Transformed Estimation for Panel Interactive Effects Models*

Cheng Hsiao† Zhentao Shi‡ Qiankun Zhou§

This version October 31, 2019

Abstract

We propose a transformed quasi-maximum likelihood estimation (QMLE) for panel models with interactive effects. The transformed estimator doesn't need to estimate the interactive effects in the model and is $\sqrt{NT}$-consistent, and is asymptotically normally distributed centered at the true value whether the regressors are exogenous or contain predetermined variables. It is computational simple, and only requires either N or T to go infinity for the large sample results to hold. The finite sample performance of the transformed QMLE is examined through extensive simulations, and we find the transformed estimator works remarkably well in our designs, regardless of whether the model is static or dynamic, whether the common factors are stationary, cointegrated or could be subject to structure change, and whether the idiosyncratic errors are homoskedastic or heteroskedastic or weakly cross-sectionally dependent.

Keywords: Panel interactive effects models, Transformed quasi-maximum likelihood estimation, Eigenvector, Eigenvalue
JEL classification: C01, C13, C23

*We would like to thank Whitney Newey and Martin Weidner for helpful discussion. Partial research support by China NSF #71631004 to the first author is also gratefully acknowledged.

†Department of Economics, University of Southern California, University Park, Los Angeles, California 90089, Email: chsiao@usc.edu, Department of Quantitative Finance, NTHU and WISE, Xiamen University, China.

‡Department of Economics, the Chinese University of Hong Kong, Sha Tin, Hong Kong SAR, China. Email: zhentao.shi@cuhk.edu.hk.

§Department of Economics, Louisiana State University, Baton Rouge, LA 70803. Email: qzhou@lsu.edu.

1 Introduction

One of the challenges for panel data modeling is to model the unobserved heterogeneity across individuals, i, and over time, t. The factor approach has been popular by assuming the unobserved individual-time varying effects, γ_{it}, as the product of r time-specific effects that are common across individuals, $\mathbf{f}_t$, and the individual-specific effects but time-invariant effects, $\boldsymbol{\lambda}_i$, i.e., $\gamma_{it} = \boldsymbol{\lambda}_i'\mathbf{f}_t$. The multiplicative form nests the traditional approach of putting the unobserved individual- and time-specific effects, γ_{it}, in additive form (e.g., Hsiao (2014)) as special cases (e.g., Bai (2009) and Hsiao (2018)). However, there does not exist a simple linear transformation to get rid of the interactive (or multiplicative) effects. The estimation of panel models with interactive effects becomes much more complicated (e.g., Ahn et al. (2001, 2013), Bai (2009), Pesaran (2006)). The consistency and asymptotic distribution of the common structural (slope) coefficients often require both the cross-sectional dimension, N, and time series dimension, T, to go to infinity. Moreover, if the explanatory variables contain predetermined variables, then some of the proposed estimators could be inconsistent or asymptotically biased (e.g., Hsiao (2018), Moon and Weidner (2015, 2017)).

In this paper, we suggest a transformed estimator for the common slope estimator without the need to estimate the interactive effect ($\boldsymbol{\lambda}_i'\mathbf{f}_t$). The consistency and asymptotic distribution of our estimator only requires either N or T to go to infinity. Moreover, there is no asymptotic bias even the explanatory variables contain predetermined variables.

The rest of the paper is organized as follows. The model is presented in Second 2. Section 3 proposes a transformed quasi-maximum likelihood estimator (TMQLE) and derives its asymptotic properties. Section 4 suggests an average estimator to further improve the efficiency of the TQMLE. Section 5 provides some Monte Carlo studies to demonstrate the desirability of our transformed estimator. Concluding remarks are in Section 6. The derivation of the asymptotic properties of our estimators and a more detailed report of the simulation results are provided in Appendix A and B.

2 The Model

We consider the panel interactive model of the form

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + v_{it}, \tag{2.1}$$

$$v_{it} = \boldsymbol{\lambda}_i'\mathbf{f}_t + u_{it}, i = 1, \ldots, N; t = 1, \ldots, T, \tag{2.2}$$

where $\mathbf{x}_{it}$ denotes a $k \times 1$ vector of observed variables, $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown constants, $\boldsymbol{\lambda}_i$ is an $r \times 1$ vector of unobserved individual-specific effects that stay constant over time, and

$\mathbf{f}_t$ is an $r \times 1$ vector of unobserved time-specific constants that stay the same across N but vary over time. The number of factors, r, is unknown to researchers.

Let $\mathbf{y}_t = (y_{1t}, \ldots, y_{Nt})'$, $\mathbf{X}_t = (\mathbf{x}_{1t}, \ldots, \mathbf{x}_{Nt})'$, $\mathbf{u}_t = (u_{1t}, \ldots, u_{Nt})'$, $\Lambda = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N)'$ and $\mathbf{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_T)'$, model (2.1) and (2.2) can be written in the form

$$\mathbf{y}_t = \mathbf{X}_t \boldsymbol{\beta} + \Lambda \mathbf{f}_t + \mathbf{u}_t, \quad t = 1, \ldots, T. \tag{2.3}$$

For model (2.3), we assume:

Assumption A1: $\mathbf{u}_t$ is independently, identically distributed over t with $E(\mathbf{u}_t | \mathbf{X}_t, \Lambda, \mathbf{f}_t) = 0$ and $E(\mathbf{u}_t \mathbf{u}_t' | \mathbf{X}_t, \Lambda, \mathbf{f}_t) = \Omega$, where Ω is nonsingular and the eigenvalues of Ω are $O(1)$. We also assume u_{it} has bounded fourth moment.

Assumption A2: $rank(\Lambda) = r$ and $\frac{1}{N}\Lambda'\Lambda = \mathbf{I}_r$.

Assumption A3: $rank(\mathbf{F}) = r$ and $plim_{T\to\infty} \frac{1}{T}\mathbf{F}'\mathbf{F} = \mathbf{D}$, and $\mathbf{D}$ is a nonsingular diagonal matrix.

Assumption A4: The sequences $\mathbf{x}_{it}$ conditional on $\mathcal{F}^{t-1}$ are fixed constants and the matrices $E(\mathbf{X}_t'\mathbf{X}_t)$ over i and $E(\mathbf{X}_i'\mathbf{X}_i)$ over t are of full rank k, where $\mathcal{F}^{t-1}$ denotes the information available up to time period $t-1$, $\mathbf{X}_t = (\mathbf{x}_{1t}, \ldots, \mathbf{x}_{Nt})'$ and $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})'$. We also assume $E\left(\|\mathbf{x}_{it}\|^4\right) < C < \infty$ for all i, t.

Assumption A1 allows $E(u_{it}u_{jt}) \neq 0$ for some i and j. However, it also restricts $\mathbf{u}_t$ to be weakly cross-sectionally correlated (e.g., Chudik et al. (2011)). Neither does it allow the variance of any component of $\mathbf{u}_t$, say u_{it}, to become dominant relative to any other u_{jt}.[1] Assumption A1 also allows $\mathbf{x}_{it}$ to contain predetermined variables such as lagged dependent variables. Assumptions A2 and A3 are the standard assumptions for a unique decomposition of $\boldsymbol{\lambda}_i'\mathbf{f}_t$ in the factor model (e.g., Anderson and Rubin (1956), Bai (2009)). Since it is the product $\boldsymbol{\lambda}_i'\mathbf{f}_t$ that affects y_{it}, not the individual component of $\boldsymbol{\lambda}_i'\mathbf{f}_t$, we shall assume Λ and $\mathbf{F}$ satisfying Assumption A2 and A3 as the true values. Assumption A4 is equivalent to the conventional full rank assumption of the explanatory variables, e.g., the matrices $\frac{1}{N}\sum_{i=1}^N \mathbf{x}_{it}\mathbf{x}_{it}'$ and $\frac{1}{T}\sum_{t=1}^T \mathbf{x}_{it}\mathbf{x}_{it}'$ are full rank. It ensures that there does not exist a nonzero $N \times 1$ constant vector $\boldsymbol{\alpha}$ such that $E(\boldsymbol{\alpha}'\mathbf{X}_t) = 0$, or a nonzero $T \times 1$ constant vector $\boldsymbol{\alpha}^*$ such that $E(\boldsymbol{\alpha}^{*\prime}\mathbf{X}_i) = 0$.[2]

[1] Assumption A1 can be considered as a restrictive version of Ahn and Horenstein's (2013) Assumption C with $\mathbf{R}_T = \mathbf{I}_T$, i.e., no serial correlation is allowed.

[2] In the event $\mathbf{X}_i$ are identical to $\mathbf{X}_j$ when $i \neq j$, we can pool the observations of $(\mathbf{y}_i, \mathbf{X}_i)$ and $(\mathbf{y}_j, \mathbf{X}_j)$ together to form a new unit $(\mathbf{y}_i^*, \mathbf{X}_i)$ to satisfy the condition of A4, where $\mathbf{y}_i^* = \frac{1}{2}(\mathbf{y}_i + \mathbf{y}_j)$. The resulting reformulated model does not affect our transformed estimator under A1 that allows heteroskedasticity and weak cross-sectional dependence.

3 The Transformed Quasi-Maximum Likelihood Estimator (TQMLE)

Under the assumption that $\boldsymbol{\lambda}_i$ and $\mathbf{f}_t$ are fixed constants, the quasi-log-likelihood function of model (2.3) takes the form of

$$l = -\frac{T}{2}\log|\Omega| - \frac{1}{2}\sum_{t=1}^{T}\left(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta} - \Lambda\mathbf{f}_t\right)' \Omega^{-1}\left(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta} - \Lambda\mathbf{f}_t\right)'. \tag{3.1}$$

Under the assumption that $\Omega = \sigma_u^2\mathbf{I}_N$, Bai (2009) proposes to estimate $(\boldsymbol{\beta},\Lambda,\mathbf{F})$ by the least squares method and shows that the least squares estimator of $\boldsymbol{\beta}$ is consistent and asymptotically normally distributed when both N and T go to infinity (denoted by $(N,T)\to\infty$). However, the least squares estimator could be asymptotically biased if $\mathbf{x}_{it}$ contains predetermined variables (Moon and Weidner (2017)).

For model (2.3), we note that when $N>r$, there exists an $N\times 1$ vector $\mathbf{w}$ that lies in the null space of Λ such that

$$\mathbf{w}'\Lambda = 0. \tag{3.2}$$

As a result, multiplying both sides of (2.3) by such a $\mathbf{w}'$ yields

$$\mathbf{w}'\mathbf{y}_t = \mathbf{w}'\mathbf{X}_t\boldsymbol{\beta} + \mathbf{w}'\mathbf{u}_t, \quad t = 1,\ldots,T. \tag{3.3}$$

Proposition 3.1 *Conditional on* $\mathbf{w}'\Lambda = 0$, *under Assumption A1-A4, and as* $T\to\infty$, *the least squares regression of (3.3),*

$$\hat{\boldsymbol{\beta}} = \left(\sum_{t=1}^{T}\mathbf{X}_t'\mathbf{w}\mathbf{w}'\mathbf{X}_t\right)^{-1}\left(\sum_{t=1}^{T}\mathbf{X}_t'\mathbf{w}\mathbf{w}'\mathbf{y}_t\right), \tag{3.4}$$

is consistent and

$$\sqrt{T}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \to_d N\left(0, \sigma_w^2\Xi_w^{-1}\right), \tag{3.5}$$

where $\sigma_w^2 = E\left(\mathbf{w}'\mathbf{u}_t\mathbf{u}_t'\mathbf{w}\right) = \mathbf{w}'\Omega\mathbf{w}$ *and* $\Xi_w = plim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t'\mathbf{w}\mathbf{w}'\mathbf{X}_t$.

However, Λ is unknown, hence $\mathbf{w}$ is unknown. To find $\mathbf{w}$ and $\boldsymbol{\beta}$, we propose to maximize the transformed quasi-log-likelihood function

$$L = -\frac{T}{2}\log\sigma_w^2 - \frac{1}{2\sigma_w^2}\sum_{t=1}^{T}\mathbf{w}'\left(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}\right)\left(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}\right)'\mathbf{w}, \tag{3.6}$$

subject to

$$\mathbf{w}'\Omega\mathbf{w} = 1, \tag{3.7}$$

where $\sigma_w^2 = E\left[\left(\mathbf{w}'\mathbf{u}_t\right)^2\right] = 1$.

Proposition 3.2 *Under Assumption A1-A4, when $N > r$, the TQMLE of $\mathbf{w}$ and $\boldsymbol{\beta}$ is equivalent to finding the $\hat{\mathbf{w}}$, which is the eigenvector corresponding to the smallest root of the $N \times N$ matrix*

$$\Omega^{-1/2}\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{y}_t - \mathbf{X}_t\hat{\boldsymbol{\beta}}\right)\left(\mathbf{y}_t - \mathbf{X}_t\hat{\boldsymbol{\beta}}\right)'\Omega^{-1/2}, \tag{3.8}$$

and $\hat{\boldsymbol{\beta}}$ that satisfies (3.4).

We note maximizing (3.6) is equivalent to finding $\hat{\mathbf{w}}$ and $\hat{\boldsymbol{\beta}}$ that minimizes

$$S(\mathbf{w}, \boldsymbol{\beta}) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{w}'(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta})(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta})'\mathbf{w}, \tag{3.9}$$

subject to (3.7). Although (3.9) does not involve $(\Lambda, \mathbf{F})$, the value of the objective function depends on $(\Lambda, \mathbf{F})$. The consistency of the TQMLE follows from the conventional proof of MLE and is provided below.

Proposition 3.3 *Under Assumption A1-A4, when $N > r$, $T \to \infty$ and $\frac{N}{T} \leq 1$, minimizing (3.9) subject to (3.7) yields $\hat{\mathbf{w}}'\Lambda \to_p 0$ and $\hat{\boldsymbol{\beta}} \to_p \boldsymbol{\beta}$. Moreover, $\hat{\mathbf{w}}$ and $\hat{\boldsymbol{\beta}}$ are asymptotically independent.*

Consequently, conditional on $\hat{\mathbf{w}}$ that satisfies $\hat{\mathbf{w}}'\Lambda = 0$, the estimator of $\boldsymbol{\beta}$ which solves (3.9) is asymptotically normally distributed as below by following the results of Proposition 3.1,

$$\left(\sum_{t=1}^{T}\mathbf{X}_t'\hat{\mathbf{w}}\hat{\mathbf{w}}'\mathbf{X}_t\right)^{1/2}\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to_d N(0, \mathbf{I}_k). \tag{3.10}$$

Since $N > r$, there could be multiple $\mathbf{w}$ that satisfies $\mathbf{w}'\Lambda = 0$. However, under Assumption A4, $\boldsymbol{\beta}$ remains uniquely determined (or locally identified) for any $\mathbf{w}$ that satisfies $\mathbf{w}'\Lambda = 0$.

Remark 3.1 *Under Assumptions A1-A4, when there are multiple $\mathbf{w}$ that maximize (3.6) subject to (3.7), the unconditional TQMLE of $\boldsymbol{\beta}$ and $\mathbf{w}$ lead to $\hat{\boldsymbol{\beta}}$, which is consistent and is asymptotically mixed normally distributed*

$$\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to_d \int_{\mathbf{w}'\Omega\mathbf{w}=1, \mathbf{w}'\Lambda=0} N\left(0, \Xi_w^{-1}\right) d\mathfrak{D}_w, \tag{3.11}$$

as $T \to \infty$, where $\mathfrak{D}_w := \lim_{T\to\infty} \underset{\mathbf{w}'\Omega\mathbf{w}=1, \mathbf{w}\in\mathcal{W}_\Lambda}{\arg\min} \mathbf{w}'\left(T^{-1}\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta})(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta})'\right)\mathbf{w}$ and $\mathcal{W}_\Lambda$ denoting the null space of Λ. Although $\hat{\boldsymbol{\beta}}$ follows an asymptotic mixed normal distribution, the standard t-statistics involving $\hat{\boldsymbol{\beta}}$ remains asymptotically normal (Jeganathan (1980, 1982)). Therefore, statistical inference of $\boldsymbol{\beta}$ follows the same standard procedure.

