Econometric Theory, **26**, 2010, 1332–1362.
doi:10.1017/S0266466609990600

ESTIMATION OF UNIT ROOT SPATIAL DYNAMIC PANEL DATA MODELS

JIHAI YU
University of Kentucky

LUNG-FEI LEE
Ohio State University

This paper examines the asymptotics of the QMLE for unit root dynamic panel data models with spatial effect and fixed effects. We consider a unit root dynamic panel data model with spatially correlated disturbances and a unit root spatial dynamic panel data model. For both models the estimate of the dynamic coefficient is $\sqrt{nT^3}$ consistent and the estimates of other parameters are $\sqrt{nT}$ consistent, and all of them are asymptotically normal. For the latter model the sum of the contemporaneous spatial effect and dynamic spatial effect converges at $\sqrt{nT^3}$ rate. We also propose a bias-correction procedure so that the asymptotic biases of those estimates are eliminated as long as $n/T^3 \to 0$.

1. INTRODUCTION

Spatial econometrics deals with the spatial interactions of economic units in physical or economic characteristic spaces. Panel data are of great interest, as dynamic structures in the time dimension can also be incorporated into the model.[1] Yu, de Jong, and Lee (2008) consider a model with the specification

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

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, v_{it} is i.i.d. $(0, \sigma_0^2)$ across i and t, W_n is an $n \times n$ spatial weights matrix with a zero diagonal that generates the spatial dependence among y_{it}'s, X_{nt} is an $n \times k_x$ matrix of nonstochastic regressors, and $\mathbf{c}_{n0}$ is an $n \times 1$ vector of fixed effects. This model is termed the spatial dynamic panel data (SDPD) model. Under stationary conditions, Yu et al. (2008) derive the asymptotic properties of the maximum

We would like to thank participants of the Far Eastern Meeting of Econometric Society 2008 (Singapore), two anonymous referees, and Guido Kuersteiner, the co-editor of this journal, for helpful comments. Lee acknowledges financial support from NSF Grant No. SES-0519204. Address correspondence to Lung-fei Lee, Department of Economics, Ohio State University, Columbus, OH 43210, USA; e-mail: lflee@econ.ohio-state.edu.

 © Cambridge University Press 2010 0266-4666/10 $15.00

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

likelihood estimator (MLE) and the quasi-maximum likelihood estimator (QMLE). Empirical applications of the spatial panel data models are found in Korniotis (2005) on U.S. state consumption growth and in Keller and Shiue (2007) on market integration.[2]

This paper investigates a nonstationary SDPD model with unit roots in the time dimension. The difference of stationarity and nonstationarity for the model with W_n being row-normalized[3] depends on whether $\gamma_0 + \lambda_0 + \rho_0 = 1$ or less than 1. From the reduced-form equation (2) below, when the eigenvalues of $A_n = (I_n - \lambda_0 W_n)^{-1}(\gamma_0 I_n + \rho_0 W_n)$ are all less than 1 in absolute value, we have the stable case in Yu et al. (2008). However, when $\gamma_0 + \lambda_0 + \rho_0 = 1$, unit eigenvalues appear in the model. In this paper, we study the nonstationary case where $\gamma_0 = 1$ and $\lambda_0 + \rho_0 = 0$ and we term it a unit root SDPD model. The particular feature of this case is that all the eigenvalues of A_n will be 1, because $A_n = (I_n - \lambda_0 W_n)^{-1}(I_n - \lambda_0 W_n) = I_n$. This unit roots model would be a generalization of the unit root dynamic panel data model in the time series literature with some spatial features. The constraint $\lambda_0 + \rho_0 = 0$ may be generated by a panel unit root model with spatially correlated disturbances, or its spatial contemporaneous lag and spatial time lag effects may happen to be cancelled out. The purpose of this paper is to consider this pure unit roots model and the QMLE.[4] There is another distinct circumstance for nonstationarity, where $\gamma_0 + \lambda_0 + \rho_0 = 1$ but $\lambda_0 + \rho_0 \neq 0$, or equivalently, $\gamma_0 \neq 1$. This case is under investigation in Yu, de Jong, and Lee (2007) and is referred to as the spatial cointegration model because it generates spatial cointegration across spatial units. Due to a sharp discontinuity in the eigenvalue structure of A_n for the two models, the asymptotics in Yu, et al. (2007) for the spatial cointegration case do not carry over to this pure unit roots case. A table in the conclusion will highlight all the differences.

The paper is organized as follows: In Section 2 we introduce the model. Section 3 studies unit root dynamic panel data with spatially correlated disturbances. Section 4 investigates the unit root SDPD model. We also propose an analytical bias correction for our estimators. Simulation results are shown in Section 5. Section 6 concludes the paper.

2. FEATURES OF THE GENERAL MODEL AND ASSUMPTIONS

Define $S_n(\lambda) = I_n - \lambda W_n$ and denote $S_n \equiv S_n(\lambda_0)$. 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

$$Y_{nt} = A_n Y_{n,t-1} + S_n^{-1} X_{nt}\beta_0 + S_n^{-1}\mathbf{c}_{n0} + S_n^{-1} V_{nt}. \tag{2}$$

By continuous substitution, for the stable case one may have Y_{nt} as an infinite sum involving all past exogenous variables and disturbances. If some of the eigenvalues of A_n are equal to 1, there will be nonstationary components in the process and it will invalidate the infinite sum. When $\gamma_0 = 1$ and $\lambda_0 + \rho_0 = 0$, A_n is equal to I_n and all the roots in (2) are 1. By denoting $\mathbb{X}_{nt} = \sum_{h=0}^{t} X_{nh}$, $\xi_{nt} = \sum_{h=0}^{t} V_{nh}$,

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

we have from (2) a finite sum starting from $t = -1$,

$$Y_{nt} = Y_{n,-1} + (t+1)S_n^{-1}\mathbf{c}_{n0} + S_n^{-1}\mathbb{X}_{nt}\beta_0 + S_n^{-1}\xi_{nt}. \tag{3}$$

For notational purposes, we define, for any $n \times 1$ vector or matrix Υ_{nt} with n rows at period t, $\tilde{\Upsilon}_{nt} = \Upsilon_{nt} - \bar{\Upsilon}_{nT}$ and $\tilde{\tilde{\Upsilon}}_{n,t-1} = \Upsilon_{n,t-1} - \bar{\Upsilon}_{nT,-1}$ for $t = 1, 2, \ldots, T$, where $\bar{\Upsilon}_{nT} = \frac{1}{T}\sum_{t=1}^{T}\Upsilon_{nt}$ and $\bar{\Upsilon}_{nT,-1} = \frac{1}{T}\sum_{t=1}^{T}\Upsilon_{n,t-1}$. With $\bar{Y}_{nT} = \frac{1}{T}\sum_{t=1}^{T}Y_{nt} = Y_{n,-1} + \frac{T+3}{2}S_n^{-1}\mathbf{c}_{n0} + \frac{1}{T}S_n^{-1}\sum_{t=1}^{T}\mathbb{X}_{nt}\beta_0 + \frac{1}{T}S_n^{-1}\sum_{t=1}^{T}\xi_{nt}$ and $\tilde{Y}_{nt} = Y_{nt} - \bar{Y}_{nT}$, we have

$$\tilde{Y}_{nt} = \tilde{t}S_n^{-1}\mathbf{c}_{n0} + S_n^{-1}\tilde{\mathbb{X}}_{nt}\beta_0 + S_n^{-1}\tilde{\xi}_{nt},$$

where $\tilde{t} = t + 1 - \frac{T+3}{2} = t - \frac{T+1}{2}$. Therefore $\tilde{Y}_{nt}$ has a time trend $\tilde{t}S_n^{-1}\mathbf{c}_{n0} + S_n^{-1}\tilde{\mathbb{X}}_{nt}\beta_0$ and a random walk $S_n^{-1}\tilde{\xi}_{nt}$. This represents the nonstationarity feature of $\tilde{Y}_{nt}$ in the model.

To analyze the model, we make the following assumptions:

Assumption 1. W_n is a nonstochastic spatial weights matrix with a zero diagonal.

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 $\mathrm{E}|v_{it}|^{4+\eta} < \infty$ for some $\eta > 0$.

Assumption 3. Here n is a nondecreasing function of T and T goes to infinity.

Assumption 4. The elements of X_{nt} are nonstochastic and bounded uniformly in n and t.

Assumption 5. $S_n(\lambda)$ is invertible for all $\lambda \in \Lambda$ where Λ is compact and the true parameter λ_0 is in the interior of Λ.

Assumption 6. Assume $\gamma_0 = 1$ and $\rho_0 + \lambda_0 = 0$.

Assumption 7. The row and column sums of W_n and $S_n^{-1}(\lambda)$ are uniformly bounded (UB) in n, and also uniformly in $\lambda \in \Lambda$ for $S_n^{-1}(\lambda)$.[5]

Assumptions 1 and 2 provide essential features of the weights matrix and disturbances of the model.[6] Assumption 2 provides a simple stochastic structure on the disturbances. The model (1) relates to unit root panel data models with spatially correlated disturbances. Extending the disturbances to be serially correlated over time may be possible, but the extended model in a panel setting will be more complicated.[7] Assumption 3 allows two cases: (i) $n \to \infty$ as $T \to \infty$; and (ii) n is fixed as $T \to \infty$. For case (i), we say that $n, T \to \infty$ simultaneously. When exogenous variables X_{nt} are included in the model, they are assumed to be uniformly bounded in Assumption 4 for simplicity. In general, X_{nt}

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

may consist of exogenous variables, say, $X_{1,nt}$, and also variables with spatial interactions, such as $W_nX_{1,nt}$. As a consequence of Assumption 5, for the system (1), Y_{nt} can be solved as in (2). Assumption 6 specifies that all the roots of A_n are equal to 1. It also allows the following possible data generating process (DGP): $Y_{nt} = Y_{n,t-1} + \mathbf{c}_{n0} + U_{nt}$ where $U_{nt} = \lambda_0 W_n U_{nt} + V_{nt}$. This is a unit root panel data process where the disturbances in U_{nt} are spatially correlated; we will pay attention to such a model in Section 3 with regressors incorporated in the main equation. Assumption 7 is originated by Kelejian and Prucha (1998, 2001) and also used in Lee (2004, 2007). The uniform boundedness of W_n and $S_n^{-1}(\lambda)$ is a condition to limit the spatial correlation across units to a manageable degree.

3. UNIT ROOT PANEL DATA MODEL WITH SPATIALLY CORRELATED DISTURBANCES

In this section we consider the following unit root dynamic panel model with spatially correlated disturbances:

$$Y_{nt} = \gamma_0 Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + U_{nt}, \tag{4}$$

$$U_{nt} = \lambda_0 W_n U_{nt} + V_{nt},$$

where $\gamma_0 = 1$. We are interested in the QMLE of γ_0, λ_0 and other parameters. The term $W_nY_{n,t-1}$ does not appear explicitly in the first equation of (4), but it appears with the coefficient $\rho_0 = -\gamma_0\lambda_0$ in the transformed equation with the transformation $(I_n - \lambda_0 W_n)$ as in (1). Hence, this model can be regarded as a special case of interest in (1) with the constraint[8] $\rho_0 = -\gamma_0\lambda_0$.

3.1. Likelihood Function

Denote $\theta = (\delta', \lambda, \sigma^2)'$, where $\delta = (\gamma, \beta')'$ and $Z_{nt} = (Y_{n,t-1}, X_{nt})$. The likelihood function of (4) is

$$\ln L_{n,T}(\theta, \mathbf{c}_n) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln\sigma^2 + T\ln|S_n(\lambda)| - \frac{1}{2\sigma^2}\sum_{t=1}^{T} V_{nt}'(\theta, \mathbf{c}_n)V_{nt}(\theta, \mathbf{c}_n), \tag{5}$$

where $V_{nt}(\theta, \mathbf{c}_n) = S_n(\lambda)(Y_{nt} - Z_{nt}\delta - \mathbf{c}_n)$. From (5), the concentrated likelihood at θ is

$$\ln L_{n,T}(\theta) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln\sigma^2 + T\ln|S_n(\lambda)| - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{V}_{nt}'(\theta)\tilde{V}_{nt}(\theta), \tag{6}$$

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

where $\tilde{V}_{nt}(\theta) = S_n(\lambda)(\tilde{Y}_{nt} - \tilde{Z}_{nt}\delta)$ and the score and information matrix are (A.1) and (A.2) in Appendix A.1. The QMLE $\hat{\theta}_{nT}$ maximizes (6). Denote

$$\mathcal{H}_{nT} = \frac{1}{nT}\sum_{t=1}^{T} \tilde{Z}_{nt}' S_n' S_n \tilde{Z}_{nt}$$

$$= \begin{pmatrix} \frac{1}{nT}\sum_{t=1}^{T} \tilde{\tilde{Y}}_{n,t-1}' S_n' S_n \tilde{\tilde{Y}}_{n,t-1} & * \\ \frac{1}{nT}\sum_{t=1}^{T} \tilde{X}_{nt}' S_n' S_n \tilde{\tilde{Y}}_{n,t-1} & \frac{1}{nT}\sum_{t=1}^{T} \tilde{X}_{nt}' S_n' S_n \tilde{X}_{nt} \end{pmatrix}, \tag{7}$$

$\Upsilon_{\theta,T} = \mathrm{diag}(T, l_{k_x}', 1, 1)$, $\Upsilon_{\delta,T} = \mathrm{diag}(T, l_{k_x}')$,

where l_{k_x} is $k_x \times 1$ vector of ones.

Assumption 8. Assume $\lim_{T\to\infty} \mathrm{E}\Upsilon_{\delta,T}^{-1}\mathcal{H}_{nT}\Upsilon_{\delta,T}^{-1}$ is nonsingular, $\lim_{n\to\infty}\left(\frac{1}{n}\ln\left|\sigma_0^2 S_n^{-1\prime} S_n^{-1}\right| - \frac{1}{n}\ln\left|\sigma_n^2(\lambda) S_n^{-1}(\lambda)' S_n^{-1}(\lambda)\right|\right) \neq 0$ for $\lambda \neq \lambda_0$, and $\lim_{T\to\infty} \frac{1}{nT^3}\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t} + \tilde{\mathbb{X}}_{nt}\beta_0)' S_n' S_n (\mathbf{c}_{n0}\tilde{t} + \tilde{\mathbb{X}}_{nt}\beta_0) \neq 0$.

The first part of Assumption 8 is for the identification of δ_0. The second part is for the identification of λ_0. The last part specifies that the nonstationary component is present asymptotically and it is dominated by the linear trend generated by either $\mathbf{c}_{n0}$ or X_{nt}.

The optimization for (6) can be simplified by focusing on the concentrated log likelihood of λ. From (6), given λ,

$$\hat{\delta}_{nT}(\lambda) = \left[\frac{1}{nT}\sum_{t=1}^{T} \tilde{Z}_{nt}' S_n'(\lambda) S_n(\lambda) \tilde{Z}_{nt}\right]^{-1}\left[\frac{1}{nT}\sum_{t=1}^{T} \tilde{Z}_{nt}' S_n'(\lambda) S_n(\lambda) \tilde{Y}_{nt}\right], \tag{8}$$

$$\hat{\sigma}_{nT}^2(\lambda) = \frac{1}{nT}\sum_{t=1}^{T}[S_n(\lambda)(\tilde{Y}_{nt} - \tilde{Z}_{nt}\hat{\delta}_{nT}(\lambda))]'[S_n(\lambda)(\tilde{Y}_{nt} - \tilde{Z}_{nt}\hat{\delta}_{nT}(\lambda))], \tag{9}$$

and the concentrated likelihood at λ is

$$\ln L_{n,T}(\lambda) = -\frac{nT}{2}(\ln 2\pi + 1) - \frac{nT}{2}\ln\hat{\sigma}_{nT}^2(\lambda) + T\ln|S_n(\lambda)|. \tag{10}$$

3.2. Asymptotics for $\hat{\theta}_{nT}$

The following theorem is on the asymptotic properties of the estimators.

