Int. J. Adv. Eng. Pure Sci. 2025, 37(UYIK 2024 Special Issue): <7-20>
DOI: 10.7240/jeps.1506329

RESEARCH ARTICLE / ARAŞTIRMA MAKALESİ

A Robust Approach using M-Estimation for Dynamic Panel Autoregressive Model

Beste Hamiye BEYAZTAŞ [1]

[1] *İstanbul Medeniyet Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, İstatistik Bölümü, İstanbul, Türkiye*

Abstract
This paper presents a robust M-estimation approach for first-order panel autoregressive models, addressing the challenges posed by high persistence levels of the autoregressive parameter and individual heterogeneity. Generalized method of moments estimators widely used in dynamic panel models exhibit substantial finite sample biases and are sensitive to weak instruments, particularly as the autoregressive parameter gets close to unity. Our proposed weighted M-estimator, which uses a power function for the scale parameter in Huber's loss function, offers a robust alternative. By minimizing the variance of model parameters through an optimal tuning parameter, our method enhances the efficiency and robustness of parameter estimates. We demonstrate the superiority of the proposed approach through several Monte-Carlo simulations and an application to hydro-electric power output data, providing comprehensive comparisons with existing generalized method of moments estimators.
Keywords: Panel autoregressive models, Robust estimation, M-estimation, Generalized method of moments

I. INTRODUCTION

Dynamic panel data (DPD) models play a pivotal role not only in econometrics but also in engineering and the natural sciences, serving as fundamental analytical tools, especially when dealing with data that evolves over time and across individual units. These models allow us to capture the temporal dynamics of data, thus facilitating the understanding of various phenomena in different fields of research, such as physics, biology, environmental science, engineering, econometrics and so on. For instance, DPD models can be used to analyze the long-term effects of greenhouse gas emissions, oceanic circulation patterns, and temperature fluctuations over time in the field of climatology. As another example, in biology, the growth and development of organisms, the spread of diseases, and the interactions between species in ecosystems can be investigated via these models, considering both short-term and long-term dynamics. The behavior of complex systems such as electrical circuits, mechanical systems, or chemical processes in engineering can also be addressed with these models to discover hidden patterns, relationships, and trends in data that may not be evident with simpler models.

The flexibility of DPD models highlights their importance as a powerful analytical tool across various disciplines. By incorporating the unobserved individual-specific effects, these models account for unobserved heterogeneity across individuals, leading to enhanced insights (cf. [1]). Additionally, the inclusion of lagged dependent variables as explanatory variables in DPD models is a crucial feature that distinguishes them from static panel data models. This feature enables them to capture both the short- and long-term dynamics of the data and allows for modeling of persistence within the data.

Estimating DPD models involves addressing several issues stemming from endogeneity, potential correlation between the individual-specific effects and the explanatory variables, and unobserved heterogeneity. Using well-known least squares (LS) techniques for dynamic models may result in obtaining inconsistent estimates of the parameters when dealing with panel data with a small time dimension. This inconsistency arises due to the presence of endogenous explanatory variables, which introduce correlation between the regressors and error terms. Even with large samples, LS techniques, such as fixed effects (LSDV) or random effects (GLS), may still exhibit bias, as noted in [2]. Furthermore, [3, 4] address the inconsistency of the the maximum likelihood estimator (MLE) when dealing with a large number of individuals (N) and a fixed number of time periods (T), which arises from the increase in parameters with the increasing number of individuals, resulting in an incidental parameter problem. This has prompted likelihood-based approaches aimed at addressing this issue, such as the conditional likelihood estimator outlined in [5], and estimators based on the the first differences, as proposed by [6, 7, 8]. Also, for a detailed discussion on the finite sample properties of the MLE within the scope of dynamic

Corresponding Author: BESTE HAMİYE BEYAZTAŞ, **Tel:** 0216 280 40 51, **E-mail:** beste.sertdemir@medeniyet.edu.tr
Submitted: 28.06.2024, **Revised:** 29.06.2024, **Accepted:** 12.07.2024

panel data models, see [9]. More recently, likelihood-based estimators for autoregressive panel data models, which are robust in the presence of heteroskedasticity, have been proposed by [10].

The primary focus in literature for the estimation of dynamic panel data models has been on a class of generalized method of moments (GMM) estimators. GMM estimation offers a flexible approach by exploiting moment conditions derived from the sample moment counterparts of population moment conditions, often referred to as orthogonality conditions, of the data-generating model (cf. [11]). As pointed out by [11], the main sources leading to the widespread use of GMM estimators include: (*i*) Providing a simple approach for demonstrating the asymptotic properties of GMM estimators, and (*ii*) The capability to construct them without specifying the complete data generation process. For comprehensive discussions of GMM estimation with a wide range of applications, see [12, 13, 14, 15]. The GMM estimators based on first-difference transformation proposed by [16, 17, 18] have led to the beginning of an extensive literature. Although the first-difference based GMM estimators yield consistent estimates for large cross-sectional size, they exhibit substantial finite sample bias, particularly when dealing with strongly persistent data and weak instruments (cf. [1, 19, 20, 21]). In order to enhance the finite sample properties of standard GMM estimators, several alternative estimators, such as level GMM (LEV) estimator of [22] and system GMM (SYS) estimator of [19], have been developed. These estimators can be considered as extensions of the standard GMM estimators by incorporating additional moment conditions derived from the level equations for LEV estimator and from the model in first differences and levels for SYS estimator. Though exploiting many instruments, leads to an improvement in the efficiency of GMM estimators and addresses weak instrument and incidental parameter issues, as noted in [23], these estimators still exhibit bias. Moreover, the SYS estimator may result in increasingly biased estimates and weak instrument issues in the presence of a large variance ratio of the individual-specific effects to the idiosyncratic errors and/or an autoregressive coefficient that is close to unity (cf. [2, 24]). The finite sample biases of the SYS estimators have been investigated by [25]. Furthermore, [26] have proposed a consistent GMM estimator with less bias in the presence an autoregressive coefficient that is close to unity.

The aim of this paper is to develop a robust M-estimator when the value of the autoregressive parameter is near unity and/or the variance of individual effects differs from the variance of the error terms, where the class of GMM estimators is highly sensitive to the increasing level of persistence and individual heterogeneity. In this paper, we propose an extension of the weighted M-estimation approach introduced by [27] to estimate the parameters of the first-order autoregressive panel data models. The proposed robust estimator, based on Huber's loss function, is obtained by weighting the M-estimator with a power function for the scale parameter. Also, the optimal value of the tuning parameter related to the loss function has been chosen with the aim of minimizing the variance of the model parameters and based on the data distribution, as in [28, 29].

The rest of the paper is organized as follows. In Section 2, we begin by presenting comprehensive information on first-order autoregressive panel data models and existing GMM estimators. Subsequently, we describe our approach to obtain the proposed M-estimator, which is weighted by a power function applied to the scale parameter in Huber's loss function. The finite sample properties of the proposed estimator are demonstrated through an extensive simulation study, and the results are compared with those of traditional GMM estimators in Section 3. In Section 4, we apply our proposed method to hydro-electric power output data to further validate its applicability. Finally, a few concluding remarks are provided in Section 5.

II. METHODOLOGY

2.1. First-Order Autoregressive Panel Model and GMM Estimators

We consider the first-order autoregressive panel model described as follows

$$y_{it} = \varphi y_{i,t-1} + \alpha_i + \varepsilon_{it};\ \ i = 1, \dots, N;\ \ t = 2, \dots, T \qquad \textbf{(1)}$$

where α_i, ε_{it} and y_{it} respectively represent unobserved individual-specific effects, idiosyncratic error terms and the response variable for an individual i observed at time t and φ is the autoregressive parameter under the stationarity assumption that $|\varphi| < 1$. For this simple DPD model, α_i's are assumed to be independent and identically distributed (iid) across individuals, with $E(\alpha_i) = 0, Var(\alpha_i) = \sigma_\alpha^2$, and ε_{it}'s are iid across time and individuals, with $E(\varepsilon_{it}) = 0, Var(\varepsilon_{it}) = \sigma_\varepsilon^2$ (cf. [19]). Also, for mean stationarity on the process, it is assumed that $E(y_{i1}\varepsilon_{it}) = 0$ and $E(\alpha_i \varepsilon_{it}) = 0$ (cf. [19]). An additional assumption developed by [30] has been imposed on initial observations as follows

$$y_{i1} = \frac{\alpha_i}{1-\varphi} + \mu_{i1} \text{ for } i = 1, \dots, N \qquad \textbf{(2)}$$

where $\mu_{i1} = \sum_{j=0}^{\infty} \varphi^j \varepsilon_{i,1-j}$ is independent of α_i. By defining $y_i = (y_{i3}, \dots, y_{iT})'$, $y_{i,-1} = \left(y_{i2}, \dots, y_{i,T-1}\right)'$

and $u_i = (u_{i3}, \dots, u_{iT})'$, Equation (2) can be expressed as

$$y_i = \varphi y_{i,-1} + u_i \tag{3}$$

where $u_{it} = \alpha_i + \varepsilon_{it}$.