When $\Omega = \sigma_u^2 \mathbf{I}_N$ (i.e., u_{it} is i.i.d over i and t with constant variance), the TQMLE of $\boldsymbol{\beta}$ and $\mathbf{w}$ is equivalent to finding $\boldsymbol{\beta}$ and $\mathbf{w}$ that minimizes

$$\mathbf{w}' \frac{1}{T} \sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}) (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta})' \mathbf{w}, \tag{3.12}$$

subject to $\mathbf{w}'\mathbf{w} = 1$.

The first order conditions for minimizing (3.12) subject to $\mathbf{w}'\mathbf{w} = 1$ are

$$\bar{\boldsymbol{\beta}} = \left(\sum_{t=1}^{T} \mathbf{X}_t' \bar{\mathbf{w}} \bar{\mathbf{w}}' \mathbf{X}_t \right)^{-1} \left(\sum_{t=1}^{T} \mathbf{X}_t' \bar{\mathbf{w}} \bar{\mathbf{w}}' \mathbf{y}_t \right), \tag{3.13}$$

where $\bar{\mathbf{w}}$ is the eigenvector corresponding to the smallest root of the determinantal equation

$$\left| \frac{1}{T} \sum_{t=1}^{T} \left(\mathbf{y}_t - \mathbf{X}_t \bar{\boldsymbol{\beta}} \right) \left(\mathbf{y}_t - \mathbf{X}_t \bar{\boldsymbol{\beta}} \right)' - \delta \mathbf{I}_N \right| = 0. \tag{3.14}$$

This suggests that when $\frac{N}{T} \leq 1$, $\bar{\mathbf{w}}$ and $\bar{\boldsymbol{\beta}}$ can be obtained by iterating between (3.13) and (3.14) from some initial estimator $\hat{\boldsymbol{\beta}}^{(0)}$ until the solution converges.

However if $\Omega \neq \sigma_u^2 \mathbf{I}_N$, then it is possible that

$$\frac{1}{T} \sum_{t=1}^{T} \bar{\mathbf{w}}' \left[\mathbf{X}_t \left(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}} \right) + \Lambda \mathbf{f}_t \right] \left[\mathbf{X}_t \left(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}} \right) + \Lambda \mathbf{f}_t \right]' \bar{\mathbf{w}} + \frac{1}{T} \sum_{t=1}^{T} \bar{\mathbf{w}}' \mathbf{u}_t \mathbf{u}_t' \bar{\mathbf{w}} \leq \frac{1}{T} \sum_{t=1}^{T} \hat{\mathbf{w}}' \mathbf{u}_t \mathbf{u}_t' \hat{\mathbf{w}}, \tag{3.15}$$

while $\boldsymbol{\beta} \neq \bar{\boldsymbol{\beta}}$, $\bar{\mathbf{w}}' \Lambda \neq 0$, and $\hat{\mathbf{w}}' \Lambda = 0$. Although we expect that the chance of such a pathetic case to arise for a given sequence of realized $(\mathbf{X}_t, \mathbf{f}_t)_{t=1}^T$ is very small, we cannot rule out such a possibility in actual estimation. To safeguard the convergent solution of (3.13) and (3.14) is not such a $\left(\bar{\boldsymbol{\beta}}, \bar{\mathbf{w}} \right)$, we need to find $\hat{\mathbf{w}}$ that corresponds to the smallest root of the $N \times N$ matrix

$$\Omega^{-1/2} \frac{1}{T} \sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}) (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta})' \Omega^{-1/2}. \tag{3.16}$$

Unfortunately, Ω is in general unknown. However, noting that the roots of

$$\left| \frac{1}{T} \sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}) (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta})' - \delta \Omega \right| = 0, \tag{3.17}$$

and

$$\left| \Omega^{-1/2} \frac{1}{T} \sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}) (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta})' \Omega^{-1/2} - \delta \mathbf{I}_N \right| = 0, \tag{3.18}$$

are identical and the smallest eigenvalue converges to 1 as $T \to \infty$. Let $\hat{\mathbf{w}}$ be the corresponding eigenvector, then (3.17) implies

$$\hat{\mathbf{w}}' \left(\frac{1}{T} \sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}) (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta})' \right) \hat{\mathbf{w}} = \delta \hat{\mathbf{w}}' \Omega \hat{\mathbf{w}}, \tag{3.19}$$

and

$$\hat{\mathbf{w}}' \Omega^{-1/2} \frac{1}{T} \sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}) (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta})' \Omega^{-1/2} \hat{\mathbf{w}} = \delta \hat{\mathbf{w}}' \hat{\mathbf{w}}. \tag{3.20}$$

As $T \to \infty$, the left hand side of (3.19) converges to

$$\hat{\mathbf{w}}' \left(\Omega + \Lambda \Lambda' \right) \hat{\mathbf{w}} = \hat{\mathbf{w}}' \Omega \hat{\mathbf{w}}, \tag{3.21}$$

when $\hat{\mathbf{w}}' \Lambda = 0$. The left hand side of (3.20) converges to

$$\hat{\mathbf{w}}' \Omega^{-1/2} \left(\Omega + \Lambda \Lambda' \right) \Omega^{-1/2} \hat{\mathbf{w}} = \hat{\mathbf{w}}' \hat{\mathbf{w}} + \delta \hat{\mathbf{w}}' \Omega^{-1/2} \Lambda \Lambda' \Omega^{-1/2} \hat{\mathbf{w}}. \tag{3.22}$$

If $\delta = 1$, the left hand side of (3.20) equals the right hand side if $\Omega^{-1/2} \hat{\mathbf{w}}$ also lies in the null space of Λ. Let $\bar{\mathbf{w}}^* = \Omega^{-1/2} \hat{\mathbf{w}}$, then $\bar{\mathbf{w}}^*$ is the eigenvector corresponding to the unit root of the determinantal equation

$$\left| \frac{1}{T} \sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}) (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta})' - \delta \mathbf{I}_N \right| = 0 \tag{3.23}$$

However, the covariance matrix (3.23) is not scale invariant. To make sure the matrix is scale invariant, we transform (3.23) into the matrix

$$\frac{1}{T} \sum_{t=1}^{T} (\mathbf{y}_t^* - \mathbf{X}_t^* \boldsymbol{\beta}) (\mathbf{y}_t^* - \mathbf{X}_t^* \boldsymbol{\beta})', \tag{3.24}$$

where

$$\mathbf{y}_t^* = \frac{1}{b} \mathbf{y}_t, \mathbf{X}_t^* = \frac{1}{b} \mathbf{X}_t \text{ with } b^2 = \frac{1}{N} tr \left[\frac{1}{T} \sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}) (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta})' \right]. \tag{3.25}$$

Therefore, we suggest the following iterative procedure to obtain the TQMLE $\left(\hat{\boldsymbol{\beta}}, \hat{\mathbf{w}} \right)$:

Computational Algorithm: *If $\Omega = \sigma_u^2 \mathbf{I}_N$, the TQMLE of $\boldsymbol{\beta}$ and $\mathbf{w}$ could be obtained by iterating between (3.13) and (3.14) when $\frac{N}{T} \leq 1$. When $\Omega \neq \sigma_u^2 \mathbf{I}_N$, but is a known constant matrix, the TQMLE can be obtained by iterating between (3.13) and the eigenvector corresponding to the smallest root of the $N \times N$ matrix (3.16) until the solution converges. When Ω is unknown, the TQMLE can be obtained through the following steps:*

Step 1: Iterate (3.13) and (3.14) until the solution converges, say $\hat{\boldsymbol{\beta}}^{(l)}$.

Step 2: Substituting $\hat{\boldsymbol{\beta}}^{(l)}$ *into (3.23). Compute* b^2 *from (3.25). Obtain the eigenvalues of (3.24), say* $(\delta_1 > \cdots > \delta_N > 0)$. *Select the eigenvector corresponding to the eigenvalue of the matrix (3.24) that is closest to 1, say* $\hat{\mathbf{w}}^{(l)}$.

Step 3: Estimate $\boldsymbol{\beta}$ *by*

$$\hat{\boldsymbol{\beta}}^{(l+1)} = \left(\sum_{t=1}^{T} \mathbf{X}_t' \hat{\mathbf{w}}^{(l)} \hat{\mathbf{w}}^{(l)\prime} \mathbf{X}_t\right)^{-1} \left(\sum_{t=1}^{T} \mathbf{X}_t' \hat{\mathbf{w}}^{(l)} \hat{\mathbf{w}}^{(l)\prime} \mathbf{y}_t\right). \tag{3.26}$$

Step 4: Repeat Step 2 and 3 until the solution converges.

The above iterative algorithm is essentially the same as the one suggested by Bai (2009) and similar with the one proposed by Gorski et al. (2007), where convergence to a local optimum for the algorithm is also provided.

Remark 3.2 *The estimation method and the asymptotic distribution of the estimator remain unchanged with heteroskedastic and weak cross-sectional dependence (or spatial dependence) of the error term* u_{it} *(e.g.,* $\sum_{j=1}^{N} |\sigma_{ij}| < M < \infty$*). It can allow any data generation process of* $\mathbf{f}_t$ *because* $\Lambda \mathbf{f}_t$ *no longer appears in the transformed equation (3.3). Moreover, the linear transformation vector* $\hat{\mathbf{w}}$ *simultaneously takes care of cross-sectional dependence due to* $\Lambda \mathbf{f}_t$ *and weakly cross-sectional dependence due to* $\mathbf{u}_t$ *as in Hsiao and Zhou (2019).*

Remark 3.3 *Both the Bai (2009) iterative scheme or the iterative scheme suggested here make use of the eigenvectors corresponding to the* $N \times N$ *covariance matrix of* $(\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta})$ *(e.g., (3.23) or (3.24)). However, there is an important difference between the two procedures. Bai's (2009) iterative scheme corresponds to finding the largest* r *eigenvalues of the covariance matrix* $(\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta})$, *while our procedure corresponds to the eigenvalues smaller than the* $(r+j)$*-th largest eigenvalues, where* $j > 0$. *Since the covariance matrix of* $(\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta})$ *converges to* $(\Omega + \Lambda\Lambda')$ *and* $\Lambda\Lambda'$ *is a positive semi-definite matrix, finding the eigenvector corresponding to the* r *largest eigenvalues requires the knowledge of the rank of* Λ *(Bai (2009)) or to assume the rank of* Λ *is greater than the true rank of* Λ *(Moon and Weidner (2015)). Our procedure corresponds to finding the eigenvector corresponding to the smallest root of* $(\Omega + \Lambda\Lambda')$. *Since* Ω *is of rank* N *and* $\Lambda\Lambda'$ *is a positive semi-definite matrix of rank* r*, the* $(r+j)$*-th eigenvalue is just an eigenvalue of* Ω *for* $j > 0$. *If* $\mathbf{X}_t$ *are correlated with* $\Lambda \mathbf{f}_t$*, the iterative procedure between* $\hat{\boldsymbol{\beta}}^{(l)}$ *and the eigenvectors corresponding to the* r *largest value of the matrix* $\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{y}_t - \mathbf{X}_t \hat{\boldsymbol{\beta}}^{(l)}\right)\left(\mathbf{y}_t - \mathbf{X}_t \hat{\boldsymbol{\beta}}^{(l)}\right)'$ *may not converge if one starts with an arbitrary initial estimator (Jiang et al. (2019)) as long as* $N > \max(r, k)$. *On the other hand, our iterative scheme will always converge due to the global identification of* $\boldsymbol{\beta}$, $E(\mathbf{u}_t | \mathbf{X}_t, \Lambda \mathbf{f}_t) = 0$, *and*

$\left(\frac{1}{T}\sum_{t=1}^{T}(\mathbf{y}_t-\mathbf{X}_t\boldsymbol{\beta})(\mathbf{y}_t-\mathbf{X}_t\boldsymbol{\beta})'\right)\hat{\mathbf{w}}=\hat{\delta}_{r+j}\Omega\hat{\mathbf{w}}=\hat{\delta}_{r+j}\hat{\mathbf{w}}$, *where* $j>0$ *for the* N *eigenvalues of the matrix (3.23) are arranged in decreasing order* $\left(\hat{\delta}_1\geq\hat{\delta}_2\geq\cdots\geq\hat{\delta}_r\geq\hat{\delta}_{r+1}\geq\cdots\geq\hat{\delta}_N>0\right)$.[3]

Remark 3.4 *When* T *is fixed and* N *is large or* $\frac{N}{T}>1$, *model (2.1) and (2.2) can be written in the form*

$$\mathbf{y}_i=\mathbf{X}_i\boldsymbol{\beta}+\mathbf{F}\boldsymbol{\lambda}_i+\mathbf{u}_i, \tag{3.27}$$

where $\mathbf{y}_i=(y_{i1},\ldots,y_{iT})'$, $\mathbf{X}_i=(\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT})'$ *and* $\mathbf{u}_i=(u_{i1},\ldots,u_{iT})'$. *Under the assumptions that*

Assumption A1′: $\mathbf{u}_i$ *is independently, identically distributed over* i *with* $E(\mathbf{u}_i|\mathbf{X}_i,\mathbf{F},\boldsymbol{\lambda}_i)=0$ *and* $E(\mathbf{u}_i\mathbf{u}_i')=\Omega^*$, *where* Ω^* *is nonsingular and the eigenvalues of* Ω^* *are* $O(1)$.

Assumption A2′: rank $(\Lambda)=r$ *and* $plim_{N\to\infty}\frac{1}{N}\Lambda'\Lambda=\mathbf{D}^*$ *where* $\mathbf{D}^*$ *is a nonsingular diagonal matrix.*

Assumption A3′: rank $(\mathbf{F})=r$ *and* $\frac{1}{T}\mathbf{F}'\mathbf{F}=\mathbf{I}_r$.

we can similarly derive the estimation of $\tilde{\mathbf{w}}^*$ *and* $\boldsymbol{\beta}$ *that minimizes*

$$\frac{1}{N}\sum_{i=1}^{N}\tilde{\mathbf{w}}^{*\prime}(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})(\mathbf{y}_i-\mathbf{X}_i\boldsymbol{\beta})'\tilde{\mathbf{w}}^*, \tag{3.28}$$

subject to $\tilde{\mathbf{w}}^{*\prime}\Omega^*\tilde{\mathbf{w}}^*=1$.

Assumption A1′ allows weak time dependence in u_{it} *but excludes lag dependent variables to appear as explanatory variables. In other words,* $\mathbf{x}_{it}$ *is strictly exogenous with respect to* u_{it}, *contrary to the fixed* N *and large* T *case, our suggested estimation method for* $\boldsymbol{\beta}$ *is consistent and asymptotically normally distributed only if* $\mathbf{x}_{it}$ *is strictly exogenous with respect to* u_{it} *(Assumption A1′). If* $\mathbf{x}_{it}$ *contains lagged dependent variable, then* $E(\mathbf{X}_i'\mathbf{u}_i)\neq 0$. *Assumption A2′-A3′ are alternative ways of normalization of* Λ *and* $\mathbf{F}$. *However, the formation (3.27) can allow structural change in the factor loading matrix* Λ.

