Maximum Likelihood Estimation of Dynamic Panel Data Models with Interactive Effects: Quasi-differencing Over Time or Across Individuals?*

Cheng Hsiao† Qiankun Zhou‡

This version May 27, 2022

Abstract

We consider the quasi maximum likelihood (MLE) estimation of dynamic panel models with interactive effects based on the Ahn et. al. (2001, 2013) quasi-differencing methods to remove the interactive effects. We show that the quasi-difference MLE (QDMLE) over time is inconsistent when $N \to \infty$ whether T is fixed or goes to infinity. On the other hand, the QDMLE is consistent and asymptotically unbiased if the difference is taken over individuals when T is large whether N is fixed or large. Monte Carlo studies are conducted to compare the performance of the QDMLE using different quasi-difference methods.

Keywords: Dynamic panel models, Interactive effects, Maximum likelihood estimation, Quasi-difference

JEL classification: C01, C13, C23

*We would like to thank the editor, a referee and Ruiqi Liu and Yonghui Zhang for helpful comments. All remaining errors are solely ours. Partial research support of China NSFgrant #771631004 and #72033008 to the first author is gratefully acknowledged.

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

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

1 Introduction

Economic models typically focus on modeling causal relationships of a few variables an investigator considers important. Panel data provide information on individual outcomes across individuals and over time. Factors affecting individual outcomes are numerous. To control the impacts of unobserved individual and time heterogeneity and to obtain valid statistical inference, various approaches have been suggested such as error component model (e.g., Hoch (1962), Kuh (1963), Balestra and Nerlov (1967)), random coefficient model (e.g., Swamy (1970), Hsiao (1974, 1975)), mixed fixed and random coefficients model (e.g., Hsiao et al (1993)), functional coefficients model (e.g., Chang et al (2016, 2021)), interactive effects model (e.g., Pesaran (2006), Bai (2009)), etc. In this paper, we consider the estimation of panel linear dynamic interactive effects models. Contrary to the approach of modeling unobserved individual- and time- specific effects in additive form (e.g., Hsiao (2014, Ch.3 and 4)), there is no simple transformation to get rid of the unobserved individual- and time-specific effects when they are in multiplicative form. Ahn, Lee and Schmidt (2001, 2003) (ALS) have suggested a quasi-difference approach to remove the interactive effects in the model. They show that when the conditional covariates are strictly exogenous, applying GMM to the quasi-differenced equations can yield consistent and asymptotically normally distributed estimator if T is fixed and $N \to \infty$. However, the moments of quasi-differenced models are nonlinear functions of the original model parameters. There could be multiple solutions satisfying the moment conditions. To rule out the inconsistent solutions, separate objective functions need to be introduced (e.g., Honore (1993), Powell (1986) for the Tobit type models). On the other hand, the likelihood approach provides a natural objective of maximizing the quasi-likelihood function. In this paper, we consider the likelihood approach of estimating linear dynamic models with interactive effects following the ALS quasi-differencing approach. Since quasi-differencing approach can be approached from either the time-differencing or the pairwise differencing perspective, we focus on comparing the asymptotic properties of quasi maximum likelihood estimates (QMLE). We show that if lagged dependent variables appear as conditional covariates in a model, the quasi-differencing over time is not consistent if T is fixed and $N \to \infty$ or both N and T go to infinity, $(N, T) \to \infty$. On the other hand, if one takes quasi-difference across individuals, then it is possible to get consistent and asymptotically normally distributed estimator whether N is fixed or large when $T \to \infty$.

We set up the model in Section 2. Section 3 introduces quasi-difference over time or across individuals. Section 4 derives the asymptotic properties of QDMLE when differencing over time period. Section 5 discusses the asymptotic properties of QDMLE when differencing across individuals. Results of Monte Carlo simulation studies are provided in Section 6. Concluding remarks are in Section 7. Proofs of asymptotic results and extension to model with multiple factors are provided in the Appendix. The extension of model with exogenous regressors is provided in the online supplement.

Throughout this paper, we assume the letter C stands for a generic finite positive constant. The notation $\rightarrow_p$ denotes convergence in probability, $\rightarrow_d$ denotes convergence in distribution, and $(N,T) \rightarrow \infty$ denotes both N and T go to infinity at the same time.

2 Model and Assumptions

We consider a simple dynamic panel model

$$y_{it} = \gamma y_{it-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + v_{it}, i = 1, \ldots, N; t = 1, \ldots, T, \tag{2.1}$$

$$v_{it} = \boldsymbol{\lambda}_i'\mathbf{f}_t + u_{it}, \tag{2.2}$$

where $|\gamma| < 1$, $\mathbf{x}_{it}$ are strictly exogenous variables with regard to u_{it} such that $E(\mathbf{x}_{it}u_{js}) = 0$, $\boldsymbol{\lambda}_i = (\lambda_{i1}, \ldots, \lambda_{ir})'$ denotes the factor loadings while $\mathbf{f}_t = (f_{1t}, \ldots, f_{rt})'$ denotes unobservable common factors, r is the number of common factors. The initial observation y_{i0} is observable with $E\left(y_{i0}^2\right) < \infty$.

We assume

Assumption A1: u_{it} is independent of $\boldsymbol{\lambda}_i$ and $\mathbf{f}_t$ and is independently and identically distributed over i and t, with mean zero, variance σ_u^2 and finite fourth moments.

Assumption A2: $\boldsymbol{\lambda}_i$ is a vector of constants over i with $N^{-1}\sum_{i=1}^N \boldsymbol{\lambda}_i\boldsymbol{\lambda}_i'$ converges to a positive definite matrix.

Assumption A3: $\mathbf{f}_t$ is a vector of constants over t with $T^{-1}\sum_{t=1}^T \mathbf{f}_t\mathbf{f}_t'$ converges to a positive definite matrix.

The i.i.d assumption on u_{it} (Assumption A1) is made to simplify algebraic derivation. In principle, we can generalize it to allow weak time and cross-sectional dependence (e.g., Bai (2009), Jiang et al (2021)). However, it does not change the asymptotic properties of QMLE, but will substantially complicates the algebra. Assumption A2 and A3 are standard assumptions for factor models (e.g., Anderson and Rubin (1956), Bai (2003, 2009)).

3 Quasi-Differencing of Panel Interactive Model

To illustrate the basic idea of the ALS quasi-differencing approach, we assume $r = 1$, then

$$v_{it} = \lambda_i' f_t + u_{it}, \tag{3.1}$$

For the single common factor structure (3.1), it is impossible to identify f_t and λ_i separately without a normalization. To avoid this problem, we follow ALS (2013) to assume λ_1 or f_1 are nonzero constants[1] and let

$$\theta_t = \frac{f_t}{f_1}, \quad t = 2, \ldots, T, \tag{3.2}$$

then

$$y_{it} - \theta_t y_{i1} = \gamma \left(y_{i,t-1} - \theta_t y_{i0} \right) + \left(\mathbf{x}'_{it} - \theta_t \mathbf{x}'_{i,1} \right) \boldsymbol{\beta} + \left(u_{it} - \theta_t u_{i1} \right), \tag{3.3}$$

for $t = 2, \ldots, T$.

The ALS idea of quasi-differencing can also be applied cross-sectionally to get rid of $\lambda'_i f_t$. Suppose $\lambda_1 \neq 0$, let

$$\phi_i = \frac{\lambda_i}{\lambda_1}, \quad i = 2, \ldots, N, \tag{3.4}$$

then taking pairwise difference yields

$$y_{it} - \phi_i y_{1t} = \gamma \left(y_{i,t-1} - \phi_i y_{1,t-1} \right) + \left(\mathbf{x}'_{it} - \phi_i \mathbf{x}'_{1t} \right) \boldsymbol{\beta} + \left(u_{it} - \phi_i u_{1t} \right), \tag{3.5}$$

for $i = 2, \ldots, N$, $t = 1, \ldots, T$.

Remark 3.1 *Although the aim of both the linear difference for the one-way fixed effects model and quasi-difference for the interactive effects model is to remove the incidental parameters, there is a fundamental difference between the two. In the one-way fixed effects model, say individual specific effects, α_i, only appears in across-sectional dimension. Taking linear linear difference over time removes the incidental parameters, α_i, from the transformed model. If the conditional covariates involve lag dependent variables, the covariance transformation (e.g., Hsiao (2014)) or the first difference (Anderson and Hsiao (1981, 1982)) or the forward demean (Arellano and Bover (1995)) or the forward and backward transformation (Han et al (2014)) etc., to remove the incidental parameters in cross-sectional dimension also creates different forms of correlations between the transformed regressors and errors of the equation. Hence, the asymptotic bias depends on the way N or T goes to infinity. On the other hand, in the case of interactive effects, the incidental parameters appears in both cross-sectional and time series dimensions. One may be able to remove the interactive effects through quasi-differencing. However, the incidental parameters issues remain. For instance, as can be seen in (3.3) and (3.5), either θ_t or ϕ_i increases with T or N.*

[1] In the extreme case when $\lambda_1 = 0$ or $f_1 = 0$, we can choose any other factor loading as long as $\lambda_i \neq 0$ and any other factor as long as $f_t \neq 0$. Then by switching index, the normalizations (3.2) and (3.4) are still valid, and the derivation will follow.

We note that from (3.3), the following moment conditions holds,

$$E\left(\mathbf{x}_i\left(u_{it}-\theta_t u_{i1}\right)\right)=0, \quad t=2,\ldots,T, \tag{3.6}$$

where $\mathbf{x}_i=\left(\mathbf{x}_{i1}',\ldots,\mathbf{x}_{iT}'\right)'$. *Ahn et al (2013) show that applying the GMM to (3.6) yields asymptotically normally distributed estimates of* γ *and* $\boldsymbol{\beta}$ *when* T *is fixed and* $N\to\infty$. *However, the moment conditions (3.6) is nonlinear in* $(\gamma,\boldsymbol{\beta},\theta_t)$. *There could be multiple solutions satisfying (3.6). We need to introduce additional conditions to rule out the irrelevant solutions. On the other hand, the QMLE of (3.3) gives a natural conditions for choosing* $(\gamma,\boldsymbol{\beta},\theta_t)$ *to maximize the likelihood function.*

Remark 3.2 *The asymptotic properties of the quasi-difference estimators are not affected by the presence of strictly exogenous variables* $\mathbf{x}_{it}$. *Therefore, for notational ease, we assume* $\boldsymbol{\beta}=0$. *Moreover, to have a clear idea of the source of difference between quasi-differencing over time or across individuals, we shall consider the case* $r=1$. *The generalization to* $r>1$ *is relegated to the appendix.*

4 Quasi-Differencing over Time

For model (2.1) with a single common factor structure (3.1), conditional on y_{i0} being fixed constants, we have the following system of equations over time periods for (3.3) for each i,

$$\mathbf{y}_i^*\left(\boldsymbol{\theta}\right)=\mathbf{y}_{i,-1}^*\left(\boldsymbol{\theta}\right)\gamma+\mathbf{u}_i^*\left(\boldsymbol{\theta}\right), \tag{4.1}$$

or

$$\dot{\mathbf{y}}_i-\boldsymbol{\theta}y_{i1}=\gamma\left(\dot{\mathbf{y}}_{i,-1}-\boldsymbol{\theta}y_{i0}\right)+\dot{\mathbf{u}}_i-\boldsymbol{\theta}u_{i1}, \tag{4.2}$$

where

$$\begin{aligned}
\underset{(T-1)\times 1}{\mathbf{y}_i^*\left(\boldsymbol{\theta}\right)} &= \begin{pmatrix} y_{i2}-\theta_2 y_{i1} \\ \vdots \\ y_{iT}-\theta_T y_{i1}\end{pmatrix}=\dot{\mathbf{y}}_i-\boldsymbol{\theta}y_{i1},\ \underset{(T-1)\times 1}{\dot{\mathbf{y}}_i}=\begin{pmatrix} y_{i2} \\ \vdots \\ y_{iT}\end{pmatrix}, \\
\underset{(T-1)\times 1}{\mathbf{y}_{i,-1}^*\left(\boldsymbol{\theta}\right)} &= \begin{pmatrix} y_{i1}-\theta_2 y_{i0} \\ \vdots \\ y_{i,T-1}-\theta_T y_{i0}\end{pmatrix}=\dot{\mathbf{y}}_{i,-1}-\boldsymbol{\theta}y_{i0},\ \underset{(T-1)\times 1}{\dot{\mathbf{y}}_{i,-1}}=\begin{pmatrix} y_{i1} \\ \vdots \\ y_{i,T-1}\end{pmatrix}, \\
\underset{(T-1)\times 1}{\mathbf{u}_i^*\left(\boldsymbol{\theta}\right)} &= \begin{pmatrix} u_{i2}-\theta_2 u_{i1} \\ \vdots \\ u_{iT}-\theta_T u_{i1}\end{pmatrix}=\dot{\mathbf{u}}_i-\boldsymbol{\theta}u_{i1},\ \underset{(T-1)\times 1}{\dot{\mathbf{u}}_i}=\begin{pmatrix} u_{i2} \\ \vdots \\ u_{iT}\end{pmatrix},
\end{aligned}$$

and $\boldsymbol{\theta} = (\theta_2, \ldots, \theta_T)'$ is a $(T-1) \times 1$ vector with unrestricted parameters.