Under the assumptions given above, we examine three commonly used GMM estimators: the first difference (DIF) GMM estimator, the LEV GMM estimator, and the SYS GMM estimator. The GMM estimators are constructed using some moment conditions, with the asymptotic covariance of these moment conditions as the weight matrix (cf. [2]). Employing two-step procedures improves the asymptotic efficiency of the standard GMM estimators. The one-step GMM estimate is derived using an initial positive semidefinite weight matrix, which is independent of estimated parameters (cf. [31]). Then, the weight matrix, which includes residuals from the one-step estimation, is used to obtain the two-step GMM estimate. Also, estimated standard errors using two-step procedure tend to show a downward bias in small samples, leading to a preference for one-step estimates with robust standard errors as noted in [1, 18, 31]. Next, we briefly discuss the one-step and two-step DIF GMM, LEV GMM and SYS GMM estimation for first-order autoregregressive panel data models.

2.1.1. First difference GMM estimator

The DIF GMM estimator transforms the model, given in Equation (3), into a system of equations in first differences to address the correlation between the lagged endogenous variable $(y_{i,-1})$ and the error term (u_i) stemming from the individual effect (α_i). Therefore, to eliminate the individual effects, [18] employ the first differences of Equation (3) as follows:

$$\Delta y_i = \varphi \Delta y_{i,-1} + \Delta u_i$$

where $\Delta y_i = \left(y_{i3} - y_{i2}, \dots, y_{iT} - y_{i,T-1}\right)'$, $\Delta y_{i,-1} = \left(y_{i2} - y_{i1}, \dots, y_{i,T-1} - y_{i,T-2}\right)'$, and $\Delta u_i = \left(u_{i3} - u_{i2}, \dots, u_{iT} - u_{i,T-1}\right)'$. By exploiting $m_D = (1/2)(T-1)(T-2)$ orthogonality conditions, $E\left(Z_i^{D'} \Delta u_i\right) = 0$, where Z_i^D denotes a $(T-2) \times m_D$ instrumental variable matrix given below,

$$Z_i^D = \begin{pmatrix} y_{i1} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & y_{i1} & y_{i2} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & y_{i1} & \cdots & y_{i,T-2} \end{pmatrix}$$

one-step DIF GMM estimator (DIF_1) of [18] for φ is calculated as

$$\hat{\varphi}_{DIF_1} = \left(\Delta y_{-1}' Z^D W^D Z^{D'} \Delta y_{-1}\right)^{-1} \Delta y_{-1}' Z^D W^D Z^{D'} \Delta y$$

where $\Delta y_{-1} = \left(\Delta y_{1,-1}', \dots, \Delta y_{N,-1}'\right)'$, $Z^D = \left(Z_1^{D'}, \dots, Z_N^{D'}\right)'$, $\Delta y = (\Delta y_1', \dots, \Delta y_N')'$, and

$$W^D = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^{D'} D Z_i^D\right)^{-1}$$

where D is a $(T-2) \times (T-2)$ square Toeplitz matrix as follows

$$D = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 2 \end{pmatrix}.$$

Using the residuals from DIF_1, $\Delta\hat{u}_i$, two-step DIF-GMM (DIF_2) estimator is obtained as

$$\hat{\varphi}_{DIF_2} = \left(\Delta y_{-1}' Z^D W_2^D Z^{D'} \Delta y_{-1}\right)^{-1} \Delta y_{-1}' Z^D W_2^D Z^{D'} \Delta y$$

where W_2^D denotes the weighting matrix as $W_2^D = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^{D'} \Delta\hat{u}_i \Delta\hat{u}_i' Z_i^D\right)^{-1}$.

[19] demonstrate that using lagged levels as instruments to obtain the first difference GMM estimator results in weak instrumental variables, and the instruments become invalid when φ is close to unity and/or when $\sigma_\alpha^2/\sigma_\varepsilon^2$ increases.

2.1.2. Level GMM estimator

The primary motivation behind level GMM estimator developed by [22] has been to remove individual effects from instrumental variables to address the correlation problem between error terms, u_i and lagged endogenous variables, $y_{i,-1}$ caused by the presence of individual effects, α_i. For consistency with a large N and a fixed time period T, LEV GMM estimator requires the mean stationarity of the process, which implies that the assumption on the initial observations is satisfied.

The level model described in Equation (3) is used to obtain LEV GMM estimator, with the assumption that the $m_L = (1/2)(T-1)(T-2)$ moment conditions, $E\left(Z_i^{L'} u_i\right) = 0$, hold. Here, Z_i^L represents a $(T-2) \times m_L$ instrumental variable matrix as follows

$$Z_i^L = \begin{pmatrix} \Delta y_{i2} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \Delta y_{i2} & \Delta y_{i3} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \Delta y_{i2} & \cdots & \Delta y_{i,T-1} \end{pmatrix}.$$

Using these orthogonal conditions, the matrix of instruments, $Z^L = \left(Z_1^{L'}, \dots, Z_N^{L'}\right)'$, and weighting matrix, W^L, described as,

$$W^L = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^{L'} Z_i^L\right)^{-1}$$

one-step LEV GMM estimator (LEV_1) of φ is obtained as follows

$$\hat{\varphi}_{LEV_1} = \left(y'_{-1}Z^L W^L Z^{L'} y_{-1}\right)^{-1} y'_{-1}Z^L W^L Z^{L'} y$$

where $y_{-1} = \left(y'_{1,-1}, \dots, y'_{N,-1}\right)'$ and $y = (y'_1, \dots, y'_N)'$.

In the second step, the weighting matrix W_2^L is constructed by using the instrument matrix, which is weighted by the fitted residuals from the LEV_1 estimator, $\hat{u}_i$'s, as follows

$$W_2^L = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^{L'} \hat{u}_i \hat{u}'_i Z_i^L\right)^{-1}$$

and the two-step LEV GMM estimator (LEV_2) for φ is computed as

$$\hat{\varphi}_{LEV_2} = \left(y'_{-1}Z^L W_2^L Z^{L'} y_{-1}\right)^{-1} y'_{-1}Z^L W_2^L Z^{L'} y$$

2.1.3. System GMM estimator

A System GMM estimator (SYS) introduced by [19] combines the moment conditions of the DIF and LEV approaches to handle weak instrument problem and enhance the efficiency of the estimator. The model incorporating all equations both in first differences and in levels can be reformulated as a system of equations as follows.

$$\begin{pmatrix}\Delta y_i \\ y_i\end{pmatrix} = \varphi \begin{pmatrix}\Delta y_{i,-1} \\ y_{i,-1}\end{pmatrix} + \begin{pmatrix}\Delta u_i \\ u_i\end{pmatrix}$$

To obtain the SYS estimator, a full set of $m_S = (1/2)(T+1)(T-2)$ moment conditions and a $2(T-2) \times m_S$ block diagonal matrix, Z_i^S, are described by the following equations, respectively.

$$E(Z_i^{S'} u_i^s) = 0 \text{ where } u_i^s = (\Delta u'_i, u'_i)'$$
$$Z_i^S = \begin{pmatrix} Z_i^D & 0 \\ 0 & Z_i^L \end{pmatrix}$$

Using the matrix of instruments, $Z^S = \left(Z_1^{S'}, \dots, Z_N^{S'}\right)'$, the one-step SYS estimator (SYS_1) of φ is obtained as

$$\hat{\varphi}_{SYS_1} = \left(y_{-1}^{s'} Z^S W^S Z^{S'} y_{-1}^s\right)^{-1} y_{-1}^{s'} Z^S W^S Z^{S'} y^s$$

where $y^s = [(\Delta y'_1, y'_1), \dots, (\Delta y'_N, y'_N)]'$, $y_{-1}^s = \left[\left(\Delta y'_{1,-1}, y'_{1,-1}\right), \dots, \left(\Delta y'_{N,-1}, y'_{N,-1}\right)\right]'$, and $W^S = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^{S'} G Z_i^S\right)^{-1}$ with $G = \begin{pmatrix} D & 0 \\ 0 & I_{T-2}\end{pmatrix}$.