4 An Average Estimator

The transformed estimator is consistent and asymptotically normally distributed either N or T goes to infinity. It has the advantage that it does not require both $(N,T)\to\infty$. In other words, the estimator converges to the true value at the speed of $O\left(T^{-1/2}\right)$ (or $O\left(N^{-1/2}\right)$). However, we have NT observations for $(y_{it},\mathbf{x}_{it}')'$. The Bai (2009) least squares estimators, although requires both $(N,T)\to\infty$, it has also the advantage that the Bai's least squares estimator converges to

[3] $\hat{\mathbf{w}}$ being an eigenvector of the matrix Ω follows from the $(r+j)$-th largest root of the determinantal equation $\left|\frac{1}{T}\sum_{t=1}^{T}(\mathbf{y}_t-\mathbf{X}_t\boldsymbol{\beta})(\mathbf{y}_t-\mathbf{X}_t\boldsymbol{\beta})'-\delta\Omega\right|=0$ is also a root of the determinantal equation $|\Omega-\delta\mathbf{I}_N|=0$.

the true value at the speed of $O\left((NT)^{-1/2}\right)$. In this section, we propose an average estimator based on the transformed model (3.3), that also converges to the true value at the speed of $O\left((NT)^{-1/2}\right)$.

We note that given N $(> r)$, under Assumption A2, there are $N-r$ orthogonal vectors, $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{N-r}$, span the null space of Λ, such that $\mathbf{w}_j'\Lambda = 0$ for $j = 1, 2, \ldots, N-r$. Then

$$\mathbf{w}_j'\mathbf{y}_t = \mathbf{w}_j'\mathbf{X}_t\boldsymbol{\beta} + \mathbf{w}_j'\mathbf{u}_t,\ j = 1, 2, \ldots, N-r,\ t = 1, \ldots, T. \tag{4.1}$$

Conditional on $\mathbf{w}_j$ for $j = 1, 2, \ldots, N-r$, the least squares estimator

$$\hat{\boldsymbol{\beta}}^{(j)} = \left(\sum_{t=1}^{T} \mathbf{X}_t'\mathbf{w}_j\mathbf{w}_j'\mathbf{X}_t\right)^{-1} \left(\sum_{t=1}^{T} \mathbf{X}_t'\mathbf{w}_j\mathbf{w}_j'\mathbf{y}_t\right), \quad j = 1, 2, \ldots, N-r, \tag{4.2}$$

is consistent and

$$\sqrt{T}\left(\hat{\boldsymbol{\beta}}^{(j)} - \boldsymbol{\beta}\right) \to_d N\left(0, \sigma_j^2\left(\Xi_w^{(j)}\right)^{-1}\right), \tag{4.3}$$

when $T \to \infty$ as shown in the Section 3, where $\sigma_j^2 = \mathbf{w}_j'\Omega\mathbf{w}_j$ and $\Xi_w^{(j)} = plim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} \mathbf{X}_t'\mathbf{w}_j\mathbf{w}_j'\mathbf{X}_t$. In other words, there are $(N-r)$ independent estimators of $\boldsymbol{\beta}$. Under the normalization condition $\mathbf{w}_j'\Omega\mathbf{w}_j = 1$, then the $(N-r)$ estimators $\hat{\boldsymbol{\beta}}^{(j)}$ are asymptotically independent because $\mathbf{w}_j'\Omega\mathbf{w}_l = 0$ for $j \neq l$. Therefore, we propose a $\sqrt{NT}$-consistent average estimator,

$$\hat{\boldsymbol{\beta}}^{Ave} = \frac{1}{N-r}\sum_{j=1}^{N-r}\hat{\boldsymbol{\beta}}^{(j)}. \tag{4.4}$$

In order to derive the asymptotic properties of the average estimator $\hat{\boldsymbol{\beta}}^{Ave}$, we let $\Xi_T^{(j)} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{X}_t'\mathbf{w}_j\mathbf{w}_j'\mathbf{w}'\mathbf{X}_t$ with $\Xi_w^{(j)} = plim_{T\to\infty}\Xi_T^{(j)}$ and $\Sigma_{NT} = \frac{1}{N-r}\sum_{j=1}^{N-r}\left(\Xi_T^{(j)}\right)^{-1}$, and $\boldsymbol{\xi}_t = \frac{1}{N-r}\sum_{j=1}^{N-r}\left(\Xi_T^{(j)}\right)^{-1}\mathbf{X}_t'\mathbf{w}_j\mathbf{w}_j'\mathbf{u}_t$. Therefore, for any fixed constant $\boldsymbol{\lambda} \in \mathbf{R}^k$ such that $\boldsymbol{\lambda}'\boldsymbol{\lambda} = 1$, we define a scalar random variable $Z_{\lambda,Nt} = \sqrt{N-r}\boldsymbol{\lambda}'\boldsymbol{\xi}_t$ and a scalar standard deviation $\sigma_{\lambda,NT} = \sqrt{\boldsymbol{\lambda}'\Sigma_{N,T}\boldsymbol{\lambda}}$. The asymptotic properties of $\hat{\boldsymbol{\beta}}^{Ave}$ is summarized in the following proposition.

Proposition 4.1 *Under Assumption A1-A4, suppose for all fixed* $\boldsymbol{\lambda}$ *and* $\varepsilon > 0$ *we have*

$$\frac{1}{T\sigma_{\lambda,N,T}^2}\sum_{t=1}^{T} E\left[Z_{\lambda,N,t}^2 \cdot \mathbf{1}\left\{|Z_{\lambda,Nt}| > \sqrt{T}\sigma_{\lambda,NT}\varepsilon\right\}\right] \to 0, \tag{4.5}$$

as $(N,T) \to \infty$*, then* $\hat{\boldsymbol{\beta}}^{Ave}$ *is consistent and*

$$\Sigma_{N,T}^{-1/2}\sqrt{(N-r)\,T}\left(\hat{\boldsymbol{\beta}}^{Ave} - \boldsymbol{\beta}\right) \to_d N\left(0, \mathbf{I}_k\right). \tag{4.6}$$

Remark 4.1 *Equation (4.5) is a Lindeberg condition for the model indexed by N and T. The Lindeberg condition is generally required for central limit theorems for i.n.i.d. observations (e.g., White (2011)).*

Remark 4.2 *For the case of fixed number of common factors, i.e., r is a finite positive constant, the results obtained in (4.6) is asymptotically equivalent to*

$$\Sigma_{N,T}^{-1/2}\sqrt{NT}\left(\hat{\boldsymbol{\beta}}^{Ave}-\boldsymbol{\beta}\right)\to_d N\left(0,\mathbf{I}_k\right). \tag{4.7}$$

Proposition 4.1 is derived under the assumption that we know $(N-r)$ orthogonal vectors that span the null space of Λ. In practice, neither Λ nor r are known. To obtain the $N-r$ orthogonal $\mathbf{w}_j$, we note that under Assumption A1-A3,

$$\frac{1}{NT}\sum_{t=1}^{T}\left(\mathbf{y}_t-\mathbf{X}_t\boldsymbol{\beta}\right)\left(\mathbf{y}_t-\mathbf{X}_t\boldsymbol{\beta}\right)'\to_p\left(\frac{1}{N}\Lambda\Lambda'+\frac{1}{N}\Omega\right). \tag{4.8}$$

It is shown by Ahn and Horenstein (2013) and Bai and Yin (1993) that the largest r eigenvalues are of order $O(1)$ and the rest are of order $O\left(1/\min\left(N,T\right)\right)$. Therefore, we propose to use the transformed estimator discussed in the Section 3 to obtain an initial consistent estimator of $\boldsymbol{\beta}$ and $\mathbf{w}$, say, $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{w}}$, then compute the sample covariance matrix

$$\mathbf{S}_{yx}=\frac{1}{NT}\sum_{t=1}^{T}\left(\mathbf{y}_t-\mathbf{X}_t\hat{\boldsymbol{\beta}}\right)\left(\mathbf{y}_t-\mathbf{X}_t\hat{\boldsymbol{\beta}}\right)', \tag{4.9}$$

and apply the Ahn and Horenstein's (2013) eigenvalue ratio test to determine r. Once r is determined, we just use the eigenvectors corresponding to the $N-r$ smallest eigenvalues of $\mathbf{S}_{yx}$ in (4.9) to transform model (2.3) and to obtain each $\hat{\boldsymbol{\beta}}^{(j)}$ using (4.2).

Remark 4.3 *Proposition 4.1 is derived under the normalization condition that $\mathbf{w}'\Lambda=0$ and $\mathbf{w}'\Omega\mathbf{w}=1$. As a matter of fact, any $(N-r)$ orthogonal factors that satisfy $\mathbf{w}'\Lambda=0$ subject to any normalization condition can transform (2.3) into (3.3), and taking the average of such $(N-r)$ estimators can achieve the $\sqrt{NT}$-consistency. However, the asymptotic independence of $E\left(\mathbf{w}_j^{*\prime}\mathbf{u}_t\mathbf{u}_t'\mathbf{w}_l^*\right)$ may not hold, hence the asymptotic covariance matrix of the resulting average estimator $\hat{\boldsymbol{\beta}}^{*Ave}$ will be more complicated.*

Remark 4.4 *One of the advantage of using the average estimator (4.4) relative to simultaneously estimating $(\boldsymbol{\beta},\boldsymbol{\Lambda},\mathbf{F})$ is that the resulting estimator also possesses $\sqrt{NT}$-consistency while there is no asymptotic bias if $\mathbf{x}_{it}$ contains lagged dependent variable.*

5 Simulations

In this section, we investigate the finite sample properties of the TQMLE for panel interactive effects models with heteroskedastic idiosyncratic errors. We consider the following data generating processes (DGPs):[4]

DGP1: (Model with two factors)

$$y_{it} = x_{1,it}\beta_1 + x_{2,it}\beta_2 + \lambda_{1,i} f_{1,t} + \lambda_{2,i} f_{2,t} + u_{it}, \tag{5.1}$$

where $\beta_1 = 1$ and $\beta_2 = 2$, and $u_{it} \sim IIDN\left(0, \sigma_{u,i}^2\right)$ with $\sigma_{u,i}^2$ being independent draws from $\left(1 + 0.5\chi^2(2)\right)$. The covariates x_{it} is generated as

$$x_{k,it} = 1 + \alpha_{ki} + c_{k1,i} f_{1t} + c_{k2,i} f_{2t} + \eta_{k,it},\ k = 1, 2. \tag{5.2}$$

DGP2: (Model with three factors)

$$y_{it} = x_{1,it}\beta_1 + x_{2,it}\beta_2 + \lambda_{1,i} f_{1,t} + \lambda_{2,i} f_{2,t} + \lambda_{3,i} f_{3,t} + u_{it}, \tag{5.3}$$

where $\beta_1 = 1$ and $\beta_2 = 2$, and x_{it} is generated the same as (5.2).

For these DGP (5.1)-(5.3), the idiosyncratic errors $\eta_{k,it}$ of $x_{k,it}$ is generated as

$$\eta_{k,it} = \rho_{k,i}\eta_{k,it-1} + v_{k,it},\ k = 1, 2,$$

with $\rho_{k,i}$ are i.i.d draws from $U(0.1, 0.9)$ for $k = 1, 2$ and $i = 1, \ldots, N$.

We also consider the following generating mechanisms of the common factors for DGP1:

DGP3: (Factors with structural change)

$$(f_{1t}, f_{2t}) \sim IIDN(0, I_2),\ t = 1, \ldots, [T/2], \tag{5.4}$$

$$\begin{cases} f_{1t} = 0.8 f_{1t-1} + \xi_{1t}, \\ f_{2t} = 0.3 f_{2t-1} + \xi_{2t}, \end{cases},\quad t = [T/2] + 1, \ldots, T,$$

where ξ_{1t} and ξ_{2t} are i.i.d $N(0, 1)$.

DGP4: (Weakly cross-sectionally correlated errors) We assume y_{it} and $x_{k,it}$ are generated as in DGP1 except that now we let u_{it} be generated as

$$u_{it} = \varepsilon_{it} + b_1\varepsilon_{i+1,t} + b_2\varepsilon_{i-1,t}, \tag{5.5}$$

where $\varepsilon_{i,t} \sim IIDN\left(0, \sigma_{\varepsilon,i}^2\right)$ with $\sigma_{\varepsilon,i}^2$ being random draws from $\left(1 + 0.5\chi^2(2)\right)$. We let $b_1 = 0.3$ and $b_2 = 0.2$.

[4]Additional simulations for DGPs with cointegrated common factors or nonnormal errors are provided in the Appendix B.

DGP5: Dynamic model (5.6) with i.i.d common factors

$$y_{it} = \rho y_{i,t-1} + x_{it}\beta + \lambda_{1,i} f_{1,t} + \lambda_{2,i} f_{2,t} + u_{it}, \tag{5.6}$$

where $\rho = 0.5$ and $\beta = 1$, respectively, and x_{it} is generated the same as (5.2).

DGP6: Dynamic model (5.6) with structural changing factors as in (5.4).

For these DGPs, we assume $\alpha_{1i}, \alpha_{2i} \sim IIDN(0,1)$, and $v_{j,it} \sim IIDN\left(0, \sigma^2_{vj,i}\right)$ for $j = 1,2,3$ and $\sigma^2_{v1,i}$, $\sigma^2_{v2,i}$ are independent draws from $\left(1 + 0.5\chi^2(2)\right)$. For the factors, we assume $f_{1t} \sim IIDN(0,1)$, $f_{2t} \sim IIDN(0,2)$ and $f_{3t} \sim IIDN(0,3)$. For the factor loadings, we assume that $\lambda_{1,i}, \lambda_{2,i}, \lambda_{3,i}$ are i.i.d draws from $N(1,1)$ and $c_{k1,i}$ and $c_{k2,i}$ are i.i.d draw from $U(0,1)$, for $i = 1,2,\ldots,N$, $k = 1,2$.

In the simulation, we let the number of replication is set to be 1000. We let $N = 10, 20, 50$ and $T = 100, 200, 500$. For these estimators of our interest, we report the mean estimates, bias, RMSE, IQR (inter-quantile range, 75%-25% percentile) and empirical size (computed as the empirical rejection frequency using 5% nominal critical value) for comparison. We use the fixed effects estimator as the initial estimator in the iteration, and the maximum number of iteration is set at 50. For the implementation of the average estimator and the Principle component analysis (PCA) estimator of Bai (2009), we first use the transformed estimation to compute the residuals for the panel interactive effects model, and then use the eigenvalue ratio test proposed by Ahn and Horenstein (2013) to determine the number of factors in the model. Finally, both the average estimator and PCA estimator are based on the estimated number of common factors. The results are summarized in Table 1-6 for DGP1-6, respectively.

From the simulation results, we observe that: (1) whatever the data generating processes, be a static or dynamic model (DGP1-4 vs 5-6), or constant or changing interactive structure (DGP1-2 vs 3 or DGP5 vs 6), or the idiosyncratic error is cross-sectionally independent or weakly dependent (DGP1-3 vs 4), the transformed maximum likelihood estimation of the slope coefficients $\boldsymbol{\beta}$, (3.4), perform well even though it is only $\sqrt{T}$-consistent. The bias is negligible and the empirical size is close to the nominal size. (2) The IQR of the average estimator is much smaller than the TQMLE, reflecting the $\sqrt{NT}$-consistency vs $\sqrt{T}$-consistency. However, the empirical size of the average estimator is not as steady as the TQMLE, so is the PCA estimator, possibly due to the misleading inference on the determination of common factors, r, that could lead to the increase of the asymptotical bias in the estimation.

