Annals of the Institute of Statistical Mathematics (2023) 75:773–798
https://doi.org/10.1007/s10463-022-00862-2

Robust density power divergence estimates for panel data models

Abhijit Mandal[1] · Beste Hamiye Beyaztas[2] · Soutir Bandyopadhyay[3]

Received: 5 February 2022 / Revised: 22 October 2022 / Accepted: 7 December 2022 /
Published online: 20 January 2023

Abstract

The panel data regression models have become one of the most widely applied statistical approaches in different fields of research, including social, behavioral, environmental sciences, and econometrics. However, traditional least-squares-based techniques frequently used for panel data models are vulnerable to the adverse effects of data contamination or outlying observations that may result in biased and inefficient estimates and misleading statistical inference. In this study, we propose a *minimum density power divergence* estimation procedure for panel data regression models with random effects to achieve robustness against outliers. The robustness, as well as the asymptotic properties of the proposed estimator, are rigorously established. The finite-sample properties of the proposed method are investigated through an extensive simulation study and an application to climate data in Oman. Our results demonstrate that the proposed estimator exhibits improved performance over some traditional and robust methods in the presence of data contamination.

Keywords Robust estimation · Minimum density power divergence · Panel data · Random-effect model

✉ Abhijit Mandal
amandal@utep.edu

[1] Department of Mathematical Sciences, University of Texas at El Paso, 500 W. University Ave, El Paso, TX 79968, USA

[2] Department of Statistics, Istanbul Medeniyet University, Unalan Mahallesi, 34700 Uskudar, Istanbul, Turkey

[3] Department of Applied Mathematics and Statistics, Colorado School of Mines, 1500 Illinois St., Golden, CO 80401, USA

1 Introduction

The advancements in applied and methodological research on panel data have been growing remarkably since the seminal paper of Balestra and Nerlove (1966). Panel data, sometimes referred to as longitudinal data, are multi-dimensional data consisting of observations collected over a period of time on the same set of cross-sectional units. This multi-dimensionality provides more information than either pure cross-sectional or pure time-series data since its grouping structure allows controlling the individual-specific heterogeneity and the intra-individual dynamics (Baltagi 2005; Hsiao 1985). Some advantages of using this type of data include more variability, more degrees of freedom, more efficiency and less collinearity between the variables, and better capability to identify and measure the effects that are not entirely detectable in a single cross-sectional or time-series data (Baltagi 2005). Moreover, one of its attractive features is that more informative data results in reliable statistical inference by improving the accuracy and precision in estimating model parameters (Beyaztas and Bandyopadhyay 2020).

The main factors affecting this excessive growth in panel data studies can be summarized as availability of panel data, greater ability for reflecting the complexity of human behavior than a single cross-sectional or time-series data, and challenging methodology (Hsiao 2007). In general, collecting panel data is more problematic than collecting cross-sectional or time-series data since such data consists of a large number of observations. However, with the latest advances in information technology and systems in different applications, generating and storing high throughput data have become much easier in recent times.

Due to the systematic differences across cross-sectional units, all pertinent information may not always be captured in regressions using aggregated time-series and pure cross-section data, and omission of such information may further lead to biased and inconsistent statistical inference results (Jirata et al. 2014). To this end, panel data models have become one of the cornerstone approaches in empirical research in fields of economics, social sciences, and medical sciences because these models allow explaining the individual behavior over time while capturing the inter-temporal dynamics. One of the most conspicuous among several attractive features of panel data models is their ability to account for the unobserved individual-specific heterogeneity. In this context, the fixed and random effects models are the most commonly used panel data regression models, including the individual-specific components. In the fixed-effects model, the unobserved heterogeneity across individuals captured by the time-invariant intercept terms is allowed to be correlated with the explanatory variables and assumed to be fixed. On the other hand, those individual effects are treated as a part of the disturbance term and controlled by the differences in the error variance components in the random-effects model. A key assumption that distinguishes these two principle models is that the individual-specific effects are assumed to be independently distributed of the explanatory variables in the random-effects model, while fixed-effects models allow for a limited form of endogeneity (Mundlak 1978). For a more detailed exposition of the research on technical details of linear panel data models and their applications, see Wallace and Hussain (1969), Maddala and Mount (1973), Mundlak (1978), Laird and Ware (1982), Cox and Hall (2002), Diggle et al. (2002), Fitzmaurice et al. (2004), Gardiner et al. (2009) and Athey et al. (2021).

Typically, panel data allows us to exploit different sources of variation: (i) variation within cross-sectional units (*within variation*), (ii) variation between cross-sectional units (*between variation*), and (iii) variation over both time and cross-sectional units (*overall variation*), and the estimation techniques differ depending on the source of variation (Kennedy 2003). For example, within estimator (also called *fixed-effects estimator*) utilizes within variation, while the generalized least squares (GLS) estimators used for random effects models take into account both within and between variations. The majority of the regression techniques rely on using the least-squares (LS)-based estimators for statistical inferences about the parameters of linear panel data models (Aquaro and Cizek 2013). However, obtaining consistent estimates of model parameters in traditional estimation techniques depends on some restrictive assumptions that may not be achieved in practice. Those assumptions such as normality and homoscedasticity of the error terms and strict exogeneity with respect to the error terms make the LS-based methods vulnerable to the adverse effects of the outliers and data contamination (Greene 2017; Kutner et al. 2004; Visek 2015). In panel data models, different types of outliers depending on the sources of contamination, i.e., vertical outliers, horizontal outliers (in the error term), and leverage points (in the explanatory variables), can arise because of the measurement error, typing error, transmission or copying error, and naturally unusual data points (Rousseeuw and van Zomeren 1990; Rousseeuw and Leroy 2003; Maronna et al. 2006; Bramati and Croux 2007; Bakar and Midi 2015). Moreover, the outliers may occur in a block form (called *block-concentrated outliers*) such that most of the outlying observations tend to be concentrated within cross-sectional units, i.e., in a few time-series (Bramati and Croux 2007). As a consequence, the LS-based approaches such as the ordinary least squares (OLS) estimator, GLS estimator, etc., may lead to substantial degradation of accuracy and erroneous estimates due to the sensitivity to outliers and/or any departure from the model assumptions. Furthermore, the outlying observations may not be determined by looking at the LS residuals or using standard outlier diagnostics due to the potential vulnerability of the complex nature of panel data to the masking effect. Although it is more crucial to have a robust method in the context of panel data models, especially in the presence of contaminated datasets, most of the efforts have been devoted to developing robust techniques for linear regression models, and the existing literature for static panel data models covers only a few approaches. Some of those robust techniques as alternatives to the fixed effects estimator have been developed by Bramati and Croux (2007) by considering high breakdown point of the well-known robust regression estimators, namely the least trimmed squares (LTS) estimator of Rousseeuw (1984) and MS estimates of Maronna and Yohai (2000). Also, a robust estimation procedure of Aquaro and Cizek (2013) in the context of linear panel data regression models with fixed effects includes to use of two different data transformations and applying the efficient weighted LS estimator of Gervini and Yohai (2002) and the reweighted LTS estimator of Cizek (2010) on the transformed data. Moreover, for fixed effects panel data models, the robust estimators proposed by Bakar and Midi (2015) have been developed by incorporating the MM-centering procedure and the within-group generalized M-estimator (WGM) of Bramati and Croux (2007). For the estimation of both fixed effects and random effects regression models, Visek (2015) has proposed a least weighted squares method based on mean-centering data, while a weighted least squares technique using MM-estimate of location has been developed by Midi and Muhammad (2018). More recently, Beyaztas and Bandyopadhyay (2020) have proposed robust versions of the OLS-based estimation procedures using weighted likelihood estimating equations methodology within the panel data regression models framework.

This study estimates model parameters using the density power divergence (DPD) based M-estimator proposed by Basu et al. (1998). The DPD is a family of statistical divergence that gives a non-negative measure of discrepancy between two densities. They are distance-like measures that equal zero only when the densities are identical. The DPD has been widely used in robust estimation, testing, and regression problems (Basu et al. 2013, 2017); also see Jana and Basu (2019) for further discussion on divergence estimators. Another family called γ-divergence with similar properties was introduced by Fujisawa and Eguchi (2008), and Fujisawa (2013). Kuchibhotla et al. (2019) provided a comparison between DPD and γ-divergence and proposed a bridge divergence.

We consider a linear panel data regression model with random effects as there are very few robust methods in this area. The main advantage of our proposed approach based on DPD is that it balances the desired efficiency and the robustness of the estimators by controlling a tuning parameter (Basu et al. 2018; Ghosh et al. 2016). We also offer an adaptive method of choosing the tuning parameter with no prior knowledge of outliers. Our method is equally good in pure data as well as heavy contamination. On the other hand, the classical rank, quantile, and trimmed-based robust techniques often sacrifice efficiency in pure data to achieve robustness (Lamarche 2010; Maciak 2021). Moreover, some methods like Beyaztas and Bandyopadhyay (2020) are only robust against random outliers, but not in the presence of cluster outliers or leverage points. Finally, the small sample performance of the DPD-based estimators is generally better than other robust estimators.

The rest of the paper is organized as follows. We introduce the linear panel data model in Sect. 2. In Sect. 3, we describe the density power divergence measure and the corresponding estimator for the linear panel data model. The theoretical properties, including the asymptotic distribution and the influence function of the proposed estimator, are presented in Sect. 4. We also propose a method to select the optimum DPD parameter by minimizing the asymptotic mean square error. Section 5 illustrates an extensive simulation study based on the proposed method and compares the results with the traditional techniques. The numerical results are further supported through a real data example from weather stations in Oman. Some concluding remarks are given in Sect. 6, and the proofs of the technical results are shown in Appendix and supplementary material.