In the second step, using the residuals ($\hat{u}_i^s$) from SYS_1, a weighting matrix, W_2^S, and the two-step SYS estimator (SYS_2) are calculated respectively as follows

$$W_2^S = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^{S'} \hat{u}_i^s \hat{u}_i^{s'} Z_i^S\right)^{-1}$$
$$\hat{\varphi}_{SYS_2} = \left(y_{-1}^{s'} Z^S W_2^S Z^{S'} y_{-1}^s\right)^{-1} y_{-1}^{s'} Z^S W_2^S Z^{S'} y^s$$

Although the SYS estimator yields more efficient estimates compared to the LEV estimator, the bias of SYS estimator significantly increases in the presence of high persistence level of the autoregressive parameter and/or when the ratio of the variance of individual effects to that of the error term deviates from unity (cf. [2]).

In this study, we consider an alternative class of robust estimators, i.e., M-estimators, since the class of GMM estimators results in distorted parameter estimates in the presence of a high persistence level of the autoregressive parameter and/or when the ratio of the variance of the individual effects to the error term variance deviates from one. Next, we describe M-estimation approach weighted by power function to estimate first-order autoregressive panel model.

2.2. A Weighted-M Estimation Using Power Function for Dispersion Parameter

The main objective of this paper is to improve efficiency using the robust M-estimation approach for processes where the weak instrument problem arises. M-estimation approaches involve the minimization of a loss function that changes slowly in the presence of abnormal residuals. The novelty of our approach relies on obtaining weighted M-estimation using a power function for the scale parameter and choosing the value of the tuning parameter in the loss function by minimizing the variance of the estimators, as in [27, 28, 29].

The power function proposed by [32] for the scale parameter, σ, can be expressed in the first-order autoregressive panel data model as follows:

$$\sigma = \tau|\varphi y_{-1}|^{\gamma}$$

where τ is an unknown dispersion parameter with an unknown parameter vector γ, and $y_{-1} = \left(y'_{1,-1}, \dots, y'_{N,-1}\right)'$. For the first-order autoregressive panel data models, the loss function, $\rho(\cdot)$ is defined as

$$\sum_{i=1}^{N}\sum_{t=2}^{T} \rho\left(\frac{y_{it} - \varphi y_{i,t-1} - \alpha_i}{\hat{\sigma}}\right)$$

where $\hat{\sigma}$ denotes an estimated value of scale parameter. To obtain the robust M-estimates of the autoregressive parameter, φ, and the individual-specific effects, α_i's, the loss function, which varies slowly in the presence of outliers, is minimized. One of the most commonly employed loss function is the Huber's loss function described as follows

$$\rho(u) = \begin{cases} \dfrac{u^2}{2} & \text{if } |u| \leq c \\ c|u| - \dfrac{c^2}{2} & \text{if } |u| > c \end{cases} \quad \textbf{(4)}$$

where *c* is a tuning constant that regulates the level of robustness and is chosen within a range of values of 0 to 3. The default value of *c* in the R package *rlm* is set to 1.345 to achieve 95% asymptotic relative efficiency under the assumption of a normally distributed data.

Differentiating the loss function defined in Equation (4) results in the following estimating equations:

$$U(\varphi;\alpha_i) = \sum_{i=1}^{N}\sum_{t=2}^{T}\left(\frac{y_{i,t-1}}{\hat{\sigma}}\right)\psi\left(\frac{y_{it} - \varphi y_{i,t-1} - \alpha_i}{\hat{\sigma}}\right) = 0$$
$$U_\alpha(\varphi;\alpha_i) = \sum_{i=1}^{N}\sum_{t=2}^{T}\left(\frac{1}{\hat{\sigma}}\right)\psi\left(\frac{y_{it} - \varphi y_{i,t-1} - \alpha_i}{\hat{\sigma}}\right) = 0$$

where $\psi(u) = min\big(c, max(u, -c)\big)$ denotes the sub-gradient function of the Huber's loss function.

Let us consider the idiosyncratic error terms in the model, $\varepsilon_{it} = y_{it} - \varphi y_{i,t-1} - \alpha_i$. Then, the solutions of this estimating functions are obtained by rewriting $U(\varphi;\alpha_i)$ and $U_\alpha(\varphi;\alpha_i)$ as in the form of the weighted score functions as follows:

$$U(\varphi;\alpha_i) = \sum_{i=1}^{N}\sum_{t=2}^{T} y_{i,t-1} W_{it}\,\bar{e}_{it} = 0$$
$$U_\alpha(\varphi;\alpha_i) = \sum_{i=1}^{N}\sum_{t=2}^{T} W_{it}\,\bar{e}_{it} = 0$$

where $\bar{e}_{it} = \frac{(y_{it} - \varphi y_{i,t-1} - \alpha_i)}{\hat{\sigma}}$ and $W_{it} = \frac{\psi(\bar{e}_{it})}{\bar{e}_{it}}$ respectively denote the Pearson residuals and weights. Then, by solving the weighted score equations, the robust estimators of φ, $\hat{\varphi}_{POWER}$, and α_i, $\hat{\alpha}_i^{POWER}$, can be calculated as follows, respectively.

$$\hat{\varphi}_{POWER} = \left(\sum_{i=1}^{N}\sum_{t=2}^{T} y'_{i,t-1} W_{it} y_{i,t-1}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=2}^{T} y'_{i,t-1} W_{it} y_{it}\right) \quad \textbf{(5)}$$

$$\hat{\alpha}_i^{POWER} = \frac{\sum_{t=2}^{T}\big(y_{it} - \hat{\varphi}_{POWER} y_{i,t-1}\big)}{T-1} \quad \textbf{(6)}$$

Here, an iterative method is required since W_{it} depends on φ, α_i, and σ. This method is based on the pseudolikelihood approach and entails fixing the model parameter φ and variance parameters (τ and γ) alternatively (cf. [27]). To estimate γ, we use a robust estimator with a breakdown point of %50 defined as

$$\sum_{i=1}^{N}\sum_{t=2}^{T}\chi\left(\frac{y_{it} - \hat{\varphi} y_{i,t-1} - \hat{\alpha}_i}{\hat{\tau}\left|\hat{\varphi} y_{i,t-1}\right|^{\gamma}}\right)\frac{\left(\left|\hat{\varphi} y_{i,t-1}\right|^{\gamma}\right)'}{\left|\hat{\varphi} y_{i,t-1}\right|^{\gamma}} = 0 \quad \textbf{(7)}$$

where $\chi(u) = min\left(\frac{u^2}{1.041^2}, 1\right) - 0.5$ is a bounded function developed by [33, 34]. Also, to estimate the dispersion parameter τ, we employ the MAD estimator expressed as

$$\hat{\tau} = \text{Median}\left\{\frac{\left|y_{it} - \hat{\varphi} y_{i,t-1} - \hat{\alpha}_i\right|}{\left|\hat{\varphi} y_{i,t-1}\right|^{\hat{\gamma}}}\right\}/0.6745 \quad \textbf{(8)}$$

In robust approaches, it is crucial to appropriately select the tuning parameter, *c*, related to the loss function, as it controls the level of robustness in the estimation. In traditional robust methods, it is important to pre-specify the tuning constant *c* in any chosen loss function based on the desired level of robustness. When the errors follow a normal distribution, the optimal value for this parameter is $+\infty$. However, in heavy-tailed distributions, then *c* should be selected as a small positive value. The selection of the tuning parameter requires careful consideration as robustness often entails a sacrifice in efficiency. The value of the tuning parameter *c* should be selected based on the potential percentage of outliers in the data or the data distribution to maximize the asymptotic efficiency of the estimators. This is because the primary goal is to enhance the efficiency of the estimators while maintaining robustness.

In this study, following the approach outlined by [27, 29], the tuning parameter that minimizes the variance of the model parameters is referred to as the optimal one. Thus, it is suggested to iteratively estimate the tuning parameter *c* from the data for values ranging between 0 and 3 in Huber's loss function and select the value that minimizes the total estimated variances of the estimators in the model.

The algorithm below provides a summary of the entire estimating process.