6 Concluding Remarks

We have suggested a transformed QMLE estimator (TQMLE) for panel models with interactive effects. The advantage of the TQMLE is that (i) computational simplicity; (ii) either N or

T goes to infinity is sufficient to obtain consistent and asymptotically normally distributed estimator; (iii) There is no need to find the dimension of $\mathbf{f}_t$ or Λ, which is usually unknown, or for the DGP of $\mathbf{x}_{it}$ to be correlated with $\mathbf{f}_t$ and to satisfy the rank condition assumed by Pesaran (2006); (vi) choosing $\mathbf{w}$ corresponding to the smallest root of (3.14) is less likely to be contaminated by the initial estimator of $\boldsymbol{\beta}$ because the solution of (3.14) could still satisfy $\mathbf{w}'\Lambda = 0$. Estimating $\mathbf{F}$ by finding the eigenvectors that correspond to the largest r eigenvalues of (3.14) could fail to converge if the initial estimator is inconsistent (Jiang et al (2019)); (v) The estimation method and the asymptotic distribution remain unchanged whether $\mathbf{X}_t$ is strictly exogenous with regard to the idiosyncratic errors u_{it} or if u_{it} is homoskedastic or heteroskedastic or weakly cross-sectional dependent. Furthermore, our linear transformation vector $\mathbf{w}$ simultaneously takes care of cross-sectional dependence due to $\Lambda\mathbf{f}_t$ and weakly cross-sectionally dependence due to $\mathbf{u}_t$ as in Hsiao and Zhou (2019). Furthermore, the efficiency of the TQMLE can be improved by taking the average of the transformed estimators based on the orthogonal eigenvectors that lie on the null space of the factor or factor loading matrices. However, to avoid possible bias due to misspecifying the dimension of common factors, we would recommend taking a conservative approach towards the empirically determined dimension, say r^*, by letting $r = r^* + c$, where c can be some constants selected by the investigator with regard to the trade-off between efficiency and asymptotic bias.[5]

[5] Given the same issue would also arise from the least squares or PCA estimator for $(\boldsymbol{\beta}, \boldsymbol{\Lambda}, \mathbf{F})$ and the Moon and Weidner's (2015) finding that selecting the dimension greater than r does not affect the consistency and asymptotic distribution, we would recommend this strategy.

Table 1: Simulation results for β_1 and β_2 of DGP 1

β_1		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	1.0006	0.0006	0.0642	0.0809	5.8%	1.0035	0.0035	0.0228	0.0306	5.6%	1.2265	0.2265	0.2325	0.0742	99%
10	200	1.0059	0.0059	0.0453	0.0593	5.2%	1.0030	0.0030	0.0158	0.0206	5.4%	1.2281	0.2281	0.2319	0.0555	100%
	500	1.0060	0.0060	0.0302	0.0365	5.2%	1.0032	0.0032	0.0101	0.0130	7%	1.2273	0.2273	0.2297	0.0452	100%
	100	0.9980	-0.0020	0.0555	0.0699	6.2%	1.0006	0.0006	0.0149	0.0203	6.1%	1.2523	0.2523	0.2560	0.0596	100%
20	200	1.0017	0.0017	0.0376	0.0486	5.3%	1.0000	0.0000	0.0100	0.0134	4.2%	1.2560	0.2560	0.2585	0.0473	100%
	500	1.0010	0.0010	0.0245	0.0333	4.8%	1.0001	0.0001	0.0064	0.0083	5.2%	1.2533	0.2533	0.2550	0.0396	100%
	100	1.0009	0.0009	0.0665	0.0830	6.3%	1.0004	0.0004	0.0112	0.0149	5.2%	1.3209	0.3209	0.3234	0.0536	100%
50	200	1.0027	0.0027	0.0452	0.0585	5.3%	1.0001	0.0001	0.0078	0.0107	5.4%	1.3207	0.3207	0.3223	0.0428	100%
	500	0.9983	-0.0017	0.0304	0.0381	5.8%	1.0005	0.0005	0.0047	0.0063	4.9%	1.3216	0.3216	0.3226	0.0341	100%
100	100	1.0050	0.0050	0.0631	0.0807	5.2%	1.0000	0.0000	0.0074	0.0104	5.3%	1.0000	0.0000	0.0063	0.0086	4.3%
	200	1.0006	0.0006	0.0434	0.0538	5.2%	0.9998	-0.0002	0.0050	0.0071	4.6%	1.3268	0.3268	0.3299	0.0379	100%
	500	1.0003	0.0003	0.0280	0.0363	5.8%	1.0001	0.0001	0.0033	0.0044	5.2%	1.3295	0.3295	0.3302	0.0300	100%
β_2		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	2.0061	0.0061	0.0658	0.0854	5.9%	2.0035	0.0035	0.0257	0.0314	5%	2.2548	0.2548	0.2603	0.0738	100%
10	200	2.0019	0.0019	0.0454	0.0572	6.1%	2.0022	0.0022	0.0161	0.0224	5%	2.2552	0.2552	0.2587	0.0581	100%
	500	2.0000	0.0000	0.0311	0.0386	5.4%	2.0011	0.0011	0.0100	0.0129	5.5%	2.2538	0.2538	0.2560	0.0445	100%
	100	2.0017	0.0017	0.0588	0.0578	5.2%	2.0005	0.0005	0.0154	0.0201	4.6%	2.2681	0.2681	0.2726	0.0667	100%
20	200	1.9997	-0.0003	0.0396	0.0516	5.4%	1.9998	-0.0002	0.0101	0.0136	4.3%	2.2668	0.2668	0.2695	0.0508	100%
	500	2.0020	0.0020	0.0277	0.0267	5.2%	2.0006	0.0006	0.0070	0.0101	5.2%	2.2680	0.2680	0.2698	0.0423	100%
	100	2.0004	0.0004	0.0610	0.0779	4.6%	1.9997	-0.0003	0.0106	0.0144	4.7%	2.2977	0.2977	0.3002	0.0514	100%
50	200	1.9987	-0.0013	0.0445	0.0559	5.5%	1.9994	-0.006	0.0072	0.0096	5.9%	2.2973	0.2973	0.2991	0.0460	100%
	500	2.0000	0.0000	0.0287	0.0357	6.2%	1.9999	-0.0001	0.0044	0.0062	5.3%	2.2981	0.2981	0.2991	0.0331	100%
100	100	2.0037	0.0037	0.0609	0.0819	5.1%	2.0002	0.0002	0.0072	0.0098	4.7%	2.0001	0.0001	0.0059	0.0080	4.3%
	200	2.0011	0.0011	0.0411	0.0556	4.1%	1.9999	-0.0001	0.0047	0.0064	4.6%	2.2786	.2786	0.2806	0.0318	100%
	500	1.9998	-0.0002	0.0250	0.0328	5.3%	2.0000	0.0000	0.0029	0.0041	4.6%	2.2810	0.2810	0.2817	0.0268	100%

Table 2: Simulation results for β_1 and β_2 of DGP 2

β_1		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	1.0041	0.0041	0.0692	0.0847	5.1%	1.0227	0.0227	0.0429	0.0488	10%	1.3630	0.3630	0.3718	0.1094	99%
10	200	1.0088	0.0088	0.0501	0.0625	6.2%	1.0232	0.0232	0.0367	0.0404	13%	1.3672	0.3672	0.3728	0.0910	100%
	500	1.0098	0.0098	0.0342	0.00425	5.5%	1.0223	0.0223	0.0326	0.0352	17%	1.3644	0.3644	0.3678	0.0693	100%
	100	1.0012	0.0012	0.0586	0.0761	4.9%	1.0008	0.0008	0.0150	0.0201	4.8%	1.2978	0.2978	0.3040	0.0832	100%
20	200	1.0005	0.0005	0.0412	0.0491	5.8%	1.0004	0.0004	0.0104	0.0139	5.7%	1.2970	0.2970	0.3011	0.0662	100%
	500	0.9995	-0.0005	0.0274	0.0346	4.6%	1.0005	0.0005	0.0066	0.0089	5.2%	1.2975	0.2975	0.2996	0.0462	100%
	100	1.0015	0.0015	0.0635	0.0777	5.7%	1.0002	0.0002	0.0114	0.0154	4.9%	1.3129	0.3129	0.3185	0.0796	100%
50	200	0.9994	-0.0006	0.0448	0.0548	6.4%	1.0000	0.0000	0.0081	0.0110	4.7%	1.3158	0.3158	0.3189	0.0583	100%
	500	0.9999	-0.0001	0.0290	0.0375	5.7%	1.0001	0.0001	0.0049	0.0068	4.9%	1.3186	0.3186	0.3204	0.0448	100%
100	100	1.0033	0.0033	0.0603	0.0767	5.6%	1.0013	0.0013	0.0106	0.0099	2.8%	1.0022	0.0022	0.0140	0.0086	3%
	200	1.0024	0.0024	0.0426	0.0531	5.3%	1.0003	0.0003	0.0049	0.0068	5.2%	1.3682	0.3682	0.3704	0.0549	100%
	500	1.0001	0.0001	0.0271	0.0360	5.5%	1.0002	0.0002	0.0030	0.0041	5.3%	1.3676	0.3676	9.3687	0.0385	100%
β_2		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	1.9966	-0.0034	0.0645	0.0770	6.5%	2.0606	0.0606	0.0718	0.0533	35%	2.2605	0.2605	0.2736	0.1099	88%
10	200	1.9953	-0.0047	0.0416	0.0541	4.4%	2.0606	0.0606	0.0675	0.0428	53%	2.2601	0.2601	0.2688	0.0910	97%
	500	2.0027	0.0027	0.0274	0.0353	5.8%	2.0601	0.0601	0.0658	0.0411	61%	2.2623	0.2623	0.2676	0.0761	100%
	100	2.0014	0.0014	0.0608	0.0796	4.2%	2.0004	0.0004	0.0167	0.0213	4.9%	2.3562	0.3562	0.3640	0.1023	100%
20	200	2.0013	0.0013	0.0428	0.0552	6%	1.9999	-0.0001	00.0115	0.0156	5%	2.3532	0.3532	0.3578	0.0801	100%
	500	2.0021	0.0021	0.0297	0.0377	5.1%	1.9998	-0.0002	0.0070	0.0095	4.9%	2.3550	0.3550	0.3578	0.0590	100%
	100	1.9999	-0.0001	0.0619	0.0768	5.1%	1.9998	-0.0002	0.0106	0.0139	4.6%	2.2789	0.2789	0.2845	0.0767	100%
50	200	2.0019	0.0019	0.0409	0.0558	5%	1.9998	-0.0002	0.0065	0.0090	5.3%	2.2817	0.2817	0.2848	0.0574	100%
	500	2.0014	0.0014	0.0261	0.0345	4.8%	1.9999	-0.0001	0.0042	0.0059	4.7%	2.2837	0.2837	0.2854	0.0416	100%
100	100	2.0028	0.0028	0.0605	0.0803	5.1%	1.9997	-0.0003	0.0082	0.0103	3.8%	2.0012	0.0012	0.0092	0.0089	3%
	200	1.9981	-0.0019	0.0446	0.0553	5.6%	1.9997	-0.0003	0.0050	0.0067	5.6%	2.3203	0.3203	0.3226	0.0520	100%
	500	2.0004	0.0004	0.0280	0.0376	4.6%	2.0003	0.0003	0.0030	0.0041	5.1%	2.3215	0.3215	0.3227	0.0381	100%

Table 3: Simulation results for β_1 and β_2 of DGP 3

β_1		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	1.0005	0.0005	0.0618	0.0778	5.4%	1.0019	0.0019	0.0236	0.0296	4.7%	1.2234	0.2234	0.2337	0.0896	92%
10	200	1.0055	0.0055	0.0441	0.0559	6.1%	1.0032	0.0032	0.0158	0.0206	6.1%	1.2271	0.2271	0.2336	0.0754	99%
	500	1.0032	0.0032	0.0292	0.0393	4.7%	1.0025	0.0025	0.0098	0.0123	6.3%	1.2266	0.2266	0.2300	0.0522	100%
	100	1.0009	0.0009	0.0547	0.0726	5.5%	1.0008	0.0008	0.0146	0.0201	4.9%	1.2467	0.2467	0.2531	0.0795	100%
20	200	1.0017	0.0017	0.0362	0.0471	5%	1.0004	0.0004	0.0105	0.0148	4.4%	1.2507	0.2507	0.2548	0.0620	100%
	500	1.0015	0.0015	0.0253	0.0338	5.3%	1.0000	0.0000	0.0064	0.0083	5.6%	1.2521	0.2521	0.2545	0.0474	100%
	100	1.0006	0.0006	0.0669	0.0829	5.3%	1.0000	0.0000	0.0117	0.0157	4.2%	1.3081	0.3081	0.3125	0.0713	100%
50	200	1.0007	0.0007	0.0458	0.0598	5.2%	1.0002	0.0002	0.0078	0.0108	4.8%	1.3123	0.2123	0.3151	0.0583	100%
	500	1.0013	0.0013	0.0287	0.0371	5.1%	1.0006	0.0006	0.0049	0.0066	5.4%	1.3179	0.3179	0.3194	0.0444	100%
100	100	0.9990	-0.0010	0.0631	0.0875	4.6%	0.9999	-0.0001	0.0075	0.0103	5.5%	1.0000	0.0000	0.0064	0.0081	3.8%
	200	1.0009	0.0009	0.0448	0.0577	5.5%	1.0001	0.0001	0.0051	0.0067	5%	1.3219	0.3219	0.3241	0.0514	100%
	500	0.9991	-0.0009	0.0274	0.0356	5.6%	1.0024	0.0024	0.0229	0.0308	5.3%	1.1350	0.1350	0.1438	0.0678	78%
β_2		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	1.9983	-0.0017	0.0655	0.0847	5.3%	2.0021	0.0021	0.0238	0.0330	4%	2.2571	0.2571	0.2661	0.0859	96%
10	200	2.0018	0.0018	0.0482	0.0578	5.4%	2.0018	0.0018	0.0166	0.0220	5.1%	2.2594	0.2594	0.2646	0.0718	100%
	500	2.0012	0.0012	0.0319	0.0414	4.6%	2.0019	0.0019	0.0102	0.0143	4.9%	2.2604	0.2604	0.2635	0.0552	100%
	100	2.0028	0.0028	0.0571	0.0730	4.4%	2.0009	0.0009	0.0157	0.0217	5.3%	2.2630	0.2630	0.2701	0.0795	99%
20	200	1.9998	-0.0002	0.0386	0.0490	5.3%	1.9998	-0.0002	0.0108	0.0147	5.2%	2.2640	0.2640	0.2680	0.0629	100%
	500	2.0011	0.0011	0.0345	0.0266	5.6%	2.0002	0.0002	0.0066	0.0089	5%	2.2693	0.2693	0.2718	0.0523	100%
	100	2.0006	0.0006	0.0648	0.0876	4.8%	1.9996	-0.0004	0.0108	0.0145	5.2%	2.2798	0.2798	0.2854	0.0739	100%
50	200	2.0012	0.0012	0.0448	0.0580	4.9%	1.9999	-0.0001	0.0073	0.0097	5.4%	2.2851	0.2851	0.2882	0.0555	100%
	500	2.0000	0.0000	0.0274	0.0377	4.5%	1.9999	-0.0001	0.0047	0.0061	5.4%	2.2903	0.2903	0.2920	0.0426	100%
100	100	2.0040	0.0040	0.0602	0.0743	5.4%	1.9996	-0.0004	0.0071	0.0091	5%	2.0000	0.0000	0.0065	0.0076	3.6%
	200	2.0013	0.0013	0.0412	0.0513	5.1%	1.9998	-0.0002	0.0047	0.0067	5%	2.2730	0.2730	0.2750	0.0429	100%
	500	1.9994	-0.0006	0.0270	0.0377	4.4%	1.9998	-0.0002	0.0234	0.324	4.8%	2.2383	0.2383	0.2441	0.0712	100%

