Journal of Econometrics 146 (2008) 118–134

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large[☆]

Jihai Yu[a], Robert de Jong[b], Lung-fei Lee[b,*]

[a] *Department of Economics, University of Kentucky, United States*

[b] *Department of Economics, Ohio State University, United States*

ARTICLE INFO

Article history:
Received 24 November 2006
Received in revised form
4 August 2008
Accepted 6 August 2008
Available online 13 August 2008

JEL classification:
C13
C23

Keywords:
Spatial autoregression
Dynamic panels
Fixed effects
Maximum likelihood estimation
Quasi-maximum likelihood estimation
Bias correction

ABSTRACT

This paper investigates the asymptotic properties of quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects, when both the number of individuals n and the number of time periods T are large. We consider the case where T is asymptotically large relative to n, the case where T is asymptotically proportional to n, and the case where n is asymptotically large relative to T. In the case where T is asymptotically large relative to n, the estimators are $\sqrt{nT}$ consistent and asymptotically normal, with the limit distribution centered around 0. When n is asymptotically proportional to T, the estimators are $\sqrt{nT}$ consistent and asymptotically normal, but the limit distribution is not centered around 0; and when n is large relative to T, the estimators are T consistent, and have a degenerate limit distribution. The estimators of the fixed effects are $\sqrt{T}$ consistent and asymptotically normal. We also propose a bias correction for our estimators. We show that when T grows faster than $n^{1/3}$, the correction will asymptotically eliminate the bias and yield a centered confidence interval.

1. Introduction

Spatial econometrics deals with the spatial interactions of economic units in cross-sectional and/or panel data. To capture correlation among cross-sectional units, the spatial autoregressive (SAR) model by Cliff and Ord (1973) has received the most attention in economics. It extends autocorrelation in times series to spatial dimensions, and captures interactions or competition among spatial units. Early development in estimation and testing is summarized in Anselin (1988), Cressie (1993), Kelejian and Robinson (1993), and Anselin and Bera (1998), among others.

Spatial correlation and dynamic settings can be extended to panel data models (Anselin, 1988; Baltagi et al., 2003). Kapoor et al. (2007) provide a rigorous theoretical framework for analysis of spatial panel methods. The model considered for estimation in Kapoor et al. (2007), is a regression panel model with SAR and error components disturbances. Baltagi et al. (2007) consider the testing of spatial and serial dependence in an extended model, where serial correlation on each spatial unit over time, in addition to spatial dependence across spatial units are allowed in the disturbances. These panel models do not incorporate time lagged dependent variables as dynamic structures in the regression equation. By allowing spatial and dynamic features in a regression model, Anselin (2001) distinguishes spatial dynamic models into four categories, namely, "pure space recursive" if only a spatial time lag is included; "time-space recursive" if an individual time lag and a spatial time lag are included; "time-space simultaneous" if an individual time lag and a contemporaneous spatial lag are specified; and "time-space dynamic" if all forms of dependence are included.

In this paper, we shall consider the maximum likelihood (ML) or, more generally, the quasi maximum likelihood (QML) estimation of the spatial dynamic panel data (SDPD) model in the general time-space dynamic category. Because the time-space dynamic category is the general one, our asymptotic analysis and results are applicable to the other three categories as special cases. As a panel model, individual effect (error components) is incorporated in the disturbances. We shall provide a rigorous

☆ We would like to thank participants of the Econometrics seminar at Ohio State University, the International Workshop on Spatial Econometrics and Statistics 2006 (Rome, Italy), the Far Eastern Meeting of Econometric Society 2006 (Beijing, China), and three anonymous referees and the co-editor, Professor Cheng Hsiao, of this journal for helpful comments. Lee acknowledges financial support from NSF Grant No. SES-0519204 and research assistantship support from the Department of Economics at Ohio State University.

* Corresponding address: Department of Economics, Ohio State University, 410 Arps Hall 1945 N High St, 43210 Columbus, OH, United States. Tel.: +1 614 292 5508; fax: +1 614 292 4192.

E-mail addresses: jihai.yu@uky.edu (J. Yu), de-jong.8@osu.edu (R. de Jong), lee.1777@osu.edu (L.-f. Lee).

0304-4076/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2008.08.002

theoretical analysis on the asymptotic properties of the ML estimator (MLE) and the QML estimator (QMLE). The asymptotics will be based on both n, the cross sectional units, and T, the time length, go to infinity, or n being a fixed finite integer, while T goes to infinity. The case with both n and T going to infinity will be the main interest.

As our model includes the dynamic panel data model without spatial dependence as a special case, estimation issues of the dynamic panel data models in the existing econometric literature are relevant. When the time dimension T is fixed, we are likely to encounter the "incidental parameters" problem discussed in Neyman and Scott (1948). This is because the introduction of fixed effects increases the number of parameters to be estimated. In a simple dynamic panel data model with fixed effects, the MLE of the autoregressive coefficient, which is also known as the within group estimator, is biased and inconsistent when n tends to infinity but T is fixed (Nickell, 1981; Hsiao, 1986). To avoid the incidental parameters problem in estimation, alternative estimation methods have been introduced. By taking time differences to eliminate the fixed effects in either the dynamic equation or the construction of instrumental variables (IV), Anderson and Hsiao (1981) show that IV methods can provide consistent estimates. Arellano and Bond (1991) and Arellano and Bover (1995) generalize Anderson and Hsiao (1981) with many more IV moments, by exploring all possible time lag values of the dependent variable in each time period. Blundell and Bond (1998) have considered system estimators, including moments of both levels and first differences in Arellano and Bond (1991) and Arellano and Bover (1995). Bun and Kiviet (2006) derive higher order asymptotic approximation of the finite sample bias for the system estimator under various circumstances, as both N and T are small or moderately large. When T is finite, additional IVs can improve the efficiency of the estimators, even though finite sample biases remain. When both n and T go to infinity, the incidental parameters issue in the MLE becomes less severe as each individual fixed effect can be consistently estimated. However, Hahn and Kuersteiner (2002) and Alvarez and Arellano (2003) have found the existence of asymptotic bias of order $O(1/T)$ in the MLE of the autoregressive parameter when both n and T tend to infinity at a proportional rate. In addition to the MLE, Alvarez and Arellano (2003) also investigate the asymptotic properties of the IV estimators in Arellano and Bond (1991). They have found the presence of asymptotic bias of a similar order to that of the MLE and the IV estimators, due to the presence of many moment conditions. The presence of asymptotic bias is an undesirable feature of these estimates.

Kiviet (1995), Hahn and Kuersteiner (2002), and Bun and Carree (2005) have constructed bias corrected estimators for the dynamic panel data model, by analytically modifying the within estimator. Hahn and Kuersteiner (2002) provide a rigorous asymptotic theory for the within estimator and their bias corrected estimator, when both n and T go to infinity with a same rate. As an alternative to the analytical bias correction, Hahn and Newey (2004) have considered also the Jackknife bias reduction approach.

For the SAR model, Kelejian and Prucha (1998) provide a theoretical foundation for asymptotic analysis for their IV estimator. Lee (2004) analyzes the asymptotic properties of the QMLE. Kapoor et al. (2007) extend their asymptotic analysis of IV and method of moments estimators to a spatial panel model with error components, where T is a fixed finite integer. To the best knowledge of the authors, there is little analytical work done on estimates of spatial dynamic models, when both n and T are large, with the exception of Korniotis (2005). The model considered in Korniotis (2005) is a time-space recursive model in that only individual time lag and spatial time lag are present, but not contemporaneous spatial lag. Fixed effects are included in the model, and this model has an empirical application to US state consumption growth. As a recursive model, the parameters including the fixed effects can be estimated by OLS (within estimator). Korniotis (2005) has also considered a bias adjusted within estimator, which generalizes that in Hahn and Kuersteiner (2002). For the dynamic spatial model considered in this paper, as the contemporaneous spatial lag is presented, the QMLEs of the parameters are nonlinear. Our asymptotic analysis is more complex, but our assumptions are more general. The asymptotics in Hahn and Kuersteiner (2002) is based on the scenario that n and T diverge at a proportional rate. Our asymptotic analysis can cover this scenario and also scenarios that n may go to infinity faster than T, and vice versa. Following the literature on bias correction, we have also considered a bias-adjusted estimator for our QMLE and its asymptotic properties. Monte Carlo experiments are conducted to provide some finite sample properties of the estimators. This paper is theoretic and does not provide an empirical application. But it is interesting to note that the empirical study on interregional trade with a historical panel data on Chinese rice price by Keller and Shiue (2007) allows own time and spatial time lags in addition to a contemporaneous spatial lag in their spatial model.[1]

This paper is organized as follows. In Section 2, we introduce the model, and explain our estimation method, which is a concentrated QML estimation. With the law of large numbers and central limit theorem for our setting developed in the Appendix, Section 3 establishes the consistency and asymptotic distributions of MLE and QMLE. We also propose an analytical bias correction for our estimators. We show that when T grows faster than $n^{1/3}$, this correction will eliminate the bias, and yield a centered confidence interval. Section 4 concludes the paper. Some useful lemmas and proofs are collected in the Appendix.[2]

2. The model and concentrated likelihood function

2.1. The model

The model considered in this paper is

$$Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + V_{nt}, \quad t = 1, 2, \ldots, T, \tag{1}$$

where $Y_{nt} = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ and $V_{nt} = (v_{1t}, v_{2t}, \ldots, v_{nt})'$ are $n \times 1$ column vectors and v_{it} is *i.i.d.*. across i and t with zero mean and variance σ_0^2, W_n is an $n \times n$ spatial weights matrix, which is predetermined and generates the spatial dependence between cross sectional units y_{it}, X_{nt} is an $n \times k_x$ matrix of nonstochastic regressors, and $\mathbf{c}_{n0}$ is $n \times 1$ column vector of fixed effects. Therefore, the total number of parameters in this model is equal to the number of individuals n plus the dimension of the common parameters $(\gamma, \rho, \beta', \lambda, \sigma^2)'$, which is $k_x + 4$.

Define $S_n \equiv S_n(\lambda_0) = I_n - \lambda_0 W_n$. Then, presuming S_n is invertible and denoting $A_n = S_n^{-1}(\gamma_0 I_n + \rho_0 W_n)$, (1) can be rewritten as $Y_{nt} = A_n Y_{n,t-1} + S_n^{-1} X_{nt}\beta_0 + S_n^{-1}\mathbf{c}_{n0} + S_n^{-1} V_{nt}$. Assuming that the infinite sums are well-defined, by continuous substitution,

$$\begin{aligned} Y_{nt} &= \sum_{h=0}^{\infty} A_n^h S_n^{-1}(\mathbf{c}_{n0} + X_{n,t-h}\beta_0 + V_{n,t-h}) \\ &= \mu_n + \mathcal{X}_{nt}\beta_0 + U_{nt}, \end{aligned} \tag{2}$$

where $\mu_n \equiv \sum_{h=0}^{\infty} A_n^h S_n^{-1}\mathbf{c}_{n0}$, $\mathcal{X}_{nt} \equiv \sum_{h=0}^{\infty} A_n^h S_n^{-1} X_{n,t-h}$, and $U_{nt} \equiv \sum_{h=0}^{\infty} A_n^h S_n^{-1} V_{n,t-h}$.

[1] However, error components have not been considered in their empirical models and no theoretic properties of the estimates are investigated in the paper.

[2] Due to space limitation, at the request of the editor and referees, some of the proofs have been condensed and removed. The detailed proofs and intermediate steps in some derivations can be found in the working paper version of this paper. The working paper under the same title is available on the web site: http://economics.sbs.ohio-state.edu/lee/.

2.2. Concentrated likelihood function

Denote $\theta = (\delta', \lambda, \sigma^2)'$ and $\zeta = (\delta', \lambda, \mathbf{c}_n')'$ where $\delta = (\gamma, \rho, \beta')'$. At the true value, $\theta_0 = (\delta_0', \lambda_0, \sigma_0^2)'$ and $\zeta_0 = (\delta_0', \lambda_0, \mathbf{c}_{n0}')'$ where $\delta_0 = (\gamma_0, \rho_0, \beta_0')'$. The likelihood function of (1) is[3]

$$\ln L_{n,T}(\theta, \mathbf{c}_n) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln \sigma^2 + T\ln|S_n(\lambda)| - \frac{1}{2\sigma^2}\sum_{t=1}^{T} V_{nt}'(\zeta)V_{nt}(\zeta), \quad (3)$$

where $V_{nt}(\zeta) = S_n(\lambda)Y_{nt} - \gamma Y_{n,t-1} - \rho W_n Y_{n,t-1} - X_{nt}\beta - \mathbf{c}_n$. Thus, $V_{nt} = V_{nt}(\zeta_0)$.

The QMLEs $\hat{\theta}_{nT}$ and $\hat{\mathbf{c}}_{nT}$ are the extreme estimators derived from the maximization of (3). When the V_{nt}'s are normally distributed, $\hat{\theta}_{nT}$ and $\hat{\mathbf{c}}_{nT}$ are the MLEs; when the V_{nt}'s are not normally distributed, $\hat{\theta}_{nT}$ and $\hat{\mathbf{c}}_{nT}$ are QMLEs. As the number of parameters goes to infinity when n goes to infinity, it is convenient to concentrate $\mathbf{c}_n$ out and focus asymptotic analysis on the estimator of θ_0 via the concentrated likelihood function. For the concentrated likelihood function, the dimension of parameter space does not change as n and/or T increase.

For notational purposes, we define $\tilde{Y}_{nt} = Y_{nt} - \bar{Y}_{nT}$ and $\tilde{Y}_{n,t-1} = Y_{n,t-1} - \bar{Y}_{nT,-1}$ for $t = 1, 2, \ldots, T$ where $\bar{Y}_{nT} = \frac{1}{T}\sum_{t=1}^{T} Y_{nt}$ and $\bar{Y}_{nT,-1} = \frac{1}{T}\sum_{t=1}^{T} Y_{n,t-1}$. Similarly, we define $\tilde{X}_{nt} = X_{nt} - \bar{X}_{nT}$ and $\tilde{V}_{nt} = V_{nt} - \bar{V}_{nT}$.

Denote $Z_{nt} = (Y_{n,t-1}, W_n Y_{n,t-1}, X_{nt})$, then from (3), using the first order condition that $\frac{\partial \ln L_{n,T}(\theta,\mathbf{c}_n)}{\partial \mathbf{c}_n} = \frac{1}{\sigma^2}\sum_{t=1}^{T} V_{nt}(\zeta)$, the concentrated estimators of $\mathbf{c}_{n0}$ given θ are $\hat{\mathbf{c}}_{nT}(\theta) = \frac{1}{T}\sum_{t=1}^{T}(S_n(\lambda)Y_{nt} - Z_{nt}\delta)$ and the concentrated likelihood is

$$\ln L_{n,T}(\theta) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln \sigma^2 + T\ln|S_n(\lambda)| - \frac{1}{2\sigma^2}\sum_{t=1}^{T} \tilde{V}_{nt}'(\zeta)\tilde{V}_{nt}(\zeta), \quad (4)$$

where $\tilde{V}_{nt}(\zeta) = S_n(\lambda)\tilde{Y}_{nt} - \tilde{Z}_{nt}\delta$ and $\tilde{Z}_{nt} = (Y_{n,t-1} - \bar{Y}_{nT,-1}, W_n Y_{n,t-1} - W_n\bar{Y}_{nT,-1}, X_{nt} - \bar{X}_{nT})$. The QMLE $\hat{\theta}_{nT}$ maximizes the concentrated likelihood function (4), and the QMLE of $\mathbf{c}_{n0}$ is $\hat{\mathbf{c}}_{nT}(\hat{\theta}_{nT})$. From (4), the first and second order derivatives of the concentrated likelihood function can be derived; see (36) and Box II in Appendix C. To analyze the asymptotic properties of (36) and Box II evaluated at true parameters, we use the law of large numbers and central limit theorem for double arrays developed in Appendix A (see Lemma 7 through Lemma 13).

3. Quasi maximum likelihood estimators and their asymptotic properties

For our analysis of the asymptotic properties of estimators, we need the following assumptions:

Assumption 1. W_n is a constant spatial weights matrix and its diagonal elements satisfy $w_{n,ii} = 0$ for $i = 1, 2, \ldots, n$.

Assumption 2. The disturbances $\{v_{it}\}$, $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$, are *i.i.d.* across i and t with zero mean, variance σ_0^2 and $E|v_{it}|^{4+\eta} < \infty$ for some $\eta > 0$.

Assumption 3. $S_n(\lambda)$ is invertible for all $\lambda \in \Lambda$. Furthermore, Λ is compact and λ_0 is in the interior of Λ.

Assumption 4. The elements of X_{nt} are nonstochastic and bounded,[4] uniformly in n and t. Also, $\lim_{T\to\infty}\frac{1}{nT}\sum_{t=1}^{T}\tilde{X}_{nt}'\tilde{X}_{nt}$ exists and is nonsingular.

Assumption 5. W_n is uniformly bounded in row and column sums in absolute value (for short, UB).[5] Also $S_n^{-1}(\lambda)$ is UB, uniformly in $\lambda \in \Lambda$.

Assumption 6. $\sum_{h=1}^{\infty}\text{abs}(A_n^h)$ is UB,[6] where $[\text{abs}(A_n)]_{ij} = |A_{n,ij}|$.

Assumption 7. n is a nondecreasing function of T and T goes to infinity.

Assumption 1 is a standard normalization assumption in spatial econometrics, and Assumption 2 provides regularity assumptions for v_{it}. Invertibility of $S_n(\lambda)$ in Assumption 3 guarantees that (2) is valid. Also, compactness is a condition for theoretical analysis. In empirical applications, where W_n is row normalized, one just searches over a parameter space on $(-1, 1)$.[7] When exogenous variables X_{nt} are included in the model, it is convenient to assume that the exogenous regressors are uniformly bounded, as in Assumption 4. Assumption 5 is originated by Kelejian and Prucha (1998, 2001) and is also used in Lee (2004, 2007). In many empirical applications, each of the rows of W_n sums to 1, which ensures that all the weights are between 0 and 1. That W_n and $S_n^{-1}(\lambda)$ are UB is a condition that limits the spatial correlation to a manageable degree. Assumption 6 combines the absolute summability condition and the UB condition of the powers of A_n, which will play an important role to derive asymptotic properties of QMLEs. This assumption is essential for the paper, because it limits the dependence between time series and between cross sectional units. In order to justify the absolute summability of A_n in (2) and Assumption 6, a sufficient condition is $\|A_n\| < 1$ where the matrix norm is the row sum norm or the column sum norm (see Horn and Johnson (1985, Corollary 5.6.16). When $\|A_n\| < 1$, $\sum_{h=0}^{\infty}A_n^h$ exists and can be defined as $(I_n - A_n)^{-1}$. If W_n is row-normalized, it seems natural to consider the parameters of spatial and temporal effects of λ, γ and ρ satisfying the constraint $|\lambda| + |\gamma| + |\rho| < 1$. This constraint has implications on Assumptions 3 and 6. First of all, it implies that $|\lambda| < 1$, and, hence $S_n(\lambda)$ is invertiable. This is so, when W_n is row-normalized, it is usually row-normalized from a symmetric matrix (Ord, 1975). In this situation, W_n is diagonalizable and all the eigenvalues ω_{ni}, $i = 1, \ldots, n$, are real and $|\omega_{ni}| \le 1$. The eigenvalues of $S_n(\lambda)$ are $1 - \lambda\omega_{ni}$, which are all different from 0 for all $\lambda \in (-1, 1)$. This implies Assumption 3 that $S_n(\lambda)$ is invertiable. The constraint $|\gamma_0| + |\rho_0| + |\lambda_0| < 1$ implies also that $\sum_{h=1}^{\infty}A_n^h$ is well-defined. As W_n is diagonalizable, $W_n = R_n D_n^* R_n^{-1}$ where $D_n^* = \text{diag}\{\omega_{n1}, \ldots, \omega_{nn}\}$ and the columns of R_n consist of all the normalized eigenvectors of W_n. Because $A_n = S_n^{-1}(\gamma_0 I_n + \rho_0 W_n)$, it follows that $A_n = R_n D_n R_n^{-1}$

[3] As T is large, we can ignore the influence of the initial condition. When T is fixed, we need to specify the initial condition if MLE is used; and we may also consider the estimation by the generalized method of moments where lagged dependent variables can be used as IVs.

