley and Sons, Chichester.
- Rousseeuw, P.J., Leroy, A.M., 2003. *Robust Regression and Outlier Detection*. John Wiley and Sons, New York.
- Rousseeuw, P.J., Yohai, V.J., 1984. Robust regression by means of s-estimators, in: Franke, J., Hardle, W., Martin, R.D. (Eds.), *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statistics*. Springer, New York.
- Rousseeuw, P.J., van Zomeren, B.C., 1990. Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association* 85, 633–639.
- Simpson, D.G., Ruppert, D., Carroll, R.J., 1992. On one-step gm estimates and stability of inferences in linear regression. *Journal of the American Statistical Association* 87, 439–450.
- Visek, J.A., 2015. Estimating the model with fixed and random effects by a robust method. *Methodology and Computing in Applied Probability* 17, 999–1014.

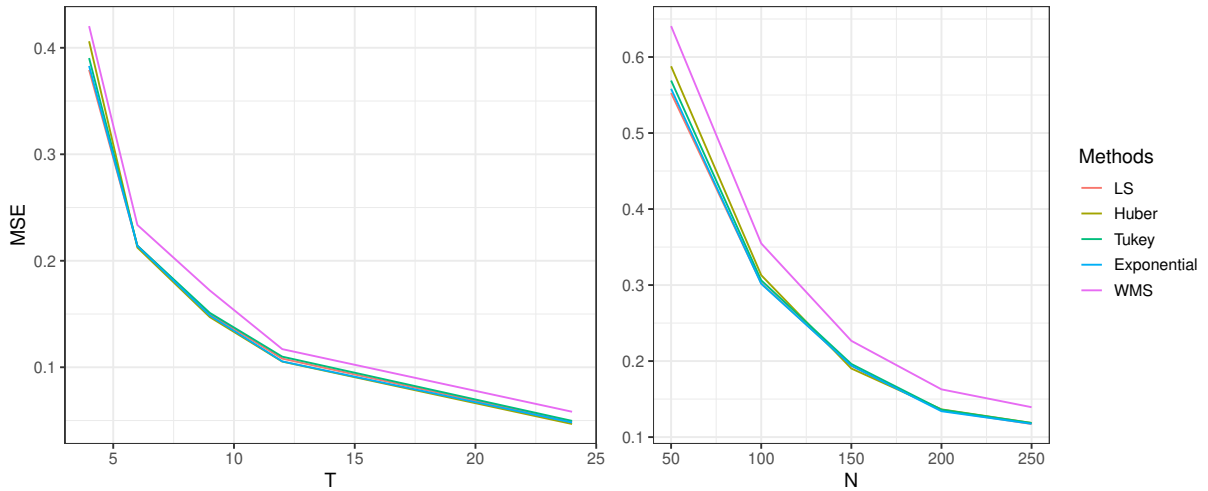
- Wagenvoort, R., Waldmann, R., 2002. On b-robust instrumental variable estimation of the linear model with panel data. *Journal of Econometrics* 106, 297–324.
- Wallace, T.D., Hussain, A., 1969. The use of error components models in combining cross section and time-series data. *Econometrica* 37, 55–72.
- Wang, N., Wang, Y.G., Hu, S., Hu, Z.H., Xu, J., Tang, H., Jin, G., 2018. Robust regression with data-dependent regularization parameters and autoregressive temporal correlations. *Environmental Modeling and Assessment* 23, 779–786.
- Wang, X., Jiang, Y., Huang, M., Zhang, H., 2013. Robust variable selection with exponential squared loss. *Journal of the American Statistical Association* 108, 632–643.
- Wang, Y.G., Lin, X., Zhu, M., Bai, Z., 2007. Robust estimation using the huber function with a data-dependent tuning constant. *Journal of Computational and Graphical Statistics* 16, 468–481.
- Wooldridge, J.M., 2002. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge.
- Yohai, V.J., 1987. High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics* 15, 642–656.