THEOREM 1. *Under Assumptions 1–8, the QMLE $\hat{\theta}_{nT}$ of (6) for the model (4) is consistent.*

When $\frac{n}{T} \to 0$, $\Upsilon_{\theta,T}\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) \xrightarrow{d} N(0, \lim_{T\to\infty}(\Sigma_{\theta_0,nT})^{-1} + \lim_{T\to\infty}[(\Sigma_{\theta_0,nT})^{-1}\Omega_{\theta_0,n}(\Sigma_{\theta_0,nT})^{-1}])$.

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

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

$$\Upsilon_{\theta,T}\sqrt{nT}\left(\hat{\theta}_{nT}-\theta_0\right)+\sqrt{k}b_{\theta_0,nT} \xrightarrow{d} N\left(0, \lim_{T\to\infty}(\Sigma_{\theta_0,nT})^{-1}+\lim_{T\to\infty}\left[(\Sigma_{\theta_0,nT})^{-1}\Omega_{\theta_0,n}(\Sigma_{\theta_0,nT})^{-1}\right]\right).$$

When $\frac{n}{T} \to \infty$, $\Upsilon_{\theta,T}T(\hat{\theta}_{nT}-\theta_0)+b_{\theta_0,nT} \xrightarrow{p} 0$.

Here, with $G_n(\lambda) = W_n S_n^{-1}(\lambda)$ *and* $G_n = W_n S_n^{-1}$,

$$b_{\theta_0,nT} = (\Sigma_{\theta_0,nT})^{-1}a_{\theta_0,n} \quad \text{where } a_{\theta,n} = \left(\frac{1}{2}, \mathbf{0}_{1\times k_x}, \frac{1}{n}\mathrm{tr}G_n(\lambda), \frac{1}{2\sigma^2}\right)', \tag{11}$$

$$\Sigma_{\theta_0,nT} = \begin{pmatrix} \frac{1}{\sigma_0^2}\mathrm{E}\Upsilon_{\delta,T}^{-1}\mathcal{H}_{nT}\Upsilon_{\delta,T}^{-1} & \mathbf{0}_{(k_x+1)\times 1} & \mathbf{0}_{(k_x+1)\times 1} \\ \mathbf{0}_{1\times(k_x+1)} & \frac{1}{n}\left[\mathrm{tr}(G_n'G_n)+\mathrm{tr}(G_n^2)\right] & \frac{1}{\sigma_0^2 n}\mathrm{tr}(G_n) \\ \mathbf{0}_{1\times(k_x+1)} & \frac{1}{\sigma_0^2 n}\mathrm{tr}(G_n) & \frac{1}{2\sigma_0^4} \end{pmatrix}, \tag{12}$$

and

$$\Omega_{\theta_0,n} = \frac{\mu_4-3\sigma_0^4}{\sigma_0^4}\begin{pmatrix} \mathbf{0}_{(k_x+1)\times(k_x+1)} & * & * \\ \mathbf{0}_{1\times(k_x+1)} & \frac{1}{n}\sum_{i=1}^n G_{n,ii}^2 & * \\ \mathbf{0}_{1\times(k_x+1)} & \frac{1}{2\sigma_0^2 n}\mathrm{tr}G_n & \frac{1}{4\sigma_0^4} \end{pmatrix}. \tag{13}$$

This result implies that $\hat{\gamma}_{nT}$ will have a rate of convergence $\sqrt{nT^3}$ with a bias of the order $O(T^{-2})$, while $\hat{\lambda}_{nT}$, $\hat{\beta}_{nT}$, and $\hat{\sigma}_{nT}^2$ will have a rate of convergence $\sqrt{nT}$ with biases of the order $O(T^{-1})$. When T is small relative to n in the sense that $\frac{n}{T} \to \infty$, the presence of $b_{\theta_0,nT}$ causes $\hat{\lambda}_{nT}$, $\hat{\beta}_{nT}$, and $\hat{\sigma}_{nT}^2$ to have the slower T^{-1} rate of convergence even though $\hat{\gamma}_{nT}$ has the T^{-2} rate of convergence. We can derive the bias-corrected estimator $\hat{\theta}_{nT}^1$ as

$$\hat{\theta}_{nT}^1 = \hat{\theta}_{nT} - \frac{\hat{B}_{nT}}{T}, \tag{14}$$

where $\hat{B}_{nT} = -[\Sigma_{\theta,nT}]^{-1}\cdot a_{\theta,n}|_{\theta=\hat{\theta}_{nT}}$. We show that, when $n/T^3 \to 0$, $\hat{\theta}_{nT}^1$ will be asymptotically centered normal even when $n/T \to \infty$.

THEOREM 2. *Under Assumptions 1–8 for the model (4), if* $\frac{n}{T^3} \to 0$,

$$\Upsilon_{\theta,T}\sqrt{nT}(\hat{\theta}_{nT}^1-\theta_0) \xrightarrow{d} N\left(0, \lim_{T\to\infty}(\Sigma_{\theta_0,nT})^{-1}+\lim_{T\to\infty}\left[(\Sigma_{\theta_0,nT})^{-1}\Omega_{\theta_0,n}(\Sigma_{\theta_0,nT})^{-1}\right]\right). \tag{15}$$

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

Hahn and Kuersteiner (2002) show that for the model $y_{it} = \gamma_0 y_{i,t-1} + \alpha_i + v_{it}$ where $\gamma_0 = 1$, $\sqrt{nT^3}(\hat{\gamma}_{nT} - \gamma_0) \xrightarrow{d} N\left(\frac{6\sigma_0^2 \lim\sqrt{\frac{n}{T}}}{\lim\frac{1}{n}\sum_{i=1}^n \alpha_i^2}, \frac{12\sigma_0^2}{\lim\frac{1}{n}\sum_{i=1}^n \alpha_i^2}\right)$ when $\lim\frac{n}{T}$ exists and is a positive finite constant. From (11) in Theorem 1 for the model (4) with spatial disturbances, the bias of $\hat{\gamma}_{nT}$ is the same as in Hahn and Kuersteiner (2002) when $\beta_0 = 0$ is imposed in the estimation and $\lambda_0 = 0$ in the DGP. To see that, from Theorem 1, due to the block diagonal matrices in (12) and (13) when $\beta_0 = 0$ is imposed, we have

$$\sqrt{nT^3}(\hat{\gamma}_{nT} - \gamma_0) + \sqrt{\frac{n}{T}} b_{\gamma_0,nT} \xrightarrow{d} N\left(0, \sigma_0^2 \lim_{T\to\infty}\left[\frac{1}{nT^3}\sum_{t=1}^T \tilde{\tilde{Y}}'_{n,t-1} S_n' S_n \tilde{\tilde{Y}}_{n,t-1}\right]^{-1}\right),$$

when the limit of $\frac{n}{T}$ is a positive finite constant. Here, $b_{\gamma_0,nT} = \left[\frac{1}{nT^3}\mathrm{E}\sum_{t=1}^T \tilde{\tilde{Y}}'_{n,t-1} S_n' S_n \tilde{\tilde{Y}}_{n,t-1}\right]^{-1} \frac{1}{2}\sigma_0^2$, which is essentially the same as in Hahn and Kuersteiner (2002) when $\lambda_0 = 0$.[9] Hence, when λ_0 happens to be 0, the asymptotic bias value and the asymptotic distribution of $\hat{\gamma}_{nT}$ are the same as those in Hahn and Kuersteiner. However, the estimation of the spatial effect will always change the value of $b_{\gamma_0,nT}$ for the unrestricted unit root SDPD model considered in the next section, even though the true spatial effect might be zero.

4. THE UNIT ROOT SDPD MODEL

In this section,[10] we consider the model (1), where $\gamma_0 = 1$ and $\lambda_0 + \rho_0 = 0$; but, for the estimation, the constraint on λ and ρ is also not imposed.

4.1. Likelihood Function

Denote $\theta = (\delta', \lambda, \sigma^2)'$, where $\delta = (\gamma, \rho, \beta')'$ and $Z_{nt} = (Y_{n,t-1}, W_n Y_{n,t-1}, X_{nt})$. The likelihood function of (1) and its concentrated likelihood are slightly different from those in Section 3. The likelihood function is now

$$\ln L_{n,T}(\theta, \mathbf{c}_n) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln\sigma^2 + T\ln|S_n(\lambda)| - \frac{1}{2\sigma^2}\sum_{t=1}^T V_{nt}'(\theta, \mathbf{c}_n) V_{nt}(\theta, \mathbf{c}_n), \tag{16}$$

where $V_{nt}(\theta, \mathbf{c}_n) = S_n(\lambda) Y_{nt} - Z_{nt}\delta - \mathbf{c}_n$. The reduced form (2) can be represented as

$$Y_{nt} = S_n^{-1}(Z_{nt}\delta + \mathbf{c}_{n0} + V_{nt}) = Z_{nt}\delta_0 + \lambda_0 G_n Z_{nt}\delta_0 + S_n^{-1}(\mathbf{c}_{n0} + V_{nt}), \tag{17}$$

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

because $S_n^{-1} = I_n + \lambda_0 G_n$. Under Assumption 6,

$$G_n \tilde{Z}_{nt}\delta_0 = W_n S_n^{-1}(\tilde{\tilde{Y}}_{n,t-1} - \lambda_0 W_n \tilde{\tilde{Y}}_{n,t-1} + \tilde{X}_{nt}\beta_0)$$
$$= W_n \tilde{\tilde{Y}}_{n,t-1} + G_n \tilde{X}_{nt}\beta_0. \tag{18}$$

The relation (18) is an important one generated by the model (1) with unit roots in addition to the structure (3). These features appear neither in the spatial cointegration model nor the stable model in Yu et al. (2007, 2008). As $G_n \tilde{Z}_{nt}\delta_0$ has the nonstationary term $W_n \tilde{\tilde{Y}}_{n,t-1}$ as one of its components, it will cause multicollinearity between $W_n \tilde{Y}_{nt}$ and $W_n \tilde{\tilde{Y}}_{n,t-1}$, which has implications on rates of convergence of estimates.

Denote $\Upsilon_{\delta,T} = \mathrm{diag}(T, T, l'_{k_x})$ and

$$\tilde{Z}_{nt}^s = \tilde{Z}_{nt}\Upsilon_{\delta,T}^{-1} = \left(\frac{1}{T}\tilde{\tilde{Y}}_{n,t-1}, \frac{1}{T}W_n\tilde{\tilde{Y}}_{n,t-1}, \tilde{X}_{nt}\right),$$

$$\mathcal{H}_{nT}^s = \begin{pmatrix} \mathcal{H}_{1,nT}^s & \mathcal{H}_{2,nT}^s \\ \mathcal{H}_{2,nT}^{s\prime} & \mathcal{H}_{3,nT}^s \end{pmatrix}, \quad \text{where} \tag{19}$$

$$\mathcal{H}_{1,nT}^s = \frac{1}{nT}\sum_{t=1}^{T} \tilde{Z}_{nt}^{s\prime}\tilde{Z}_{nt}^s,\ \mathcal{H}_{2,nT}^s = \frac{1}{nT}\sum_{t=1}^{T} \tilde{Z}_{nt}^{s\prime} G_n \tilde{X}_{nt}\beta_0, \text{ and}$$

$$\mathcal{H}_{3,nT}^s = \frac{1}{nT}\sum_{t=1}^{T} (G_n \tilde{X}_{nt}\beta_0)' G_n \tilde{X}_{nt}\beta_0.$$

Assumptions 8′. Assume $\lim_{T\to\infty} \mathrm{E}\mathcal{H}_{nT}^s$ is nonsingular, or $\lim_{n\to\infty}(\frac{1}{n}\ln|\sigma_0^2 S_n^{-1\prime}S_n^{-1}| - \frac{1}{n}\ln|\sigma_n^2(\lambda)S_n^{-1}(\lambda)'S_n^{-1}(\lambda)|) \neq 0$ for $\lambda \neq \lambda_0$ and $\lim_{T\to\infty}\frac{1}{nT}\sum_{t=1}^{T}\tilde{X}'_{nt}\tilde{X}_{nt}$ exists and is nonsingular. Additionally, $\lim_{T\to\infty}\frac{1}{nT^3}\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t} + \tilde{\mathbb{X}}_{nt}\beta_0)'(\mathbf{c}_{n0}\tilde{t} + \tilde{\mathbb{X}}_{nt}\beta_0) \neq 0$.

The first part of Assumption 8′ states that (i) $G_n \tilde{X}_{nt}\beta_0$ is not multicollinear with $\tilde{Z}_{nt}^s$ asymptotically or (ii) the spatial effect can be identified from disturbances of the reduced-form equation (17). Statement (i) is related to the identification of the model parameters from the regression component of the reduced form equation. In the event of multicollinearity, the regression equation would not provide enough information for parameter identification, in particular, λ_0, and then the identification of λ_0 via the reduced form disturbances will be needed. The last part of Assumption 8′ corresponds to that of Assumption 8 for the restricted model.

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

4.2. Asymptotics for $\hat{\theta}_{nT}$

Denote

$$\Xi_T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & T^{-1} & 0 & -1 & 0 \\ 0 & 0 & I_{k_x} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{20}$$

We have the following theorem on the asymptotic properties of the estimators.

THEOREM 3. *Under Assumptions 1–7 and 8′, the QMLE $\hat{\theta}_{nT}$ of (17) for the model (1) is consistent.*

When $\frac{n}{T} \to 0$,

$$\begin{aligned} &\operatorname{diag}(T, 1, l'_{k_x}, 1, 1)\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) \\ &\quad \xrightarrow{d} N(0, \lim_{T\to\infty}(\Xi_T[(\Sigma^s_{\theta_0,nT})^{-1} + (\Sigma^s_{\theta_0,nT})^{-1}\Omega_{\theta_0,n}(\Sigma^s_{\theta_0,nT})^{-1}]\Xi'_T)). \end{aligned}$$

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

$$\begin{aligned} &\operatorname{diag}(T, 1, l'_{k_x}, 1, 1)\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) + \sqrt{k}b_{\theta_0,nT} \\ &\quad \xrightarrow{d} N(0, \lim_{T\to\infty}(\Xi_T[(\Sigma^s_{\theta_0,nT})^{-1} + (\Sigma^s_{\theta_0,nT})^{-1}\Omega_{\theta_0,n}(\Sigma^s_{\theta_0,nT})^{-1}]\Xi'_T)). \end{aligned}$$

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

$$\operatorname{diag}(T, 1, l'_{k_x}, 1, 1)T(\hat{\theta}_{nT} - \theta_0) + b_{\theta_0,nT} \xrightarrow{p} 0.$$

Here,

$$b_{\theta_0,nT} \equiv \Xi_T \cdot [\Sigma^s_{\theta_0,nT}]^{-1} \cdot a_{\theta_0,n},$$

$$a_{\theta,n} = \left(\frac{1}{2n}\operatorname{tr}(S_n^{-1}(\lambda)), \frac{1}{2n}\operatorname{tr}(G_n(\lambda)), \mathbf{0}_{1\times k_x}, \frac{1}{n}\operatorname{tr}(G_n(\lambda)), \frac{1}{2\sigma^2}\right)', \tag{21}$$

$$\begin{aligned} \Sigma^s_{\theta_0,nT} = &\frac{1}{\sigma_0^2}\begin{pmatrix} \mathrm{E}\mathcal{H}^s_{nT} & \mathbf{0}_{(k_x+3)\times 1} \\ \mathbf{0}_{1\times(k_x+3)} & 0 \end{pmatrix} \\ &+ \begin{pmatrix} \mathbf{0}_{(k_x+2)\times(k_x+2)} & \mathbf{0}_{(k_x+2)\times 1} & \mathbf{0}_{(k_x+2)\times 1} \\ \mathbf{0}_{1\times(k_x+2)} & \frac{1}{n}\left[\operatorname{tr}(G'_nG_n) + \operatorname{tr}(G_n^2)\right] & \frac{1}{\sigma_0^2 n}\operatorname{tr}(G_n) \\ \mathbf{0}_{1\times(k_x+2)} & \frac{1}{\sigma_0^2 n}\operatorname{tr}(G_n) & \frac{1}{2\sigma_0^4} \end{pmatrix}, \end{aligned} \tag{22}$$

