Received: 10 September 2019 | Revised: 16 July 2020 | Accepted: 27 July 2020

DOI: 10.1002/sim.8732

RESEARCH ARTICLE

Statistics in Medicine WILEY

Robust estimation for linear panel data models

Beste Hamiye Beyaztas[1] | **Soutir Bandyopadhyay**[2]

[1]Department of Statistics, Istanbul Medeniyet University, Istanbul, Turkey

[2]Department of Applied Mathematics, Statistics Colorado School of Mines, Golden, Colorado, USA

Correspondence
Beste Hamiye Beyaztas, Department of Statistics, Faculty of Engineering and Natural Sciences, Istanbul Medeniyet University, Istanbul, Turkey.
Email: beste.sertdemir@medeniyet.edu.tr

Funding information
National Science Foundation, Grant/Award Number: NSF DMS-1811384

In different fields of applications including, but not limited to, behavioral, environmental, medical sciences, and econometrics, the use of panel data regression models has become increasingly popular as a general framework for making meaningful statistical inferences. However, when the ordinary least squares (OLS) method is used to estimate the model parameters, presence of outliers may significantly alter the adequacy of such models by producing biased and inefficient estimates. In this work, we propose a new, weighted likelihood based robust estimation procedure for linear panel data models with fixed and random effects. The finite sample performances of the proposed estimators have been illustrated through an extensive simulation study as well as with an application to blood pressure dataset. Our thorough study demonstrates that the proposed estimators show significantly better performances over the traditional methods in the presence of outliers and produce competitive results to the OLS based estimates when no outliers are present in the dataset.

KEYWORDS
fixed effects, least squares, panel data, random effects, robust estimation, weighted likelihood

1 | INTRODUCTION

Panel data, also known as longitudinal data in biological sciences, are two-dimensional data in which cross-sectional measurements are observed over time. These type of data typically allow us to take into account the unobserved individual-specific heterogeneity as well as the intra-individual dynamics (cf. References 1 and 2 for more details), and therefore, in general, are more informative and yield more degrees of freedom, less collinearity between the variables and more efficiency than a single cross-sectional or time-series data, thereby improving the accuracy and precision in the inference of model parameters.

Since the seminal paper of Reference 3, panel data have received growing attention in many empirical and methodological studies. As pointed out in Reference 4, the main sources leading to the improvements in panel data studies include (a) increased availability of such data, (b) better capability to model the complexity of human behavior than a pure cross-section or time series data, and (c) demanding methodology. In this context, the linear panel data regression models have become most widely applied statistical methods to analyze two-dimensional data in many fields, such as econometrics, biostatistics, and so on. For a comprehensive review on static linear panel data models and its applications in different areas, see References 1 and 5-14, and the references therein.

In panel data studies, three main sources of variability are generally considered, namely, (a) the within variation, that is, the variation from observation to observation in each of cross-sectional unit, (b) the between variation, that is, the variation in observations from an individual unit to another individual unit, and (c) the overall variation, that is, the variation over both dimensions, since panel data include the information over two dimensions, cross-sectional and time

series (cf. References 15 and 16 for details). The statistical appeal of panel data models typically lies in the fact that these models focus particularly on explaining within variations over time and provide controls over individual heterogeneity. The most commonly used panel data models are fixed and random effects models (cf. References 15 and 17). In the fixed effect approach, subject-specific means (individual-specific effects, individual heterogeneity), which belong to each cross-sectional unit, are assumed to be fixed and are included as time-invariant intercept terms in the regression model, while these may vary across subjects. On the other hand, in random effect models, individual heterogeneity is explained by the differences in the error variance components. As noted in Reference 10, the main difference between the fixed and random effects models is that the fixed effects model assumes that the time-invariant characteristics of individuals are correlated with the covariates, whereas random effects model does not allow such correlation.

Typically, one uses ordinary least squares (OLS) methods for making statistical inferences regarding the parameters of linear panel data regression models. However, to obtain consistent estimates of the model parameters, traditional estimation techniques require some assumptions such as strict exogeneity with respect to the error terms and homoscedasticity of the error terms, which are rarely fulfilled in practice. Hence, the classical OLS estimators may considerably be affected due to any departure from the model assumptions as well as the presence of outliers. The outlying observations are generally masked due to the complex nature of the data and not directly detectable using standard outlier diagnostics. Moreover, the OLS based estimators are highly sensitive to the leverage points due to the distortions being caused by the outliers in the covariates. Thus, the well-known estimators, such as generalized least squares (GLS) estimator for random effects model and fixed effects estimators based on several transformations, may lead us to incorrect and unreliable results. To overcome these issues, Reference 18 has considered alternatives to the fixed effect estimator for the purpose of building highly robust procedures with high breakdown point. More recently, Reference 19 has proposed a new estimation procedure based on two different data transformations by applying standard robust estimation methods in the fixed effects linear panel data framework. A robust algorithm based on the idea of weighting down the large order statistics of squared residuals has been proposed in Reference 20 to obtain reliable estimates of the model parameters. To the best of our knowledge, only a few studies considering the robustness of conventional estimation methods are available in the context of static linear panel data models; see, for instance, References 18-22.

This article aims to study the impacts of outlying observations on the OLS based estimation methods (such as between, pooled OLS, fixed effects and random effects estimators as discussed in Section 2) in linear panel data models and suggest robust alternatives to these estimation procedures. The proposed weighted likelihood based estimators, based on weighted likelihood estimating equations introduced in Reference 23, produce more robust estimates compared to their traditional counterparts in the presence of outlier(s) or in case of any departure from model assumptions and their asymptotic properties are equivalent to the OLS based techniques when no outliers are present in the data. In this study, we focus on the impacts of several types of outliers including vertical outliers and leverage points (random and concentrated) on the estimation procedures. Monte Carlo experiments under different data generating processes and contamination schemes are used to compare the finite sample performances of the proposed estimators and traditional OLS based estimators. The numerical results support that the proposed methods yield more accurate and precise estimates compared to the OLS estimators when the data have outliers.

The rest of the article is organized as follows. We start with providing details about the static linear panel data models and discuss the OLS based estimation methods commonly used to estimate the parameters (cf. Section 2.1). In Section 2.2, we describe the estimation method based on weighted likelihood and subsequently propose the robust counterparts of the OLS estimators. The finite sample properties of the proposed methods are illustrated through an extensive simulation study and the results are compared with traditional estimation methods in Section 3. To further validate the applicability of our proposed methods, we apply those to blood pressure data. The results are presented in Section 4.

2 | LINEAR PANEL DATA MODELS

Let us consider the linear panel data regression model with a random sample as follows.

$$y_{it} = x_{it}'\beta + \alpha_i + \epsilon_{it}, \quad i = 1, \ldots, N \quad t = 1, \ldots, T,$$

where the subscript i represents an individual observed at time t, α_i s are the unobserved individual-specific effects (time-invariant characteristics), β is a $K \times 1$ vector of coefficients and an element of the parameter space Θ, y_{it} and x_{it}s are the response variable and the K-dimensional vector of explanatory variables, respectively, and ϵ_{it}s are the

independent and identically distributed (iid) error terms with $E(\varepsilon_{it}|x_{i1}, \ldots, x_{iT}, \alpha_i) = 0$, $E(\varepsilon_{it}^2|x_{i1}, \ldots, x_{iT}, \alpha_i) = \sigma_\varepsilon^2$ and $E(\varepsilon_{it}\varepsilon_{is}|x_{i1}, \ldots, x_{iT}, \alpha_i) = 0$ for $t \neq s$. The above panel data regression model can be represented in matrix form as follows.

$$y = \alpha \otimes e_T + X\beta + \varepsilon,$$

where $y = (y_1, \ldots, y_N)'$ is an $NT \times 1$ vector obtained by stacking observations $y_i = (y_{i1}, \ldots, y_{iT})'$ for individual $i = 1, \ldots, N$, $X = (x_1, \ldots, x_N)'$ is an $NT \times K$ matrix of regressors with $x_i = (x_{i1}', \ldots, x_{iT}')'$, α is an $N \times 1$ vector consisting of the individual effects α_i for $i = 1 \ldots N$, e_T is a $T \times 1$ vector of ones, and $\otimes$ denotes the kronecker product.

In fixed effects models, only variation within each cross-sectional unit is exploited (cf. References 15-17). Thus, in the presence of small or no within variation, the coefficients of the regressors in fixed effects models cannot be correctly estimated or identified, as noted in Reference 24. The fixed effects models generally allow for possible correlations between individual-specific unobservable effects and independent variables by including dummy variables for different intercepts, allowing a limited form of endogeneity (cf. Reference 24) while yielding unbiased estimates of the regression parameters (cf. References 6,16,17, and 25). The information on both within and between variations are included by the random effects models. In random effects models, the individual-specific effects are being included in the model as a part of the disturbance, and these are required to be uncorrelated with the explanatory variables and the error terms (cf. References 6 and 9). In particular, if α_i is assumed to be random then, the random effects model can be formulated as follows.

$$y_{it} = x_{it}'\beta + \alpha_i + \varepsilon_{it} = x_{it}'\beta + \nu_{it}, \quad \alpha_i \sim \text{iid}(0, \sigma_\alpha^2), \quad \varepsilon_{it} \sim \text{iid}(0, \sigma_\varepsilon^2), \tag{1}$$

where $\nu_{it} = \alpha_i + \varepsilon_{it}$ denotes a compound error term with $\sigma_\nu^2 = \sigma_\alpha^2 + \sigma_\varepsilon^2$ and $cov(\nu_{it}, \nu_{is}) = \sigma_\alpha^2$ for $t \neq s$. α_is are assumed to be uncorrelated with ε_{it} and x_{it}.

The pooled regression model

$$y_{it} = \alpha + x_{it}'\beta + \varepsilon_{it},$$

is a restricted type of panel data model such that the regression coefficients, that is, α and β, have the common values to all cross-sectional units for all time periods as noted in References 17 and 26.