1. Obtain initial M-estimates $\hat{\varphi}_0$ and $\hat{\alpha}_i^0 = \frac{\sum_{t=2}^{T}(y_{it} - \hat{\varphi}_0 y_{i,t-1})}{T-1}$ for $i = 1, \dots, N$ using the default value of *c* under the assumption of a constant variance $\left|\varphi y_{i,t-1}\right|^{\gamma} = 1$.
2. Calculate the residuals of this model by fixing $\hat{\varphi} = \hat{\varphi}_0$ and $\hat{\alpha}_i = \hat{\alpha}_i^0$ for $i = 1, \dots, N$ and obtain the robust estimates of the variance parameters, $(\hat{\gamma}, \hat{\tau})$, using Equations (7)-(8).
3. Update $\hat{\varphi}$ and $\hat{\alpha}_i$'s with the Equations (5)-(6), i.e., using weighted M-estimators, by setting the robust estimates of variance parameters, $(\hat{\gamma}, \hat{\tau})$, obtained in the previous step.

4. Repeat steps 2-3 till the optimal value of the tuning parameter, c, is obtained from a finite set of values ranging from 0 to 3. The best tuning parameter yields the smallest sum of the estimated variance of the model parameters.
5. Finally calculate the robust estimates of φ and α_i's using the best tuning constant.

III. NUMERICAL RESULTS

In this section, we carry out a comprehensive simulation study to investigate the performance of our proposed estimator (referred to as "POWER" in the graphs), and we compare our results with the one- and two-step GMM estimators: DIF, LEV and SYS. Note that the subscripts used in the abbreviations of the GMM estimators indicate the number of steps applied in the GMM estimation. For example, DIF_1 is used to show the performance of the one-step DIF estimator, while DIF_2 is associated with the two-step DIF estimator. To assess the finite-sample properties of each estimator, we consider three different scenarios: (*i*) different sample sizes, (*ii*) varying levels of individual heterogeneity, and (*iii*) varying levels of the persistence. For the data generation process (DGP), dynamic panel autoregressive model of order one given in Equation (1) is considered.

To assess the performance of the proposed and conventional GMM estimators using S = 1000 simulations, we calculate the bias and the root mean squared error (RMSE) of the estimator of autoregressive parameter, $\hat{\varphi}$, as follows:

$$\text{Bias} = \frac{1}{S}\sum_{s=1}^{S}(\hat{\varphi}^s - \varphi)$$

$$\text{RMSE} = \sqrt{\frac{1}{S}\sum_{s=1}^{S} Var(\hat{\varphi}^s) + Bias^2(\hat{\varphi}^s)}$$

where $\hat{\varphi}^s, s = 1, \dots, S$ are the estimates of the autoregressive parameter, φ, obtained from the simulated samples. Throughout the experiments, the initial observations are generated considering the initial conditions developed by [30] as $y_{i1} = \frac{\alpha_i}{1-\varphi} + \mu_{i1}$ where $\mu_{i1} \sim N(0,1)$, independent of α_i and ε_{it}.

3.1. Sample Sizes

To investigate the performance of the estimators under different sample sizes, the individual effects and the errors are generated from a normal distribution with mean zero and equal variances of $\sigma_\alpha^2 = 0.25$ and $\sigma_\varepsilon^2 = 0.25$, respectively. This implies that the ratio of the variance of the individual effect to that of the error term, $r = \frac{\sigma_\alpha^2}{\sigma_\varepsilon^2}$, kept fixed at $r = 1$. Thus, we exclude the examination of individual heterogeneity and focus solely on investigating the impact of sample sizes on the performance of the estimators in this subsection. To this end, we consider the increasing values of the cross-sectional dimension $N = 25, 50, 100, 200, 500$, with a fixed time period $T = 8$, and the increasing values of the time dimension $T = 4, 8, 12, 18, 24$ while keeping the number of cross-sectional units fixed at $N = 100$. The RMSE and bias of the estimators are compared when a moderate level of persistence is present with an autoregressive parameter $\varphi = 0.5$. Figure 1 illustrates the calculated RMSE and bias of both the conventional GMM estimators and the proposed weighted M-estimator with a power function as the cross-sectional dimension increases. Figure 2 displays the simulation results for the increasing time periods. In general, the RMSEs and bias of all the methods tend to decrease with increasing *N* and *T*, as expected. These figures demonstrates that the RMSEs calculated for the one- and two-step LEV estimators are largest relative to the other estimators, while the one- and two-step DIF estimators have the largest negative bias values in all cases. Also, although our proposed weighted M-estimator with a power function produces slightly increasing RMSEs and bias with increasing *T*, our proposed method (POWER) yields the lowest RMSE and bias values, particularly when considering small values of *T*. It is conspicuous from these figures that our proposed weighted M-estimator outperforms all the conventional GMM estimators in all situations.

3.2. Levels of Individual Heterogeneity

In this subsection, we focus on the impacts of different values of the variance ratio, $r = \frac{\sigma_\alpha^2}{\sigma_\varepsilon^2}$, which indicates the level of individual heterogeneity, on the performance of the estimators because conventional GMM estimators result in obtaining distorted estimates of the parameters as heterogeneity level increases. Thus, we investigate the robustness against increasing heterogeneity for our proposed estimator, following the experimental design used by [1]. To consider the scenarios where the variance ratio is $r < 1, r = 1$, and $r > 1$, we select four different pairs of σ_α^2 and σ_ε^2 as $(\sigma_\alpha^2, \sigma_\varepsilon^2) = (0.25,0.50)$, $(0.25,0.25), (2.5,0.5), (2.5,0.25)$ when $(N,T) = (100,8)$. We report the results for the selected values of autoregressive parameter $\varphi = 0.1, 0.3, 0.5, 0.7, 0.9$, considering a range from weak persistence to strong persistence.

The plots of the calculated RMSEs of the estimators versus the variance ratio, $r = 0.5, 1, 5, 10$, are given in Figure 3, whereas Figure 4 presents the bias values of the estimators versus *r*. Figure 3 illustrates that under weak and moderate levels of persistence, all the estimators, including our proposed estimator, tend to exhibit increasing RMSEs as the variance ratio *r* rises.

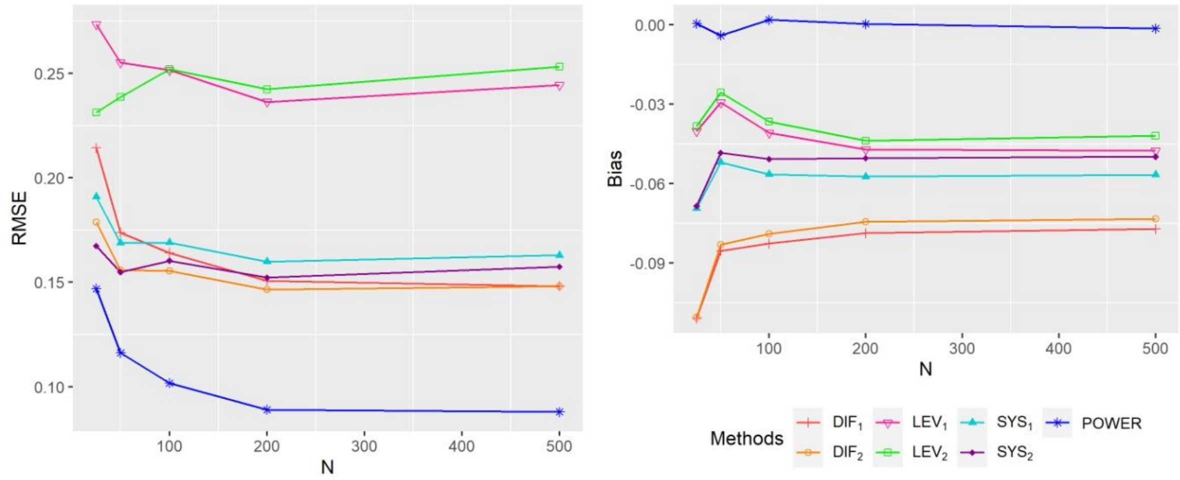

Figure 1: RMSE and Bias of conventional GMM estimators and the proposed weighted M-estimator with a power function when N = 25, 50, 100, 200, 500 with T = 8, r = 1, and $\varphi = 0.5$.