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

$$\Omega_{\theta_0,n} = \frac{\mu_4 - 3\sigma_0^4}{\sigma_0^4} \cdot \begin{pmatrix} \mathbf{0}_{(k_x+2)\times(k_x+2)} & \mathbf{0}_{(k_x+2)\times 1} & \mathbf{0}_{(k_x+2)\times 1} \\ \mathbf{0}_{1\times(k_x+2)} & \frac{1}{n}\sum_{i=1}^n G_{n,ii}^2 & \frac{1}{2\sigma_0^2 n}\mathrm{tr}G_n \\ \mathbf{0}_{1\times(k_x+2)} & \frac{1}{2\sigma_0^2 n}\mathrm{tr}G_n & \frac{1}{4\sigma_0^4} \end{pmatrix}. \tag{23}$$

Additionally, denote $e_2 = (0, 1, \mathbf{0}_{1\times k_x}, 0, 0)'$,

$$\sqrt{nT^3}(\hat{\rho}_{nT} + \hat{\lambda}_{nT}) + e_2'\sqrt{\frac{n}{T}}b_{\theta_0,nT} + O_p\left(\max\left(\frac{1}{\sqrt{T}}, \sqrt{\frac{n}{T^3}}\right)\right)$$

$$\xrightarrow{d} N(0, \lim_{T\to\infty}(e_2'[(\Sigma^s_{\theta_0,nT})^{-1} + (\Sigma^s_{\theta_0,nT})^{-1}\Omega_{\theta_0,n}(\Sigma^s_{\theta_0,nT})^{-1}]e_2)). \tag{24}$$

Hence, $\hat{\gamma}_{nT}$ is $\sqrt{nT^3}$ consistent and $\hat{\lambda}_{nT}$, $\hat{\rho}_{nT}$ are $\sqrt{nT}$ consistent. Additionally, $\hat{\lambda}_{nT} + \hat{\rho}_{nT}$ is $\sqrt{nT^3}$ consistent. The QMLE has the bias $-\frac{1}{T}\mathrm{diag}(T^{-1}, 1, l_{k_x}', 1, 1)b_{\theta_0,nT}$ and the confidence interval will not center properly for $\hat{\theta}_{nT}$ when $\frac{n}{T} \to k$ where $0 < k < \infty$. The bias-corrected estimator is in the form of (14) where $\hat{B}_{nT} = \mathrm{diag}(T^{-1}, 1, l_{k_x}', 1, 1)\cdot[-\Xi_T\cdot(\Sigma^s_{\theta,nT})^{-1}\cdot a_{\theta,n}]|_{\theta=\hat{\theta}_{nT}}$. This bias reduction procedure will be valuable for all estimates in $\hat{\theta}_{nT}$, even if the bias of $\hat{\gamma}_{nT}$ is already of the small order $O(T^{-2})$. We show that, when $n/T^3 \to 0$, $\hat{\theta}^1_{nT}$ in (14) will be asymptotically centered normal even when $n/T \to \infty$.

THEOREM 4. *Under Assumptions 1–7 and 8′ for the model (1), if* $\frac{n}{T^3} \to 0$,

$$\mathrm{diag}(T, 1, l_{k_x}', 1, 1)\sqrt{nT}(\hat{\theta}^1_{nT} - \theta_0)$$

$$\xrightarrow{d} N(0, \lim_{T\to\infty}(\Xi_T[(\Sigma^s_{\theta_0,nT})^{-1} + (\Sigma^s_{\theta_0,nT})^{-1}\Omega_{\theta_0,n}(\Sigma^s_{\theta_0,nT})^{-1}]\Xi_T')). \tag{25}$$

Compared to Hahn and Kuersteiner (2002), even though $\rho_0 = \lambda_0 = 0$ in the DGP so that $\frac{1}{n}\mathrm{tr}G_n = 0$ in $a_{\theta_0,n}$, $\mathrm{E}\mathcal{H}^s_{nT}$ in the information matrix is not diagonal in general, as ρ_0 and λ_0 are included in the estimation. Hence, the bias of $\hat{\gamma}_{nT}$ is different from that in Hahn and Kuersteiner for this unrestricted model. Similarly, the asymptotic variance of $\hat{\gamma}_{nT}$ is also different from that in Hahn and Kuersteiner.

5. SIMULATION

We conduct a small Monte Carlo experiment to evaluate the performance of the MLEs and the bias-corrected estimators of the unrestricted model. We generate samples from (1) and use $\theta_0^a = (1, -0.5, 1, 0.5, 1)'$ or $\theta_0^b = (1, -0.1, 1, 0.1, 1)'$, where $\theta_0 = (\gamma_0, \rho_0, \beta_0', \lambda_0, \sigma_0^2)'$; X_{nt}, $\mathbf{c}_{n0}$, and V_{nt} are generated from independent normal distributions; and the spatial weights matrix is a rook matrix.[11] We use

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

$T = 20$ and $T = 50$, and $n = 25$, $n = 49$, and $n = 196$. For each set of generated sample observations, we calculate the MLE $\hat{\theta}_{nT}$ and evaluate its bias, $\hat{\theta}_{nT} - \theta_0$. We then construct the bias-corrected estimator $\hat{\theta}^1_{nT}$ and evaluate the bias, $\hat{\theta}^1_{nT} - \theta_0$. We do this 1,000 times to see if the bias is reduced on average by using the analytical bias-correction procedure, i.e., to compare $\frac{1}{1000}\sum_{i=1}^{1000}(\hat{\theta}_{nT} - \theta_0)_i$ with $\frac{1}{1000}\sum_{i=1}^{1000}(\hat{\theta}^1_{nT} - \theta_0)_i$. With two different values of θ_0 for each n and T, finite sample properties of both estimators are summarized in Tables 1 and 2, where Table 1 is for the performance of the estimators before bias correction and Table 2 is for the performance after 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 biases, but they are reduced with the bias-corrected estimators. Even though $\hat{\gamma}^1_{nT}$ has a bias of the order of $O(T^{-2})$ while the biases of the other parameters are $O(1/T)$, we observe in this Monte Carlo study that the magnitude of bias of $\hat{\gamma}_{nT}$ may not be smaller than those of the other estimates. This might be due to its finite sample property, where the coefficient of its order could be larger than those of the others. The bias-correction procedure turns out to be helpful for $\hat{\gamma}_{nT}$, in addition to the other parameter estimates. The bias-corrected estimator $\hat{\gamma}^1_{nT}$ will eliminate the bias of order $O(T^{-2})$ and the bias corrected estimators for other parameters will eliminate the bias of order $O(T^{-1})$. Also, the 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.[12]

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 with θ_0^a and θ_0^b will be smaller and the variance will be smaller; and 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 orders of estimators depend on T but not on n, and the variance of the estimators is of the order $O(\frac{1}{nT^3})$ for the estimate of γ_0 and $O(\frac{1}{nT})$ for estimates of other parameters. Also, for different values of θ_0, the biases become larger when θ_0 is the larger θ_0^a, and the variances do not change much.

We also run small simulations to check the consequences due to misspecification of the disturbances, where the true disturbances might have serial correlation or spatial autocorrelation. To avoid unnecessary tables, we report the results for $n = 25$ and $T = 20$ in Table 3. From Misspecification 1 with serial correlated disturbances, we can see that serial correlation in the disturbances does not cause the bias problem. This is similar to the time series literature due to the deterministic trend component in the DGP. From Misspecification 2 with spatially correlated disturbances, we can see that the spatial correlation in the disturbances will cause a severe bias problem, mostly for λ_0 and ρ_0. However, for the

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

TABLE 1. Performance of estimators before bias correction

	T	n	θ_0		γ	ρ	β	λ	σ^2	$\rho+\lambda$
(1)	20	25	θ_0^a	Bias	−0.0135	0.0104	−0.0098	−0.0048	−0.0648	0.0056
				SD	0.0075	0.0399	0.0451	0.0384	0.0604	0.0128
				RMSE	0.0162	0.0539	0.0630	0.0512	0.1023	0.0190
				CP	0.5720	0.9490	0.9330	0.9500	0.7670	0.8850
(2)	20	25	θ_0^b	Bias	−0.0137	−0.0003	−0.0106	0.0020	−0.0664	0.0017
				SD	0.0073	0.0487	0.0446	0.0473	0.0591	0.0138
				RMSE	0.0163	0.0648	0.0626	0.0630	0.1018	0.0198
				CP	0.5440	0.9530	0.9360	0.9500	0.7610	0.9080
(3)	20	49	θ_0^a	Bias	−0.0127	0.0097	−0.0062	−0.0039	−0.0588	0.0058
				SD	0.0052	0.0293	0.0322	0.0283	0.0434	0.0089
				RMSE	0.0139	0.0408	0.0446	0.0389	0.0804	0.0137
				CP	0.3320	0.9350	0.9310	0.9410	0.6940	0.8590
(4)	20	49	θ_0^b	Bias	−0.0127	0.0006	−0.0069	0.0009	−0.0601	0.0015
				SD	0.0051	0.0356	0.0319	0.0347	0.0425	0.0097
				RMSE	0.0139	0.0486	0.0442	0.0477	0.0805	0.0137
				CP	0.3040	0.9430	0.9280	0.9390	0.6800	0.9200
(5)	20	196	θ_0^a	Bias	−0.0121	0.0079	−0.0065	0.0022	−0.0552	0.0057
				SD	0.0026	0.0151	0.0161	0.0146	0.0218	0.0044
				RMSE	0.0124	0.0224	0.0231	0.0210	0.0603	0.0081
				CP	0.0040	0.8940	0.9220	0.9240	0.2790	0.7200
(6)	20	196	θ_0^b	Bias	−0.0121	0.0006	−0.0069	0.0015	−0.0560	0.0010
				SD	0.0025	0.0183	0.0160	0.0179	0.0213	0.0048
				RMSE	0.0124	0.0256	0.0230	0.0253	0.0607	0.0068
				CP	0.0040	0.9350	0.9180	0.9260	0.2640	0.9250
(7)	50	25	θ_0^a	Bias	−0.0024	0.0036	0.0004	−0.0028	−0.0240	0.0008
				SD	0.0020	0.0243	0.0286	0.0242	0.0398	0.0034
				RMSE	0.0035	0.0328	0.0389	0.0325	0.0590	0.0048
				CP	0.7960	0.9560	0.9440	0.9550	0.8790	0.9340
(8)	50	25	θ_0^b	Bias	−0.0024	−0.0018	−0.0001	0.0018	−0.0251	0.0000
				SD	0.0020	0.0300	0.0283	0.0299	0.0390	0.0037
				RMSE	0.0035	0.0404	0.0385	0.0402	0.0585	0.0052
				CP	0.7750	0.9540	0.9480	0.9580	0.8700	0.9300
(9)	50	49	θ_0^a	Bias	−0.0023	0.0043	−0.0011	−0.0033	−0.0217	0.0010
				SD	0.0014	0.0180	0.0204	0.0179	0.0285	0.0024
				RMSE	0.0029	0.0248	0.0281	0.0245	0.0437	0.0035
				CP	0.6240	0.9520	0.9340	0.9490	0.8700	0.9140
(10)	50	49	θ_0^b	Bias	−0.0024	−0.0002	−0.0016	0.0004	−0.0228	0.0002
				SD	0.0014	0.0220	0.0202	0.0219	0.0279	0.0026
				RMSE	0.0029	0.0295	0.0278	0.0294	0.0435	0.0036
				CP	0.5930	0.9540	0.9350	0.9510	0.8550	0.9390
(11)	50	196	θ_0^a	Bias	−0.0022	0.0030	−0.0009	−0.0019	−0.0211	0.0011
				SD	0.0007	0.0093	0.0102	0.0092	0.0143	0.0012
				RMSE	0.0023	0.0136	0.0141	0.0134	0.0274	0.0019
				CP	0.0910	0.9170	0.9460	0.9340	0.6620	0.8550
(12)	50	196	θ_0^b	Bias	−0.0022	−0.0010	−0.0012	0.0012	−0.0218	0.0003
				SD	0.0007	0.0113	0.0101	0.0113	0.0140	0.0013
				RMSE	0.0023	0.0153	0.0139	0.0152	0.0276	0.0018
				CP	0.0780	0.9480	0.9440	0.9530	0.6350	0.9490

Note: $\theta_0^a = (1, -0.5, 1, 0.5, 1)$ and $\theta_0^b = (1, -0.1, 1, 0.1, 1)$.

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

TABLE 2. Performance of estimators after bias correction

	T	n	θ_0		γ	ρ	β	λ	σ^2	$\rho+\lambda$
(1)	20	25	θ_0^a	Bias	−0.0001	0.0045	−0.0029	−0.0047	−0.0180	−0.0002
				SD	0.0077	0.0402	0.0462	0.0386	0.0634	0.0131
				RMSE	0.0104	0.0533	0.0634	0.0515	0.0885	0.0179
				CP	0.9430	0.9490	0.9450	0.9470	0.9030	0.9410
(2)	20	25	θ_0^b	Bias	−0.0001	−0.0017	−0.0037	0.0018	−0.0198	0.0001
				SD	0.0075	0.0491	0.0457	0.0476	0.0621	0.0142
				RMSE	0.0102	0.0647	0.0629	0.0632	0.0872	0.0194
				CP	0.9410	0.9630	0.9440	0.9480	0.8950	0.9410
(3)	20	49	θ_0^a	Bias	0.0000	0.0038	0.0001	−0.0039	−0.0118	−0.0001
				SD	0.0053	0.0295	0.0330	0.0284	0.0456	0.0091
				RMSE	0.0072	0.0400	0.0448	0.0391	0.0637	0.0124
				CP	0.9460	0.9470	0.9380	0.9410	0.9130	0.9440
(4)	20	49	θ_0^b	Bias	0.0000	−0.0007	−0.0005	0.0009	−0.0131	0.0002
				SD	0.0052	0.0358	0.0327	0.0349	0.0446	0.0099
				RMSE	0.0070	0.0487	0.0444	0.0479	0.0628	0.0134
				CP	0.9500	0.9470	0.9400	0.9390	0.9080	0.9390
(5)	20	196	θ_0^a	Bias	0.0000	0.0020	−0.0004	−0.0022	−0.0080	−0.0002
				SD	0.0026	0.0152	0.0165	0.0147	0.0228	0.0045
				RMSE	0.0035	0.0215	0.0227	0.0211	0.0316	0.0060
				CP	0.9470	0.9320	0.9400	0.9250	0.9200	0.9520
(6)	20	196	θ_0^b	Bias	−0.0000	−0.0017	−0.0008	0.0015	−0.0088	−0.0001
				SD	0.0025	0.0184	0.0164	0.0180	0.0224	0.0049
				RMSE	0.0035	0.0257	0.0225	0.0254	0.0312	0.0066
				CP	0.9550	0.9410	0.9360	0.9230	0.9110	0.9460
(7)	50	25	θ_0^a	Bias	0.0001	0.0026	0.0017	−0.0028	−0.0045	−0.0002
				SD	0.0020	0.0244	0.0288	0.0242	0.0406	0.0035
				RMSE	0.0028	0.0328	0.0392	0.0325	0.0560	0.0047
				CP	0.9500	0.9580	0.9470	0.9560	0.9340	0.9420
(8)	50	25	θ_0^b	Bias	0.0001	−0.0020	0.0011	0.0018	−0.0056	−0.0001
				SD	0.0020	0.0301	0.0285	0.0300	0.0398	0.0038
				RMSE	0.0027	0.0405	0.0387	0.0403	0.0552	0.0052
				CP	0.9450	0.9550	0.9480	0.9580	0.9330	0.9400
(9)	50	49	θ_0^a	Bias	0.0000	0.0032	0.0001	−0.0033	−0.0021	−0.0000
				SD	0.0014	0.0180	0.0206	0.0179	0.0291	0.0024
				RMSE	0.0019	0.0247	0.0282	0.0246	0.0399	0.0033
				CP	0.9590	0.9510	0.9370	0.9510	0.9320	0.9460
(10)	50	49	θ_0^b	Bias	0.0000	−0.0004	−0.0005	0.0004	−0.0032	0.0000
				SD	0.0014	0.0221	0.0204	0.0220	0.0285	0.0026
				RMSE	0.0019	0.0296	0.0279	0.0295	0.0392	0.0035
				CP	0.9600	0.9530	0.9350	0.9500	0.9270	0.9480
(11)	50	196	θ_0^a	Bias	−0.0000	0.0019	0.0003	−0.0019	−0.0016	0.0000
				SD	0.0007	0.0093	0.0103	0.0093	0.0145	0.0012
				RMSE	0.0009	0.0134	0.0141	0.0134	0.0198	0.0016
				CP	0.9580	0.9320	0.9430	0.9340	0.9420	0.9580
(12)	50	196	θ_0^b	Bias	−0.0000	−0.0012	−0.0000	0.0012	−0.0022	0.0000
				SD	0.0007	0.0114	0.0102	0.0113	0.0143	0.0013
				RMSE	0.0009	0.0153	0.0140	0.0153	0.0194	0.0018
				CP	0.9590	0.9480	0.9490	0.9510	0.9450	0.9510

Note: $\theta_0^a = (1, -0.5, 1, 0.5, 1)$ and $\theta_0^b = (1, -0.1, 1, 0.1, 1)$.