Let

$$\underset{(T-1)\times(T-1)}{\Omega} = E\left(\mathbf{u}_i^* \mathbf{u}_i^{*\prime}\right) = E\left[\left(\dot{\mathbf{u}}_i - \boldsymbol{\theta} u_{i1}\right)\left(\dot{\mathbf{u}}_i - \boldsymbol{\theta} u_{i1}\right)'\right]$$
$$= \sigma_u^2\left(\mathbf{I}_{T-1} + \boldsymbol{\theta}\boldsymbol{\theta}'\right), \tag{4.3}$$

with

$$\Omega^{-1} = \frac{1}{\sigma_u^2}\left(\mathbf{I}_{T-1} - \frac{\boldsymbol{\theta}\boldsymbol{\theta}'}{1+\boldsymbol{\theta}'\boldsymbol{\theta}}\right). \tag{4.4}$$

The quasi-log-likelihood function of (4.1) is given by

$$\log L\left(\gamma, \boldsymbol{\theta}, \sigma_u^2\right) = -\frac{N}{2}\log|\Omega| - \frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{y}_i^*(\boldsymbol{\theta}) - \mathbf{y}_{i,-1}^*(\boldsymbol{\theta})\gamma\right)'\Omega^{-1}\left(\mathbf{y}_i^*(\boldsymbol{\theta}) - \mathbf{y}_{i,-1}^*(\boldsymbol{\theta})\gamma\right). \tag{4.5}$$

The first order conditions for maximizing (4.5) with respect to $(\gamma, \boldsymbol{\theta}, \sigma_u^2)$ are

$$\frac{\partial \log L}{\partial \gamma} = \sum_{i=1}^{N}\mathbf{y}_{i,-1}^*\left(\hat{\boldsymbol{\theta}}\right)'\hat{\Omega}^{-1}\left(\mathbf{y}_i^*\left(\hat{\boldsymbol{\theta}}\right) - \mathbf{y}_{i,-1}^*\left(\hat{\boldsymbol{\theta}}\right)\hat{\gamma}\right) = 0, \tag{4.6}$$

and

$$\frac{\partial \log L}{\partial \boldsymbol{\theta}} = -N\sigma_u^2\hat{\Omega}^{-1}\hat{\boldsymbol{\theta}} + N\sigma_u^2\hat{\Omega}^{-1}\mathbf{S}\left(\hat{\boldsymbol{\theta}}\right)\hat{\Omega}^{-1}\hat{\boldsymbol{\theta}}$$
$$+\hat{\Omega}^{-1}\sum_{i=1}^{N}\left[\left(y_{i1} - y_{i0}\hat{\gamma}\right)\left[\dot{\mathbf{y}}_i - \dot{\mathbf{y}}_{i,-1}\hat{\gamma}\right] - \hat{\boldsymbol{\theta}}\left(y_{i1} - y_{i0}\hat{\gamma}\right)^2\right]$$
$$= 0, \tag{4.7}$$

where $\hat{\Omega} = \frac{1}{\hat{\sigma}_u^2}\mathbf{S}\left(\hat{\boldsymbol{\theta}}\right)$ with

$$\mathbf{S}\left(\hat{\boldsymbol{\theta}}\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{y}_i^*\left(\hat{\boldsymbol{\theta}}\right) - \mathbf{y}_{i,-1}^*\left(\hat{\boldsymbol{\theta}}\right)\hat{\gamma}\right)\left(\mathbf{y}_i^*\left(\hat{\boldsymbol{\theta}}\right) - \mathbf{y}_{i,-1}^*\left(\hat{\boldsymbol{\theta}}\right)\hat{\gamma}\right)', \tag{4.8}$$

and

$$\hat{\sigma}_u^2 = \frac{1}{N(T-1)}\sum_{i=1}^{N}\left(\mathbf{y}_i^* - \mathbf{y}_{i,-1}^*\hat{\gamma}\right)'\left(\mathbf{I}_{T-1} + \hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}'\right)^{-1}\left(\mathbf{y}_i^* - \mathbf{y}_{i,-1}^*\hat{\gamma}\right). \tag{4.9}$$

Conditional on $\boldsymbol{\theta}$, from (4.6), we obtain

$$\hat{\gamma}_{MLE}^{TS} = \left(\sum_{i=1}^{N}\mathbf{y}_{i,-1}^*(\boldsymbol{\theta})'\Omega^{-1}\mathbf{y}_{i,-1}^*(\boldsymbol{\theta})\right)^{-1}\sum_{i=1}^{N}\left(\mathbf{y}_{i,-1}^*(\boldsymbol{\theta})'\Omega^{-1}\mathbf{y}_i^*(\boldsymbol{\theta})\right), \tag{4.10}$$

where ^{TS} refers to using normalization (3.2) to apply the quasi-difference over time period.

We note that conditional on $\boldsymbol{\theta}$,

$$\hat{\gamma}_{MLE}^{TS} - \gamma = \left(\frac{1}{NT} \sum_{i=1}^{N} \mathbf{y}_{i,-1}^{*}(\boldsymbol{\theta})' \, \Omega^{-1} \mathbf{y}_{i,-1}^{*}(\boldsymbol{\theta}) \right)^{-1} \frac{1}{NT} \sum_{i=1}^{N} \mathbf{y}_{i,-1}^{*}(\boldsymbol{\theta})' \, \Omega^{-1} \mathbf{u}_{i}^{*}(\boldsymbol{\theta}), \tag{4.11}$$

and the denominator converges to

$$\frac{1}{NT} \sum_{i=1}^{N} \mathbf{y}_{i,-1}^{*}(\boldsymbol{\theta})' \, \Omega^{-1} \mathbf{y}_{i,-1}^{*}(\boldsymbol{\theta}) \to_p \frac{\bar{\sigma}_{\lambda}^{2} \bar{\sigma}_{\gamma,f}^{2}}{\sigma_u^2} + \frac{1}{1-\gamma^2}, \tag{4.12}$$

where $\bar{\sigma}_{\lambda}^{2} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \lambda_i^2$ and $\bar{\sigma}_{\gamma,f}^{2} = \lim_{T\to\infty} \frac{1}{T} \boldsymbol{\omega}' \left(\mathbf{I}_{T-1} + \boldsymbol{\theta}\boldsymbol{\theta}' \right)^{-1} \boldsymbol{\omega}$ with $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_{T-1})'$ and $\omega_t = \sum_{s=1}^{t} \gamma^{t-s} f_s$.

The numerator of (4.11) converges to

$$\begin{aligned} tr\left[\Omega(\boldsymbol{\theta})^{-1} \sigma_u^2 \boldsymbol{\theta} \left(1, \gamma, \ldots, \gamma^{T-2}\right) \right] &= tr\left[\left(\mathbf{I}_{T-1} - \frac{\boldsymbol{\theta}\boldsymbol{\theta}'}{1+\boldsymbol{\theta}'\boldsymbol{\theta}} \right) \boldsymbol{\theta} \left(1, \gamma, \ldots, \gamma^{T-2}\right) \right] \\ &= \frac{1}{1+\boldsymbol{\theta}'\boldsymbol{\theta}} tr\left[\boldsymbol{\theta} \left(1, \gamma, \ldots, \gamma^{T-2}\right) \right] \\ &= \frac{1}{1+\boldsymbol{\theta}'\boldsymbol{\theta}} tr \left(1, \gamma, \ldots, \gamma^{T-2}\right) \boldsymbol{\theta}' \\ &= \frac{1}{1+\boldsymbol{\theta}'\boldsymbol{\theta}} \sum_{t=2}^{T} \theta_t \gamma^{t-2} = O\left(\frac{1}{T}\right). \end{aligned} \tag{4.13}$$

Equations (4.12) and (4.13) imply that the numerator of (4.11) is $O_p\left(\frac{1}{T}\right)$, it does not go to zero when T is fixed no matter how large N is. In other words, $\hat{\gamma}_{MLE}^{TS}$ is inconsistent if T is fixed and $N \to \infty$.

Although conditional on $\boldsymbol{\theta}$, the QDMLE of γ is consistent when $T \to \infty$, if $\frac{N}{T} \to a \neq 0$ as $(N,T) \to \infty$,

$$E\left(\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbf{y}_{i,-1}^{*}(\boldsymbol{\theta})' \, \Omega^{-1} \mathbf{u}_{i}^{*}(\boldsymbol{\theta}) \right) = -\frac{\boldsymbol{\theta}'\Psi\boldsymbol{\theta}}{1+\boldsymbol{\theta}'\boldsymbol{\theta}} \sqrt{a} + o(1), \tag{4.14}$$

where Ψ is defined in (A.8).

Unfortunately, $\hat{\boldsymbol{\theta}}$ is inconsistent even $(N,T) \to \infty$. We note that conditional on γ, (4.7) leads to

$$
\begin{aligned}
\hat{\boldsymbol{\theta}} &= \left[\left(\frac{1}{N}\sum_{i=1}^{N}\left(y_{i1}-y_{i0}\gamma\right)^2+\sigma_u^2\right)\mathbf{I}_{T-1}-\frac{1}{1+\hat{\boldsymbol{\theta}}'\hat{\boldsymbol{\theta}}}\mathbf{S}\right]^{-1}\times\frac{1}{N}\sum_{i=1}^{N}\left(y_{i1}-y_{i0}\gamma\right)\left(\dot{\mathbf{y}}_i-\dot{\mathbf{y}}_{i,-1}\gamma\right)\\
&= \left(\frac{1}{N}\sum_{i=1}^{N}\left(y_{i1}-y_{i0}\gamma\right)^2+\sigma_u^2\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\left(y_{i1}-y_{i0}\gamma\right)\left[\boldsymbol{\theta}\left(y_{i1}-\gamma y_{i0}\right)+\left(\dot{\mathbf{u}}_i-\boldsymbol{\theta}u_{i1}\right)\right]+o_p(1)\\
&= \boldsymbol{\theta}+\left(\frac{1}{N}\sum_{i=1}^{N}\left(y_{i1}-y_{i0}\gamma\right)^2+\sigma_u^2\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\left(y_{i1}-y_{i0}\gamma\right)\left(\dot{\mathbf{u}}_i-\boldsymbol{\theta}u_{i1}\right)-\boldsymbol{\theta}\sigma_u^2\right)+o_p(1)\\
&= \boldsymbol{\theta}-\frac{2\sigma_u^2}{\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(y_{i1}-y_{i0}\gamma\right)^2+\sigma_u^2}\boldsymbol{\theta}+o_p(1)\\
&\nrightarrow {}_p\boldsymbol{\theta}. \qquad (4.15)
\end{aligned}
$$

Since the score function evaluated at the true $(\gamma,\boldsymbol{\theta})$,

$$
\frac{1}{NT}\begin{pmatrix}\frac{\partial\log L}{\partial\gamma}\\ \frac{\partial\log L}{\partial\boldsymbol{\theta}}\end{pmatrix}\neq 0 \quad \text{as } (N,T)\to\infty, \tag{4.16}
$$

the QDMLE is inconsistent whether T is fixed or $T\to\infty$.

It then follows that

Theorem 4.1 *Under Assumption A1-A3, conditional on y_{i0} being fixed constants, the QDMLE over time for model (2.1) with single common factor structure (3.1) is inconsistent when $N\to\infty$ whether T is fixed or $T\to\infty$. If consistent estimator for $\boldsymbol{\theta}$ can be found, the QDMLE is consistent if $T\to\infty$. However, if $\frac{N}{T}\to a\neq 0<\infty$, then*

$$
\sqrt{NT}\left(\hat{\gamma}_{MLE}^{TS}-\left(\gamma-\frac{1}{T}\frac{b}{k_1}\right)\right)\to_d N\left(0,\frac{k_2}{k_1^2}\right), \tag{4.17}
$$

where b denotes the bias term defined in (A.10), k_1 and k_2 are given by

$$
k_1=\frac{\bar{\sigma}_\lambda^2\bar{\sigma}_{\gamma,f}^2}{\sigma_u^2}+\frac{1}{1-\gamma^2}, k_2=\frac{\bar{\sigma}_\lambda^2\tilde{\sigma}_{\gamma,f}^2}{\sigma_u^2}+\frac{1}{1-\gamma^2}, \tag{4.18}
$$

where $\bar{\sigma}_\lambda^2=\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\lambda_i^2$, $\bar{\sigma}_{\gamma,f}^2=\lim_{T\to\infty}\frac{1}{T}\boldsymbol{\omega}'\left(\mathbf{I}_{T-1}+\boldsymbol{\theta}\boldsymbol{\theta}'\right)^{-1}\boldsymbol{\omega}$ and $\tilde{\sigma}_{\gamma,f}^2=\lim_{T\to\infty}\frac{1}{T}\boldsymbol{\omega}'\boldsymbol{\omega}$ with $\boldsymbol{\omega}=\left(\omega_1,\ldots,\omega_{T-1}\right)'$ and $\omega_t=\sum_{s=1}^{t}\gamma^{t-s}f_s$.

Remark 4.2 *There are many different ways to implement the ALS quasi-difference approach. For instance, we can use the long difference as (3.3). We can also use the first difference by letting $\theta_t^*=\frac{f_t}{f_{t-1}}$, then*

$$
y_{it}-\theta_t^*y_{i,t-1}=\gamma\left(y_{i,t-1}-\theta_t^*y_{i,t-2}\right)+\left(\mathbf{x}_{it}'-\theta_t^*\mathbf{x}_{i,t-1}'\right)\boldsymbol{\beta}+\left(u_{it}-\theta_t^*u_{i,t-1}\right), \quad t=2,\ldots,T. \tag{4.19}
$$

Or the backward and forward differencing proposed by Han et al (2014). In principle, one can derive the QDMLE with corresponding transformed likelihood function. However, there is an advantage of using the long difference, the variance-covariance matrix of the transformed error terms has the form of standard random effects one way error component model. The variance-covariance matrix of the errors has the form (4.3) and known pattern of its inverse (4.4). On the other hand, if we use (4.19), the variance-covariance matrix of the system $\tilde{\mathbf{u}}_i^* = (u_{i2} - \theta_2^* u_{i1}, \ldots, u_{iT} - \theta_T^* u_{i,T-1})'$ *takes the form*

$$\sigma_u^2 \begin{pmatrix} \left(1+\theta_2^{*2}\right) & -\theta_2^* & 0 & 0 \\ -\theta_2^* & \left(1+\theta_3^{*2}\right) & \ddots & 0 \\ 0 & \ddots & \ddots & -\theta_T^* \\ 0 & 0 & -\theta_T^* & \left(1+\theta_T^{*2}\right) \end{pmatrix}. \tag{4.20}$$

The element of its inverse matrix consists of all $(\theta_2^*, \ldots, \theta_T^*)$, *and its computation is much more complicated. Similarly for the backward and forward transformation proposed by Han et al (2014).*

However, we expect the asymptotic properties of the different forms of transformations remain the same. The QDMLE remains inconsistent whether T *is fixed or goes to infinity as* $N \to \infty$. *For instance, in the case of (4.19),* $\theta_t^* = \frac{f_t}{f_{t-1}}$. *If* θ_t *can not be consistently estimated either* T *is fixed or large no matter how large* N *is, then there is no reason to expect* θ_t^* *will be consistently estimated.*

5 Quasi-Differencing across Individuals

Rewrite (3.5) in vector form as

$$\mathbf{y}_t^*(\boldsymbol{\phi}) = \mathbf{y}_{t-1}^*(\boldsymbol{\phi})\,\gamma + \mathbf{u}_t^*(\boldsymbol{\phi}), \quad t = 1, \ldots, T, \tag{5.1}$$

or

$$\dot{\mathbf{y}}_t - \boldsymbol{\phi} y_{1t} = \gamma\left(\dot{\mathbf{y}}_{t-1} - \boldsymbol{\phi} y_{1,t-1}\right) + \dot{\mathbf{u}}_t - \boldsymbol{\phi} u_{1t}, \tag{5.2}$$

where

$$
\begin{aligned}
\underset{(N-1)\times 1}{\mathbf{y}_t^*(\boldsymbol{\phi})} &= \begin{pmatrix} y_{2t}-\phi_2 y_{1t} \\ \vdots \\ y_{Nt}-\phi_N y_{1t} \end{pmatrix} = \dot{\mathbf{y}}_t - \boldsymbol{\phi} y_{1t}, \ \underset{(N-1)\times 1}{\dot{\mathbf{y}}_t} = \begin{pmatrix} y_{2t} \\ \vdots \\ y_{Nt} \end{pmatrix}, \\
\underset{(N-1)\times 1}{\mathbf{y}_{t-1}^*(\boldsymbol{\phi})} &= \begin{pmatrix} y_{2,t-1}-\phi_2 y_{1,t-1} \\ \vdots \\ y_{N,t-1}-\phi_N y_{1,t-1} \end{pmatrix} = \dot{\mathbf{y}}_{t-1} - \boldsymbol{\phi} y_{1,t-1}, \ \underset{(N-1)\times 1}{\mathbf{y}_{t-1}} = \begin{pmatrix} y_{2,t-1} \\ \vdots \\ y_{N,t-1} \end{pmatrix}, \\
\underset{(N-1)\times 1}{\mathbf{u}_t^*(\boldsymbol{\phi})} &= \begin{pmatrix} u_{2t}-\phi_2 u_{1t} \\ \vdots \\ u_{Nt}-\phi_N u_{1t} \end{pmatrix} = \dot{\mathbf{u}}_t - \boldsymbol{\phi} u_{1t}, \ \underset{(N-1)\times 1}{\dot{\mathbf{u}}_t} = \begin{pmatrix} u_{2t} \\ \vdots \\ u_{Nt} \end{pmatrix},
\end{aligned}
$$

and $\boldsymbol{\phi} = (\phi_2, \ldots, \phi_N)'$ is a $(N-1)\times 1$ vector with unrestricted parameters.