[4] If X_{nt} is allowed to be stochastic and unbounded, appropriate moment conditions can be imposed instead.

[5] We say a (sequence of $n \times n$) matrix P_n is uniformly bounded in row and column sums if $\sup_{n\ge1}\|P_n\|_\infty < \infty$ and $\sup_{n\ge1}\|P_n\|_1 < \infty$, where $\|P_n\|_\infty \equiv \sup_{1\le i,j\le n}\sum_{j=1}^{n}|p_{ij,n}|$ is the row sum norm and $\|P_n\|_1 = \sup_{1\le i,j\le n}\sum_{i=1}^{n}|p_{ij,n}|$ is the column sum norm.

[6] This assumption has effectively ruled out some cases, and, hence, imposed limited dependence across units or time series. For example, if $\lambda_{0n} = 1 - 1/n$ under $n \to \infty$, it is a near unit root case for a cross sectional spatial autoregressive model and S_n^{-1} will not be UB. For spatial dynamic panel model, if $\lambda_0 + \rho_0 + \gamma_0 = 1$, Y_{nt} might have deterministic trends as well as a nonstationary stochastic component (see Yu et al. (2007) for detail).

[7] For the case W_n is not row normalized but its eigenvalues are real, Λ can be a closed interval contained in $(-1/|\omega_{n,\min}|, 1/\omega_{n,\max})$ where $\omega_{n,\min}$ and $\omega_{n,\max}$ are the minimum and maximum eigenvalues of W_n (Anselin, 1988).

where $D_n = \text{diag}\{\frac{\gamma_0+\rho_0\omega_{n1}}{1-\lambda_0\omega_{n1}}, \ldots, \frac{\gamma_0+\rho_0\omega_{nn}}{1-\lambda_0\omega_{nn}}\}$ is the eigenvalue matrix of A_n. When $|\lambda_0| + |\gamma_0| + |\rho_0| < 1$, it is easy to show that $|\frac{\gamma_0+\rho_0\omega_{ni}}{1-\lambda_0\omega_{ni}}| < 1$ for all $i = 1, \ldots, n$. Thus, $\sum_{h=0}^{\infty} A_n^h = \sum_{h=0}^{\infty} R_n D_n^h R_n^{-1} = R_n(I_n - D_n)^{-1} R_n^{-1}$ is a well defined matrix. Assumption 6 imposes stronger convergence of this series in term of absolute values and assumes UB as $n \to \infty$. Assumption 7 allows two cases: (i) $n \to \infty$ as $T \to \infty$; (ii) n is fixed as $T \to \infty$. Because (ii) is similar to a vector autoregressive (VAR) model with restricted coefficients, our main interest is in (i); but our analysis is applicable to both cases. If Assumption 7 holds, then we say that $n, T \to \infty$ simultaneously.

3.1. Consistency of the concentrated estimator $\hat{\theta}_{nT}$

For the concentrated log likelihood function (4) divided by the sample size nT, the corresponding expected value function is $Q_{n,T}(\theta) = E \max_{\mathbf{c}_n} \frac{1}{nT} \ln L_{n,T}(\theta, \mathbf{c}_n)$, which is

$$Q_{n,T}(\theta) = \frac{1}{nT} E \ln L_{n,T}(\theta) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 + \frac{1}{n}\ln|S_n(\lambda)| - \frac{1}{2\sigma^2} E \frac{1}{nT} \sum_{t=1}^{T} \tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta). \quad (5)$$

To show the consistency of $\hat{\theta}_{nT}$, we need the following uniform convergence result.

Claim 1. *Let Θ be any compact parameter space. Then under Assumptions 1–7, $\frac{1}{nT}\ln L_{n,T}(\theta) - Q_{n,T}(\theta) \xrightarrow{p} 0$ uniformly in $\theta \in \Theta$ and $Q_{n,T}(\theta)$ is uniformly equicontinuous for $\theta \in \Theta$.*

For local identification, a sufficient condition (but not necessary) is that the information matrix $\Sigma_{\theta_0,nT}$, where $\Sigma_{\theta_0,nT} = -E\left(\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'}\right)$, is nonsingular and $-E\left(\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}(\theta)}{\partial\theta\partial\theta'}\right)$ has full rank for any θ in some neighborhood $N(\theta_0)$ of θ_0 (see Rothenberg (1971)). Denote $G_n \equiv W_n S_n^{-1}$ and $\mathcal{H}_{nT} = \frac{1}{nT}\sum_{t=1}^{T}(\tilde{Z}_{nt}, G_n\tilde{Z}_{nt}\delta_0)'(\tilde{Z}_{nt}, G_n\tilde{Z}_{nt}\delta_0)$ which is $(k_x+3)\times(k_x+3)$, $\Sigma_{\theta_0,nT}$ is derived in Appendix C as

$$\Sigma_{\theta_0,nT} = \frac{1}{\sigma_0^2}\begin{pmatrix} E\mathcal{H}_{nT} & * \\ \mathbf{0}_{1\times(k_x+3)} & 0 \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{(k_x+2)\times(k_x+2)} & * & * \\ \mathbf{0}_{1\times(k_x+2)} & \frac{1}{n}\left[\text{tr}(G_n'G_n) + \text{tr}(G_n^2)\right] & * \\ \mathbf{0}_{1\times(k_x+2)} & \frac{1}{\sigma_0^2 n}\text{tr}(G_n) & \frac{1}{2\sigma_0^4} \end{pmatrix} + O\left(\frac{1}{T}\right), \quad (6)$$

which is nonsingular for large enough T if $E\mathcal{H}_{nT}$ is nonsingular in the limit or $\frac{1}{n}(\text{tr}\, G_n'G_n + \text{tr}\, G_n^2 - \frac{2(\text{tr}\, G_n)^2}{n})$ is nonzero (see Appendix D for proof). Also, its rank does not change in a small neighborhood of θ_0 (see (41)). When $\lim_{T\to\infty} E\mathcal{H}_{nT}$ is nonsingular, the global identification of the parameters is shown in Theorem 1. When $\lim_{T\to\infty} E\mathcal{H}_{nT}$ is singular, global identification can still be obtained from Theorem 2 via a condition on the variance structure of the model. Denote $\sigma_n^2(\lambda) = \frac{\sigma_0^2}{n}\text{tr}(S_n'^{-1}S_n'(\lambda)S_n(\lambda)S_n^{-1})$.

Theorem 1. *Under Assumptions 1–7, if $\lim_{T\to\infty} E\mathcal{H}_{nT}$ is nonsingular, θ_0 is globally identified and $\hat{\theta}_{nT} \xrightarrow{p} \theta_0$.*

Theorem 2. *Under Assumptions 1–7, θ_0 is globally identified and $\hat{\theta}_{nT} \xrightarrow{p} \theta_0$ if $\lim_{n\to\infty}\left(\frac{1}{n}\ln\left|\sigma_0^2 S_n^{-1\prime}S_n^{-1}\right| - \frac{1}{n}\ln\left|\sigma_n^2(\lambda)S_n^{-1}(\lambda)'S_n^{-1}(\lambda)\right|\right) \neq 0$ for $\lambda \neq \lambda_0$.*[8]

[8] When n is finite, the condition is $\frac{1}{n}\ln|\sigma_0^2 S_n'^{-1}S_n^{-1}| - \frac{1}{n}\ln|\sigma_n^2(\lambda)S_n'^{-1}(\lambda)S_n^{-1}(\lambda)| \neq 0$ for $\lambda \neq \lambda_0$.

3.2. Distribution of QMLEs

The asymptotic distribution of the QMLE $\hat{\theta}_{nT}$ can be derived from the Taylor expansion of $\frac{\partial \ln L_{n,T}(\hat{\theta}_{nT})}{\partial\theta}$ around θ_0. At θ_0, from (35) and (36), the first order derivative of the concentrated likelihood function at θ_0 is in (37) of Appendix C, which involves both linear and quadratic functions of $\tilde{V}_{nt}$. Also, from (2),

$$\tilde{Z}_{nt} = \tilde{Z}_{nt}^* - (\bar{U}_{nT,-1},\ W_n\bar{U}_{nT,-1},\ \mathbf{0}_{n\times k_x}), \quad (7)$$

where $\tilde{Z}_{nt}^* = ((\tilde{\bar{\mathcal{X}}}_{n,t-1} + U_{n,t-1}), (W_n\tilde{\bar{\mathcal{X}}}_{n,t-1} + W_nU_{n,t-1}), \tilde{X}_{nt})$ with $\tilde{\bar{\mathcal{X}}}_{n,t-1} = \mathcal{X}_{n,t-1} - \bar{\mathcal{X}}_{nT,-1}$. Hence, $\tilde{Z}_{nt}$ has two components: one is $\tilde{Z}_{nt}^*$, which is uncorrelated with V_{nt}; the other is $-(\bar{U}_{nT,-1},\ W_n\bar{U}_{nT,-1},\ \mathbf{0}_{n\times k_x})$, which is correlated with V_{nt} when $t \le T-1$.

Hence, from the first order condition in (37) and the decomposition of $\tilde{Z}_{nt}$ in (7), $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta} = \frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^*(\theta_0)}{\partial\theta} - \Delta_{nT}$ where

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^*(\theta_0)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\tilde{Z}_{nt}^{*\prime}V_{nt} \\ \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}(G_n\tilde{Z}_{nt}^*\delta_0)'V_{nt} + \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}(V_{nt}'G_n'V_{nt} - \sigma_0^2\,\text{tr}\, G_n) \\ \frac{1}{2\sigma_0^4}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}(V_{nt}'V_{nt} - n\sigma_0^2) \end{pmatrix}, \quad (8)$$

and

$$\Delta_{nT} = \begin{pmatrix} \frac{1}{\sigma_0^2}\sqrt{\frac{T}{n}}(\bar{U}_{nT,-1},\ W_n\bar{U}_{nT,-1}, \mathbf{0}_{n\times k_x})'\bar{V}_{nT} \\ \frac{1}{\sigma_0^2}\sqrt{\frac{T}{n}}(G_n(\bar{U}_{nT,-1}, W_n\bar{U}_{nT,-1},\ \mathbf{0}_{n\times k_x})\delta_0)'\bar{V}_{nT} + \frac{1}{\sigma_0^2}\sqrt{\frac{T}{n}}\bar{V}_{nT}'G_n'\bar{V}_{nT} \\ \frac{1}{2\sigma_0^4}\sqrt{\frac{T}{n}}\bar{V}_{nT}'\bar{V}_{nT} \end{pmatrix}. \quad (9)$$

This decomposition is useful, as the second component has isolated the source of possible asymptotic bias of $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta}$, due to the estimation of the fixed effects. As is derived in Appendix C, the variance matrix of $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^*(\theta_0)}{\partial\theta}$ is equal to

$$E\left(\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^*(\theta_0)}{\partial\theta}\cdot\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^*(\theta_0)}{\partial\theta'}\right) = \Sigma_{\theta_0,nT} + \Omega_{\theta_0,nT} + O\left(\frac{1}{T}\right), \quad (10)$$

and $\Omega_{\theta_0,nT} = \frac{\mu_4 - 3\sigma_0^4}{\sigma_0^4}\begin{pmatrix} \mathbf{0}_{(k_x+2)\times(k_x+2)} & * & * \\ \mathbf{0}_{1\times(k_x+2)} & \frac{1}{n}\sum_{i=1}^{n}G_{n,ii}^2 & * \\ \mathbf{0}_{1\times(k_x+2)} & \frac{1}{2\sigma_0^2 n}\text{tr}\, G_n & \frac{1}{4\sigma_0^4} \end{pmatrix}$ is a symmetric matrix, with μ_4 being the fourth moment of v_{it}, where $G_{n,ii}$ is the (i,i) entry of G_n. When V_{nt} are normally distributed, $\Omega_{\theta_0,nT} = 0$ because $\mu_4 - 3\sigma_0^4 = 0$ for a normal distribution. Denote $\Sigma_{\theta_0} = \lim_{T\to\infty}\Sigma_{\theta_0,nT}$ and $\Omega_{\theta_0} = \lim_{T\to\infty}\Omega_{\theta_0,nT}$, then,

$$\lim_{T\to\infty} E\left(\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^*(\theta_0)}{\partial\theta}\cdot\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^*(\theta_0)}{\partial\theta'}\right) = \Sigma_{\theta_0} + \Omega_{\theta_0}. \quad (11)$$

The asymptotic distribution of $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^*(\theta_0)}{\partial\theta}$ can be derived from the central limit theorem for martingale difference arrays

$$\varphi_n(\theta) = \begin{pmatrix} \frac{1}{n}\operatorname{tr}\left(\left(\sum_{h=0}^{\infty} A_n^h(\theta)\right) S_n^{-1}(\lambda)\right) \\ \frac{1}{n}\operatorname{tr}\left(W_n \left(\sum_{h=0}^{\infty} A_n^h(\theta)\right) S_n^{-1}(\lambda)\right) \\ \mathbf{0}_{k_x\times 1} \\ \frac{1}{n}\gamma \operatorname{tr}(G_n(\lambda)\left(\sum_{h=0}^{\infty} A_n^h(\theta)\right) S_n^{-1}(\lambda)) + \frac{1}{n}\rho \operatorname{tr}(G_n(\lambda) W_n \left(\sum_{h=0}^{\infty} A_n^h(\theta)\right) S_n^{-1}(\lambda)) + \frac{1}{n}\operatorname{tr} G_n(\lambda) \\ \frac{1}{2\sigma^2} \end{pmatrix}$$

Box I.

(Lemma 13). For the term Δ_{nT}, from Lemma 9 and Lemma 11, $\Delta_{nT} = \sqrt{\frac{n}{T}}\varphi_n + O(\sqrt{\frac{n}{T^3}}) + O_p(\frac{1}{\sqrt{T}})$ where $\varphi_n = \varphi_n(\theta_0)$ is $O(1)$ with the equation in Box I.

When $\gamma_0 = \rho_0 = 0, \varphi_n = ((\operatorname{tr} S_n^{-1})/n, (\operatorname{tr} G_n)/n, \mathbf{0}_{1\times k_x}, (\operatorname{tr} G_n)/n, 1/(2\sigma_0^2))'$. When $\lambda_0 = \rho_0 = 0$, we have $S_n = I_n$, $G_n = W_n$ and $\varphi_n = (1/(1-\gamma_0), (\operatorname{tr} W_n)/n, \mathbf{0}_{1\times k_x}, (\operatorname{tr} W_n)/n, 1/(2\sigma_0^2))'$. If $\lambda_0 = \rho_0 = 0$ is imposed in the estimation so that we estimate $Y_{nt} = \gamma_0 Y_{n,t-1} + \mathbf{c}_{n0} + V_{nt}$, the leading asymptotic bias term will be the same as that of Hahn and Kuersteiner (2002).

Assumption 8. $\lim_{T\to\infty} E\mathcal{H}_{nT}$ is nonsingular or $\lim_{n\to\infty}\frac{1}{n}(\operatorname{tr} G_n'G_n + \operatorname{tr} G_n^2 - \frac{2(\operatorname{tr} G_n)^2}{n}) \neq 0$.

Assumption 8 is a condition for the nonsingularity of the limiting information matrix Σ_{θ_0} in addition to the global identification in Theorems 1 and 2. When $\lim_{T\to\infty} E\mathcal{H}_{nT}$ is singular, as long as we have $\lim_{n\to\infty}\frac{1}{n}(\operatorname{tr} G_n'G_n + \operatorname{tr} G_n^2 - \frac{2(\operatorname{tr} G_n)^2}{n}) \neq 0$, the limiting information matrix Σ_{θ_0} is still nonsingular (see Appendix D).

Claim 2. *Under Assumptions 1–8,* $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta} + \Delta_{nT} \xrightarrow{d} N(0, \Sigma_{\theta_0} + \Omega_{\theta_0})$, *where* $\Delta_{nT} = \sqrt{\frac{n}{T}}\varphi_n + O_p\left(\max\left(\sqrt{\frac{n}{T^3}}, \sqrt{\frac{1}{T}}\right)\right)$ *from (9) with* φ_n *from Box I. When* $\{v_{it}\}$, $i = 1, 2, \ldots, n$ *and* $t = 1, 2, \ldots, T$, *are normal,* $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta} + \Delta_{nT} \xrightarrow{d} N(0, \Sigma_{\theta_0})$.

Also, under Assumptions 1–8, we have $\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}(\theta)}{\partial\theta\partial\theta'} - \frac{1}{nT}\frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'} = \|\theta-\theta_0\|\cdot O_p(1)$ and $\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'} - \frac{\partial^2 Q_{n,T}(\theta_0)}{\partial\theta\partial\theta'} = O_p\left(\frac{1}{\sqrt{nT}}\right)$ (see (38) and (39)). Hence, for the Taylor expansion $\sqrt{nT}(\hat\theta_{nT}-\theta_0) = \left(-\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}(\bar\theta_{nT})}{\partial\theta\partial\theta'}\right)^{-1}\cdot\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta}$, we have $-\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}(\bar\theta_{nT})}{\partial\theta\partial\theta'} = \Sigma_{\theta_0,nT} + O_p\left(\max\left(\sqrt{\frac{1}{nT}}, \frac{1}{T}\right)\right)$ (see Proof for Theorem 3 in Appendix D for details). Combined with Claim 2, we have the following theorem for the distribution of $\hat\theta_{nT}$.