2 Linear panel data models

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

$$y_{it} = x_{it}^T\beta + \alpha_i + \varepsilon_{it}, \quad i = 1, 2, \dots, N; \; t = 1, 2, \dots, T, \tag{1}$$

where the subscript i represents an individual observed at time t. Here, α_i's are the unobserved individual-specific effects (time-invariant characteristics),β is a $K \times 1$ vector of regression coefficients, 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 $E\left(\varepsilon_{it}|x_{i1}, \dots, x_{iT}, \alpha_i\right) = 0$, $E\left(\varepsilon_{it}^2|x_{i1}, \dots, x_{iT}, \alpha_i\right) = \sigma_\varepsilon^2$ and $E\left(\varepsilon_{it}\varepsilon_{is}|x_{i1}, \dots, x_{iT}, \alpha_i\right) = 0$ for $t \neq s$. The above panel data regression model can be represented in matrix form as

$$y = \alpha \otimes e_T + X\beta + \varepsilon, \tag{2}$$

where $y = \left(y_1, \dots, y_N\right)^T$ is an $NT \times 1$ vector obtained by stacking observations $y_i = \left(y_{i1}, \dots, y_{iT}\right)^T$ for individual $i = 1, \dots, N$. The $NT \times K$ matrix $X = \left(x_1, \dots, x_N\right)^T$ is formed with regressors $x_i = \left(x_{i1}^T, \dots, x_{iT}^T\right)^T$, α is an $N \times 1$ vector consisting of the individual effects α_i, e_T is a $T \times 1$ vector of ones and $\otimes$ denotes the Kronecker product.

If α_i is assumed to be random, then the random effects model can be succinctly written as

$$y_{it} = x_{it}^T\beta + \alpha_i + \varepsilon_{it} = x_{it}^T\beta + v_{it}, \quad \alpha_i \sim \text{iid}(0, \sigma_\alpha^2), \quad \varepsilon_{it} \sim \text{iid}(0, \sigma_\varepsilon^2), \tag{3}$$

where $v_{it} = \alpha_i + \varepsilon_{it}$ denotes a compound error term with $\sigma_v^2 = \sigma_\alpha^2 + \sigma_\varepsilon^2$ and $\text{cov}\left(v_{it}, v_{it'}\right) = \sigma_\alpha^2$. α_i's are assumed to be uncorrelated with ε_{it} and x_{it}. We further assume that α_i and ϵ_{it} are normally distributed for all i and t. Note that the outcomes from an individual are correlated as $\text{cov}\left(y_{it}, y_{it'}|X\right) = \sigma_\alpha^2$ for $t \neq t'$. One may also consider a specific structure of auto-correlation. It will change the form of Ω defined in the next section; however, all subsequent results can be derived in the same way.

One may build a robust model by assuming a heavy-tailed error distribution. However, as the outlying distribution is unknown in practice, it is often difficult to fit a suitable heavy-tailed distribution. Moreover, the corresponding estimator may be inefficient in pure data with no outlier. We use an M-estimator based on the DPD measure instead of the traditional least-squares or likelihood-based methods to avoid these problems. Although we assume normality of the error distribution, we allow the error term and the other parts of the panel data model to deviate considerably to form a contaminated model (see Lemma 1). The proposed estimator automatically down-weights outliers without first detecting them and ensures it converges to the target parameter.

3 Density power divergence

Let us consider a family of models $\{F_\theta, \theta \in \Theta\}$ with density f_θ. We denote $\mathcal{G}$ as the class of all distributions having densities with respect to the Lebesgue measure. Suppose $G \in \mathcal{G}$ is the true distribution with density g. Then, the density power divergence (DPD) measure between the model density f_θ and the true density g is defined as

$$d_\gamma(f_\theta, g) = \begin{cases} \int_y \left\{ f_\theta^{1+\gamma}(y) - \left(1 + \frac{1}{\gamma}\right) f_\theta^\gamma(y) g(y) + \frac{1}{\gamma} g^{1+\gamma}(y) \right\} dy, & \text{for } \gamma > 0, \\ \int_y g(y) \log \left(\frac{g(y)}{f_\theta(y)} \right) dy, & \text{for } \gamma = 0, \end{cases} \tag{4}$$

where γ is a tuning parameter (Basu et al. 1998). Note that G is not necessarily a member of the model family F_θ. Further, for $\gamma = 0$, the DPD measure is obtained as a limiting case of $\gamma \to 0^+$, and is the same as the Kullback–Leibler (KL) divergence. Generally, given a parametric model, we estimate θ by minimizing the DPD measure with respect to θ over its parametric space Θ. We call the estimator the *minimum power divergence estimator* (MDPDE). It is well-known that for $\gamma = 0$, minimization of the KL-divergent is equivalent to maximization of the log-likelihood function. Thus, the MLE can be considered a special case of the MDPDE when $\gamma = 0$.

Let $\theta = (\beta^T, \sigma_\alpha^2, \sigma_\epsilon^2)^T$ denote the parameter of the random effects model given in Eq. (3). We define the conditional probability density for disturbance terms, $\varepsilon_i + \alpha_i e_T = y_i - x_i\beta$ as

$$f_\theta(y_i|x_i) = (2\pi)^{-\frac{T}{2}} |\Omega|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2}(y_i - x_i\beta)^T \Omega^{-1} (y_i - x_i\beta) \right\}, \tag{5}$$

where

$$\Omega = E\left(v_i v_i^T\right) = \sigma_\varepsilon^2 I_T + \sigma_\alpha^2 e_T e_T^T \tag{6}$$

and I_T is the identity matrix of dimension T. In this case, we introduce the DPD measure based on the conditional density $f_\theta(y|x)$ as

$$d_\gamma(f_\theta, g) = \begin{cases} \int_x \int_y \left\{ f_\theta^{1+\gamma}(y|x) - \left(1 + \frac{1}{\gamma}\right) f_\theta^\gamma(y|x) g(y|x) + \frac{1}{\gamma} g^{1+\gamma}(y|x) \right\} h(x) dy dx, & \text{for } \gamma > 0, \\ \int_x \int_y g(y|x) \log \left(\frac{g(y|x)}{f_\theta(y|x)} \right) h(x) dy dx, & \text{for } \gamma = 0, \end{cases} \tag{7}$$

where $h(x)$ is the marginal probability density function of X and $g(y|x)$ is the true conditional density of Y given X.

For $\gamma > 0$, the DPD measure can empirically be written as

$$\hat{d}_\gamma(f_\theta, g) = \frac{1}{N} \sum_{i=1}^N \int_y f_\theta^{1+\gamma}(y|x_i) dy - \frac{1+\gamma}{N\gamma} \sum_{i=1}^N f_\theta^\gamma(y_i|x_i) + c(\gamma), \tag{8}$$

where $c(\gamma) = \frac{1}{N\gamma} \sum_{i=1}^{N} \int_y g_i^{1+\gamma}(y|x_i)dy$ does not depend on θ. The MDPDE of θ is then obtained by minimizing $\hat{d}_\gamma(f_\theta, g)$ over $\theta \in \Theta_0$. Alternatively, one may construct an iterative algorithm from the estimating equations given in Appendix. If some values of the response variable are missing, we may start with suitable initial values, such as the median, for the missing observations. Then, we get the first-stage MDPDE by minimizing the empirical measure or solving the estimating equations. In the next step, we predict the missing values using the fitted model. Finally, we continue the iteration until convergence.

Note that if the i-th observation is an outlier, then the value of $f_\theta(y_i|x_i)$ is very small compared to other samples. In that case, the second term of Equation (8) is negligible when $\gamma > 0$, thus the corresponding MDPDE becomes robust against outlier. On the other hand, when $\gamma = 0$, the KL divergent can be written as $\hat{d}_\gamma(f_\theta, g) = -\sum_{i=1}^{N} \log f_\theta(y_i|x_i) + d$, where d is independent of θ. For an outlying observation, the KL divergence measure diverges as $f_\theta(y_i|x_i) \to 0$. Therefore, the MLE breaks down in the presence of outliers as they dominate the loss function. In fact, the tuning parameter γ controls the trade-off between efficiency and robustness of the MDPDE— robustness measure increases if γ increases, but at the same time efficiency decreases.

It is important to define the 'true' θ or the target parameter when the data generating distribution $G(y|x)$ does not belong to the model family $\{F_\theta, \theta \in \Theta\}$. The true value of θ is defined with respect to the DPD measure that minimizes $d_\gamma(f_\theta, g)$ over $\theta \in \Theta$. So, in general, it depends on γ. A completely arbitrary distribution G does not make any sense in practical applications, so we assume G is a contaminated distribution close to the model density, where a small proportion (p) of data is from an arbitrary outlying distribution. Thus, the probability density function can be written as $g(y|x) = (1-p)f_{\theta_0}(y|x) + p\chi(y|x)$, for some $\theta_0 \in \Theta$ and χ being the outlying density. Then, the following lemma ensures that the target parameter is θ_0 under some assumptions.

Lemma 1 *Consider a contaminated model $g(y|x) = (1-p)f_{\theta_0}(y|x) + p\chi(y|x)$. Suppose there exist a small number $\gamma_0 > 0$, such that $\eta(\gamma) = \int_x \int_y f_\theta^\gamma(y|x)\chi(y|x)h(x)dxdy$ is sufficiently small for $\gamma > \gamma_0$, then the target parameter that minimizes the DPD measure $d_\gamma(f_\theta, g)$ is θ_0 for all values of $\gamma > \gamma_0$.*

A small value of $\eta(\gamma)$ ensures that χ is an outlying distribution or the effective mass of χ lies at the tail of the model distribution f_{θ_0} (Fujisawa and Eguchi 2008). If $\eta(\gamma)$ is sufficiently small, then under the contaminated model, $d_\gamma(f_\theta, g)$ is close to $d_\gamma(f_\theta, f_{\theta_0})$. Thus, the true value of θ is always θ_0 for $\gamma > \gamma_0$. Therefore, we keep the target parameter free of γ in the subsequent sections.