We note that for model (5.1), for any $s \neq t$,

$$
\begin{aligned}
E\left(\mathbf{u}_t^*(\boldsymbol{\phi})\,\mathbf{u}_s^*(\boldsymbol{\phi})'\right) &= E\left[\left(\dot{\mathbf{u}}_t - \boldsymbol{\phi} u_{1t}\right)\left(\dot{\mathbf{u}}_s' - \boldsymbol{\phi}' u_{1s}\right)\right] \\
&= E\left(\dot{\mathbf{u}}_t \dot{\mathbf{u}}_s'\right) - E\left(\dot{\mathbf{u}}_t u_{1s}\right)\boldsymbol{\phi}' - \boldsymbol{\phi} E\left(u_{1t}\dot{\mathbf{u}}_s'\right) + \boldsymbol{\phi}\boldsymbol{\phi}' E\left(u_{1t}u_{1s}\right) \\
&= 0, \qquad (5.3)
\end{aligned}
$$

under serially uncorrelated assumption (A1). The variance-covariance matrix of (5.1) takes the form,

$$
\begin{aligned}
\underset{(N-1)\times(N-1)}{\tilde{\Omega}} &= E\left(\mathbf{u}_t^*(\boldsymbol{\phi})\,\mathbf{u}_t^*(\boldsymbol{\phi})'\right) = E\left[\left(\dot{\mathbf{u}}_t - \boldsymbol{\phi} u_{1t}\right)\left(\dot{\mathbf{u}}_t' - \boldsymbol{\phi}' u_{1t}\right)\right] \\
&= \sigma_u^2\left(\mathbf{I}_{N-1} + \boldsymbol{\phi}\boldsymbol{\phi}'\right), \qquad (5.4)
\end{aligned}
$$

with

$$
\tilde{\Omega}^{-1} = \frac{1}{\sigma_u^2}\left(\mathbf{I}_{N-1} - \frac{\boldsymbol{\phi}\boldsymbol{\phi}'}{1+\boldsymbol{\phi}'\boldsymbol{\phi}}\right). \qquad (5.5)
$$

The quasi-log-likelihood function of (5.1) is given by

$$
\log L = -\frac{T}{2}\log\left|\tilde{\Omega}\right| - \frac{1}{2}\sum_{t=1}^{T}\left(\mathbf{y}_t^*(\boldsymbol{\phi}) - \mathbf{y}_{t-1}^*(\boldsymbol{\phi})\gamma\right)'\tilde{\Omega}^{-1}\left(\mathbf{y}_t^*(\boldsymbol{\phi}) - \mathbf{y}_{t-1}^*(\boldsymbol{\phi})\gamma\right). \qquad (5.6)
$$

Conditional on $\boldsymbol{\phi}$,

$$
\hat{\gamma}_{MLE}^{CS} = \left(\sum_{t=1}^{T}\mathbf{y}_{t-1}^*(\boldsymbol{\phi})'\tilde{\Omega}^{-1}\mathbf{y}_{t-1}^*(\boldsymbol{\phi})\right)^{-1}\left(\sum_{t=1}^{T}\mathbf{y}_{t-1}^*(\boldsymbol{\phi})'\tilde{\Omega}^{-1}\mathbf{y}_t^*(\boldsymbol{\phi})\right), \qquad (5.7)
$$

where ^{CS} refers to using normalization (3.4) to apply the quasi-difference over cross sectional dimension.

To see that the above QDMLE across individuals is asymptotically unbiased, we note that

$$\sqrt{NT}\left(\hat{\gamma}_{MLE}^{CS}-\gamma\right)=\left[\frac{1}{NT}\sum_{t=1}^{T}\mathbf{y}_{t-1}^{*}\left(\boldsymbol{\phi}\right)'\tilde{\Omega}^{-1}\mathbf{y}_{t-1}^{*}\left(\boldsymbol{\phi}\right)\right]^{-1}\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\mathbf{y}_{t-1}^{*}\left(\boldsymbol{\phi}\right)'\tilde{\Omega}^{-1}\mathbf{u}_{t}^{*}\left(\boldsymbol{\phi}\right), \quad (5.8)$$

where $\mathbf{y}_t^*=\mathbf{y}_{t-1}^*\gamma+\mathbf{u}_t^*=\sum_{j=0}^{\infty}\gamma^j\mathbf{u}_{t-j}^*$ since $|\gamma|<1$. Thus, as $(N,T)\to\infty$, the denominator of (5.8) converges

$$\frac{1}{NT}\sum_{t=1}^{T}E\left[\mathbf{y}_{t-1}^{*}\left(\boldsymbol{\phi}\right)'\tilde{\Omega}^{-1}\mathbf{y}_{t-1}^{*}\left(\boldsymbol{\phi}\right)\right]=\frac{1}{NT}\sum_{t=1}^{T}tr\left\{\tilde{\Omega}^{-1}E\left[\mathbf{y}_{t-1}^{*}\left(\boldsymbol{\phi}\right)\mathbf{y}_{t-1}^{*}\left(\boldsymbol{\phi}\right)'\right]\right\}\to\frac{1}{1-\gamma^{2}}, \quad (5.9)$$

because

$$E\left[\mathbf{y}_{t}^{*}\left(\boldsymbol{\phi}\right)\mathbf{y}_{t}^{*}\left(\boldsymbol{\phi}\right)'\right]=\frac{1}{1-\gamma^{2}}E\left[\mathbf{u}_{t}^{*}\left(\boldsymbol{\phi}\right)\mathbf{u}_{t}^{*}\left(\boldsymbol{\phi}\right)'\right]=\frac{1}{1-\gamma^{2}}\tilde{\Omega}. \quad (5.10)$$

For the numerator of (5.8), we notice that

$$\begin{aligned}E\left(\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\mathbf{y}_{t-1}^{*}\left(\boldsymbol{\phi}\right)'\tilde{\Omega}^{-1}\mathbf{u}_{t}^{*}\left(\boldsymbol{\phi}\right)\right) &= \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}E\left[\mathbf{y}_{t-1}^{*}\left(\boldsymbol{\phi}\right)'\tilde{\Omega}^{-1}\mathbf{u}_{t}^{*}\left(\boldsymbol{\phi}\right)\right]\\ &= \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}tr\left\{\tilde{\Omega}^{-1}E\left[\mathbf{u}_{t}^{*}\left(\boldsymbol{\phi}\right)\mathbf{y}_{t-1}^{*}\left(\boldsymbol{\phi}\right)'\right]\right\}\\ &= 0, \qquad (5.11)\end{aligned}$$

which suggests that the QDMLE (5.7) is asymptotically unbiased conditional on $\boldsymbol{\phi}$, and the asymptotical unbiasedness of QDMLE is independent of the way how (N,T) go to infinity.

Moreover, we can show that

$$\begin{aligned}E\left(\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\mathbf{y}_{t-1}^{*}\left(\boldsymbol{\phi}\right)'\tilde{\Omega}^{-1}\mathbf{u}_{t}^{*}\left(\boldsymbol{\phi}\right)\right)^{2} &= \frac{1}{NT}\sum_{s,t}E\left[\mathbf{y}_{t-1}^{*}\left(\boldsymbol{\phi}\right)'\tilde{\Omega}^{-1}\mathbf{u}_{t}^{*}\left(\boldsymbol{\phi}\right)\mathbf{u}_{s}^{*}\left(\boldsymbol{\phi}\right)'\tilde{\Omega}^{-1}\mathbf{y}_{s-1}^{*}\left(\boldsymbol{\phi}\right)\right]\\ &= \frac{1}{NT}\sum_{t=1}^{T}E\left[\mathbf{y}_{t-1}^{*}\left(\boldsymbol{\phi}\right)'\tilde{\Omega}^{-1}\mathbf{u}_{t}^{*}\left(\boldsymbol{\phi}\right)\mathbf{u}_{t}^{*}\left(\boldsymbol{\phi}\right)'\tilde{\Omega}^{-1}\mathbf{y}_{t-1}^{*}\left(\boldsymbol{\phi}\right)\right]\\ &= \frac{1}{NT}\sum_{t=1}^{T}E\left[\mathbf{y}_{t-1}^{*}\left(\boldsymbol{\phi}\right)'\tilde{\Omega}^{-1}\mathbf{y}_{t-1}^{*}\left(\boldsymbol{\phi}\right)\right]\\ &\to \frac{1}{1-\gamma^{2}}, \qquad (5.12)\end{aligned}$$

as $(N,T)\to\infty$. Furthermore, we can show the fourth moments of the numerator of (5.8) goes to zero by following Bao and Ullah (2010).

Combining the results of (5.9)-(5.12), we can show that

Theorem 5.1 *Under assumption A1-A3, the pairwise QDMLE (5.7) for model (2.1) with single common factor structure (3.1) is consistent and asymptotically unbiased as $N \to \infty$ whether T is fixed or $T \to \infty$, and*

$$\sqrt{NT}\left(\hat{\gamma}_{MLE}^{CS} - \gamma\right) \to_d N\left(0, 1-\gamma^2\right), \tag{5.13}$$

as $(N,T) \to \infty$.

The pairwise QDMLE depends on the simultaneous solution of the first order conditions (5.6) which can be computational cumbersome. One way to obtain a feasible solution is through the following iteration.

Step 1: Conditional on some initial estimates of $\boldsymbol{\phi}$, apply the least squares estimation to (3.5) to obtain initial estimates of γ and $\boldsymbol{\beta}$.

Step 2: Conditional on γ and $\boldsymbol{\beta}$, minimizing $\sum_{t=1}^{T}\left(\hat{u}_{it}^{*} - \phi_i \hat{u}_{it}^{*}\right)^2$ to obtain the least squares estimator of ϕ_i for $i = 2, \ldots, N$ where $\hat{u}_{it}^{*}$ denotes the residuals based on the initial estimates of γ and $\boldsymbol{\beta}$.

Step 3: Iterate between Step 1 and 2 until the solutions of $(\gamma, \boldsymbol{\beta}, \boldsymbol{\phi})$ converge.[2]

Step 4: Substitute the convergent solutions of $(\gamma, \boldsymbol{\beta}, \boldsymbol{\phi})$ as initial estimates to implement the interactive procedure to obtain the pairwise QDMLE.

6 Monte Carlo Simulation

In this section, we investigate the finite sample properties of the QDMLE for panels with cross-sectional dependence. The data generating process (DGP) is given by

$$y_{it} = \gamma y_{i.t-1} + \lambda_i f_t + u_{it}, \tag{6.1}$$

We generate $f_t \sim IIDN\left(3,1\right)$ for $t = 1, \ldots, T$ and $\lambda_i \sim IIDN\left(2, 0.3\right)$ for $t = 1, \ldots, N$. We assume the error term $u_{it} \sim IIDN\left(0,1\right)$ for $i = 1, 2, \ldots, N$, $t = 1, \ldots, T$, and they are independent of λ_i and f_t for all i and t.

The true value of γ is set at $\gamma = 0.1, 0.5, 0.9$. We let $N = 50, 100, 500$ and $T = 50, 100, 500, 1000$. The number of replication is 1000 times, and the maximum number of iteration of feasible QDMLE is 100. We report mean estimates, bias, RMSE and size comparison for these two estimators using the nominal 5% significance level, QDMLE (4.10) using normalization (3.2) and MLE-CS refers to QDMLE (5.7) using normalization (3.4).

[2]Steps 1-3 can be viewed as an iterative procedure to obtain the nonlinear least squares estimation of (3.5).

The simulation results are summarized in Tables 1-3. For the infeasible QDMLE estimators, which use the true pre-specified factors or factor loadings (or $\boldsymbol{\phi}_i$) in the estimation. The pairwise QDMLE has negligible bias and the actual size is close to the nominal size whatever sample configurations of N and T are, even for cases where the lag coefficient is on the boundary (e.g., $\gamma = 0.1$ or $\gamma = 0.9$). However, the feasible pairwise QDMLE that depends on estimated $\boldsymbol{\phi}_i$, $\hat{\boldsymbol{\phi}}_i$, although has negligible bias, could have significant size distortion when T is small due to the instability of $\hat{\boldsymbol{\phi}}_i$. Nevertheless, the size distortion declines with the increase of T. On the other hand, the QDMLE over time has significant bias and significant size distortion even in the infeasible case. In all, the simulation results appears to support the findings that QDMLE over time is asymptotically biased while QDMLE pairwise is asymptotically unbiased.

7 Conclusion

Panel data blend inter-individual difference and intra-individual dynamics provide means to identify impacts of omitted time-invariant and individual-invariant variables. If the unobserved individual-specific and time-specific effects are in additive form, they can be easily removed through a covariance transformation (e.g., Hsiao (2014, Ch3 and Ch4)). On the other hand, if they are in multiplicative form, linear transformations cannot remove them. Ahn, Lee and Schmidt (2001, 2013) have suggested a quasi-differencing method to remove the interactive effects. They show the resulting GMM method is consistent and asymptotically unbiased if the regressor are strictly exogenous. In this paper, we show that the quasi-differencing MLE over time period for a dynamic panel model is inconsistent when either T is fixed and N is large or both N and T are large. On the other hand, if we take quasi-differencing across individuals, the resulting quasi-maximum likelihood estimator is consistent and asymptotically unbiased if $T \to \infty$ whether N is fixed or $N \to \infty$.

Table 1: Simulation results for γ of DGP (6.1) when $\gamma = 0.1$

T		50				100				500				1000			
		feasible MLE		infeasible MLE		feasible MLE		infeasible MLE		feasible MLE		infeasible MLE		feasible MLE		infeasible MLE	
N		TS	CS	TS	CS	TS	CS	TS	CS	TS	CS	TS	CS	TS	CS	TS	CS
	Mean	0.1529	0.0773	0.5224	0.0986	0.1475	0.0890	0.5224	0.0994	0.1554	0.0974	0.5247	0.0998	0.1544	0.0988	0.5249	0.0999
50	Bias	0.0529	-0.0227	0.4224	-0.0014	0.0475	-0.0110	0.4224	-0.0006	0.0554	-0.0026	0.4257	-0.0002	0.0544	-0.0012	0.4249	-0.0001
	RMSE	0.0570	0.0304	0.4224	0.0199	0.0500	0.0185	0.4224	0.0141	0.0558	0.0071	0.4247	0.0064	0.0547	0.0047	0.4249	0.0046
	size	72%	19%	100%	4.7%	86%	11%	100%	5.5%	100%	5.2%	100%	5%	100%	5.8%	100%	5%
	Mean	0.1237	0.0778	0.5194	0.0994	0.1391	0.0896	0.5263	0.0997	0.1488	0.0978	0.5234	0.0999	0.1512	0.0900	0.5247	0.0999
100	Bias	0.0237	-0.0222	0.4194	-0.0006	0.0391	-0.0104	0.4263	-0.003	0.0488	-0.0022	0.4234	-0.0001	0.0512	-0.0100	0.4247	-0.0001
	RMSE	0.0286	0.0268	0.4194	0.0147	0.0404	0.0141	0.4263	0.0100	0.0490	0.0051	0.4234	0.0045	0.0513	0.0035	0.4247	0.0032
	size	28%	30%	100%	5.5%	96%	19%	100%	3.8%	100%	6%	100%	5.7%	100%	5.2%	100%	5.4%
	Mean	0.1170	0.0795	0.5127	0.1002	0.1244	0.0900	0.5193	0.0999	0.1424	0.0980	0.5232	0.1000	0.1405	0.0900	0.5229	0.0999
500	Bias	0.0170	-0.0205	0.4127	0.0002	0.0244	-0.0100	0.4193	-0.0001	0.0424	-0.0020	0.4232	0.0000	0.0405	-0.0010	0.4229	-0.0001
	RMSE	0.0182	0.0213	0..428	0.0065	0.0249	0.0110	0.4193	0.0044	0.0425	0.0028	0.4232	0.0019	0.0405	0.0017	0.4229	0.0014
	size	74%	91%	100%	5.4%	99%	58%	100%	5.8%	100%	18%	100%	5.2%	100%	11%	100%	5.1%

Notes: 1. "feasible MLE" refers to the feasible QDMLE using the iterative procedure, "infeasible MLE" refers to the QDMLE using the true prespecified factors or factor loadings to construct the variance-covariance matrix.