Finally, before we describe the between regression models, let $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$, $\bar{x}_i = T^{-1}\sum_{t=1}^{T} x_{it}$, and $\bar{\varepsilon}_i = T^{-1}\sum_{t=1}^{T} \varepsilon_{it}$, respectively, denote the time averages of y_{it}, x_{it}, and ε_{it} for the ith cross-sectional unit. By considering the N linear regression models based on the time averages of each cross-sectional unit, the between model is defined as follows:

$$\bar{y}_i = \alpha_i + \bar{x}_i\beta + \bar{\varepsilon}_i. \tag{2}$$

The between regressions are frequently used to investigate the long-run relationships by ignoring all the information owing to intra-subject variability (cf. References 24, 27, and 28). For example, Reference 29 has examined the elasticity of demand for some countries and compared the estimates obtained using within and between regressions. The results obtained from the between country model can be interpreted as long run effects, whereas the short run effects are captured by the within country regression model. Additionally, Reference 27 has reported the results of elasticity estimates for short run price and long run price, and compared the estimates in terms of mean, standard deviation, and root mean square error (RMSE) criteria. It has been emphasized that the between estimator has a better performance according to the RMSE criterion for estimates of long run elasticity price than that of short run price elasticity.

Next, we briefly discuss the commonly used estimation procedures for above mentioned linear panel data models.

2.1 | Traditional estimation methods

The estimation procedures commonly applied in linear panel data models discussed above, can be examined within the scope of OLS estimation as noted in Reference 28. The OLS-based estimation techniques mainly rely on the type of variations (cf. Reference 16).

The pooled OLS estimator is simply the implementation of the OLS method to the linear model on the pooled data across two dimensions by completely disregarding the panel structure of the data. Therefore, $\hat{\beta}_{pols}$, the pooled OLS

estimator of β, can be obtained as follows.

$$\hat{\beta}_{pols} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}' x_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}' y_{it} \right).$$

As noted in References 6, 14, and 24, the pooled OLS method provides consistent estimates of the parameters for random effects and pooled regression models under the independence assumption of explanatory variables and error terms. On the other hand, it is severely biased and inconsistent for the fixed effects model due to the inclusion of the individual-specific effects, which are correlated with the explanatory variables. Also, while investigating the bias and efficiency of some well-known panel data estimators in observational health studies, Reference 30 raised two main concerns for pooled estimator, namely, the heteroscedastic error terms and the bias caused by the omitted individual-specific effects.

The estimation of the fixed effects model requires the time-demeaned data. The fixed effects transformed model for the mean-centered data is obtained as follows.

$$\ddot{y}_{it} = \ddot{x}_{it}' \beta + \ddot{\varepsilon}_{it},$$

where $\ddot{y}_{it} = y_{it} - \bar{y}_i$, $\ddot{x}_{it} = x_{it} - \bar{x}_i$, and $\ddot{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_i$, so that the individual-specific effects have been eliminated. Under the assumptions of fixed effects model, $\hat{\beta}_{fe}$, the fixed effects estimator of β, can be obtained as

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

The fixed effects estimator (also known as the within estimator) is consistent for the fixed effects model when time dimension, T gets large (cf. Reference 1). Furthermore, as noted in Reference 31, the precisions of the fixed effects estimates are significantly affected when the independent variables vary greatly across individual units and simultaneously exhibit small variation over time for each individual.

The fixed effects least squares method has several shortcomings. First, the fixed effects estimators suffer from the incidental parameter problem (see, Reference 32 for more details). If the cross-sectional dimension, N, is significantly large, the fixed effects estimators of individual-specific effects become biased and inconsistent because of the increasing number of these parameters (cf. References 1, 6, and 33). Furthermore, the fixed effects method is incapable of estimating the coefficients of the time-invariant variables since it only considers the variation within cross-sections. Therefore, the explanatory power of the model decreases with less efficient estimates (see References 24 and 31 for more details). Another possible drawback is that several dummies used for time-invariant variables such as gender, race, geographic location, education, religion cause to aggravate collinearity among the regressors as noted in Reference 1. Furthermore, Reference 34 emphasizes that the fixed effects methods underestimate the model parameters and result in drastically biased inference since the measurement errors get magnified in within dimension. The random effects model compensates for some of these problems encountered in fixed effects least squares.

The GLS method is used for estimating random effect model to deal with the autocorrelation in the error terms caused by the individual-specific effects. This method entails the quasi demeaning transformation of the variables to obtain the homoscedastic variance-covariance matrix for achieving efficiency, as noted in References 28 and 35. As emphasized in Reference 1, the GLS method asymptotically provides the best linear unbiased estimator if the variance-covariance matrix of the disturbance term is known. Also, Reference 28 points out that it produces the equivalent estimates of β to the OLS method on the quasi-demeaned data. The quasi-demeaning transformation contains subtracting the time-averages, weighted by using the variances of the idiosyncratic errors and individual effects, from the original variables. The transformed version of the random effects model is expressed as,

$$\tilde{y}_{it} = \tilde{x}_{it}' \beta + \tilde{v}_{it},$$

where $\theta = 1 - \left[\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_a^2} \right]^{1/2}$, $\tilde{v}_{it} = v_{it} - \theta \bar{v}_i$, $\tilde{y}_{it} = y_{it} - \theta \bar{y}_i$, and $\tilde{x}_{it} = x_{it} - \theta \bar{x}_i$ with the time averages $\bar{y}_i$ and $\bar{x}_i$. By running OLS method on the transformed model, the GLS estimator (also called as random effects estimator), $\hat{\beta}_{re}$ can be obtained

as follows.

$$\hat{\beta}_{re} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}'_{it} \tilde{x}_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}'_{it} \tilde{y}_{it} \right).$$

Note that the fixed effects and pooled OLS estimators can be obtained by employing OLS method on the transformed model for $\theta = 1$ with $\sigma_\epsilon^2 = 0$ and $\theta = 0$ with $\sigma_\alpha^2 = 0$, respectively, as the special cases of the above mentioned GLS estimator.

When the random effects model as in Equation (1) is appropriate, both fixed effects and random effects estimators are consistent but random effects method provides more efficient estimates, with high explanatory power, compared to the fixed effects method. This is due to fact that the random effects estimators have the advantages of using both within and between variations, hence, these can be viewed as a weighted average of the between and fixed effects estimators (cf. References 16 and 24). However, there is a trade-off between bias and efficiency, and the random effects method is more vulnerable to omitted variable bias than the fixed effects method, as noted in Reference 31. In case of no omitted variables, the random effects model is generally preferred over the fixed effects model because it allows for estimating the effects of time-invariant variables; see Reference 16.

The between regression models (2) include the information reflected in the differences between cross-sections. A large between variation generally indicates the differences in means of the variables over time for each subject as noted in Reference 29. By employing the OLS method on the between regression model, the between estimator, $\hat{\beta}_{be}$, is obtained as follows.

$$\hat{\beta}_{be} = (\bar{x}'_i \bar{x}_i)^{-1} (\bar{x}'_i \bar{y}_i).$$

Although the between estimator generates consistent results for the pooled and random effects models, it is rarely preferred in practice since the pooled and random effects estimators yield more efficient results compared to the between estimator, see Reference 24. Furthermore, as noted in Reference 17, it can be used to estimate the effects of time-invariant variables in fixed effects model, but with biased estimates of the effects of both time-invariant and time-variant variables.

In spite of the fact that all of the traditional methods discussed above suffer heavily due to the presence of outlying observations, the existing literature on the robust methods to estimate static panel data models is fairly limited. Recently, a few different approaches within the robust estimation framework for the fixed effects panel data models have been developed by utilizing the generalized M-estimation and least trimmed squares (LTS) techniques; see, for example, References 18 and 19. Reference 18 defined the robust versions of fixed effects estimator with high breakdown point by extending some known robust regression estimators, such as LTS estimator of Reference 36 and a combination of M and S estimates of Reference 37. Another robust estimation approach has been proposed in Reference 19 based on two different data transformations by employing the efficient weighted least squares estimator of Reference 38 and the reweighted LTS estimator of Reference 39 in the context of linear regression model.

Next, in Section 2.2, we propose robust alternatives of the OLS based estimation procedures, which are highly sensitive to the presence of outliers, erroneous observations and any departure from the distributional assumptions on the error terms. Our approach is primarily based on the weighted likelihood estimating equations methodology introduced in Reference 23. The main idea behind the weighted likelihood methodology is to replace the maximum likelihood (ML) equations with weighted score equations, in which the weights come from minimum disparity estimation as in Reference 40, for obtaining efficient estimates and reducing the effects of outliers on the score equations. Note that the ML method (and its weighted version) provides a flexible framework for the purposes of likelihood based model specification testing and estimation in the presence of endogeneity problem leading to correlation between regressors and error terms. Also, it eliminates the incidental parameters problem over time; see Reference 41. Furthermore, the ML estimator is equivalent to GLS estimator under the assumptions of homoscedasticity, no-autocorrelation, and normally distributed error terms (cf. Reference 42).

2.2 | New robust weighted likelihood based estimation procedure

As defined earlier, let $\{y_1, \ldots, y_N\}'$ be an iid sample of $NT \times 1$ vector, and $\{x_1, \ldots, x_N\}'$ denote an $NT \times K$ matrix of predictors with $x_i = \{x'_{i1}, \ldots, x'_{iT}\}'$. Let us consider the random effects model given in Equation (1) with density function

$f = f(y_{it}; x_{it}, \beta)$ and define the joint probability density for disturbance terms, $\varepsilon_i + \alpha_i e_T = y_i - x_i\beta$, as given below.

$$f(\varepsilon_i + \alpha_i e_T) = (2\pi)^{-\frac{T}{2}} |\mathbf{\Omega}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y_i - x_i\beta)'\mathbf{\Omega}^{-1}(y_i - x_i\beta)\right\}.$$

Then, under the assumption of normally distributed ν_{it} and α_i terms, the log likelihood function utilizing the log likelihood contribution for cross-sectional unit i, $\mathcal{L}_i(\beta) = \log f(y_i; x_i, \beta)$, can be expressed as

$$\begin{aligned}\log \mathcal{L}(\beta, \sigma_\varepsilon^2, \sigma_\alpha^2) &= \sum_{i=1}^{N} \mathcal{L}_i(\beta) = \sum_{i=1}^{N}\left[\sum_{t=1}^{T} \log f(y_{it}; x_{it}, \beta)\right], \\ &= -\frac{NT}{2}\log(2\pi) - \frac{N}{2}\log|\mathbf{\Omega}| - \frac{1}{2}\sum_{i=1}^{N}(y_i - x_i\beta)'\mathbf{\Omega}^{-1}(y_i - x_i\beta),\end{aligned}$$

where $\mathbf{\Omega} = E(\nu_i\nu_i') = \sigma_\varepsilon^2\mathbf{I}_T + \sigma_\alpha^2 e_T e_T'$ denote a $T \times T$ matrix, with $\mathbf{I}_T$ being an identity matrix of dimension T. Let V denote the full $NT \times NT$ variance-covariance matrix of compound error terms ν_{it}, that is,

$$V = \begin{pmatrix} \mathbf{\Omega} & 0 & \dots & 0 \\ 0 & \mathbf{\Omega} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{\Omega} \end{pmatrix} = \mathbf{I}_N \otimes \mathbf{\Omega}.$$

The ML estimator of the unknown parameter vector, $\hat{\beta}$, is obtained by solving the score functions

$$\arg\max_{\beta \in \mathbb{R}^K} \prod_{i=1}^{N} f(y_i; x_i, \beta) = \arg\min_{\beta \in \mathbb{R}^K} \sum_{i=1}^{N} r_i^2(\beta) = \arg\min_{\beta \in \mathbb{R}^K} \sum_{i=1}^{N} (y_i - x_i\beta)'\mathbf{\Omega}^{-1}(y_i - x_i\beta),$$

where $r(\beta) = y_{it} - x_{it}'\beta$ denote the error terms. In a similar way, for fixed effects model, the log-likelihood function can be constructed using the fixed-effects transformed version of the model and the variance of $\ddot{\varepsilon}_{it}$, $E(\ddot{\varepsilon}_{it}^2) = \sigma_\varepsilon^2(1 - 1/T)$ in obtaining joint probability density function for homoscedastic error terms $\ddot{\varepsilon}_{it}$ (see Reference 14).