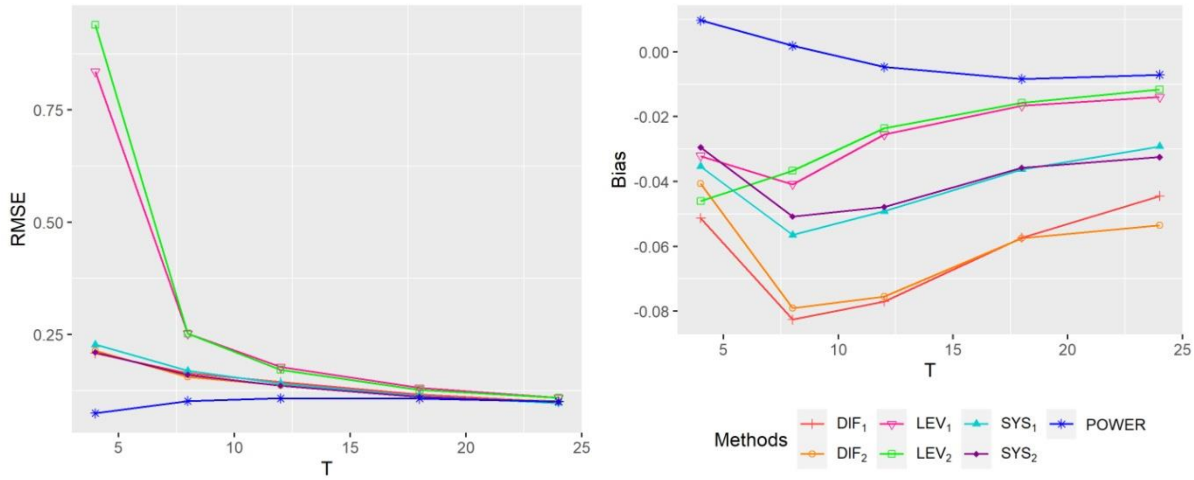

Figure 2: RMSE and Bias of conventional GMM estimators and the proposed weighted M-estimator with a power function when T = 4, 8, 12, 18, 24 with N = 100, r = 1 and $\varphi = 0.5$.

For weak and moderate levels of persistence, the LEV-GMM estimators have the highest RMSEs among the other estimators. On the other hand, when $\varphi = 0.9$, indicating strong persistence, all the methods except the DIF-GMM methods produce decreasing values of RMSEs as r increases, with the DIF-GMM estimators providing the largest RMSEs in that scenario. Our proposed weighted-M estimator with a power function, yields the smallest RMSE values as r increases for all values of the autoregressive parameter considered in this study. Figure 4 demonstrates that the DIF estimators have the largest bias values among all estimators and exhibit increasing bias as the level of heterogeneity increases. This indicates that the DIF estimators are more efficient than the LEV estimators at the weak and moderate levels of persistence when taking into account both performance metrics, i.e., RMSE and bias. The bias values obtained by the proposed POWER estimator are the smaller than those of the class of conventional GMM estimators for the weak and moderate level of persistence, even as the level of heterogeneity increases. At $\varphi = 0.9$, both LEV-GMM and our proposed estimator yield competitive results when bias values are evaluated. In general, our results demonstrate that that proposed POWER estimator outperforms conventional GMM estimators as the heterogeneity level increases across all persistence levels.

3.3. Level of Persistence

Following the same experimental design used in the previous subsection, we examine the finite sample performance of the estimators against increasing persistence level. To this end, we plot the RMSEs and bias values of the estimators versus the various levels of persistence in Figures 5-6. In Figure 5, both the proposed and conventional GMM estimators result in decreasing RMSE values as the persistence level increases across all heterogeneity levels considered. Under weak heterogeneity ($r = 0.5$ and $r = 1$), the one- and two-step LEV-GMM estimators have higher RMSEs than the other GMM estimators.

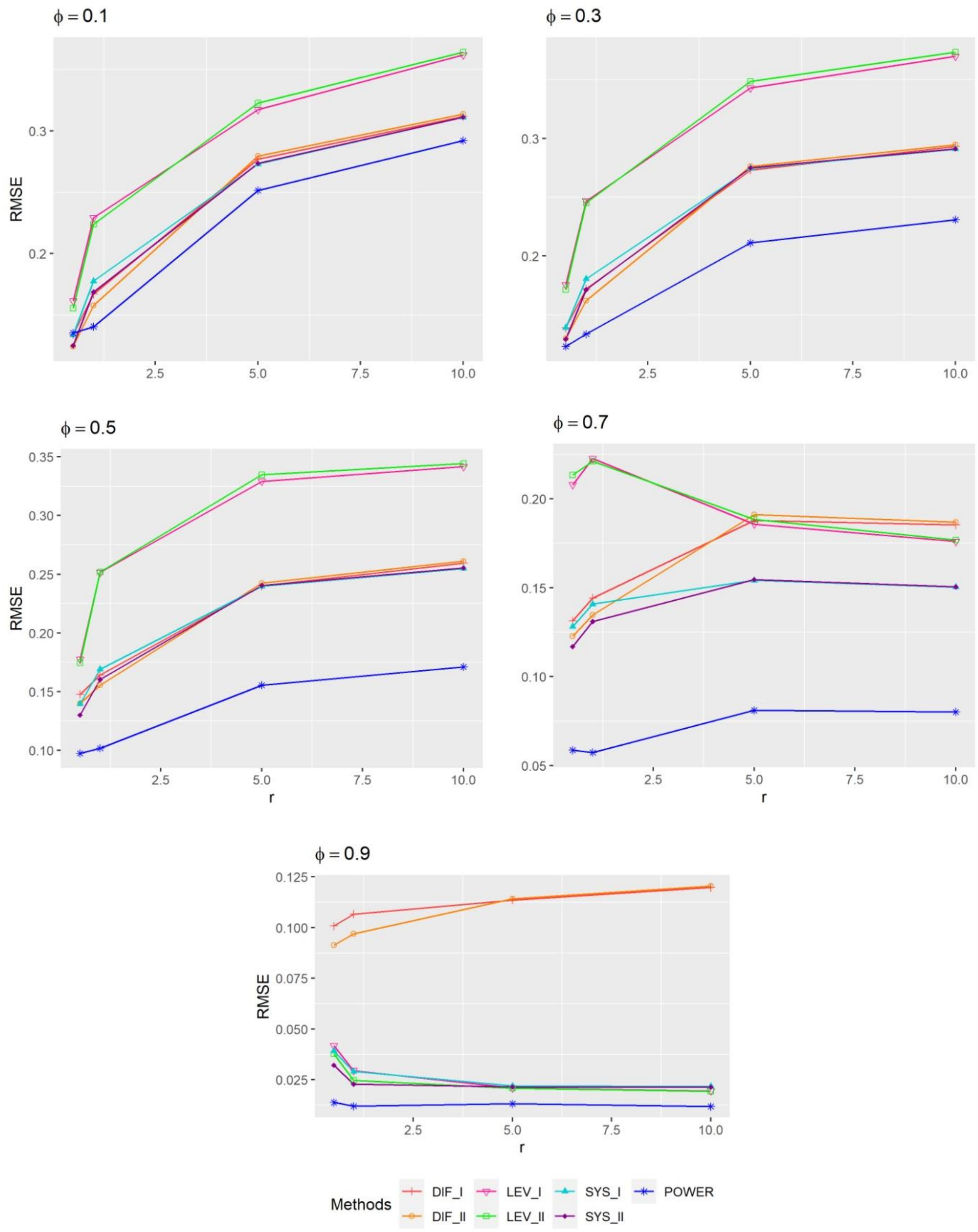

Figure 3: RMSE of conventional GMM estimators and the proposed weighted M-estimator with a power function for different values of $\varphi = 0.1, 0.3, 0.5, 0.7, 0.9$ when (N, T) = (100, 8) and r = 0.5, 1, 5, 10.