2. TS refers to QDMLE (4.10) using normalization (3.2), CS refers to QDMLE (5.7) using normalization (3.4);

3. The size is calculated at $H_0 : \gamma_0 = 0.1$.

Table 2: Simulation results for γ of DGP (6.1) when $\gamma = 0.5$

	T	50				100				500				1000			
		feasible MLE		infeasible MLE		feasible MLE		infeasible MLE		feasible MLE		infeasible MLE		feasible MLE		infeasible MLE	
N		TS	CS	TS	CS	TS	CS	TS	CS	TS	CS	TS	CS	TS	CS	TS	CS
	Mean	0.6962	0.4693	0.4986	0.4988	0.6905	0.4852	0.4990	0.4994	0.6813	0.4970	0.4999	0.5000	0.6799	0.4985	0.4999	0.4999
50	Bias	0.1962	-0.0307	-0.0014	-0.0012	0.1905	-0.0148	-0.0010	-0.0006	0.1813	-0.0030	-0.0001	0.0000	0.1799	-0.0015	-0.0001	-0.0001
	RMSE	0.1964	0.0356	0.0042	0.0171	0.1096	0.0197	0.0031	0.0123	0.1813	0.0064	0.0013	0.0055	0.1799	0.0042	0.0009	0.0040
	size	100%	39%	6.8%	5.1%	100%	22%	6.9%	4.9%	100%	7.2%	5%	5%	100%	7.4%	5.4%	5.3%
	Mean	0.6875	0.4698	0.4982	0.4993	0.6773	0.4848	0.4992	0.4999	0.6747	0.4973	0.4998	0.4999	0.6807	0.4986	0.4999	0.4999
100	Bias	0.1875	-0.0302	-0.0018	-0.0007	0.1773	-0.0142	-0.0008	-0.0001	0.1747	-0.0027	-0.0002	-0.0001	0.1807	-0.0014	-0.0001	-0.0001
	RMSE	0.1877	0.0333	0.0037	0.0128	0.1774	0.0167	0.0022	0.0087	0.1747	0.0048	0.0010	0.0039	0.1807	0.0032	0.0006	0.0028
	size	100%	56%	7.6%	5.7%	100%	36%	6.9%	4.7%	100%	10%	6.1%	5.4%	100%	6.8%	4.4%	5.2%
	Mean	0.6577	0.4718	0.4983	0.5003	0.6383	0.4862	0.4992	0.5000	0.6634	0.4973	0.4999	0.5000	0.6632	0.4986	0.4999	0.5000
500	Bias	0.1577	-0.0282	-0.0017	0.0003	0.1383	-0.0138	-0.0008	0.0000	0.1634	-0.0027	-0.0001	0.0000	0.1632	-0.0014	-0.0001	0.0000
	RMSE	0.1578	0.0287	0.0022	0.0057	0.1383	0.0144	0.0013	0.0038	0.1634	0.0032	0.0004	0.0017	0.1632	0.0018	0.0003	0.0012
	size	100%	100%	23%	5.1%	100%	92%	13%	5.8%	100%	36%	5.7%	4.6%	100%	18%	6.4%	4.6%

Notes: The size is calculated at $H_0 : \gamma_0 = 0.5$. Refer to notes of Table 1.

Table 3: Simulation results for γ of DGP (6.1) when $\gamma = 0.9$

	T	50				100				500				1000			
		feasible MLE		infeasible MLE		feasible MLE		infeasible MLE		feasible MLE		infeasible MLE		feasible MLE		infeasible MLE	
N		TS	CS	TS	CS	TS	CS	TS	CS	TS	CS	TS	CS	TS	CS	TS	CS
	Mean	0.9917	0.8556	0.9909	0.8994	0.9922	0.8796	0.9966	0.8994	0.9877	0.8965	0.9897	0.9001	0.9878	0.8982	0.9918	0.9000
50	Bias	0.0917	-0.0444	0.0909	-0.0006	0.0922	-0.0204	0.0966	-0.0006	0.0877	-0.0035	0.0897	0.0001	0.0878	-0.0018	0.0918	0.0000
	RMSE	0.0917	0.0458	0.0909	0.0085	0.0922	0.0218	0.0966	0.0062	0.0877	0.0045	0.0897	0.0027	0.0878	0.0027	0.0918	0.0020
	size	100%	96%	100%	4.8%	100%	73%	100%	5.2%	100%	22%	100%	5.4%	100%	13%	100%	5%
	Mean	0.9934	0.8563	1.0046	0.8996	0.9889	0.8804	0.9979	0.9001	0.9877	0.8966	0.9913	0.9000	0.9885	0.8983	0.9928	0.9000
100	Bias	0.0934	-0.0437	0.1046	-0.0004	0.0889	-0.0196	0.0979	0.0001	0.0877	-0.0034	0.0913	0.0000	0.0885	-0.0017	0.0928	0.0000
	RMSE	0.0934	0.0446	0.1046	0.0063	0.0889	0.0203	0.0979	0.0044	0.0877	0.0039	0.0913	0.0019	0.0885	0.0022	0.0928	0.0014
	size	100%	99%	100%	5.5%	100%	96%	100%	5%	100%	41%	100%	4.9%	100%	21%	100%	4.7%
	Mean	0.9854	0.8584	0.9792	0.9002	0.9854	0.8810	0.9909	0.9001	0.9871	0.8965	0.9917	0.9000	0.9875	0.8983	0.0928	0.9000
500	Bias	0.0854	-0.0416	0.0742	0.0002	0.0854	-0.0190	0.0909	0.0001	0.0871	-0.0035	0.0917	0.0000	0.0875	-0.0017	0.0928	0.0000
	RMSE	0.0854	0.0418	0.0742	0.0028	0.0854	0.0192	0.0909	0.0019	0.0871	0.0036	0.0197	0.0009	0.0875	0.0017	0.0928	0.0006
	size	100%	100%	100%	5.4%	100%	100%	100%	4%	100%	97%	100%	4.6%	100%	76%	100%	5.1%

Notes: The size is calculated at $H_0 : \gamma_0 = 0.9$. Refer to notes of Table 1.

References

[1] 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.

[2] 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.

[3] Anderson, T.W., 1971, *The Statistical Analysis of Time Series*, John Wiley & Sons, Inc.

[4] Anderson, T.W., and C. Hsiao, 1981, Estimation of dynamic models with error components, *Journal of the American Statistical Association* 76, 598–606.

[5] Anderson, T.W., and C. Hsiao, 1982, Formulation and estimation of dynamic models using panel data, *Journal of Econometrics* 18, 47–82.

[6] Anderson, T.W. and Rubin, H., 1956, Statistical Inference in Factor Analysis. In: Neyman, J., Ed., *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 5, Berkeley, 111-150.

[7] Arellano, M., and O. Bover, 1995, Another look at the instrumental variable estimation of error-components models, *Journal of Econometrics* 68, 29-51.

[8] Bai J., 2003, Inferential theory for factor models of large dimensions, *Econometrica,* 71, 135-171.

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

[10] Balestra, P., and M. Nerlove, 1966, Pooling Cross-Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas, *Econometrica*, 34, 585–612.

[11] Bao, Y., and A. Ullah, 2010, Expectation of quadratic forms in normal and nonnormal variables with applications, *Journal of Statistical Planning and Inference* 140, 1193–1205.

[12] Chang Y., Y. Choi, C. Kim, J.I. Miller and J. Park, 2016, Disentangling Temporal Patterns in Elasticities: A Functional Coefficient Panel Anaylsis of Electricity Demand, *Energy Economics*, 60, 232-243

[13] Chang Y., Y. Choi, C. Kim, Z. Miller and J. Park, 2021, Forecasting regional long-run energy demand: A functional coefficient panel approach, *Energy Economics*, 96, 105-117.

[14] Fang, Y., K. Loparo, and X. Feng, 1994, Inequalities for the trace of matrix product, *IEEE Transactions on Automatic Control* 39, 2489-2490.

[15] Han, C., P. Phillips, and D. Sul, 2014, X-Differencing and Dynamic Panel Model Estimation, Econometric Theory 30, 201-251.

[16] Hoch I., 1962, Estimation of Production Function Parameters Combining Time-Series and Cross-Section Data, *Econometrica*, 30, 34–53.

[17] Honore, B.E., 1993, Orthogonality conditions for Tobit models with fixed effects and lagged dependent variables, *Journal of Econometrics* 59, 35-61.

[18] Hsiao, C. 1974, Statistical Inference for a Modelwith Both Random Cross-Sectional and Time Effects, *International Economic Review*, 15, 12–30.

[19] Hsiao, C. 1975, Some Estimation Methods for a Random Coefficients Model, *Econometrica,* 43, 305–25.

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

[21] Hsiao, C., T.W. Appelbe, and C.R. Dineen, 1993, A General Framework for Panel Data Analysis – with an Application to Canadian Customer Dialed Long Distance Service, *Journal of Econometrics,* 59, 63–86.

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

[23] Hsiao, C., and Q. Zhou, 2018, Incidental Parameters, Initial Conditions and Sample Size in Statistical Inference for Dynamic Panel Data Models, *Journal of Econometrics* 207, 114-128.

[24] Jiang, B.,Yang, Y., Gao, J., and C. Hsiao, 2021, Recursive estimation in large panel data models:theory and practice. *Journal of Econometrics* 224, 439-465.

[25] Kuh, E., 1963, The Validity of Cross Sectionally Estimated Behavior Equations in Time Series Applications, *Econometrica*, 27, 197–214.

[26] Magnus, J.R., and H. Neudecker, 1999, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 2nd Edition, Wiley.

[27] Pesaran, M.H., 2009, Estimation and Inference in Large Heterogeneous Panels with Cross-Section Dependence, *Econometrica*, 74, 967–1012.

[28] Powell, J., 1986, Symmetrically Trimmed Least Squares Estimation for Tobit Models, *Econometrica* 54, 1435-1460.

[29] Swamy, P., 1970, Efficient Inference in a Random Coefficient Regression Model, *Econometrica*, 38, 311–23.

Appendix: Mathematical Derivations

A. Derivation of Theorem 4.1

Conditional on $\boldsymbol{\theta}$, for (4.11), we have

$$\sqrt{NT}\left(\hat{\gamma}_{MLE}^{TS}-\gamma\right)=\left(\frac{1}{NT}\sum_{i=1}^{N}\mathbf{y}_{i,-1}^{*\prime}\Omega^{-1}\mathbf{y}_{i,-1}^{*}\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbf{y}_{i,-1}^{*\prime}\Omega^{-1}\mathbf{u}_{i}^{*}. \tag{A.1}$$

where

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\mathbf{y}_{i,-1}^{*\prime}\Omega^{-1}\mathbf{y}_{i,-1}^{*}\\
=\;&\frac{1}{\sigma_u^2}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{y}_{i,-1}^{\prime}\left(\mathbf{I}_{T-1}-\boldsymbol{\theta}\left(1+\boldsymbol{\theta}'\boldsymbol{\theta}\right)^{-1}\boldsymbol{\theta}'\right)\mathbf{y}_{i,-1}-\frac{1}{\sigma_u^2}\frac{1}{1+\boldsymbol{\theta}'\boldsymbol{\theta}}\frac{2}{NT}\sum_{i=1}^{N}\boldsymbol{\theta}'\mathbf{y}_{i,-1}y_{i0}\\
&+\frac{1}{\sigma_u^2}\frac{\boldsymbol{\theta}'\boldsymbol{\theta}}{1+\boldsymbol{\theta}'\boldsymbol{\theta}}\frac{1}{NT}\sum_{i=1}^{N}y_{i0}^2\\
=\;&\frac{1}{\sigma_u^2}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{y}_{i,-1}^{\prime}\left(\mathbf{I}_{T-1}-\boldsymbol{\theta}\left(1+\boldsymbol{\theta}'\boldsymbol{\theta}\right)^{-1}\boldsymbol{\theta}'\right)\mathbf{y}_{i,-1}+O_p\left(\frac{1}{T}\right)\\
=\;&\frac{1}{\sigma_u^2}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{y}_{i,-1}^{\prime}\mathbf{y}_{i,-1}-\frac{1}{\sigma_u^2}\frac{1}{NT}\sum_{i=1}^{N}\mathbf{y}_{i,-1}^{\prime}\boldsymbol{\theta}\left(1+\boldsymbol{\theta}'\boldsymbol{\theta}\right)^{-1}\boldsymbol{\theta}'\mathbf{y}_{i,-1}+O_p\left(\frac{1}{T}\right),
\end{aligned}\tag{A.2}$$

where the penultimate identity holds since $\boldsymbol{\theta}'\boldsymbol{\theta}=O\left(T\right).$

By continuous substitution,

$$y_{it}=\gamma^{t}y_{i0}+\lambda_i\sum_{s=1}^{t}\gamma^{t-s}f_s+\sum_{s=1}^{t}\gamma^{t-s}u_{is}, \tag{A.3}$$

then we have

$$\begin{aligned}
&\lim_{(N,T)\to\infty}\frac{1}{NT}\sum_{i=1}^{N}E\left(\mathbf{y}_{i,-1}^{\prime}\mathbf{y}_{i,-1}\right)\\
=\;&\lim_{(N,T)\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\gamma^{2t}E\left(y_{i0}^2\right)+\lim_{(N,T)\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\lambda_i^2\left(\sum_{s=1}^{t}\gamma^{t-s}f_s\right)^2\\
&+\lim_{(N,T)\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\left(\sum_{s=1}^{t}\gamma^{t-s}u_{is}\right)^2\\
=\;&\bar{\sigma}_\lambda^2\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T-1}\left(\sum_{s=1}^{t}\gamma^{t-s}f_s\right)^2+\frac{\sigma_u^2}{1-\gamma^2},
\end{aligned}\tag{A.4}$$

where $\bar{\sigma}_\lambda^2 = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \lambda_i^2$.