In order to construct asymptotically consistent, weighted versions of the estimation equations, we next introduce some definitions and notations on weighted likelihood methodology.

Let $M_\beta = \{m_\beta(\cdot; \sigma_\nu); \sigma_\nu \in \mathbb{R}^+\}$ denote a parametric family of distributions for the theoretical error terms $r_i(\beta)$. We define $f^*(\cdot)$, the kernel density estimator based on the empirical distribution $\hat{F}_N$ of the observed values of the residuals $r_i(\hat{\beta})$, $i = 1, \dots, N$, and $m_\beta^*(\cdot;\cdot)$, the smoothed model density, for $r_i(\hat{\beta})$ as follows

$$\begin{aligned} f^*(r_i(\hat{\beta})) &= \int k(r_i(\hat{\beta}); t, h) d\hat{F}_N(t), \text{ and} \\ m_\beta^*(r_i(\hat{\beta}); \hat{\sigma}_\nu) &= \int k(r_i(\hat{\beta}); t, h) dM_\beta(t; \hat{\sigma}_\nu), \end{aligned}$$

where $M_\beta(\cdot; \sigma_\nu)$ is the distribution function for density $m_\beta(\cdot; \sigma_\nu)$ and $k(r; t, h)$ is a kernel density with bandwidth h. In this study, the normal kernel density with variance h^2, $k(r; t, h) = \frac{exp(-(r-t)^2/2h^2)}{\sqrt{2\pi}h}$, is used. Note that the bandwidth parameter h is chosen as $h = c\sigma_\nu$ where c is a constant term independent of the scale of the model so that outlying points will receive very small weights (cf. Reference 43). For the normal model, choosing the smoothing parameter based on the parameter c in determining the level of downweighting ensures that the weighted likelihood estimating equations become location and scale equivariant as noted in Reference 43. We then define the Pearson residuals as follows:

$$\delta(r_i(\hat{\beta})) = \frac{f^*(r_i(\hat{\beta}))}{m_\beta^*(r_i(\hat{\beta}); \hat{\sigma}_\nu)} - 1.$$

Based on the above, the weighted likelihood estimators of β and σ_v are obtained by solving the following estimating equations.

$$\sum_{i=1}^{N} \omega(r_i(\hat{\beta}); M_\beta, \hat{F}_N) s(r_i(\beta); \sigma_v) = 0, \tag{3}$$

$$\sum_{i=1}^{N} \omega(r_i(\hat{\beta}); M_\beta, \hat{F}_N) s_{\sigma_v}(r_i(\beta); \sigma_v) = 0, \tag{4}$$

where

$$s(r_i(\beta); \sigma_v) = \frac{\partial}{\partial \beta} \log f(y_{it}; x_{it}, \beta, \sigma_v) = \frac{\partial}{\partial \beta} \log m_\beta(r_i(\beta); \sigma_v), \text{ and}$$
$$s_{\sigma_v}(r_i(\beta); \sigma_v) = \frac{\partial}{\partial \sigma_v} \log f(y_{it}; x_{it}, \beta, \sigma_v) = \frac{\partial}{\partial \sigma_v} \log m_\beta(r_i(\beta); \sigma_v)$$

are the usual score functions and

$$\omega(r_i(\hat{\beta}); M_\beta, \hat{F}_N) = \omega_i = \min\left\{1, \frac{[A(\delta(r_i(\hat{\beta}))) + 1]^+}{\delta(r_i(\hat{\beta})) + 1}\right\},$$

where $[\ .\ ]^+$ and $A(.)$ denote the positive part of a function and the residual adjustment function (RAF) as described in Reference 40 (eg, Hellinger RAF $A(\delta) = 2[(\delta+1)^{1/2} - 1]$), respectively. When $A(\delta(r_i(\hat{\beta}))) = \delta(r_i(\hat{\beta}))$, the weights $\omega(r_i(\hat{\beta}); M_\beta, \hat{F}_N) = 1$, and this leads to produce ML estimates of the parameters (cf. References 44 and 45).

In weighted likelihood methodology, the usual score equations based on ML model are replaced by the weighted score equations to estimate model parameters. The weighted score equations defined above use the weights expressed as a function of Pearson residuals, $\delta(r_i(\hat{\beta})) = \frac{f^*(r_i(\hat{\beta}))}{m^*_\beta(r_i(\hat{\beta}); \hat{\sigma}_v)} - 1$. The weight function reflects the discordance between assumed model density and an estimate of true model density as noted in Reference 46. If the model is correctly specified in the absence outlying observations, then δ converges with probability 1 to 0 and thus, the weight function assigns a value close to 1. However, if the data involve outlying observations, large Pearson residuals are produced and the weight function assigns small weights to the outlying points depending on the level of discordance between the kernel density estimate of the model $f^*(\cdot)$ and the smoothed model density $m^*_\beta(\cdot;\cdot)$. Thus, the proposed estimators obtained using weighted likelihood estimating equations defined in Equations (3) and (4) will be robust in presence of outliers and/or contamination in the data due to use of weighted residuals.

An algorithm using resampling techniques has been proposed by Reference 43 to find the roots of the weighted likelihood estimating equations. They suggest to use of data-driven starting values to create a reasonable search region that includes all reasonable solutions having high probability in parameter space. To this end, the sub-samples with fixed dimension, which are sufficiently large for obtaining the ML estimates of parameters β, are drawn without replacement from the data. Then, the ML estimates of β, $\hat{\beta}^*_b$ for $b = 1, \ldots, B$ are obtained for each bootstrap sample. Finally, each of these estimates is used as an initial value in the iterative re-weighting algorithm for obtaining the roots of weighted likelihood estimating equations. $B = 30$ bootstrap sub-samples are created, and the maximum number of iterations are determined as 500 in our simulation studies (as in the default values of `R` package `wle`).

The ML method can be considered as a minimum distance (minimum disparity) method and growing attention has been paid to construct a parallel method of estimation which has the similar or same efficiency properties with the ML method until the late 1970s. Reference 47 has focused the robustness properties of density based minimum distance estimation methods and demonstrated asymptotic first-order efficiency of the estimator which minimizes the Hellinger distance between a kernel density estimator and a density from the model family within the continuous parametric models framework. The robustness of our proposed estimators is based on using the parallel minimum disparity measure in obtaining weight function for which downweight the outlying observations in the data. One of the main advantages related to the robustness properties in the minimum disparity estimation is that the presence of the valid objective function allows to investigate the breakdown point of the estimates as a measure of the robust global property. The breakdown properties of the estimators based on weighted likelihood estimating equations are examined by References 43 and 46

using the stability property of the estimating equations. The root selection method plays a very crucial role in determining the theoretical breakdown properties of the estimators when an estimating equation has multiple roots (cf. Reference 43). To achieve the robust global property, a root is chosen based on using minimum parallel disparity measure defined as follows

$$\rho_G(f^*, m^*_\beta) = \int G(\delta(x)) m^*_\beta(x) dx,$$

where G is a thrice differentiable convex function defined on $[-1, \infty)$ with $G(0) = 0$. Reference 40 has indicated that the choice of RAF may have a great impact on the robustness and efficiency of the corresponding estimators in the class of minimum disparity type methods. The function $G(\delta) = 2((\delta + 1)^{1/2} - 1)^2$ is the squared Hellinger distance in our proposed approach. Under differentiability and regularity conditions, $\hat{\beta}_i$ for $i = 1,2, \ldots, \ell$ is obtained as a root of the minimum disparity estimating equation

$$\int A(\delta(x)) \nabla m^*_\beta(x) dx = 0,$$

where ∇ denote the gradient with respect to β, $A(\delta) = G'(\delta)(1 + \delta) - G(\delta)$, G' representing the derivative of G, and $\delta(x) + 1 = f^*(x)/m^*_\beta(x)$. The parallel disparity measures obtained for each $\hat{\beta}_i$ where $i = 1,2, \ldots, \ell$, $\rho_G(f^*, m^*_{\hat{\beta}_1}(x)), \rho_G(f^*, m^*_{\hat{\beta}_2}(x)), \ldots, \rho_G(f^*, m^*_{\hat{\beta}_\ell}(x))$ are examined. Then, the proposed estimators based on weighted likelihood estimating equations can achieve the highest asymptotic breakdown point of 1/2 by selecting a root providing the minimum value of disparity measure as shown below.

$$\hat{\beta}_\omega = \underset{\hat{\beta}_i, i=1,2,\ldots,l}{\arg\min}\, \rho_G\left(f^*, m^*_{\hat{\beta}_i}(x)\right).$$

Let β_0, $\hat{\beta} = \underset{\beta \in \mathbb{R}^K}{\arg\min} \sum_{i=1}^N r_i^2(\beta)$ and $\hat{\beta}_\omega = \underset{\beta \in \mathbb{R}^K}{\arg\min} \sum_{i=1}^N \omega_i r_i^2(\beta)$ denote the true value of the parameters, the ML (or GLS) and weighted likelihood estimators of the parameters, respectively. For the linear panel data model with fixed effects and random effects, the conditions required for the existence of solutions and asymptotic normality of the proposed estimators are as follows (cf. References 43 and 48):

A1. The weight function $\omega(\delta)$ is a nonnegative, bounded and differentiable function with respect to δ.
A2. The weight function $\omega(\delta)$ is regular with bounded $\omega'(\delta)(\delta + 1)$, where prime denotes the derivative.