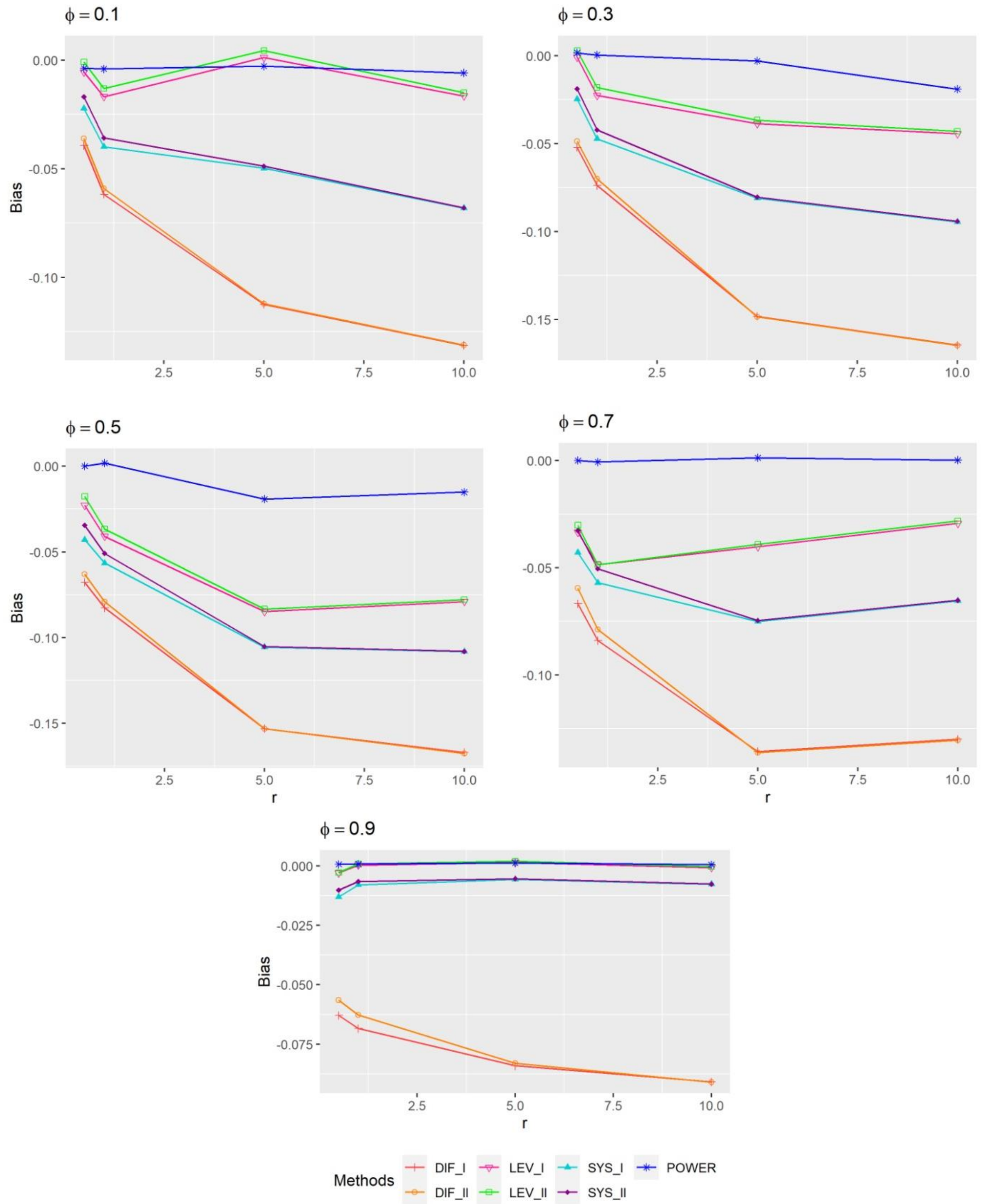

Figure 4: Bias of conventional GMM estimators and the proposed weighted M-estimator with a power function for different values of $\varphi = 0.1, 0.3, 0.5, 0.7, 0.9$ when $(N, T) = (100, 8)$ and $r = 0.5, 1, 5, 10$.

However, their performance tends to converge closely with that of the DIF-GMM and SYS-GMM estimators as the persistence level increases ($\varphi = 0.9$). Furthermore, as the heterogeneity level increases ($r = 5$ and $r = 10$), LEV-GMM estimators yield competitive results to SYS-GMM estimators for moderate and strong levels of persistence ($\varphi = 0.7$ and $\varphi = 0.9$). The POWER estimator, proposed in this study, demonstrates superior performance overall, except for the case $r = 0.5$ and $\varphi = 0.1$. This is despite the fact that the two-step SYS estimator achieves the lowest RMSEs among all GMM estimators, particularly with higher values of the autoregressive parameter. Based on Figure 6, it can be observed that the largest negative bias values are obtained by the one-step DIF-GMM estimator.

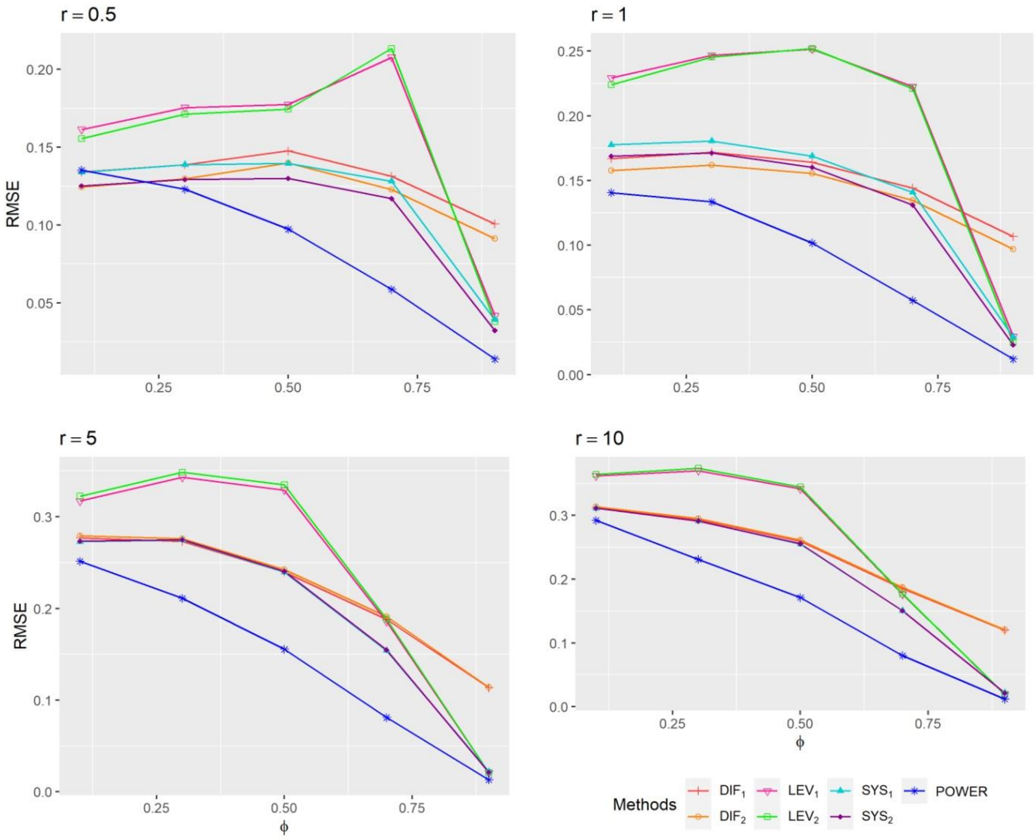

Figure 5: RMSE of conventional GMM estimators and the proposed weighted M-estimator with a power function for different levels of variance ratio r = 0.5, 1, 5, 10 when (N, T) = (100, 8).