Theorem 3. *Under Assumptions 1–8,*

$$\sqrt{nT}\left(\hat\theta_{nT} - \left(\theta_0 - \frac{\varphi_{\theta_0,nT}}{T}\right)\right) + O_p\left(\max\left(\sqrt{\frac{n}{T^3}}, \sqrt{\frac{1}{T}}\right)\right) \xrightarrow{d} N\left(0, \Sigma_{\theta_0}^{-1}\left(\Sigma_{\theta_0} + \Omega_{\theta_0}\right)\Sigma_{\theta_0}^{-1}\right), \quad (12)$$

where $\varphi_{\theta_0,nT} = \Sigma_{\theta_0,nT}^{-1}\varphi_n$ *is* $O(1)$.
When $\frac{n}{T}\to 0$,

$$\sqrt{nT}(\hat\theta_{nT}-\theta_0) \xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0}+\Omega_{\theta_0})\Sigma_{\theta_0}^{-1}). \quad (13)$$

When $\frac{n}{T}\to k<\infty$,

$$\sqrt{nT}(\hat\theta_{nT}-\theta_0) + \sqrt{k}\varphi_{\theta_0,nT} \xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0}+\Omega_{\theta_0})\Sigma_{\theta_0}^{-1}). \quad (14)$$

When $\frac{n}{T}\to\infty$,

$$T(\hat\theta_{nT}-\theta_0) + \varphi_{\theta_0,nT} \xrightarrow{p} 0. \quad (15)$$

Additionally, if $\{v_{it}\}$, $i = 1, 2, \ldots, n$ *and* $t = 1, 2, \ldots, T$, *are normal, (12) becomes*

$$\sqrt{nT}(\hat\theta_{nT}-\theta_0) + \sqrt{\frac{n}{T}}\varphi_{\theta_0,nT} + O_p\left(\max\left(\sqrt{\frac{n}{T^3}}, \sqrt{\frac{1}{T}}\right)\right) \xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1}). \quad (16)$$

Hence, $\hat\theta_{nT}$ is consistent, but has a bias of the order $O(T^{-1})$. For the distribution of $\hat\theta_{nT}$, when T is relatively large, the QMLEs are $\sqrt{nT}$ consistent and asymptotically properly centered normal; when n is asymptotically proportional to T, the estimators are $\sqrt{nT}$ consistent and asymptotically normal, but the limit distribution does not center around the truth; when n is relatively large, the estimators are T consistent and have a degenerate distribution.

The estimators of fixed effects are $\sqrt{T}$ consistent and asymptotically centered normal, as shown below.

Theorem 4. *Assume that the elements of* $\mathbf{c}_{n0}$ *are bounded. Then under Assumptions 1–8, for* $i = 1, 2, \ldots, n$, $\sqrt{T}\left(\hat c_{i,nT} - c_{i,0}\right) \xrightarrow{d} N(0, \sigma_0^2)$ *and they are asymptotically independent with each other.*

3.3. Bias reduction

From (12), the QMLE $\hat\theta_{nT}$ has the bias $-\frac{1}{T}\varphi_{\theta_0,nT}$ and the confidence interval is not centered when $\frac{n}{T}\to k$ where $0<k<\infty$. Furthermore, when T is small relative to n in the sense that $\frac{n}{T}\to\infty$, the presence of $\varphi_{\theta_0,nT}$ causes $\hat\theta_{nT}$ to have the slower rate T^{-1} of convergence. An analytical bias reduction procedure is to correct the bias $B_{nT} = -\varphi_{\theta_0,nT}$, by constructing an estimator $\hat B_{nT}$ and defining the bias corrected estimator as

$$\hat\theta_{nT}^1 = \hat\theta_{nT} - \frac{\hat B_{nT}}{T}. \quad (17)$$

From Theorem 3, $B_{nT} = -\Sigma_{\theta_0,nT}^{-1}\varphi_n$ where $\varphi_n = \varphi_n(\theta_0)$ from Box I, and we may choose[9]

$$\hat B_{nT} = \left[\left(E\left(\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}(\theta)}{\partial\theta\partial\theta'}\right)\right)^{-1}\varphi_n(\theta)\right]\Bigg|_{\theta=\hat\theta_{nT}}. \quad (18)$$

[9] An asymptotically equivalent alternative way is to replace $\Sigma_{\theta_0,nT}^{-1}$ by the empirical Hessian matrix of the concentrated log likelihood function.

Table 1
Performance of estimators before bias correction

	T	n	θ_0		γ	ρ	β	λ	σ^2
(1)	10	49	θ_0^a	Bias	−0.0628	−0.0031	−0.0077	−0.0024	−0.1168
				SD	0.0322	0.0591	0.0452	0.0477	0.0566
				RMSE	0.0733	0.0807	0.0635	0.0667	0.1352
				CP	0.5020	0.9430	0.9290	0.9300	0.4530
(2)	10	49	θ_0^b	Bias	−0.0701	−0.0080	−0.0111	−0.0105	−0.1193
				SD	0.0322	0.0570	0.0453	0.0457	0.0567
				RMSE	0.0792	0.0779	0.0641	0.0639	0.1372
				CP	0.4050	0.9300	0.9230	0.9370	0.4330
(3)	10	196	θ_0^a	Bias	−0.0625	−0.0036	−0.0076	−0.0024	−0.1105
				SD	0.0161	0.0304	0.0226	0.0246	0.0285
				RMSE	0.0647	0.0417	0.0320	0.0344	0.1146
				CP	0.0310	0.9380	0.9260	0.9260	0.0580
(4)	10	196	θ_0^b	Bias	−0.0691	−0.0067	−0.0109	−0.0091	−0.1129
				SD	0.0160	0.0292	0.0226	0.0236	0.0285
				RMSE	0.0710	0.0405	0.0329	0.0322	0.1169
				CP	0.0130	0.9300	0.9140	0.9320	0.0530
(5)	50	49	θ_0^a	Bias	−0.0121	−0.0018	−0.0008	0.0005	−0.0220
				SD	0.0141	0.0260	0.0202	0.0213	0.0280
				RMSE	0.0221	0.0350	0.0278	0.0288	0.0433
				CP	0.8460	0.9460	0.9370	0.9480	0.8590
(6)	50	49	θ_0^b	Bias	−0.0132	−0.0024	−0.0009	−0.0006	−0.0221
				SD	0.0139	0.0243	0.0203	0.0201	0.0281
				RMSE	0.0224	0.0327	0.0279	0.0269	0.0435
				CP	0.8310	0.9530	0.9340	0.9580	0.8570
(7)	50	196	θ_0^a	Bias	−0.0122	−0.0002	−0.0004	0.0012	−0.0211
				SD	0.0071	0.0134	0.0101	0.0110	0.0140
				RMSE	0.0148	0.0182	0.0139	0.0149	0.0271
				CP	0.5990	0.9410	0.9450	0.9470	0.6530
(8)	50	196	θ_0^b	Bias	−0.0133	−0.0008	−0.0005	0.0004	−0.0212
				SD	0.0070	0.0125	0.0101	0.0103	0.0141
				RMSE	0.0156	0.0171	0.0140	0.0141	0.0273
				CP	0.5040	0.9430	0.9480	0.9480	0.6640

$\theta_0^a = (0.2, 0.2, 1, 0.2, 1)$ and $\theta_0^b = (0.3, 0.3, 1, 0.3, 1)$.

We show that when $T/n^{1/3} \to \infty$, $\hat{\theta}_{nT}^1$ is $\sqrt{nT}$ consistent and asymptotically centered normal, even when $n/T \to \infty$. For the bias corrected estimator, we need the following additional assumption.

Assumption 9. $\sum_{h=0}^{\infty} A_n^h(\theta)$ and $\sum_{h=1}^{\infty} hA_n^{h-1}(\theta)$ are uniformly bounded in either row sum or column sums, uniformly in a neighborhood of θ_0.

Assumption 9 can be justified by Lemma 14. Our result for the bias corrected estimator is in Theorem 5.

Theorem 5. *If $T/n^{1/3} \to \infty$, under Assumptions 1–9, $\sqrt{nT}(\hat{\theta}_{nT}^1 - \theta_0) \xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0} + \Omega_{\theta_0})\Sigma_{\theta_0}^{-1})$.*

Hence, if T grows faster than $n^{1/3}$, the analytical bias correction will give us estimators that are asymptotically normal and centered around θ_0. For the case $\frac{n}{T} \to k$, $\hat{\theta}_{nT}^1$ has removed the asymptotic bias $\varphi_{\theta_0,nT}$. Note that $\frac{n}{T} \to k$ implies $T/n^{1/3} \to \infty$. For the case $\frac{n}{T} \to \infty$, as long as $T/n^{1/3} \to \infty$, $\hat{\theta}_{nT}^1$ is $\sqrt{nT}$ consistent, which is also an improvement upon the T consistency of $\hat{\theta}_{nT}$. Thus, $\hat{\theta}_{nT}^1$ might have better performance in economic applications, especially when n is much larger than T.

3.4. Monte Carlo results

We conduct a small Monte Carlo experiment to evaluate the performance of our MLEs and the bias corrected estimators. We generate samples from (1) and use $\theta_0^a = (0.2, 0.2, 1, 0.2, 1)'$, $\theta_0^b = (0.3, 0.3, 1, 0.3, 1)'$ where $\theta_0 = (\gamma_0, \rho_0, \beta_0', \lambda_0, \sigma_0^2)'$, and X_{nt}, $\mathbf{c}_{n0}$ and V_{nt} are generated from independent normal distributions[10] and the spatial weights matrix we use is a rook matrix. We use $T = 10$ and $T = 50$, and $n = 49$ and $n = 196$. For each set of generated sample observations, we calculate the MLE $\hat{\theta}_{nT}$ and evaluate the bias $\hat{\theta}_{nT} - \theta_0$; we then construct the bias corrected estimator $\hat{\theta}_{nT}^1$ and evaluate the bias $\hat{\theta}_{nT}^1 - \theta_0$. We do this 1000 times to see if the bias is reduced on average by using the analytical bias correction procedure,[11] i.e., to compare $\frac{1}{1000}\sum_{i=1}^{1000}(\hat{\theta}_{nT} - \theta_0)_i$ with $\frac{1}{1000}\sum_{i=1}^{1000}(\hat{\theta}_{nT}^1 - \theta_0)_i$. With two different values of θ_0 for each n and T, finite sample properties of both estimators are summarized in Table 1 and Table 2, where Table 1 is for the performance of the estimators before bias correction and Table 2 is for the performance after the bias correction. For each case, we report the bias (Bias), standard deviation (SD), root mean square error (RMSE) and coverage probability (CP).

We can see that both estimators have some bias, but the bias corrected estimators reduce those biases which are originally larger. This is consistent with our asymptotic analysis, because the bias corrected estimators will eliminate the bias of order $O(T^{-1})$. Also, bias reduction is achieved while there is no significant increase in the variance of the estimators. Before bias correction, the CPs of the estimators under 95% confidence level have lower values due to the bias, especially when n is relatively large. After bias correction, the CPs are close to the specified 95% confidence level.

For different cases of n and T, we can see that for each given n, when T is larger, the biases of two sets of estimators will be smaller and the variance will be smaller; for each given T, when n is larger, the biases of two sets of estimators will be nearly the same, but the variance will be smaller. This is consistent with our theoretical prediction, because the bias is of the order $O(T^{-1})$ and the variance

[10] We generated the spatial panel data with $20 + T$ periods and then take the last T periods as our sample. The initial value is generated as $N(0, I_n)$ in the simulation. We have also generated the data with a much longer history $1000+T$ and the results are similar.

[11] For $n = 196$ and $T = 50$, each iteration takes about 3 s on average using a desktop with 4G memory and duo 2.66 GHz CPU.

Table 2
Performance of estimators after bias correction

	T	n	θ_0		γ	ρ	β	λ	σ^2
(1)	10	49	θ_0^a	Bias	−0.0039	−0.0005	−0.0001	−0.0008	−0.0287
				SD	0.0338	0.0623	0.0474	0.0483	0.0623
				RMSE	0.0467	0.0857	0.0650	0.0671	0.0911
				CP	0.9270	0.9260	0.9320	0.9360	0.8600
(2)	10	49	θ_0^b	Bias	−0.0038	0.0036	0.0004	−0.0039	−0.0322
				SD	0.0337	0.0606	0.0475	0.0459	0.0623
				RMSE	0.0470	0.0855	0.0653	0.0642	0.0921
				CP	0.9130	0.8970	0.9340	0.9220	0.8510
(3)	10	196	θ_0^a	Bias	−0.0040	−0.0011	−0.0000	−0.0009	−0.0217
				SD	0.0169	0.0320	0.0237	0.0249	0.0314
				RMSE	0.0237	0.0441	0.0322	0.0346	0.0484
				CP	0.9120	0.9240	0.9380	0.9270	0.8160
(4)	10	196	θ_0^b	Bias	−0.0035	0.0027	0.0003	−0.0037	−0.0250
				SD	0.0168	0.0310	0.0237	0.0237	0.0314
				RMSE	0.0236	0.0436	0.0322	0.0328	0.0497
				CP	0.9110	0.9020	0.9390	0.9370	0.7950
(5)	50	49	θ_0^a	Bias	−0.0001	−0.0018	−0.0005	0.0005	−0.0025
				SD	0.0143	0.0263	0.0204	0.0213	0.0286
				RMSE	0.0197	0.0355	0.0280	0.0289	0.0393
				CP	0.9400	0.9460	0.9370	0.9460	0.9300
(6)	50	49	θ_0^b	Bias	−0.0002	−0.0019	−0.0004	−0.0002	−0.0026
				SD	0.0141	0.0246	0.0205	0.0201	0.0287
				RMSE	0.0194	0.0332	0.0280	0.0269	0.0395
				CP	0.9410	0.9470	0.9360	0.9570	0.9270
(7)	50	196	θ_0^a	Bias	−0.0002	−0.0001	−0.0001	0.0013	−0.0015
				SD	0.0071	0.0136	0.0102	0.0110	0.0143
				RMSE	0.0097	0.0184	0.0140	0.0149	0.0194
				CP	0.9430	0.9380	0.9440	0.9470	0.9430
(8)	50	196	θ_0^b	Bias	−0.0003	−0.0003	−0.0001	0.0007	−0.0017
				SD	0.0070	0.0127	0.0102	0.0104	0.0144
				RMSE	0.0096	0.0173	0.0140	0.0141	0.0195
				CP	0.9420	0.9410	0.9450	0.9440	0.9440

$\theta_0^a = (0.2, 0.2, 1, 0.2, 1)$ and $\theta_0^b = (0.3, 0.3, 1, 0.3, 1)$.

of the estimators is of the order $O(\frac{1}{nT})$. Also, for different values of θ_0, the biases become larger when θ_0 is larger, and the variances do not change much.

We also run the simulation when V_{nt} is generated from independent exponential distribution with unit variance (demeaned by the population mean). The disturbances are skewed. In order not to produce more tables unnecessarily, the Monte Carlo simulation is conducted only for the parameter vector θ_0^a. From Table 3, we can see that the bias correction can improve the performance of estimators, even for non-gaussian error terms. By comparing the corresponding estimates in Table 3 with those in Tables 1 and 2 under normal disturbances, we see that the biases and SDs are similar except that the SDs for the estimates of σ_0^2 in Table 3 are relatively larger.

Finally, we conduct a simulation to compare the performance of estimators when we use both the SDPD model and VAR model ($n = 9$ and $T = 200$). For the SDPD process without exogenous variable, $Y_{nt} = A_n Y_{n,t-1} + S_n^{-1}\mathbf{c}_{n0} + S_n^{-1} V_{nt}$, which can be considered as a restricted form of the VAR process $Y_{nt} = \Phi_n Y_{n,t-1} + \alpha_{n0} + \epsilon_{nt}$, where Φ_n is $n \times n$ coefficient matrix, ϵ_{nt} is $N(0, \Sigma_\epsilon)$ for each t and is independent over time. When the true data generating process (DGP) is SDPD, we use $(\gamma_0, \rho_0, \lambda_0) = (0.2, 0.2, 0.2)$, W_n is a 9×9 queen matrix, $\mathbf{c}_{n0}$ and V_{nt} are generated from independent normal distributions. When the true DGP is VAR, the 9×9 coefficient matrix Φ_n is designed to have eigenvalues smaller than 1 in absolute value,[12] α_{n0} and ϵ_{nt} are generated from independent normal distributions. Given a DGP, we first use the SDPD model to get the bias corrected estimators $(\hat{\gamma}_{nT}^1, \rho_{nT}^1, \lambda_{nT}^1)$ and get $\hat{A}_n = (I_n - \hat{\lambda}_{nT}^1 W_n)^{-1}(\hat{\gamma}_{nT}^1 I_n + \hat{\rho}_{nT}^1 W_n)$, then, we use the VAR model to get $\hat{\Phi}_n$. We do this 1000 times to compare the Biases, SDs and RMSEs of each element in $\hat{A}_n$ with its corresponding element in $\hat{\Phi}_n$ (there are in total $9 \times 9 = 81$ elements). The results are in Table 4 where the X axis denotes 81 elements of vectorized A_n or Φ_n and Y axis denotes the corresponding values of Biases, SDs and RMSEs. We can see that when the true DGP is SDPD, the restricted SDPD estimators outperform unrestricted VAR estimators, mainly due to the small SDs of the restricted estimates. When the true DGP is VAR, the restricted estimates have larger biases for some parameters, and overall, they have some larger RMSEs than those of the unrestricted VAR estimates.

4. Conclusion

In this paper, we derived the properties of QMLEs of spatial dynamic panel data with fixed effects, and with special attention to the asymptotics when both n and T are large. Estimates of the fixed effects are $\sqrt{T}$ consistent and asymptotically normally distributed. For distribution of the common parameters, where T is asymptotically large relative to n, the estimators are $\sqrt{nT}$ consistent and asymptotically normal, with the limiting distribution centered around 0; when n is asymptotically proportional to T, the estimators are $\sqrt{nT}$ consistent and asymptotically normal, but the limiting distribution is not centered around 0; and when n is large relative to T, the estimators are T consistent, and have a degenerate limiting distribution. We also propose a bias correction for our estimators. We show that when T grows faster than $n^{1/3}$, the correction will eliminate the bias of order $O(T^{-1})$ and yield a centered confidence interval. The contribution of this paper is that it establishes the asymptotic properties of QMLEs and bias-corrected estimators of the spatial dynamic panel model, when both n and T are large.

[12] Each element of the 9×9 coefficient matrix Φ_n is generated from uniform distribution (0, 1). We row normalize the coefficient matrix (so that none of the eigenvalues will be greater than 1 in absolute value) and then multiply it with 0.8 so that all the eigenvalues will be smaller than 1 in absolute value.

Table 3
Performance of estimators under non-normality

	T	n		γ	ρ	β	λ	σ^2
Before bias correction								
(1)	10	49	Bias	−0.0606	−0.0065	−0.0064	−0.0032	−0.1162
			SD	0.0321	0.0590	0.0451	0.0477	0.1035
			RMSE	0.0715	0.0798	0.0637	0.0663	0.1850
			CP	0.5230	0.9590	0.9260	0.9330	0.6800
(2)	10	196	Bias	−0.0621	−0.0019	−0.0089	−0.0002	−0.1124
			SD	0.0161	0.0304	0.0226	0.0246	0.0530
			RMSE	0.0644	0.0418	0.0326	0.0349	0.1294
			CP	0.0310	0.9400	0.9160	0.9140	0.4390
(3)	50	49	Bias	−0.0116	−0.0005	−0.0006	−0.0003	−0.0223
			SD	0.0141	0.0260	0.0202	0.0213	0.0561
			RMSE	0.0220	0.0351	0.0276	0.0294	0.0798
			CP	0.8660	0.9410	0.9480	0.9440	0.9010
(4)	50	196	Bias	−0.0121	−0.0009	−0.0001	0.0012	−0.0214
			SD	0.0071	0.0134	0.0101	0.0110	0.0282
			RMSE	0.0148	0.0182	0.0136	0.0151	0.0427
			CP	0.5930	0.9470	0.9540	0.9340	0.8720
After bias correction								
(5)	10	49	Bias	−0.0019	−0.0042	0.0014	−0.0016	−0.0279
			SD	0.0337	0.0622	0.0473	0.0482	0.1018
			RMSE	0.0466	0.0846	0.0651	0.0668	0.1548
			CP	0.9180	0.9400	0.9350	0.9340	0.8410
(6)	10	196	Bias	−0.0035	0.0006	−0.0014	0.0013	−0.0237
			SD	0.0169	0.0321	0.0237	0.0249	0.0522
			RMSE	0.0235	0.0442	0.0324	0.0351	0.0798
			CP	0.9120	0.9240	0.9320	0.9200	0.8570
(7)	50	49	Bias	0.0003	−0.0004	−0.0003	−0.0002	−0.0028
			SD	0.0143	0.0263	0.0204	0.0213	0.0560
			RMSE	0.0196	0.0355	0.0277	0.0295	0.0774
			CP	0.9330	0.9360	0.9510	0.9430	0.9380
(8)	50	196	Bias	−0.0002	−0.0008	0.0002	0.0012	−0.0018
			SD	0.0071	0.0136	0.0102	0.0110	0.0282
			RMSE	0.0098	0.0184	0.0137	0.0152	0.0382
			CP	0.9450	0.9450	0.9550	0.9340	0.9400

We use $\theta_0^a = (0.2, 0.2, 1, 0.2, 1)$.