Let $\tilde{s}(x; \beta) = \nabla m^*_\beta(x)/m^*_\beta(x)$ and $s(x; \beta) = \nabla m_\beta(x)/m_\beta(x)$ where $m^*_\beta(x)$ and $m_\beta(x)$ denote the smoothed model and true model, respectively.

A3. For every $\beta_0 \in \Theta$, there is a neighborhood $N(\beta_0)$ such that for $\beta \in N(\beta_0)$, $M_i(x)$ for $i = 1,2,3,4$, where $E_{\beta_0}[M_i(X)] < \infty$, are the bounds for the quantities $|\tilde{s}(x; \beta)s'(x; \beta)|$, $|\tilde{s}^2(x; \beta)s(x; \beta)|$, $|\tilde{s}'(x; \beta)s(x; \beta)|$, and $|s''(x; \beta)|$.
A4. $E_{\beta_0}[\tilde{s}^2(x; \beta)s^2(x; \beta)] < \infty$.
A5. The Fisher information is finite; $I(\beta) = E_\beta[s^2(x; \beta)] < \infty$.
A6. (i) $\int |\nabla m_\beta(x)/m^*_\beta(x)| dx = \int |m_\beta(x)s(x; \beta)/m^*_\beta(x)| dx < \infty$.
(ii) $\int |\tilde{s}(x; \beta)s(x; \beta)| \left[\frac{m_\beta(x)}{m^*_\beta(x)}\right] dx < \infty$.
(iii) $\int |s'(x; \beta)| \left[\frac{m_\beta(x)}{m^*_\beta(x)}\right] dx < \infty$.
A7. The kernel density function $k(X; t, h)$ is bounded for all x by a finite constant $M(h)$ that may depend on the smoothing parameter h but not on t or x.

Under the assumption that the model is correctly specified and the conditions A1-A7 given above hold, the asymptotic equivalence of $\hat{\beta}_\omega$ and $\hat{\beta}$ are given as follows:

$$\sqrt{N}(\hat{\beta}_\omega - \hat{\beta}) = o_p(1) \quad as \quad N \to \infty. \tag{5}$$

To prove Equation (5), we first present the following equations from Reference 49, which are needed for completeness.

$$\sqrt{N}\left|\frac{1}{N}\sum_{i=1}^{N}\omega(\delta(r_i))s(x_i,y_i;\beta_0)-\frac{1}{N}\sum_{i=1}^{N}s(x_i,y_i;\beta_0)\right|\xrightarrow{p}0 \quad as \quad N\to\infty \tag{6}$$

$$\left|\frac{1}{N}\sum_{i=1}^{N}\nabla_\beta\{\omega(\delta(r_i))s(x_i,y_i;\beta)\}|_{\beta=\beta_0}-\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\beta}s(x_i,y_i;\beta_0)\right|\xrightarrow{p}0 \quad as \quad N\to\infty, \tag{7}$$

and

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2}{\partial\beta^2}(\omega(\delta(r_i))s(x_i,y_i;\beta))|_{\beta=\beta^*}=\mathcal{O}_p(1), \tag{8}$$

where β^* is the initial value between true value β_0 and $\hat{\beta}_\omega$. Then, using the fact that $\sup_i|\hat{\omega}_i-1|\xrightarrow{p}0$ (cf. Reference 44) and Equations (6)-(8) and observing that $\int\omega(\delta(r))(\nabla_\beta s(x,y;\beta))dF(x,y)$ is a continuous function in β, the proof of Equation (5) follows from theorem 1 of Reference 44.

3 | NUMERICAL RESULTS

In this section, we present results from an extensive simulation study to assess the finite sample properties of the proposed and conventional estimators. The robustness performances of the proposed procedures are examined via three different scenarios: (a) different sample sizes, (b) different error distributions, and (c) various types of outliers. 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.)

The following static linear panel data model is considered for the data generation processes (DGP)

$$y_{it}=x_{it}'\beta+\alpha_i+\varepsilon_{it},\quad i=1,\ldots,N,\quad t=1,\ldots,T,$$

where ε_{it}s are assumed to be iid N($\mu=0,\sigma^2=1$). The vector of regression coefficients is chosen as $\beta'=(\beta_1,\beta_2)=(2.4,-1.2)$. For both the fixed effects model (DGP-I) and random effects model (DGP-II) specifications, the explanatory variables x_{itK} for $K=1,2$ are generated as follows

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

where χ_2^2 represents the chi-squared distribution with 2 degrees of freedom. One of the variables is generated from an asymmetric distribution to avoid completely symmetric experimental design as in Reference 19.

In random effects model specification, α_is are assumed to be iid N($\mu=0,\sigma^2=1$). For the fixed effects model where the experimental design is similar to those of Reference 19, the individual-specific effects are generated depending on the predictors x_{it} through $\gamma=(2,4)'$ and on $\eta_i\sim U(0,12)$ to ensure homogeneity of the variances of deterministic γ and random parts η_i of α_i as follows.

$$\alpha_i=\sum_{t=1}^{T}x_{it}'\gamma/\sqrt{T}+\eta_i.$$

Throughout the experiments, $S=1000$ simulations are performed to estimate the model coefficients and calculate the performance metrics. To evaluate the performance of the methods previously described, we calculate the squared errors (SE): $SE=\|\hat{\beta}^s-\beta\|^2$ and mean squared errors (MSE): $MSE=\frac{1}{S}\sum_{s=1}^{S}\|\hat{\beta}^s-\beta\|^2$, where $\hat{\beta}^s$, $s=1,\ldots,S$, denote the estimates obtained from S simulated samples. The bandwidth of the kernel $k(\cdot)$ in Equation (3) is chosen using the `wle.smooth` function in the R package `wle`.

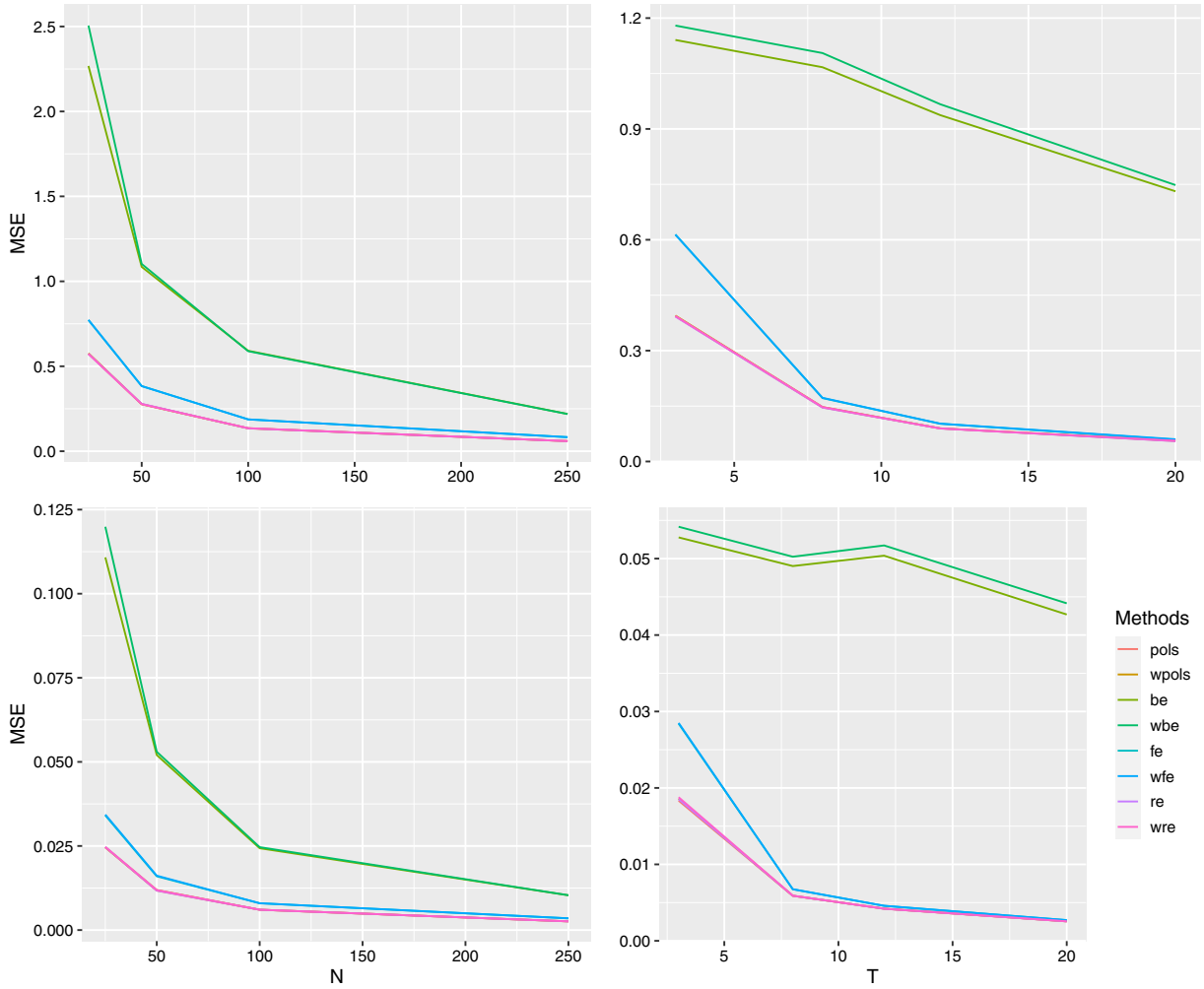

FIGURE 1 Calculated MSEs of all estimators under N($\mu = 0, \sigma^2 = 1$) errors. The data were generated under DGP-I (first row) and DGP-II (second row) when $N = 25, 50, 100, 250$ and $T = 4$ (first column) and $T = 3, 8, 12, 20$ and $N = 50$ [Colour figure can be viewed at wileyonlinelibrary.com]

3.1 | Sample sizes

Different values of cross-sectional dimension N and time dimension T are considered to investigate the effect of panel sizes on the performances of our proposed estimators. In particular, we consider $N = 25, 50, 100, 250$ for fixed time period $T = 4$, and $T = 3, 8, 12, 20$ for fixed cross-sectional dimension $N = 50$. We compare the MSE values of the estimators under standard normal errors, $\varepsilon_{it} \sim$ N($\mu = 0, \sigma^2 = 1$). The simulation results are presented in Figure 1. In this figure, the lines representing the performances of the pooled OLS and weighted pooled OLS estimators do not appear since they overlap with the lines of random effects and weighted random effects estimators, respectively. Also, the fixed effects and random effects estimators have almost same performances with their weighted versions. In general, the proposed weighted likelihood based estimators perform similarly with their traditional counterparts for both DGPs. These results also confirm that the proposed methods are consistent with the original least squares based estimators when N and/or T goes to infinity.