Table 4: Simulation results for β_1 and β_2 of DGP 4

β_1		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	0.9971	-0.0029	0.0685	0.0854	6.1%	1.0024	0.0024	0.0229	0.0308	5.3%	1.1350	0.1350	0.1438	0.0678	78%
10	200	0.9996	-0.0004	0.0499	0.0601	5.5%	1.0016	0.0016	0.0159	0.0219	4.3%	1.1346	0.1345	0.1403	0.0556	93%
	500	0.9989	-0.0011	0.0314	0.0411	4.8%	1.0015	0.0015	0.0103	0.0140	4.7%	1.1366	0.1366	0.1401	0.0413	99%
	100	0.9988	-0.0012	0.0562	0.0717	6.5%	1.0005	0.0005	0.0143	0.0191	5.6%	1.2475	0.2475	0.2515	0.0602	100%
20	200	0.9971	-0.0029	0.0406	0.0530	4.8%	0.9996	-0.0004	0.0098	0.0136	4.5%	1.2458	0.2458	0.2484	0.0492	100%
	500	0.9981	-0.0019	0.0257	0.0331	5.3%	1.0000	0.0000	0.0062	0.0082	5.2%	1.2464	0.2464	0.2482	0.0401	100%
	100	1.0010	0.0010	0.06300	0.0824	4.7%	0.9991	-0.0009	0.0117	0.0148	5.9%	1.2849	0.2849	0.2873	0.0496	100%
50	200	0.9994	-0.0006	0.0469	0.0579	5.7%	0.9991	-0.0009	0.0080	0.0107	5.7%	1.2873	0.2873	0.2888	0.0420	100%
	500	0.9988	-0.0012	0.0312	0.0254	5.9%	0.9989	-0.0011	0.0050	0.0064	6.1%	1.2865	0.2865	0.2876	0.0337	100%
100	100	1.0039	0.0039	0.0597	0.0784	4.9%	0.9993	-0.0007	0.0081	0.0106	5.3%	0.9992	-0.0008	00069	0.0092	5.1%
	200	0.9992	-0.0008	0.0447	0.0566	5.1%	0.9994	-0.0006	0.0054	0.0072	5.4%	1.3196	0.3196	0.3209	0.0380	100%
	500	0.9986	-0.0014	0.0285	0.0353	6.1%	0.9995	-0.0005	0.0032	0.0045	4.5%	1.3216	0.3216	0.3223	0.0306	100%

β_2		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	1.9964	-0.0036	0.0716	0.0926	5.9%	1.9998	-0.0002	0.0234	0.0324	4.8%	2.2383	0.2383	0.2.441	0.0712	100%
10	200	2.0035	0.0035	0.0511	0.0643	5.5%	1.9991	-0.0009	0.0153	0.0203	4.1%	2.2364	0.2364	0.2398	0.0533	100%
	500	2.0054	0.0054	0.0318	0.0400	5.9%	1.9983	-0.0017	0.0097	0.0125	4.7%	2.2373	0.2373	0.2394	0.0428	100%
	100	2.0002	0.0002	0.0569	0.0704	5.2%	1.9996	-0.0004	0.0148	0.0195	5%	2.3221	0.3221	0.3258	0.0669	100%
20	200	1.9984	-0.0016	0.0411	0.0522	5.2%	1.9990	-0.0010	0.0105	0.0146	5.5%	2.3235	0.3235	0.3259	0.0559	100%
	500	1.9981	-0.0019	0.0292	0.0364	5.7%	1.9993	-0.0007	0.0065	0.0085	5%	2.3228	0.3228	0.3246	0.0449	100%
	100	2.0004	0.0004	0.0638	0.0825	5.6%	1.9985	-0.0015	0.0113	0.0152	4.8%	2.3351	0.3351	0.3378	0.0561	100%
50	200	2.0005	0.0005	0.0474	0.0611	5.8%	1.9987	-0.0013	0.0082	0.0111	4.3%	2.3382	0.3382	0.3398	0.0431	100%
	500	1.9989	-0.0011	0.0299	0.0405	5.2%	1.9992	-0.0008	0.0050	0.0069	5.1%	2.3378	0.3378	0.3389	0.0363	100%
100	100	2.0001	0.0001	0.0568	0.0735	5.5%	1.9993	-0.0007	0.0072	0.0095	5.3%	1.9995	-0.0005	0.0063	0.0085	5.6%
	200	2.0002	0.0002	0.0425	0.0570	5.2%	1.9997	-0.0003	0.0051	0.0069	4.9%	2.2702	0.2702	0.2715	0.0349	100%
	500	2.0010	0.0010	0.0176	0.0234	5.7%	1.9997	-0.0003	0.0031	0.0039	4.5%	2.2704	0.2704	0.2711	0.0252	100%

Table 5: Simulation results for ρ and β of DGP 5

ρ		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	0.4976	-0.0024	0.0372	0.0513	4.7%	0.4970	-0.0030	0.0134	0.0164	6.5%	0.4364	-0.0636	0.0717	0.0434	47%
10	200	0.4947	-0.0053	0.0302	0.0400	4.2%	0.4961	-0.0039	0.0121	0.0152	6.2%	0.4169	-0.0831	0.0875	0.0358	86%
	500	0.4963	-0.0037	0.0188	0.0250	5.1%	0.4976	-0.0024	0.0071	0.086	6.6%	0.4181	0.0819	0.0839	0.0236	99%
	100	0.4972	-0.0028	0.0376	0.0518	4.9%	0.4968	-0.0032	0.0100	0.0130	5.9%	0.4234	-0.0766	0.0851	0.0543	55%
20	200	0.4987	-0.0013	0.0273	0.0351	4.9%	0.4982	-0.0018	0.0074	0.0093	6.2%	0.4274	-0.0726	0.0776	0.0365	76%
	500	0.4992	-0.0008	0.0170	0.0221	5.4%	0.4990	-0.0010	0.0044	0.0056	6.3%	0.4279	-0.0721	0.0741	0.0231	99%
	100	0.4914	-0.0086	0.0469	0.0615	5.1%	0.4950	-0.0050	0.0087	0.0095	10%	0.4227	-0.0773	0.0855	0.0489	55%
50	200	0.4975	-0.0025	0.0322	0.0419	5.4%	0.4976	-0.0024	0.0055	0.0069	7.3%	0.4260	-0.0740	0.0786	0.0358	80%
	500	0.4995	-0.0005	0.0199	0.0273	5.1%	0.4990	-0.0010	0.0032	0.0040	6.8%	0.4269	-0.0731	0.0750	0.0229	99%
100	100	0.4891	-0.0109	0.0482	0.0595	5.5%	0.4952	-0.0048	0.0068	0.0066	17%	0.4996	-0.0004	0.0046	0.0063	5%
	200	0.4968	-0.0032	0.0313	0.0434	5.1%	0.4979	-0.0021	0.0039	0.0045	10%	0.4237	-0.0763	0.0802	0.0336	86%
	500	0.4994	-0.0006	0.0192	0.0262	5.2%	0.4992	-0.0008	0.0022	0.0026	6.5%	0.4244	-0.0756	0.0773	0.0226	99%

β		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	1.0046	0.0046	0.0679	0.0897	4.9%	1.0043	0.0043	0.0269	0.0360	4.7%	1.3352	0.3352	0.3395	0.0716	100%
10	200	1.0105	0.0105	0.0580	0.0760	6.4%	1.0090	0.0090	0.0253	0.0280	5.7%	1.4374	0.4374	0.4405	0.0723	100%
	500	1.0085	0.0085	0.0388	0.0497	5.2%	1.0053	0.0053	0.0142	0.0183	7%	1.4377	0.4377	0.4393	0.0483	100%
	100	1.0025	0.0025	0.0673	0.0912	5.5%	1.0038	0.0038	0.0195	0.0264	5.4%	1.4276	0.4276	0.4321	0.0849	100%
20	200	1.0035	0.0035	0.0487	0.0625	6.1%	1.0016	0.0016	0.0137	0.0197	4.6%	1.4233	0.4233	0.4258	0.0626	100%
	500	1.0021	0.0021	0.0310	0.0430	4.9%	1.0013	0.0013	0.0083	0.0111	4.2%	1.4247	0.4247	0.4260	0.0433	100%
	100	1.0093	0.0093	0.0833	0.1109	5.1%	1.0069	0.0069	0.0163	0.0204	8.2%	1.4730	0.4730	0.4764	0.0772	100%
50	200	1.0015	0.0015	0.0585	0.0765	5.8%	1.0027	0.0027	0.0106	0.0142	6%	1.4736	0.4736	0.4756	0.0585	100%
	500	1.0012	0.0012	0.0365	0.0469	5.8%	1.0017	0.0017	0.0063	0.0083	6.2%	1.4758	0.4758	0.4766	0.0378	100%
100	100	1.0070	0.0070	0.0745	0.0748	4.7%	1.0053	0.0053	0.0108	0.0128	8.4%	1.0005	0.0005	0.0090	0.0122	5.5%
	200	1.0054	0.0054	0.0535	0.0707	5.1%	1.0024	0.0024	0.0069	0.0090	5.6%	1.5003	0.5003	0.5020	0.0533	100%
	500	1.0015	0.0015	0.0336	0.0453	4.6%	1.0007	0.0007	0.0042	0.0054	5.1%	1.5016	0.5016	0.5024	0.0375	100%

Table 6: Simulation results for ρ and β of DGP 6

ρ		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	0.4930	-0.0070	0.0422	0.0589	5.4%	0.4938	-0.0062	0.0163	0.0207	6.4%	0.5747	0.0747	0.0916	0.0709	27%
10	200	0.4956	-0.0044	0.0296	0.0395	5.2%	0.4961	-0.0039	0.0112	0.0153	5.7%	0.5823	0.0823	0.0910	0.0519	54%
	500	0.4968	-0.0032	0.0189	0.0244	5.9%	0.4975	-0.0025	0.0070	0.0091	5.7%	0.5850	0.0850	0.0888	0.0355	92%
	100	0.4968	-0.0032	0.0399	0.0550	4.6%	0.4949	-0.0051	0.0106	0.0129	8.4%	0.6081	0.1081	0.1220	0.0781	47%
20	200	0.4978	-0.0022	0.0266	0.0370	4.7%	0.4978	-0.0022	0.0070	0.0090	6.4%	0.6158	0.1158	0.1226	0.0549	81%
	500	0.4986	-0.0014	0.0163	0.0219	5.6%	0.4990	-0.0010	00042	0.0053	7%	0.6205	0.1205	0.1235	0.0365	100%
	100	0.4897	-0.0103	0.0456	0.0572	5.8%	0.4932	-0.0068	0.0097	0.0097	18%	0.6025	0.1205	0.1165	0.0554	100%
50	200	0.4970	-0.0030	0.0312	0.0433	4.9%	0.4969	-0.0031	0.0055	0.0065	10%	0.6136	0.1136	0.1214	0.0613	74%
	500	0.4987	-0.0013	0.0203	0.0278	5.3%	0.4987	-0.0013	0.0032	0.0040	7.8%	0.6162	0.1162	0.1194	0.0391	100%
100	100	0.4862	-0.0138	0.0473	0.0599	5.6%	0.4934	-00066	0.0081	0.0066	28%	0.4983	-0.0017	0.0058	0.0060	2.6%
	200	0.4954	-0.0046	0.0302	0.0407	5.4%	0.4969	-0.0031	0.0045	0.0043	16%	0.6186	0.1186	0.1.261	0.0607	78%
	500	0.4990	-0.0010	0.0197	0.0269	5%	0.4988	-0.0012	0.0023	0.0027	9.2%	0.6217	0.1217	0.1250	0.0388	99%

β		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	1.0142	0.0142	0.0801	0.1021	5.9%	1.0113	0.0113	0.0310	0.0386	6.1%	1.1703	0.1703	0.1925	0.1154	52%
10	200	1.0100	0.0100	0.0594	0.0771	6.1%	1.0080	0.0080	0.0225	0.0270	6.6%	1.1588	0.1588	0.1730	0.0916	66%
	500	1.0100	0.0100	0.0390	0.0515	6.2%	1.0066	0.0066	0.0142	0.0175	7.9%	1.1551	0.1551	0.1616	0.0609	92%
	100	1.0038	0.0038	0.0661	0.0885	5.9%	1.0055	0.0055	0.0195	0.0251	6.2%	1.1499	0.1499	0.1706	0.1105	47%
20	200	1.0017	0.0017	0.0486	0.0638	5.5%	1.0024	0.0024	0.0134	0.0174	5.1%	1.1446	0.1446	0.1577	0.0855	64%
	500	1.0030	0.0030	0.0327	0.0464	4.5%	1.0016	0.0016	0.0082	0.0104	5.8%	1.1425	0.1425	0.1489	0.0561	91%
	100	1.0092	0.0092	0.0802	0.1062	5.5%	1.0084	0.0084	0.0167	0.0197	9.7%	1.2105	0.2105	0.2240	0.1021	79%
50	200	1.0031	0.0031	0.0585	0.0785	4.7%	1.0036	0.0036	0.0103	0.0133	6.1%	1.2052	0.2052	0.2138	0.0797	92%
	500	1.0025	0.0025	0.0375	0.0511	5.1%	1.0024	0.0024	0.0066	0.0080	6.8%	1.2069	0.2069	0.2110	0.0534	100%
100	100	1.0094	0.0094	0.0767	0.1006	5.3%	1.0071	0.0071	0.0118	0.0129	11%	1.0025	0.0025	0.0097	0.0118	5%
	200	1.0065	0.0065	0.0554	0.0751	4.6%	1.0033	0.0033	0.0073	0.0084	8.5%	1.2118	0.2118	0.2202	0.0805	93%
	500	1.0011	0.0011	0.0530	0.0531	4.8%	1.0013	0.0013	0.0042	0.0051	5.7%	1.2116	0.2116	0.2151	0.0512	100%

References

[1] Ahn, S.C., and A.R., Horenstein, 2013, Eigenvalue ratio test for the number of factors. *Econometrica* 81, 1203–1227.

[2] Ahn, S.C., Y.H. Lee and P. Schmidt, 2001, GMM estimation of linear panel data models with time-varying individual effects, *Journal of Econometrics* 101, 219-255.

[3] Ahn, S.C., Y.H. Lee and P. Schmidt, 2013, Panel data models with multiple time-varying individual effects, *Journal of Econometrics* 174, 1–14.

[4] Anderson, T.W., and H. Rubin. 1956, Statistical Inference in Factor Analysis, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Volume 5, 111-150.

[5] Bai, J., 2009, Panel data models with interactive effects, *Econometrica* 77, 1229-1279.

[6] Bai, Z., and Y.Q., Yin, 1993, Limit of the Smallest Eigenvalue of a Large Dimensional Sample Covariance Matrix, *The Annals of Probability*, 21, 1275–1294.

[7] Chudik, A., M.H., Pesaran and E. Tosetti, 2011, Weak and Strong Cross Section Dependence and Estimation of Large Panels, The Econometrics Journal 14, 45-90.

[8] Gorski, J., F. Pfeuffer and K. Klamroth, 2007, Biconvex sets and optimization with biconvex functions: a survey and extensions, *Mathematical Methods of Operations Research* 66, 373–407.

[9] Hsiao, C., 2014, *Analysis of Panel Data*, 3rd Edition, Cambridge University Press.

[10] Hsiao, C., 2018, Panel Models with Interactive Effects, *Journal of Econometrics* 206, 645-673.