Our asymptotic analysis in this paper has focused on the spatial dynamic model with fixed effects, but the remaining disturbances are *i.i.d.* across spatial units. We expect that our asymptotic analysis can be easily extended to dynamic panel models with error components, and spatially and serially correlated disturbances. The spatial panel data model in Baltagi et al. (2007) is an example. Their model is a regression panel model with serial correlation and spatial dependence in disturbances: $Y_{nt} = X_{nt}\beta_0 + c_{n0} + \epsilon_{nt}$ where $\epsilon_{nt} = \lambda_0 W_n \epsilon_{nt} + U_{nt}$ and $U_{nt} = \gamma_0 U_{n,t-1} + V_{nt}$. Denote the n-dimensional vector of total disturbances $\eta_{nt} = c_{n0} + \epsilon_{nt}$. The disturbance process implies the structure $\eta_{nt} = \lambda_0 W_n \eta_{nt} + \gamma_0 \eta_{n,t-1} - \gamma_0 \lambda_0 W_n \eta_{n,t-1} + c_{n0}^* + V_{nt}$, where $c_{n0}^* = (1-\gamma_0)(I_n - \lambda_0 W_n)c_{n0}$. The process of η_{nt} is in the form of our dynamic model when c_{n0}^* is treated as fixed effects and with nonlinear constraints on the spatial and dynamic coefficients. Hence, our theory can be easily adopted to cover the estimation of this model for the case with T (and n) goes to infinity.

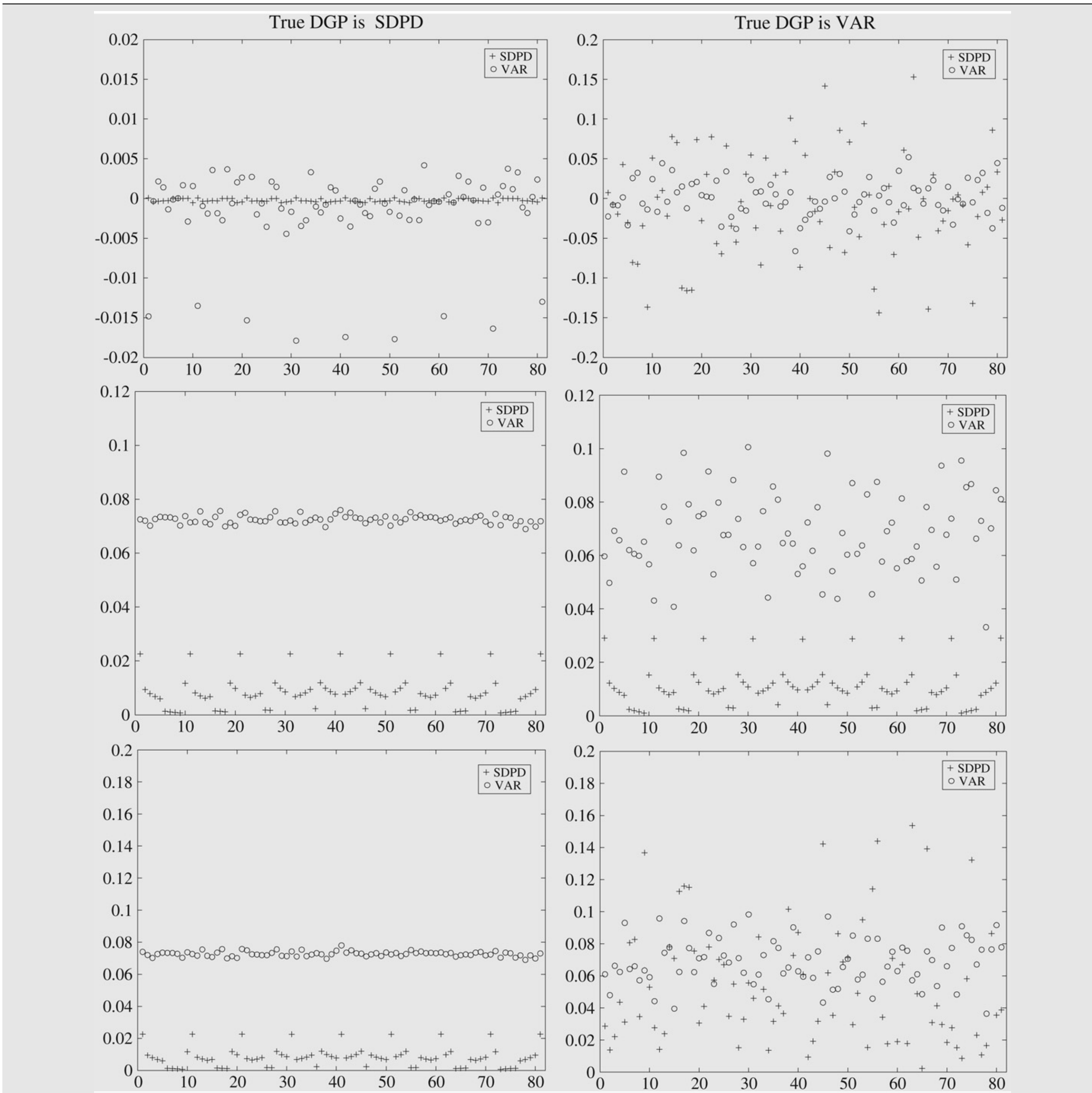
The dynamic panel model analyzed in this paper allows individual-invariant, time-varying exogenous variables in the equation, but it does not incorporate cross-section dependence due to unobserved macroeconomic variables or shocks. Such a cross-section dependence has been considered in some recent panel time series models, e.g., Phillips and Sul (2003) and Pesaran (2006), among others. As an extension of this paper, Lee and Yu (2007) have considered the ML estimation of the SDPD model with both (additive) individual and time fixed effects. By estimating both the individual and time fixed effects, the asymptotic bias problem becomes more severe. With only individual fixed effects, for the case that $\frac{n}{T} \to 0$, as shown in this paper, the QMLE of θ is asymptotically normal centered at 0 (without an asymptotic bias). However, with both individual and time fixed effects, there exists an asymptotic bias of order $O(\frac{1}{n})$. So, contrary to the model with only individual effects, for the model with both individual and time effects, an asymptotic bias of order either $O(\frac{1}{n})$ or $O(\frac{1}{T})$ exists. Lee and Yu (2007) have also constructed a bias-corrected estimator which can remove such biases, but will require conditions that both T/n^3 and n/T^3 go to zero. The model in this paper with only individual effects is of interest in its own as it includes the scenarios of a fixed finite n or $\frac{n}{T} \to 0$.[13] Under such scenarios, the spatial dynamic model can be regarded as a structural vector autoregressive model with restricted coefficients.

For future research, it may be of interest to model common and persistent shocks directly as in Phillips and Moon (1999), Phillips and Sul (2003) and Pesaran (2006) in a random component or factor structural framework with the spatial setting. In addition, as in Korniotis (2007), the model may be extended to accommodate endogenous control variables. With endogenous control variables, a possible estimation method is the generalized method of moments, if proper instrumental variables can be found. The method of maximum likelihood may also be possible if the model is expanded into a simultaneous equation system. These extensions are of interest, as those features can be important in many macroeconomic applications.[14] In addition to the above extension, it may also be of interest to extend the model to incorporate high order contemporaneous spatial lags and spatial time lags. With high order spatial lags, the ML approach is not computationally

[13] Lee and Yu (2007) have found a data transformation approach, which can avoid the additional bias caused by the time effects. However, the transformed approach is valid only for spatial weights matrices with row-normalization.

[14] We appreciate referees for pointing out these important features in empirical macroeconomics models.

Table 4
Biases (1st row), SDs (2nd row) and RMSEs (3rd row) under different DGPs



practical. A practical approach may be based on the generalized method of moments. For the cross section model with high order spatial lags, the generalized method of moments has been considered in Lee and Liu (2007). A possible generalization to the estimation of spatial dynamic panel models remains to be seen.

Appendix A. Some basic lemmas

Let $V_{nt} = (v_{1t}, v_{2t}, \ldots, v_{nt})'$ be $n \times 1$ column vector. We assume that $\{v_{it}\}$, $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$, are *i.i.d.* across i and t with zero mean, variance σ_0^2 and $E\,|v_{it}|^{4+\eta} < \infty$ for some $\eta > 0$.

Denote

$$\mathbb{U}_{nt} = \sum_{h=1}^{\infty} P_{nh} V_{n,t+1-h}, \qquad \mathbb{W}_{nt} = \sum_{h=1}^{\infty} Q_{nh} V_{n,t+1-h}, \tag{19}$$

where $\{P_{nh}\}_{h=1}^{\infty}$ and $\{Q_{nh}\}_{h=1}^{\infty}$ are sequences of $n \times n$ nonstochastic square matrices. Denote $\tilde{\mathbb{U}}_{nt} = \mathbb{U}_{nt} - \bar{\mathbb{U}}_{nT}$ where $\bar{\mathbb{U}}_{nT} = \left(\sum_{t=1}^{T} \mathbb{U}_{nt}\right)/T$, and $\tilde{\tilde{\mathbb{U}}}_{n,t-1} = \mathbb{U}_{n,t-1} - \bar{\mathbb{U}}_{nT,-1}$ where $\bar{\mathbb{U}}_{nT,-1} = \left(\sum_{t=0}^{T-1} \mathbb{U}_{nt}\right)/T$. Also $\tilde{\mathbb{W}}_{nt}$, $\tilde{\tilde{\mathbb{W}}}_{n,t-1}$ and $\tilde{V}_{nt}$ are similarly defined. Below, we state the law of large numbers and central limit theorem useful to derive the asymptotic properties of our estimators. Let D_{nt} be $n \times 1$ vector of uniformly bounded constants for all n and t and let B_{1n} and B_{2n} be $n \times n$ nonstochastic square matrices. We first list the basic assumptions needed for those lemmas.

Assumption A1. The disturbances $\{v_{it}\}$, $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$, are *i.i.d.* across i and t with zero mean, variance σ_0^2 and $E|v_{it}|^{4+\eta} < \infty$ for some $\eta > 0$.

Assumption A2. $P_{nh} = B_{1n}P_n^h$ and $Q_{nh} = B_{2n}Q_n^h$ where P_n^h and Q_n^h are the P_n and Q_n to the power of h. Furthermore, B_{1n}, B_{2n}, $\sum_{h=1}^{\infty}\text{abs}(P_n^h)$ and $\sum_{h=1}^{\infty}\text{abs}(Q_n^h)$ are UB, where $[\text{abs}(P_n)]_{ij} = |P_{n,ij}|$.

Assumption A3. The elements of $n\times 1$ vector D_{nt} are nonstochastic and bounded, uniformly in n and t.

Assumption A4. n is a nondecreasing function of T.

Lemma 1. *With $\mathbb{U}_{nt}$ and $\mathbb{W}_{nt}$ in (19), $\bar{\mathbb{U}}_{nT} = \sum_{h=1}^{\infty}\ddot{P}_{nh}V_{n,T+1-h}$ and $\bar{\mathbb{W}}_{nT} = \sum_{h=1}^{T}\ddot{Q}_{nh}V_{n,T+1-h}$ where*

$$\ddot{P}_{nh} = \begin{cases} \frac{1}{T}(P_{n1} + P_{n2} + \cdots + P_{nh}) = \frac{1}{T}\sum_{g=1}^{h}P_{ng} & \text{for } h \le T \\ \frac{1}{T}\sum_{g=1}^{T}P_{n,h-T+g} & \text{for } h > T, \end{cases} \tag{20}$$

and $\ddot{Q}_{nh}$ has the same pattern. Furthermore, $\sum_{h=1}^{\infty}\ddot{P}_{nh} = \sum_{h=1}^{\infty}P_{nh}$, and $\sum_{h=1}^{\infty}\ddot{Q}_{nh} = \sum_{h=1}^{\infty}Q_{nh}$.

Lemma 2. *Under Assumption A1, for $t \ge s$, $E(\mathbb{U}_{nt}\mathbb{W}_{ns}') = \sigma_0^2\left(\sum_{h=1}^{\infty}P_{n,t-s+h}Q_{nh}'\right)$ and $E(\mathbb{U}_{nt}'\mathbb{W}_{ns}) = \sigma_0^2\text{tr}\left(\sum_{h=1}^{\infty}P_{n,t-s+h}'Q_{nh}\right)$.*

Lemma 3. *Under Assumption A1, $E(V_{nt}'B_{1n}V_{ns})(V_{ng}'B_{2n}V_{nh})$ is equal to $(\mu_4 - 3\sigma_0^4)\sum_{i=1}^{n}B_{1,ii}B_{2,ii} + \sigma_0^4(\text{tr}B_{1n} \times \text{tr}B_{2n} + \text{tr}B_{1n}B_{2n} + \text{tr}B_{1n}B_{2n}')$ for $t = s = g = h$; $\sigma_0^4\text{tr}B_{1n} \times \text{tr}B_{2n}$ for $t = s \ne g = h$; $\sigma_0^4\text{tr}(B_{1n}B_{2n}')$ for $t = g \ne s = h$; $\sigma_0^4\text{tr}(B_{1n}B_{2n})$ for $t = h \ne s = g$; and 0 otherwise.*

Lemma 4. *Under Assumption A1, for $t \ge s$,*

$$Cov(\mathbb{U}_{nt}'\mathbb{W}_{nt}, \mathbb{U}_{ns}'\mathbb{W}_{ns}) = (\mu_4 - 3\sigma_0^4)\sum_{h=1}^{\infty}\sum_{i=1}^{n}(P_{n,t-s+h}'Q_{n,t-s+h})_{ii} \times (P_{nh}'Q_{nh})_{ii} + \sigma_0^4\text{tr}\left[\left(\sum_{h=1}^{\infty}P_{nh}P_{n,t-s+h}'\right)\left(\sum_{h=1}^{\infty}Q_{n,t-s+h}Q_{nh}'\right) + \left(\sum_{h=1}^{\infty}Q_{nh}P_{n,t-s+h}'\right)\left(\sum_{h=1}^{\infty}Q_{n,t-s+h}P_{nh}'\right)\right].$$

Lemma 5. *Suppose B_n, C_{nh} and D_{nh} are $n \times n$ square matrices with all elements being non-negative, and B_n, $\sum_{h=1}^{\infty}C_{nh}$ and $\sum_{h=1}^{\infty}D_{nh}$ are UB. Then, $\sum_{h=1}^{\infty}C_{nh}B_nD_{nh}$ is UB.*

Lemma 6. *Under Assumptions A1, A2 and A4, $Var(\sum_{t=1}^{T}\mathbb{U}_{nt}'\mathbb{W}_{nt}) = O(nT)$.*

Lemma 7. *Under Assumptions A1, A2 and A4,*

$$\frac{1}{nT}\sum_{t=1}^{T}\mathbb{U}_{nt}'\mathbb{W}_{nt} - E\left(\frac{1}{nT}\sum_{t=1}^{T}\mathbb{U}_{nt}'\mathbb{W}_{nt}\right) = O_p\left(\frac{1}{\sqrt{nT}}\right), \tag{21}$$

$$\frac{1}{n}\bar{\mathbb{U}}_{nT}'\bar{\mathbb{W}}_{nT} - E\left(\frac{1}{n}\bar{\mathbb{U}}_{nT}'\bar{\mathbb{W}}_{nT}\right) = O_p\left(\frac{1}{\sqrt{nT^2}}\right), \tag{22}$$

$$\frac{1}{nT}\sum_{t=1}^{T}\tilde{\mathbb{U}}_{nt}'\tilde{\mathbb{W}}_{nt} - E\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{\mathbb{U}}_{nt}'\tilde{\mathbb{W}}_{nt}\right) = O_p\left(\frac{1}{\sqrt{nT}}\right), \tag{23}$$

where $E(\frac{1}{nT}\sum_{t=1}^{T}\mathbb{U}_{nt}'\mathbb{W}_{nt}) = \frac{\sigma_0^2}{n}\text{tr}\left(\sum_{h=1}^{\infty}P_{nh}'Q_{nh}\right) = O(1)$ and $E(\frac{1}{n}\bar{\mathbb{U}}_{nT}'\bar{\mathbb{W}}_{nT}) = \frac{\sigma_0^2}{n}\text{tr}\left(\sum_{h=1}^{\infty}\ddot{P}_{nh}'\ddot{Q}_{nh}\right) = O(\frac{1}{T})$ where $\ddot{P}_{nh}$ and $\ddot{Q}_{nh}$ are defined in (20).

Lemma 8. *Under Assumptions A1–A4, $\frac{1}{nT}\sum_{t=1}^{T}\tilde{D}_{nt}'\tilde{\mathbb{U}}_{nt} = \frac{1}{nT}\sum_{t=1}^{T}\tilde{D}_{nt}'\mathbb{U}_{nt} = O_p\left(\frac{1}{\sqrt{nT}}\right)$, and $\frac{1}{nT}\sum_{t=1}^{T}\bar{D}_{nt}'\bar{\mathbb{U}}_{nt} = O_p\left(\frac{1}{\sqrt{nT}}\right)$.*

Lemma 9. *Under Assumptions A1 and A4, for an $n \times n$ nonstochastic UB matrix B_n,*

$$\frac{1}{nT}\sum_{t=1}^{T}V_{nt}'B_nV_{nt} - E\left(\frac{1}{nT}\sum_{t=1}^{T}V_{nt}'B_nV_{nt}\right) = O_p\left(\frac{1}{\sqrt{nT}}\right), \tag{24}$$

$$\frac{1}{n}\bar{V}_{nT}'B_n\bar{V}_{nT} - E\left(\frac{1}{n}\bar{V}_{nT}'B_n\bar{V}_{nT}\right) = O_p\left(\frac{1}{\sqrt{nT^2}}\right), \tag{25}$$

$$\frac{1}{nT}\sum_{t=1}^{T}\tilde{V}_{nt}'B_n\tilde{V}_{nt} - E\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{V}_{nt}'B_n\tilde{V}_{nt}\right) = O_p\left(\frac{1}{\sqrt{nT}}\right), \tag{26}$$

where $E(\frac{1}{nT}\sum_{t=1}^{T}V_{nt}'B_nV_{nt}) = \frac{1}{n}\sigma_0^2\text{tr}(B_n) = O(1)$ and $E(\frac{1}{n}\bar{V}_{nT}'B_n\bar{V}_{nT}) = \frac{1}{nT}\sigma_0^2\text{tr}(B_n) = O\left(\frac{1}{T}\right)$.