3.2 | Error distributions

Three different error distributions, namely, N($\mu = 0, \sigma^2 = 1$), Student's t distribution with 5 degrees of freedom (t_5), and double exponential distribution with rate 1 ($DExp(1)$) are considered to evaluate the influence of error

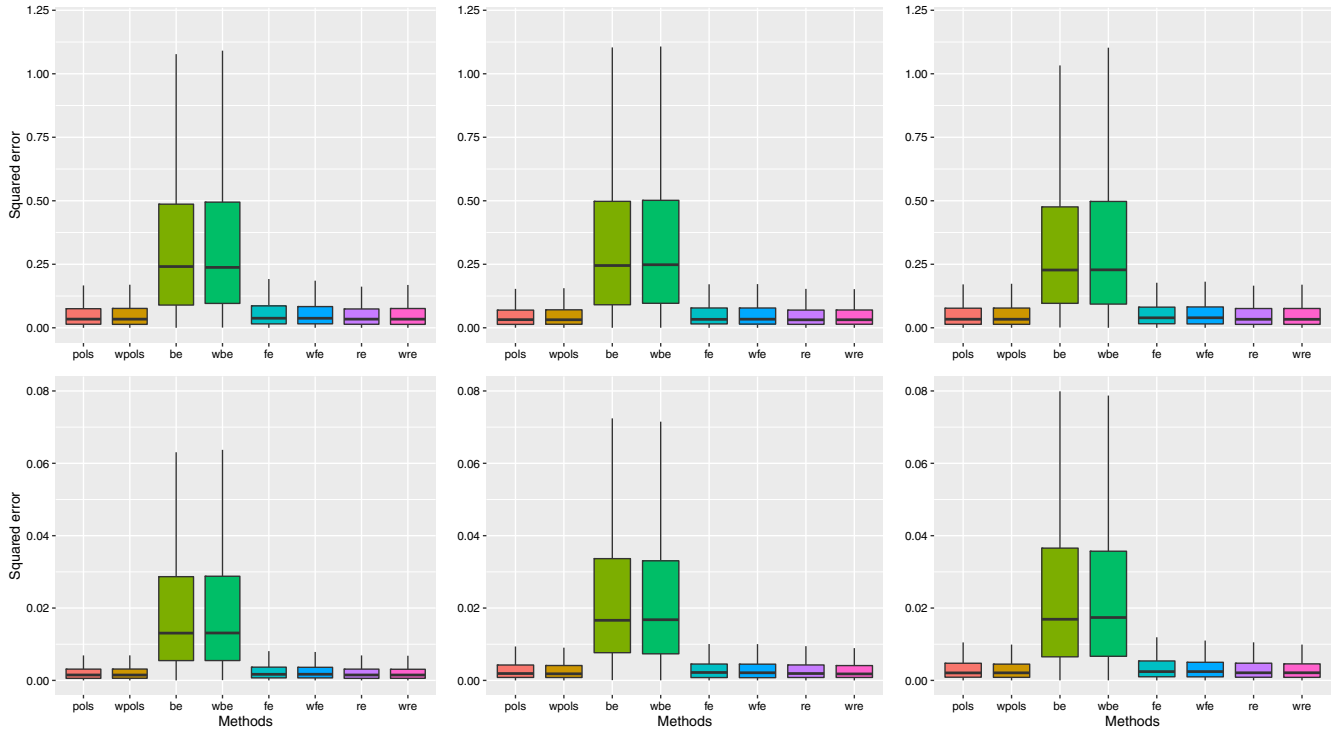

FIGURE 2 Calculated squared errors of all estimators: pooled OLS (pols), pooled weighted likelihood (wpols), between (be), weighted between (wbe), fixed-effects (fe), weighted fixed-effects (wfe), random effects (re), and weighted random effects (wre) under different error distributions: N($\mu = 0, \sigma^2 = 1$) (first column), t_5 (second column), and $DExp(1)$ (third column). The data were generated under DGP-I (first row) and DGP-II (second row) when $N = 100$, $T = 10$ [Colour figure can be viewed at wileyonlinelibrary.com]

distributions on the estimation methods. The SE values of the estimators are calculated for three pairs of values of cross-sectional sizes and time periods: $(N,T) = (250, 5)$,$(100, 10)$, and $(30, 20)$. Since our conclusions do not vary significantly with different choices of panel sizes, therefore to save space, we report only the results for $N = 100$ and $T = 10$. Figure 2 illustrates the results indicating that the estimators have similar performances under different error distributions.

3.3 | Outliers

In this section, the finite sample properties of the estimation procedures are investigated in the presence of different types of outliers. Throughout the simulations, the panel size is chosen to be 240 with two levels of cross-sectional sizes and time periods, namely, $N_1 = 120$, $T_1 = 2$ and $N_2 = 80$, $T_2 = 3$. Two different levels of contamination (5% and 10%) are considered by setting the number of outliers as $m = 12$ and $m = 24$. The contaminated data is generated by two different ways as in References 18 and 19: (a) outliers are randomly allocated over all observations, and (b) half of the observations within individual units are contaminated such that outlying observations are concentrated in some blocks. Based on the above framework, the following contamination schemes are considered.

1. Random vertical outliers (y_{it}^r) are obtained by replacing randomly selected original values of the response variable with the points coming from Uniform distribution $y_{it}^r \sim U(20, 80)$.
2. To generate the random leverage points, first the randomly selected values of the response variable are contaminated by using the same rule as in first scheme. Then, the values of the explanatory variables corresponding to the contaminated values of the response variable are generated from a Normal distribution N($\mu = 8, \sigma^2 = 4$).
3. Concentrated vertical outliers (y_{it}^c) are inserted into the randomly selected blocks of the original values of response variable by replacing them by the random values from Uniform distribution $y_{it}^c \sim U(79, 80)$.

4. To generate the concentrated leverage points, the randomly selected blocks of the original values of response variable are contaminated following the same rule as in the third scheme. Then, the blocks including the original values of the explanatory variables corresponding to the contaminated blocks of the response variable are replaced by the observations from a normal distribution $N(\mu = 8, \sigma^2 = 4)$ as in second scheme.

Note that the proportion of contaminated values per cross-sectional unit constitute at least a half of observations over time periods for concentrated vertical outliers and concentrated leverage points. For the representative plots of the contamination schemes mentioned above, please see figure 1 of Reference 18. The simulation results are given in Table 1. Our results clearly demonstrate that the proposed estimators outperform the conventional least-square based estimators in all situations. Note that the performances of the conventional estimators can be severely degraded in the presence of outliers depending on the types and levels of contaminations. On the other hand, the performances of the proposed estimators are quite less sensitive to the choice of contamination level and/or scheme compared to traditional ones. We further see that, compared to the presence of vertical outliers, the traditional estimators produce more biased and less efficient results when the data is contaminated by the leverage points, in general. The MSE values calculated for traditional estimators significantly increase with increasing level of contamination in the presence of vertical outliers under both DGP-I and DGP-II. It can further be seen that the MSE values of the traditional procedures are mostly affected in case of the concentrated outliers compared to the other schemes. Moreover, there are substantial differences among the proposed weighted estimators. For example, the weighted fixed-effects and between estimators ($\hat{\beta}_{wfe}$ and $\hat{\beta}_{wbe}$) are sensitive to the concentrated leverage points at %10 contamination, while the weighted random-effects estimator ($\hat{\beta}_{wre}$) is not severely affected by the different choices of contamination schemes and levels under both DGPs. Also, $\hat{\beta}_{wre}$ and $\hat{\beta}_{wpols}$ have the similar performances in the presence of outliers regardless of the types and levels of contaminations. All the proposed estimators ($\hat{\beta}_{wpols}$, $\hat{\beta}_{wbe}$, $\hat{\beta}_{wfe}$, and $\hat{\beta}_{wre}$) exhibit improved performances over the conventional counterparts, especially for DGP-II.

To justify the superiority of the proposed methods further, we compare the power of significance tests of the regression coefficients. For the comparisons, the significance level γ is set to 0.05 to calculate the power of significance testing of individual coefficients, which is defined as follows.

$$P_K = \frac{1}{S}\sum_{s=1}^{S} \mathbf{1}\left\{ Q(\gamma/2) \leq \frac{\hat{\beta}_K^s}{\hat{s}_K^s} \leq Q(1-\gamma/2) \right\},$$

where $\mathbf{1}(\cdot)$ and $Q(\cdot)$ denote the indicator function and the quantiles of standard normal distribution, respectively, and $\hat{s}_K$ represents the standard error of the estimate for $K = 1, 2$.

Table 2 reports the simulated powers of the significance tests for the individual coefficients $\beta_1 = 2.4$ and $\beta_2 = -1.2$, respectively. The results demonstrate that the powers of the significance tests of β_1 are similar for both proposed and traditional estimators when the data are contaminated by random vertical outliers at 5% contamination. On the other hand, the significance tests based on the proposed estimators have significantly better power values than those of classical methods for β_2. The proposed methods are less affected by increasing number of outliers and yield a large gain in power of significance testing in almost all cases.