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

TABLE 3. Performance of estimators under misspecifications of disturbances

T	n	θ_0		γ	ρ	β	λ	σ^2	$\rho+\lambda$
Misspecification 1: serial correlation									
(1) 20	25	θ_0^a	Bias	−0.0132	0.0104	−0.0098	−0.0049	−0.0509	0.0055
(before bias correction)			SD	0.0074	0.0400	0.0454	0.0385	0.0613	0.0126
			RMSE	0.0162	0.0547	0.0635	0.0521	0.0981	0.0199
			CP	0.5650	0.9500	0.9370	0.9480	0.8140	0.8360
(1′) 20	25	θ_0^a	Bias	−0.0001	0.0047	−0.0031	−0.0049	−0.0034	−0.0001
(after bias correction)			SD	0.0076	0.0403	0.0465	0.0387	0.0644	0.0129
			RMSE	0.0111	0.0542	0.0639	0.0523	0.0902	0.0191
			CP	0.8890	0.9530	0.9440	0.9470	0.9160	0.8940
Misspecification 2: spatial error									
(2) 20	25	θ_0^a	Bias	−0.0133	−0.0888	−0.0259	0.0931	−0.0812	0.0043
(before bias correction)			SD	0.0074	0.0359	0.0448	0.0341	0.0597	0.0126
			RMSE	0.0160	0.0977	0.0660	0.100.	0.1114	0.0184
			CP	0.5780	0.2950	0.8920	0.2180	0.6980	0.9020
(2′) 20	25	θ_0^a	Bias	−0.0001	−0.0935	−0.0191	0.0933	−0.0354	−0.0002
(after bias correction)			SD	0.0076	0.0361	0.0459	0.0343	0.0627	0.0129
			RMSE	0.0103	0.1020	0.0651	0.1009	0.0919	0.0175
			CP	0.9500	0.2400	0.9200	0.2240	0.8610	0.9430
Misspecification 3: spatial error (restricted model w/o W_nX_{nt})									
(3) 20	25	θ_0^a	Bias	−0.0152	0.1323	−0.0105	−0.1360	0.0203	−0.0037
(before bias correction)			SD	0.0076	0.0450	0.0469	0.0433	0.0653	0.0141
			RMSE	0.0176	0.1408	0.0660	0.1435	0.0904	0.0207
			CP	0.5020	0.1630	0.9300	0.1160	0.9430	0.8980
(3′) 20	25	θ_0^a	Bias	−0.0003	0.1355	−0.0030	−0.1356	0.0712	−0.0002
(after bias correction)			SD	0.0077	0.0453	0.0480	0.0436	0.0686	0.0145
			RMSE	0.0105	0.1438	0.0663	0.1432	0.1122	0.0199
			CP	0.9450	0.1400	0.9370	0.1140	0.8350	0.9310
Misspecification 4: spatial error (restricted model with W_nX_{nt})									
(4) 20	25	θ_0^a	Bias	−0.0143	0.0031	−0.0100	−0.0091	−0.0669	−0.0059
(before bias correction)			SD	0.0072	0.0491	0.0447	0.0471	0.0608	0.0136
			RMSE	0.0166	0.0656	0.0627	0.0634	0.1040	0.0201
			CP	0.5090	0.9580	0.9360	0.9520	0.7610	0.8970
(4′) 20	25	θ_0^a	Bias	−0.0002	0.0088	−0.0029	−0.0002	0.0088	−0.0029
(after bias correction)			SD	0.0074	0.0492	0.0458	0.0074	0.0492	0.0458
			RMSE	0.0101	0.0654	0.0631	0.0101	0.0654	0.0631
			CP	0.9420	0.9630	0.9430	0.9420	0.9630	0.9430

Note: $\theta_0^a = (1, -0.5, 1, 0.5, 1)$.

restricted model where $\gamma_0 = 1$ and $\lambda_0 = 0.5$ in (4), we can see from Misspecifications 3 and 4 that adding the regressor W_nX_{nt} will resolve the misspecification problem.[13]

To check the consequences of misspecification of the spatial weights matrix, we also conduct a small Monte Carlo study. The true weights matrix is the rook matrix we have used in Tables 1 and 2. We have two misspecifications of the weights

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

TABLE 4. Performance of estimators under misspecifications of weights matrix

T	n	θ_0		γ	ρ	β	λ	σ^2	$\rho+\lambda$
Misspecification 1: underspecification									
(1) 20	25	θ_0^a	Bias	−0.0144	0.1976	0.0235	−0.1935	0.0852	0.0041
(before bias correction)			SD	0.0076	0.0352	0.0484	0.0340	0.0697	0.0106
			RMSE	0.0173	0.2008	0.0705	0.1963	0.1235	0.0160
			CP	0.5230	0	0.9170	0	0.7790	0.8850
(1′) 20	25	θ_0^a	Bias	−0.0002	0.1937	0.0309	−0.1934	0.1395	0.0003
(after bias correction)			SD	0.0078	0.0354	0.0496	0.0342	0.0732	0.0109
			RMSE	0.0110	0.1970	0.0735	0.1965	0.1635	0.0151
			CP	0.9060	0	0.8890	0	0.4950	0.9270
average likelihood: -568.3 vs. -538.1 under true, none of them greater than the true ones									
Misspecification 2: overspecification									
(2) 20	25	θ_0^a	Bias	−0.0138	0.0087	0.0176	−0.0042	0.0358	0.0044
(before bias correction)			SD	0.0074	0.0479	0.0472	0.0461	0.0663	0.0152
			RMSE	0.0165	0.0657	0.0674	0.0631	0.0971	0.0226
			CP	0.5390	0.9470	0.9220	0.9520	0.9120	0.8940
(2′) 20	25	θ_0^a	Bias	−0.0001	0.0039	0.0348	−0.0039	0.0875	−0.0000
(after bias correction)			SD	0.0076	0.0482	0.0484	0.0463	0.0696	0.0156
			RMSE	0.0105	0.0655	0.0699	0.0634	0.1237	0.0214
			CP	0.9260	0.9520	0.9030	0.9510	0.7430	0.9390
average likelihood: −555.7 vs. -538.1 under true, 0.1% of them greater than the true ones									

Note: $\theta_0^a = (1, -0.5, 1, 0.5, 1)$.

matrix: One is the underspecification with a circle matrix, where each unit has two neighbors; the other is the overspecification with a queen matrix.[14] To avoid unnecessary tables, we report the results for $n = 25$ and $T = 20$ in Table 4. By comparing the results with Tables 1 and 2, we can see that both misspecifications have poorer performances compared to the correct specification. However, underspecification produces much more bias than overspecification of the weights matrix. The comparison of maximized log likelihood values provides a valuable model selection criterion for the models.

6. CONCLUSION

In this paper we derive the asymptotic properties of QMLEs for the unit root dynamic panel data model with spatially correlated disturbances and the unit root SDPD model. We also propose a bias-correction procedure for both cases so that the QMLEs are properly centered normal. For both models, when the exogenous variables are relevant or fixed effects are presented in the DGP, the estimate of the dynamic coefficient is $\sqrt{nT^3}$ consistent, the estimates of other parameters are $\sqrt{nT}$ consistent, and all are asymptotically normal. For the unit root SDPD model, the sum of the contemporaneous spatial effect and dynamic spatial effect

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

TABLE 5. Summary of different SDPD models

	Stable	Spatial cointegration	Unit root
Parameters	$\gamma_0+\lambda_0+\rho_0<1$	$\gamma_0+\lambda_0+\rho_0=1$ and $\gamma_0\neq 1$	$\gamma_0+\lambda_0+\rho_0=1$ and $\gamma_0=1$
Eigenvalues of A_n	All less than 1 in absolute value	Some unit roots	All unit roots
Dynamics	Stable	Cointegration Cointegrating matrix I_n-W_n	No cointegration
Consistency rates	$\sqrt{nT}$ for $\hat{\gamma}_{nT},\hat{\lambda}_{nT},\hat{\rho}_{nT}$	$\sqrt{nT}$ for $\hat{\gamma}_{nT},\hat{\lambda}_{nT},\hat{\rho}_{nT}$, $\sqrt{nT^3}$ for $\hat{\gamma}_{nT}+\hat{\lambda}_{nT}+\hat{\rho}_{nT}$	$\sqrt{nT}$ for $\hat{\lambda}_{nT},\hat{\rho}_{nT}$ $\sqrt{nT^3}$ for $\hat{\gamma}_{nT}$ and $\hat{\lambda}_{nT}+\hat{\rho}_{nT}$
Information matrix	Nonsingular	Singular	Nonsingular after rescaling
Bias magnitude	$O(\frac{1}{T})$ for $\hat{\gamma}_{nT},\hat{\lambda}_{nT},\hat{\rho}_{nT}$	$O(\frac{1}{T})$ for $\hat{\gamma}_{nT},\hat{\lambda}_{nT},\hat{\rho}_{nT}$	$O(\frac{1}{T^2})$ for $\hat{\gamma}_{nT}$ $O(\frac{1}{T})$ for $\hat{\lambda}_{nT},\hat{\rho}_{nT}$

will converge at $\sqrt{nT^3}$ rate. In Table 5 we compare the unit root SDPD model with the spatial cointegration case and the stable case in Yu et al. (2007, 2008).

From the comparison in Table 5, we see that the various cases imply some different asymptotics of the estimates of parameters. These differences are due to different asymptotic behaviors of the models. The unit eigenvalues of the unit root SDPD model are not linked to the eigenvalues of the spatial weights matrix. On the contrary, for the spatial cointegration model, the unit eigenvalues correspond exactly to those unit eigenvalues of the spatial weights matrix via a well defined relation. For the unit roots SDPD model, the outcomes of the different spatial units do not show comovements. However, one can show that for the spatial cointegration model, the outcomes of different spatial units can be cointegrated with a reduced rank, where the rank is the number of eigenvalues of W_n different from 1.

Because of the different asymptotics and implications of the unit root SDPD model and the spatial cointegration model, model specification tests for unit roots will be of special interest. Unit root tests in time series as well as in panels have a large body of literature (for a recent survey, see Choi, 2006). The literature on unit root tests covers panels with cross-section dependence generated by unobserved common time variables or factors (e.g., Moon and Perron, 2004). As the spatial interactions in this paper generate cross-section correlation of a different nature, it would be of interest to understand the asymptotic properties of unit root tests in the presence of spatial correlation. See Baltagi et al. (2007) for some Monte Carlo evidence. For either the unit root panel model or the unit root SDPD model in this paper, the Wald test for $\gamma_0=1$ is a unit root test.

The panel model investigated in this paper is on spatial and dynamic autoregressive structures, where the regressors are assumed to be stationary or exogenous if present. It would be of interest to consider cases where regressors are

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

unit root processes and the panel has cointegrating relations. Such a model will generalize the panel cointegration model considered in Kao and Chiang (2000), Pedroni (2000), and Phillips and Moon (1999) with spatial correlation and dynamics. The implication of spatial correlation on the estimation of cointegration relations will be of interest. These issues will be investigated in future research.

NOTES

1. See, e.g., Baltagi, Song, and Koh (2003), Korniotis (2005), Kapoor, Kelejian, and Prucha (2007), Baltagi, Song, Jung, and Koh (2007), and Yu, de Jong, and Lee (2007, 2008).

2. In Yu and Lee (2009) we apply the SDPD model to study a regional growth convergence model with U.S. state data. The 48 contiguous states are spatially correlated due to technological spillover and factor mobility; with panel data available for the personal state income, we can study the convergence rate taking into account the spatial correlation.

3. W_n is row-normalized if $W_n l_n = l_n$, where l_n is an n-dimensional vector of ones. This feature of W_n is useful to characterize the nonstationarity in Yu et al. (2007).

4. An alternative to the maximum likelihood (ML) or the QML method is the generalized method of moments (GMM). With a finite and fixed number of moments, the asymptotic properties of a GMM estimator can be derived without additional difficulty given the basic asymptotics in this paper. However, as in a dynamic panel data model, many instrumental variables (IVs) and moments may exist as T increases. With many IVs, the asymptotic properties of such a GMM estimation need to be investigated. The GMM approach is of particular interest when T is much smaller than n, under which the asymptotic results in this paper might not hold. In our preliminary investigation of the GMM estimation for the SDPD, there is Monte Carlo evidence that the MLE and some carefully designed GMM estimates are compatible, but the MLE can be better than other GMM estimators. In the implementation, the ML or the QML method is much more straightforward. MLE with bias correction has been shown to be efficient for the dynamic panel data model in Hahn and Kuersteiner (2002).

5. We say that the row and column sums of a (sequence of $n \times n$) matrix $P_n = [p_{ij,n}]$ are uniformly bounded in n if $\sup_{1\le i\le n, n\ge 1} \sum_{j=1}^{n} |p_{ij,n}| < \infty$ and $\sup_{1\le j\le n, n\ge 1} \sum_{i=1}^{n} |p_{ij,n}| < \infty$.

6. Assumption 1 is slightly less restrictive than Assumption 1 in Yu et al. (2007) because the row normalization requirement is not needed for analysis.

7. The problem of serial correlation in disturbances is of interest. But as in the usual panel model with or without dynamic, the estimation of models with serial correlation needs special treatment. For example, the ML method becomes complicated (Anderson and Hsiao, 1982). We leave such an extended model for future research. We expect that MLE will still be consistent due to a deterministic trend component in the DGP. This is also true in time series where serial correlation will cause inconsistency in the estimate of a stationary AR(1) parameter while the parameter estimate may still be consistent for a nonstationary AR(1) model. See Table 3 for Monte Carlo results on possible consequences of this misspecification of the disturbances.

8. In addition, there are also constraints on the coefficients across the regressors X_{nt} and $W_n X_{nt}$ in the transformed equation.

9. From the DGP (4) without X_{nt}, we have $\tilde{Y}_{nt} = \tilde{\iota}\mathbf{c}_{n0} + S_n^{-1}\tilde{\xi}_{nt}$. Hence, $\lim_{T\to\infty} \frac{1}{nT^3} \sum_{t=1}^{T} \tilde{\tilde{Y}}'_{n,t-1} S'_n S_n \tilde{\tilde{Y}}_{n,t-1}$ is equal to $\frac{1}{12n} \mathbf{c}'_{n0} S'_n S_n \mathbf{c}_{n0}$ because $\lim_{T\to\infty} \frac{1}{nT^3} \sum_{t=1}^{T} \tilde{\xi}'_{n,t-1} \tilde{\xi}_{n,t-1} = 0$ from Lemma 8 in Yu et al. (2007). When $\lambda_0 = 0$, $\frac{1}{n} \mathbf{c}'_{n0} S'_n S_n \mathbf{c}_{n0}$ is reduced to $\frac{1}{n} \mathbf{c}'_{n0} \mathbf{c}_{n0}$, which is $\frac{1}{n} \sum_{i=1}^{n} \alpha_i^2$ in Hahn and Kuersteiner (2002). However, when $\lambda_0 \neq 0$, the bias term is not equal to that in Hahn and Kuersteiner in general.