[11] Hsiao, C., and J. Zhang, 2015. IV, GMM or MLE to estimate dynamic panel data models. *Journal of Econometrics* 187, 312-322.

[12] Hsiao, C., and Q. Zhou, 2015, Statistical inference for panel dynamic simultaneous equations models. *Journal of Econometrics* 189, 383-396.

[13] Hsiao, C., and Q. Zhou, 2018, Incidental parameters, initial conditions and sample size for dynamic panel data models. *Journal of Econometrics*, 207, 114-128.

[14] Hsiao, C., and Q. Zhou, 2019, Panel Parametric, Semi-parametric and Nonparametric Construction of Counterfactuals, *Journal of Applied Econometrics* 34, 463-481.

[15] Jeganathan, P., 1980, An extension of a result of L. LeCam concerning asymptotic normality, *Sankhyā* Series A., 42, 146-160.

[16] Jeganathan, P., 1982, On some asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal, *Sankhyā* Series A., 44, 173-212.

[17] Jiang, B., Y. Yang, J. Gao and C. Hsiao, 2019, Recursive Estimation in Large Panel Data Models: Theory and Practice, working paper.

[18] Moon, R., and M, Weidner, 2015, Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects, *Econometrica* 83, 1543-1579.

[19] Moon, R., and M, Weidner, 2017, Dynamic Linear Panel Regression Models with Interactive Fixed Effects, *Econometric Theory* 33, 158-195.

[20] Pesaran, M.H., 2006, Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure, *Econometrica* 74, 967-1012.

[21] Phillips, P.C.B, 1989. Partially identified econometric models. *Econometric Theory*, 5(2), pp.181-240.

[22] Phillips, P.C.B, and R. Moon, 1999, Linear Regression Limit Theory for Nonstationary Panel Data, *Econometrica* 67, 1057-1111.

[23] Rothenberg, T., 1971, Identification in Parametric Models, *Econometrica* 39, 577-591.

[24] White, H., 2001, *Asymptotic Theory for Econometricians,* Revised Edition, Emerald Group Publishing Limited.

Appendix: Mathematical Derivations and Additional Simulation Results

This appendix contains the mathematical derivations of the main results in the paper. Let $tr(\mathbf{A})$ denotes the trace of matrix $\mathbf{A}$, $\lambda_{\min}(\mathbf{A})$ denotes the minimum eigenvalue of $\mathbf{A}$, and the norm of the matrix $\mathbf{A}$ is defined as $\|\mathbf{A}\| = \sqrt{tr(\mathbf{A}'\mathbf{A})}$. In what follows, we assume there exists a generic finite positive constant C whose value doesn't depend on N and T, and may change case by case.

We first provide the proofs of the results in the main paper and then provide Additional Simulation Results.

A. Proofs of the Propositions

Proof of Proposition 3.1. The proof follows straightforwardly from the application of standard central limit theorem (e.g., White (2001)), since under Assumption A1 and A4, $\mathbf{w}'\mathbf{u}_t$ is uncorrelated with $\mathbf{w}'\mathbf{X}_t$ and is i.i.d. over t with $E(\mathbf{w}'\mathbf{u}_t) = 0$. ■

Proof of Proposition 3.2. The Lagrangian of (3.6) subject to (3.7) takes the form

$$L = -\frac{T}{2}\log\sigma_w^2 - \frac{1}{2\sigma_w^2}\sum_{t=1}^{T}\mathbf{w}'(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta})(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta})'\mathbf{w} + \eta\left(\mathbf{w}'\Omega\mathbf{w} - 1\right). \tag{A.1}$$

The first order conditions are (3.4) and

$$\hat{\sigma}_w^2 = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{\mathbf{w}}'\left(\mathbf{y}_t - \mathbf{X}_t\hat{\boldsymbol{\beta}}\right)\right)^2 = 1, \tag{A.2}$$

$$\left[\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{y}_t - \mathbf{X}_t\hat{\boldsymbol{\beta}}\right)\left(\mathbf{y}_t - \mathbf{X}_t\hat{\boldsymbol{\beta}}\right)' - \delta\Omega\right]\hat{\mathbf{w}} = 0, \tag{A.3}$$

where $\delta = \frac{1}{T}\eta\hat{\sigma}_w^2$. The eigenvector of (A.3) is identical to the eigenvector of (3.8). Substituting (A.2) and (A.3) into the log-likelihood function (3.6) yields

$$L_C = -\frac{T}{2}(\ln\delta + 1), \tag{A.4}$$

which is maximized by letting $\mathbf{w}$ correspond to the smallest root of (3.8) or (A.3). ■

Proof of Proposition 3.3. Since as $T \to \infty$,

$$\begin{aligned} S &= \frac{1}{T}\sum_{t=1}^{T} \hat{\mathbf{w}}' \left(\mathbf{y}_t - \mathbf{X}_t\hat{\boldsymbol{\beta}}\right)\left(\mathbf{y}_t - \mathbf{X}_t\hat{\boldsymbol{\beta}}\right)' \hat{\mathbf{w}} \\ &= \frac{1}{T}\sum_{t=1}^{T} \hat{\mathbf{w}}' \left[\mathbf{X}_t\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) + \Lambda\mathbf{f}_t + \mathbf{u}_t\right]\left[\mathbf{X}_t\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) + \Lambda\mathbf{f}_t + \mathbf{u}_t\right]' \hat{\mathbf{w}} \\ &= \frac{1}{T}\sum_{t=1}^{T} \hat{\mathbf{w}}' \left[\mathbf{X}_t\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) + \Lambda\mathbf{f}_t\right]\left[\mathbf{X}_t\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) + \Lambda\mathbf{f}_t\right]' \hat{\mathbf{w}} + \hat{\mathbf{w}}'\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{u}_t\mathbf{u}_t'\right)\hat{\mathbf{w}} + o_p(1) \\ &= S^*\left(\hat{\mathbf{w}}, \hat{\boldsymbol{\beta}}\right) + \hat{\mathbf{w}}'\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{u}_t\mathbf{u}_t'\right)\hat{\mathbf{w}} + o_p(1). \end{aligned} \quad \text{(A.5)}$$

where the third equality follows from Assumption A1 as $\mathbf{u}_t$ is mean independent of $(\mathbf{X}_t, \mathbf{f}_t)$, and

$$S^*\left(\hat{\mathbf{w}}, \hat{\boldsymbol{\beta}}\right) = \frac{1}{T}\sum_{t=1}^{T} \hat{\mathbf{w}}' \left[\mathbf{X}_t\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) + \Lambda\mathbf{f}_t\right]\left[\mathbf{X}_t\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) + \Lambda\mathbf{f}_t\right]' \hat{\mathbf{w}} \geq 0, \quad \text{(A.6)}$$

where $\left(\hat{\mathbf{w}}, \hat{\boldsymbol{\beta}}\right)$ is a solution of $S^*(\mathbf{w}, \boldsymbol{\beta})$ subject to (3.7). We note that the objective function (A.6) is nonconvex in $\left(\hat{\mathbf{w}}, \hat{\boldsymbol{\beta}}, \Lambda, \mathbf{F}\right)$.

We argue the consistency of the estimators by contraposition. Obviously, the minimal value of S^* is achieved if $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$ and $\hat{\mathbf{w}}'\Lambda = 0$. Now suppose $\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right\| \geq \epsilon > 0$, then $S^*\left(\hat{\mathbf{w}}, \hat{\boldsymbol{\beta}}\right) = 0$ if and only if

$$-\hat{\mathbf{w}}'\mathbf{X}_t\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) = \hat{\mathbf{w}}'\Lambda\mathbf{f}_t. \quad \text{(A.7)}$$

Let $\hat{\mathbf{w}}'\Lambda = \mathbf{c}'$, and $-\hat{\mathbf{w}}'\mathbf{X}_t\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) = \mathbf{a}_t$, (A.7) is rewritten as

$$\mathbf{c}'\mathbf{f}_t = \mathbf{a}_t, \quad \text{for each } t = 1, \ldots, T. \quad \text{(A.8)}$$

Since both $\mathbf{c}$ and $\mathbf{a}_t$ are constants given $\mathbf{X}_t$ and $\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)$, for given t, $\mathbf{f}_t$ is also a given constant vector, not an arbitrary constant vector. The sequence of realized $\mathbf{f}_t$ that can satisfy (A.8) for every t if and only if $\mathbf{c} = \mathbf{0}$ and $\mathbf{a}_t = 0$. If $\mathbf{c} = \mathbf{0}$, then $\hat{\mathbf{w}}'\Lambda = 0$. Substituting $\hat{\mathbf{w}}'\Lambda = 0$ into (A.6) yields

$$S^*\left(\hat{\mathbf{w}}, \hat{\boldsymbol{\beta}}\right) = \left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)'\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t'\hat{\mathbf{w}}\hat{\mathbf{w}}'\mathbf{X}_t\right)\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) \geq \epsilon^2\lambda_{\min}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t'\hat{\mathbf{w}}\hat{\mathbf{w}}'\mathbf{X}_t\right). \quad \text{(A.9)}$$

Under Assumption A4

$$rank\left(plim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t'\hat{\mathbf{w}}\hat{\mathbf{w}}'\mathbf{X}_t\right) = k, \quad \text{(A.10)}$$

and it implies that $S^*\left(\hat{\mathbf{w}},\hat{\boldsymbol{\beta}}\right)$ in (A.9) strictly positive. It violates the fact that $\hat{\boldsymbol{\beta}}$ is the minimizer, as the minimum of $S^*(\mathbf{w},\boldsymbol{\beta})=0$ can be achieved only when $\hat{\boldsymbol{\beta}}\to_p\boldsymbol{\beta}$.

Furthermore, when $\hat{\boldsymbol{\beta}}\to_p\boldsymbol{\beta}$, (A.5) converges to

$$\hat{\mathbf{w}}'\left(\Omega+\Lambda\Lambda'\right)\hat{\mathbf{w}}\geq\hat{\mathbf{w}}'\Omega\hat{\mathbf{w}}\ \ (=1), \tag{A.11}$$

where the equality holds iff $\hat{\mathbf{w}}'\Lambda=0$.

In order to show that $\hat{\mathbf{w}}$ and $\hat{\boldsymbol{\beta}}$ are asymptotically independent, we first decompose

$$\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{y}_t-\mathbf{X}_t\hat{\boldsymbol{\beta}}\right)\left(\mathbf{y}_t-\mathbf{X}_t\hat{\boldsymbol{\beta}}\right)'=\frac{1}{T}\sum_{t=1}^{T}\left(\Lambda\mathbf{f}_t+\mathbf{u}_t\right)\left(\Lambda\mathbf{f}_t+\mathbf{u}_t\right)'+\mathbf{R}_\beta, \tag{A.12}$$

where $\mathbf{R}_\beta=-\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\left(\Lambda\mathbf{f}_t+\mathbf{u}_t\right)'-\frac{1}{T}\sum_{t=1}^{T}\left(\Lambda\mathbf{f}_t+\mathbf{u}_t\right)\left(\mathbf{X}_t(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\right)'+\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{X}_t'=o_p(1)$ by the consistency of $\hat{\boldsymbol{\beta}}$.

We note that the eigenvector is estimated as

$$\begin{aligned}\hat{\mathbf{w}}&=\underset{\mathbf{w}'\Omega\mathbf{w}=1}{\arg\min}\,\mathbf{w}'\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{y}_t-\mathbf{X}_t\hat{\boldsymbol{\beta}}\right)\left(\mathbf{y}_t-\mathbf{X}_t\hat{\boldsymbol{\beta}}\right)'\mathbf{w}\\&=\underset{\mathbf{w}'\Omega\mathbf{w}=1}{\arg\min}\,\mathbf{w}'\left[\frac{1}{T}\sum_{t=1}^{T}\left(\Lambda\mathbf{f}_t+\mathbf{u}_t\right)\left(\Lambda\mathbf{f}_t+\mathbf{u}_t\right)'\right]\mathbf{w}+\mathbf{w}'\mathbf{R}_\beta\mathbf{w}.\end{aligned}$$

Since $\hat{\boldsymbol{\beta}}$ is only involved in R_β, the asymptotic distribution of $\hat{\mathbf{w}}$ is determined by $\frac{1}{T}\sum_{t=1}^{T}\left(\Lambda\mathbf{f}_t+\mathbf{u}_t\right)\left(\Lambda\mathbf{f}_t+\mathbf{u}_t\right)'$ but not affected by the realization of $\hat{\boldsymbol{\beta}}$. In other words,

$$\begin{aligned}\hat{\mathbf{w}}|\hat{\boldsymbol{\beta}}&\to_d\lim_{T\to\infty}\underset{\mathbf{w}'\Omega\mathbf{w}=1}{\arg\min}\,\mathbf{w}'\left[\frac{1}{T}\sum_{t=1}^{T}\left(\Lambda\mathbf{f}_t+\mathbf{u}_t\right)\left(\Lambda\mathbf{f}_t+\mathbf{u}_t\right)'\right]\mathbf{w}\\&\sim\lim_{T\to\infty}\underset{\mathbf{w}'\Omega\mathbf{w}=1}{\arg\min}\,\mathbf{w}'\left[\frac{1}{T}\sum_{t=1}^{T}\left(\Lambda\mathbf{f}_t^*+\mathbf{u}_t^*\right)\left(\Lambda\mathbf{f}_t^*+\mathbf{u}_t^*\right)'\right]\mathbf{w}\end{aligned}$$

where $(\mathbf{f}_t^*,\mathbf{u}_t^*)$ is an independent copy of $(\mathbf{f}_t,\mathbf{u}_t)$. An "independent copy" means that $(\mathbf{f}_t^*,\mathbf{u}_t^*)$ follows the same distribution as $(\mathbf{f}_t,\mathbf{u}_t)$, satisfies Assumption A1-A4, but $(\mathbf{f}_t^*,\mathbf{u}_t^*)$ and $(\mathbf{f}_t,\mathbf{u}_t)$ are statistically independent. Therefore, $\hat{\mathbf{w}}$ and $\hat{\boldsymbol{\beta}}$ is asymptotically independent. ■

Proof of Proposition 4.1. We note that the explicit solution $\hat{\boldsymbol{\beta}}^{(j)}=\left(\Xi_T^{(j)}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t'\mathbf{w}_j\mathbf{w}_j'\mathbf{y}_t$

implies that the average estimator can be rewritten as

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^{Ave} - \boldsymbol{\beta} &= \frac{1}{N-r}\sum_{j=1}^{N-r}\left(\hat{\boldsymbol{\beta}}^{Ave} - \boldsymbol{\beta}\right) \\
&= \frac{1}{N-r}\sum_{j=1}^{N-r}\left[(\Xi_T^{(j)})^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t'\mathbf{w}_j\mathbf{w}_j'\mathbf{u}_t\right] \\
&= \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{N-r}\sum_{j=1}^{N-r}(\Xi_T^{(j)})^{-1}\mathbf{X}_t'\mathbf{w}_j\mathbf{w}_j'\mathbf{u}_t\right] \\
&= \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\xi}_t.
\end{aligned}$$