4 | CASE STUDY

In this section, we study the performances of the proposed and traditional OLS based estimators with a case study, the blood pressure data set. The dataset, which is consisted of a total of 2400 observations ($N = 200$, $T = 12$) with three variables: pulse rate, systolic pressure, and diastolic pressure, are collected from male and female patients (during hospitalization) by ambulatory blood pressure monitors. Let x_{it1} = systolic pressure, x_{it2} = diastolic pressure, and y_{it} = pulse rate, we conduct a linear panel data regression model: $y_{it} = x_{it}'\beta + \alpha_i + \varepsilon_{it}$ where $\beta' = (\beta_1, \beta_2)$, $i = 1, \ldots, 200$, and $t = 1, \ldots, 12$. The scatterplots of the response variable against the explanatory variables are presented in Figure 3. It is evident form the scatterplots that both datasets include outlying observations and the number of outliers in blood pressure data for females seems larger than those of males. The estimates of individual coefficients and standard errors of the estimates obtained for this dataset are shown in Table 3. For both datasets gathered from 200 male and 200 female patients, our proposed weighted procedures yield slightly more efficient estimates than the OLS method, in general. The proposed $\hat{\beta}_{wbe}$ have significantly better performances for estimating parameters compared to its traditional version.

TABLE 1 Calculated MSEs of all estimators for $N_1 = 120$, $T_1 = 2$ and $N_2 = 80$, $T_2 = 3$ in the presence of 5% and 10% random and concentrated contamination by setting the number of outliers as $m = 12$ and $m = 24$

	Random contamination				**Concentrated contamination**			
DGP-I	**Vertical outliers**		**Leverage points**		**Vertical outliers**		**Leverage points**	
$(N_1, T_1) = (120, 2)$	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**
$\hat{\beta}_{pols}$	0.786	1.302	0.764	1.384	0.941	1.606	1.583	2.878
$\hat{\beta}_{wpols}$	0.274	0.316	0.265	0.310	0.242	0.239	0.255	0.270
$\hat{\beta}_{be}$	1.617	2.557	1.602	2.716	1.941	3.251	3.139	5.966
$\hat{\beta}_{wbe}$	0.674	1.472	0.698	1.396	0.527	0.511	0.521	3.184
$\hat{\beta}_{fe}$	1.593	2.670	1.535	2.768	1.848	3.234	3.333	5.589
$\hat{\beta}_{wfe}$	0.717	1.732	0.702	1.710	0.515	0.534	0.588	4.438
$\hat{\beta}_{re}$	0.789	1.307	0.770	1.390	0.939	1.615	1.590	2.894
$\hat{\beta}_{wre}$	0.277	0.317	0.266	0.313	0.245	0.249	0.259	0.272
$(N_2, T_2) = (80, 3)$								
$\hat{\beta}_{pols}$	0.771	1.404	0.827	1.304	0.729	1.196	1.542	2.755
$\hat{\beta}_{wpols}$	0.276	0.309	0.254	0.310	0.248	0.245	0.255	0.258
$\hat{\beta}_{be}$	2.472	4.144	2.376	3.947	2.400	3.844	4.933	8.647
$\hat{\beta}_{wbe}$	1.409	3.548	1.401	3.364	0.797	0.873	1.105	7.580
$\hat{\beta}_{fe}$	1.219	2.016	1.233	2.014	1.070	1.627	2.385	4.387
$\hat{\beta}_{wfe}$	0.614	1.193	0.575	1.172	0.419	0.466	0.622	1.859
$\hat{\beta}_{re}$	0.778	1.407	0.824	1.310	0.725	1.190	1.554	2.767
$\hat{\beta}_{wre}$	0.276	0.311	0.254	0.313	0.251	0.250	0.260	0.261
	Random contamination				**Concentrated contamination**			
DGP-II	**Vertical outliers**		**Leverage points**		**Vertical outliers**		**Leverage points**	
$(N_1, T_1) = (120, 2)$	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**
$\hat{\beta}_{pols}$	0.705	1.401	0.765	1.334	0.795	1.574	1.588	2.975
$\hat{\beta}_{wpols}$	0.011	0.011	0.011	0.012	0.010	0.011	0.011	0.011
$\hat{\beta}_{be}$	1.550	2.919	1.541	2.830	1.675	3.378	3.280	6.452
$\hat{\beta}_{wbe}$	0.025	0.083	0.025	0.099	0.022	0.024	0.025	0.711
$\hat{\beta}_{fe}$	1.313	2.921	1.530	2.528	1.684	3.122	3.237	5.803
$\hat{\beta}_{wfe}$	0.024	0.112	0.023	0.109	0.020	0.025	0.025	1.515
$\hat{\beta}_{re}$	0.706	1.406	0.765	1.325	0.784	1.571	1.603	2.970
$\hat{\beta}_{wre}$	0.011	0.013	0.012	0.013	0.010	0.013	0.012	0.013
$(N_2, T_2) = (80, 3)$								
$\hat{\beta}_{pols}$	0.692	1.293	0.736	1.486	0.559	1.054	1.573	3.115
$\hat{\beta}_{wpols}$	0.012	0.011	0.012	0.011	0.010	0.011	0.012	0.012
$\hat{\beta}_{be}$	2.270	4.117	2.233	4.315	1.803	3.855	5.020	9.655
$\hat{\beta}_{wbe}$	0.042	2.898	0.040	3.011	0.035	0.036	0.040	7.867
$\hat{\beta}_{fe}$	1.042	1.985	1.128	2.206	0.869	1.450	2.361	4.700
$\hat{\beta}_{wfe}$	0.018	0.381	0.020	0.386	0.017	0.017	0.019	0.726
$\hat{\beta}_{re}$	0.692	1.289	0.741	1.488	0.554	1.042	1.580	3.137
$\hat{\beta}_{wre}$	0.012	0.011	0.012	0.012	0.011	0.012	0.013	0.015

TABLE 2 Calculated powers of all estimators when $N_1 = 120$, $T_1 = 2$ and $N_2 = 80$, $T_2 = 3$

DGP-I	Random contamination				Concentrated contamination				DGP-II	Random contamination				Concentrated contamination			
(N_1, T_1)	Vertical outliers		Leverage points		Vertical outliers		Leverage points		(N_1, T_1)	Vertical outliers		Leverage points		Vertical outliers		Leverage points	
(120, 2)	5%	10%	5%	10%	5%	10%	5%	10%	(120, 2)	5%	10%	5%	10%	5%	10%	5%	10%
$\hat{\beta}_{1,pols}$	1.000	0.989	1.000	0.984	1.000	0.988	0.984	0.804	$\hat{\beta}_{1,pols}$	1.000	0.993	1.000	0.990	1.000	0.987	0.989	0.784
$\hat{\beta}_{1,wpols}$	1.000	0.998	1.000	0.999	1.000	1.000	1.000	1.000	$\hat{\beta}_{1,wpols}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\hat{\beta}_{1,be}$	0.978	0.843	0.978	0.855	0.961	0.791	0.792	0.512	$\hat{\beta}_{1,be}$	0.985	0.810	0.987	0.822	0.976	0.764	0.802	0.463
$\hat{\beta}_{1,wbe}$	1.000	0.993	0.999	0.996	1.000	0.999	1.000	0.943	$\hat{\beta}_{1,wbe}$	1.000	0.999	1.000	0.998	1.000	1.000	1.000	0.995
$\hat{\beta}_{1,fe}$	0.970	0.844	0.975	0.835	0.966	0.823	0.797	0.527	$\hat{\beta}_{1,fe}$	0.979	0.831	0.976	0.852	0.972	0.852	0.798	0.465
$\hat{\beta}_{1,wfe}$	1.000	0.976	0.998	0.975	1.000	0.999	1.000	0.835	$\hat{\beta}_{1,wfe}$	1.000	0.995	1.000	0.995	1.000	1.000	1.000	0.965
$\hat{\beta}_{1,re}$	1.000	0.980	0.999	0.977	1.000	0.987	0.984	0.779	$\hat{\beta}_{1,re}$	1.000	0.992	1.000	0.987	1.000	0.988	0.990	0.783
$\hat{\beta}_{1,wre}$	1.000	0.991	0.999	0.993	1.000	0.999	1.000	0.958	$\hat{\beta}_{1,wre}$	1.000	0.999	1.000	0.997	1.000	1.000	1.000	0.986
$\hat{\beta}_{2,pols}$	0.293	0.179	0.292	0.191	0.284	0.173	0.162	0.099	$\hat{\beta}_{2,pols}$	0.342	0.197	0.325	0.171	0.331	0.176	0.188	0.096
$\hat{\beta}_{2,wpols}$	0.790	0.794	0.780	0.795	0.828	0.826	0.818	0.813	$\hat{\beta}_{2,wpols}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\hat{\beta}_{2,be}$	0.180	0.105	0.182	0.129	0.179	0.110	0.114	0.089	$\hat{\beta}_{2,be}$	0.211	0.118	0.203	0.125	0.176	0.115	0.127	0.094
$\hat{\beta}_{2,wbe}$	0.701	0.593	0.688	0.610	0.754	0.767	0.765	0.691	$\hat{\beta}_{2,wbe}$	1.000	0.974	1.000	0.970	1.000	1.000	1.000	0.959
$\hat{\beta}_{2,fe}$	0.185	0.129	0.171	0.125	0.142	0.111	0.102	0.073	$\hat{\beta}_{2,fe}$	0.185	0.131	0.197	0.105	0.201	0.119	0.106	0.069
$\hat{\beta}_{2,wfe}$	0.484	0.421	0.458	0.368	0.494	0.540	0.561	0.419	$\hat{\beta}_{2,wfe}$	1.000	0.959	1.000	0.946	1.000	1.000	1.000	0.869
$\hat{\beta}_{2,re}$	0.295	0.187	0.290	0.184	0.287	0.176	0.157	0.100	$\hat{\beta}_{2,re}$	0.343	0.197	0.320	0.169	0.337	0.175	0.183	0.097
$\hat{\beta}_{2,wre}$	0.790	0.792	0.771	0.782	0.824	0.819	0.807	0.764	$\hat{\beta}_{2,wre}$	1.000	0.999	1.000	0.997	1.000	1.000	1.000	0.986
DGP-I	**Random contamination**				**Concentrated contamination**				**DGP-II**	**Random contamination**				**Concentrated contamination**			
(N_2, T_2)	Vertical outliers		Leverage points		Vertical outliers		Leverage points		(N_2, T_2)	Vertical outliers		Leverage points		Vertical outliers		Leverage points	
(80, 3)	5%	10%	5%	10%	5%	10%	5%	10%	(80, 3)	5%	10%	5%	10%	5%	10%	5%	10%
$\hat{\beta}_{1,pols}$	1.000	0.989	1.000	0.992	1.000	0.999	0.985	0.840	$\hat{\beta}_{1,pols}$	1.000	0.996	1.000	0.992	1.000	1.000	0.983	0.789
$\hat{\beta}_{1,wpols}$	1.000	1.000	1.000	1.000	1.000	0.999	1.000	0.999	$\hat{\beta}_{1,wpols}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\hat{\beta}_{1,be}$	0.883	0.652	0.897	0.696	0.927	0.749	0.640	0.385	$\hat{\beta}_{1,be}$	0.908	0.676	0.911	0.638	0.973	0.753	0.599	0.369
$\hat{\beta}_{1,wbe}$	0.998	0.957	0.997	0.962	0.999	0.999	0.998	0.816	$\hat{\beta}_{1,wbe}$	1.000	0.955	1.000	0.961	1.000	1.000	0.999	0.786
$\hat{\beta}_{1,fe}$	0.993	0.935	0.991	0.935	0.997	0.978	0.914	0.638	$\hat{\beta}_{1,fe}$	0.998	0.897	0.995	0.910	1.000	0.990	0.920	0.581
$\hat{\beta}_{1,wfe}$	0.999	0.993	0.997	0.995	0.999	0.998	1.000	0.946	$\hat{\beta}_{1,wfe}$	1.000	0.983	1.000	0.985	1.000	1.000	0.999	0.916