Also, for the second term of (A.2), we have

$$\frac{1}{NT}\sum_{i=1}^N \mathbf{y}_{i,-1}'\boldsymbol{\theta}\left(1+\boldsymbol{\theta}'\boldsymbol{\theta}\right)^{-1}\boldsymbol{\theta}'\mathbf{y}_{i,-1} = \frac{T}{1+\boldsymbol{\theta}'\boldsymbol{\theta}}\frac{1}{N}\sum_{i=1}^N\left(\frac{1}{T}\mathbf{y}_{i,-1}'\boldsymbol{\theta}\right)^2, \tag{A.5}$$

where as $T\to\infty$, $T\left(1+\boldsymbol{\theta}'\boldsymbol{\theta}\right)^{-1}$ will converge to a constant and

$$\frac{1}{T}\mathbf{y}_{i,-1}'\boldsymbol{\theta} = \lambda_i\frac{1}{T}\sum_{t=1}^{T-1}\theta_{t+1}\sum_{s=1}^t\gamma^{t-s}f_s + o_p(1) = O_p(1),$$

it follows that

$$\begin{aligned}\frac{1}{NT}\sum_{i=1}^N \mathbf{y}_{i,-1}'\boldsymbol{\theta}\left(1+\boldsymbol{\theta}'\boldsymbol{\theta}\right)^{-1}\boldsymbol{\theta}'\mathbf{y}_{i,-1} &= \frac{T}{1+\boldsymbol{\theta}'\boldsymbol{\theta}}\frac{1}{N}\sum_{i=1}^N\left(\lambda_i\frac{1}{T}\sum_{t=1}^{T-1}\theta_{t+1}\sum_{s=1}^t\gamma^{t-s}f_s\right)^2 + o_p(1)\\ &= \bar{\sigma}_\lambda^2\frac{1}{T\left(1+\boldsymbol{\theta}'\boldsymbol{\theta}\right)}\left(\sum_{t=1}^{T-1}\theta_{t+1}\sum_{s=1}^t\gamma^{t-s}f_s\right)^2 + o_p(1) = O_p(1)\end{aligned} \tag{A.6}$$

Combining (A.4) and (A.6), and letting $\omega_t = \sum_{s=1}^t\gamma^{t-s}f_s$ (which is $O(1)$) and $\boldsymbol{\omega} = (\omega_1,\ldots,\omega_{T-1})'$ yields (4.12).

By similar manipulaiton and taking account of u_{it} is i.i.d, we can show (4.13).

Similarly, for $\frac{1}{\sqrt{NT}}\sum_{i=1}^N \mathbf{y}_{i,-1}^{*\prime}\Omega^{-1}\mathbf{u}_i^*$, we have

$$\begin{aligned}\frac{1}{\sqrt{NT}}\sum_{i=1}^N \mathbf{y}_{i,-1}^{*\prime}\Omega^{-1}\mathbf{u}_i^* &= \frac{1}{\sigma_u^2}\frac{1}{\sqrt{NT}}\sum_{i=1}^N\left(\mathbf{y}_{i,-1}'-\boldsymbol{\theta}'y_{i0}\right)\left(\mathbf{I}_{T-1}-\boldsymbol{\theta}\left(1+\boldsymbol{\theta}'\boldsymbol{\theta}\right)^{-1}\boldsymbol{\theta}'\right)\left(\mathbf{u}_i-\boldsymbol{\theta}u_{i1}\right)\\ &= \frac{1}{\sigma_u^2}\frac{1}{\sqrt{NT}}\sum_{i=1}^N\mathbf{y}_{i,-1}'\mathbf{u}_i - \frac{1}{\sigma_u^2}\frac{1}{1+\boldsymbol{\theta}'\boldsymbol{\theta}}\frac{1}{\sqrt{NT}}\sum_{i=1}^N\mathbf{y}_{i,-1}'\boldsymbol{\theta}\boldsymbol{\theta}'\mathbf{u}_i\\ &\quad -\frac{1}{\sigma_u^2}\frac{1}{1+\boldsymbol{\theta}'\boldsymbol{\theta}}\frac{1}{\sqrt{NT}}\sum_{i=1}^N\mathbf{y}_{i,-1}'u_{i1}\boldsymbol{\theta} - \frac{1}{\sigma_u^2}\frac{1}{1+\boldsymbol{\theta}'\boldsymbol{\theta}}\frac{1}{\sqrt{NT}}\sum_{i=1}^N\boldsymbol{\theta}'y_{i0}\mathbf{u}_i\\ &\quad +\frac{1}{\sigma_u^2}\frac{\boldsymbol{\theta}'\boldsymbol{\theta}}{1+\boldsymbol{\theta}'\boldsymbol{\theta}}\frac{1}{\sqrt{NT}}\sum_{i=1}^N y_{i0}u_{i1}\\ &= A_1+A_2+A_3+A_4+A_5,\end{aligned}$$

where A_1 will contribute to the limiting distribution.

For A_2, it will contribute to the asymptotical bias since

$$E\left(A_2\right) = -\frac{\boldsymbol{\theta}'\Psi\boldsymbol{\theta}}{1+\boldsymbol{\theta}'\boldsymbol{\theta}}\sqrt{\frac{N}{T}}, \tag{A.7}$$

by following the derivation in the main paper and conditional on $\boldsymbol{\theta}$,

$$
\begin{aligned}
E\left[\dot{\mathbf{u}}_i \dot{\mathbf{y}}_{i,-1}'\right] &= E\begin{pmatrix} u_{i2}y_{i1} & u_{i2}y_{i2} & \cdots & u_{i2}y_{i,T-1} \\ u_{i3}y_{i1} & u_{i3}y_{i2} & \cdots & u_{i3}y_{i,T-1} \\ \vdots & \vdots & \ddots & \vdots \\ u_{iT}y_{i1} & u_{iT}y_{i2} & \cdots & u_{iT}y_{i,T-1} \end{pmatrix} \\
&= \sigma_u^2 \begin{pmatrix} 0 & 1 & \cdots & \gamma^{T-3} \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1 \\ 0 & 0 & \cdots & 0 \end{pmatrix} = \sigma_u^2 \Psi. \qquad \text{(A.8)}
\end{aligned}
$$

Also, it can be shown that $Var\left(A_2\right) = O\left(\frac{1}{T}\right) = o\left(1\right)$. Thus, we obtain

$$
A_2 \to_p -\frac{\boldsymbol{\theta}'\Psi\boldsymbol{\theta}}{1+\boldsymbol{\theta}'\boldsymbol{\theta}}\sqrt{\frac{N}{T}},
$$

as $(N,T) \to \infty$.

For A_3, it can be shown to be $o_p\left(1\right)$ as $(N,T)\to\infty$ since

$$
\begin{aligned}
E\left(A_3\right) &= -\frac{1}{\sigma_u^2}\frac{1}{1+\boldsymbol{\theta}'\boldsymbol{\theta}}\frac{1}{\sqrt{NT}}\sum_{i=1}^N E\left(\mathbf{y}_{i,-1}'u_{i1}\right)\boldsymbol{\theta} \\
&= -\frac{1}{1+\boldsymbol{\theta}'\boldsymbol{\theta}}\frac{1}{\sqrt{NT}}\sum_{i=1}^N\left(1,\gamma,\ldots,\gamma^{T-2}\right)\boldsymbol{\theta} \\
&= O\left(\frac{N^{1/2}}{T^{3/2}}\right) = o\left(1\right),
\end{aligned}
$$

as long as $\frac{N}{T^3} \to 0$ (which is true if $\frac{N}{T} \to a \neq 0$ as $(N,T)\to\infty$) and because $1+\boldsymbol{\theta}'\boldsymbol{\theta} = O\left(T\right)$ and $E\left(A_3^2\right) = O\left(\frac{1}{T}\right) + O\left(\frac{N}{T^2}\right) = o\left(1\right)$, which implies $A_3 = o_p\left(1\right)$.

For A_4, we have

$$
E\left(A_4\right) = -\frac{1}{\sigma_u^2}\frac{1}{1+\boldsymbol{\theta}'\boldsymbol{\theta}}\frac{1}{\sqrt{NT}}\sum_{i=1}^N E\left(y_{i0}\boldsymbol{\theta}'\mathbf{u}_i\right) = 0,
$$

and $E\left(A_4^2\right) = O\left(\frac{1}{T^2}\right)$, which implies $A_4 = o_p\left(1\right)$.

For A_5, we have

$$
E\left(A_5\right) = \frac{1}{\sigma_u^2}\frac{\boldsymbol{\theta}'\boldsymbol{\theta}}{1+\boldsymbol{\theta}'\boldsymbol{\theta}}\frac{1}{\sqrt{NT}}\sum_{i=1}^N E\left(y_{i0}u_{i1}\right) = 0,
$$

and

$$
\begin{aligned}
E\left(A_5^2\right) &= \frac{1}{\sigma_u^4}\left(\frac{\boldsymbol{\theta}'\boldsymbol{\theta}}{1+\boldsymbol{\theta}'\boldsymbol{\theta}}\right)^2 \frac{1}{NT}\sum_{i,j} E\left(y_{i0}u_{i1}y_{j0}u_{j1}\right) \\
&= \frac{1}{\sigma_u^4}\left(\frac{\boldsymbol{\theta}'\boldsymbol{\theta}}{1+\boldsymbol{\theta}'\boldsymbol{\theta}}\right)^2 \frac{1}{NT}\sum_{i} E\left(y_{i0}^2u_{i1}^2\right) = O\left(\frac{1}{T}\right),
\end{aligned}
$$

which gives $A_5 = o_p(1)$.

Combining these result yields (4.14). Following the derivation of Anderson (1971, Ch 5), we can show that

$$
\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} \mathbf{y}_{i,-1}'\mathbf{u}_i \to_d N\left(0, \frac{\sigma_y^2}{\sigma_u^2}\right),
$$

where

$$
\sigma_y^2 = \frac{1}{T}\sum_{t=1}^{T} E\left(y_{it}^2\right) \to \bar{\sigma}_\lambda^2\tilde{\sigma}_{\gamma,f}^2 + \frac{\sigma_u^2}{1-\gamma^2},
$$

$\tilde{\sigma}_{\gamma,f}^2 = \lim_{T\to\infty}\frac{1}{T}\boldsymbol{\omega}'\boldsymbol{\omega}$ and $\boldsymbol{\omega} = (\omega_1,\ldots,\omega_{T-1})'$ with $\omega_t = \sum_{s=1}^{t}\gamma^{t-s}f_s$.

Equivalently, we have

$$
\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} \mathbf{y}_{i,-1}^{*}(\boldsymbol{\theta})'\,\Omega(\boldsymbol{\theta})^{-1}\,\mathbf{u}_i^{*}(\boldsymbol{\theta}) \to_d N\left(-b\sqrt{\frac{N}{T}}, \frac{\bar{\sigma}_\lambda^2}{\sigma_u^2}\tilde{\sigma}_{\gamma,f}^2 + \frac{1}{1-\gamma^2}\right), \tag{A.9}
$$

where

$$
b = \lim_{T\to\infty}\frac{\boldsymbol{\theta}'\Psi\boldsymbol{\theta}}{1+\boldsymbol{\theta}'\boldsymbol{\theta}}, \tag{A.10}
$$

denoting the bias term with $\boldsymbol{\theta} = (\theta_2,\ldots,\theta_T)'$ is a $(T-1)\times 1$ vector with unrestricted parameters and Ψ is provided in (A.8), $\bar{\sigma}_\lambda^2$ is defined as before and $\tilde{\sigma}_{\gamma,f}^2 = \lim_{T\to\infty}\frac{1}{T}\boldsymbol{\omega}'\boldsymbol{\omega}$ with $\boldsymbol{\omega} = (\omega_1,\ldots,\omega_{T-1})'$ and $\omega_t = \sum_{s=1}^{t}\gamma^{t-s}f_s$.

Combining (A.6) and (A.9) yields Theorem 4.1 as required.

B. QDMLE for Model with Multiple Factors

When $r > 1$, the regression error v_{it} takes the form of

$$
v_{it} = \boldsymbol{\lambda}_i'\mathbf{f}_t + u_{it}, \tag{A.11}
$$

where $\boldsymbol{\lambda}_i = (\lambda_{i1},\ldots,\lambda_{ir})'$ denotes the factor loadings while $\mathbf{f}_t = (f_{1t},\ldots,f_{rt})'$ denotes unobservable common factors. Let $\mathbf{v}_i = (v_{i1},\ldots,v_{iT})'$, then (A.11) can be rewritten in vector form as

$$
\mathbf{v}_i = \mathbf{F}\boldsymbol{\lambda}_i + \mathbf{u}_i, \quad i = 1,2,\ldots,N, \tag{A.12}
$$

where $\mathbf{F} = (\mathbf{f}_1,\ldots,\mathbf{f}_T)'$ denotes the $T\times r$ matrix of r time-specific common factors, $\mathbf{f}_t$, over time and $\mathbf{u}_i = (u_{i1},\ldots,u_{iT})'$. Similarly, let $\mathbf{v}_t = (v_{1t},\ldots,v_{Nt})'$, then (A.11) can be rewritten in vector form as

$$
\mathbf{v}_t = \Lambda\mathbf{f}_t + \mathbf{u}_t, \tag{A.13}
$$

where $\Lambda = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N)'$ denotes the $N \times r$ matrix of factor loading and $\mathbf{u}_t = (u_{1t}, \ldots, u_{Nt})'$.

For model (2.1) with multiple common factor structure (A.11), it can be stacked in vector form as

$$\mathbf{y}_i = \gamma \mathbf{y}_{i,-1} + \mathbf{F}\boldsymbol{\lambda}_i + \mathbf{u}_i, \tag{A.14}$$

where $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$, $\mathbf{y}_{i,-1} = (y_{i0}, \ldots, y_{i,T-1})'$ and $\mathbf{u}_i = (u_{i1}, \ldots, u_{iT})'$.

Note that $\mathbf{F}$ and $\boldsymbol{\lambda}_i$ in (A.12) or (A.14) are separately identified since for any full rank matrix $\mathbf{C}$, we have

$$\mathbf{F}\boldsymbol{\lambda}_i = \mathbf{F}\mathbf{C}\mathbf{C}^{-1}\boldsymbol{\lambda}_i = \mathbf{F}^*\boldsymbol{\lambda}_i^*, \tag{A.15}$$

which is the typical rotation problem for models with interactive effects (Bai (2009)). To avoid this problem, following ALS (2013), we use the normalization condition $\mathbf{F} = (-\mathbf{I}_r, \Xi')'$, where Ξ is a $(T-r) \times r$ matrix of unrestricted parameters. Define the $T \times (T-r)$ matrix

$$\mathbf{H}(\boldsymbol{\vartheta}) = (\Xi, \mathbf{I}_{T-r})' = [\mathbf{h}_1(\boldsymbol{\vartheta}_1), \mathbf{h}_2(\boldsymbol{\vartheta}_2), \ldots, \mathbf{h}_{T-r}(\boldsymbol{\vartheta}_{T-r})], \tag{A.16}$$

where $\boldsymbol{\vartheta} = vec(\Xi)$, $\boldsymbol{\vartheta}_j$ is the j-th column of Ξ', and $\mathbf{h}_j(\boldsymbol{\vartheta}_j)$ is the j-th column of $\mathbf{H}(\boldsymbol{\vartheta})$ for $j = 1, \ldots, T-r$. Given the definition of $\mathbf{H}(\boldsymbol{\vartheta})$, we have

$$\mathbf{H}(\boldsymbol{\vartheta})'\mathbf{F} = (\Xi, \mathbf{I}_{T-r})\left(-\mathbf{I}_r, \Xi'\right)' = 0_{(T-r)\times r}. \tag{A.17}$$

As a result, we can remove the interactive effects in (A.14) by multiplying $\mathbf{H}(\boldsymbol{\vartheta})'$ in both sides of (A.14), which in turn gives

$$\mathbf{H}(\boldsymbol{\vartheta})'\mathbf{y}_i = \mathbf{H}(\boldsymbol{\vartheta})'\mathbf{y}_{i,-1}\gamma + \mathbf{H}(\boldsymbol{\vartheta})'\mathbf{u}_i, \quad i = 1, \ldots, N. \tag{A.18}$$

For the transformed errors, we have

$$\mathbf{H}(\boldsymbol{\vartheta})'\mathbf{u}_i = \begin{pmatrix} \mathbf{h}_1(\boldsymbol{\vartheta}_1)'\mathbf{u}_i \\ \mathbf{h}_2(\boldsymbol{\vartheta}_2)'\mathbf{u}_i \\ \vdots \\ \mathbf{h}_{T-r}(\boldsymbol{\vartheta}_{T-r})'\mathbf{u}_i \end{pmatrix} = \begin{pmatrix} u_{i,r+1} + \boldsymbol{\vartheta}_1'\mathbf{u}_{i,1,r} \\ u_{i,r+2} + \boldsymbol{\vartheta}_2'\mathbf{u}_{i,1,r} \\ \vdots \\ u_{i,T} + \boldsymbol{\vartheta}_{T-r}'\mathbf{u}_{i,1,r} \end{pmatrix} = \mathbf{u}_{i,\underline{r+1,T}} + \Xi\mathbf{u}_{i,\underline{1,r}},$$

with $\mathbf{u}_{i,\underline{1,r}} = (u_{i1}, \ldots, u_{ir})'$ and $\mathbf{u}_{i,\underline{r+1,T}} = (u_{i,r+1}, \ldots, u_{iT})'$.