Under Assumption A1, the random variable $\boldsymbol{\xi}_t$ is i.n.i.d. across t and so does its linear combination $Z_{\lambda,N,t}$. The mean $E\left[Z_{\lambda,N,t}\right] = 0$ and the variance

$$\begin{aligned}
E\left[Z_{\lambda,N,t}^2\right] &= (N-r)^{-1}\boldsymbol{\lambda}'\left(\sum_{j=1}^{N-r}E\left[(\Xi_T^{(j)})^{-1}\mathbf{X}_t'\mathbf{w}_j\mathbf{w}_j'\mathbf{u}_t\mathbf{u}_t'\mathbf{w}_j\mathbf{w}_j'\mathbf{X}_t(\Xi_T^{(j)})^{-1}\right]\right)\boldsymbol{\lambda} \\
&\quad + (N-r)^{-1}\boldsymbol{\lambda}'\left(\sum_{j=1}^{N-r}\sum_{l=1,l\neq j}^{N-r}E\left[(\Xi_T^{(j)})^{-1}\mathbf{X}_t'\mathbf{w}_j\mathbf{w}_j'\mathbf{u}_t\mathbf{u}_t'\mathbf{w}_l\mathbf{w}_l'\mathbf{X}_t(\Xi_T^{(j)})^{-1}\right]\right)\boldsymbol{\lambda} \\
&= \frac{1}{(N-r)}\boldsymbol{\lambda}'\left(\sum_{j=1}^{N-r}(\Xi_T^{(j)})^{-1}\mathbf{X}_t'\mathbf{w}_j\mathbf{w}_j'\mathbf{X}_t(\Xi_T^{(j)})^{-1}\right)\boldsymbol{\lambda},
\end{aligned}$$

where the second equality follows by $\mathbf{w}_j'\Omega\mathbf{w}_j = 1$ and $\mathbf{w}_j'\Omega\mathbf{w}_l = 0$ for $j \neq l$. Since $\boldsymbol{\xi}_t$ are independent across t, then

$$\begin{aligned}
Var\left(\sum_{t=1}^{T}Z_{\lambda,N,t}\right) &= \frac{1}{(N-r)}\boldsymbol{\lambda}'\left(\sum_{t=1}^{T}\sum_{j=1}^{N-r}(\Xi_T^{(j)})^{-1}\left(\mathbf{X}_t'\mathbf{w}_j\mathbf{w}_j'\mathbf{X}_t\right)(\Xi_T^{(j)})^{-1}\right)\boldsymbol{\lambda} \\
&= \frac{T}{(N-r)}\boldsymbol{\lambda}'\left(\sum_{j=1}^{N-r}(\Xi_T^{(j)})^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{X}_t'\mathbf{w}_j\mathbf{w}_j'\mathbf{X}_t\right)(\Xi_T^{(j)})^{-1}\right)\boldsymbol{\lambda} \\
&= \frac{T}{(N-r)}\boldsymbol{\lambda}'\left(\sum_{j=1}^{N-r}(\Xi_T^{(j)})^{-1}\right)\boldsymbol{\lambda} \\
&= T\boldsymbol{\lambda}'\Sigma_{NT}\boldsymbol{\lambda} = T\sigma_{\lambda,NT}^2.
\end{aligned}$$

Consequently, we obtain

$$\begin{aligned}
\sqrt{(N-r)\,T}\frac{\boldsymbol{\lambda}'\left(\hat{\boldsymbol{\beta}}^{Ave}-\boldsymbol{\beta}\right)}{\sigma_{\lambda,NT}} &= \frac{1}{\sqrt{T}\sigma_{\lambda,NT}}\sum_{t=1}^{T}\sqrt{N-r}\boldsymbol{\lambda}'\boldsymbol{\xi}_t \\
&= \frac{1}{\sqrt{T}\sigma_{\lambda,NT}}\sum_{t=1}^{T}Z_{\lambda,Nt},
\end{aligned}$$

where by the Lindeberg condition (4.5), we invoke the Lindeberg-Feller central limit theorem for i.n.i.d sequence of random variables $T^{-1/2}\sigma_{\lambda,N,T}^{-1}z_{\lambda,N,t}$ and conclude (e.g., White (2011))

$$\frac{1}{\sqrt{T}\sigma_{\lambda,N,T}}\sum_{t=1}^{T}Z_{\lambda,N,t}\rightarrow_d N\left(0,1\right),$$

as $N,T\rightarrow\infty$. Since k is a fixed constant, the Cramér-Wold-device implies

$$\Sigma_{NT}^{-1/2}\sqrt{(N-r)\,T}\left(\hat{\boldsymbol{\beta}}^{Ave}-\boldsymbol{\beta}\right)\rightarrow_d N(0,\mathbf{I}_k),$$

as required. ■

B. Additional Simulation Results

Using the same notations in the Monte Carlo Simulation section, here we consider five additional DGPs to investigate the finite sample performance of the transformed estimation.

For DGP A1-A4, we assume y_{it} is

$$y_{it}=x_{1,it}\beta_1+x_{2,it}\beta_2+\lambda_{1,i}f_{1,t}+\lambda_{2,i}f_{2,t}+u_{it}, \tag{B.1}$$

where $\beta_1=1$ and $\beta_2=2$, and $u_{it}\sim IIDN\left(0,\sigma_{u,i}^2\right)$ with $\sigma_{u,i}^2$ being independent draws from $\left(1+0.5\chi^2\left(2\right)\right)$. The covariates x_{it} is generated as

DGP A1:

$$x_{k,it}=1+\alpha_{ki}+c_{k1,i}\lambda_{1,i}+c_{k2,i}\lambda_{2,i}+\eta_{k,it},\ k=1,2. \tag{B.2}$$

DGP A2:

$$x_{k,it}=1+\alpha_{ki}+\eta_{k,it},\ k=1,2. \tag{B.3}$$

The idiosyncratic errors $\eta_{k,it}$ of $x_{k,it}$ is generated as

$$\eta_{k,it}=\rho_{k,i}\eta_{k,it-1}+v_{k,it},\ k=1,2,$$

with $\rho_{k,i}$ are i.i.d draws from $U\left(0.1,0.9\right)$ for $k=1,2$ and $i=1,\ldots,N$. We also assume $\alpha_{1i},\alpha_{2i}\sim IIDN\left(0,1\right)$, and $v_{j,it}\sim IIDN\left(0,\sigma_{vj,i}^2\right)$ for $j=1,2,3$ and $\sigma_{v1,i}^2$, $\sigma_{v2,i}^2$ are independent draws from $\left(1+0.5\chi^2(2)\right)$. For the factors, we assume $f_{1t} \sim IIDN(0,1)$ and $f_{2t} \sim IIDN(0,2)$. For the factor loadings, we assume that $\lambda_{1,i}, \lambda_{2,i}$ are i.i.d draws from $N(1,1)$ and $c_{k1,i}$ and $c_{k2,i}$ are i.i.d draw from $U(0,1)$, for $i = 1, 2, \ldots, N$, $k = 1, 2$.

DGP A3: (Cointegrated factors) We assume y_{it} and $x_{k,it}$ are generated as in DGP1 in the main paper, except now we let the factors are generated as

$$\begin{aligned} f_{1t} &= f_{2t} + \xi_{1t}, \\ f_{2t} &= 0.5 f_{2t-1} + \xi_{2t}, \end{aligned} \tag{B.4}$$

where ξ_{1t} and ξ_{2t} are i.i.d $N(0,1)$.

DGP A4: (Non-normal errors) We assume y_{it} and $x_{k,it}$ are generated as in DGP1 in the main paper, except now we let u_{it} be generated as

$$u_{it} \sim IID\chi^2(1) - 1,$$

for $i = 1, \ldots, N; t = 1, \ldots, T$.

DGP A5: Dynamic model (5.6) with cointegrated factors as in (B.4).

The simulation results of DGP A1-A5 are summarized in Table A1-A5. Similar finding as in the main paper can be observed from these simulations, i.e., the TQMLE performs quite robust, regardless of whether model is static or dynamic, or the generation of the covariates x_{it}, or whether the common factors are cointegrated, or whether is not normally distributed. The empirical size is also very close to the nominal value in all our designs. For the average estimator, even if it is more efficient than the TQMLE with a much smaller IQR, while the size is not as steady as the TQMLE, suggesting the possibility that its performance depends on the determination of the number of unknown common factors or the asymptotical bias in the estimation.

Table A1: Simulation results for β_1 and β_2 of DGP A1

β_1		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	0.9953	-0.0047	0.0666	0.0831	5.8%	1.0005	0.0005	0.0220	0.0304	5.5%	0.9974	-0.0026	0.0422	0.0567	4.2%
10	200	0.9994	-0.0006	0.0433	0.0549	5.7%	1.0000	0.0000	0.0150	0.0214	4.7%	0.9994	-0.0006	0.0295	0.0386	5.6%
	500	1.0009	0.0009	0.0278	0.0354	5.5%	1.0004	0.0004	0.0095	0.0129	4.7%	0.9999	-0.0001	0.0191	0.0243	5.1%
	100	0.9958	-0.0042	0.0608	0.0758	5.4%	1.0002	0.0002	0.0146	0.0201	6%	0.9964	-0.0036	0.0358	0.0447	5.7%
20	200	1.0017	0.0017	0.0401	0.0503	6.2%	0.9998	-0.0002	0.0099	0.0134	5.3%	1.0017	0.0017	0.0247	0.0315	4.7%
	500	1.0001	0.0001	0.0247	0.0325	5.2%	1.0001	0.0001	0.0064	0.0081	4.6%	0.9997	-0.0003	0.0166	0.0215	5.7%
	100	0.9987	-0.0013	0.0701	0.0847	6.1%	0.9997	-0.0003	0.0109	0.0147	5.5%	1.0007	0.0007	0.0314	0.0412	5%
50	200	1.0012	0.0012	0.0479	0.0573	5.9%	0.9995	-0.0005	0.0075	0.0102	5.2%	1.0006	0.0006	0.0226	0.0276	5.3%
	500	0.9973	-0.0027	0.0294	0.0360	5.9%	1.0001	0.0001	0.0046	0.0061	5.8%	0.9998	-0.0002	0.0140	0.0176	5.4%
100	100	1.0007	0.0007	0.0681	0.0835	5.4%	0.9999	-0.0001	0.0074	0.0096	5.3%	1.0000	0.0000	0.0060	0.0079	5%
	200	1.0012	0.0012	0.0474	0.0668	4.7%	0.9999	-0.0001	0.0050	0.0071	5.1%	0.9990	-0.0010	0.0185	0.0237	5.3%
	500	0.9996	-0.0004	0.0303	0.0364	6.5%	1.0000	0.0000	0.0032	0.0043	4.8%	0.9995	-0.0005	0.0120	0.0163	4.9%
β_2		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	2.0048	0.0048	0.0681	0.0891	6.2%	2.0006	0.0006	0.0239	0.0306	6%	1.9992	-0.0008	0.0421	0.0528	6.4%
10	200	2.0010	0.0010	0.0454	0.0573	5.5%	2.0002	0.0002	0.0156	0.0216	5.1%	1.9991	-0.0009	0.0296	0.0392	4.6%
	500	1.9981	-0.0019	0.0307	0.0380	5.6%	1.9991	-0.0009	0.0099	0.0129	6.2%	1.9999	-0.0001	0.0188	0.0232	5.9%
	100	2.0007	0.0007	0.0632	0.0846	4.4%	2.0002	0.0002	0.0152	0.0195	5.3%	1.9988	-0.0012	0.0378	0.0490	4.9%
20	200	1.9990	-0.0010	0.0426	0.0511	5.3%	1.9994	-0.0006	0.0099	0.0139	4.8%	2.0015	0.0015	0.0278	0.0358	5.4%
	500	2.0007	0.0007	0.0276	0.0350	5.2%	2.0003	0.0003	0.0068	0.0098	5.2%	1.9994	-0.0006	0.0175	0.0229	4.7%
	100	2.0039	0.0039	0.0680	0.0860	5.2%	2.0000	0.0000	0.0107	0.0139	5.4%	2.0013	0.0013	0.0307	0.0384	5.4%
50	200	1.9968	-0.0032	0.0479	0.0608	5.5%	1.9996	-0.0004	0.0074	0.0100	5.4%	2.0007	0.0007	0.0214	0.0291	5.7%
	500	2.0005	0.0005	0.0293	0.0370	5.3%	2.0002	0.0002	0.0045	0.0063	4.8%	2.0000	0.0000	0.0142	0.0193	5.5%
100	100	1.9993	-0.0007	0.0620	0.0754	5.3%	2.0003	0.0003	0.0070	0.0096	4.5%	2.0000	0.0000	0.0057	0.0075	5.4%
	200	1.9976	-0.0024	0.0445	0.0548	5.9%	1.9999	-0.0001	0.0046	0.0062	4.6%	1.9991	-0.0009	0.0162	0.0195	5.6%
	500	1.9992	-0.0008	0.0273	0.0333	5.6%	2.0000	0.0000	0.0029	0.0038	5.5%	1.9999	-0.0001	0.0104	0.0131	4.8%

Table A2: Simulation results for β_1 and β_2 of DGP A2

β_1		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	0.9962	-0.0038	0.0654	0.0789	5.3%	1.0003	0.0003	0.0228	0.0310	5.2%	0.9973	-0.0027	0.0445	0.0610	4.6%
10	200	1.0012	0.0012	0.0436	0.0512	5.2%	1.0001	0.0001	0.0157	0.0214	5.9%	0.9996	-0.0004	0.0317	0.0410	6%
	500	1.0013	0.0013	0.0287	0.0360	5.4%	1.0005	0.0005	0.0098	0.0131	5%	0.9998	-0.0002	0.0202	0.0261	4.4%
	100	0.9983	-0.0017	0.0605	0.0769	5.2%	1.0004	0.0004	0.0148	0.0200	6.2%	0.9980	-0.0020	0.0311	0.0423	4.8%
20	200	1.0013	0.0013	0.0389	0.0501	6.3%	0.9997	-0.0003	0.0101	0.0139	5%	1.0004	0.0004	0.0225	0.0304	5.5%
	500	0.9998	-0.0002	0.0247	0.0341	4.8%	1.0000	0.0000	0.0064	0.0086	4.6%	0.9997	-0.0003	0.0142	0.0187	5.2%
	100	0.9996	-0.0004	0.0720	0.0864	6.5%	0.9999	-0.0001	0.0111	0.0157	5.8%	0.9993	-0.0007	0.0214	0.0279	5.5%
50	200	0.9992	-0.0008	0.0489	0.0609	5.5%	0.9996	-0.0004	0.0079	0.0106	4.8%	0.9993	-0.0007	0.0148	0.0194	5.5%
	500	0.9986	-0.0014	0.0314	0.0379	6.4%	1.0001	0.0001	0.0047	0.0063	4.9%	0.9997	-0.0003	0.0095	0.0122	5.1%
100	100	0.9999	-0.0001	0.0691	0.0835	6.1%	1.0001	0.0001	0.0075	0.0103	4.8%	0.9997	-0.0003	0.0062	0.0085	4.7%
	200	1.0013	0.0013	0.0474	0.0614	6%	0.9999	-0.0001	0.0051	0.0070	5.8%	1.0000	0.0000	0.0110	0.0144	5.4%
	500	1.0007	0.0007	0.0301	0.0398	5.7%	1.0001	0.0001	0.0033	0.0044	5.4%	0.9998	-0.0002	0.0069	0.0092	5.7%
β_2		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	2.0037	0.0037	0.0673	0.0868	5.7%	2.0010	0.0010	0.0241	0.0311	5.6%	1.9989	-0.0011	0.0441	0.0566	5.7%
10	200	2.0002	0.0002	0.0453	0.0578	4.8%	2.0000	0.0000	0.0159	0.0221	5%	1.9989	-0.0011	0.0309	0.0429	4.1%
	500	1.9982	-0.0018	0.0305	0.0384	4.6%	1.9992	-0.0008	0.0098	0.0126	5.5%	1.9999	-0.0001	0.0201	0.0257	5.4%
	100	2.0024	0.0024	0.0643	0.0836	5.7%	2.0001	0.0001	0.0153	0.0203	4.4%	2.0008	0.0008	0.0327	0.0446	4.9%
20	200	1.9980	-0.0020	0.0423	0.0531	5.8%	1.9995	-0.0005	0.0100	0.0136	4.7%	1.9994	-0.0006	0.0242	0.0317	4%
	500	2.0007	0.0007	0.0282	0.0359	5.4%	2.0003	0.0003	0.0069	0.0101	4.4%	1.9993	-0.0007	0.0154	0.0209	5.1%
	100	2.0033	0.0033	0.0649	0.0855	4.5%	2.0000	0.0000	0.0107	0.0149	4.9%	1.9995	-0.0005	0.0203	0.0272	4.9%
50	200	1.9971	-0.0029	0.0464	0.0579	5.6%	1.9997	-0.0003	0.0073	0.0098	5.7%	1.9996	-0.0004	0.0153	0.0201	4.6%
	500	2.0013	0.0013	0.0298	0.0369	5.6%	2.0002	0.0002	0.0045	0.0064	4.6%	2.0003	0.0003	0.0094	0.0125	4.6%
100	100	2.0003	0.0003	0.0604	0.0781	5%	2.0003	0.0003	0.0071	0.0097	4.7%	2.0001	0.0001	0.0059	0.0080	4.8%
	200	1.9990	-0.0010	0.0454	0.0587	4.9%	1.9999	-0.0001	0.0047	0.0064	5%	1.9999	-0.0001	0.0105	0.0139	5.1%
	500	1.9997	-0.0003	0.0272	0.0356	5.3%	2.0001	0.0001	0.0029	0.0041	4.7%	2.0002	0.0002	0.0069	0.0092	5.1%