Remark 1 The empirical DPD measure $\hat{d}_\gamma(f_\theta, g)$ in Eq. (8) converges to $d_\gamma(f_\theta, g)$. So, under the assumption given in Lemma 1, the global minimizer of $d_\gamma(f_\theta, g)$ is at $\theta = \theta_0$. It can be further shown that $d_\gamma(f_\theta, g)$ is convex in the neighborhood of θ_0. Therefore, the algorithm for finding the MDPDE is a convex optimization problem. We observed in our numerical studies that the convergence rate is very fast, even in small sample sizes. However, it is worth mentioning that $d_\gamma(f_\theta, g)$ has a local minimizer at the center of the outlying distribution $\chi(y|x)$. But this is assumed to be far away from the center of the target distribution $f_{\theta_0}(y|x)$. So as long as the initial value of the parameter belongs to the effective range of the target distribution, the algorithm converges to the global minimizer. Thus, another robust estimator, such as the weighted likelihood estimator (WLE) of Beyaztas and Bandyopadhyay (2020), should be used as an initial value in the iterative optimization algorithm.

Remark 2 We consider a balanced panel data model as it is one of the most widely used models in different applications. However, we follow the same approach to generalize it when the data is unbalanced. Suppose the i-th individual is observed at n_i time points, where $i = 1, 2, \dots, k$ and $N = \sum_{t=1}^{k} n_i$ is the total sample size. Then, for $\gamma > 0$, the empirical DPD measure in Eq. (8) can be replaced by $\widehat{d}_\gamma(f_\theta, g) = \frac{1}{N}\sum_{i=1}^{k}\sum_{t=1}^{n_i}\int_y f_\theta^{1+\gamma}(y|x_{it})dy - \frac{1+\gamma}{N\gamma}\sum_{i=1}^{k}\sum_{t=1}^{n_i} f_\theta^{\gamma}(y_{it}|x_{it}) + c(\gamma)$, where $c(\gamma) = \frac{1}{N\gamma}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\int_y g_i^{1+\gamma}(y|x_{ij})dy$.

Finally, it is worth mentioning that the estimation procedure for the fixed-effects model is straightforward from the current work. For a fixed-effect model, the covariance matrix defined in Eq. (6) will be simply $\Omega = \sigma_\epsilon^2 I_T$ as σ_α vanishes. Thus, the calculations will be much simpler, and all theoretical properties will follow in the same way. Our method can also be extended to a dynamic panel data model, where $y_{it} = \rho y_{i,t-1} + x_{it}^T\beta + \alpha_i + \epsilon_{it}$ and ρ being the lag parameter. In this case, $y_i|x_i$ is also a multivariate normal distribution. However, the form of the covariance matrix Ω will change. We also have an additional parameter ρ in this model. So, all calculations will be modified accordingly.

4 Asymptotic distribution of the MDPDE

In this section, we present the asymptotic distribution of the MDPDE when the data generating distribution G is not necessarily in the model family. In most practical applications, the total number of time components (T) is fixed, so we derive the asymptotic distribution as $N \to \infty$. However, the same techniques are used when $T \to \infty$ for fixed N or both $T \to \infty$ and $N \to \infty$. Let us define the score function as $u_\theta(y_i|x_i) = \frac{\partial}{\partial\theta}\log f_\theta(y_i|x_i)$ (see Appendix 1 for more details). For $i = 1, 2, \dots, N$, we define

$$
\begin{aligned}
J^{(i)} &= \int_y u_\theta(y|x_i)u_\theta^T(y|x_i)f_\theta^{1+\gamma}(y|x_i)dy + \int_y \Big\{ I_\theta(y|x_i) - \gamma u_\theta(y|x_i)u_\theta^T(y|x_i) \Big\} \\
&\quad \times \Big\{ g(y|x_i) - f_\theta(y|x_i) \Big\} f_\theta^{\gamma}(y|x_i)dy, \\
K^{(i)} &= \int_y u_\theta(y|x_i)u_\theta^T(y|x_i)f_\theta^{2\gamma}(y|x_i)g(y|x_i)dy - \xi^{(i)}\xi^{(i)T}, \\
I_\theta(y|x_i) &= -\frac{\partial}{\partial\theta}u_\theta(y|x_i), \quad \xi^{(i)} = \int_y u_\theta(y|x_i)f_\theta^{\gamma}(y|x_i)g(y|x_i)dy.
\end{aligned} \tag{9}
$$

We further define $J = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} J^{(i)}$, $K = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} K^{(i)}$. The following theorem gives the consistency and the asymptotic distribution of the MDPDE.

Theorem 1 *Under the regularity conditions* (A1)–(A3) *of the Appendix, with probability tending to* 1 *as* $N \to \infty$, *there exists* $\widehat{\theta}$, *such that*

(i) $\widehat{\theta}$ *is consistent for* θ, *and*
(ii) *the asymptotic distribution of* $\widehat{\theta}$ *is given by*

$$\sqrt{N}(\widehat{\theta} - \theta) \sim N_{K+2}(0, J^{-1}KJ^{-1}), \tag{10}$$

where N_{K+2} denotes a $(K+2)$-dimensional normal distribution.

Proof The proof of the theorem is given in Appendix. □

Appendix gives J and K matrices at the model, i.e., when $g = f_\theta$. Further calculations show that the variance of each component of $\widehat{\beta}$ increases as γ increases. Therefore, the efficiency of the MDPDE decreases as γ increases—the MLE being the most efficient estimator. However, our simulation studies indicate that the loss of efficiency is very small. We use the expressions of J and K matrices to get the optimum value of γ by minimizing the MSE of the MDPDE.

4.1 Influence Function of the MDPDE

We further access the extent of the resistance to outliers of our proposed estimator using the influence function approach of Hampel et al. (1986). It measures the rate of asymptotic bias of an estimator to infinitesimal contamination in the distribution. A bounded influence function suggests that the corresponding estimator is robust against extreme outliers. Note that the MDPDE is an M-estimator (Huber 1981) as the estimating equation can be written as $\sum_i \Psi_\theta(y_i|x_i) = 0$, where

$$\Psi_\theta(y_i|x_i) = u_\theta(y_i|x_i) f_\theta^\gamma(y_i|x_i) - \int_y u_\theta(y|x_i) f_\theta^{1+\gamma}(y|x_i)dy. \tag{11}$$

This is obtained by differentiating $\widehat{d}_\gamma(f_\theta, g)$ with respect to θ from Eq. (8). Let $G(y|x)$ be the true conditional distribution function Y given X, and $\theta = T_\gamma(G)$ be functional for the MDPDE. Following Basu et al. (1998), the influence function of the MDPDE is given by

$$IF((x, y), T_\gamma, G) = J^{-1}\left\{u_\theta(y|x) f_\theta^\gamma(y|x) - \xi^{(i)}\right\}, \tag{12}$$

where J is evaluated at the model when $g = f_\theta$, and $\xi^{(i)}$, given in Eq. (39), is a fixed vector that does not depend on index i.

Note that the score function $u_\theta(y|x)$ in Eq. (33) is unbounded in both x and y. As a result, the influence function of the MLE, i.e., the MDPDE with $\gamma = 0$, is unbounded for the panel data regression model. On the other hand, $u_\theta(y|x)f_\theta^\gamma(y|x)$ is bounded in y when $\gamma > 0$ as the corresponding terms can be written as $y\exp(y^2)$. So, the influence function of the MDPDE of θ is bounded in y when $\gamma > 0$. Moreover, $IF((x,y), T_\gamma, G)$ tends to zero as $|y| \to \infty$, indicating a redescending effect for large vertical outliers. The higher the value of γ, the larger the down-weighting effect to the outliers. However, this influence function could still be unbounded in x, but only for small values of $|y|$ and simultaneously large values of $||x||$. This implies that good leverage points have the strongest effect on the MDPDE. But, if the leverage points are also vertical outliers, then the MDPDE will not be sensitive to those observations.

4.2 Choice of the optimum γ

One important use of the asymptotic distribution of the MDPDE is in selecting the optimum value of the DPD parameter γ. As β is the main parameter of interest in the panel data model, we choose γ that is optimum in terms of robustness and efficiency of $\hat{\beta}$. In practice, the user may work with a fixed value of γ depending on the desired level of robustness measure at the cost of efficiency. Alternatively, we may select a data-driven optimum γ. Following Warwick and Jones (2005), we minimize the mean square error (MSE) of $\hat{\beta}$ to obtain the optimum value of γ adaptively. Suppose Σ_β is the asymptotic variance of β obtained from Theorem 1, assuming that the true distribution belongs to the model family. Let $\hat{\Sigma}_\beta$ be the estimate of Σ_β. The empirical estimate of the MSE, as the function of a pilot estimator β^P, is given by

$$\widehat{MSE}(\gamma) = (\hat{\beta} - \beta^P)^T(\hat{\beta} - \beta^P) + \operatorname{tr}(\hat{\Sigma}_\beta). \tag{13}$$

In particular, we recommend that a robust estimator, such as the MDPDE with $\gamma = 0.5$, be used as a pilot estimator, i.e., $\beta^P = \hat{\beta}_{\gamma=0.5}$. Our first stage optimum parameter is the value of γ that minimizes $\widehat{MSE}(\gamma)$, keeping β^P fixed. One may also iterate this process by taking the current optimum MDPDE as the next stage's pilot estimator and proceeding until convergence. The iterative algorithm eliminates the sensitivity in the initial value of β^P, see Basak et al. (2021) for more details. In our numerical examples, we have used this iterative procedure. Alternatively, the optimum γ may be obtained from the robust versions of the Akaike information criterion (AIC) or Mallow's C_p as proposed in Mandal and Ghosh (2019) or the Hyvarinen score of Sugasawa and Yonekura (2021).

Lemma 1 shows that the target parameter is the same for all γ for the contaminated model. Moreover, Theorem 1 proves that all $\hat{\beta}$ converge to the target parameter. However, their small sample performance is very different depending on the contaminated proportion (p) and closeness of the contaminated distribution (χ) to the model distribution (f_{θ_0}). Thus, selecting the DPD parameter γ in finite samples is important to get the best performance.