Similarly, we have

$$\mathbf{H}(\boldsymbol{\vartheta})'\mathbf{y}_i = \begin{pmatrix} \mathbf{h}_1(\boldsymbol{\vartheta}_1)'\mathbf{y}_i \\ \mathbf{h}_2(\boldsymbol{\vartheta}_2)'\mathbf{y}_i \\ \vdots \\ \mathbf{h}_{T-r}(\boldsymbol{\vartheta}_{T-r})'\mathbf{y}_i \end{pmatrix} = \begin{pmatrix} y_{i,r+1} + \boldsymbol{\vartheta}_1'\mathbf{y}_{i,1,r} \\ y_{i,r+2} + \boldsymbol{\vartheta}_2'\mathbf{y}_{i,1,r} \\ \vdots \\ y_{i,T} + \boldsymbol{\vartheta}_{T-r}'\mathbf{y}_{i,1,r} \end{pmatrix} = \mathbf{y}_{i,\underline{r+1,T}} + \Xi\mathbf{y}_{i,\underline{1,r}},$$

with $\mathbf{y}_{i,\underline{1,r}} = (y_{i1}, \ldots, y_{ir})'$ and $\mathbf{y}_{i,\underline{r+1,T}} = (y_{i,r+1}, \ldots, y_{iT})'$, and

$$\mathbf{H}(\boldsymbol{\vartheta})'\mathbf{y}_{i,-1} = \begin{pmatrix} \mathbf{h}_1(\boldsymbol{\vartheta}_1)'\mathbf{y}_{i,-1} \\ \mathbf{h}_2(\boldsymbol{\vartheta}_2)'\mathbf{y}_{i,-1} \\ \vdots \\ \mathbf{h}_{T-r}(\boldsymbol{\vartheta}_{T-r})'\mathbf{y}_{i,-1} \end{pmatrix} = \begin{pmatrix} y_{i,r} + \boldsymbol{\vartheta}_1'\mathbf{y}_{i,0,r-1} \\ y_{i,r+2} + \boldsymbol{\vartheta}_2'\mathbf{y}_{i,0,r-1} \\ \vdots \\ y_{i,T-1} + \boldsymbol{\vartheta}_{T-r}'\mathbf{y}_{i,0,r-1} \end{pmatrix} = \mathbf{y}_{i,\underline{r,T-1}} + \Xi\mathbf{y}_{i,\underline{0,r-1}},$$

with $\mathbf{y}_{i,\underline{0,r-1}} = (y_{i0}, \ldots, y_{i,r-1})'$ and $\mathbf{y}_{i,\underline{r,T-1}} = (y_{i,r+1}, \ldots, y_{iT})'$.

For the transformed model (A.18), the variance-covariance matrix is given by

$$\begin{aligned}
\underset{(T-r)\times(T-r)}{\Omega_{TS}} &= E\left[\mathbf{H}(\boldsymbol{\vartheta})' \mathbf{u}_i \mathbf{u}_i' \mathbf{H}(\boldsymbol{\vartheta})\right] = E\left[\left(\mathbf{u}_{i,r+1,T} + \Xi \mathbf{u}_{i,1,r}\right)\left(\mathbf{u}_{i,r+1,T} + \Xi \mathbf{u}_{i,1,r}\right)'\right] \\
&= \sigma_u^2 \left(\mathbf{I}_{T-r} + \Xi\Xi'\right). \qquad \text{(A.19)}
\end{aligned}$$

In order to obtain an explicit solution of $\left(\mathbf{I}_{T-r} + \Xi\Xi'\right)^{-1}$, assume Ξ has a singular value decomposition such that $\Xi = \mathbf{UDV}$ where $\mathbf{U}$ and $\mathbf{V}$ are $(T-r)\times(T-r)$ and $r \times r$ orthogonal matrices (Magnus and Neudecker (1999)), respectively, and $\mathbf{D}$ is the $(T-r)\times r$ matrix of singular values of Ξ on its diagonal

$$\mathbf{D} = \begin{pmatrix} diag\left(d_1, d_2, \ldots, d_r\right) \\ 0 \end{pmatrix},$$

where $d_1 \geq d_2 \geq \cdots \geq d_r$. Then it is obvious that $\Xi\Xi' = \mathbf{UDD'U'}$, and $\mathbf{DD'}$ is the $(T-r)\times(T-r)$ diagonal matrix

$$\mathbf{DD}' = \begin{pmatrix} diag\left(d_1^2, d_2^2, \ldots, d_r^2\right) & 0 \\ 0 & 0 \end{pmatrix},$$

and

$$\mathbf{I}_{T-r} + \Xi\Xi' = \mathbf{U}\left(\mathbf{I}_{T-r} + \mathbf{DD}'\right)\mathbf{U}',$$

thus

$$\left(\mathbf{I}_{T-r} + \Xi\Xi'\right)^{-1} = \mathbf{U}\begin{pmatrix} \mathring{\mathbf{D}} & 0 \\ 0 & \mathbf{I}_{T-2r} \end{pmatrix}\mathbf{U}',$$

with $\mathring{\mathbf{D}} = diag\left(\frac{1}{1+d_1^2}, \frac{1}{1+d_2^2}, \ldots, \frac{1}{1+d_r^2}\right)$.

As a result, we have

$$\begin{aligned}
\mathbf{H}(\boldsymbol{\vartheta})'\left(\mathbf{I}_{T-r} + \Xi\Xi'\right)^{-1}\mathbf{H}(\boldsymbol{\vartheta}) &= \begin{pmatrix} \mathbf{V'D'U'} \\ \mathbf{I}_{T-r} \end{pmatrix}\mathbf{U}\begin{pmatrix} \mathring{\mathbf{D}} & 0 \\ 0 & \mathbf{I}_{T-2r} \end{pmatrix}\mathbf{U}'\left(\mathbf{UDV}, \mathbf{I}_{T-r}\right) \\
&= \begin{pmatrix} \mathbf{V'D'} \\ \mathbf{U} \end{pmatrix}\begin{pmatrix} \mathring{\mathbf{D}} & 0 \\ 0 & \mathbf{I}_{T-2r} \end{pmatrix}\left(\mathbf{DV}, \mathbf{U}'\right). \qquad \text{(A.20)}
\end{aligned}$$

The quasi-log-likelihood function of (A.18) is given by

$$\log L(\gamma, \boldsymbol{\vartheta}) = -\frac{N}{2}\log\left|\Omega_{TS}\right| - \frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{y}_i - \mathbf{y}_{i,-1}\gamma\right)'\mathbf{H}(\boldsymbol{\vartheta})\,\Omega_{TS}{}^{-1}\mathbf{H}(\boldsymbol{\vartheta})'\left(\mathbf{y}_i - \mathbf{y}_{i,-1}\gamma\right). \quad \text{(A.21)}$$

Conditional on $\boldsymbol{\vartheta}$, the FOC for γ is

$$\frac{\partial \log L(\gamma, \boldsymbol{\vartheta})}{\partial \gamma} = \sum_{i=1}^{N}\mathbf{y}_{i,-1}'\mathbf{H}(\boldsymbol{\vartheta})\,\Omega_{TS}{}^{-1}\mathbf{H}(\boldsymbol{\vartheta})'\left(\mathbf{y}_i - \mathbf{y}_{i,-1}\gamma\right) = 0, \qquad \text{(A.22)}$$

which yields

$$\hat{\gamma}_{MLE}^{TS,M} = \left(\sum_{i=1}^{N} \mathbf{y}_{i,-1}' \mathbf{H}(\boldsymbol{\vartheta}) \, \Omega_{TS}{}^{-1} \mathbf{H}(\boldsymbol{\vartheta})' \, \mathbf{y}_{i,-1} \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{y}_{i,-1}' \mathbf{H}(\boldsymbol{\vartheta}) \, \Omega_{TS}{}^{-1} \mathbf{H}(\boldsymbol{\vartheta})' \, \mathbf{y}_{i} \right), \quad \text{(A.23)}$$

where ^{TS,M} refers to using normalization (A.17) to remove the multiple interactive effects.

The denominator of (A.23) divided by NT can be shown to converge to a nonzero constant. For the numerator, we can observe that

$$\begin{aligned} E\left[\mathbf{y}_{i,-1}' \mathbf{H}(\boldsymbol{\vartheta}) \, \Omega_{TS}{}^{-1} \mathbf{H}(\boldsymbol{\vartheta})' \, \mathbf{u}_i\right] &= tr\left\{\mathbf{H}(\boldsymbol{\vartheta}) \, \Omega_{TS}{}^{-1} \mathbf{H}(\boldsymbol{\vartheta})' \, E\left[\mathbf{u}_i \mathbf{y}_{i,-1}'\right]\right\} \\ &= tr\left\{\mathbf{H}(\boldsymbol{\vartheta}) \left(\mathbf{I}_{T-r} + \Xi\Xi'\right)^{-1} \mathbf{H}(\boldsymbol{\vartheta})' \, \dot{\Psi}\right\}, \end{aligned}$$

by using the derivation of (A.8) and

$$E\left[\mathbf{u}_i \mathbf{y}_{i,-1}'\right] = \sigma_u^2 \begin{pmatrix} 0 & 1 & \cdots & \gamma^{T-2} \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1 \\ 0 & 0 & \cdots & 0 \end{pmatrix} = \sigma_u^2 \dot{\Psi}. \quad \text{(A.24)}$$

Using the results of (A.20) yields

$$\begin{aligned} tr\left\{\mathbf{H}(\boldsymbol{\vartheta}) \left(\mathbf{I}_{T-r} + \Xi\Xi'\right)^{-1} \mathbf{H}(\boldsymbol{\vartheta})' \, \dot{\Psi}\right\} &= tr\left[\begin{pmatrix} \mathring{\mathbf{D}} & 0 \\ 0 & \mathbf{I}_{T-2r} \end{pmatrix} \left(\mathbf{DV}, \mathbf{U}'\right) \dot{\Psi} \begin{pmatrix} \mathbf{V}'\mathbf{D}' \\ \mathbf{U} \end{pmatrix}\right] \\ &\leq Ctr\left[\dot{\Psi} \begin{pmatrix} \mathbf{V}'\mathbf{D}' \\ \mathbf{U} \end{pmatrix} \left(\mathbf{DV}, \mathbf{U}'\right)\right] \\ &\leq Ctr\left[\begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{pmatrix} \begin{pmatrix} \mathbf{V}'\mathbf{D}'\mathbf{DV} & \mathbf{V}'\mathbf{D}'\mathbf{U}' \\ \mathbf{UDV} & \mathbf{I}_{T-r} \end{pmatrix}\right] \end{aligned}$$

where we partition $\dot{\Psi}$ as follows

$$\Psi_{11} = \begin{pmatrix} 0 & 1 & \gamma & \cdots & \gamma^{r-2} \\ 0 & 0 & 1 & \cdots & \gamma^{r-3} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}_{r\times r}, \Psi_{12} = \begin{pmatrix} \gamma^{r-1} & \gamma^{r} & ... & \gamma^{T-2} \\ \gamma^{r-2} & \gamma^{r-1} & ... & \gamma^{T-3} \\ \vdots & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \vdots \\ 1 & \gamma & \cdots & \gamma^{T-r-1} \end{pmatrix}_{r\times(T-r)},$$

$$\Psi_{21} = 0_{(T-r)\times r}, \Psi_{22} = \begin{pmatrix} 0 & 1 & \gamma & \cdots & \gamma^{T-r-2} \\ 0 & 0 & 1 & \cdots & \gamma^{T-r-3} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}_{(T-r)\times(T-r)},$$

then

$$\begin{pmatrix} \Psi_{11} & \Psi_{12} \\ 0 & \Psi_{22} \end{pmatrix} \begin{pmatrix} \mathbf{V}'\mathbf{D}'\mathbf{D}\mathbf{V} & \mathbf{V}'\mathbf{D}'\mathbf{U}' \\ \mathbf{U}\mathbf{D}\mathbf{V} & \mathbf{I}_{T-r} \end{pmatrix} = \begin{pmatrix} \Psi_{11}\mathbf{V}'\mathbf{D}'\mathbf{D}\mathbf{V} + \Psi_{12}\mathbf{U}\mathbf{D}\mathbf{V} & \Psi_{11}\mathbf{V}'\mathbf{D}'\mathbf{U}' + \Psi_{12} \\ \Psi_{22}\mathbf{U}\mathbf{D}\mathbf{V} & \Psi_{22} \end{pmatrix}.$$

Taking trace, we obtain

$$\begin{aligned} tr\left[\begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{pmatrix} \begin{pmatrix} \mathbf{V}'\mathbf{D}'\mathbf{D}\mathbf{V} & \mathbf{V}'\mathbf{D}'\mathbf{U}' \\ \mathbf{U}\mathbf{D}\mathbf{V} & \mathbf{I}_{T-r} \end{pmatrix}\right] &= tr(\Psi_{11}\mathbf{V}'\mathbf{D}'\mathbf{D}\mathbf{V} + \Psi_{12}\mathbf{U}\mathbf{D}\mathbf{V}) + tr(\Psi_{22}) \\ &= tr(\Psi_{11}\mathbf{V}'\mathbf{D}'\mathbf{D}\mathbf{V}) + tr(\Psi_{12}\Xi), \end{aligned} \tag{A.25}$$

since $tr(\Psi_{22}) = 0$.