**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  $m = 12, 24$

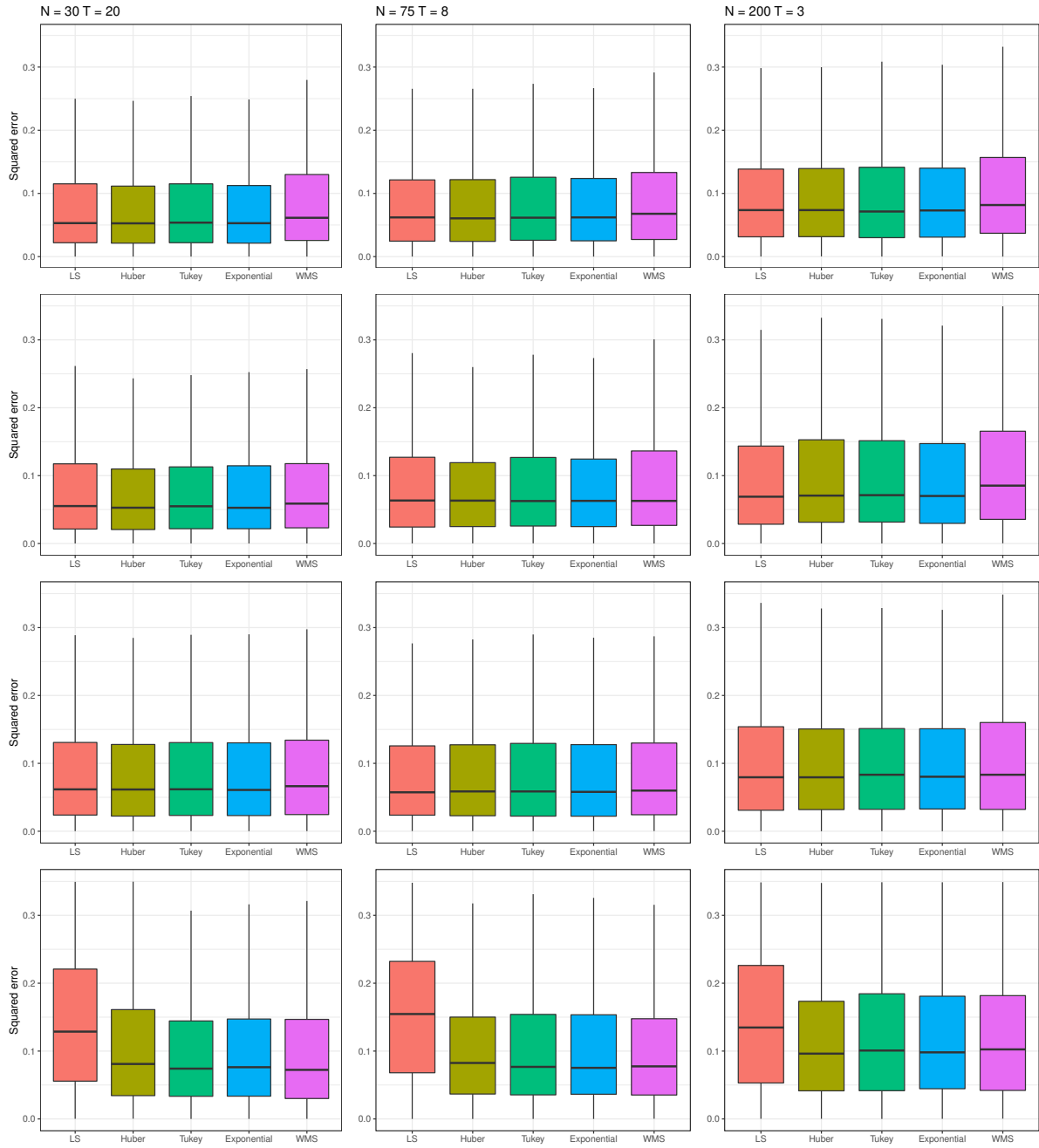
$(N_1, T_1) = (120, 2)$		Random contamination				Concentrated contamination			
Method		Vertical outliers		Leverage points		Vertical outliers		Leverage points	
		5%	10%	5%	10%	5%	10%	5%	10%
LS		1.651	2.579	7.688	9.116	2.985	5.916	25.231	30.581
Huber		0.710	0.941	0.678	0.942	0.746	0.987	0.771	0.963
Tukey		0.609	0.842	0.614	0.828	0.577	0.642	0.569	0.630
Exponential		0.637	0.848	0.618	0.823	0.595	0.668	0.602	0.673
WMS		0.762	1.571	0.727	1.547	1.009	3.951	1.102	3.644
$(N_2, T_2) = (80, 3)$									
LS		1.172	1.982	7.464	8.830	2.467	4.150	25.422	30.490
Huber		0.575	0.806	0.606	0.824	0.599	0.902	0.652	0.922
Tukey		0.522	0.713	0.559	0.719	0.467	0.649	0.530	0.642
Exponential		0.517	0.706	0.558	0.709	0.477	0.695	0.542	0.704
WMS		0.510	0.755	0.522	0.773	0.460	0.926	0.485	0.947

**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  $m = 12, 24$

$(N_1, T_1) = (120, 2)$		Random contamination				Concentrated contamination			
Method		Vertical outliers		Leverage points		Vertical outliers		Leverage points	
		5%	10%	5%	10%	5%	10%	5%	10%
LS		4.815	4.918	5.134	5.227	4.923	5.128	6.078	6.332
Huber		4.744	4.788	4.722	4.753	4.738	4.764	4.702	4.761
Tukey		4.739	4.782	4.718	4.744	4.722	4.734	4.687	4.730
Exponential		4.740	4.784	4.718	4.742	4.724	4.736	4.690	4.734
WMS		4.748	4.838	4.727	4.801	4.766	4.992	4.732	4.968
$(N_2, T_2) = (80, 3)$									
LS		5.560	5.645	6.014	6.081	5.650	5.815	7.008	7.325
Huber		5.508	5.529	5.531	5.540	5.480	5.533	5.482	5.507
Tukey		5.503	5.518	5.526	5.529	5.468	5.510	5.471	5.483
Exponential		5.503	5.518	5.526	5.528	5.469	5.515	5.472	5.489
WMS		5.503	5.524	5.523	5.536	5.470	5.547	5.466	5.522