5 Numerical results

To investigate the performances of our proposed method, we conducted an extensive simulation study under different sample sizes and different types of outliers. We compare the performance with the weighted likelihood estimator (WLE) of Beyaztas and Bandyopadhyay (2020) as a robust alternative method and the generalized least squares (GLS) and ordinary least squares (OLS) estimators. The GLS estimator is asymptotically equivalent to the MLE (Cameron and Trivedi 2005). As the robustness properties of the MDPDE depend on the choice of the tuning parameter, we have taken five fixed values of $\gamma = 0.05, 0.1, 0.2, 0.3$ and 0.4 along with the data-driven adaptive choice that minimizes the MSE as discussed in Sect. 4.2. The pilot estimator is modified iteratively until convergence.

We consider the random-effects model given in Eq. (3). The vector of regression coefficients is taken as $\beta = (2.0, 2.4, -1.2, 1.6, -0.5)^T$, where the first component is the intercept term. The individual-specific effects α_is are generated from the standard normal distribution. The explanatory variables x_{itk} for $k = 2, 3, 4, 5$ are generated as follows:

$$x_{itk} \sim \begin{cases} \chi_2^2 - 2 & \text{for } k = 2, \\ N(0, 1) & \text{for } k > 2, \end{cases} \tag{14}$$

where χ_2^2 represents the chi-square distribution with 2 degrees of freedom. One regressor is generated from a skewed distribution to avoid a symmetric experimental design.

To evaluate the performance of each estimator based on $S = 1,000$ simulations, we compute the mean squared errors (MSEs) of $\widehat{\beta}$ given by

$$MSE = \frac{1}{S} \sum_{s=1}^{S} \left\| \widehat{\beta}^s - \beta \right\|^2, \tag{15}$$

where $\widehat{\beta}^s$ is the estimate obtained from the s-th replication and β is the true value of the parameter. As all estimators are root-N consistent, i.e., $\sqrt{N}(\widehat{\beta} - \beta)$ is bounded in probability, the MSE values presented here are multiplied by N to show that they converge to a constant. Thus, in this section, we denote MSE as the MSE of $\sqrt{N}\widehat{\beta}$. In addition, the outlier deleted mean prediction error (MPE) is calculated to assess the predictive performance of the methods under consideration. We define MPE as

$$MPE = \frac{1}{S} \sum_{s=1}^{S} \left[\frac{1}{NT - m^{(s)}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(1 - c_{it}^{(s)}\right) \left(y_{it}^{(s)} - \widehat{y}_{it}^{(s)}\right)^2 \right], \tag{16}$$

where $y_{it}^{(s)}$ and $\widehat{y}_{it}^{(s)}$ are the actual and predicted value of the (i, t)-th element of y, respectively. The indicator variable $c_{it}^{(s)}$ is 1 if $y_{it}^{(s)}$ is an outlier, and 0 otherwise. Here, $m^{(s)}$ denotes the total number of outliers in the s-th simulated sample. We assume that there is no outlier in pure data, i.e., $c_{it}^{(s)} = 0$ for all i, t and s. We further calculated the MPE based on 1,000 new (pure) data from model (1).

5.1 Sample sizes

The performance of the estimators is examined under standard normal errors $\varepsilon_{it} \sim \mathrm{N}(0,1)$ for the increasing number of cross-sectional units ($N = 25, 50, 100, 200$) by keeping time period fixed at $T = 10$ and for the increasing values of time periods ($T = 5, 15, 20, 25$) when the number of cross-sectional units is fixed at $N = 100$. Table 1 illustrates the simulation results, and the best values are marked in boldface. For fixed N and T, the performance of all estimators are very similar. However, the OLS estimator has slightly low efficiency in terms of the MSE of $\widehat{\beta}$ as it considers a common effect $\alpha_i = \alpha$ in the model. Theoretically, it is shown that the MDPDE loses efficiency as γ increases. It is also observed in this table, although the difference is very small. The mean and standard deviation (SD) of the data-dependent optimum DPD parameter γ are also reported in this table. In most cases, they are close to zero, and therefore, their MSE and MPEs are almost identical to the GLS estimator. To save time in large-scale simulation, we used a gradient-based algorithm to find optimum γ instead of a grid search. We noticed that for some large values of T, the rate of change in the MSE is very low before a sudden jump, so the mean and standard deviation of the optimum γ parameter, produced by the gradient-based algorithm, are large. As a result, the corresponding optimum DPD has a slightly higher MSE. However, a grid search will produce improved results from the optimum DPD in this situation. The table shows that when T is fixed, the MSEs of $\sqrt{N}\widehat{\beta}$ for an estimator seem to be a constant for all values of N. It justifies that all estimators, including our proposed one, are root-N consistent.

5.2 Outliers

In the following simulation studies, the robustness properties of the estimators are evaluated in the presence of different types of outliers. The cross-sectional size $N = 100$ and time periods $T = 5$ are chosen for the panel size consisting of 500 observations. Three levels of contamination (p) are considered as 5%, 7.5%, and 10%. Outliers are inserted using two different ways to generate contaminated data—random contamination and concentrated contamination, as explained in Bramati and Croux (2007). The random contamination is obtained by distributing outlying data points randomly over all observations. On the other hand, the outliers clustered in all time points of some randomly selected blocks when generating concentrated contamination. The contamination schemes considered are explained below based on the types of outliers.

1. The random outliers in the y-direction, namely random vertical outliers, are obtained by replacing the original standard normal errors with $\varepsilon_{it} \sim \mathrm{N}(10,1)$ in the panel data model given in Eq. (3).
2. Concentrated vertical outliers are generated by substituting all errors in the selected random blocks by $\varepsilon_{it} \sim \mathrm{N}(10,1)$.
3. To obtain random leverage points, i.e., random contamination in both y-direction and x-direction, the same rule in the second scheme is applied to generate the

Table 1 MSEs of $\sqrt{N}\widehat{\beta}$ (first rows), MPEs from the existing data (second rows), and MPEs from 1,000 new data (third rows) of all estimators when $N = 25, 50, 100, 200$ for a fixed time dimension $T = 10$ (first 4 columns) and $T = 5, 15, 20, 25$ for a fixed cross-sectional dimension $N = 100$ (last 4 columns)

	N=25	N=50	N=100	N=200	T=5	T=15	T=20	T=25
	1.4938	1.5482	1.5000	1.4736	1.9981	1.3418	1.3026	1.1913
DPD(0.05)	1.9179	1.9530	1.9602	1.9595	1.9492	1.9616	1.9599	1.9636
	2.0607	1.9646	2.0111	1.9866	2.0356	2.0261	1.9575	2.0326
	1.5489	1.5988	1.5532	1.5197	2.0455	1.3920	1.3703	1.2511
DPD(0.1)	1.9198	1.9539	1.9606	1.9598	1.9496	1.9622	1.9605	1.9641
	2.0631	1.9657	2.0116	1.9869	2.0361	2.0266	1.9581	2.0332
	1.6714	1.7135	1.6723	1.6283	2.1781	1.4841	1.4492	1.3026
DPD(0.2)	1.9245	1.9562	1.9617	1.9603	1.9508	1.9632	1.9613	1.9647
	2.0684	1.9682	2.0129	1.9875	2.0376	2.0275	1.9590	2.0337
	1.7412	1.7794	1.7392	1.6909	2.3084	1.5138	1.4842	2.7105
DPD(0.3)	1.9273	1.9575	1.9624	1.9606	1.9521	1.9635	1.9615	1.9694
	2.0713	1.9696	2.0136	1.9878	2.0390	2.0278	1.9593	2.0478
	1.7653	1.8030	1.7640	1.7148	2.4056	1.6223	2.4667	1.3295
DPD(0.4)	1.9283	1.9580	1.9626	1.9607	1.9531	1.9636	1.9636	1.9828
	2.0724	1.9701	2.0139	1.9879	2.0401	2.0289	1.9694	2.0465
	1.4835	1.5439	1.4989	1.4692	1.9821	1.3418	1.3550	**1.0734**
DPD(Opt.)	1.9176	1.9528	1.9601	1.9595	1.9492	1.9615	1.9678	1.9665
	2.0602	1.9645	2.0110	**1.9866**	**2.0354**	2.0261	1.9616	**2.0321**
Mean Opt. γ	0.0104	0.0078	0.0071	0.0066	0.0069	0.0100	0.3331	0.1446
SD Opt. γ	0.0182	0.0141	0.0122	0.0291	0.0121	0.0218	0.1728	0.1711
	1.7896	1.8128	1.7535	1.7510	2.5147	1.5286	1.4313	1.2809
OLS	**1.9032**	**1.9457**	**1.9567**	**1.9577**	**1.9430**	**1.9589**	**1.9577**	**1.9617**
	2.0750	1.9713	2.0143	1.9884	2.0418	2.0285	1.9590	2.0337
	1.4680	**1.5266**	**1.4768**	**1.4567**	**1.9762**	**1.3291**	**1.2655**	1.1582
GLS	1.9170	1.9526	1.9601	1.9595	1.9491	1.9614	1.9597	1.9633
	2.0596	**1.9641**	**2.0108**	**1.9866**	**2.0354**	**2.0260**	**1.9571**	2.0322
	1.4858	1.5345	1.4773	1.4557	1.9851	1.3287	1.2655	1.1566
WLE	1.9171	1.9526	1.9600	1.9595	1.9491	1.9614	1.9596	1.9633
	2.0603	1.9643	2.0108	1.9866	2.0355	2.0260	1.9571	2.0322

The best performance is marked in bold font

error variable. Then, 50% values of the regressors corresponding to the outlying points are replaced by $x_{itk} \sim \mathrm{N}(5, 1)$.