Table A3: Simulation results for β_1 and β_2 of DGP A3

β_1		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	1.0034	0.0034	0.0589	0.0790	4.4%	1.0024	0.0024	0.0224	0.0304	4.5%	1.2406	0.2406	0.2466	0.0745	99%
10	200	1.0085	0.0085	0.0449	0.0578	5.4%	1.0035	0.0085	0.0162	0.0225	5.2%	1.2450	0.2450	0.2490	0.0589	100%
	500	1.0043	0.0043	0.0297	0.0391	3.8%	1.0026	0.0026	0.0097	0.0132	5.7%	1.2450	0.2450	0.2480	0.0528	100%
	100	1.0010	0.0010	0.0539	0.0680	5.2%	1.0087	0.0087	0.0196	0.0234	7.9%	1.3705	0.3705	0.3758	0.0833	100%
20	200	1.0008	0.0008	0.0395	0.0491	5.2%	1.0075	0.0075	0.0146	0.0170	8.8%	1.3710	0.3710	0.3746	0.0749	100%
	500	1.0006	0.0006	0.0235	0.0305	5%	1.0070	0.0070	0.0104	0.0103	16%	1.3697	0.3697	0.3721	0.0594	100%
	100	1.0014	0.0014	0.0649	0.0824	5.1%	1.0072	0.0072	0.0154	0.0191	8.4%	1.4617	0.4617	0.4650	0.0733	100%
50	200	1.0019	0.0019	0.0464	0.0594	5.3%	1.0066	0.0066	0.0119	0.0131	9.5%	1.4710	0.4710	0.4729	0.0606	100%
	500	1.0007	0.0007	0.0299	0.0379	5.5%	1.0064	0.0064	0.0090	0.0083	17%	1.4719	0.4719	0.4730	0.0435	100%
100	100	1.0037	0.0037	0.0624	0.0766	6%	0.9988	-0.0012	0.0080	0.0108	5.1%	0.9982	-0.0018	0.0073	0.0092	6.2%
	200	1.0034	0.0034	0.0443	0.0572	5.8%	0.9964	-0.0036	0.0069	0.0072	9.8%	1.5404	0.5404	0.5420	0.0541	100%
	500	1.0004	0.0004	0.0272	0.0349	5.1%	0.9968	-0.0032	0.0050	0.0051	15%	1.5435	0.5435	0.5443	0.0429	100%

β_2		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	2.0064	0.0064	0.0707	0.0768	5.3%	2.0020	0.0020	0.0244	0.0328	5.5%	2.3763	0.3763	0.3812	0.0848	100%
10	200	2.0025	0.0025	0.0491	0.0608	5.7%	2.0016	0.0016	0.0162	0.0224	5.4%	2.3816	0.3816	0.3847	0.0665	100%
	500	2.0023	0.0023	0.0310	0.0403	5.2%	2.0020	0.0020	0.0103	0.0141	4.9%	2.3818	0.3818	0.3838	0.0540	100%
	100	2.0045	0.0045	0.0571	0.0569	5.8%	1.9905	-0.0095	0.0191	0.0215	8.5%	2.4762	0.4762	0.4811	0.0907	100%
20	200	1.9996	-0.0004	0.0395	0.0526	5.6%	1.9905	-0.0095	0.0155	0.0159	12%	2.4841	0.4841	0.4874	0.0764	100%
	500	2.0006	0.0006	0.0262	0.0335	5.4%	1.9910	-0.0090	0.0117	0.0099	22%	2.4833	0.4833	0.4854	0.0624	100%
	100	1.9980	-0.0020	0.0602	0.0829	4.8%	1.9933	-0.0067	0.0155	0.0190	7.5%	2.4460	0.4460	0.4489	0.0674	100%
50	200	1.9997	-0.0003	0.0432	0.0553	5%	1.9936	-0.0064	0.0116	0.0133	11%	2.4503	0.4503	0.4521	0.0545	100%
	500	1.9984	-0.0016	0.00289	0.0401	5.5%	1.9938	-0.0062	0.0086	0.0081	17%	2.4552	0.4552	0.4565	0.0461	100%
100	100	2.0025	0.0025	0.0584	0.0784	5.4%	2.0006	0.0006	0.0073	0.0096	4.1%	1.9998	-0.0002	0.0068	0.0083	5%
	200	2.0006	0.0006	0.0409	0.0535	5%	2.0021	0.0021	0.0058	0.0073	6.2%	2.4277	0.4277	0.4291	0.0465	100%
	500	1.9992	-0.0008	0.0269	0.0360	5.5%	2.0021	0.0021	0.0040	0.0047	8.6%	2.4302	0.4302	0.4309	0.0324	100%

Table A4: Simulation results for β_1 and β_2 of DGP A4

β_1		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	0.9920	-0.0080	0.1007	0.1197	6.2%	0.9987	-0.0013	0.0331	0.0441	4.3%	0.9991	-0.0009	0.0459	0.0614	4.2%
10	200	0.9964	-0.0036	0.0700	0.0871	4.5%	0.9989	-0.0011	0.0223	0.0306	5.2%	0.9993	-0.0007	0.0328	0.0441	5.2%
	500	0.9976	-0.0024	0.0419	0.0535	5.6%	1.0003	0.0003	0.0138	0.0186	5.2%	1.0004	0.0004	0.0209	0.0281	5.1%
	100	0.9957	-0.0043	0.0984	0.1225	5%	0.9990	-0.0010	0.0225	0.0298	4.6%	1.0015	0.0015	0.0369	0.0502	4.8%
20	200	0.9953	-0.0047	0.0755	0.0919	5.8%	1.0008	0.0008	0.0149	0.0194	5.1%	1.0006	0.0006	0.0265	0.0343	5.8%
	500	0.9936	-0.0064	0.0481	0.0620	4.9%	0.9999	-0.0001	0.0095	0.0129	4.9%	0.9998	-0.0002	0.0166	0.0219	4.9%
	100	1.0005	0.0005	0.0977	0.1209	5.5%	1.0002	0.0002	0.0147	0.0192	5.2%	0.9994	-0.0006	0.0332	0.0440	4.8%
50	200	0.9995	-0.0005	0.0753	0.1009	5.1%	0.9996	-0.0004	0.0104	0.0132	6%	0.9995	-0.0005	0.0238	0.0329	5%
	500	0.9958	-0.0042	0.0586	0.0731	5.6%	1.0002	0.0002	0.0064	0.0086	4.2%	1.0005	0.0005	0.0149	0.0202	5.8%
100	100	0.9987	-0.0013	0.0900	0.1151	5.3%	1.0006	0.0006	0.0104	0.0141	4.6%	1.0001	0.0001	0.0081	0.0113	4.6%
	200	0.9952	-0.0048	0.0692	0.0860	5.4%	0.9999	-0.0001	0.0071	0.0100	5.2%	0.9988	-0.0012	0.0192	0.0254	5.5%
	500	0.9974	-0.0026	0.0492	0.0597	5.3%	0.9999	-0.0001	0.0043	0.0057	5.1%	1.0004	0.0004	0.0120	0.0167	4.5%
β_2		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	1.9990	-0.0010	0.1083	0.1325	4.8%	1.9992	-0.0008	0.0339	0.0465	5.3%	2.0028	0.0028	0.0473	0.0636	5.3%
10	200	1.9957	-0.0043	0.0713	0.0878	5.3%	1.9995	-0.0005	0.0236	0.0318	5.1%	1.9999	-0.0001	0.0324	0.0414	5.3%
	500	1.9954	-0.0046	0.0451	0.0550	6.1%	1.9993	-0.0007	0.0143	0.0194	5.3%	1.9998	-0.0002	0.02010	0.0271	4.6%
	100	1.9937	-0.0063	0.1008	0.1266	5.9%	2.0000	0.0000	0.0230	0.0308	5.2%	2.0016	0.0016	0.0399	0.0500	4.4%
20	200	1.9951	-0.0049	0.0774	0.0972	6.6%	2.0002	0.0002	0.0153	0.0214	4.8%	2.0001	0.0001	0.0286	0.0372	5.7%
	500	1.9961	-0.0039	0.0446	0.0571	5.4%	1.9996	-0.0004	0.0099	0.0138	5.1%	1.9992	-0.0008	0.0176	0.0234	4.8%
	100	1.9985	-0.0015	0.0933	0.1215	6.3%	2.0002	0.0002	0.0146	0.0193	5.7%	1.9985	-0.0015	0.0316	0.0433	4.7%
50	200	1.9960	-0.0040	0.0726	0.0930	5.2%	2.0000	0.0000	0.0102	0.0139	5%	1.9998	-0.0002	0.0226	0.0292	5%
	500	1.9977	-0.0023	0.0547	0.0714	6.1%	2.0001	0.0001	0.0063	0.0091	5%	2.0002	0.0002	0.0144	0.0184	4.8%
100	100	1.9992	-0.0008	0.0857	0.1023	5.9%	2.0004	0.0004	0.0097	0.0136	4.9%	2.0001	0.0001	0.0077	0.0103	5.4%
	200	2.0002	0.0002	0.0653	0.0862	4.9%	1.9997	-0.0003	0.0066	0.0089	5.3%	1.9993	-0.0007	0.0176	0.0232	4.3%
	500	2.0000	0.0000	0.0459	0.0568	5.5%	2.0000	0.0000	0.0040	0.0054	5.4%	2.0004	0.0004	0.0105	0.0138	4.7%

Table A5: Simulation results for ρ and β of DGP A5

ρ		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	0.4958	-0.0042	0.0417	0.0532	5.8%	0.4955	-0.0045	0.0153	0.0197	6.7%	0.5154	0.0154	0.0367	0.0471	7.5%
10	200	0.4979	-0.0021	0.0279	0.0364	5.1%	0.4981	-0.0019	0.0140	0.0132	6.3%	0.5197	0.0197	0.0298	0.0298	15%
	500	0.4990	-0.0010	0.0190	0.0256	5.4%	0.4986	-0.0014	0.0066	0.0089	5.7%	0.5207	0.0207	0.0258	0.0211	27%
	100	0.4955	-0.0045	0.0391	0.0522	5.4%	0.4997	-0.0003	0.0095	0.0133	5.1%	0.5800	0.0800	0.0908	0.0611	48%
20	200	0.4977	-0.0023	0.0244	0.0316	5.5%	0.5013	0.0013	0.0066	0.0090	5.3%	0.5898	0.0898	0.0909	0.0396	83%
	500	0.4990	-0.0010	0.0160	0.0203	5.8%	0.5027	0.0027	0.0049	0.0053	10%	0.5863	0.0863	0.0885	0.0260	99%
	100	0.4923	-0.0077	0.0427	0.0583	4.7%	0.4945	-0.0055	0.0086	0.0085	13%	0.5772	0.0772	0.0872	0.0529	49%
50	200	0.4979	-0.0021	0.0294	0.0402	5.3%	0.4979	-0.0021	0.0049	0.0058	7.7%	0.5809	0.0809	0.0856	0.0386	82%
	500	0.4991	-0.0009	0.0187	0.0245	4.2%	0.4994	-0.0006	0.0030	0.0040	5.6%	0.5829	0.0829	0.0820	0.0248	99%
100	100	0.4886	-0.0114	0.0457	0.0584	5.3%	0.4965	-0.0035	0.0058	0.0063	12%	0.5126	0.0126	0.0166	0.0147	20%
	200	0.4969	-0.0031	0.0290	0.0388	5.2%	0.4991	-0.0009	0.0032	0.0040	6.6%	0.5668	0.0668	0.0722	0.0360	70%
	500	0.4993	-0.0007	0.0180	0.0237	5%	0.5004	0.0004	0.0020	0.0025	5.4%	0.5706	0.0706	0.0726	0.0227	99%
β		TQMLE					Average Estimator					PCA				
N	T	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size	Estimate	Bias	RMSE	IQR	size
	100	1.0094	0.0094	0.0779	0.0974	6.4%	1.0091	0.0091	0.0311	0.0387	5.6%	1.3594	0.3594	0.3647	0.0849	100%
10	200	1.0069	0.0069	0.0556	0.0748	5.1%	1.0048	0.0048	0.0215	0.0282	6%	1.3554	0.3554	0.3586	0.0652	100%
	500	1.0061	0.0061	0.0380	0.0503	5.5%	1.0044	0.0044	0.0134	0.0176	5.8%	1.3530	0.3530	0.3549	0.0497	100%
	100	1.0051	0.0051	0.0679	0.0935	5.1%	1.0011	0.0011	0.0198	0.0263	4.1%	1.4268	0.4268	0.4338	0.1009	100%
20	200	1.0024	0.0024	0.0463	0.0611	3.9%	1.0007	0.0007	0.0141	0.0194	5.1%	1.4281	0.4281	0.4324	0.0838	6.1%
	500	1.0030	0.0030	0.0325	0.0428	5.3%	0.9993	-0.0007	0.0085	0.0112	6.1%	1.4297	0.4297	0.4322	0.0614	100%
	100	1.0114	0.0114	0.0802	0.1028	5.6%	1.0129	0.0129	0.0208	0.0232	12%	1.5117	0.5117	0.5172	0.0972	100%
50	200	1.0053	0.0053	0.0573	0.0778	5.2%	1.0097	0.0097	0.0149	0.0158	13%	1.5181	0.5181	0.5209	0.0762	100%
	500	1.0008	0.0008	0.0381	0.0499	5.3%	1.0077	0.0077	0.0106	0.0098	17%	1.5209	0.5209	0.5227	0.0575	100%
100	100	1.0087	0.0087	0.0711	0.0952	5.2%	1.0016	0.0016	0.0097	0.0130	5.5%	0.9791	-0.0209	0.0271	0.0239	22%
	200	1.0045	0.0045	0.0537	0.0695	5%	0.9985	-0.0015	0.0069	0.0092	5.1%	1.5952	0.5952	0.5976	0.0739	5.3%
	500	1.0005	0.0005	0.0327	0.0420	4.9%	0.9973	-0.0027	0.0050	0.0057	9%	1.5987	0.5987	0.5999	0.0498	100%