Using the result of Fang et al (1994, p.2489), we have

$$\begin{aligned} |tr(\Psi_{11}\mathbf{V}'\mathbf{D}'\mathbf{D}\mathbf{V}) &\leq (|\lambda_{\max}(\frac{\Psi_{11}+\Psi_{11}'}{2})| + |\lambda_{\min}(\frac{\Psi_{11}+\Psi_{11}'}{2})|)tr(\mathbf{V}'\mathbf{D}'\mathbf{D}\mathbf{V}) \\ &= (|\lambda_{\max}(\frac{\Psi_{11}+\Psi_{11}'}{2})| + |\lambda_{\min}(\frac{\Psi_{11}+\Psi_{11}'}{2})|)\sum_{i=1}^{r} d_i^2 \\ &= O(1), \end{aligned}$$

the last equation follows the fact $\frac{\Psi_{11}+\Psi_{11}'}{2}$ is a $r \times r$ matrix with bounded elements, so it has bounded smallest and largest eigenvalues. For the last term of (A.25), we notice that the diagonal element $a_{ii}, i = 1, 2, \ldots, r$, of $\Psi_{12}\Xi$ is given by

$$a_{ii} = \frac{1}{\gamma^{i+1}} \sum_{t=1}^{T-r} \gamma^{r+t} \theta_{ti},$$

thus

$$\begin{aligned} |a_{ii}| &\leq \frac{1}{\gamma^{i+1}} \sum_{t=1}^{T-r} \gamma^{r+t} |\theta_{ti}| \leq \frac{1}{\gamma} \sum_{t=1}^{T-r} \gamma^{t} |\theta_{ti}| \\ &= O(1), \end{aligned}$$

as $T \to \infty$, which in turn yields

$$\begin{aligned} |tr(\Psi_{12}\Xi)| &= \left|\sum_{i=1}^{r} a_{ii}\right| \leq \sum_{i=1}^{r} |a_{ii}| \\ &= O(1), \end{aligned}$$

as long as r is finite.

Combining these results we obtain

$$tr\left\{\mathbf{H}(\boldsymbol{\vartheta})\left(\mathbf{I}_{T-r} + \Xi\Xi'\right)^{-1}\mathbf{H}(\boldsymbol{\vartheta})'\dot{\Psi}\right\} = O(1), \tag{A.26}$$

as $T \to \infty$.

Substituting (A.26) into (A.23), we get $\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \left(\mathbf{y}_{i,-1}' \mathbf{H}(\boldsymbol{\vartheta}) \Omega_{TS}(\boldsymbol{\vartheta})^{-1} \mathbf{H}(\boldsymbol{\vartheta})' \mathbf{u}_i \right) = O_p\left(\sqrt{\frac{N}{T}} \right)$, i.e., the QDMLE for γ using normalization (A.17) is asymptotically biased of order $\sqrt{\frac{N}{T}}$.

Alternatively, model (2.1) with (A.11) can be rewritten in vector form as

$$\mathbf{y}_t = \gamma \mathbf{y}_{t-1} + \Lambda \mathbf{f}_t + \mathbf{u}_t, \quad t = 1, \ldots, T, \tag{A.27}$$

where $\mathbf{y}_t = (y_{1t}, \ldots, y_{Nt})'$, $\mathbf{y}_{t-1} = (y_{1,t-1}, \ldots, y_{N,t-1})'$, $\mathbf{u}_t = (u_{1t}, \ldots, u_{Nt})'$ and $\Lambda = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N)'$ is a $N \times r$ matrix of factor loading.

For model (A.27), we can apply the quasi-difference pairwise proposed before. To this end, let the factor loading matrix Λ satisfies the normalization such that $\Lambda = (-\mathbf{I}_r, \Psi')'$, where Ψ is an $(N-r) \times r$ matrix of unrestricted parameters. Define the $N \times (N-r)$ matrix

$$\mathbf{G}(\boldsymbol{\psi}) = (\Psi, \mathbf{I}_{N-r})' = \left[\mathbf{g}_1(\boldsymbol{\psi}_1), \mathbf{g}_2(\boldsymbol{\psi}_2), \ldots, \mathbf{g}_{N-r}(\boldsymbol{\psi}_{N-r}) \right], \tag{A.28}$$

where $\boldsymbol{\psi} = vec(\Psi)$, $\boldsymbol{\psi}_j$ is the j-th column of Ψ', and $\mathbf{g}_j(\boldsymbol{\psi}_j)$ is the j-th column of $\mathbf{G}(\boldsymbol{\psi})$ for $j = 1, \ldots, N-r$. Given the definition of $\mathbf{G}(\boldsymbol{\psi})$, we have

$$\mathbf{G}(\boldsymbol{\psi})' \Lambda = (\Psi, \mathbf{I}_{N-r}) \left(-\mathbf{I}_r, \Psi' \right)' = 0_{(N-r) \times r}. \tag{A.29}$$

As a result, we can remove the interactive effects in (A.27) by multiplying $\mathbf{G}(\boldsymbol{\psi})'$ in both sides of (A.27), which in turn gives

$$\mathbf{G}(\boldsymbol{\psi})' \mathbf{y}_t = \gamma \mathbf{G}(\boldsymbol{\psi})' \mathbf{y}_{t-1} + \mathbf{G}(\boldsymbol{\psi})' \mathbf{u}_t, \quad t = 1, \ldots, T, \tag{A.30}$$

or

$$\check{\mathbf{y}}_t(\boldsymbol{\psi}) = \gamma \check{\mathbf{y}}_{t-1}(\boldsymbol{\psi}) + \check{\mathbf{u}}_t(\boldsymbol{\psi}), \quad t = 1, \ldots, T, \tag{A.31}$$

where $\check{\mathbf{y}}_t(\boldsymbol{\psi}) = \mathbf{G}(\boldsymbol{\psi})' \mathbf{y}_t$, $\check{\mathbf{y}}_{t-1}(\boldsymbol{\psi}) = \mathbf{G}(\boldsymbol{\psi})' \mathbf{y}_{t-1}$ and $\check{\mathbf{u}}_t(\boldsymbol{\psi}) = \mathbf{G}(\boldsymbol{\psi})' \mathbf{u}_t$.

For the transformed errors, we have

$$\check{\mathbf{u}}_t(\boldsymbol{\psi}) = \mathbf{G}(\boldsymbol{\psi})' \mathbf{u}_t = \begin{pmatrix} \mathbf{g}_1(\boldsymbol{\psi}_1)' \mathbf{u}_t \\ \mathbf{g}_2(\boldsymbol{\psi}_2)' \mathbf{u}_t \\ \vdots \\ \mathbf{g}_{N-r}(\boldsymbol{\psi}_{N-r})' \mathbf{u}_t \end{pmatrix} = \begin{pmatrix} u_{r+1,t} + \boldsymbol{\psi}_1' \mathbf{u}_{1,r,t} \\ u_{r+2,t} + \boldsymbol{\psi}_2' \mathbf{u}_{1,r,t} \\ \vdots \\ u_{N,t} + \boldsymbol{\psi}_{N-r}' \mathbf{u}_{1,r,t} \end{pmatrix} = \mathbf{u}_{\underline{r+1,N},t} + \Psi \mathbf{u}_{\underline{1,r},t},$$

with $\mathbf{u}_{\underline{1,r},t} = (u_{1t}, \ldots, u_{rt})'$ and $\mathbf{u}_{\underline{r+1,N},t} = (u_{r+1,t}, \ldots, u_{Nt})'$.

Similarly, we have

$$\check{\mathbf{y}}_t(\boldsymbol{\psi}) = \mathbf{G}(\boldsymbol{\psi})' \mathbf{y}_t = \begin{pmatrix} \mathbf{g}_1(\boldsymbol{\psi}_1)' \mathbf{y}_t \\ \mathbf{g}_2(\boldsymbol{\psi}_2)' \mathbf{y}_t \\ \vdots \\ \mathbf{g}_{N-r}(\boldsymbol{\psi}_{N-r})' \mathbf{y}_t \end{pmatrix} = \begin{pmatrix} y_{r+1,t} + \boldsymbol{\psi}_1' \mathbf{y}_{1,r,t} \\ y_{r+2,t} + \boldsymbol{\psi}_2' \mathbf{y}_{1,r,t} \\ \vdots \\ y_{N,t} + \boldsymbol{\psi}_{N-r}' \mathbf{y}_{1,r,t} \end{pmatrix} = \mathbf{y}_{\underline{r+1,N},t} + \Psi \mathbf{y}_{\underline{1,r},t},$$

with $\mathbf{y}_{\underline{1,r},t} = (y_{1t}, \ldots, y_{rt})'$ and $\mathbf{y}_{\underline{r+1,N},t} = (y_{r+1,t}, \ldots, y_{Nt})'$, and

$$\check{\mathbf{y}}_{t-1}(\boldsymbol{\psi}) = \mathbf{G}(\boldsymbol{\psi})' \mathbf{y}_{t-1} = \begin{pmatrix} \mathbf{g}_1(\boldsymbol{\psi}_1)' \mathbf{y}_{t-1} \\ \mathbf{g}_2(\boldsymbol{\psi}_2)' \mathbf{y}_{t-1} \\ \vdots \\ \mathbf{g}_{N-r}(\boldsymbol{\psi}_{N-r})' \mathbf{y}_{t-1} \end{pmatrix} = \begin{pmatrix} y_{r+1,t-1} + \boldsymbol{\psi}_1' \mathbf{y}_{1,r,t-1} \\ y_{r+2,t-1} + \boldsymbol{\psi}_2' \mathbf{y}_{1,r,t-1} \\ \vdots \\ y_{N,t-1} + \boldsymbol{\psi}_{N-r}' \mathbf{y}_{1,r,t-1} \end{pmatrix} = \mathbf{u}_{\underline{r+1,N},t-1} + \Psi \mathbf{u}_{\underline{1,r},t-1},$$

with $\mathbf{y}_{\underline{1,r},t-1} = (y_{1,t-1}, \ldots, y_{r,t-1})'$ and $\mathbf{y}_{\underline{r+1,N},t-1} = (y_{r+1,t-1}, \ldots, y_{N,t-1})'$.

For the transformed model (A.30), the variance-covariance matrix is given by

$$\underset{(N-r)\times(N-r)}{\Omega_{CS}} = \sigma_u^2 \left(\mathbf{I}_{N-r} + \Psi\Psi' \right), \tag{A.32}$$

and

$$\Omega_{CS}^{-1} = \frac{1}{\sigma_u^2} \left(\mathbf{I}_{N-r} - \Psi \left(\mathbf{I}_r + \Psi'\Psi \right)^{-1} \Psi' \right). \tag{A.33}$$

The quasi-log-likelihood function of (A.30) is given by

$$\log L(\gamma, \boldsymbol{\psi}) = -\frac{N}{2} \log |\Omega_{CS}| - \frac{1}{2} \sum_{t=1}^{T} (\mathbf{y}_t - \gamma \mathbf{y}_{t-1})' \mathbf{G}(\boldsymbol{\psi})' \Omega_{CS}^{-1} \mathbf{G}(\boldsymbol{\psi}) (\mathbf{y}_t - \gamma \mathbf{y}_{t-1}). \tag{A.34}$$

Conditional on $\boldsymbol{\psi}$, the FOC for γ is

$$\frac{\partial \log L(\gamma, \boldsymbol{\psi})}{\partial \gamma} = \sum_{t=1}^{T} \mathbf{y}_{t-1}' \mathbf{G}(\boldsymbol{\psi})' \Omega_{CS}^{-1} \mathbf{G}(\boldsymbol{\psi}) (\mathbf{y}_t - \gamma \mathbf{y}_{t-1}) = 0, \tag{A.35}$$

which yields

$$\hat{\gamma}_{MLE}^{CS,M} = \left(\sum_{t=1}^{T} \mathbf{y}_{t-1}' \mathbf{G}(\boldsymbol{\psi})' \Omega_{CS}^{-1} \mathbf{G}(\boldsymbol{\psi}) \mathbf{y}_{t-1} \right)^{-1} \sum_{t=1}^{T} \left(\mathbf{y}_{t-1}' \mathbf{G}(\boldsymbol{\psi})' \Omega_{CS}^{-1} \mathbf{G}(\boldsymbol{\psi}) \mathbf{y}_t \right), \tag{A.36}$$

where ^{CS,M} refers to using normalization (A.29) to remove the multiple interactive effects.

For this QDMLE (A.36), we have

$$\begin{aligned} \sqrt{NT} \left(\hat{\gamma}_{MLE}^{CS,M} - \gamma \right) &= \left(\frac{1}{NT} \sum_{t=1}^{T} \mathbf{y}_{t-1}' \mathbf{G}(\boldsymbol{\psi})' \Omega_{CS}^{-1} \mathbf{G}(\boldsymbol{\psi}) \mathbf{y}_{t-1} \right)^{-1} \\ &\quad \times \left(\frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \mathbf{y}_{t-1}' \mathbf{G}(\boldsymbol{\psi})' \Omega_{CS}^{-1} \mathbf{G}(\boldsymbol{\psi}) \mathbf{u}_t \right). \end{aligned} \tag{A.37}$$

For the asymptotic distribution of $\hat{\gamma}_{MLE}^{CS,M}$, we notice that the transformed system (A.31) is stationary under assumption that $|\gamma| < 1$. Thus, as $(N, T) \to \infty$, the denominator of (A.37) converges to

$$\begin{aligned}\frac{1}{NT}\sum_{t=1}^{T}E\left[\mathbf{y}_{t-1}'\mathbf{G}\left(\boldsymbol{\psi}\right)'\Omega_{CS}{}^{-1}\mathbf{G}\left(\boldsymbol{\psi}\right)\mathbf{y}_{t-1}\right] &= \frac{1}{NT}\sum_{t=1}^{T}E\left[\check{\mathbf{y}}_{t-1}\left(\boldsymbol{\psi}\right)'\Omega_{CS}{}^{-1}\check{\mathbf{y}}_{t-1}\left(\boldsymbol{\psi}\right)\right] \\ &\to \frac{1}{1-\gamma^2}. \qquad \text{(A.38)}\end{aligned}$$

For the numerator, we can observe that

$$\begin{aligned}E\left[\mathbf{y}_{t-1}'\mathbf{G}\left(\boldsymbol{\psi}\right)'\Omega_{CS}{}^{-1}\mathbf{G}\left(\boldsymbol{\psi}\right)\mathbf{u}_t\right] &= tr\left\{E\left[\mathbf{y}_{t-1}'\mathbf{G}\left(\boldsymbol{\psi}\right)'\Omega_{CS}{}^{-1}\mathbf{G}\left(\boldsymbol{\psi}\right)\mathbf{u}_t\right]\right\} \\ &= tr\left\{\mathbf{G}\left(\boldsymbol{\psi}\right)'\Omega_{CS}{}^{-1}\mathbf{G}\left(\boldsymbol{\psi}\right)E\left(\mathbf{u}_t\mathbf{y}_{t-1}'\right)\right\} \\ &= 0, \qquad \text{(A.39)}\end{aligned}$$

by using the fact that $\mathbf{u}_t$ is independent of $\mathbf{y}_{t-1}$. As a result, we can claim that the QDMLE (A.36) is asymptotically unbiased in the sense that

$$E\left[\sqrt{NT}\left(\hat{\gamma}_{MLE}^{CS,M}-\gamma\right)\right]=0.$$

which suggests that the QDMLE (A.36) is asymptotically unbiased, and the asymptotical unbiasedness of QDMLE is independent of the way how (N,T) go to infinity.