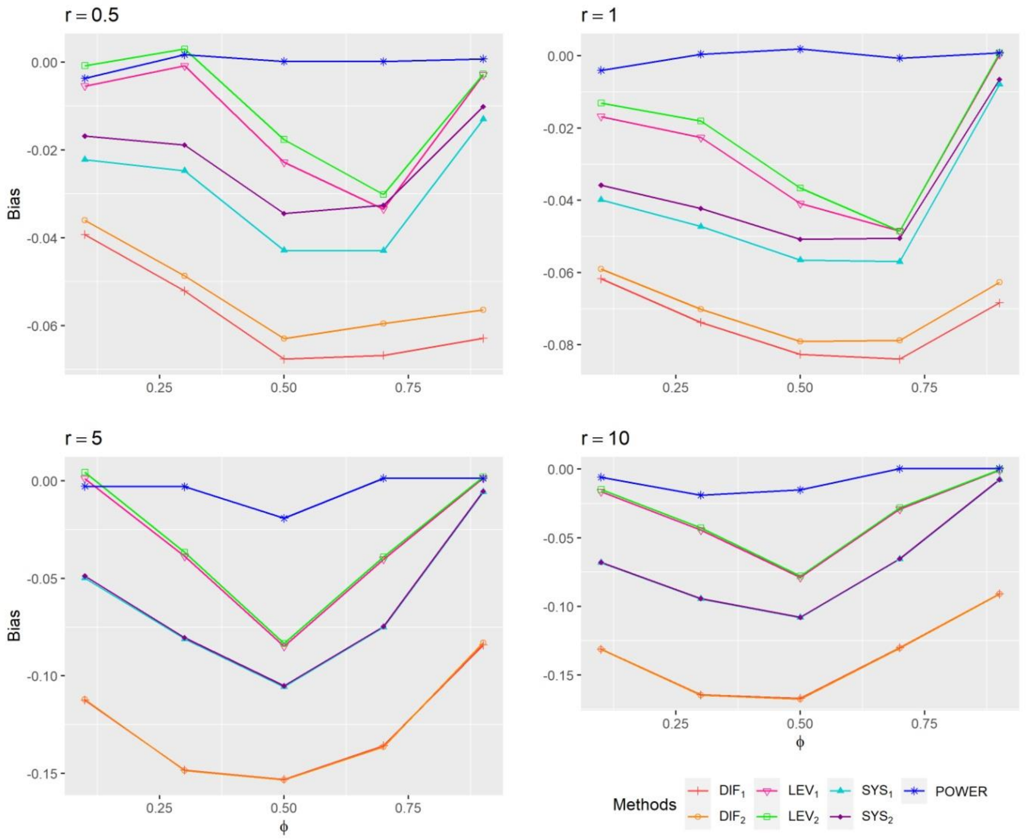

Figure 6: Bias of conventional GMM estimators and the proposed weighted M-estimator with a power function for different levels of variance ratio r = 0.5, 1, 5, 10 when (N, T) = (100, 8).

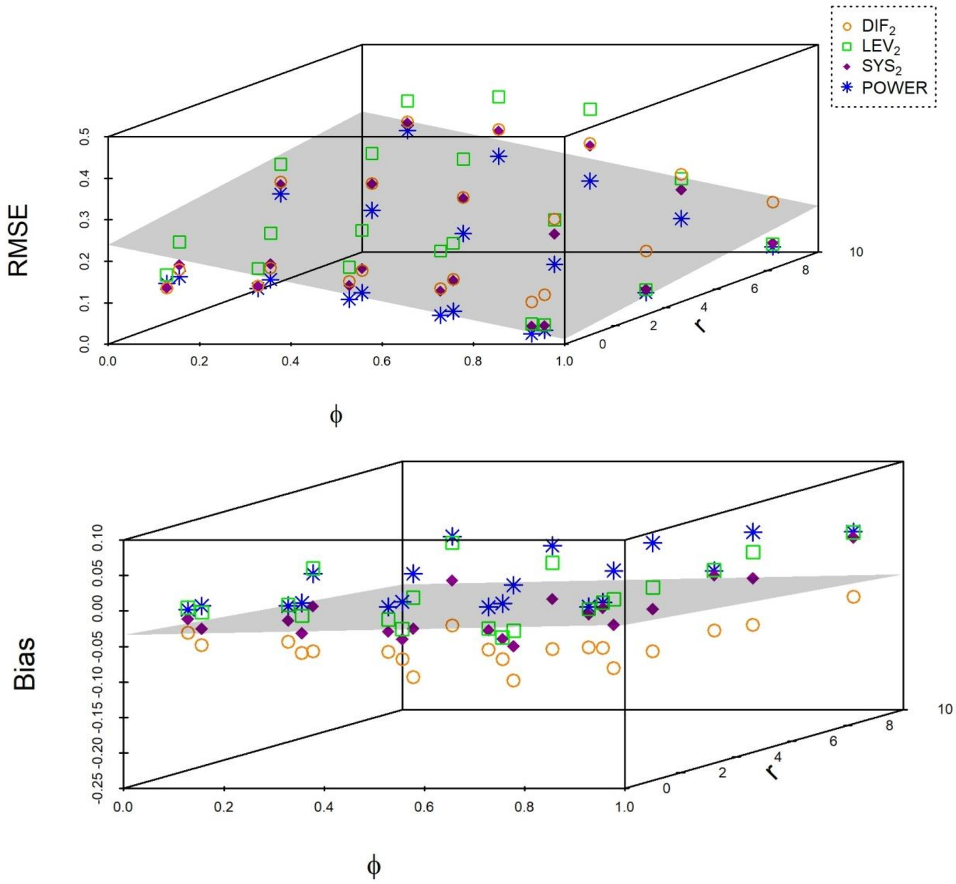

Figure 7: 3D-scatterplots of RMSE and Bias for conventional GMM estimators and the proposed weighted M-estimator with a power function across φ = 0.1, 0.3, 0.5, 0.7, 0.9 and r = 0.5, 1, 5, 10 when (N, T) = (100, 8).

For all GMM estimators, the absolute bias values increase until a moderate level of persistence. However, they tend to produce decreasing bias in absolute value when $\varphi = 0.7$ and $\varphi = 0.9$. Indeed, the autoregressive parameter value of 0.5 acts as a critical threshold for all methods. In all scenarios, our proposed POWER estimator demonstrates improved performance over conventional GMM estimators based on both performance metrics, except for cases of weak persistence and weak heterogeneity level ($\varphi = 0.1$ and $r = 0.5$).

The simulation results for different levels of heterogeneity and persistence are depicted by a three dimensional scatter plot (3D-scatter plot) in Figure 7. Figure 7 presents that the RMSE and bias of the estimators plotted as functions of the autoregressive parameter φ and variance ratio r. An increase in the variance ratio r have a more significant impact on RMSE compared to an increase in the autoregressive parameter φ. The larger RMSE values are obtained for higher values of r when the level of persistence is weak and moderate. The bias values of the GMM estimators are often negative, whereas the proposed estimator yields bias values that are closest to zero. In summary, our proposed POWER estimator is considerably less affected by increasing level of heterogeneity and/or persistence.

IV. CASE STUDY

In this section, we employ the proposed robust procedure to examine the monthly hydro-electric power output data (100 million kwh), available from the National Bureau of Statistics of China at https://data.stats.gov.cn/english/easyquery.htm. This dataset includes 220 observations (N = 22, T = 10), representing a cross section of 22 regions across China from March 2023 to December 2023. Table 1 presents the regions of China included in this study. Figure 8 displays the marginal distributions of the hydro-electric power output across 22 regions of China throughout different months. This figure suggests that there is heterogeneity in the hydro-electric power output among different regions of China, with some regions including outliers in their hydro-electric power output.

A first-order panel autoregressive model, given in Equation (1), is fitted to the data. For this data, y_{it} represents the hydro-electric power output, and the indices $i = 1, \dots, 22$ and $t = 1, \dots, 10$ denote the regions of China and months, respectively.

Table 1: List of regions of China

Beijing	Jilin	Fujian	Hainan	Qinghai
Hebei	Heilongjiang	Jiangxi	Chongqing	Ningxia
Shanxi	Jiangsu	Shandong	Tibet	
Inner Mongolia	Zhejiang	Henan	Shaanxi	
Liaoning	Anhui	Guangxi	Gansu	

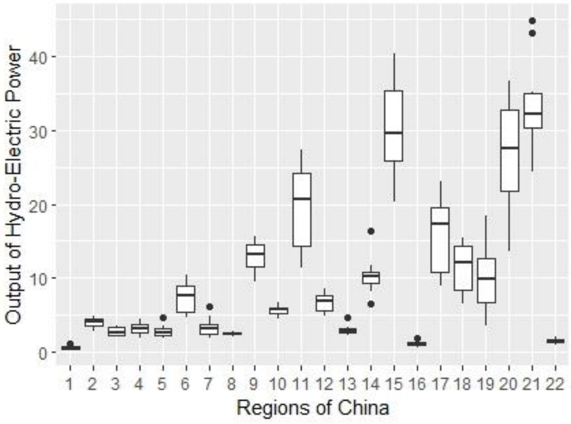

Figure 8: Boxplots of the hydro-electric power output across 22 regions of China

To assess the predictive performance of all estimators, we calculate the trimmed mean prediction errors due to the lack of prior knowledge about outliers in empirical data analysis. To this end, a specified percentage (e.g., p%) of observations is trimmed by considering those with the highest squared prediction errors, followed by computing the mean prediction error for the remaining portion of the dataset. The mean prediction error (MPE), excluding outliers, is calculated as defined below

$$\mathrm{MPE} = \frac{1}{22 \times 10 - m} \sum_{i=1}^{22} \sum_{t=1}^{10} (1 - \mathrm{I}_{it})(y_{it} - \hat{y}_{it})^2$$

where y_{it} and $\hat{y}_{it}$ respectively denote the observed and predicted values of the hydro-electric power output, and $m = \frac{p}{100} \times 22 \times 10$ represents the number of outliers in the dataset. Here, I_{it} represents an indicator variable that takes the value 1 if y_{it} is an outlying observation, and 0 otherwise. The estimates of the autoregressive parameter and trimmed MPEs are reported in Table 2.