10. In this section some notations are similar to those in Section 3 but might have different meanings.

11. A rook matrix presents a square tessellation with a connectivity of four for the inner fields on the chessboard and two and three for the corner and border fields, respectively.

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

12. As the Wald test for $\gamma_0 = 1$ can be regarded as a test of unit root in our model, the accuracy of the CP for the bias-corrected estimator indicates that such a unit root test has an accurate size.

13. For this restricted model $Y_{nt} = Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + V_{nt}$ where $V_{nt} = \psi_0 W_n V_{nt} + \varepsilon_{nt}$ with elements of ε_{nt} being i.i.d., the transformed equation is $Y_{nt} = \psi_0 W_n Y_{nt} + Y_{nt} - \psi_0 W_n Y_{n,t-1} + X_{nt}\beta_0 - \psi_0 W_n X_{nt}\beta_0 + \ddot{\mathbf{c}}_{n0} + \varepsilon_{nt}$, where $\ddot{\mathbf{c}}_{n0} = (I_n - \psi_0 W_n)\mathbf{c}_{n0}$ is the transformed fixed effects. We note that while the sample observations are generated by this restricted model, the estimates in Misspecifications 3 and 4 of Table 3 are based on the likelihood function of the unrestricted model.

14. A queen matrix presents a square tessellation with a connectivity of eight for the inner fields on the chessboard and three and five for the corner and border fields, respectively.

REFERENCES

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

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

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

Choi, I. (2006) Nonstationary panels. In T.C. Mills & K. Patterson (eds.), *Palgrave Handbook of Econometrics*, vol. 1, Ch. 13, Econometric Theory, pp. 511–539. Palgrave Macmillan.

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

Kao, C. & M.H. Chiang (2000) On the estimation and inference of a cointegrated regression in panel data. In B.H. Baltagi (ed.), *Advances in Econometrics* 15, pp. 179–222. JAI.

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

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

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

Korniotis, G.M. (2005) A Dynamic Panel Estimator with Both Fixed and Spatial Effects. Working paper, University of Notre Dame.

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

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

Moon, H.R. & B. Perron (2004) Testing for a unit root in panels with dynamic factors. *Journal of Econometrics* 122, 81–126.

Pedroni, P. (2000) Fully modified OLS for heterogeneous cointegrated panels and the case of purchasing power parity. In B.H. Baltagi (ed.), *Advances in Econometrics* 15, pp. 93–130. JAI.

Phillips, P.C.B. & H. Moon (1999) Linear regression limit theory for nonstationary panel data. *Econometrica* 67, 1057–1112.

Yu, J., R. de Jong, & L.F. Lee (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, Ohio State University.

Yu, J., R. de Jong, & L.F. Lee (2008) Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics* 146, 118–134.

Yu, J. & L.F. Lee (2009) Convergence: A Spatial Dynamic Panel Data Approach. Working paper, Ohio State University.

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

APPENDIX A: Algebra for Section 3

A.1. First and Second Conditions for (6). For the concentrated likelihood function (6), the score and the information matrix are

$$\frac{\partial \ln L_{n,T}(\theta_0)}{\partial \theta} = \begin{pmatrix} \frac{\partial \ln L_{n,T}(\theta_0)}{\partial \delta} \\ \frac{\partial \ln L_{n,T}(\theta_0)}{\partial \lambda} \\ \frac{\partial \ln L_{n,T}(\theta_0)}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_0^2}\sum_{t=1}^T \tilde{Z}_{nt}' S_n' \tilde{V}_{nt} \\ \frac{1}{\sigma_0^2}\sum_{t=1}^T (\tilde{V}_{nt}' G_n' \tilde{V}_{nt} - \sigma_0^2 \mathrm{tr} G_n) \\ \frac{1}{2\sigma_0^4}\sum_{t=1}^T (\tilde{V}_{nt}' \tilde{V}_{nt} - n\sigma_0^2) \end{pmatrix}, \tag{A.1}$$

and

$$-\mathrm{E}\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial \theta \partial \theta'} = \begin{pmatrix} \frac{1}{\sigma_0^2}\mathrm{E}\mathcal{H}_{nT} & \mathbf{0}_{(k_x+1)\times 1} & \mathbf{0}_{(k_x+1)\times 1} \\ \mathbf{0}_{1\times(k_x+1)} & \frac{1}{n}\left[\mathrm{tr}(G_n' G_n) + \mathrm{tr}(G_n^2)\right] & \frac{1}{\sigma_0^2 n}\mathrm{tr}(G_n) \\ \mathbf{0}_{1\times(k_x+1)} & \frac{1}{\sigma_0^2 n}\mathrm{tr}(G_n) & \frac{1}{2\sigma_0^4} \end{pmatrix} \tag{A.2}$$

$$-\mathrm{E}\begin{pmatrix} \mathbf{0}_{(k_x+1)\times(k_x+1)} & \frac{1}{\sigma_0^2 nT}\sum_{t=1}^T \tilde{Z}_{nt}'[S_n' G_n + W_n']\tilde{V}_{nt} & \frac{1}{\sigma_0^4 nT}\sum_{t=1}^T \tilde{Z}_{nt}' S_n' \tilde{V}_{nt} \\ * & 0 & \frac{1}{\sigma_0^2 nT}\mathrm{tr}(G_n) \\ * & * & \frac{1}{T\sigma_0^4} \end{pmatrix},$$

where $\mathcal{H}_{nT}$ is in (7). Therefore, the elements of $-\mathrm{E}\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial \theta \partial \theta'}$ have different orders as $\tilde{Z}_{nt}$ has the element $\tilde{\tilde{Y}}_{n,t-1}$. After rescaling, using Lemmas 8 and 9 in Yu et al. (2007), we have $\Upsilon_{\theta,T}^{-1}\left(-\mathrm{E}\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial \theta \partial \theta'}\right)\Upsilon_{\theta,T}^{-1} = \Sigma_{\theta_0,nT} + O(T^{-1})$, where $\Sigma_{\theta_0,nT}$ is in (12). We can decompose $\tilde{Z}_{nt}$ as

$$\tilde{Z}_{nt} = Z_{nt}^* - (S_n^{-1}\bar{\xi}_{nT,-1},\ \mathbf{0}_{n\times k_x}), \tag{A.3}$$

where $Z_{nt}^* = (Y_{n,t-1}^*, \tilde{X}_{nt})$ and $Y_{n,t-1}^* = \tilde{t}_{-1}\mathbf{c}_n + \tilde{\tilde{\mathbb{X}}}_{n,t-1}\beta_0 + S_n^{-1}\xi_{n,t-1}$ with $\tilde{t}_{-1} = t - 1 - \frac{T+1}{2}$. Hence, $\tilde{Z}_{nt}$ is decomposed into two parts: Z_{nt}^*, uncorrelated with V_{nt}, and $(S_n^{-1}\bar{\xi}_{nT,-1},\ \mathbf{0}_{n\times k_x})$, correlated with V_{nt}. Correspondingly,

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial \theta} = \frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^*(\theta_0)}{\partial \theta} - \Delta_{\theta_0,nT}, \tag{A.4}$$

where

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^*(\theta_0)}{\partial \theta} = \begin{pmatrix} \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^T Z_{nt}^{*\prime} S_n' V_{nt} \\ \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^T (V_{nt}' G_n' V_{nt} - \sigma_0^2 \mathrm{tr} G_n) \\ \frac{1}{2\sigma_0^4}\frac{1}{\sqrt{nT}}\sum_{t=1}^T (V_{nt}' V_{nt} - n\sigma_0^2) \end{pmatrix}, \tag{A.5}$$

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

and

$$\Delta_{\theta_0,nT} = \begin{pmatrix} \frac{1}{\sigma_0^2}\sqrt{\frac{T}{n}}(\bar{\xi}_{nT,-1},\ \mathbf{0}_{n\times k_x})'\bar{V}_{nT} \\ \frac{1}{\sigma_0^2}\sqrt{\frac{T}{n}}\bar{V}_{nT}'G_n'\bar{V}_{nT} \\ \frac{1}{2\sigma_0^4}\sqrt{\frac{T}{n}}\bar{V}_{nT}'\bar{V}_{nT} \end{pmatrix}. \tag{A.6}$$

Similar to Yu et al. (2008), the variance matrix of $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^*(\theta_0)}{\partial\theta}$ is

$$\mathrm{E}\left(\Upsilon_{\theta,T}^{-1}\frac{1}{nT}\frac{\partial \ln L_{n,T}^*(\theta_0)}{\partial\theta}\cdot\frac{\partial \ln L_{n,T}^*(\theta_0)}{\partial\theta'}\Upsilon_{\theta,T}^{-1}\right) = \Sigma_{\theta_0,nT} + \Omega_{\theta_0,n} + O\left(T^{-1}\right). \tag{A.7}$$

When V_{nt} are normally distributed, $\Omega_{\theta_0,n} = \mathbf{0}$ because $\mu_4 - 3\sigma_0^4 = 0$ for a normal distribution.

A.2. About $\hat{\sigma}_{nT}^2(\lambda)$ for (10). From (9), we have

$$\hat{\sigma}_{nT}^2(\lambda) = \frac{1}{nT}\sum_{t=1}^{T}\tilde{V}_{nt}'S_n'^{-1}S_n'(\lambda)S_n(\lambda)S_n^{-1}\tilde{V}_{nt} + R_{1,nT}(\lambda), \tag{A.8}$$

$$\frac{\partial\hat{\sigma}_{nT}^2(\lambda)}{\partial\lambda} = -\frac{2}{nT}\sum_{t=1}^{T}\tilde{V}_{nt}'G_n'S_n(\lambda)S_n^{-1}\tilde{V}_{nt} + \frac{\partial R_{1,nT}(\lambda)}{\partial\lambda}, \tag{A.9}$$

$$\frac{\partial^2\hat{\sigma}_{nT}^2(\lambda)}{\partial\lambda^2} = \frac{2}{nT}\sum_{t=1}^{T}\tilde{V}_{nt}'G_n'G_n\tilde{V}_{nt} + \frac{\partial^2 R_{1,nT}(\lambda)}{\partial\lambda^2}, \tag{A.10}$$

where

$$\begin{aligned} R_{1,nT}(\lambda) &= -\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}'S_n'(\lambda)S_n(\lambda)S_n^{-1}\tilde{V}_{nt}\right)'\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}'S_n'(\lambda)S_n(\lambda)\tilde{Z}_{nt}\right)^{-1} \\ &\quad\times\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}'S_n'(\lambda)S_n(\lambda)S_n^{-1}\tilde{V}_{nt}\right) \\ &= -\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}^{s\prime}S_n'(\lambda)S_n(\lambda)S_n^{-1}\tilde{V}_{nt}\right)'\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}^{s\prime}S_n'(\lambda)S_n(\lambda)\tilde{Z}_{nt}^{s}\right)^{-1} \\ &\quad\times\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}^{s\prime}S_n'(\lambda)S_n(\lambda)S_n^{-1}\tilde{V}_{nt}\right), \end{aligned}$$

because $\tilde{Z}_{nt}^{s} = \tilde{Z}_{nt}\Upsilon_{\delta,T}^{-1}$. Using Lemmas 8 and 9 in Yu et al. (2007), $R_{1,nT}(\lambda) = O_p\left(\max\left(\frac{1}{nT},\frac{1}{T^2}\right)\right)$ uniformly in λ, and so are their derivatives with respect to λ.

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

Hence, using $\frac{1}{nT}\sum_{t=1}^{T}\tilde{V}_{nt}'\mathcal{B}_n\tilde{V}_{nt}=\sigma_0^2\frac{1}{n}\mathrm{tr}\mathcal{B}_n+O_p\left(\max\left(\frac{1}{\sqrt{nT}},\frac{1}{T}\right)\right)$ for any $n\times n$ non-stochastic square matrix $\mathcal{B}_n$ from Lemma 9 in Yu et al. (2007),

$$\hat{\sigma}_{nT}^2(\lambda)=\frac{1}{nT}\sum_{t=1}^{T}\tilde{V}_{nt}'S_n'^{-1}S_n'(\lambda)S_n(\lambda)S_n^{-1}\tilde{V}_{nt}+O_p\left(\max\left(\frac{1}{nT},\frac{1}{T^2}\right)\right), \tag{A.11a}$$

$$\sqrt{nT}\frac{\partial\hat{\sigma}_{nT}^2(\lambda_0)}{\partial\lambda}=-\frac{2}{\sqrt{nT}}\sum_{t=1}^{T}\tilde{V}_{nt}'G_n'\tilde{V}_{nt}+O_p\left(\max\left(\frac{1}{\sqrt{nT}},\sqrt{\frac{n}{T^3}}\right)\right), \tag{A.11b}$$

$$\frac{\partial^2\hat{\sigma}_{nT}^2(\lambda)}{\partial\lambda^2}=\frac{2\sigma_0^2}{n}\mathrm{tr}(G_n'G_n)+O_p\left(\max\left(\frac{1}{\sqrt{nT}},\frac{1}{T}\right)\right), \tag{A.11c}$$

where the orders of the residual terms are all uniformly in λ.

APPENDIX B: Proofs for Theorems in Section 3

Proof of Theorem 1.

Consistency. The concentrated log likelihood for the restricted model is $\ln L_{n,T}(\lambda)$ in (10). This log likelihood function can be compared with that of a simple pure spatial dynamic process below. It is known that, for the log likelihood for pure spatial dynamic process $Y_{nt}=\lambda_0W_nY_{nt}+V_{nt}$ for $t=1,\ldots,T$,

$$\ln\check{L}_{n,T}(\lambda)=-\frac{n}{2}(\ln 2\pi+1)-\frac{n}{2}\ln\check{\sigma}_{nT}^2(\lambda)+T\ln|S_n(\lambda)|,$$

$$\check{\sigma}_{nT}^2(\lambda)=\frac{1}{nT}\sum_{t=1}^{T}\tilde{V}_{nt}'S_n'^{-1}S_n'(\lambda)S_n(\lambda)S_n^{-1}\tilde{V}_{nt}.$$

Under $\lim_{n\to\infty}(\frac{1}{n}\ln|\sigma_0^2S_n^{-1\prime}S_n^{-1}|-\frac{1}{n}\ln|\sigma_n^2(\lambda)S_n^{-1}(\lambda)'S_n^{-1}(\lambda)|)\neq 0$ for $\lambda\neq\lambda_0$, the maximizer $\check{\lambda}_{nT}$ is globally identified and consistent (Thm. 2 in Yu et al., 2008).

From $\hat{\sigma}_{nT}^2(\lambda)$ in (A.11a), $\frac{1}{nT}\ln L_{n,T}(\lambda)-\frac{1}{nT}\ln\check{L}_{n,T}(\lambda)=o_p(1)$ uniformly in λ. By using Lemma A.6 in Lee (2007), the maximizer $\hat{\lambda}_{nT}$ of $\frac{1}{nT}\ln L_{n,T}(\lambda)$ will also be consistent. Hence, $\hat{\lambda}_{nT}-\lambda_0\xrightarrow{p}0$. Using (A.11a), we have $\hat{\sigma}_{nT}^2(\hat{\lambda}_{nT})-\sigma_0^2\xrightarrow{p}0$. Also, from (8), using $Z_{nt}^s=Z_{nt}\Upsilon_{\delta,T}^{-1}$ where $\Upsilon_{\delta,T}=\mathrm{diag}(T,l_{k_x}')$, we have

$$\Upsilon_{\delta,T}[\hat{\delta}_{nT}(\hat{\lambda}_{nT})-\delta_0]=\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}^{s\prime}S_n'(\hat{\lambda}_{nT})S_n(\hat{\lambda}_{nT})\tilde{Z}_{nt}^{s}\right)^{-1}\times\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}^{s\prime}S_n'(\hat{\lambda}_{nT})S_n(\hat{\lambda}_{nT})S_n^{-1}\tilde{V}_{nt}\right).$$

Hence, $\hat{\delta}_{nT}(\hat{\lambda}_{nT})$ is also consistent such that $\Upsilon_{\delta,T}(\hat{\delta}_{nT}(\hat{\lambda}_{nT})-\delta_0)\xrightarrow{p}0$.