Online supplementary material to "**Estimation of Dynamic Panel Data Models with Interactive Effects: Quasi-differencing Over Time or Across Individuals?**"

Cheng Hsiao, Qiankun Zhou

May 2022

This online supplement includes the mathematical proofs of the QDMLE for Model with both Exogenous Regressors and Interactive Effects. In the main text, we consider the QDMLE for dynamic panels without exogenous variables. Here we show that the QDMLE can be easily extended for dynamic panels with exogenous variables. Consider the general model

$$\begin{aligned} y_{it} &= \gamma y_{it-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + \boldsymbol{\lambda}_i'\mathbf{f}_t + u_{it} \\ &= \mathbf{z}_{it}'\boldsymbol{\delta} + \boldsymbol{\lambda}_i'\mathbf{f}_t + u_{it}, \end{aligned} \tag{OA.1}$$

where $\mathbf{z}_{it} = \left(y_{i,t-1}, \mathbf{x}_{it}'\right)'$, $\boldsymbol{\delta} = \left(\gamma, \boldsymbol{\beta}'\right)'$ and $\mathbf{x}_{it}$ is a $k \times 1$ vector of strictly exogenous variables with respect to u_{it}.

For the general model (OA.1), it can be stacked in vector form as

$$\mathbf{y}_i = \mathbf{Z}_i\boldsymbol{\delta} + \mathbf{F}\boldsymbol{\lambda}_i + \mathbf{u}_i, \quad i = 1, \ldots, N, \tag{OA.2}$$

where $\mathbf{Z}_i = \left(\mathbf{z}_{i1}, \ldots, \mathbf{z}_{iT}\right)'$, $\mathbf{y}_i, \mathbf{F}$, $\boldsymbol{\lambda}_i$ and $\mathbf{u}_i$ are defined in (A.14). Using the same normalization (A.17), if we multiply $\mathbf{H}(\boldsymbol{\vartheta})'$ in both sides of (OA.2), then we have

$$\mathbf{H}(\boldsymbol{\vartheta})'\mathbf{y}_i = \mathbf{H}(\boldsymbol{\vartheta})'\mathbf{Z}_i\boldsymbol{\delta} + \mathbf{H}(\boldsymbol{\vartheta})'\mathbf{u}_i, \quad i = 1, \ldots, N, \tag{OA.3}$$

then it is obvious that the $\mathbf{H}(\boldsymbol{\vartheta})'\mathbf{y}_i$ and $\mathbf{H}(\boldsymbol{\vartheta})'\mathbf{u}_i$ have the similar structure as in (A.18), and

$$\mathbf{H}(\boldsymbol{\vartheta})'\mathbf{Z}_i = \begin{pmatrix} \mathbf{h}_1(\boldsymbol{\vartheta}_1)'\mathbf{Z}_i \\ \mathbf{h}_2(\boldsymbol{\vartheta}_2)'\mathbf{Z}_i \\ \vdots \\ \mathbf{h}_{T-r}(\boldsymbol{\vartheta}_{T-r})'\mathbf{Z}_i \end{pmatrix} = \mathbf{Z}_{i,\underline{r+1,T}} + \Xi\mathbf{Z}_{i,\underline{1,r}},$$

with $\mathbf{Z}_{i,\underline{1,r}} = \left(\mathbf{z}_{i1}, \ldots, \mathbf{z}_{ir}\right)'$ and $\mathbf{Z}_{i,\underline{r+1,T}} = \left(\mathbf{z}_{i,r+1}, \ldots, \mathbf{z}_{iT}\right)'$.

For model (OA.3), the variance-covariance matrix is given by (A.19), and the quasi-log-likelihood function of (OA.3) is given by

$$\log L(\boldsymbol{\delta}, \boldsymbol{\vartheta}) = -\frac{N}{2}\log\left|\Omega_{TS}\right| - \frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{y}_i - \mathbf{Z}_i\boldsymbol{\delta}\right)'\mathbf{H}(\boldsymbol{\vartheta})'\Omega_{TS}{}^{-1}\mathbf{H}(\boldsymbol{\vartheta})\left(\mathbf{y}_i - \mathbf{Z}_i\boldsymbol{\delta}\right). \tag{OA.4}$$

Conditional on $\boldsymbol{\vartheta}$, the FOC for $\boldsymbol{\delta}$ is

$$\frac{\partial \log L(\boldsymbol{\delta}, \boldsymbol{\vartheta})}{\partial \boldsymbol{\delta}} = \sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{H}(\boldsymbol{\vartheta})'\Omega_{TS}{}^{-1}\mathbf{H}(\boldsymbol{\vartheta})\left(\mathbf{y}_i - \mathbf{Z}_i\boldsymbol{\delta}\right) = 0, \tag{OA.5}$$

which yields

$$\hat{\boldsymbol{\delta}}_{MLE}^{TS} = \left(\sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{H}(\boldsymbol{\vartheta})' \Omega_{TS}{}^{-1} \mathbf{H}(\boldsymbol{\vartheta}) \mathbf{Z}_i \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{H}(\boldsymbol{\vartheta})' \Omega_{TS}{}^{-1} \mathbf{H}(\boldsymbol{\vartheta}) \mathbf{y}_i \right), \tag{OA.6}$$

where ^{TS} refers to using normalization (A.17) to remove the multiple interactive effects.

For this QDMLE (OA.6), we have

$$\sqrt{NT}\left(\hat{\boldsymbol{\delta}}_{MLE}^{TS} - \boldsymbol{\delta}\right) = \left(\frac{1}{NT} \sum_{i=1}^{N} \mathbf{Z}_i' \mathbf{H}(\boldsymbol{\vartheta})' \Omega_{TS}{}^{-1} \mathbf{H}(\boldsymbol{\vartheta}) \mathbf{Z}_i \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \left(\mathbf{Z}_i' \mathbf{H}(\boldsymbol{\vartheta})' \Omega_{TS}{}^{-1} \mathbf{H}(\boldsymbol{\vartheta}) \mathbf{u}_i \right), \tag{OA.7}$$

where the denominator can be shown to converge to a nonsingular matrix. For the numerator, we can observe that

$$\mathbf{Z}_i' \mathbf{H}(\boldsymbol{\vartheta})' \Omega_{TS}{}^{-1} \mathbf{H}(\boldsymbol{\vartheta}) \mathbf{u}_i = \begin{pmatrix} \mathbf{y}_{i,-1}' \mathbf{H}(\boldsymbol{\vartheta})' \Omega_{TS}{}^{-1} \mathbf{H}(\boldsymbol{\vartheta}) \mathbf{u}_i \\ \mathbf{X}_i' \mathbf{H}(\boldsymbol{\vartheta})' \Omega_{TS}{}^{-1} \mathbf{H}(\boldsymbol{\vartheta}) \mathbf{u}_i \end{pmatrix},$$

and $E\left[\mathbf{u}_i \mathbf{y}_{i,-1}'\right] = \sigma_u^2 \dot{\Psi}$ with $\dot{\Psi}$ is given by (A.24), then

$$E\left[\mathbf{y}_{i,-1}' \mathbf{H}(\boldsymbol{\vartheta})' \Omega_{TS}{}^{-1} \mathbf{H}(\boldsymbol{\vartheta}) \mathbf{u}_i\right] = O(1),$$

by following the above derivation. It can also be show that

$$E\left(\mathbf{X}_i' \mathbf{H}(\boldsymbol{\vartheta})' \Omega_{TS}{}^{-1} \mathbf{H}(\boldsymbol{\vartheta}) \mathbf{u}_i\right) = E \begin{pmatrix} \mathbf{x}_{1i}' \mathbf{H}(\boldsymbol{\vartheta})' \Omega_{TS}{}^{-1} \mathbf{H}(\boldsymbol{\vartheta}) \mathbf{u}_i \\ \vdots \\ \mathbf{x}_{ki}' \mathbf{H}(\boldsymbol{\vartheta})' \Omega_{TS}{}^{-1} \mathbf{H}(\boldsymbol{\vartheta}) \mathbf{u}_i \end{pmatrix} = 0,$$

under the strictly exogenous assumption of $\mathbf{x}_{it}$, where $\mathbf{x}_{ji} = (x_{j,i1}, x_{j,i2}, \ldots, x_{j,iT})$ for $j = 1, \ldots, k$. Consequently, we have

$$E\left(\mathbf{Z}_i' \mathbf{H}(\boldsymbol{\vartheta})' \Omega_{TS}{}^{-1} \mathbf{H}(\boldsymbol{\vartheta}) \mathbf{u}_i\right) = \begin{pmatrix} O(1) \\ \mathbf{0} \end{pmatrix},$$

which in turn yields

$$\sqrt{NT}\left(\hat{\boldsymbol{\delta}}_{MLE}^{TS} - \boldsymbol{\delta}\right) = O_p\left(\sqrt{\frac{N}{T}}\right),$$

i.e., the QDMLE of $\boldsymbol{\delta}$ is asymptotically biased of order $\sqrt{\frac{N}{T}}$, which is the same order as in the model without exogenous variables.

Alternatively, (OA.1) can be stacked in vector form as

$$\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\delta} + \Lambda \mathbf{f}_t + \mathbf{u}_t, \quad t = 1, \ldots, T, \tag{OA.8}$$

where $\mathbf{Z}_t = (\mathbf{z}_{1t}, \dots, \mathbf{z}_{Nt})'$, $\mathbf{y}_t$, $\mathbf{u}_t$, and Λ are defined in (A.27).

Using the same normalization (A.29), if we multiply $\mathbf{G}(\boldsymbol{\psi})'$ in both sides of (OA.8), then we have

$$\mathbf{G}(\boldsymbol{\psi})' \mathbf{y}_t = \mathbf{G}(\boldsymbol{\psi})' \mathbf{Z}_t \boldsymbol{\delta} + \mathbf{G}(\boldsymbol{\psi})' \mathbf{u}_t, \quad t = 1, \dots, T, \tag{OA.9}$$

where $\mathbf{G}(\boldsymbol{\psi})' \mathbf{y}_t$ and $\mathbf{G}(\boldsymbol{\psi})' \mathbf{u}_t$ are defined in (A.30), and

$$\mathbf{G}(\boldsymbol{\psi})' \mathbf{Z}_t = \begin{pmatrix} \mathbf{g}_1(\boldsymbol{\psi}_1)' \mathbf{Z}_t \\ \mathbf{g}_2(\boldsymbol{\psi}_2)' \mathbf{Z}_t \\ \vdots \\ \mathbf{g}_{N-r}(\boldsymbol{\psi}_{N-r})' \mathbf{Z}_t \end{pmatrix} = \mathbf{Z}_{\underline{r+1,N},t} + \Psi \mathbf{Z}_{\underline{1,r},t},$$

where $\mathbf{Z}_{\underline{1,r},t} = (\mathbf{z}_{1t}, \dots, \mathbf{z}_{rt})'$ and $\mathbf{Z}_{\underline{r+1,N},t} = (\mathbf{z}_{r+1,t}, \dots, \mathbf{z}_{Nt})'$.

For the transformed model (OA.9), the variance-covariance matrix is given by (A.32), and the quasi-log-likelihood function of (OA.9) is given by

$$\log L(\gamma, \boldsymbol{\psi}) = -\frac{N}{2} \log |\Omega_{CS}| - \frac{1}{2} \sum_{t=1}^{T} (\mathbf{y}_t - \mathbf{Z}_t \boldsymbol{\delta})' \mathbf{G}(\boldsymbol{\psi})' \Omega_{CS}^{-1} \mathbf{G}(\boldsymbol{\psi}) (\mathbf{y}_t - \mathbf{Z}_t \boldsymbol{\delta}). \tag{OA.10}$$

Conditional on $\boldsymbol{\psi}$, the FOC for γ is

$$\frac{\partial \log L(\boldsymbol{\delta}, \boldsymbol{\psi})}{\partial \boldsymbol{\delta}} = \sum_{t=1}^{T} \mathbf{Z}_t' \mathbf{G}(\boldsymbol{\psi})' \Omega_{CS}^{-1} \mathbf{G}(\boldsymbol{\psi}) (\mathbf{y}_t - \mathbf{Z}_t \boldsymbol{\delta}) = 0, \tag{OA.11}$$

which yields

$$\hat{\boldsymbol{\delta}}_{MLE}^{CS} = \left(\sum_{t=1}^{T} \mathbf{Z}_t' \mathbf{G}(\boldsymbol{\psi})' \Omega_{CS}^{-1} \mathbf{G}(\boldsymbol{\psi}) \mathbf{Z}_t \right)^{-1} \sum_{t=1}^{T} \left(\mathbf{Z}_t' \mathbf{G}(\boldsymbol{\psi})' \Omega_{CS}^{-1} \mathbf{G}(\boldsymbol{\psi}) \mathbf{y}_t \right), \tag{OA.12}$$

where ^{CS} refers to using normalization (A.29) to remove the multiple interactive effects.

For this QDMLE (OA.12), we have

$$\sqrt{NT} \left(\hat{\boldsymbol{\delta}}_{MLE}^{CS} - \boldsymbol{\delta} \right) = \left(\frac{1}{NT} \sum_{t=1}^{T} \mathbf{Z}_t' \mathbf{G}(\boldsymbol{\psi})' \Omega_{CS}^{-1} \mathbf{G}(\boldsymbol{\psi}) \mathbf{Z}_t \right)^{-1} \left(\frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{t=1}^{T} \mathbf{Z}_t' \mathbf{G}(\boldsymbol{\psi})' \Omega_{CS}^{-1} \mathbf{G}(\boldsymbol{\psi}) \mathbf{u}_t \right). \tag{OA.13}$$

For the numerator of (OA.13), we can observe that

$$\mathbf{Z}_t' \mathbf{G}(\boldsymbol{\psi})' \Omega_{CS}(\boldsymbol{\psi})^{-1} \mathbf{G}(\boldsymbol{\psi}) \mathbf{u}_t = \begin{pmatrix} \mathbf{y}_{t-1}' \mathbf{H}(\boldsymbol{\vartheta})' \Omega_{TS}^{-1} \mathbf{H}(\boldsymbol{\vartheta}) \mathbf{u}_t \\ \mathbf{X}_t' \mathbf{H}(\boldsymbol{\vartheta})' \Omega_{TS}^{-1} \mathbf{H}(\boldsymbol{\vartheta}) \mathbf{u}_t \end{pmatrix},$$

and thus

$$\begin{aligned} E\left[\mathbf{Z}_t' \mathbf{G}(\boldsymbol{\psi})' \Omega_{CS}^{-1} \mathbf{G}(\boldsymbol{\psi}) \mathbf{u}_t \right] &= E \begin{pmatrix} \mathbf{y}_{t-1}' \mathbf{H}(\boldsymbol{\vartheta})' \Omega_{CS}^{-1} \mathbf{H}(\boldsymbol{\vartheta}) \mathbf{u}_t \\ \mathbf{X}_t' \mathbf{H}(\boldsymbol{\vartheta})' \Omega_{CS}^{-1} \mathbf{H}(\boldsymbol{\vartheta}) \mathbf{u}_t \end{pmatrix} \\ &= 0, \end{aligned} \tag{OA.14}$$

the last identity holds since

$$E\left(\mathbf{u}_t \mathbf{y}_{t-1}'\right) = 0 \text{ and } E\left(\mathbf{u}_t \mathbf{X}_t'\right) = 0.$$

by using the fact that $\mathbf{u}_t$ is independent of $\mathbf{y}_{t-1}$ and $\mathbf{X}_t$. As a result, we can claim that the QDMLE (OA.12) is asymptotically unbiased in the sense that

$$E\left[\sqrt{NT}\left(\hat{\boldsymbol{\delta}}_{MLE}^{CS} - \boldsymbol{\delta}\right)\right] = 0.$$

which suggests that the QDMLE (A.36) is asymptotically unbiased, and the asymptotical unbiasedness of QDMLE is independent of the way how (N, T) go to infinity.