Lemma 10. *Under Assumption A1, $E([(\mathbb{U}_{n,t-1})_i]^4) = (\mu_4 - 3\sigma_0^4)\sum_{h=1}^{\infty}\sum_{j=1}^{n}[(P_{nh})_{ij}]^4 + 3\sigma_0^4[\sum_{h=1}^{\infty}(P_{nh}P_{nh}')_{ii}]^2$.*

Lemma 11. *Under Assumptions A1, A2 and A4, $\sqrt{\frac{T}{n}}(\bar{\mathbb{U}}_{nT,-1}'\bar{V}_{nT} - E(\bar{\mathbb{U}}_{nT,-1}'\bar{V}_{nT})) = O_p\left(\frac{1}{\sqrt{T}}\right)$ where $\sqrt{\frac{T}{n}}E\left(\bar{\mathbb{U}}_{nT,-1}'\bar{V}_{nT}\right) = \sqrt{\frac{n}{T}}\frac{1}{n}\sigma_0^2\text{tr}\left(\sum_{h=1}^{\infty}P_{nh}\right) + O\left(\sqrt{\frac{n}{T^3}}\right)$ when $T \to \infty$.*

Lemma 12. *Let B_n^- denote the lower diagonal matrix constructed from a symmetric B_n by deleting the diagonal and the upper triangle entries. Under Assumptions A1 and A2, if B_n is UB and K_n is an n-dimensional nonstochastic vector with all its elements uniformly bounded, then*

(a) $\frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}(\sum_{j=1}^{i-1}b_{nij}v_{jt})^2 - \frac{1}{2}\sigma_0^2[\text{tr}(B_n^2) - \text{vec}_D'(B_n)\text{vec}_D(B_n)] = \frac{1}{nT}\sum_{t=1}^{T}[V_{nt}'B_n^{-'}B_n^-V_{nt} - \sigma_0^2\text{tr}(B_n^{-'}B_n^-)] = O_p\left(\frac{1}{\sqrt{nT}}\right)$.

(b) $\frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}k_{ni}(\sum_{j=1}^{i-1}b_{nij}v_{jt}) = \frac{1}{nT}\sum_{t=1}^{T}K_n'B_n^-V_{nt} = O_p\left(\frac{1}{\sqrt{nT}}\right)$.

(c) $\frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}u_{n,t-1,i}(\sum_{j=1}^{i-1}b_{nij}v_{jt}) = \frac{1}{nT}\sum_{t=1}^{T}\mathbb{U}_{n,t-1}'B_n^-V_{nt} = O_p\left(\frac{1}{\sqrt{nT}}\right)$.

(d) $\frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}k_{ni}u_{n,t-1,i} = \frac{1}{nT}\sum_{t=1}^{T}K_n'\mathbb{U}_{n,t-1} = O_p\left(\frac{1}{\sqrt{nT}}\right)$

where $\text{vec}_D(B_n)$ is the n-dimensional column vector formed by the the diagonal elements of B_n.

For the central limit theorem that follows, we will consider the following form:

$$Q_{nT} = \sum_{t=1}^{T}\left(\mathbb{U}_{n,t-1}'V_{nt} + D_{nt}'V_{nt} + V_{nt}'B_nV_{nt} - \sigma_0^2\text{tr}\,B_n\right) = \sum_{t=1}^{T}\sum_{i=1}^{n}z_{nt,i},$$

where B_n is an arbitrary $n \times n$ symmetric UB matrix,[15] and $z_{nt,i} = (u_{i,t-1} + d_{nti})v_{it} + b_{n,ii}(v_{it}^2 - \sigma_0^2) + 2(\sum_{j=1}^{i-1}b_{n,ij}v_{jt})v_{it}$. Then, the

[15] The assumption that B_n is symmetric is maintained w.l.o.g. since $V_{nt}'B_nV_{nt} = V_{nt}'[(B_n + B_n')/2]V_{nt}$.

mean and variance of Q_{nT} are $\mu_{Q_{nT}} = 0$ and

$$\sigma^2_{Q_{nT}} = T\sigma_0^4 \mathrm{tr}\left(\sum_{h=1}^{\infty} P'_{nh}P_{nh}\right) + \sigma_0^2 \sum_{t=1}^{T} D'_{nt}D_{nt} + T\left(\left(\mu_4 - 3\sigma_0^4\right)\sum_{i=1}^{n} b^2_{n,ii} + 2\sigma_0^4 \mathrm{tr}(B_n^2)\right) + 2\mu_3 \sum_{t=1}^{T}\sum_{i=1}^{n} d_{nti}b_{n,ii},$$

where $\mu_s = Ev^s_{it}$ for $s = 3, 4$, $b_{n,ii}$'s are diagonal elements of B_n and d_{nti} is the ith element of D_{nt}.

Lemma 13. *Under Assumptions A1–A4 and that B_n is UB, if the sequence $\frac{1}{nT}\sigma^2_{Q_{nT}}$ is bounded away from zero, then, $\frac{Q_{nT}}{\sigma_{Q_{nT}}} \xrightarrow{d} N(0, 1)$.*

Lemma 14. *If $\sup_{n\geq 1}\|A_n(\theta_0)\|_\infty < 1$ (resp: $\sup_{n\geq 1}\|A_n(\theta_0)\|_1 < 1$), then the row sum (resp: column sum) of $\sum_{h=0}^{\infty} A_n^h(\theta)$ and $\sum_{h=1}^{\infty} hA_n^{h-1}(\theta)$ are bounded uniformly in n and in a neighborhood of θ_0.*

Proof for Lemma 13. [16]We are going to use the CLT of the martingale difference array in Gänsler and Stute (1977, p. 365), to prove our CLT (see also Pötscher and Prucha (1997), p. 235). Consider the σ-field

$$\mathcal{F}_{n,t,i} = \sigma(v_{11}, v_{21}, \ldots, v_{n1}, \ldots, v_{1,t-1}, \ldots, v_{n,t-1}, v_{1t}, \ldots, v_{it}), \quad (27)$$

then $E(z_{nt,i}|\mathcal{F}_{n,t,i-1}) = 0$ and $E(z_{nt,i}|\mathcal{F}_{n,t-1,n}) = 0$. As a convention, define $\mathcal{F}_{n,t,0} = \mathcal{F}_{n,t-1,n}$. Thus, $\{z_{nt,i}, \mathcal{F}_{n,t,i}, 1 \leq t \leq T, 1 \leq i \leq n\}$ forms a martingale difference array. To see explicitly that this is a difference array, let $j = n(t-1) + i$ for $1 \leq i \leq n$ and $1 \leq t \leq T$. Thus, j takes integer values from 1 to J where $J = nT$. The σ-field $\mathcal{F}_{n,t,i}$ can be renamed as $\mathcal{F}_{J,j}$ and $z_{Jj} = z_{nt,i}$. As $E(z_{J,j}|\mathcal{F}_{J,j-1}) = 0$ because $E(z_{nt,i}|\mathcal{F}_{n,t,i-1}) = 0$ and $E(z_{nt,i}|\mathcal{F}_{n,t-1,n}) = 0$, $\{z_{nt,i}, \mathcal{F}_{n,t,i}\} = \{z_{J,j}, \mathcal{F}_{J,j-1}\}$ is a martingale difference array. Denote $z^*_{Jj} = z^*_{nt,i} = \frac{z_{nt,i}}{\sigma_{Q_{nT}}}$, where $z_{nt,i} = (u_{i,t-1} + d_{nti})v_{it} + b_{n,ii}(v^2_{it} - \sigma_0^2) + 2(\sum_{j=1}^{i-1} b_{n,ij}v_{jt})v_{it}$, we will apply the martingale CLT to $\sum_{j=1}^{nT} z_{Jj} = \sum_{t=1}^{T}\sum_{i=1}^{n} z_{nt,i}$. The sufficient conditions are (i) $\frac{1}{\sigma^{2+\delta}_{Q_{nT}}}\sum_{t=1}^{T}\sum_{i=1}^{n} E|z_{nt,i}|^{2+\delta} \to 0$ and (ii) $\frac{1}{\sigma^2_{Q_{nT}}}\sum_{t=1}^{T}\sum_{i=1}^{n} E(z^2_{nt,i}|\mathcal{F}_{nt,i-1}) \xrightarrow{p} 1$.

To show (i): For any $p > 0$ and $q > 0$ such that $\frac{1}{p} + \frac{1}{q} = 1$, as

$$|z_{nt,i}| \leq (|u_{n,t-1,i}| + |d_{nt,i}|)|v_{it}| + |b_{nii}|^{\frac{1}{p}}|b_{nii}|^{\frac{1}{q}} \times |v^2_{it} - \sigma_0^2| + 2|v_{it}| \cdot \sum_{j=1}^{i-1} |b_{nij}|^{\frac{1}{p}}|b_{nij}|^{\frac{1}{q}}|v_{jt}|,$$

the Holder inequality implies that

$$|z_{nt,i}| \leq \left[\left(|u_{n,t-1,i}| + |d_{nt,i}|\right)^p + \sum_{j=1}^{i} |b_{nij}|\right]^{\frac{1}{p}} \times \left[|v_{it}|^q + |b_{nii}| \cdot |v^2_{it} - \sigma_0^2|^q + 2^q|v_{it}|^q \cdot \left(\sum_{j=1}^{i-1} |b_{nij}||v_{jt}|^q\right)\right]^{\frac{1}{q}}.$$

Hence,

$$E|z_{nt,i}|^q \leq E\left[(|u_{n,t-1,i}| + |d_{nt,i}|)^p + \sum_{j=1}^{i} |b_{nij}|\right]^{\frac{q}{p}} \times \left[E|v_{it}|^q + |b_{nii}| \cdot E|v^2_{it} - \sigma_0^2|^q + 2^q E|v_{it}|^q \cdot \left(\sum_{j=1}^{i-1} |b_{nij}|E|v_{jt}|^q\right)\right].$$

Because the fourth and more moments of v_{it} exist, by taking $q = 2 + \delta$ for some small δ, there exists a constant $c_1 > 0$ such that $E|z_{nt,i}|^q \leq c_1 E[(|u_{n,t-1,i}| + |d_{nti}|)^p + \sum_{j=1}^{i} |b_{n,ij}|]^{\frac{q}{p}}$. By the c_r-inequality and because B_n is UB, there exist constants $c_2 > 0$, $c_3 > 0$ and $c_4 > 0$ such that

$$\left[(|u_{n,t-1,i}| + |d_{nti}|)^p + \sum_{j=1}^{i} |b_{n,ij}|\right]^{\frac{q}{p}} \leq 2^{\frac{q}{p}-1}\left[(|u_{n,t-1,i}| + |d_{nti}|)^q + c_3\right] \leq 2^{\frac{q}{p}-1}[2^{q-1}(|u_{n,t-1,i}|^q + |d_{nti}|^q) + c_3] \leq c_2|u_{n,t-1,i}|^{2+\delta} + c_4,$$

as $q = 2 + \delta$ implies $\frac{q}{p} = 1 + \delta$. As $E|u_{nt,i}|^4 = O(1)$ uniformly in n, t and i (from Lemma 10), it follows that $E|z_{nt,i}|^{2+\delta} \leq c_1c_2E|u_{n,t-1,i}|^{2+\delta} + c_1c_4 = O(1)$ uniformly. Because $\sigma^{2+\delta}_{QnT} = O[(nT)^{1+\frac{\delta}{2}}]$, one has $\frac{1}{\sigma^{2+\delta}_{Q_{nT}}}\sum_{t=1}^{T}\sum_{i=1}^{n} E|z_{nt,i}|^{2+\delta} = O\left(\frac{1}{(nT)^{\frac{\delta}{2}}}\right)$, which goes to zero. This proves (i).

To show (ii): Because $z_{nt,i} = (u_{n,t-1,i} + d_{nti} + 2\sum_{j=1}^{i-1} b_{nij}v_{jt})v_{it} + b_{nii}(v^2_{it} - \sigma_0^2)$, it implies that

$$E(z^2_{nt,i}|\mathcal{F}_{nt,i-1}) = \sigma_0^2\left(u_{n,t-1,i} + d_{nti} + 2\sum_{j=1}^{i-1} b_{nij}v_{jt}\right)^2 + (\mu_4 - \sigma_0^4)b^2_{nii} + 2\mu_3 b_{nii}\left(u_{n,t-1,i} + d_{nti} + 2\sum_{j=1}^{i-1} b_{nij}v_{jt}\right),$$

as $E(v_{it}(v^2_{it} - \sigma_0^2)) = \mu_3$ and $E(v^2_{it} - \sigma_0^2)^2 = \mu_4 - \sigma_0^4$. Therefore,

$$\sum_{t=1}^{T}\sum_{i=1}^{n} E(z^2_{nt,i}|\mathcal{F}_{nt,i-1}) = \sigma_0^2\sum_{t=1}^{T}\sum_{i=1}^{n}(u_{n,t-1,i} + 2\sum_{j=1}^{i-1} b_{nij}v_{jt})^2 + 2\sum_{t=1}^{T}\sum_{i=1}^{n}[\sigma_0^2 d_{nti} + \mu_3 b_{nii}]\left(u_{n,t-1,i} + 2\sum_{j=1}^{i-1} b_{nij}v_{jt}\right) + (\mu_4 - \sigma_0^4)T\sum_{i=1}^{n} b^2_{nii} + 2\mu_3\sum_{t=1}^{T}\sum_{i=1}^{n} b_{nii}d_{nti} + \sigma_0^2\sum_{t=1}^{T}\sum_{i=1}^{n} d^2_{nti}.$$

This can be compared with $\sigma^2_{Q_{nT}}$, which can be rewritten as

$$\sigma^2_{Q_{nT}} = T\sigma_0^4 \mathrm{tr}\left(\sum_{h=1}^{\infty} P'_{nh}P_{nh}\right) + 2\sigma_0^4 T\left[\mathrm{tr}(B_n^2) - \sum_{i=1}^{n} b^2_{nii}\right] + T(\mu_4 - \sigma_0^4)\sum_{i=1}^{n} b^2_{nii} + 2\mu_3\sum_{t=1}^{T}\sum_{i=1}^{n} d_{nti}b_{nii} + \sigma_0^2\sum_{t=1}^{T} D'_{nt}D_{nt}.$$

From these, we can see that (ii) follows from the results in Lemmas 7 and 12. ■

[16] Proofs for other lemmas of this Appendix and those of Appendix B can be found on the working paper version under the same title via the web site: http://economics.sbs.ohio-state.edu/lee/.

$$\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}(\theta)}{\partial\theta\partial\theta'} = -\frac{1}{nT}\begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^T \tilde{Z}_{nt}'\tilde{Z}_{nt} & \frac{1}{\sigma^2}\sum_{t=1}^T \tilde{Z}_{nt}'W_n\tilde{Y}_{nt} & \frac{1}{\sigma^4}\sum_{t=1}^T \tilde{Z}_{nt}'\tilde{V}_{nt}(\zeta) \\ * & \frac{1}{\sigma^2}\sum_{t=1}^T \left((W_n\tilde{Y}_{nt})'W_n\tilde{Y}_{nt} + \mathrm{tr}(G_n^2(\lambda))\right) & \frac{1}{\sigma^4}\sum_{t=1}^T (W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\zeta) \\ * & * & -\frac{nT}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{t=1}^T \tilde{V}_{nt}'(\zeta)\tilde{V}_{nt}(\zeta) \end{pmatrix}$$

Box II.

Appendix B. Lemmas for some statistics in the model

Denote $Z_{nt} = (Y_{n,t-1}, W_nY_{n,t-1}, X_{nt})$, we provide some lemmas related to $\tilde{Z}_{nt}, \bar{Z}_{nT}$ and $\tilde{V}_{nt}, \bar{V}_{nT}$.

Lemma 15. *Under Assumptions 1–7, for an $n \times n$ nonstochastic UB matrix B_n,*

$$\frac{1}{nT}\sum_{t=1}^T \tilde{Z}_{nt}'B_n\tilde{Z}_{nt} - E\frac{1}{nT}\sum_{t=1}^T \tilde{Z}_{nt}'B_n\tilde{Z}_{nt} = O_p\left(\frac{1}{\sqrt{nT}}\right), \tag{28}$$

$$\frac{1}{nT}\sum_{t=1}^T \tilde{Z}_{nt}'B_n\tilde{V}_{nt} - E\frac{1}{nT}\sum_{t=1}^T \tilde{Z}_{nt}'B_n\tilde{V}_{nt} = O_p\left(\frac{1}{\sqrt{nT}}\right), \tag{29}$$

$$\frac{1}{nT}\sum_{t=1}^T \tilde{V}_{nt}'B_n\tilde{V}_{nt} - E\frac{1}{nT}\sum_{t=1}^T \tilde{V}_{nt}'B_n\tilde{V}_{nt} = O_p\left(\frac{1}{\sqrt{nT}}\right), \tag{30}$$

where $E\frac{1}{nT}\sum_{t=1}^T \tilde{Z}_{nt}'B_n\tilde{Z}_{nt}$ is $O(1)$, $E\frac{1}{nT}\sum_{t=1}^T \tilde{Z}_{nt}'B_n\tilde{V}_{nt}$ is $O\left(\frac{1}{T}\right)$ and $E\frac{1}{nT}\sum_{t=1}^T \tilde{V}_{nt}'B_n\tilde{V}_{nt}$ is $O(1)$.

Lemma 16. *Under Assumptions 1–7, for an $n \times n$ nonstochastic UB matrix B_n,*

$$\frac{1}{n}\bar{Z}_{nT}'B_n\bar{Z}_{nT} - E\frac{1}{n}\bar{Z}_{nT}'B_n\bar{Z}_{nT} = O_p\left(\frac{1}{\sqrt{nT}}\right), \tag{31}$$

$$\frac{1}{n}\bar{Z}_{nT}'B_n\bar{V}_{nT} - E\frac{1}{n}\bar{Z}_{nT}'B_n\bar{V}_{nT} = O_p\left(\frac{1}{\sqrt{nT}}\right), \tag{32}$$

$$\frac{1}{n}\bar{V}_{nT}'B_n\bar{V}_{nT} - E\frac{1}{n}\bar{V}_{nT}'B_n\bar{V}_{nT} = O_p\left(\frac{1}{\sqrt{nT^2}}\right), \tag{33}$$

where $E\frac{1}{n}\bar{Z}_{nT}'B_n\bar{Z}_{nT}$ is $O(1)$, $E\frac{1}{n}\bar{Z}_{nT}'B_n\bar{V}_{nT}$ is $O\left(\frac{1}{T}\right)$ and $E\frac{1}{n}\bar{V}_{nT}'B_n\bar{V}_{nT}$ is $O\left(\frac{1}{T}\right)$.