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

Asymptotic Distribution. From the Taylor expansion, for $\bar{\theta}_{nT}$ lying between $\hat{\theta}_{nT}$ and θ_0,

$$\Upsilon_{\theta,T}\sqrt{nT}(\hat{\theta}_{nT}-\theta_0)=\left(\Upsilon_{\theta,T}^{-1}\frac{1}{nT}\frac{\partial \ln L_{n,T}(\bar{\theta}_{nT})}{\partial\theta\partial\theta'}\Upsilon_{\theta,T}^{-1}\right)^{-1}\times\left(\Upsilon_{\theta,T}^{-1}\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^{*}(\theta_0)}{\partial\theta}-\Upsilon_{\theta,T}^{-1}\Delta_{\theta_0,nT}\right), \tag{B.1}$$

where $\frac{\partial \ln L_{n,T}^{*}(\theta_0)}{\partial\theta}$ and $\Delta_{\theta_0,nT}$ are in (A.5) and (A.6).

The $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^{*}(\theta_0)}{\partial\theta}$ is a combination of the linear and quadratic forms of V_{nt}. After being rescaled,

$$\Upsilon_{\theta,T}^{-1}\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta}+\Upsilon_{\theta,T}^{-1}\Delta_{\theta_0,nT}\xrightarrow{d}N(0,\lim_{T\to\infty}(\Sigma_{\theta_0,nT}+\Omega_{\theta_0,n})), \tag{B.2}$$

using the CLT in Yu et al. (2007). For $\Upsilon_{\theta,T}^{-1}\Delta_{\theta_0,nT}$, we have $\frac{1}{n}\bar{\xi}'_{nT,-1}\mathcal{B}_n\bar{V}_{nT}=\mathrm{E}\big(\frac{1}{n}\bar{\xi}'_{nT,-1}\mathcal{B}\bar{V}_{nT}\big)+O_p\big(\frac{1}{\sqrt{n}}\big)$ where $\mathrm{E}\big(\frac{1}{n}\bar{\xi}'_{nT,-1}\mathcal{B}_n\bar{V}_{nT}\big)=\sigma_0^2\frac{T-1}{2T}\frac{1}{n}\mathrm{tr}\mathcal{B}_n$ is $O(1)$ and $\frac{1}{n}\bar{V}'_{nT}\mathcal{B}_n\bar{V}_{nT}=\frac{1}{nT}\sigma_0^2\mathrm{tr}\mathcal{B}_n+O_p\big(\frac{1}{\sqrt{nT^2}}\big)$ for any $n\times n$ nonstochastic matrix $\mathcal{B}_n$. Hence, $\Upsilon_{\theta,T}^{-1}\Delta_{\theta_0,nT}=\sqrt{\frac{n}{T}}a_{\theta_0,n}+O_p\big(\frac{1}{\sqrt{T}}\big)$ where $a_{\theta_0,n}$ is in (11). Similar to Yu et al. (2008), we have $-\Upsilon_{\theta,T}^{-1}\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}(\bar{\theta}_{nT})}{\partial\theta\partial\theta'}\Upsilon_{\theta,T}^{-1}=\|\bar{\theta}_{nT}-\theta_0\|\cdot O_p(1)+O_p\big(\frac{1}{\sqrt{nT}}\big)+\Sigma_{\theta_0,nT}$. Because (i) $\|\bar{\theta}_{nT}-\theta_0\|=o_p(1)$ as $\hat{\theta}_{nT}$ is consistent and (ii) $\Sigma_{\theta_0,nT}$ is the nonsingular in the limit by Assumption 8, $-\Upsilon_{\theta,T}^{-1}\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}(\bar{\theta}_{nT})}{\partial\theta\partial\theta'}\Upsilon_{\theta,T}^{-1}$ is invertible for large T. Hence, from (B.1), $\Upsilon_{\theta,T}\sqrt{nT}(\hat{\theta}_{nT}-\theta_0)=O_p(1)\cdot\left(O_p(1)+O\left(\sqrt{\frac{n}{T}}\right)\right)$, which implies that

$$\Upsilon_{\theta,T}(\hat{\theta}_{nT}-\theta_0)=O_p\left(\max\left(\sqrt{\frac{1}{nT}},\frac{1}{T}\right)\right). \tag{B.3}$$

Hence, $\Upsilon_{\theta,T}(\bar{\theta}_{nT}-\theta_0)=O_p\left(\max\left(\sqrt{\frac{1}{nT}},\frac{1}{T}\right)\right)$, which implies that

$$\left(-\Upsilon_{\theta,T}^{-1}\frac{1}{nT}\frac{\partial^2 \ln L_{nT}(\bar{\theta}_{nT})}{\partial\theta\partial\theta'}\Upsilon_{\theta,T}^{-1}\right)^{-1}=(\Sigma_{\theta_0,nT})^{-1}+O_p\left(\max\left(\sqrt{\frac{1}{nT}},\frac{1}{T}\right)\right). \tag{B.4}$$

Consequently,

$$\Upsilon_{\theta,T}\sqrt{nT}(\hat{\theta}_{nT}-\theta_0)=\left((\Sigma_{\theta_0,nT})^{-1}+O_p\left(\max\left(\sqrt{\frac{1}{nT}},\frac{1}{T}\right)\right)\right)\cdot\left(\Upsilon_{\theta,T}^{-1}\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^{*}(\theta_0)}{\partial\theta}-\Upsilon_{\theta,T}^{-1}\Delta_{\theta_0,nT}\right),$$

which, in turn, implies the result in Theorem 1. ■

Proof for Theorem 2. As $\hat{\theta}_{nT}$ has the asymptotic distribution in Theorem 1, we will have (15) if $\sqrt{\frac{n}{T}}(-\hat{B}_{nT}-(\Sigma_{\theta_0,nT})^{-1}a_n(\theta_0))\xrightarrow{p}0$ and $\frac{n}{T^3}\to 0$. The difference

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

$(\Sigma_{\theta,nT})^{-1}|_{\theta=\hat{\theta}_{nT}} - (\Sigma_{\theta_0,nT})^{-1}$ has a similar order in (B.4). Thus we have

$$-\hat{B}_{nT} - (\Sigma_{\theta_0,nT})^{-1}a_n(\theta_0) = (\Sigma_{\theta_0,nT})^{-1}\left(a_n(\hat{\theta}_{nT}) - a_n(\theta_0)\right) + a_n(\hat{\theta}_{nT}) \times O_p\left(\max\left(\frac{1}{T}, \frac{1}{\sqrt{nT}}\right)\right).$$

From $a_n(\theta)$ in (11), $\frac{\partial a_n(\bar{\theta}_{nT})}{\partial \theta'}$ will be uniformly bounded in probability and $a_n(\hat{\theta}_{nT})$ is $O_p(1)$ because $a_n(\theta_0)$ is $O(1)$. Hence, using (B.3), $-\hat{B}_{nT} - (\Sigma_{\theta_0,nT})^{-1}a_n(\theta_0) = O_p\left(\max\left(\frac{1}{T}, \frac{1}{\sqrt{nT}}\right)\right)$. This implies that $\sqrt{\frac{n}{T}}(-\hat{B}_{nT} - (\Sigma_{\theta_0,nT})^{-1}a_n(\theta_0)) \xrightarrow{p} 0$ under $\frac{n}{T^3} \to 0$. ■

APPENDIX C: Algebra for Section 4

C.1. A Lemma. Denote

$$\mathcal{H}_{nT} = \frac{1}{nT}\sum_{t=1}^{T}(\tilde{Z}_{nt}, G_n\tilde{Z}_{nt}\delta_0)'(\tilde{Z}_{nt}, G_n\tilde{Z}_{nt}\delta_0) = \begin{pmatrix} \mathcal{H}_{1,nT} & \mathcal{H}_{2,nT} \\ \mathcal{H}'_{2,nT} & \mathcal{H}_{3,nT} \end{pmatrix}, \tag{C.1}$$

where

$$\mathcal{H}_{1,nT} = \frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}'_{nt}\tilde{Z}_{nt}, \mathcal{H}_{2,nT} = \frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}'_{nt}G_n\tilde{Z}_{nt}\delta_0, \mathcal{H}_{3,nT} = \frac{1}{nT}\sum_{t=1}^{T}(G_n\tilde{Z}_{nt}\delta_0)'G_n\tilde{Z}_{nt}\delta_0. \tag{C.2}$$

Hence, $\mathcal{H}_{nT}$ is the covariance matrix of the explanatory variables of (17) after being demeaned from time average. By using Lemmas 8 and 9 in Yu et al. (2007), we have the following lemma.

LEMMA 1. *Under Assumptions 1–7, for (C.2) and (19), we have*

$$(\mathcal{H}_{3,nT} - \mathcal{H}'_{2,nT}\mathcal{H}^{-1}_{1,nT}\mathcal{H}_{2,nT}) = (\mathcal{H}^s_{3,nT} - [\mathcal{H}^s_{2,nT}]'[\mathcal{H}^s_{1,nT}]^{-1}[\mathcal{H}^s_{2,nT}]),$$

$$(\mathrm{E}\mathcal{H}_{3,nT} - [\mathrm{E}\mathcal{H}_{2,nT}]'[\mathrm{E}\mathcal{H}_{1,nT}]^{-1}[\mathrm{E}\mathcal{H}_{2,nT}]) = (\mathrm{E}\mathcal{H}^s_{3,nT} - [\mathrm{E}\mathcal{H}^s_{2,nT}]'[\mathrm{E}\mathcal{H}^s_{1,nT}]^{-1}[\mathrm{E}\mathcal{H}^s_{2,nT}]),$$

$$\frac{1}{nT}\sum_{t=1}^{T}(\delta'_0\tilde{Z}'_{nt}G'_n - \mathcal{H}'_{2,nT}\mathcal{H}^{-1}_{1,nT}\tilde{Z}'_{nt})\mathcal{B}_n\tilde{V}_{nt} = \frac{1}{nT}\sum_{t=1}^{T}(\beta'_0\tilde{X}'_{nt}G'_n - \mathcal{H}^{s\prime}_{2,nT}[\mathcal{H}^s_{1,nT}]^{-1}Z^{s\prime}_{nt})'\mathcal{B}_n\tilde{V}_{nt},$$

$$(\mathcal{H}_{3,nT} - \mathcal{H}'_{2,nT}\mathcal{H}^{-1}_{1,nT}\mathcal{H}_{2,nT}) - (\mathrm{E}\mathcal{H}_{3,nT} - \mathrm{E}\mathcal{H}'_{2,nT}[\mathrm{E}\mathcal{H}_{1,nT}]^{-1}\mathrm{E}\mathcal{H}_{2,nT}) = O_p\left(\frac{1}{\sqrt{nT}}\right),$$

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

$$\frac{1}{nT}\sum_{t=1}^{T}(\delta_0'\tilde{Z}_{nt}'G_n' - \mathcal{H}_{2,nT}'\mathcal{H}_{1,nT}^{-1}\tilde{Z}_{nt}')\mathcal{B}_n\tilde{V}_{nt}$$
$$-\mathrm{E}\frac{1}{nT}\sum_{t=1}^{T}(\delta_0'\tilde{Z}_{nt}'G_n' - \mathcal{H}_{2,nT}'\mathcal{H}_{1,nT}^{-1}\tilde{Z}_{nt}')\mathcal{B}_n\tilde{V}_{nt} = O_p\left(\frac{1}{\sqrt{nT}}\right),$$

where $(\mathrm{E}\mathcal{H}_{3,nT} - \mathrm{E}\mathcal{H}_{2,nT}'(\mathrm{E}\mathcal{H}_{1,nT})^{-1}\mathrm{E}\mathcal{H}_{2,nT})$ *is* $O(1)$ *and* $\mathrm{E}\frac{1}{nT}\sum_{t=1}^{T}(\delta_0'\tilde{Z}_{nt}'G_n' - \mathcal{H}_{2,nT}'\mathcal{H}_{1,nT}^{-1}\tilde{Z}_{nt}')\mathcal{B}_n\tilde{V}_{nt}$ *is* $O(\frac{1}{T})$.

C.2. About $\hat{\sigma}_{nT}^2(\lambda)$**.** Denote $\mathcal{H}_{\lambda,nT} = \mathcal{H}_{3,nT} - \mathcal{H}_{2,nT}'\mathcal{H}_{1,nT}^{-1}\mathcal{H}_{2,nT}$, $\mathcal{V}_{z,nT} = \frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}'S_n(\lambda)S_n^{-1}\tilde{V}_{nt}$ and $\mathcal{X}_{z,nt} = G_n\tilde{Z}_{nt}\delta_0 - \tilde{Z}_{nt}\mathcal{H}_{1,nT}^{-1}\mathcal{H}_{2,nT}$ for notational simplicity in this section. For the unit root SDPD model in Section 4, given λ, the concentrated estimate $\hat{\delta}_{nT}(\lambda)$ and $\hat{\sigma}_{nT}^2(\lambda)$ are

$$\hat{\delta}_{nT}(\lambda) = \left[\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}'\tilde{Z}_{nt}\right]^{-1}\left[\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}'S_n(\lambda)\tilde{Y}_{nt}\right], \tag{C.3}$$

$$\hat{\sigma}_{nT}^2(\lambda) = \frac{1}{nT}\sum_{t=1}^{T}(S_n(\lambda)\tilde{Y}_{nt} - \tilde{Z}_{nt}\hat{\delta}_{nT}(\lambda))'(S_n(\lambda)\tilde{Y}_{nt} - \tilde{Z}_{nt}\hat{\delta}_{nT}(\lambda)). \tag{C.4}$$

Hence,

$$\hat{\sigma}_{nT}^2(\lambda) = (\lambda_0-\lambda)^2\mathcal{H}_{\lambda,nT} + \frac{1}{nT}\sum_{t=1}^{T}\tilde{V}_{nt}'S_n'^{-1}S_n'(\lambda)S_n(\lambda)S_n^{-1}\tilde{V}_{nt}$$
$$+2(\lambda_0-\lambda)\frac{1}{nT}\sum_{t=1}^{T}\mathcal{X}_{z,nt}'S_n(\lambda)S_n^{-1}\tilde{V}_{nt} - \mathcal{V}_{z,nT}'\mathcal{H}_{1,nT}^{-1}\mathcal{V}_{z,nT}, \tag{C.5}$$

$$\frac{\partial\hat{\sigma}_{nT}^2(\lambda)}{\partial\lambda} = 2(\lambda-\lambda_0)\mathcal{H}_{\lambda,nT} - \frac{2}{nT}\sum_{t=1}^{T}\tilde{V}_{nt}'G_n'S_n(\lambda)S_n^{-1}\tilde{V}_{nt}$$
$$-2(\lambda_0-\lambda)\frac{1}{nT}\sum_{t=1}^{T}\mathcal{X}_{z,nt}'G_n\tilde{V}_{nt} - \frac{2}{nT}\sum_{t=1}^{T}\mathcal{X}_{z,nt}'S_n(\lambda)S_n^{-1}\tilde{V}_{nt}$$
$$+2\mathcal{V}_{z,nT}'\mathcal{H}_{1,nT}^{-1}\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}'G_n\tilde{V}_{nt}\right), \tag{C.6}$$

$$\frac{\partial^2\hat{\sigma}_{nT}^2(\lambda)}{\partial\lambda^2} = 2\mathcal{H}_{\lambda,nT} + \frac{2}{nT}\sum_{t=1}^{T}\tilde{V}_{nt}'G_n'G_n\tilde{V}_{nt} + \frac{4}{nT}\sum_{t=1}^{T}\mathcal{X}_{z,nt}'G_n\tilde{V}_{nt}$$
$$-2\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}'G_n\tilde{V}_{nt}\right)'\mathcal{H}_{1,nT}^{-1}\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}'G_n\tilde{V}_{nt}\right). \tag{C.7}$$

Additionally, denoting $\mathcal{H}_{\lambda,nT}^s = \mathcal{H}_{3,nT}^s - \mathcal{H}_{2,nT}^{s\prime}\mathcal{H}_{1,nT}^{s-1}\mathcal{H}_{2,nT}^s$, $\mathcal{V}_{z,nT}^s = \frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}^{s\prime}S_n(\lambda)S_n^{-1}\tilde{V}_{nt}$, and $\mathcal{X}_{x,nt}^s = G_n\tilde{X}_{nt}\beta_0 - \tilde{Z}_{nt}^s\mathcal{H}_{1,nT}^{s-1}\mathcal{H}_{2,nT}^s$ for simplicity, by using