These three sets of simulation results are presented in Table 2. They clearly indicate that the proposed estimator MDPDE (except $\gamma = 0.05$ and 0.1) outperforms the conventional estimators (OLS and GLS) and the robust WLE in all cases. The OLS and GLS methods result in obtaining severely distorted estimates of the parameters in the presence of outliers and, as a result, produce the largest MSE and MPE. The WLE works well in the presence of random contamination. But it breaks down when

Table 2 MSEs of $\sqrt{N}\hat{\beta}$ (first rows), MPEs from the existing data (second rows), and MPEs from 1,000 new (pure) data (third rows) of all estimators in the presence of $p = 5\%, 7.5\%, 10\%$ vertical outliers at random location (first 3 columns), concentrated panels (middle 3 columns) and concentrated vertical outliers with 50% leverage points (last 3 columns). In all cases, $N = 100$ and $T = 5$

	p = 5%	p =7.5%	p = 10%	p = 5%	p = 7.5%	p = 10%	p = 5%	p = 7.5%	p = 10%
	13.4765	38.1250	75.8345	10.6989	38.0571	67.1386	11.1601	35.5186	63.1344
DPD(0.05)	2.0909	2.3483	2.6965	2.0699	2.3354	2.6421	2.0698	2.3325	2.5907
	2.0480	2.4248	2.7367	2.0626	2.4278	2.5485	2.1032	2.3095	2.5805
	2.8505	13.8408	44.0286	2.3205	9.4933	31.6068	2.6046	10.6402	31.5572
DPD(0.1)	1.9866	2.1019	2.3799	1.9839	2.0432	2.2777	1.9843	2.0741	2.2714
	1.9601	2.1789	2.4151	2.0020	2.1424	2.2070	2.0049	2.0585	2.2496
	2.7694	3.3847	3.5993	2.2429	2.3260	2.5641	2.3662	2.3958	2.3923
DPD(0.2)	1.9872	1.9992	1.9888	1.9833	1.9721	1.9835	1.9830	1.9878	1.9784
	1.9621	2.0732	2.0042	2.0037	2.0710	1.9403	2.0013	1.9769	1.9300
	2.8655	3.4848	3.6596	2.3659	2.4298	2.6748	2.4694	2.5141	2.4955
DPD(0.3)	1.9883	2.0002	1.9898	1.9846	1.9733	1.9846	1.9842	1.9890	1.9795
	1.9633	2.0743	2.0048	2.0051	2.0722	1.9416	2.0023	1.9783	1.9310
	2.9224	3.5323	3.6840	2.4565	2.5080	2.7559	2.5497	2.6014	2.5816
DPD(0.4)	1.9889	2.0007	1.9901	1.9856	1.9742	1.9855	1.9852	1.9899	1.9804
	1.9639	2.0748	2.0051	2.0061	2.0731	1.9425	2.0031	1.9792	1.9318
	2.6908	**3.3042**	**3.5551**	**2.1520**	**2.2661**	**2.5012**	**2.3118**	**2.3327**	**2.3446**
DPD(Opt.)	**1.9859**	**1.9981**	**1.9879**	**1.9825**	**1.9716**	**1.9830**	**1.9822**	**1.9873**	**1.9780**
	1.9606	**2.0723**	**2.0037**	**2.0022**	**2.0704**	**1.9393**	**2.0009**	**1.9759**	**1.9297**
Mean Opt. γ	0.1257	0.1416	0.1582	0.1265	0.1412	0.1504	0.1295	0.1461	0.1544
SD Opt. γ	0.0087	0.0114	0.0161	0.0145	0.0163	0.0189	0.0152	0.0182	0.0201
	30.6475	64.1862	109.3536	30.6206	70.8888	107.7174	60.5823	90.9375	109.7130
OLS	2.2588	2.6070	3.0266	2.2588	2.6498	3.0334	2.5676	2.9081	3.0775

Table 2 (continued)

	p = 5%	p =7.5%	p = 10%	p = 5%	p = 7.5%	p = 10%	p = 5%	p = 7.5%	p = 10%
	2.2101	2.6890	3.0755	2.2527	2.7666	2.9599	2.6068	2.9008	3.0790
	30.3918	63.9904	109.1771	27.1822	65.5772	101.4264	26.5718	59.0418	91.4344
GLS	2.2618	2.6096	3.0290	2.2354	2.6121	2.9876	2.2230	2.5713	2.8743
	2.2069	2.6865	3.0733	2.2093	2.7029	2.8826	2.2672	2.5471	2.8733
	4.5184	5.6706	5.8394	23.4396	60.9042	96.2465	22.0408	52.8901	84.7819
WLE	1.9988	2.0138	2.0056	2.1982	2.5649	2.9351	2.1776	2.5084	2.8083
	1.9907	2.0962	2.0262	2.1753	2.6562	2.8321	2.2197	2.4848	2.8057

The best performance is marked in bold font outliers are clustered, or there are leverage points. For small values of γ, the MDPDEs are not robust against outliers. So DPD(0.05) and DPD(0.1) have large MSE and MPE; however, they are smaller than the GLS estimator. On the other hand, considering both performance metrics MSE and MPE, other DPD estimators are almost insensitive to different choices of contamination schemes and levels. Moreover, the MSE and MPE are somehow similar to the corresponding values in Table 1 for the pure data. It suggests that the performance of the MDPDE for large values of γ are robust against outliers. The data-dependent optimum MDPDE automatically selects a high value of γ where the estimated MSE is minimized. We notice that the mean value of the optimum γ increases with the contamination proportion.

Remark 3 In Table 1, we notice that the efficiency of the DPD estimators decreases as γ increases. On the other hand, Table 2 shows that initially, robustness increases, then again, it drops. The optimum value of γ is close to zero in pure data and between 0.1 and 0.2 under contamination. So, we have not increased γ beyond 0.4. The performance of MDPDE with $\gamma = 1$ further deteriorates in terms of both efficiency and robustness. In Tables 1 and 2, we have taken five different values of $\gamma = 0.05, 0.1, 0.2, 0.3$ and 0.4. The MLE is obtained when $\gamma = 0$, and Cameron and Trivedi (2005) mentioned that the GLS estimator is asymptotically equivalent to the MLE. So we have verified this result using simulation but omitted the $\gamma = 0$ case for redundancy.

Other than the above contamination schemes, we have used different sets of N and T, different error distributions (e.g., chi-square and t-distribution), placed the center of outliers at various points and observed a similar pattern in those cases. These outcomes suggest that the performance of the MDPDE is almost identical with the efficient classical methods, like GLS, in the pure data. On the other hand, in contaminated data, the MDPDEs with large values of γ yield an accurate and precise estimate of the parameters even when the WLE fails. The data-dependent optimum MDPDE successfully produces the optimum performance and adequately balances the efficiency in pure data and robustness properties in the contaminated data. As in real-data analysis, we generally do not have prior knowledge of the proportion and size of outliers, an adaptive choice of the DPD tuning parameter plays an important role. This simulation study also shows substantial indication and evidence of the theoretical robustness properties, and root-N consistency result of the MDPDE derived in this paper.

5.3 Case study

In this section, we apply the proposed methodology to analyze the Oman weather dataset, which is available on the National Center for Statistics & Information at https://data.gov.om/bixytwb/weather. This dataset consists of 660 observations ($N = 55$, $T = 12$) covering a cross section of 55 stations across Oman over the period January 2018 to December 2018. The list of weather stations is given in

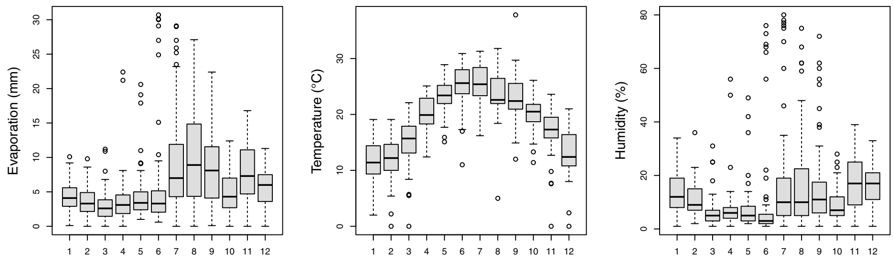

Fig. 1 Box-plots of the monthly evaporation (left panel), temperature (middle panel) and humidity (right panel) for Oman weather data

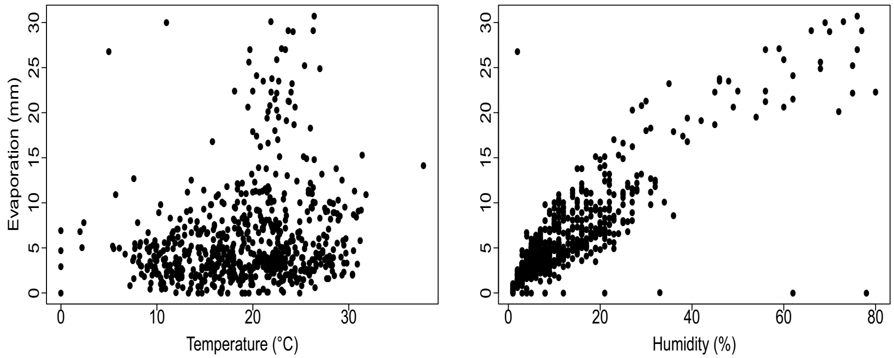

Fig. 2 Scatter plots of the monthly temperature versus evaporation (left panel) and humidity versus evaporation (right panel) for Oman weather data

Table 4 of Supplementary Material J. Our interest is in the relationships between the monthly minimum evaporation (mm) and two regressors—minimum temperature (°C) and minimum humidity (%). For this dataset, the following panel data regression model is conducted:

$$E_{it} = \beta_0 + \beta_1 Temp_{it} + \beta_2 H_{it} + v_{it} \quad i = 1, 2, \dots, 55; t = 1, 2, \dots, 12, \tag{17}$$

where E denotes the monthly minimum evaporation (mm) as a response variable, $Temp$ and H, respectively, represent the minimum temperature (°C) and minimum humidity (%) as explanatory variables. The marginal distributions of these three variables over different months are shown in Fig. 1. The scatter plots of monthly evaporation versus the temperature and humidity are given in Fig. 2. The first plot shows a cluster on the top that may arise due to unusual points in the response or explanatory variables. There are also a few large values of humidity in the second plot; moreover, some observations are very much isolated from the main region.