From (7), $\tilde{Z}_{nt} = \tilde{Z}_{nt}^* - (\bar{U}_{nT,-1}, W_n\bar{U}_{nT,-1}, \mathbf{0})$ where $\tilde{Z}_{nt}^* = ((\tilde{\bar{\mathcal{X}}}_{n,t-1} + U_{n,t-1}), (W_n\tilde{\bar{\mathcal{X}}}_{n,t-1} + W_nU_{n,t-1}), \tilde{X}_{nt})$ with $\tilde{\bar{\mathcal{X}}}_{n,t-1} = \mathcal{X}_{n,t-1} - \bar{\mathcal{X}}_{nT,-1}$. Hence Z_{nt} has two components: one is $\tilde{Z}_{nt}^*$, uncorrelated with V_{nt}; the other is $-(\bar{U}_{nT,-1}\ W_n\bar{U}_{nT,-1}\ \mathbf{0})$, correlated with V_{nt} when $t \le T-1$. Following is a lemma related to $\tilde{Z}_{nt}^*$ and Z_{nt}.

Lemma 17. *Under Assumptions 1–7, for an $n \times n$ nonstochastic UB matrix B_n, $E\frac{1}{nT}\sum_{t=1}^T \tilde{Z}_{nt}'B_n\tilde{Z}_{nt} - E\frac{1}{nT}\sum_{t=1}^T \tilde{Z}_{nt}^{*\prime}B_n\tilde{Z}_{nt}^* = O\left(\frac{1}{T}\right)$ where $E\frac{1}{nT}\sum_{t=1}^T \tilde{Z}_{nt}^{*\prime}B_n\tilde{Z}_{nt}^*$ is $O(1)$.*

Lemma 18. *Under Assumptions 1–7, if the elements of $\mathbf{c}_{n0}$ are uniformly bounded, then the elements of $\frac{1}{T}\sum_{t=1}^T((G_n\mathbf{c}_{n0} + G_nZ_{nt}\delta_0)_i, (Z_{nt})_i)$ are $O_p(1)$ uniformly in n and i, where $(G_n\mathbf{c}_{n0} + G_nZ_{nt}\delta_0)_i$ is the ith element of $(G_n\mathbf{c}_{n0} + G_nZ_{nt}\delta_0)$ and $(Z_{nt})_i$ is the ith row of Z_{nt}.*

Appendix C. Concentrated QMLEs

C.1. Reduced form of (1)

As $Z_{nt} = (Y_{n,t-1}, W_nY_{n,t-1}, X_{nt})$, the reduced form of (1) can be represented as

$$\begin{aligned} Y_{nt} &= S_n^{-1}Z_{nt}\delta_0 + S_n^{-1}(\mathbf{c}_{n0} + V_{nt}) \\ &= Z_{nt}\delta_0 + \lambda_0G_nZ_{nt}\delta_0 + S_n^{-1}(\mathbf{c}_{n0} + V_{nt}), \end{aligned} \tag{34}$$

for $t = 1, 2, \ldots, T$ because $S_n^{-1} = I_n + \lambda_0G_n$. This implies that

$$\tilde{Y}_{nt} = S_n^{-1}\tilde{Z}_{nt}\delta_0 + S_n^{-1}\tilde{V}_{nt} = \tilde{Z}_{nt}\delta_0 + \lambda_0G_n\tilde{Z}_{nt}\delta_0 + S_n^{-1}\tilde{V}_{nt}. \tag{35}$$

C.2. The first and second order conditions

For the concentrated likelihood function (4), the first order derivatives are

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}(\theta)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^T \tilde{Z}_{nt}'\tilde{V}_{nt}(\zeta) \\ \frac{1}{\sigma^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^T \left((W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\zeta) - \mathrm{tr}\,G_n(\lambda)\right) \\ \frac{1}{2\sigma^4}\frac{1}{\sqrt{nT}}\sum_{t=1}^T (\tilde{V}_{nt}'(\zeta)\tilde{V}_{nt}(\zeta) - n\sigma^2) \end{pmatrix}, \tag{36}$$

and the second order derivatives are given in Box II.

Hence,

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^T \tilde{Z}_{nt}'\tilde{V}_{nt} \\ \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^T (G_n\tilde{Z}_{nt}\delta_0)'\tilde{V}_{nt} + \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^T (\tilde{V}_{nt}'G_n'\tilde{V}_{nt} - \sigma_0^2\mathrm{tr}\,G_n) \\ \frac{1}{2\sigma_0^4}\frac{1}{\sqrt{nT}}\sum_{t=1}^T (\tilde{V}_{nt}'\tilde{V}_{nt} - n\sigma_0^2) \end{pmatrix}, \tag{37}$$

and the information matrix is equal to the equation in Box III.

Using Lemma 16, $\Sigma_{\theta_0,nT}^{(2)}$ is $O\left(\frac{1}{T}\right)$. Hence, $\Sigma_{\theta_0,nT} = \Sigma_{\theta_0,nT}^{(1)} + O\left(\frac{1}{T}\right)$.

C.3. The variance of the gradient

From (8), as $\tilde{Z}_{nt}^*$ is uncorrelated with V_{nt}, we have the equation in Box IV.

For the first matrix, it is equal to $\Sigma_{\theta_0,nT} + O\left(\frac{1}{T}\right)$ using Lemma 17. For the second matrix, $E\sum_{t=1}^T \tilde{Z}_{nt}^* = \mathbf{0}_{n\times(k_x+2)}$ and

$$\Sigma_{\theta_0,nT} = -E\left(\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'}\right) = \Sigma^{(1)}_{\theta_0,nT} - \Sigma^{(2)}_{\theta_0,nT} \quad \text{where}$$

$$\Sigma^{(1)}_{\theta_0,nT} = \begin{pmatrix} \frac{1}{\sigma_0^2 nT}E\sum_{t=1}^T \tilde{Z}_{nt}'\tilde{Z}_{nt} & * & * \\ \frac{1}{\sigma_0^2 nT}E\sum_{t=1}^T (G_n\tilde{Z}_{nt}\delta_0)'\tilde{Z}_{nt} & \frac{1}{\sigma_0^2 nT}E\sum_{t=1}^T (G_n\tilde{Z}_{nt}\delta_0)'G_n\tilde{Z}_{nt}\delta_0 + \frac{1}{n}\left[\operatorname{tr}(G_n'G_n)+\operatorname{tr}(G_n^2)\right] & * \\ \mathbf{0}_{1\times(k_x+2)} & \frac{1}{\sigma_0^2 n}\operatorname{tr}(G_n) & \frac{1}{2\sigma_0^4} \end{pmatrix}$$

$$\text{and}\quad \Sigma^{(2)}_{\theta_0,nT} = \begin{pmatrix} \mathbf{0}_{(k_x+2)\times(k_x+2)} & * & * \\ \frac{1}{\sigma_0^2 n}E(G_n\bar{V}_{nT})'\bar{Z}_{nT} & \frac{2}{\sigma_0^2 n}E\left[(G_n\bar{Z}_{nT}\delta_0)'G_n\bar{V}_{nT}\right] + \frac{1}{nT}\operatorname{tr}(G_n'G_n) & * \\ \frac{1}{\sigma_0^4 n}E\left(\bar{Z}_{nT}'\bar{V}_{nT}\right)' & \frac{1}{\sigma_0^4 n}E\left[(G_n\bar{Z}_{nT}\delta_0)'\bar{V}_{nT}\right] + \frac{1}{\sigma_0^2 nT}\operatorname{tr}(G_n) & \frac{1}{T}\frac{1}{\sigma_0^4} \end{pmatrix}$$

Box III.

$$E\left(\frac{1}{\sqrt{nT}}\frac{\partial \ln L^*_{n,T}(\theta_0)}{\partial\theta}\cdot\frac{1}{\sqrt{nT}}\frac{\partial \ln L^*_{n,T}(\theta_0)}{\partial\theta'}\right)$$
$$= \begin{pmatrix} \frac{1}{\sigma_0^2 nT}E\sum_{t=1}^T \tilde{Z}_{nt}^{*\prime}\tilde{Z}^*_{nt} & * & * \\ \frac{1}{\sigma_0^2 nT}E\sum_{t=1}^T (G_n\tilde{Z}^*_{nt}\delta_0)'\tilde{Z}^*_{nt} & \frac{1}{\sigma_0^2 nT}E\sum_{t=1}^T (G_n\tilde{Z}^*_{nt}\delta_0)'G_n\tilde{Z}^*_{nt}\delta_0 + \frac{1}{n}\left[\operatorname{tr}(G_n'G_n)+\operatorname{tr}(G_n^2)\right] & * \\ \mathbf{0}_{1\times(k_x+2)} & \frac{1}{\sigma_0^2 n}\operatorname{tr}(G_n) & \frac{1}{2\sigma_0^4} \end{pmatrix}$$
$$+ \begin{pmatrix} \mathbf{0}_{(k_x+2)\times(k_x+2)} & * & * \\ \frac{\mu_3}{\sigma_0^4 nT}\sum_{i=1}^n G_{n,ii}E\left(\sum_{t=1}^T \tilde{Z}^*_{nt}\right)_i & \frac{2\mu_3}{\sigma_0^4 nT}\sum_{i=1}^n G_{n,ii}E\left(\sum_{t=1}^T G_n\tilde{Z}^*_{nt}\delta_0\right)_i + \frac{\mu_4-3\sigma_0^4}{\sigma_0^4 n}\sum_{i=1}^n G^2_{n,ii} & * \\ \frac{\mu_3}{2\sigma_0^6 nT}l_n'E\sum_{t=1}^T \tilde{Z}^*_{nt} & \frac{1}{2\sigma_0^6 nT}\mu_3 l_n'E\sum_{t=1}^T G_n\tilde{Z}^*_{nt}\delta_0 + \frac{\mu_4-3\sigma_0^4}{2\sigma_0^6 n}\operatorname{tr}G_n & \frac{\mu_4-3\sigma_0^4}{4\sigma_0^8} \end{pmatrix}$$

Box IV.

$E\sum_{t=1}^T G_n\tilde{Z}^*_{nt}\delta_0 = \mathbf{0}_{n\times 1}$. Hence, $E(\frac{1}{\sqrt{nT}}\frac{\partial \ln L^*_{n,T}(\theta_0)}{\partial\theta}\cdot\frac{1}{\sqrt{nT}}\frac{\partial \ln L^*_{n,T}(\theta_0)}{\partial\theta'}) = \Sigma_{\theta_0,nT} + \Omega_{\theta_0,nT} + O\left(\frac{1}{T}\right)$ where

$$\Omega_{\theta_0,nT} = \frac{\mu_4-3\sigma_0^4}{\sigma_0^4}\begin{pmatrix} \mathbf{0}_{(k_x+2)\times(k_x+2)} & * & * \\ \mathbf{0}_{1\times(k_x+2)} & \frac{1}{n}\sum_{i=1}^n G^2_{n,ii} & * \\ \mathbf{0}_{1\times(k_x+2)} & \frac{1}{2\sigma_0^2 n}\operatorname{tr}G_n & \frac{1}{4\sigma_0^4} \end{pmatrix}.$$

When V_{nt} are normally distributed, $\Omega_{\theta_0,nT} = \mathbf{0}_{(k_x+4)\times(k_x+4)}$ because $\mu_4 - 3\sigma_0^4 = 0$ for a normal distribution.

C.4. About $-\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}(\theta)}{\partial\theta\partial\theta'}$, $-\frac{1}{nT}\frac{\partial^2 \ln L_{nT}(\theta)}{\partial\theta\partial\theta'}$, $-\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}(\theta_0)}{\partial\theta\partial\theta'}$ and $-\frac{1}{nT}\frac{\partial^2 \ln L_{nT}(\theta_0)}{\partial\theta\partial\theta'}$

Denote $\|\theta-\theta_0\|$ as the Euclidean norm of $\theta-\theta_0$, and Θ_1 as a neighborhood of θ_0. We have

$$-\frac{1}{nT}\frac{\partial^2 \ln L_{nT}(\theta)}{\partial\theta\partial\theta'} - \left(-\frac{1}{nT}\frac{\partial^2 \ln L_{nT}(\theta_0)}{\partial\theta\partial\theta'}\right) = \|\theta-\theta_0\|\cdot O_p(1), \tag{38}$$

$$\left(-\frac{1}{nT}\frac{\partial^2 \ln L_{nT}(\theta_0)}{\partial\theta\partial\theta'}\right) - \Sigma_{\theta_0,nT} = O_p\left(\frac{1}{\sqrt{nT}}\right), \tag{39}$$

$$\sup_{\theta\in\Theta}\left|-\frac{1}{nT}\frac{\partial^2 \ln L_{nT}(\theta)}{\partial\theta\partial\theta'} - \left(-\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}(\theta)}{\partial\theta\partial\theta'}\right)\right|_{ij} = O_p\left(\frac{1}{\sqrt{nT}}\right), \tag{40}$$

and

$$\sup_{\theta\in\Theta_1}\left|-\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}(\theta)}{\partial\theta\partial\theta'} - \Sigma_{\theta_0,nT}\right|_{ij} = \sup_{\theta\in\Theta_1}\|\theta-\theta_0\|\cdot O(1), \tag{41}$$

for all $i, j = 1, 2, \ldots, k_x+4$.

Proof for (38). The detailed expressions of each entry of the difference $-\frac{1}{nT}\frac{\partial^2 \ln L_{nT}(\theta)}{\partial\theta\partial\theta'} - (-\frac{1}{nT}\frac{\partial^2 \ln L_{nT}(\theta_0)}{\partial\theta\partial\theta'})$ are straightforward from Box II. First, $\frac{1}{n}\operatorname{tr}(G_n^2(\lambda) - G_n^2) = \frac{1}{n}tr[(G_n(\bar{\lambda}))^3](\lambda-\lambda_0)$ where $\bar{\lambda}$ lies between λ and λ_0. As $\frac{1}{n}tr[(G_n(\lambda))^3]$ is UB by Lemma A.7 in Lee (2004), $\frac{1}{n}\operatorname{tr}(G_n^2(\lambda) - G_n^2)$ is of the order $|\lambda-\lambda_0|\cdot O(1)$. Second, as $\tilde{V}_{nt}(\zeta) = \tilde{V}_{nt} - (\lambda-\lambda_0)W_n\tilde{Y}_{nt} - \tilde{Z}_{nt}(\delta-\delta_0)$ and $W_n\tilde{Y}_{nt} = G_n\tilde{Z}_{nt}\delta_0 + G_n\tilde{V}_{nt}$, using Lemma 15, all the entries in the above matrices difference are of the same order as $\|\theta-\theta_0\|$, multiplied by stochastic terms of orders not larger than $O_p(1)$. Hence, $-\frac{1}{nT}\frac{\partial^2 \ln L_{nT}(\theta)}{\partial\theta\partial\theta'} - (-\frac{1}{nT}\frac{\partial^2 \ln L_{nT}(\theta_0)}{\partial\theta\partial\theta'}) = \|\theta-\theta_0\|\cdot O_p(1)$. ■

Proof for (39). As $\Sigma_{\theta_0,nT} = -E\frac{1}{nT}\frac{\partial^2 \ln L_{nT}(\theta_0)}{\partial\theta\partial\theta'}$, all the entries of the difference $(-\frac{1}{nT}\frac{\partial^2 \ln L_{nT}(\theta_0)}{\partial\theta\partial\theta'}) - \Sigma_{\theta_0,nT}$ have zero means. The detailed expressions of the entries are immediate from Box II evaluated at θ_0. Using Lemma 15, all the entries in above difference are of the order $O_p\left(\frac{1}{\sqrt{nT}}\right)$. ■

Proof for (40). Again, all the detailed expressions of entries of the difference $-\frac{1}{nT}\frac{\partial^2 \ln L_{nT}(\theta)}{\partial\theta\partial\theta'} - (-\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}(\theta)}{\partial\theta\partial\theta'})$ follow from Box II. As $\tilde{V}_{nt}(\zeta) = \tilde{V}_{nt} - (\lambda - \lambda_0)W_n\tilde{Y}_{nt} - \tilde{Z}_{nt}(\delta - \delta_0)$ and $W_n\tilde{Y}_{nt} = G_n\tilde{Z}_{nt}\delta_0 + G_n\tilde{V}_{nt}$, by Lemma 15, we have $\sup_{\theta\in\Theta}\left|-\frac{1}{nT}\frac{\partial^2 \ln L_{nT}(\theta)}{\partial\theta\partial\theta'} - \Sigma_{\theta_0,nT}\right|_{ij} = O_p\left(\frac{1}{\sqrt{nT}}\right)$ because Θ is bounded. ■

Proof for (41). The entries of $-\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}(\theta)}{\partial\theta\partial\theta'} - \Sigma_{\theta_0,nT}$ are all differences in expectations, which are of orders no larger than $O(1)$ by Lemma 15; hence, we have $\sup_{\theta\in\Theta_1}\left|-\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}(\theta)}{\partial\theta\partial\theta'} - \Sigma_{\theta,nT}\right|_{ij} = \sup_{\theta\in\Theta_1}\|\theta - \theta_0\| \cdot O(1)$ because Θ_1 is bounded. ■

Appendix D. Proofs for claims and theorems

Proof of Claim 1. To prove $\frac{1}{nT}\ln L_{n,T}(\theta) - Q_{n,T}(\theta) \xrightarrow{p} 0$ uniformly in θ in any compact parameter space Θ:

From $\tilde{V}_{nt}(\zeta) = \tilde{Y}_{nt} - \lambda W_n\tilde{Y}_{nt} - \tilde{Z}_{nt}\delta$, we have $\tilde{V}_{nt}(\zeta) = \tilde{V}_{nt} - (\lambda - \lambda_0)W_n\tilde{Y}_{nt} - \tilde{Z}_{nt}(\delta - \delta_0)$. Hence,

$$\begin{aligned}\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta) &= \tilde{V}'_{nt}\tilde{V}_{nt} + (\lambda - \lambda_0)^2(W_n\tilde{Y}_{nt})'W_n\tilde{Y}_{nt} \\ &\quad + (\delta - \delta_0)'\tilde{Z}'_{nt}\tilde{Z}_{nt}(\delta - \delta_0) \\ &\quad + 2(\lambda - \lambda_0)(W_n\tilde{Y}_{nt})'\tilde{Z}_{nt}(\delta - \delta_0) - 2(\lambda - \lambda_0) \\ &\quad \times (W_n\tilde{Y}_{nt})'\tilde{V}_{nt} - 2(\delta - \delta_0)'\tilde{Z}'_{nt}\tilde{V}_{nt}, \end{aligned} \tag{42}$$

where, using $W_n\tilde{Y}_{nt} = G_n\tilde{Z}_{nt}\delta_0 + G_n\tilde{V}_{nt}$,

$$\begin{aligned}(W_n\tilde{Y}_{nt})'W_n\tilde{Y}_{nt} &= (G_n\tilde{Z}_{nt}\delta_0)'(G_n\tilde{Z}_{nt}\delta_0) \\ &\quad + 2(G_n\tilde{Z}_{nt}\delta_0)'G_n\tilde{V}_{nt} + (G_n\tilde{V}_{nt})'G_n\tilde{V}_{nt}.\end{aligned}$$