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

Lemma 1, we have

$$\hat{\sigma}_{nT}^2(\lambda) = (\lambda_0-\lambda)^2\mathcal{H}_{\lambda,nT}^s + \frac{1}{nT}\sum_{t=1}^{T}\tilde{V}_{nt}'S_n'^{-1}S_n'(\lambda)S_n(\lambda)S_n^{-1}\tilde{V}_{nt} \tag{C.8}$$

$$+2(\lambda_0-\lambda)\frac{1}{nT}\sum_{t=1}^{T}\mathcal{X}_{x,nt}^{s\prime}S_n(\lambda)S_n^{-1}\tilde{V}_{nt} - \mathcal{V}_{z,nT}^{s\prime}\mathcal{H}_{1,nT}^{s-1}\mathcal{V}_{z,nT}^{s}.$$

Hence, using Lemmas 8 and 9 in Yu et al. (2007), we have

$$\hat{\sigma}_{nT}^2(\lambda) = \sigma_0^2 + |\lambda-\lambda_0|^2\cdot O_p(1) + |\lambda-\lambda_0|\cdot O_p(1) + O_p\left(\max\left(\frac{1}{\sqrt{nT}},\frac{1}{T}\right)\right), \tag{C.9a}$$

$$\frac{\partial\hat{\sigma}_{nT}^2(\lambda)}{\partial\lambda} = -\sigma_0^2\frac{2}{n}\mathrm{tr}G_n + |\lambda-\lambda_0|\cdot O_p(1) + O_p\left(\max\left(\frac{1}{\sqrt{nT}},\frac{1}{T}\right)\right), \tag{C.9b}$$

$$\frac{\partial^2\hat{\sigma}_{nT}^2(\lambda)}{\partial\lambda^2} = 2(\mathcal{H}_{3,nT} - \mathcal{H}_{2,nT}'\mathcal{H}_{1,nT}^{-1}\mathcal{H}_{2,nT})$$

$$+2\sigma_0^2\frac{1}{n}\mathrm{tr}G_n'G_n + O_p\left(\max\left(\frac{1}{\sqrt{nT}},\frac{1}{T}\right)\right). \tag{C.9c}$$

Furthermore, using Lemma 1, at $\lambda=\lambda_0$, we have

$$\sqrt{nT}\frac{\partial\hat{\sigma}_{nT}^2(\lambda_0)}{\partial\lambda}$$

$$= -\frac{2}{\sqrt{nT}}\sum_{t=1}^{T}\tilde{V}_{nt}'G_n'\tilde{V}_{nt} - \frac{2}{\sqrt{nT}}\sum_{t=1}^{T}(\beta_0'\tilde{X}_{nt}'G_n' - [\mathcal{H}_{2,nT}^s]'[\mathcal{H}_{1,nT}^s]^{-1}\tilde{Z}_{nt}^{s\prime})\tilde{V}_{nt}$$

$$+O_p\left(\max\left(\frac{1}{\sqrt{nT}},\frac{1}{T},\sqrt{\frac{n}{T^3}}\right)\right). \tag{C.10}$$

APPENDIX D: Proofs of Theorems in Section 4

Proof of Theorem 3. In the following, we show the consistency of the QMLE $\hat{\lambda}_{nT}$ for λ_0, then derive the distributions of the QMLEs of the remaining parameters.

Consistency of $\hat{\lambda}_{nT}$. We have the concentrated likelihood (10), where $\hat{\sigma}_{nT}^2(\lambda)$ has the form (C.8). The optimization problem is similar to Yu et al. (2008), where the regressors in the reduced form (17) are Z_{nt}^s and $G_n\tilde{X}_{nt}\beta_0$ instead. Hence, under the identification condition in Assumption 8′, we get the consistency of $\hat{\lambda}_{nT}$.

Asymptotic Distribution of $\hat{\lambda}_{nT}$. The distribution of $\hat{\lambda}_{nT}$ is such that

$$\sqrt{nT}(\hat{\lambda}_{nT}-\lambda_0) + \sqrt{\frac{n}{T}}b_{\lambda_0,nT} + O_p\left(\max\left(\sqrt{\frac{n}{T^3}},\frac{1}{\sqrt{T}}\right)\right)$$

$$\xrightarrow{d} N(0, \lim_{T\to\infty}(\Sigma_{\lambda_0,nT}^{-1} + \Sigma_{\lambda_0,nT}^{-2}\Omega_{\lambda_0,nT})), \tag{D.1}$$

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

where

$$b_{\lambda_0,nT} = \Sigma^{-1}_{\lambda_0,nT} \cdot a_{\lambda_0,nT},$$

$$a_{\lambda_0,nT} = -\frac{1}{2}\left[([\mathcal{H}^s_{1,nT}]^{-1}[\mathcal{H}^s_{2,nT}])_1 \frac{1}{n}\mathrm{tr}(S_n^{-1}) + ([\mathcal{H}^s_{1,nT}]^{-1}[\mathcal{H}^s_{2,nT}])_2 \frac{1}{n}\mathrm{tr}(G_n)\right], \tag{D.2}$$

$$\Sigma_{\lambda_0,nT} = \frac{1}{\sigma_0^2}(\mathcal{H}^s_{3,nT} - [\mathcal{H}^s_{2,nT}]'[\mathcal{H}^s_{1,nT}]^{-1}[\mathcal{H}^s_{2,nT}]) + \frac{1}{n}\left(\mathrm{tr}G_n' G_n + \mathrm{tr}G_n^2 - \frac{2(\mathrm{tr}G)^2}{n}\right), \tag{D.3}$$

$$\Omega_{\lambda_0,nT} = \frac{\mu_4 - 3\sigma_0^4}{\sigma_0^4}\frac{1}{n}\sum_{i=1}^{n}\left[G_{n,ii} - \frac{\mathrm{tr}G_n}{n}\right]^2, \tag{D.4}$$

and $([\mathcal{H}^s_{1,nT}]^{-1}[\mathcal{H}^s_{2,nT}])_1$, $([\mathcal{H}^s_{1,nT}]^{-1}[\mathcal{H}^s_{2,nT}])_2$ are, respectively, the first and second elements of the column vector $[\mathcal{H}^s_{1,nT}]^{-1}[\mathcal{H}^s_{2,nT}]$.

To prove this, from the concentrated log likelihood for this model, which is similar to (10), we have

$$\frac{1}{nT}\frac{\partial \ln L_{n,T}(\lambda)}{\partial \lambda} = -\frac{1}{2\hat{\sigma}^2_{nT}(\lambda)}\frac{\partial \hat{\sigma}^2_{nT}(\lambda)}{\partial \lambda} - \frac{1}{n}\mathrm{tr}G_n(\lambda), \tag{D.5}$$

$$\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}(\lambda)}{\partial \lambda^2} = -\frac{1}{2\hat{\sigma}^4_{n,T}(\lambda)}\left[\frac{\partial^2 \hat{\sigma}^2_{nT}(\lambda)}{\partial \lambda^2}\hat{\sigma}^2_{nT}(\lambda) - \left(\frac{\partial \hat{\sigma}^2_{nT}(\lambda)}{\partial \lambda}\right)^2\right] - \frac{1}{n}\mathrm{tr}(G_n^2(\lambda)).$$

By plugging (C.5) at λ_0, (C.9), and (C.10) into (D.5), we have

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}(\lambda_0)}{\partial \lambda} = \frac{1}{\sqrt{nT}}\frac{\partial \ln L^*_{n,T}(\lambda_0)}{\partial \lambda} - \Delta_{\lambda_0,nT} + O_p\left(\max\left(\frac{1}{\sqrt{T}}, \sqrt{\frac{n}{T^3}}\right)\right),$$

where

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L^*_{n,T}(\lambda_0)}{\partial \lambda} = \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}(\beta_0'\tilde{X}_{nt}'G_n' - \mathcal{H}^{s\prime}_{2,nT}[\mathcal{H}^s_{1,nT}]^{-1}Z^{s*\prime}_{nt})V_{nt} + \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}V_{nt}'\left(G_n' - \frac{1}{n}\mathrm{tr}G_n \cdot I_n\right)V_{nt}, \tag{D.6}$$

and

$$\Delta_{\lambda_0,nT} = -\frac{1}{\sigma_0^2}\sqrt{\frac{T}{n}}\left[\mathcal{H}^{s\prime}_{2,nT}(\mathcal{H}^s_{1,nT})^{-1}\left(\frac{1}{T}S_n^{-1}\bar{\xi}_{nT,-1}, \frac{1}{T}G_n\bar{\xi}_{nT,-1}, \mathbf{0}_{n\times k_x}\right)\right]'\bar{V}_{nT}. \tag{D.7}$$

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

Here, $Z_{nt}^{s*} = (\frac{1}{T}(\tilde{\iota}_{-1}S_n^{-1}\mathbf{c}_{n0} + S_n^{-1}\tilde{\bar{\mathbb{X}}}_{n,t-1}\beta_0 + S_n^{-1}\xi_{n,t-1}),\ \frac{1}{T}(\tilde{\iota}_{-1}G_n\mathbf{c}_{n0} + G_n\tilde{\bar{\mathbb{X}}}_{n,t-1}\beta_0 + G_n\xi_{n,t-1}),\ \tilde{X}_{nt})$ in (D.6) is the component of $\tilde{Z}_{nt}^s$ uncorrelated with V_{nt} such that

$$\tilde{Z}_{nt}^s = Z_{nt}^{s*} - \left(\frac{1}{T}S_n^{-1}\bar{\xi}_{nT,-1}, \frac{1}{T}G_n\bar{\xi}_{nT,-1}, \mathbf{0}_{n\times k_x}\right). \tag{D.8}$$

Under Assumption 8′, so that $\Sigma_{\lambda_0,nT}$ is nonzero, we have

$$\sqrt{nT}(\hat{\lambda}_{nT} - \lambda_0) = \Sigma_{\lambda_0,nT}^{-1}\left(\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^*(\lambda_0)}{\partial\lambda} - \Delta_{\lambda_0,nT}\right)$$

$$+ O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{T}, \sqrt{\frac{n}{T^3}}\right)\right). \tag{D.9}$$

For $\Delta_{\lambda_0,nT}$ in (D.7), using Lemmas 8 and 9 in Yu et al. (2007), we have

$$\Delta_{\lambda_0,nT} = \sqrt{\frac{n}{T}}a_{\lambda_0,nT} + O_p\left(\max\left(\sqrt{\frac{n}{T^3}}, \frac{1}{\sqrt{T}}\right)\right), \tag{D.10}$$

where $a_{\lambda_0,nT}$ is in (D.2). For $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^*(\lambda_0)}{\partial\lambda}$, it is asymptotically normal implied by the CLT in Yu et al. (2008). To calculate the limit variance, using uncorrelatedness of Z_{nt}^{s*} and V_{nt}, we have

$$Cov\left(\frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\left(V_{nt}'\left(G_n' - \sigma_0^2\frac{\mathrm{tr}G_n}{n}I_n\right)V_{nt}\right),\right.$$

$$\left.\frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}(\beta_0'\tilde{X}_{nt}'G_n' - [\mathcal{H}_{2,nT}^s]'[\mathcal{H}_{1,nT}^s]_{nt}^{-1}Z_{nt}^{s*\prime})V_{nt}\right)$$

$$= \frac{\mu_3}{\sigma_0^4}\frac{1}{nT}\sum_{i=1}^{n}\left(G_{n,ii} - \frac{\mathrm{tr}G_n}{n}\right)\left(\mathrm{E}\sum_{t=1}^{T}(G_n\tilde{X}_{nt}\beta_0 - Z_{nt}^{s*}[\mathcal{H}_{1,nT}^s]^{-1}\mathcal{H}_{2,nT}^s)\right)_i = 0$$

because $\sum_{t=1}^{T}G_n\tilde{X}_{nt}\beta_0 = \mathbf{0}$ and $\mathrm{E}\sum_{t=1}^{T}Z_{nt}^{s*} = \mathbf{0}$. Hence,

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^*(\lambda_0)}{\partial\lambda} \xrightarrow{d} N(0, \lim_{T\to\infty}(\Sigma_{\lambda_0,nT} + \Omega_{\lambda_0,nT})) \tag{D.11}$$

where $\Sigma_{\lambda_0,nT}$ and $\Omega_{\lambda_0,nT}$ are defined in (D.3) and (D.4). Also, from Lemma 1 we have

$$\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}(\bar{\lambda}_{nT})}{\partial\lambda^2} = -\frac{1}{n}\left(\mathrm{tr}G_n'G_n + \mathrm{tr}G_n^2 - \frac{2(\mathrm{tr}G_n)^2}{n}\right)$$

$$- \frac{1}{\sigma_0^2}(\mathcal{H}_{3,nT}^s - [\mathcal{H}_{2,nT}^s]'[\mathcal{H}_{1,nT}^s]^{-1}[\mathcal{H}_{2,nT}^s]) + o_p(1). \tag{D.12}$$

From (D.12), (D.10), and (D.11), we have $\sqrt{nT}(\hat{\lambda}_{nT} - \lambda_0) = O_p(1)\cdot\left(O_p(1) + O_p\left(\sqrt{\frac{n}{T}}\right)\right)$. Hence,

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

$$\hat{\lambda}_{nT} - \lambda_0 = O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{T}\right)\right). \tag{D.13}$$

Using (D.12), (D.10), and (D.11) again, we get the result in (D.1).

Asymptotic Distribution of $\hat{\theta}_{nT}$. Denote $J_2 = (0, 1, \mathbf{0}_{1\times k_x})'$, then $\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}'\tilde{Z}_{nt}\right)^{-1}$ $\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}'W_n\tilde{\tilde{Y}}_{n,t-1}\right) = J_2$. Hence, from (C.3), using $S_n(\lambda)\tilde{Y}_{nt} = [I_n - (\lambda-\lambda_0)G_n]$ $\tilde{Z}_{nt}\delta_0 + S_n(\lambda)S_n^{-1}\tilde{V}_{nt}$, we have

$$\begin{aligned}
&\sqrt{nT}(\hat{\delta}_{nT}(\hat{\lambda}_{nT}) - \delta_0) \\
&\quad = \Upsilon_{\delta,T}^{-1}\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}^{s\prime}\tilde{Z}_{nt}^{s}\right)\left(\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\tilde{Z}_{nt}^{s\prime}V_{nt}\right) - \sqrt{nT}(\hat{\lambda}_{nT}-\lambda_0)J_2 \\
&\qquad -\Upsilon_{\delta,T}^{-1}\sqrt{nT}(\hat{\lambda}_{nT}-\lambda_0)\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}^{s\prime}\tilde{Z}_{nt}^{s}\right)\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}^{s\prime}(G_n\tilde{X}_{nt}\beta_0)\right) + R_{\hat{\delta}_{nT}},
\end{aligned} \tag{D.14}$$

where $R_{\hat{\delta}_{nT}} = -\Upsilon_{\delta,T}^{-1}\sqrt{nT}(\hat{\lambda}_{nT}-\lambda_0)(\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}^{s\prime}\tilde{Z}_{nt}^{s})^{-1}(\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}^{s\prime}G_nV_{nt})$, $\Upsilon_{\delta,T} = \mathrm{diag}(T, T, l_{k_x}')$, and $\tilde{Z}_{nt}^{s} = \tilde{Z}_{nt}\Upsilon_{\delta,T}^{-1}$. As $(\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}^{s\prime}\tilde{Z}_{nt}^{s})^{-1}(\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}_{nt}^{s\prime}G_nV_{nt})$ is $O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{T}\right)\right)$ and $\sqrt{nT}(\hat{\lambda}_{nT}-\lambda_0)$ is $O_p\left(\max\left(1, \sqrt{\frac{n}{T}}\right)\right)$ from (D.13), we have $R_{\hat{\delta}_{nT}} = \Upsilon_{\delta,T}^{-1}\cdot O_p\left(\max\left(\frac{1}{T}, \frac{1}{\sqrt{nT}}, \sqrt{\frac{n}{T^3}}\right)\right)$.