In real data, as we do not have prior knowledge of outliers, trimmed mean prediction errors are used to evaluate the performance of different estimators. For this, we trimmed a certain percentage (say $p\%$) of observations based on the highest squared prediction errors and calculated the average. The estimated individual coefficients,

Table 3 The estimates of individual coefficients and MPEs of all estimators for different trimmings

Trimming	Estimates	DPD(Opt.)	OLS	GLS	WLE
	$\widehat{\beta}_0$	−2.4570	−1.6281	−1.7009	−2.6841
	$\widehat{\beta}_1$	0.1764	0.1732	0.1860	0.2053
Full Data	$\widehat{\beta}_2$	0.4039	0.3429	0.3293	0.3867
	$\widehat{\sigma}_\alpha$	0.4674	–	1.1945	1.8627
	$\widehat{\sigma}_\epsilon$	1.5860	7.3529	6.0516	6.4275
$p = 20\%$	MPE	0.9351	1.0918	1.1727	1.0208
	Increase	–	16.75%	25.41%	9.16%
$p = 10\%$	MPE	1.5704	1.6937	1.8000	1.6208
	Increase	–	7.85%	14.62%	3.21%
$p = 5\%$	MPE	2.1579	2.2965	2.4379	2.1718
	Increase	–	6.42%	12.97%	0.64%

trimmed MPEs, and the percentage of trimmed MPE increased over the MDPDE are reported in Table 3. Three different values of trimming percentages are considered as $p = 20\%, 10\%$, and 5%. The optimum value of the DPD parameter based on the iterative method, discussed in Sect. 4.2, comes out to be $\gamma = 0.1439$ for this data set. From Table 3, we observe that the MDPDE with the optimum γ parameter yields the smallest MPE for all trimming percentages. The higher the trimming proportions, the better the performance from the MDPDE. For example, in the 20% trimming case, the MDPDE gives 16.75%, 25.41%, and 8.07% better prediction than the OLS, GLS, and WLE, respectively, for the remaining 80% observations. By construction, the OLS estimator produces the least MPE in the linear panel data regression model when all observations are considered. Thus, the corresponding regression line moves closer to the outliers by sacrificing the prediction power for good observations. On the other hand, the MDPDE better fits those observations close to the fitted model and is least affected by outliers. Being a robust method, the WLE gives a relatively smaller MPE; however, the MDPDE shows further improvement. All the results clearly demonstrate that the proposed estimator has considerably better predictive ability than the traditional and robust WLE methods.

6 Conclusions

We have proposed a robust procedure for estimating the parameters of the linear panel data regression model with random effects using the density power divergence. The efficiency and robustness properties of the proposed estimator MDPDE are controlled by a tuning parameter that can be estimated adaptively based on a given data set. The consistency and the asymptotic distribution of the MDPDE are theoretically derived. The influence function of the estimator indicates a redescending effect for vertical outliers. Our simulation studies show that the MDPDE with the optimum tuning parameter produces results almost as efficient as the GLS method in pure data, while at the same time, it outperforms even the existing robust techniques in the presence of outliers. The real data example also confirms that the MDPDE gives an excellent fit to the major part of the data, whereas the GLS fit move toward the outliers.

Appendix

A The Estimating Equations

Using Equation (17) of Supplementary Material, the DPD measure in Eq. (8) can be simplified as

$$\begin{aligned}\widehat{d}_\gamma(f_\theta, g) &= (2\pi)^{-\frac{T\gamma}{2}}|\Omega|^{-\frac{\gamma}{2}}(1+\gamma)^{-\frac{1}{2}} - \frac{1+\gamma}{N\gamma}\sum_{i=1}^{N} f_\theta^\gamma(y_i|x_i) + c(\gamma)\\ &= (2\pi)^{-\frac{T\gamma}{2}}|\Omega|^{-\frac{\gamma}{2}}(1+\gamma)^{-\frac{1}{2}}\left[1 - \frac{(1+\gamma)^{3/2}}{N\gamma}\sum_{i=1}^{N}\exp\left[-\frac{\gamma}{2}B_i\right]\right] + c(\gamma),\end{aligned} \tag{18}$$

where $B_i = (y_i - x_i\beta)^T\Omega^{-1}(y_i - x_i\beta)$. Using the Sherman–Morrison formula (Sherman and Morrison 1950), we get

$$\Omega^{-1} = \frac{1}{\sigma_\epsilon^2}I_T - \frac{\sigma_\alpha^2 e_T e_T^T}{\sigma_\epsilon^2(\sigma_\epsilon^2 + T\sigma_\alpha^2)}, \quad |\Omega| = \sigma_\epsilon^{2(T-1)}(\sigma_\epsilon^2 + T\sigma_\alpha^2). \tag{19}$$

It further simplifies B_i as follows

$$B_i = \frac{1}{\sigma_\epsilon^2}\sum_{t=1}^{T}(y_{it} - x_{it}\beta)^2 - \frac{\sigma_\alpha^2}{\sigma_\epsilon^2(\sigma_\epsilon^2 + T\sigma_\alpha^2)}\left\{\sum_{t=1}^{T}(y_{it} - x_{it}\beta)\right\}^2. \tag{20}$$

The estimating equations of θ is obtained from equation $\frac{\partial}{\partial\theta}\widehat{d}_\gamma(f_\theta, g) = 0$, and the equations corresponding to β, σ_α^2 and σ_ϵ^2 are simplified as

$$\begin{aligned}&\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}(y_{it} - x_{it}\beta)\exp\left[-\frac{\gamma}{2}B_i\right] = \frac{T\sigma_\alpha^2}{(\sigma_\epsilon^2 + T\sigma_\alpha^2)}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{x}_i(y_{it} - x_{it}\beta)\exp\left[-\frac{\gamma}{2}B_i\right],\\ &\gamma T|\Omega|^{-1}\sigma_\epsilon^{2(T-1)}\left\{(1+\gamma)^{-\frac{1}{2}} - \frac{1+\gamma}{N\gamma}\sum_{i=1}^{N}\exp\left[-\frac{\gamma}{2}B_i\right]\right\}\\ &\quad = -\frac{(1+\gamma)}{N(\sigma_\epsilon^2 + T\sigma_\alpha^2)^2}\sum_{i=1}^{N}\exp\left[-\frac{\gamma}{2}B_i\right]\left\{\sum_{t=1}^{T}(y_{it} - x_{it}\beta)\right\}^2,\\ &\gamma T|\Omega|^{-1}\sigma_\epsilon^{2(T-2)}\left\{\sigma_\epsilon^2 + (T-1)\sigma_\alpha^2\right\}\left\{(1+\gamma)^{-\frac{1}{2}} - \frac{1+\gamma}{N\gamma}\sum_{i=1}^{N}\exp\left[-\frac{\gamma}{2}B_i\right]\right\}\\ &\quad = \frac{(1+\gamma)}{N}\sum_{i=1}^{N}\exp\left[-\frac{\gamma}{2}B_i\right]\left[-\frac{1}{\sigma_\epsilon^4}\sum_{t=1}^{T}(y_{it} - x_{it}\beta)^2 + \frac{\sigma_\alpha^2(2\sigma_\epsilon^2 + T\sigma_\alpha^2)}{\sigma_\epsilon^4(\sigma_\epsilon^2 + T\sigma_\alpha^2)^2}\left\{\sum_{t=1}^{T}(y_{it} - x_{it}\beta)\right\}^2\right],\end{aligned} \tag{21}$$

where $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$. The MDPDE of θ is obtained by solving the above system of equations. One may use an iterative algorithm for this purpose or directly minimize the DPD measure in Eq. (8) with respect to $\theta \in \Theta_0$.

B Regularity Conditions

For the asymptotic distribution of the MDPDE, we need the following assumptions:

(A1) The true density $g(y|x)$ is supported over the entire real line $\mathbb{R}$.
(A2) There is an open subset $\omega \in \Theta_0$ containing the best fitting parameter θ such tat J is positive definite for all $\theta \in \omega$.
(A3) There exist functions $M_{jkl}(x, y)$ such tat $|\partial^3 \exp[(y - x\beta)^T \Omega^{-1}(y - x\beta)]/\partial\theta_j\partial\theta_k\partial\theta_l| \leq M_{jkl}(x, y)$ for all $\theta \in \omega$, where $\int_x \int_y |M_{jkl}(x, y)|g(y|x)h(x)dydx < \infty$ for all j, k and l.

Note that these regularity conditions hold good for the contaminated model, defined in Lemma 1, when $\eta(\gamma)$ is sufficiently small.

C Proof of Theorem 1

Proof The proof of the first part closely follows the consistency of the maximum likelihood estimator with the line of modifications as given in Theorem 3.1 of Ghosh and Basu (2013). For brevity, we only present the detailed proof of the second part.