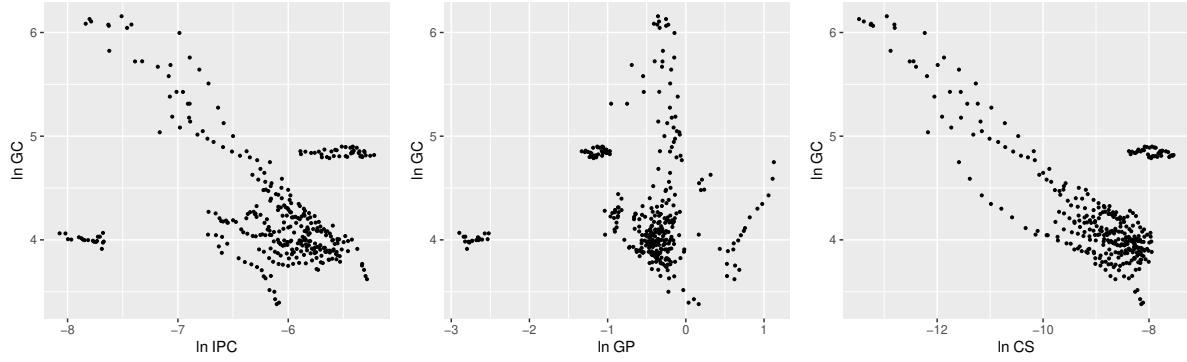


**Figure 1:** The MSEs of all estimators when  $T = 4, 6, 9, 12, 24$  for fixed cross-sectional dimension  $N = 50$  and  $N = 50, 100, 150, 200, 250$  for fixed time dimension  $T = 3$  and  $\epsilon_{it} \sim N(\mu = 0, \sigma^2 = 1)$

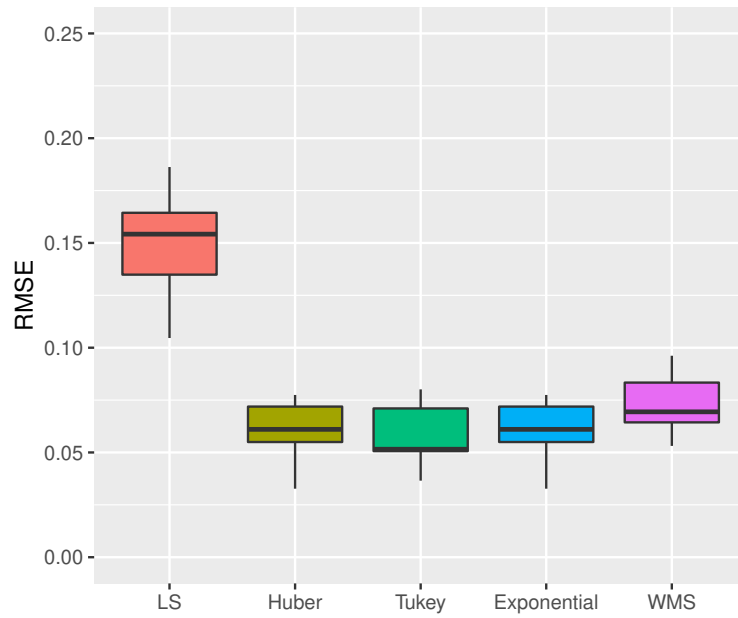


**Figure 2:** The boxplots of the SEs of all estimators under different error distributions when  $(N_1, T_1) = (30, 20)$ ,  $(N_2, T_2) = (75, 8)$  and  $(N_3, T_3) = (200, 3)$ . First row:  $N(0, 1)$  distribution, second row:  $t_5$  distribution, third row:  $\chi^2_{(4)}$  distribution and fourth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





**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.



**Figure 4:** The boxplots of the RMSEs of predicted and observed values of the logarithm of the gasoline consumption for 3 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	$\beta_1$	$\beta_2$	$\beta_3$
LS	-1.043 (0.054)	-0.264 (0.066)	0.113 (0.021)
Huber	0.579 (0.048)	-0.317 (0.029)	-0.559 (0.019)
Tukey	0.479 (0.040)	-0.302 (0.024)	-0.451 (0.016)
Exponential	0.482 (0.027)	-0.307 (0.018)	-0.452 (0.012)
WMS	0.902 (0.017)	-0.356 (0.014)	-0.651 (0.011)