From (C.8), using Lemma 1 and (D.13), we have $\hat{\sigma}_{nT}^2(\hat{\lambda}_{nT}) = \frac{1}{nT}\sum_{t=1}^{T}\tilde{V}_{nt}'S_n^{\prime-1}S_n'(\hat{\lambda}_{nT})$ $S_n(\hat{\lambda}_{nT})S_n^{-1}\tilde{V}_{nt} + O_p\left(\max\left(\frac{1}{nT}, \frac{1}{T^2}\right)\right)$. As $S_n(\lambda)S_n^{-1} = I_n - (\lambda-\lambda_0)G_n$, using $\frac{1}{nT}$ $\sum_{t=1}^{T}\tilde{V}_{nt}'G_n'G_n\tilde{V}_{nt} = \sigma_0^2\frac{\mathrm{tr}G_n}{n} + O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{T}\right)\right)$ and $\sqrt{nT}(\hat{\lambda}_{nT}-\lambda_0) = O_p\left(\max\left(1, \sqrt{\frac{n}{T}}\right)\right)$, we have

$$\begin{aligned}
\sqrt{nT}\left(\hat{\sigma}_{n,T}^2(\hat{\lambda}_{nT}) - \sigma_0^2\right) &= \frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\left(\tilde{V}_{nt}'\tilde{V}_{nt} - n\sigma_0^2\right) \\
&\quad -2(\sqrt{nT}(\hat{\lambda}_{nT}-\lambda_0))\sigma_0^2\frac{\mathrm{tr}G_n}{n} + R_{\hat{\sigma}_{nT}^2},
\end{aligned} \tag{D.15}$$

where $R_{\hat{\sigma}_{nT}^2} = O_p\left(\max\left(\frac{1}{T}, \frac{1}{\sqrt{nT}}, \sqrt{\frac{n}{T^3}}\right)\right)$.

Also, from (D.9), we have

$$\begin{aligned}
&\sqrt{nT}(\hat{\lambda}_{nT}-\lambda_0) \\
&\quad = \left(\Sigma_{\lambda_0,nT}\right)^{-1}\frac{1}{\sigma_0^2}\left(\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\tilde{V}_{nt}'(G_n' - \frac{1}{n}\mathrm{tr}G_n\cdot I_n)\tilde{V}_{nt}\right) \\
&\qquad + \left(\Sigma_{\lambda_0,nT}\right)^{-1}\frac{1}{\sigma_0^2}\left(\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}(\beta_0'\tilde{X}_{nt}'G_n' - [\mathcal{H}_{2,nT}^{s}]'[\mathcal{H}_{1,nT}^{s}]^{-1}\tilde{Z}_{nt}^{s\prime})\tilde{V}_{nt}\right) + R_{\hat{\lambda}_{nT}},
\end{aligned} \tag{D.16}$$

where $R_{\hat{\lambda}_{nT}} = O_p\left(\max\left(\frac{1}{T}, \frac{1}{\sqrt{nT}}, \sqrt{\frac{n}{T^3}}\right)\right)$.

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

Hence, from (D.14), (D.15), and (D.16), define

$$B_{1,nT} = \begin{pmatrix} I_{k_x+2} & \Upsilon_{\delta,T}^{-1}[\mathcal{H}_{1,nT}^{s}]^{-1}\mathcal{H}_{2,nT}^{s} + J_2 & \mathbf{0}_{(k_x+2)\times 1} \\ \mathbf{0}_{1\times(k_x+2)} & 1 & 0 \\ \mathbf{0}_{1\times(k_x+2)} & \frac{2}{n}\sigma_0^2 \mathrm{tr}G_n & 1 \end{pmatrix},$$

$$B_{2,nT} = \begin{pmatrix} \Upsilon_{\delta,T}^{-1}[\mathcal{H}_{1,nT}^{s}]^{-1}\sigma_0^2 & \mathbf{0}_{(k_x+2)\times 1} & \mathbf{0}_{(k_x+2)\times 1} \\ \mathbf{0}_{1\times(k_x+2)} & \Sigma_{\lambda_0,nT}^{-1} & 0 \\ \mathbf{0}_{1\times(k_x+2)} & 0 & 2\sigma_0^4 \end{pmatrix}, \text{and}$$

$$B_{3,nT} = \begin{pmatrix} I_{k_x+2} & \mathbf{0}_{(k_x+2)\times 1} & \mathbf{0}_{(k_x+2)\times 1} \\ -\mathcal{H}_{2,nT}^{s\prime}[\mathcal{H}_{1,nT}^{s}]^{-1} & 1 & -\frac{2}{n}\sigma_0^2 \mathrm{tr}G_n \\ \mathbf{0}_{1\times(k_x+2)} & 0 & 1 \end{pmatrix};$$

we have

$$\sqrt{nT}\begin{pmatrix} \hat{\delta}_{n,T}(\hat{\lambda}_{nT}) - \delta_0 \\ \hat{\lambda}_{nT} - \lambda_0 \\ \hat{\sigma}_{nT}^2(\hat{\lambda}_{nT}) - \sigma_0^2 \end{pmatrix} = C_{1,nT} \times \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{nT}^{s*}(\theta_0)}{\partial \theta} + R_{\hat{\theta}_{nT}},$$

where $C_{1,nT} = B_{1,nT}^{-1} \times B_{2,nT} \times B_{3,nT}$, $B_{1,nT}^{-1} =$

$$\begin{pmatrix} I_{k_x+2} & -\Upsilon_{\delta,T}^{-1}[\mathcal{H}_{1,nT}^{s}]^{-1}\mathcal{H}_{2,nT}^{s} - J_2 & \mathbf{0}_{(k_x+2)\times 1} \\ \mathbf{0}_{1\times(k_x+2)} & 1 & 0 \\ \mathbf{0}_{1\times(k_x+2)} & -\frac{2}{n}\sigma_0^2 \mathrm{tr}G_n & 1 \end{pmatrix}, \text{ and}$$

$$R_{\hat{\theta}_{nT}} \equiv B_{1,nT}^{-1} \cdot (R_{\hat{\delta}_{nT}}', R_{\hat{\lambda}_{nT}}, R_{\hat{\sigma}_{nT}^2})' = \begin{pmatrix} R_{\hat{\delta}_{nT}} - (\Upsilon_{\delta,T}^{-1}[\mathcal{H}_{1,nT}^{s}]^{-1}\mathcal{H}_{2,nT}^{s} + J_2)R_{\hat{\lambda}_{nT}} \\ R_{\hat{\lambda}_{nT}} \\ R_{\hat{\sigma}_{nT}^2} - \frac{2}{n}\sigma_0^2 \mathrm{tr}G_n R_{\hat{\lambda}_{nT}} \end{pmatrix}. \tag{D.17}$$

For $\Sigma_{\theta_0,nT}^{s}$ in (22), we can also express it as

$$[\Sigma_{\theta_0,nT}^{s}]^{-1} = B_{3,nT}' \times \begin{pmatrix} [\mathcal{H}_{1,nT}^{s}]^{-1}\sigma_0^2 & \mathbf{0}_{(k_x+2)\times 1} & \mathbf{0}_{(k_x+2)\times 1} \\ \mathbf{0}_{1\times(k_x+2)} & \Sigma_{\lambda_0,nT}^{-1} & 0 \\ \mathbf{0}_{1\times(k_x+2)} & 0 & 2\sigma_0^4 \end{pmatrix} \times B_{3,nT}.$$

Denote

$$J = \begin{pmatrix} \mathbf{0}_{(k_x+2)\times(k_x+2)} & J_2 & \mathbf{0}_{(k_x+2)\times 1} \\ \mathbf{0}_{2\times(k_x+2)} & \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 1} \end{pmatrix}.$$

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

By comparison, we can see that $C_{1,nT} = (I-J)\cdot \mathrm{diag}(\Upsilon_{\delta,T}^{-1}, 1, 1)\cdot[\Sigma^s_{\theta_0,nT}]^{-1}$. We have

$$\mathrm{diag}(T, 1, l'_{k_x}, 1, 1)\cdot R_{\hat{\theta}_{nT}} = O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{T}, \sqrt{\frac{n}{T^3}}\right)\right), \tag{D.18}$$

because $\Upsilon_{\delta,T} R_{\hat{\delta}_{nT}}$, $R_{\hat{\lambda}_{nT}}$, and $R_{\hat{\sigma}^2_{nT}}$ are all $O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{T}, \sqrt{\frac{n}{T^3}}\right)\right)$. Hence, we have

$$\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) = (I-J)\cdot \mathrm{diag}(\Upsilon_{\delta,T}^{-1}, 1, 1)[\Sigma^s_{\theta_0,nT}]^{-1}\frac{1}{\sqrt{nT}}\frac{\partial \ln L^s_{nT}(\theta_0)}{\partial\theta} + R_{\hat{\theta}_{nT}}. \tag{D.19}$$

As $\mathrm{diag}(T, 1, l'_{k_x}, 1, 1)\cdot(I-J)\cdot\mathrm{diag}(\Upsilon_{\delta,T}^{-1}, 1, 1) = \Xi_T$ in (20), using (D.18), we have

$$\begin{aligned}\mathrm{diag}(T, 1, l'_{k_x}, 1, 1)\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) = {} & \Xi_T[\Sigma^s_{\theta_0,nT}]^{-1}\frac{1}{\sqrt{nT}}\frac{\partial \ln L^s_{nT}(\theta_0)}{\partial\theta} \\ & + O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{T}, \sqrt{\frac{n}{T^3}}\right)\right).\end{aligned} \tag{D.20}$$

Here, $\frac{1}{\sqrt{nT}}\frac{\partial \ln L^s_{n,T}(\theta_0)}{\partial\theta} = \frac{1}{\sqrt{nT}}\frac{\partial \ln L^{s*}_{n,T}(\theta_0)}{\partial\theta} - \Delta_{\theta_0,nT}$, where $\frac{1}{\sqrt{nT}}\frac{\partial \ln L^{s*}_{nT}(\theta_0)}{\partial\theta}$ and $\Delta_{\theta_0,nT}$ are

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L^{s*}_{nT}(\theta_0)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^T Z^{s*\prime}_{nt} V_{nt} \\ \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^T (V'_{nt}G'_n V_{nt} - \sigma_0^2 \mathrm{tr}G_n) \\ \quad + \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^T (G_n\tilde{X}_{nt}\beta_0)' V_{nt} \\ \frac{1}{2\sigma_0^4}\frac{1}{\sqrt{nT}}\sum_{t=1}^T \left(V'_{nt}V_{nt} - n\sigma_0^2\right) \end{pmatrix}, \tag{D.21}$$

and

$$\begin{aligned}\Delta_{\theta_0,nT} &= \begin{pmatrix} \frac{1}{\sigma_0^2}\sqrt{\frac{T}{n}}(\frac{1}{T}S_n^{-1}\bar{\xi}_{nT,-1}, \frac{1}{T}G_n\bar{\xi}_{nT,-1}, \mathbf{0}_{n\times k_x})'\bar{V}_{nT} \\ \frac{1}{\sigma_0^2}\sqrt{\frac{T}{n}}\bar{V}'_{nT}G'_n\bar{V}_{nT} \\ \frac{1}{2\sigma_0^4}\sqrt{\frac{T}{n}}\bar{V}'_{nT}\bar{V}_{nT} \end{pmatrix} \\ &= \sqrt{\frac{n}{T}}a_{\theta_0,n} + O_p\left(\max\left(\sqrt{\frac{n}{T^3}}, \frac{1}{\sqrt{T}}\right)\right),\end{aligned} \tag{D.22}$$

with Z^{s*}_{nt} in (D.8). Using the CLT in Yu et al. (2008), we have

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L^{s*}_{n,T}(\theta_0)}{\partial\theta} \xrightarrow{d} N\left(0, \lim_{T\to\infty}[\Sigma^s_{\theta_0,nT} + \Omega_{\theta_0,n}]\right), \tag{D.23}$$

where $\Sigma^s_{\theta_0,nT}$ is in (22) and $\Omega_{\theta_0,n}$ is in (23). For $\Delta_{\theta_0,nT}$, using Lemmas 8 and 9 in Yu et al. (2007), we have (D.22). Therefore, we have the result in Theorem 3.

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press

Also, multiplying both sides of (D.19) by $(0, 1, \mathbf{0}_{1\times k_x}, 1, 0)$ and denoting $e_2 = (0, 1, \mathbf{0}_{1\times k_x}, 0, 0)'$, we have

$$\sqrt{nT^3}(\hat{\rho}_{nT} + \hat{\lambda}_{nT} - \rho_0 - \lambda_0)$$
$$= e_2'[\Sigma^s_{\theta_0,nT}]^{-1}\frac{1}{\sqrt{nT}}\frac{\partial \ln L^s_{nT}(\theta_0)}{\partial\theta} + T\cdot(0, 1, \mathbf{0}_{1\times k_x}, 1, 0)R_{\hat{\theta}_{nT}},$$

because $T\cdot(0, 1, \mathbf{0}_{1\times k_x}, 1, 0)\cdot(I - J)\cdot\mathrm{diag}(\Upsilon^{-1}_{\delta,T}, 1, 1) = e_2'$. From (D.17), we can have $T\cdot(0, 1, \mathbf{0}_{1\times k_x}, 1, 0)R_{\hat{\theta}_{nT}} = O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{T}, \sqrt{\frac{n}{T^3}}\right)\right)$. Hence, we have the result in (24). ■

Proof of Theorem 4. We will have the result in (25) if $\sqrt{\frac{n}{T}}(\Xi_T\cdot[\Sigma^s_{\hat{\theta}_{nT},nT}]^{-1}\cdot a_{\hat{\theta}_{nT},n} - \Xi_T\cdot[\Sigma^s_{\theta_0,nT}]^{-1}\cdot a_{\theta_0,n}) \xrightarrow{p} 0$ and $\frac{n}{T^3}\to 0$. The condition $\frac{n}{T^3}\to 0$ is an assumption. For the remaining condition, we have

$$\left(\Xi_T\cdot[\Sigma^s_{\hat{\theta}_{nT},nT}]^{-1}\cdot a_{\hat{\theta}_{nT},n} - \Xi_T\cdot[\Sigma^s_{\theta_0,nT}]^{-1}\cdot a_{\theta_0,n}\right)$$
$$= \left(\Xi_T\cdot[\Sigma^s_{\theta_0,nT}]^{-1}\cdot\left(a_{\hat{\theta}_{nT},n} - a_{\theta_0,n}\right)\right) + a_{\hat{\theta}_{nT},n}\cdot O_p\left(\max\left(\frac{1}{T}, \frac{1}{\sqrt{nT}}\right)\right). \quad \textbf{(D.24)}$$

From $a_n(\theta)$ in (21), $\frac{\partial a_n(\bar{\theta}_{nT})}{\partial\theta'}$ will be uniformly bounded in probability and $a_n(\hat{\theta}_{nT})$ is $O_p(1)$. Hence, using $\hat{\theta}_{nT} - \theta_0 = O_p\left(\max\left(\frac{1}{T}, \frac{1}{\sqrt{nT}}\right)\right)$, $\Xi_T\cdot[\Sigma^s_{\hat{\theta}_{nT},nT}]^{-1}\cdot a_{\hat{\theta}_{nT},n} - \Xi_T\cdot[\Sigma^s_{\theta_0,nT}]^{-1}\cdot a_{\theta_0,n} = O_p\left(\max\left(\frac{1}{T}, \frac{1}{\sqrt{nT}}\right)\right)$. This implies that $\sqrt{\frac{n}{T}}(\Xi_T\cdot[\Sigma^s_{\hat{\theta}_{nT},nT}]^{-1}\cdot a_{\hat{\theta}_{nT},n} - \Xi_T\cdot[\Sigma^s_{\theta_0,nT}]^{-1}\cdot a_{\theta_0,n}) \xrightarrow{p} 0$ under $\frac{n}{T^3}\to 0$. ■

https://doi.org/10.1017/S0266466609990600 Published online by Cambridge University Press