Let $\hat{\theta}$ be the MDPDE of θ. Then

$$\frac{\partial}{\partial\theta}\hat{d}_\gamma(f_\theta, g) = \frac{\partial}{\partial\theta}\left[\frac{1}{N}\sum_{i=1}^{N}\int_y f_\theta^{1+\gamma}(y|x_i)dy - \frac{1+\gamma}{N\gamma}\sum_{i=1}^{N} f_\theta^\gamma(y_i|x_i)\right] = 0. \tag{22}$$

Thus, it can be written as the estimating equation of an M-estimator as follows

$$\sum_{i=1}^{N} \Psi_{\hat{\theta}}(y_i|x_i) = 0, \tag{23}$$

where

$$\Psi_\theta(y_i|x_i) = u_\theta(y_i|x_i)f_\theta^\gamma(y_i|x_i) - \int_y u_\theta(y|x_i)f_\theta^{1+\gamma}(y_i|x_i)dy. \tag{24}$$

Let θ_g be the true value of θ, then $E\left(\sum_{i=1}^{N}\Psi_{\theta_g}(y_i|x_i)\right) = 0$ gives

$$\sum_{i=1}^{N}\left[\int_y u_{\theta_g}(y|x_i)f_{\theta_g}^\gamma(y|x_i)g(y|x_i)dy - \int_y u_{\theta_g}(y|x_i)f_{\theta_g}^{1+\gamma}(y_i|x_i)dy\right] = 0. \tag{25}$$

Taking a Taylor series expansion of Eq. (23), we get

$$\frac{1}{N}\sum_{i=1}^{N}\Psi_{\theta_g}(y_i|x_i) + \frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\theta}\Psi_\theta(y_i|x_i)\Big|_{\theta=\theta_g}(\hat{\theta}-\theta_g) + R_N = 0,$$
$$\text{or } \sqrt{N}(\hat{\theta}-\theta_g) = -\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\theta}\Psi_\theta(y_i|x_i)\Big|_{\theta=\theta_g}\right]^{-1}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Psi_{\theta_g}(y_i|x_i) + \sqrt{N}R_N\right], \tag{26}$$

where R_N is the remainder term. Using the weak law of large numbers (WLLN), we have

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\theta}\Psi_\theta(y_i|x_i)\\
&\xrightarrow{p} \lim_{N\to\infty} E\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\theta}\Psi_\theta(y_i|x_i)\right]\\
&\xrightarrow{p} \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\left[\frac{\partial}{\partial\theta}\left(u_\theta f_\theta^\gamma - \int u_\theta f_\theta^{1+\gamma}\right)\right]\\
&\xrightarrow{p} \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E\left[-I_\theta f_\theta^\gamma + \gamma u_\theta u_\theta^T f_\theta^\gamma - \int\left\{-I_\theta f_\theta^{1+\gamma} + (1+\gamma)u_\theta u_\theta^T f_\theta^{1+\gamma}\right\}\right]\\
&\xrightarrow{p} \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[-\int I_\theta f_\theta^\gamma g + \gamma\int u_\theta u_\theta^T f_\theta^\gamma g + \int I_\theta f_\theta^{1+\gamma} - (1+\gamma)\int u_\theta u_\theta^T f_\theta^{1+\gamma}\right]\\
&\xrightarrow{p} -\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left[\int u_\theta u_\theta^T f_\theta^{1+\gamma} + \int\left(I_\theta - \gamma u_\theta u_\theta^T\right)(g-f_\theta)f_\theta^\gamma\right].
\end{aligned} \tag{27}$$

So

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\theta}\Psi_\theta(y_i|x_i)\Big|_{\theta=\theta_g} \xrightarrow{p} -\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}J^{(i)} = -J. \tag{28}$$

From Eq. (25), we get

$$\begin{aligned}
&E\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Psi_{\theta_g}(y_i|x_i)\right]\\
&= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\int_y u_{\theta_g}(y|x_i)f_{\theta_g}^\gamma(y|x_i)g(y|x_i)dy - \int_y u_{\theta_g}(y|x_i)f_{\theta_g}^{1+\gamma}(y_i|x_i)dy\right]\\
&= 0.
\end{aligned} \tag{29}$$

Now,

$$V\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Psi_{\theta_g}(y_i|x_i)\right] = \frac{1}{N}\sum_{i=1}^{N}V\Big[\Psi_{\theta_g}(y_i|x_i)\Big]$$
$$= \frac{1}{N}\sum_{i=1}^{N}\left[\int_y u_{\theta_g}(y|x_i)u_{\theta_g}^T(y|x_i)f_{\theta_g}^{2\gamma}(y|x_i)g(y|x_i)dy - \xi^{(i)}\xi^{(i)T}\right]$$
$$= \frac{1}{N}\sum_{i=1}^{N}K^{(i)}. \tag{30}$$

Following Section 5 of Ferguson (1996) or Section 2.7 of Lehmann (1999) and using Eqs. (29) and (30), the central limit theorem (CLT) for the independent but not identical random variables gives

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Psi_{\theta_g}(y_i|x_i) \overset{a}{\sim} N(0,K). \tag{31}$$

Under regularity condition (A3), it can be easily shown that the reminder term $\sqrt{N}R_N = o_p(1)$. Therefore, combining Eqs. (28) and (31), we get from Eq. (26)

$$\sqrt{N}(\hat{\theta}-\theta_g) \overset{a}{\sim} N\big(0, J^{-1}KJ^{-1}\big). \tag{32}$$

This completes the proof. □

D *J* and *K* Matrices at the model

Let us write the score function as

$$u_\theta(y_i|x_i) = (u_\beta^T(y_i|x_i), u_{\sigma_\alpha^2}^T(y_i|x_i), u_{\sigma_\epsilon^2}^T(y_i|x_i))^T. \tag{33}$$

Suppose $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T}x_{it}$. Then, it can be shown that

$$u_\beta(y_i|x_i) = \frac{\partial}{\partial\beta}\log f_\theta(y_i|x_i) = \frac{1}{\sigma_\epsilon^2}\sum_{t=1}^{T}x_{it}(y_{it}-x_{it}\beta) - \frac{T\bar{x}_i\sigma_\alpha^2}{\sigma_\epsilon^2(\sigma_\epsilon^2+T\sigma_\alpha^2)}\sum_{t=1}^{T}(y_{it}-x_{it}\beta),$$
$$u_{\sigma_\alpha^2}(y_i|x_i) = \frac{\partial}{\partial\sigma_\alpha^2}\log f_\theta(y_i|x_i) = -\frac{T}{2(\sigma_\epsilon^2+T\sigma_\alpha^2)} + \frac{1}{2(\sigma_\epsilon^2+T\sigma_\alpha^2)^2}\left\{\sum_{t=1}^{T}(y_{it}-x_{it}\beta)\right\}^2,$$
$$u_{\sigma_\epsilon^2}(y_i|x_i) = \frac{\partial}{\partial\sigma_\epsilon^2}\log f_\theta(y_i|x_i) = -\frac{1}{2(\sigma_\epsilon^2+T\sigma_\alpha^2)}\big[(T-1)\sigma_\epsilon^2(\sigma_\epsilon^2+T\sigma_\alpha^2)+1\big]$$
$$+\frac{1}{2\sigma_\epsilon^4}\sum_{t=1}^{T}(y_{it}-x_{it}\beta)^2 - \frac{\sigma_\alpha^2(2\sigma_\epsilon^2+T\sigma_\alpha^2)}{2\sigma_\epsilon^4(\sigma_\epsilon^2+T\sigma_\alpha^2)^2}\left\{\sum_{t=1}^{T}(y_{it}-x_{it}\beta)\right\}^2. \tag{34}$$

Note that if the true distribution $g(y|x)$ is a member of the model family $f_\theta(y|x)$ for some $\theta \in \Theta_0$, then

$$
\begin{aligned}
J^{(i)} &= \int_y u_\theta(y|x_i)u_\theta^T(y|x_i)f_\theta^{1+\gamma}(y|x_i)dy,\\
K^{(i)} &= \int_y u_\theta(y|x_i)u_\theta^T(y|x_i)f_\theta^{2\gamma+1}(y|x_i)dy - \xi^{(i)}\xi^{(i)T},\\
\xi^{(i)} &= \int_y u_\theta(y|x_i)f_\theta^{\gamma+1}(y|x_i)dy.
\end{aligned}
\tag{35}
$$

In this case, the symmetric matrix $J^{(i)}$ can be partitioned as

$$
J^{(i)} = \begin{bmatrix} J_\beta^{(i)} & J_{\beta,\,\sigma_\alpha^2}^{(i)} & J_{\beta,\,\sigma_\epsilon^2}^{(i)} \\ \cdot & J_{\sigma_\alpha^2}^{(i)} & J_{\sigma_\alpha^2,\,\sigma_\epsilon^2}^{(i)} \\ \cdot & \cdot & J_{\sigma_\epsilon^2}^{(i)} \end{bmatrix},
\tag{36}
$$

where

$$
\begin{aligned}
J_\beta^{(i)} &= M\sigma_\epsilon^{-4}\left[\sigma_\epsilon^2\sum_{t=1}^T x_{it}x_{it}^T + T^2\sigma_\alpha^2\left(\frac{T\sigma_\alpha^2}{(\sigma_\epsilon^2+T\sigma_\alpha^2)} - 1\right)\bar{x}_i\bar{x}_i^T\right],\\
J_{\sigma_\alpha^2}^{(i)} &= \frac{MT^2(\gamma^2+2)}{4(1+\gamma)(\sigma_\epsilon^2+T\sigma_\alpha^2)^2},\\
J_{\sigma_\epsilon^2}^{(i)} &= \frac{MT^2(\gamma-1)\left[\sigma_\epsilon^2+(T-1)\sigma_\alpha^2\right]^2}{4\sigma_\epsilon^4(\sigma_\epsilon^2+T\sigma_\alpha^2)^2}\\
&\quad + \frac{MT}{4\sigma_\epsilon^8}\left[(T+2)\sigma_\epsilon^4 + 2(T+2)\sigma_\epsilon^2\sigma_\alpha^2 + 3T\sigma_\alpha^4\right] + \frac{3MT^2\sigma_\alpha^4(2\sigma_\epsilon^2+T\sigma_\alpha^2)^2}{4\sigma_\epsilon^8(1+\gamma)(\sigma_\epsilon^2+T\sigma_\alpha^2)^2}\\
&\quad - \frac{TM(1+\gamma)\sigma_\alpha^2(2\sigma_\epsilon^2+T\sigma_\alpha^2)}{2\sigma_\epsilon^8(\sigma_\epsilon^2+T\sigma_\alpha^2)^2}\left[(T+2)\sigma_\epsilon^4 + (T^2+2T+3)\sigma_\epsilon^2\sigma_\alpha^2 + 3(T^2-T+1)\sigma_\alpha^4\right],\\
J_{\beta,\,\sigma_\alpha^2}^{(i)} &= 0,\ \ J_{\beta,\,\sigma_\epsilon^2}^{(i)} = 0,\\
J_{\sigma_\alpha^2,\,\sigma_\epsilon^2}^{(i)} &= \frac{TM(1+\gamma)}{4\sigma_\epsilon^4(\sigma_\epsilon^2+T\sigma_\alpha^2)^2}\left[2(T+1)\sigma_\epsilon^4 + (2T^2+T+3)\sigma_\epsilon^2\sigma_\alpha^2 + 3(T^2-T+1)\sigma_\alpha^4\right]\\
&\quad - \frac{3MT^2\sigma_\alpha^2(2\sigma_\epsilon^2+T\sigma_\alpha^2)}{4\sigma_\epsilon^4(1+\gamma)(\sigma_\epsilon^2+T\sigma_\alpha^2)^2} - \frac{MT^2\left[(T-1)\sigma_\alpha^2+\sigma_\epsilon^2\right]}{2\sigma_\epsilon^2(\sigma_\epsilon^2+T\sigma_\alpha^2)^2},
\end{aligned}
\tag{37}
$$