Using Lemma 15,

$$\frac{1}{nT}\sum_{t=1}^{T}\tilde{V}'_{nt}\tilde{V}_{nt} - E\frac{1}{nT}\sum_{t=1}^{T}\tilde{V}'_{nt}\tilde{V}_{nt} \xrightarrow{p} 0,$$
$$\frac{1}{nT}\sum_{t=1}^{T}(W_n\tilde{Y}_{nt})'W_n\tilde{Y}_{nt} - E\frac{1}{nT}\sum_{t=1}^{T}(W_n\tilde{Y}_{nt})'W_n\tilde{Y}_{nt} \xrightarrow{p} 0,$$
$$\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}'_{nt}\tilde{Z}_{nt} - E\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}'_{nt}\tilde{Z}_{nt} \xrightarrow{p} 0,$$
$$\frac{1}{nT}\sum_{t=1}^{T}(W_n\tilde{Y}_{nt})'\tilde{V}_{nt} - E\frac{1}{nT}\sum_{t=1}^{T}(W_n\tilde{Y}_{nt})'\tilde{V}_{nt} \xrightarrow{p} 0,$$
$$\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}'_{nt}\tilde{V}_{nt} - E\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}'_{nt}\tilde{V}_{nt} \xrightarrow{p} 0,$$
$$\frac{1}{nT}\sum_{t=1}^{T}(W_n\tilde{Y}_{nt})'\tilde{Z}_{nt} - E\frac{1}{nT}\sum_{t=1}^{T}(W_n\tilde{Y}_{nt})'\tilde{Z}_{nt} \xrightarrow{p} 0.$$

As λ and δ are bounded in Θ, we have $\frac{1}{nT}\sum_{t=1}^{T}\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta) - \frac{1}{nT}E\sum_{t=1}^{T}\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta) \xrightarrow{p} 0$ uniformly in θ in Θ. Also, $\frac{1}{nT}\ln L_{n,T}(\theta) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 + \frac{1}{n}\ln|S_n(\lambda)| - \frac{1}{2\sigma^2 nT}\sum_{t=1}^{T}\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta)$ and $Q_{n,T}(\theta) = E\frac{1}{nT}\ln L_{n,T}(\theta)$. Using the fact that σ^2 is bounded away from zero in Θ,

$$\frac{1}{nT}\ln L_{n,T}(\theta) - Q_{n,T}(\theta) = -\frac{1}{2\sigma^2}\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta) - \frac{1}{nT}E\sum_{t=1}^{T}\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta)\right) \xrightarrow{p} 0 \text{ uniformly in } \theta.$$

To prove $Q_{n,T}(\theta)$ is uniformly equicontinuous in θ in any compact parameter space Θ:

We have $Q_{nT}(\theta) = E\frac{1}{nT}\ln L_{n,T}(\theta) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 + \frac{1}{n}\ln|S_n(\lambda)| - \frac{1}{2\sigma^2 nT}E\sum_{t=1}^{T}\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta)$. As $\tilde{V}_{nt}(\zeta) = S_n(\lambda)S_n^{-1}\tilde{Z}_{nt}\delta_0 - \tilde{Z}_{nt}\delta + S_n(\lambda)S_n^{-1}\tilde{V}_{nt}$,

$$\begin{aligned}E\frac{1}{nT}\sum_{t=1}^{T}\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta) &= \frac{1}{nT}E\sum_{t=1}^{T}(S_n(\lambda)S_n^{-1}\tilde{Z}_{nt}\delta_0 - \tilde{Z}_{nt}\delta)' \\ &\quad \times (S_n(\lambda)S_n^{-1}\tilde{Z}_{nt}\delta_0 - \tilde{Z}_{nt}\delta) + \frac{1}{n}\frac{T-1}{T}\sigma_0^2\operatorname{tr}(S_n^{-1'}S'_n(\lambda)S_n(\lambda)S_n^{-1}) \\ &\quad + \frac{2}{nT}E\sum_{t=1}^{T}(S_n(\lambda)S_n^{-1}\tilde{Z}_{nt}\delta_0 - \tilde{Z}_{nt}\delta)'S_n(\lambda)S_n^{-1}\tilde{V}_{nt}.\end{aligned} \tag{43}$$

The third term $\frac{2}{nT}E\sum_{t=1}^{T}(S_n(\lambda)S_n^{-1}\tilde{Z}_{nt}\delta_0 - \tilde{Z}_{nt}\delta)'S_n(\lambda)S_n^{-1}\tilde{V}_{nt}$ is $O\left(\frac{1}{T}\right)$ according to Lemma 15, and the order $O\left(\frac{1}{T}\right)$ is uniformly in θ in Θ, because it is a polynomial function in θ and Θ is a bounded set. The first term is equal to $(\delta' - \delta'_0, \lambda - \lambda_0)E\mathcal{H}_{nT}(\delta' - \delta'_0, \lambda - \lambda_0)'$ using $S_n(\lambda)S_n^{-1} = I_n - (\lambda - \lambda_0)G_n$; the second term is equal to $\frac{T-1}{T}\sigma_n^2(\lambda)$ where $\sigma_n^2(\lambda) = \frac{\sigma_0^2}{n}\operatorname{tr}(S_n'^{-1}S'_n(\lambda)S_n(\lambda)S_n^{-1})$, which are all polynomial functions of θ. To prove $Q_{n,T}(\theta)$ is uniformly equicontinuous in θ, the following are sufficient: (1) $\ln\sigma^2$ is uniformly continuous; (2) $\frac{1}{n}\ln|S_n(\lambda)|$ is uniformly equicontinuous; (3) $(\delta' - \delta'_0, \lambda - \lambda_0)\mathcal{H}_{nT}(\delta' - \delta'_0, \lambda - \lambda_0)'$ is uniformly equicontinuous; (4) $\sigma_n^2(\lambda)$ is uniformly equicontinuous.

(1) is obvious, because σ^2 is bounded away from zero in Θ. For (2), $\frac{1}{n}\ln|S_n(\lambda_2)| - \frac{1}{n}\ln|S_n(\lambda_1)| = \frac{1}{n}\operatorname{tr}\left(W_nS_n^{-1}\left(\bar{\lambda}\right)\right)(\lambda_2 - \lambda_1)$ where $\bar{\lambda}$ lies between λ_2 and λ_1. As $S_n^{-1}(\lambda)$ is UB, uniformly in $\theta \in \Theta$, $\frac{1}{n}\operatorname{tr}\left(W_nS_n^{-1}\left(\bar{\lambda}\right)\right)$ is bounded, and, hence, $\frac{1}{n}\ln|S(\lambda)|$ is uniformly equicontinuous. For (3), because δ and λ are bounded and because $E\mathcal{H}_{nT}$ is $O(1)$, the result follows. For (4), $\sigma_n^2(\lambda_2) - \sigma_n^2(\lambda_1) = \frac{\sigma_0^2}{n}\operatorname{tr}(S_n'^{-1}S'_n(\lambda_2)S_n(\lambda_2)S_n^{-1}) - \frac{\sigma_0^2}{n}\operatorname{tr}(S_n'^{-1}S'_n(\lambda_1)S_n(\lambda_1)S_n^{-1}) = \sigma_0^2[(\lambda_2 - \lambda_1)(\lambda_2 + \lambda_1 - 2\lambda_0)\frac{\operatorname{tr}G'_nG_n}{n} - (\lambda_2 - \lambda_1)\frac{\operatorname{tr}(G'_n + G_n)}{n}]$ by using $S_n(\lambda)S_n^{-1} = I_n - (\lambda - \lambda_0)G_n$. As G'_nG_n and G_n are UB, $\sigma_n^2(\lambda)$ is uniformly equicontinuous. ■

Proof of nonsingularity of the information matrix. We can prove the result, by using an argument by contradiction (similar to Lee (2004)). For $\Sigma_{\theta_0} \equiv \lim_{T\to\infty}\Sigma_{\theta_0,nT}$, where $\Sigma_{\theta_0,nT}$ is (6), we need to prove that $\Sigma_{\theta_0}\alpha = 0$ implies $\alpha = 0$ where $\alpha = (\alpha'_1, \alpha_2, \alpha_3)'$, α_2, α_3 are scalars and α_1 is $(k_x + 2) \times 1$ vector. If this is true, then, columns of Σ_{θ_0} would be linear independent, and Σ_{θ_0} would be nonsingular. Denote $\mathcal{H}_\delta = \lim_{T\to\infty}\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}'_{nt}\tilde{Z}_{nt}$, $\mathcal{H}_{\delta\lambda} = \lim_{T\to\infty}\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}'_{nt}G_n\tilde{Z}_{nt}\delta_0$, $\mathcal{H}_{\lambda\delta} = \mathcal{H}'_{\delta\lambda}$ and $\mathcal{H}_\lambda = \lim_{T\to\infty}\frac{1}{nT}\sum_{t=1}^{T}(G_n\tilde{Z}_{nt}\delta_0)'G_n\tilde{Z}_{nt}\delta_0$, then

$$\Sigma_{\theta_0} = \frac{1}{\sigma_0^2} \times \begin{pmatrix} E\mathcal{H}_\delta & E\mathcal{H}_{\delta\lambda} & \mathbf{0}_{(k_x+2)\times 1} \\ E\mathcal{H}_{\lambda\delta} & E\mathcal{H}_\lambda + \lim_{n\to\infty}\frac{\sigma_0^2}{n}\left[\operatorname{tr}(G'_nG_n) + \operatorname{tr}(G_n^2)\right] & \lim_{n\to\infty}\frac{1}{n}\operatorname{tr}(G_n) \\ \mathbf{0}_{1\times(k_x+2)} & \lim_{n\to\infty}\frac{1}{n}\operatorname{tr}(G_n) & \frac{1}{2\sigma_0^2}\end{pmatrix}.$$

Hence, $\Sigma_{\theta_0}\alpha = 0$ implies

$$\frac{1}{\sigma_0^2}E\mathcal{H}_\delta \times \alpha_1 + \frac{1}{\sigma_0^2}E\mathcal{H}_{\delta\lambda} \times \alpha_2 = 0,$$

$$\frac{1}{\sigma_0^2}E\mathcal{H}_{\lambda\delta}\times\alpha_1+\left(\frac{1}{\sigma_0^2}E\mathcal{H}_{\lambda}+\lim_{n\to\infty}\frac{1}{n}\left[\mathrm{tr}(G_n'G_n)+\mathrm{tr}(G_n^2)\right]\right)\times\alpha_2 + \lim_{n\to\infty}\frac{1}{\sigma_0^2 n}\mathrm{tr}(G_n)\times\alpha_3=0,$$

$$\lim_{n\to\infty}\frac{1}{\sigma_0^2 n}\mathrm{tr}(G_n)\times\alpha_2+\frac{1}{2\sigma_0^4}\times\alpha_3=0.$$

The first and third equations imply, respectively, $\alpha_1=-(E\mathcal{H}_\delta)^{-1}E\mathcal{H}_{\delta\lambda}\times\alpha_2$ and $\alpha_3=-2\lim_{n\to\infty}\frac{\sigma_0^2}{n}\mathrm{tr}(G_n)\times\alpha_2$. By eliminating α_1 and α_3, the second equation becomes $\{(\frac{1}{\sigma_0^2}\left(E\mathcal{H}_\lambda-E\mathcal{H}_{\lambda\delta}(E\mathcal{H}_\delta)^{-1}E\mathcal{H}_{\delta\lambda}\right))+\lim_{n\to\infty}\frac{1}{n}[\mathrm{tr}(G_n'G_n)+\mathrm{tr}(G_n^2) - 2\frac{\mathrm{tr}^2(G_n)}{n}]\}\times\alpha_2=0$. Because $\mathrm{tr}(G_n'G_n)+\mathrm{tr}(G_n^2)-2\frac{\mathrm{tr}^2(G_n)}{n}=\frac{1}{2}\mathrm{tr}\left[(C_n'+C_n)(C_n'+C_n)'\right]\geq 0$ where $C_n=G_n-\frac{\mathrm{tr}\,G_n}{n}I_n$, combined with the condition that $\lim_{T\to\infty}E\mathcal{H}_{nT}$ is nonsingular, we have $\alpha_2=0$ and hence $\alpha=0$. ■

Proof of Theorem 1. As $E\sum_{t=1}^T\tilde{V}_{nt}'\tilde{V}_{nt}=n(T-1)\sigma_0^2$, at θ_0, (5) implies $E\ln L_{n,T}(\theta_0)=-\frac{nT}{2}\ln 2\pi-\frac{nT}{2}\ln\sigma_0^2+T\ln|S_n|-\frac{n(T-1)}{2}$. Denote $\sigma_n^2(\lambda)=\frac{\sigma_0^2}{n}\mathrm{tr}(S_n^{-1\prime}S_n'(\lambda)S_n(\lambda)S_n^{-1})$. By using $S_n(\lambda)S_n^{-1}=I_n+(\lambda_0-\lambda)G_n$ for (43), it follows that

$$\begin{aligned}&\frac{1}{nT}E\ln L_{n,T}(\theta)-\frac{1}{nT}E\ln L_{n,T}(\theta_0)\\&\quad=-\frac{1}{2}(\ln\sigma^2-\ln\sigma_0^2)+\frac{1}{n}\ln|S_n(\lambda)|-\frac{1}{n}\ln|S_n|\\&\qquad-\left(\frac{1}{2\sigma^2}\frac{1}{nT}\sum_{t=1}^T E\tilde{V}_{nt}'(\zeta)\tilde{V}_{nt}(\zeta)-\frac{T-1}{2T}\right)\\&\quad=T_{1,n}(\lambda,\sigma^2)-\frac{1}{2\sigma^2}T_{2,n,T}(\delta,\lambda)+o(1)\end{aligned}$$

where $T_{1,n}(\lambda,\sigma^2)=-\frac{1}{2}(\ln\sigma^2-\ln\sigma_0^2)+\frac{1}{n}\ln|S_n(\lambda)|-\frac{1}{n}\ln|S_n|-\frac{1}{2\sigma^2}(\sigma_n^2(\lambda)-\sigma^2)$ and

$$T_{2,n,T}(\delta,\lambda)=\frac{1}{nT}\sum_{t=1}^T E\left\{(\tilde{Z}_{nt}(\delta_0-\delta)+(\lambda_0-\lambda)G_n\tilde{Z}_{nt}\delta_0)'\times(\tilde{Z}_{nt}(\delta_0-\delta)+(\lambda_0-\lambda)G_n\tilde{Z}_{nt}\delta_0)\right\}.$$

Consider the process $Y_{nt}=\lambda_0W_nY_{nt}+V_{nt}$ for a period t, the log likelihood function of this process is $\ln L_{p,n}(\lambda,\sigma^2)=-\frac{n}{2}\ln 2\pi-\frac{n}{2}\ln\sigma^2+\ln|S_n(\lambda)|-\frac{1}{2\sigma^2}(S_n(\lambda)Y_{nt})'S_n(\lambda)Y_{nt}$. Let $E_p(\cdot)$ be the expectation operator for Y_{nt} based on this process. It follows that $E_p(\frac{1}{n}\ln L_{p,n}(\lambda,\sigma^2))-E_p(\frac{1}{n}\ln L_{p,n}(\lambda_0,\sigma_0^2))=-\frac{1}{2}(\ln\sigma^2-\ln\sigma_0^2)+\frac{1}{n}\ln|S_n(\lambda)|-\frac{1}{n}\ln|S_n(\lambda_0)|-\frac{1}{2\sigma^2}(\sigma_n^2(\lambda)-\sigma^2)$, which equals $T_{1,n}(\lambda,\sigma^2)$. By the information inequality, $\ln L_{p,n}(\lambda,\sigma^2)-\ln L_{p,n}(\lambda_0,\sigma_0^2)\leq 0$. Thus, $T_{1,n}(\lambda,\sigma^2)\leq 0$ for any (λ,σ^2). Also, $T_{2,n,T}(\delta,\lambda)$ is a quadratic function of δ and λ. Under the condition that $\lim_{T\to\infty}E\mathcal{H}_{nT}$ is nonsingular, $T_{2,n,T}(\delta,\lambda)>0$ whenever $(\delta,\lambda)\neq(\delta_0,\lambda_0)$, so, (δ,λ) is globally identified. Given λ_0, σ_0^2 is the unique maximizer of $T_{1,n}(\lambda_0,\sigma^2)$. Hence, $(\delta,\lambda,\sigma^2)$ is globally identified.

Combined with uniform convergence and equicontinuity in Claim 1, the consistency follows. ■

Proof of Theorem 2. From proof of Theorem 1, $\frac{1}{nT}E\ln L_{n,T}(\theta)-\frac{1}{nT}E\ln L_{n,T}(\theta_0)=T_{1,n}(\lambda,\sigma^2)-\frac{1}{2\sigma^2}T_{2,n,T}(\delta,\lambda)+o(1)$. When $\lim_{T\to\infty}E\mathcal{H}_{nT}$ is singular, δ_0 and λ_0 cannot be identified from $T_{2,n,T}(\delta,\lambda)$. Global identification requires that the limit of $T_{1,n}(\lambda,\sigma^2)$ is strictly less than zero. As $T_{1,n}(\lambda,\sigma^2)\leq 0$ by the information inequality, $T_{1,n}(\lambda,\sigma^2)\neq 0$ is equivalent to $\frac{1}{n}\ln\left|\sigma_0^2S_n^{-1}S_n^{-1\prime}\right|\neq\frac{1}{n}\ln\left|\sigma_n^2(\lambda)S_n^{-1}(\lambda)S_n^{-1\prime}(\lambda)\right|$ (see Lee (2004, Proof of Theorem 4.1)). After λ_0 and σ_0^2 are identified, given λ_0, δ_0 can be identified from $T_{2,n,T}(\delta,\lambda)$. Combined with uniform convergence and equicontinuity in Claim 1, the consistency follows. ■

Proof of Claim 2. From (7), $\tilde{Z}_{nt}=\tilde{Z}_{nt}^*-(\bar{U}_{nT,-1},W_n\bar{U}_{nT,-1},\mathbf{0}_{n\times k_x})$, which has two components: one is $\tilde{Z}_{nt}^*$, uncorrelated with V_{nt}; the other is $-(\bar{U}_{nT,-1},\ W_n\bar{U}_{nT,-1},\ \mathbf{0}_{n\times k_x})$, correlated with V_{nt} when $t\leq T-1$. Correspondingly, $\frac{1}{\sqrt{nT}}\frac{\partial\ln L_{n,T}(\theta_0)}{\partial\theta}=\frac{1}{\sqrt{nT}}\frac{\partial\ln L_{n,T}^*(\theta_0)}{\partial\theta}-\Delta_{nT}$ where $\frac{1}{\sqrt{nT}}\frac{\partial\ln L_{n,T}^*(\theta_0)}{\partial\theta}$ is in (8) and Δ_{nT} is in (9). For $\frac{1}{\sqrt{nT}}\frac{\partial\ln L_{n,T}^*(\theta_0)}{\partial\theta}$, the CLT of martingale difference arrays (Lemma 13) can be applied. For Δ_{nT}, using Lemma 9 and Lemma 11, it is equal to $\sqrt{\frac{n}{T}}\varphi_n+O\left(\sqrt{\frac{n}{T^3}}\right)+O_p\left(\sqrt{\frac{1}{T}}\right)$ where φ_n is $O(1)$ in Box I. ■