(Continues)

TABLE 2 (Continued)

DGP-I	Random contamination				Concentrated contamination				DGP-II	Random contamination				Concentrated contamination			
(N_2, T_2)	Vertical outliers		Leverage points		Vertical outliers		Leverage points		(N_2, T_2)	Vertical outliers		Leverage points		Vertical outliers		Leverage points	
(80, 3)	5%	10%	5%	10%	5%	10%	5%	10%	(80, 3)	5%	10%	5%	10%	5%	10%	5%	10%
$\hat{\beta}_{1,re}$	0.999	0.984	0.997	0.988	0.999	0.997	0.987	0.812	$\hat{\beta}_{1,re}$	1.000	0.979	1.000	0.978	1.000	1.000	0.982	0.724
$\hat{\beta}_{1,wre}$	0.999	0.996	0.997	0.996	0.999	0.998	1.000	0.964	$\hat{\beta}_{1,wre}$	1.000	0.983	1.000	0.986	1.000	1.000	0.999	0.917
$\hat{\beta}_{2,pols}$	0.302	0.179	0.293	0.193	0.365	0.232	0.157	0.110	$\hat{\beta}_{2,pols}$	0.347	0.176	0.349	0.186	0.426	0.248	0.185	0.116
$\hat{\beta}_{2,wpols}$	0.806	0.793	0.819	0.793	0.804	0.825	0.814	0.829	$\hat{\beta}_{2,wpols}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\hat{\beta}_{2,be}$	0.155	0.092	0.106	0.094	0.163	0.105	0.102	0.078	$\hat{\beta}_{2,be}$	0.139	0.083	0.135	0.097	0.177	0.101	0.090	0.072
$\hat{\beta}_{2,wbe}$	0.640	0.487	0.623	0.476	0.682	0.729	0.710	0.408	$\hat{\beta}_{2,wbe}$	1.000	0.522	1.000	0.544	1.000	1.000	0.999	0.481
$\hat{\beta}_{2,fe}$	0.214	0.147	0.219	0.149	0.295	0.188	0.129	0.093	$\hat{\beta}_{2,fe}$	0.242	0.123	0.254	0.155	0.303	0.184	0.138	0.096
$\hat{\beta}_{2,wfe}$	0.481	0.293	0.492	0.313	0.625	0.620	0.612	0.202	$\hat{\beta}_{2,wfe}$	1.000	0.569	1.000	0.601	1.000	1.000	0.999	0.500
$\hat{\beta}_{2,re}$	0.307	0.187	0.290	0.189	0.363	0.231	0.157	0.103	$\hat{\beta}_{2,re}$	0.345	0.171	0.342	0.179	0.430	0.256	0.175	0.106
$\hat{\beta}_{2,wre}$	0.806	0.783	0.813	0.786	0.801	0.819	0.804	0.791	$\hat{\beta}_{2,wre}$	1.000	0.983	1.000	0.986	1.000	1.000	0.999	0.917

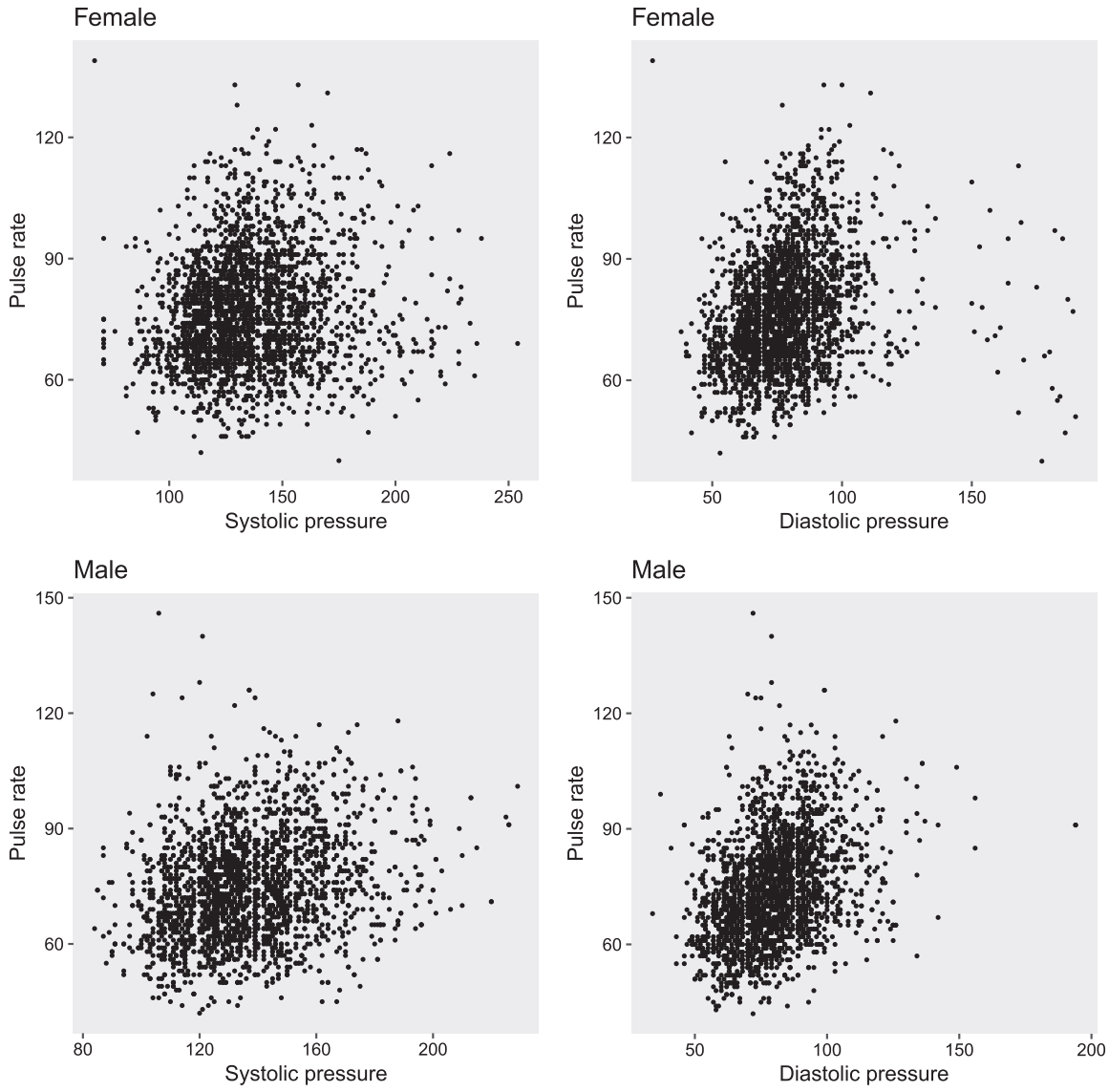

FIGURE 3 Scatter plots of the pulse rates against the systolic and diastolic blood pressures for female and male patients

Moreover, we compared the predictive performances of the OLS and proposed methods. In doing so, the datasets are divided into the following two parts; the model is constructed based on the randomly selected 150 male and female patients, and the pulse rates of the remaining 50 patients are predicted using the estimated model parameters. This process is repeated 100 times, and for each time, the MSE for the predicted and observed pulse rates, $MSE = \frac{1}{50\times12}\sum_{i=1}^{50}\sum_{t=1}^{12}(y_{it}-\hat{y}_{it})^2$, are computed. The results are presented in Figure 4. This figure shows that both methods have similar MSE values. This is due to fact that the number of outliers is relatively small for the datasets with sample size $N \times T = 2400$.