and

$$
M = (2\pi)^{-\frac{T\gamma}{2}}(1+\gamma)^{-\frac{T+2}{2}}\sigma_\epsilon^{-\gamma(T-1)}(\sigma_\epsilon^2+T\sigma_\alpha^2)^{-\frac{\gamma}{2}}.
\tag{38}
$$

Similarly, $\xi^{(i)}$ can be partitioned as $\xi^{(i)} = \left(\xi_\beta^{(i)T}, \xi_{\sigma_\alpha^2}^{(i)}, \xi_{\sigma_\epsilon^2}^{(i)}\right)^T$, and it is shown that

$$
\xi_\beta^{(i)} = 0, \quad \xi_{\sigma_\alpha^2}^{(i)} = -\frac{MT\gamma}{2(\sigma_\epsilon^2+T\sigma_\alpha^2)}, \text{ and } \xi_{\sigma_\epsilon^2}^{(i)} = -\frac{MT\gamma\left[\sigma_\epsilon^2+(T-1)\sigma_\alpha^2\right]}{2\sigma_\epsilon^2(\sigma_\epsilon^2+T\sigma_\alpha^2)}.
\tag{39}
$$

Note that if we write the matrix $J^{(i)}$ as a function of γ, i.e., $J^{(i)} \equiv J^{(i)}(\gamma)$, then we have

$$K^{(i)} = J^{(i)}(2\gamma) - \xi^{(i)}\xi^{(i)T}. \tag{40}$$

Moreover, $\xi_\beta^{(i)}$ is constant for all values of $i = 1, 2, \cdots, N$. Therefore, K can be written as

$$K = \frac{1}{N}\sum_{i=1}^{N} J^{(i)}(2\gamma) - \xi^{(i)}\xi^{(i)T}. \tag{41}$$

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10463-022-00862-2.

Acknowledgements The authors gratefully acknowledge the comments of two anonymous referees, which led to an improved version of the manuscript.

References

Aquaro, M., Cizek, P. (2013). One-step robust estimation of fixed-effects panel data models. *Computational Statistics and Data Analysis, 57*(1), 536–548.

Athey, S., Bayati, M., Doudchenko, N., et al. (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 1–15.

Bakar, N. M. A., Midi, H. (2015). Robust centering in the fixed effect panel data model. *Pakistan Journal of Statistics, 31*(1), 33–48.

Balestra, P., Nerlove, M. (1966). Pooling cross-section and time series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica, 34*(3), 585–612.

Baltagi, B. H. (2005). *Econometric analysis of panel data*. Chichester: John Wiley and Sons.

Basak, S., Basu, A., Jones, M. (2021). On the 'optimal' density power divergence tuning parameter. *Journal of Applied Statistics, 48*(3), 536–556.

Basu, A., Ghosh, A., Mandal, A., et al. (2017). A Wald-type test statistic for testing linear hypothesis in logistic regression models based on minimum density power divergence estimator. *Electronic Journal of Statistics, 11*(2), 2741–2772.

Basu, A., Harris, I. R., Hjort, N. L., et al. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika, 85*(3), 549–559.

Basu, A., Mandal, A., Martin, N., et al. (2013). Testing statistical hypotheses based on the density power divergence. *Annals of the Institute of Statistical Mathematics, 65*(2), 319–348.

Basu, A., Mandal, A., Martin, N., et al. (2018). Testing composite hypothesis based on the density power divergence. *Sankhya, Ser. B*, 80(2), 222–262.

Beyaztas, B. H., Bandyopadhyay, S. (2020). Robust estimation for linear panel data models. *Statistics in Medicine, 39*(29), 4421–4438.

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

Cameron, A. C., Trivedi, P. K. (2005). *Microeconometrics*: *Methods and applications*. New York Cambridge University Press.

Cizek, P. (2010). Reweighted least trimmed squares: an alternative to one-step estimators. *CentER Discussion Paper Series 91/2010*.

Cox, D. R., Hall, P. (2002). Estimation in a simple random effects model with nonnormal distributions. *Biometrika, 89*(4), 831–840.

Diggle, P. J., Heagerty, P., Liang, K.-Y., et al. (2002). *Analysis of Longitudinal Data*. United Kingdom Oxford University Press.

Ferguson, T. S. (1996). *A course in large sample theory*. Texts in Statistical Science Series. London Chapman & Hall.

Fitzmaurice, G. M., Laird, N. M., Ware, J. H. (2004). *Applied longitudinal analysis*. New York: John Wiley and Sons.

Fujisawa, H. (2013). Normalized estimating equation for robust parameter estimation. *Electronic Journal of Statistics, 7*, 1587–1606.

Fujisawa, H., Eguchi, S. (2008). Robust parameter estimation with a small bias against heavy contamination. *Journal of Multivariate Analysis, 99*(9), 2053–2081.

Gardiner, J. C., Luo, Z., Roman, L. A. (2009). Fixed effects, random effects and gee: What are the differences? *Statistics in Medicine, 28*(2), 221–239.

Gervini, D., Yohai, V. J. (2002). A class of robust and fully efficient regression estimators. *The Annals of Statistics, 30*(2), 583–616.

Ghosh, A., Basu, A. (2013). Robust estimation for independent non-homogeneous observations using density power divergence with applications to linear regression. *Electronic Journal of Statistics, 7*, 2420–2456.

Ghosh, A., Mandal, A., Martin, N., et al. (2016). Influence analysis of robust Wald-type tests. *Journal of Multivariate Analysis, 147*, 102–126.

Greene, W. H. (2017). *Econometric analysis*. New York: Prentice Hall.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., et al. (1986). *Robust statistics*: *The approach based on influence functions*. New York: John Wiley & Sons Inc.

Hsiao, C. (1985). Benefits and limitations of panel data. *Econometric Reviews, 4*(1), 121–174.

Hsiao, C. (2007). Panel data analysis - advantages and challenges. *Test, 16*(1), 1–22.

Huber, P. J. (1981). *Robust statistics*. John Wiley & Sons, Inc., New York. Wiley Series in Probability and Mathematical Statistics.

Jana, S., Basu, A. (2019). A characterization of all single-integral, non-kernel divergence estimators. *IEEE Transactions on Information Theory, 65*(12), 7976–7984.

Jirata, M. T., Chelule, J. C., Odhiambo, R. O. (2014). Deriving some estimators of panel data regression models with individual effects. *International Journal of Science and Research, 3*(5), 53–59.

Kennedy, P. (2003). *A guide to econometrics*. Cambridge: The MIT Press.

Kuchibhotla, A. K., Mukherjee, S., Basu, A. (2019). Statistical inference based on bridge divergences. *Annals of the Institute of Statistical Mathematics, 71*(3), 627–656.

Kutner, M. H., Nachtsheim, C. J., Neter, J. (2004). *Applied linear regression models*. New York: McGraw-Hill Education.

Laird, N. M., Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics, 38*(4), 963–974.

Lamarche, C. (2010). Robust penalized quantile regression estimation for panel data. *Journal of Econometrics, 157*(2), 396–408.

Lehmann, E. L. (1999). *Elements of large-sample theory*. New York: Springer Texts in Statistics. Springer-Verlag.

Maciak, M. (2021). Quantile LASSO with changepoints in panel data models applied to option pricing. *Econometrics and Statistics, 20*, 166–175.

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*(342), 324–328.

Mandal, A., Ghosh, S. (2019). Robust variable selection criteria for the penalized regression. *arXiv preprint* arXiv:1912.12550.

Maronna, R. A., Martin, R. D., Yohai, V. J. (2006). *Robust statistics. Theory and methods*. New York: John Wiley and Sons.

Maronna, R. A., and Yohai, V. J. (2000). Robust regression with both continuous and categorical predictors. *Journal of Statistical Planning and Inference, 89*(1–2), 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*(17), 7156–7161.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica, 46*(1), 69–85.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association, 79*(388), 871–880.

Rousseeuw, P. J., Leroy, A. M. (2003). *Robust regression and outlier detection*. New York: John Wiley and Sons.

Rousseeuw, P. J., van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association, 85*(41), 633–639.

Sherman, J., Morrison, W. J. (1950). Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics, 21*(1), 124–127.

Sugasawa, S., Yonekura, S. (2021). On selection criteria for the tuning parameter in robust divergence. *Entropy, 23*(9), 1147.

Visek, J. A. (2015). Estimating the model with fixed and random effects by a robust method. *Methodology and Computing in Applied Probability, 17*(4), 999–1014.

Wallace, T. D., Hussain, A. (1969). The use of error components models in combining cross section and time-series data. *Econometrica, 37*(1), 55–72.

Warwick, J., Jones, M. (2005). Choosing a robustness tuning parameter. *Journal of Statistical Computation and Simulation, 75*(7), 581–588.