Proof of Theorem 3. According to the Taylor expansion, $\sqrt{nT}(\hat{\theta}_{nT}-\theta_0)=\left(-\frac{1}{nT}\frac{\partial^2\ln L_{n,T}(\bar{\theta}_{nT})}{\partial\theta\partial\theta'}\right)^{-1}\cdot\left(\frac{1}{\sqrt{nT}}\frac{\partial\ln L_{n,T}^*(\theta_0)}{\partial\theta}-\Delta_{nT}\right)$ where $\frac{1}{\sqrt{nT}}\frac{\partial\ln L_{n,T}^*(\theta_0)}{\partial\theta}\xrightarrow{d}N(0,\Sigma_{\theta_0}+\Omega_{\theta_0})$, $\Delta_{nT}=\sqrt{\frac{n}{T}}\varphi_n+O\left(\sqrt{\frac{n}{T^3}}\right)+O_p\left(\sqrt{\frac{1}{T}}\right)$ with $\varphi_n=O(1)$ and $\bar{\theta}_{nT}$ lies between θ_0 and $\hat{\theta}_{nT}$. As $-\frac{1}{nT}\frac{\partial^2\ln L_{n,T}(\bar{\theta}_{nT})}{\partial\theta\partial\theta'}=\left(-\frac{1}{nT}\frac{\partial^2\ln L_{n,T}(\bar{\theta}_{nT})}{\partial\theta\partial\theta'}-\left(-\frac{1}{nT}\frac{\partial^2\ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'}\right)\right)+\left(-\frac{1}{nT}\frac{\partial^2\ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'}-\Sigma_{\theta_0,nT}\right)+\Sigma_{\theta_0,nT}$ where the first term is $\left\|\bar{\theta}_{nT}-\theta_0\right\|\cdot O_p(1)$ from (38) and the second term is $O_p\left(\frac{1}{\sqrt{nT}}\right)$ from (39), $-\frac{1}{nT}\frac{\partial^2\ln L_{n,T}(\bar{\theta}_{nT})}{\partial\theta\partial\theta'}=\left\|\bar{\theta}_{nT}-\theta_0\right\|\cdot O_p(1)+O_p\left(\frac{1}{\sqrt{nT}}\right)+\Sigma_{\theta_0,nT}$. Because $\left\|\bar{\theta}_{nT}-\theta_0\right\|=o_p(1)$ and $\Sigma_{\theta_0,nT}$ is nonsingular in the limit, $-\frac{1}{nT}\frac{\partial^2\ln L_{n,T}(\bar{\theta}_{nT})}{\partial\theta\partial\theta'}$ is invertible for large n and T and $\left(-\frac{1}{nT}\frac{\partial^2\ln L_{n,T}(\bar{\theta}_{nT})}{\partial\theta\partial\theta'}\right)^{-1}$ is $O_p(1)$. Then, $\sqrt{nT}(\hat{\theta}_{nT}-\theta_0)=O_p(1)\cdot\left(O_p(1)+O\left(\sqrt{\frac{n}{T}}\right)\right)$, which implies that

$$\hat{\theta}_{nT}-\theta_0=O_p\left(\max\left(\sqrt{\frac{1}{nT}},\frac{1}{T}\right)\right). \tag{44}$$

Hence, $\sqrt{nT}(\hat{\theta}_{nT}-\theta_0)=\left(\Sigma_{\theta_0,nT}+O_p\left(\max\left(\sqrt{\frac{1}{nT}},\frac{1}{T}\right)\right)\right)^{-1}\cdot\left(\frac{1}{\sqrt{nT}}\frac{\partial\ln L_{n,T}^*(\theta_0)}{\partial\theta}-\Delta_{nT}\right)$. Using the fact that[17]

$$\left(\Sigma_{\theta_0,nT}+O_p\left(\max\left(\sqrt{\frac{1}{nT}},\frac{1}{T}\right)\right)\right)^{-1} =\Sigma_{\theta_0,nT}^{-1}+O_p\left(\max\left(\sqrt{\frac{1}{nT}},\frac{1}{T}\right)\right), \tag{45}$$

we have $\sqrt{nT}(\hat{\theta}_{nT}-\theta_0)=\Sigma_{\theta_0,nT}^{-1}\cdot\frac{1}{\sqrt{nT}}\frac{\partial\ln L_{n,T}^*(\theta_0)}{\partial\theta}+O_p\left(\max\left(\sqrt{\frac{1}{nT}},\frac{1}{T}\right)\right)\cdot\frac{1}{\sqrt{nT}}\frac{\partial\ln L_{n,T}^*(\theta_0)}{\partial\theta}-\Sigma_{\theta_0,nT}^{-1}\cdot\Delta_{nT}-O_p\left(\max\left(\sqrt{\frac{1}{nT}},\frac{1}{T}\right)\right)\cdot\Delta_{nT}$, which implies that $\sqrt{nT}(\hat{\theta}_{nT}-\theta_0)+\Sigma_{\theta_0,nT}^{-1}\cdot\Delta_{nT}+O_p\left(\max\left(\sqrt{\frac{1}{nT}},\frac{1}{T}\right)\right)\Delta_{nT}=(\Sigma_{\theta_0,nT}^{-1}+o_p(1))\cdot\frac{1}{\sqrt{nT}}\frac{\partial\ln L_{n,T}^*(\theta_0)}{\partial\theta}$. As $\Sigma_{\theta_0}=\lim_{T\to\infty}\Sigma_{\theta_0,nT}$ exists, then using Claim 2 and $\Delta_{nT}=\sqrt{\frac{n}{T}}\varphi_n+O\left(\sqrt{\frac{n}{T^3}}\right)+O_p\left(\sqrt{\frac{1}{T}}\right)$, $\sqrt{nT}(\hat{\theta}_{nT}-$

[17] For two nonsingular matrices C_k and D_k with $C_k-D_k=O_p(T^{-\eta})$ for $\eta>0$, we have $C_k^{-1}-D_k^{-1}=C_k^{-1}(D_k-C_k)D_k^{-1}=O_p(T^{-\eta})$.

$\theta_0) + \sqrt{\frac{n}{T}}\Sigma_{\theta_0,nT}^{-1}\varphi_n + O_p\left(\max\left(\sqrt{\frac{n}{T^3}}, \sqrt{\frac{1}{T}}\right)\right) \xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0} + \Omega_{\theta_0})\Sigma_{\theta_0}^{-1})$. ■

Proof of Theorem 4. From the first order condition $\frac{\partial \ln L_{n,T}(\theta,\mathbf{c}_n)}{\partial \mathbf{c}_n} = \frac{1}{\sigma^2}\sum_{t=1}^T V_{nt}(\zeta)$, we have $\hat{\mathbf{c}}_{nT}(\theta) = \frac{1}{T}\sum_{t=1}^T(S_n(\lambda)Y_{nt} - Z_{nt}\delta)$. As $S_nY_{nt} = Z_{nt}\delta_0 + \mathbf{c}_{n0} + V_{nt}$ and $S_n(\lambda)S_n^{-1} = I_n - (\lambda - \lambda_0)G_n$, it implies that $\hat{\mathbf{c}}_{nT}(\theta) = \frac{1}{T}\sum_{t=1}^T((I_n - (\lambda-\lambda_0)G_n)(Z_{nt}\delta_0 + \mathbf{c}_{n0} + V_{nt}) - Z_{nt}\delta)$. Hence, for each fixed effect,

$$\hat{c}_{i,nT}(\hat{\theta}_{nT}) - c_{i,0} = -\frac{1}{T}\sum_{t=1}^T((G_n\mathbf{c}_{n0} + G_nZ_{nt}\delta_0)_i\,,(Z_{nt})_i) \times \begin{pmatrix}\hat{\lambda}_{nT} - \lambda_0\\ \hat{\delta}_{nT} - \delta_0\end{pmatrix} + \frac{1}{T}\sum_{t=1}^T\left\{\left(I_n - (\hat{\lambda}_{nT} - \lambda_0)G_n\right)V_{nt}\right\}_i.$$

As elements of $\frac{1}{T}\sum_{t=1}^T((G_n\mathbf{c}_{n0} + G_nZ_{nt}\delta_0)_i\,,(Z_{nt})_i)$ are $O_p(1)$ uniformly in n and i by Lemma 18 and $\hat{\theta}_{nT} - \theta_0 = O_p\left(\max\left(\sqrt{\frac{1}{nT}}, \frac{1}{T}\right)\right)$ by Theorem 3, the dominant term of $\hat{c}_{i,nT}(\hat{\theta}_{nT}) - c_{i,0}$ would be $\frac{1}{T}\sum_{t=1}^T v_{it}$. So, for each fixed effect, $\sqrt{T}\left(\hat{c}_{i,nT}(\hat{\theta}_{nT}) - c_0\right) \xrightarrow{d} N(0, \sigma_0^2)$ and they are independent from each other asymptotically. ■

Proof of Theorem 5. Theorem 3 states that $\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) + \sqrt{\frac{n}{T}}\Sigma_{\theta_0,nT}^{-1}\varphi_n + O_p\left(\max\left(\sqrt{\frac{n}{T^3}}, \sqrt{\frac{1}{T}}\right)\right) \xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0} + \Omega_{\theta_0})\Sigma_{\theta_0}^{-1})$. As $\hat{\theta}_{nT}^1 = \hat{\theta}_{nT} + \frac{1}{T}(-\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}(\theta_{nT})}{\partial\theta\partial\theta'})^{-1}\varphi_n(\hat{\theta}_{nT})$, $\sqrt{nT}(\hat{\theta}_{nT}^1 - \theta_0) \xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0} + \Omega_{\theta_0})\Sigma_{\theta_0}^{-1})$ if $\sqrt{\frac{n}{T}}((-\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}(\hat{\theta}_{nT})}{\partial\theta\partial\theta'})^{-1}\varphi_n(\hat{\theta}_{nT}) - \Sigma_{\theta_0,nT}^{-1}\varphi_n(\theta_0)) \xrightarrow{p} 0$ and $\frac{n}{T^3} \to 0$. Assuming that $\frac{n}{T^3} \to 0$, we are going to prove that

$$\sqrt{\frac{n}{T}}\left(\left(-\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}(\hat{\theta}_{nT})}{\partial\theta\partial\theta'}\right)^{-1}\varphi_n(\hat{\theta}_{nT}) - \Sigma_{\theta_0,nT}^{-1}\varphi_n(\theta_0)\right) \xrightarrow{p} 0. \quad (46)$$

From (44) and (45), $-\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}(\hat{\theta}_{nT})}{\partial\theta\partial\theta'} = \Sigma_{\theta_0,nT}^{-1} + O_p(\max(\frac{1}{T}, \frac{1}{\sqrt{nT}}))$. Hence,

$$\sqrt{\frac{n}{T}}\left(\left(-\frac{1}{nT}E\frac{\partial^2 \ln L_{nT}(\hat{\theta}_{nT})}{\partial\theta\partial\theta'}\right)^{-1}\varphi_n(\hat{\theta}_{nT}) - \Sigma_{\theta_0,nT}^{-1}\varphi_n(\theta_0)\right) = \sqrt{\frac{n}{T}}\left(\Sigma_{\theta_0,nT}^{-1}\left(\varphi_n(\hat{\theta}_{nT}) - \varphi_n(\theta_0)\right)\right) + \sqrt{\frac{n}{T}}\varphi_n(\hat{\theta}_{nT}) \times O_p\left(\max\left(\frac{1}{T}, \frac{1}{\sqrt{nT}}\right)\right).$$

As $\hat{\theta}_{nT} - \theta_0 = O_p\left(\max\left(\frac{1}{T}, \frac{1}{\sqrt{nT}}\right)\right)$ and $\varphi_n(\theta_0)$ is $O(1)$, according to the Taylor expansion of $\varphi_n(\hat{\theta}_{nT})$ in Box I around $\varphi_n(\theta_0)$, to prove (46) is reduced to prove that elements of $\frac{\partial\varphi_n(\bar{\theta}_{nT})}{\partial\theta'} < \infty$ where $\bar{\theta}_{nT}$ lies between $\hat{\theta}_{nT}$ and θ_0. As $A_n(\theta) = S_n^{-1}(\lambda)(\gamma I_n + \rho W_n)$, we have $\frac{\partial A_n(\theta)}{\partial\gamma} = S_n^{-1}(\lambda)$, $\frac{\partial A_n(\theta)}{\partial\rho} = S_n^{-1}(\lambda)W_n$, $\frac{\partial A_n(\theta)}{\partial\beta_i} = 0$ for $i = 1, 2, \ldots, k_x$ and $\frac{\partial A_n(\theta)}{\partial\lambda} = S_n^{-1}(\lambda)W_nS_n^{-1}(\lambda)(\gamma I_n + \rho W_n)$. Because[18] $\frac{\partial A_n^h(\theta)}{\partial\theta'} = hA_n^{h-1}(\theta)\frac{\partial A_n(\theta)}{\partial\theta'}$ for $h \geq 1$, $\sum_{h=1}^\infty\frac{\partial A_n^h(\theta)}{\partial\theta'} = \sum_{h=1}^\infty hA_n^{h-1}(\theta)\frac{\partial A_n(\theta)}{\partial\theta'}$. As (1) $\sum_{h=0}^\infty A_n^h(\theta)$ and $\sum_{h=1}^\infty hA_n^{h-1}(\theta)$ are uniformly bounded in either row sum or column sum, uniformly in a neighborhood of θ_0, (2) $S_n^{-1}(\lambda)$ is UB, also uniformly in λ in a neighborhood of λ_0 and (3) W_n is UB, it follows that the elements of $\frac{\partial\varphi_n(\theta)}{\partial\theta'}$ will be uniformly bounded in a neighborhood of θ_0. As $\bar{\theta}_{nT}$ converges in probability to θ_0, elements of $\frac{\partial\varphi_n(\bar{\theta}_{nT})}{\partial\theta'}$ are $O_p(1)$. ■

[18] This can be proved by mathematical induction. Step (i) For $h = 2$, $\frac{\partial A_n^2(\theta)}{\partial\lambda} = A_n(\theta)\frac{\partial A_n(\theta)}{\partial\lambda} + \frac{\partial A_n(\theta)}{\partial\lambda}A_n(\theta)$. Using $W_nS_n^{-1}(\lambda) = S_n^{-1}(\lambda)W_n$, $\frac{\partial A_n(\theta)}{\partial\lambda}A_n(\theta) = A_n(\theta)\frac{\partial A_n(\theta)}{\partial\lambda}$. So, $\frac{\partial A_n^2(\theta)}{\partial\lambda} = 2A_n(\theta)\frac{\partial A_n(\theta)}{\partial\lambda}$. Step (ii) Suppose $\frac{\partial A_n^h(\theta)}{\partial\lambda} = hA_n^{h-1}(\theta)\frac{\partial A_n(\theta)}{\partial\lambda}$, then $\frac{\partial A_n^{h+1}(\theta)}{\partial\lambda} = hA_n^{h-1}(\theta)\frac{\partial A_n(\theta)}{\partial\lambda}A_n(\theta) + A_n^h(\theta)\frac{\partial A_n(\theta)}{\partial\lambda} = (h+1)A_n^h(\theta)\frac{\partial A_n(\theta)}{\partial\lambda}$. Same arguments can be applied to other components of $\frac{\partial A_n^h(\theta)}{\partial\theta}$.

References

Alvarez, J., Arellano, M., 2003. The time series and cross-section asymptotics of dynamic panel data estimators. Econometrica 71, 1121–1159.

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

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic, The Netherlands.

Anselin, L., 2001. Spatial econometrics. In: Baltagi, Badi H. (Ed.), A Companion to Theoretical Econometrics. Blackwell Publishers Lte, Massachusetts, Chapter 14.

Anselin, L., Bera, A.K., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah, A., Giles, D.E.A. (Eds.), Handbook of Applied Economics Statistics. Marcel Dekker, New York.

Arellano, M., Bond, O., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies 58, 277–297.

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

Baltagi, B., Song, S.H., Kon, W., 2003. Testing panel data regression models with spatial error correlation. Journal of Econometrics 117, 123–150.

Baltagi, B., Song, S.H., Jung, B.C., Kon, W., 2007. Testing for serial correlation, spatial autocorrelation and random effects using panel data. Journal of Econometrics 140, 5–51.

Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87, 115–143.

Bun, M., Carree, M., 2005. Bias-corrected estimation in dynamic panel data models. Journal of Business & Economic Statistics 3 (2), 200-210(11).

Bun, M.J.G., Kiviet, J.F., 2006. The effects of dynamic feedbacks and LS and MM estimator accuracy in panel data models. Journal of Econometrics 132, 409–444.

Cliff, A.D., Ord, J.K., 1973. Spatial Autocorrelation. Pion Ltd., London.

Cressie, N., 1993. Statistics for Spatial Data. Wiley, New York.

Gänsler, P., Stute, W., 1977. Wahrscheinlichkeitstheorie. Springer Verlag, New York.

Hahn, J., Kuersteiner, G., 2002. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are Large. Econometrica 70 (4), 1639–1657.

Hahn, J., Newey, W., 2004. Jackknife and analytical bias reduction for nonlinear panel models. Econometrica 72 (4), 1295–1319.

Horn, R., Johnson, C., 1985. Matrix Algebra. Cambridge University Press.

Hsiao, C., 1986. Analysis of Panel Data. Cambridge University Press.

Kapoor, M., Kelejian, H.H., Prucha, I.R., 2007. Panel data models with spatially correlated error components. Journal of Econometrics 140, 97–130.

Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance. Journal of Real Estate Finance and Economics 17 (1), 99–121.

Kelejian, H.H., Prucha, I.R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. Journal of Econometrics 104, 219–257.

Kelejian, H.H., Robinson, D., 1993. A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a vounty expenditure model. Papers in Regional Science 72, 297–312.

Keller, W., Shiue, C.H., 2007. The origin of spatial interaction. Journal of Econometrics 140, 304–332.

Kiviet, J., 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. Journal of Econometrics 68, 81–126.

Korniotis, G.M., 2005. A dynamic panel estimator with both fixed and spatial effects, Manuscript, Department of Finance, University of Notre Dame.

Korniotis, G.M., 2007. Estimating panel models with internal and external habit formation. Manuscript, Board of Governors of the Federal Reserve System, Washington, DC.

Lee, L.F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models. Econometrica 72 (6), 1899–1925.

Lee, L.F., 2007. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. Journal of Econometrics 137, 489–514.

Lee, L.F, Liu, X., 2007. Efficient GMM estimation of high order spatail autoregressive models with autoregressive disturbances. Working Paper, Department of Economics, The Ohio State University.

Lee, L.F., Yu, J., 2007. A spatial dynamic panel data model with both time and individual fixed effects. Working Paper, The Ohio State University.

Neyman, J., Scott, E., 1948. Consistent estimates based on partially consistent observations. Econometrica 16 (1), 1–32.

Nickell, S.J., 1981. Biases in dynamic models with fixed effects. Econometrica 49, 1417–1426.

Ord, J.K., 1975. Estimation methods for models of spatial interaction. Journal of the American Statistical Association 70, 120–297.
Pesaran, M.H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74, 967–1012.
Phillips, P.C.B., Moon, H.R., 1999. Linear regression limit theory for nonstationary panel data. Econometrica 67 (5), 1057–1111.
Phillips, P.C.B., Sul, D., 2003. Dynamic panel estimation and homogeneity tests under cross section dependence. Econometric Journal 6, 217–259.
Pötscher, B.M., Prucha, I.R., 1997. Dynamic Nonlinear Econometric Models, Asymptotic Theory. Springer, New York.
Rothenberg, T.J., 1971. Identification in parametric models. Econometrica 39 (3), 577–591.
Yu, J., de Jong, R., Lee, L.-F., 2007. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large: A nonstationary case. Working Paper, The Ohio State University.