Table 2: The estimates of autoregressive parameter φ obtained by the conventional GMM estimators and the proposed weighted M-estimator with a power function, along with MPEs for various trimming percentages, p.

Estimator	**Full Data**		**MPE p = 5%**	**MPE p = 10%**	**MPE p = 15%**	**MPE p = 20%**
	$\hat{\varphi}$	**MPE p = 0%**				
POWER	0.9946	10.7786	5.2866	2.9803	1.8708	1.2068
DIF_1	0.4749	61.9876	38.7937	26.0449	17.6564	12.5514
DIF_2	0.4738	62.2326	38.9658	26.1661	17.7376	12.6114
LEV_1	0.9535	10.8289	5.5147	3.0137	1.9170	1.2982
LEV_2	0.9535	10.8296	5.5145	3.0136	1.9169	1.2982
SYS_1	0.9357	11.0653	5.7130	3.1685	1.9643	1.3014
SYS_2	0.9356	11.0666	5.7141	3.1694	1.9645	1.3015

We consider trimming percentages as $p = 5\%, 10\%, 15\%,$ and 20%, with the value zero representing the full dataset. From Table 2, it is evident that the estimated value of the autoregressive parameter to be greater than 0.9, indicating strong persistence within the dataset. The strong persistence is evident since the DIF-GMM estimators result in the largest MPEs with estimated values of the parameter is considerably high, approaching unity, for the proposed POWER estimator. Additionally, the GMM estimators, with the exception of the DIF-GMM estimators, have estimated the autoregressive autoregressive parameter near the moderate level of 0.5. As the trimming percentages increase, all estimators exhibit improved predictive performance with decreasing MPEs. The proposed POWER estimator produces the lowest MPE among all estimators considered across all trimming percentages. Although the smallest MPEs are provided by the LEV-GMM estimators among the GMM estimators, our proposed POWER estimator yields better predictions, ranging from 0.47% to 7.57% across various trimming percentages, compared to the LEV-GMM estimators. Moreover, the proposed estimator produces more accurate predictions, ranging from 2.65% to 8.08, compared to SYS-GMM estimators. All empirical findings clearly indicate that the proposed POWER estimator is robust in the presence of strong persistence and heterogeneity, demonstrating substantially improved predictive ability compared to conventional GMM methods.

V. CONCLUSION

In this study, we propose a robust weighted M-estimation approach for first-order dynamic panel autoregressive model, which considerably enhances the precision and robustness of parameter estimates under challenging conditions of high persistence and individual heterogeneity. The proposed estimator is constructed by weighting the M-estimator with a power function for the scale parameter used in the Huber’s loss function, with optimization of the tuning parameter by minimizing the variance of the model parameters. Through extensive simulations and empirical data analysis, we have demonstrated that our method outperforms traditional GMM estimators, particularly in the presence of weak instruments and near-unity autoregressive parameters. This robust estimator provides a valuable tool for researchers across various fields, including econometrics, biology, and engineering, facilitating more accurate analysis of dynamic panel data.

REFERENCES

[1] Youssef, A. H. & Abonazel, M. R. (2017). Alternative GMM estimators for first-order autoregressive panel model: An improving efficiency approach. *Communications in Statistics-Simulation and Computation*, 46(4), 3112-3128.

[2] Bun, M. J. G. & Kiviet, J. F. (2006). The effects of dynamic feedbacks on LS and MM estimator accuracy in panel data models. *Journal of Econometrics*, 132(2), 409-444.

[3] Neyman, J. & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16(1), 1-32.

[4] Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49(6), 1417-1426.

[5] Lancaster, T. (2002). Orthogonal parameters and panel data. *Review of Economic Studies*, 69(3), 647-666.

[6] Hsiao, C., Pesaran, M. H. & Tahmiscioglu, A. K. (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics*, 109(1), 107-150.

[7] Binder, M., Hsiao, C. & Pesaran, M. H. (2005). Estimation and inference in short panel vector autoregressions with unit roots and cointegration. *Econometric Theory*, 21(4), 795-837.

[8] Hayakawa, K. & Pesaran, M. H. (2015). Robust standard errors in transformed likelihood estimation of dynamic panel data models with cross-sectional heteroskedasticity. *Journal of Econometrics*, 188(1), 111-134.

[9] Bun M. J. G., Carree, M. A. & Juodis, A. (2017). On Maximum likelihood estimation of dynamic panel data models. *Oxford Bulletin of Economics and Statistics*, 79(4), 463-494.

[10] Alvarez, J. & Arellano, M. (2022). Robust likelihood estimation of dynamic panel data models. *Journal of Econometrics*, 226(1), 21-61.

[11] Hansen, L. P. (2010). Generalized method of moments estimation. In: *Macroeconometrics and Time Series Analysis*, Durlauf, S. N., Blume, L. E. (eds), Palgrave Macmillan, London, p. 105-118.

[12] Cochrane, J. H. (2001). Asset pricing. Princeton University Press, Princeton, New Jersey.

[13] Arellano, M. (2003). Panel data econometrics. Oxford University Press, New York.

[14] Hall, A. R. (2005). Generalized method of moments. Oxford University Press, New York

[15] Singleton, K. J. (2006). Empirical dynamic asset pricing: Model specification and econometric assessment. Princeton University Press, Princeton, New Jersey.

[16] Anderson, T. W. & Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, 18(1), 47-82.

[17] Holtz-Eakin, D., Newey, W. & Rosen, H. S. (1988). Estimating vector autoregressions with panel data. *Econometrica*, 56(6), 1371-1395.

[18] Arellano, M. & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, 58(2), 277-297.

[19] Blundell, R. & Bond, S. (1998). Initial conditions andmoment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1), 115-143.

[20] Alonso-Borrego, C. & Arellano, M. (1999). Symmetrically normalized instrumental variable estimation using panel data. *Journal of Business & Economic Statistics*, 17(1), 36-49.

[21] Han, C., Phillips, P. C. B. & Sul, D. (2014). X-differencing and dynamic panel model estimation. *Econometric Theory*, 30(1), 201-251.

[22] Arellano, M. & Bover, O. (1995). Another look at the instrumental variable estimation of error components models. *Journal of Econometrics*, 68(1), 29-51.

[23] Ahn, S. C. & Schmidt, P. (1995). Efficient estimation of models for dynamic panel data. *Journal of Econometrics*, 68(1), 5-27.

[24] Bun, M. J. G. & Windmeijer, F. (2010). The weak instrument problem of the system GMM estimator in dynamic panel data models. *Econometrics Journal*, 13(1), 95-126.

[25] Hayakawa, K. (2007). Small sample bias properties of the system GMM estimator in dynamic panel data models. *Economics Letters*, 95(1), 32-38.

[26] Han, C. & Philips, P. (2010). GMM estimation for dynamic panels with fixed effects and strong instruments at unity. *Econometric Theory*, 26(1), 119-151.

[27] Callens, A., Wang, Y. G., Fu, L. & Liquet, B. (2021). Robust estimation procedure for autoregressive models with heterogeneity. *Environmental Modeling & Assessment*, 26, 313-323.

[28] Wang, Y. G., Lin, X., Zhu, M. & Bai, Z. (2007). Robust estimation using the Huber function with a data-dependent tuning constant. *Journal of Computational and Graphical Statistics*, 16(2), 468-481.

[29] Wang, N., Wang, Y. G., Hu, S., Hu, Z. H., Xu, J., Tang, H. & Jin, G. (2018). Robust regression with data-dependent regularization parameters and autoregressive temporal correlations. *Environmental Modeling & Assessment*, 23, 779-786.

[30] Alvarez, J. & Arellano, M. (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica*, 71(4), 1121-1159.

[31] Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics*, 126(1), 25-51.

[32] Box, G. E. P. & Hill, W. J. (1974). Correcting inhomogeneity of variance with power transformation weighting. *Technometrics*, 16(3), 385-389.

[33] Croux, C. (1994). Efficient high-breakdown M-estimators of scale. *Statistics and Probability Letters*, 19(5), 371–379.

[34] Bianco, A., Boente, G. & Di Rienzo, J. (2000). Some results for robust GM-based estimators in heteroscedastic regression models. *Journal of Statistical Planning and Inference*, 89(1-2), 215-242.