5 | CONCLUSIONS

In this article, we propose asymptotically valid, robust estimation procedures to obtain parameter estimates in linear panel data models with fixed and random effects. The proposed approaches are based on using weighted likelihood methodology. The finite sample performances of the proposed methods are examined through extensive simulation studies and a real-world example, and the results are compared with existing methods. Our records show that the proposed procedures have similar performance with existing estimation methods when the data have no outliers and/or under different error distributions. However, our proposed method produces more accurate and efficient parameter

TABLE 3 Estimates of individual coefficients (upper rows) and standard errors of the estimates (lower rows) for blood pressure data of female and male patients

Female patients				Male patients			
$\hat{\beta}_{1,pols}$	-0.109 (0.015)	$\hat{\beta}_{2,pols}$	0.414 (0.024)	$\hat{\beta}_{1,pols}$	-0.030 (0.017)	$\hat{\beta}_{2,pols}$	0.335 (0.024)
$\hat{\beta}_{1,wpols}$	-0.114 (0.014)	$\hat{\beta}_{2,wpols}$	0.423 (0.023)	$\hat{\beta}_{1,wpols}$	-0.022 (0.016)	$\hat{\beta}_{2,wpols}$	0.334 (0.022)
$\hat{\beta}_{1,be}$	-0.250 (0.052)	$\hat{\beta}_{2,be}$	0.583 (0.092)	$\hat{\beta}_{1,be}$	-0.118 (0.061)	$\hat{\beta}_{2,be}$	0.352 (0.086)
$\hat{\beta}_{1,wbe}$	-0.255 (0.014)	$\hat{\beta}_{2,wbe}$	0.589 (0.025)	$\hat{\beta}_{1,wbe}$	-0.116 (0.016)	$\hat{\beta}_{2,wbe}$	0.346 (0.023)
$\hat{\beta}_{1,fe}$	0.101 (0.015)	$\hat{\beta}_{2,fe}$	0.214 (0.022)	$\hat{\beta}_{1,fe}$	0.076 (0.017)	$\hat{\beta}_{2,fe}$	0.301 (0.023)
$\hat{\beta}_{1,wfe}$	0.098 (0.013)	$\hat{\beta}_{2,wfe}$	0.216 (0.019)	$\hat{\beta}_{1,wfe}$	0.081 (0.016)	$\hat{\beta}_{2,wfe}$	0.299 (0.021)
$\hat{\beta}_{1,re}$	0.070 (0.015)	$\hat{\beta}_{2,re}$	0.241 (0.021)	$\hat{\beta}_{1,re}$	0.061 (0.017)	$\hat{\beta}_{2,re}$	0.307 (0.023)
$\hat{\beta}_{1,wre}$	0.065 (0.013)	$\hat{\beta}_{2,wre}$	0.246 (0.018)	$\hat{\beta}_{1,wre}$	0.066 (0.015)	$\hat{\beta}_{2,wre}$	0.307 (0.020)

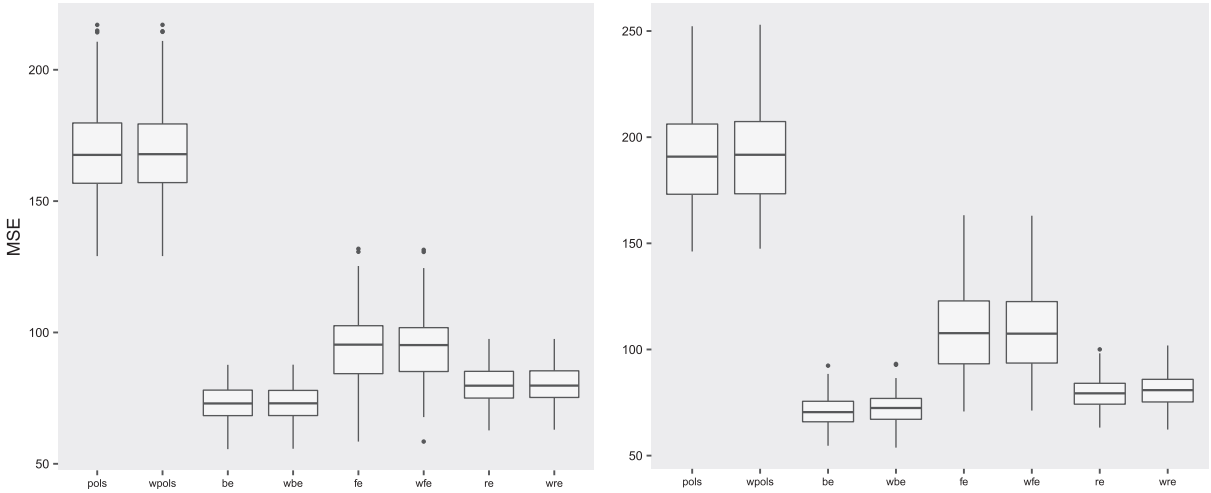

FIGURE 4 Calculated MSE values for the blood pressure dataset: male (first column) and female (second column). Methods: OLS based pooled regression (pols), between regression (be), fixed effects model (fe), random effects model (re), and their weighted likelihood counterparts; wpols, wbe, wfe, and wre

estimates with better power compared to the traditional OLS methods when outliers are presented in the data. As a part of our future research, we will investigate whether the proposed methods can also be used to estimate the parameters in linear dynamic panel data models as an alternative to the generalized method of moments.

ORCID

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

REFERENCES

1. Baltagi BH. *Econometric Analysis of Panel Data*. Chichester, UK: John Wiley and Sons; 2005.
2. Hsiao C. Benefits and limitations of panel data. *Econ Rev*. 1985;4(1):121-174.
3. Balestra P, Nerlove M. Pooling cross-section and time series data in the estimation of a dynamic model: the demand for natural gas. *Econometrica*. 1966;34(3):585-612.
4. Hsiao C. Panel data analysis-advantages and challenges. *TEST*. 2007;16(1):1-22.
5. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. New York, NY: John Wiley and Sons; 2004.
6. Greene WH. *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall; 2003.
7. Gardiner JC, Luo Z, Roman LA. Fixed effects, random effects and gee: what are the differences? *Stat Med*. 2009;28(2):221-239.
8. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics*. 1982;38(4):963-974.
9. Maddala GS, Mount TD. A comparative study of alternative estimators for variance components models used in econometric applications. *J Am Stat 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. Diggle PJ, Liang K-Y, Heagerty P, Zeger SL. *Analysis of Longitudinal Data*. Oxford, UK: Oxford University Press; 2002.
12. Hill RC, Griffiths WE, Lim GC. *Principles of Econometrics*. Hoboken, NJ: John Wiley and Sons; 2007.
13. Wallace TD, Hussain A. The use of error components models in combining cross section and time-series data. *Econometrica*. 1969;37(1):55-72.
14. Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press; 2002.
15. Bălă RM, Prada EM. Migration and private consumption in Europe: a panel data analysis. *Proc Econ Financ*. 2014;10:141-149.
16. Kennedy P. *A Guide to Econometrics*. Cambridge, MA: The MIT Press; 2003.
17. Zhang L. The use of panel data models in higher education policy studies. *Higher Education: Handbook of Theory and Research*. Dordrecht, Netherlands: Springer; 2010.
18. Bramati MC, Croux CP. Robust estimators for the fixed effects panel data model. *Econ J*. 2007;10(3):521-540.
19. Aquaro M, Cizek P. One-step robust estimation of fixed-effects panel data models. *Comput Stat Data An*. 2013;57(1):536-548.
20. Visek JA. Estimating the model with fixed and random effects by a robust method. *Methodol Comput Appl*. 2015;17(4):999-1014.
21. Namur VV, Luneburg JW. Robust estimation of linear fixed effects panel data models with an application to the exporter productivity premium. *J Econ Stat*. 2011;231(4):546-557.
22. Wagenvoort R, Waldmann R. On b-robust instrumental variable estimation of the linear model with panel data. *J Econ*. 2002;106(2):297-324.
23. Markatou M, Basu A, Lindsay B. Weighted likelihood estimating equations: the discrete case with applications to logistic regression. *J Stat Plan Infer*. 1997;57(2):215-232.
24. Cameron AC, Trivedi PK. *Microeconometrics Using Stata*. Texas: Stata Press; 2009.
25. Lindeboom M, Portrait F, van den Berg GJ. An econometric analysis of the mental-health effects of major events in the life of older individuals. *Health Econ*. 2002;11(6):505-520.
26. Hsiao C. *Analysis of Panel Data*. Cambridge, UK: Cambridge University Press; 2003.
27. Baltagi BH, Griffin JM. Short and long run effects in pooled models. *Int Econ Rev*. 1984;25(3):631-645.
28. Croissant Y, Millo G. Panel data econometrics in R: the PLM package. *J Stat Softw*. 2008;27(2):1-43.
29. Houthakker HS. New evidence on demand elasticities. *Econometrica*. 1965;33(2):277-288.
30. Dieleman JL, Templin T. Random-effects, fixed-effects and the within-between specification for clustered data in observational health studies: a simulation study. *Plos One*. 2014;9(10):e110257.
31. Allison PD. *Fixed Effects Regression Models*. Los Angeles: SAGE; 2009.
32. Neyman J, Scott EL. Consistent estimation from partially consistent observations. *Econometrica*. 1948;16(1):1-32.
33. Lancaster T. The incidental parameter problem since 1948. *J Econ*. 2000;95(2):391-413.
34. Griliches Z, Hausman JA. Errors in variables in panel data. *J Econ*. 1986;31(1):93-118.
35. Jirata MT, Chelule JC, Odhiambo RO. Deriving some estimators of panel data regression models with individual effects. *Int J Sci Res*. 2014;3(5):53-59.
36. Rousseeuw PJ. Least median of squares regression. *J Am Stat Assoc*. 1984;79(388):871-880.
37. Maronna RA, Yohai VJ. Robust regression with both continuous and categorical predictors. *J Stat Plan Infer*. 2000;89(1-2):197-214.
38. Gervini D, Yohai VJ. A class of robust and fully efficient regression estimators. *Ann Stat*. 2002;30(2):583-616.
39. Cizek P. Reweighted least trimmed squares: an alternative to one-step estimators. Center Discussion Paper Series 91/2010; 2010.
40. Lindsay B. Efficiency versus robustness: the case for minimum hellinger distance and related methods. *Ann Stat*. 1994;22(2):1018-1114.
41. Bai J, Li K. Theory and methods of panel data models with interactive effects. *Ann Stat*. 2014;42(1):142-170.
42. Aitken AC. On least squares and linear combinations of observations. *Proc Royal Soc Edinburgh*. 1935;55:42-48.
43. Markatou M, Basu A, Lindsay B. Weighted likelihood estimating equations with a bootstrap root search. *J Am Stat Assoc*. 1998;93(442):740-750.
44. Agostinelli C. Robust stepwise regression. *J Appl Stat*. 2002;29(6):825-840.
45. Agostinelli C, Markatou M. Test of hypotheses based on the weighted likelihood methodology. *Stat Sin*. 2001;11(2):499-514.
46. Markatou M. Robust statistical inference: weighted likelihoods or usual m-estimation? *Commun Stat-Theor M*. 1996;25(11):2597-2613.
47. Beran R. Minimum Hellinger distance estimates for parametric models. *Ann Stat*. 1977;5(3):445-463.
48. Agostinelli C, Markatou M. A one-step robust estimator for regression based on the weighted likelihood reweighting scheme. *Stat Probabil Lett*. 1998;37(4):341-350.
49. Basu A, Markatou M, Lindsay BG. *Weighted Likelihood Estimating Equations: The Continuous Case*. New York, NY: Department of Statistics, Columbia University; 1995.

How to cite this article: Hamiye Beyaztas B, Bandyopadhyay S. Robust estimation for linear panel data models. *Statistics in Medicine*. 2020;39:4421–4438. https://doi.org/10.1002/sim.8732