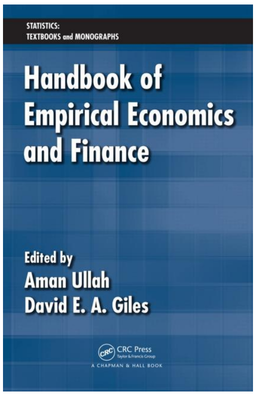


# Handbook of Empirical Economics and Finance

Ullah Aman, E. A. Giles David

## A Unified Estimation Approach for Spatial Dynamic Panel Data Models

Publication details
https://test.routledgehandbooks.com/doi/10.1201/b10440-15
Lee Lung-fei, Yu Jihai
**Published online on: 20 Dec 2010**

# 14

## *A Unified Estimation Approach for Spatial Dynamic Panel Data Models: Stability, Spatial Co-integration, and Explosive Roots*

**Lung-fei Lee and Jihai Yu**

**CONTENTS**

14.1 Introduction ... 397
14.2 The Model ... 398
 14.2.1 The DGP ... 398
 14.2.2 Data Transformation ... 401
 14.2.3 The Log-Likelihood Function ... 403
14.3 Asymptotic Properties of QMLE ... 404
 14.3.1 Consistency ... 405
 14.3.2 Asymptotic Distribution ... 407
 14.3.3 Bias Correction ... 409
 14.3.4 Testing ... 410
14.4 Monte Carlo Results ... 411
14.5 Conclusion ... 417
Appendices ... 418
References ... 432

### 14.1 Introduction

In recent decades, there is growing literature on the estimation of dynamic panel data models (see Phillips and Moon 1999; Hahn and Kuersteiner 2002; Alvarez and Arellano 2003; Hahn and Newey 2004, etc.). For the panel data with spatial interactions, Kapoor, Kelejian, and Prucha (2007) extend the asymptotic analysis of the method of moments estimators to a spatial panel model with error components, where $T$ is finite. Baltagi, Song, Jung, and Koh (2007) consider the testing of spatial and serial dependence in an extended model, where serial correlation on each spatial unit over time and spatial dependence across spatial units are allowed in the disturbances. Su and Yang (2007) study the dynamic panel data with spatial error and random effects. These panel models specify the spatial correlation by including spatially correlated disturbances but do not incorporate a spatial autoregressive term in the regression equation. With large $n$ and moderate or large $T$, Korniotis (2005) studies a time-space recursive model where only an individual time lag and a spatial time lag are present but not a contemporaneous spatial lag. A general model could be the spatial dynamic panel data (SDPD) where a contemporaneous spatial lag is also included. Yu, de Jong, and Lee (2007, 2008) and Yu and Lee (2010) study, respectively, the spatial cointegration, stable, and unit root SDPD models, where the individual time lag, spatial time lag and contemporaneous spatial lag are all included.

When the SDPD model has time dummy effects, we might need to transform the data to reduce the possible bias caused by the estimation of time effects (see Lee and Yu, 2010a), especially, when $n$ is proportional to $T$, or $n$ is small relative to $T$. Yu, de Jong, and Lee (2007) have a different bias correction procedures from that of the stable case in Yu, de Jong, and Lee (2008). In this chapter, we propose a data transformation approach based on a spatial difference operator, which can eliminate the time dummy effects as well as possible unstable and/or explosive components. After the data transformation, we can estimate the model by the method of maximum likelihood (ML) or quasi-maximum likelihood (QML) similar to Yu, de Jong, and Lee (2008), where there are neither time dummy effects, nor unstable and explosive components. We derive the asymptotics for the ML estimator (MLE) and QML estimator (QMLE). We propose a bias correction procedure that can be applied to different types of DGPs.

This chapter is organized as follows. In Section 14.2, the model is presented. We show that the stochastic process can be decomposed into stable, unstable or explosive, and time components. A spatial difference operator motivated by the spatial co-integration can provide a unified data transformation to eliminate the time component and the possible unstable or explosive components. We explain our method of estimation, which is a concentrated QML. Section 14.3 establishes the consistency and asymptotic distribution of the QMLE of the unified transformation approach. A bias correction procedure is also proposed. A Monte Carlo study is conducted in Section 14.4 to investigate finite sample performance of the estimators under different DGPs, and also the power of hypothesis testing of spatial co-integration using this unified approach. Section 14.5 concludes the chapter. Some useful lemmas and proofs are collected in the appendices.

## 14.2 The Model

### 14.2.1 The DGP

Consider the general SDPD model:

$$Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + V_{nt}, \quad t = 1, 2, \ldots, T, \tag{14.1}$$

where $Y_{nt} = (y_{1t}, y_{2t}, ..., y_{nt})'$ and $V_{nt} = (v_{1t}, v_{2t}, ..., v_{nt})'$ are $n \times 1$ column vectors, and $v_{it}$ is i.i.d. across $i$ and $t$ with zero mean and variance $\sigma_0^2$. $W_n$ is an $n \times n$ nonstochastic spatial weights matrix, $X_{nt}$ is an $n \times k$ matrix of nonstochastic regressors, $\mathbf{c}_{n0}$ is an $n \times 1$ column vector of individual fixed effects, $\alpha_{t0}$ is a scalar of time effect, and $l_n$ is an $n \times 1$ column vector of ones.[1] Therefore, the total number of parameters in this model is equal to the sum of the number of individuals $n$ and the number of time periods $T$, plus the dimension of the common parameters $(\gamma, \rho, \beta', \lambda, \sigma^2)'$ which is $k + 4$. In practice, $W_n$ is usually row-normalized with zero diagonals. A row-normalized $W_n$ has the property $W_n l_n = l_n$. The row-normalization of $W_n$ ensures that all the weights are between 0 and 1 and weighting operations can be interpreted as an average of the neighboring values. In this chapter, the row-normalization feature is imposed for our estimation approach.

Define $S_n(\lambda) = I_n - \lambda W_n$ and $S_n \equiv S_n(\lambda_0) = I_n - \lambda_0 W_n$. Then, presuming that $S_n$ is invertible and denoting $A_n = S_n^{-1}(\gamma_0 I_n + \rho_0 W_n)$, Equation 14.1 can be rewritten as

$$Y_{nt} = A_n Y_{n,t-1} + S_n^{-1} X_{nt}\beta_0 + S_n^{-1}\mathbf{c}_{n0} + \alpha_{t0} S_n^{-1} l_n + S_n^{-1} V_{nt}. \tag{14.2}$$

In the SDPD model, when all the eigenvalues of $A_n$ are smaller than 1, we have the stable case. When some eigenvalues of $A_n$ are equal to 1 but not all being 1, we have the spatial co-integration case. When some of them are greater than 1, we have the explosive case. Let $\varpi_n = \text{diag}\{\varpi_{n1}, \varpi_{n2}, ..., \varpi_{nn}\}$ be the $n \times n$ diagonal eigenvalues matrix of $W_n$ such that $W_n = R_n \varpi_n R_n^{-1}$, where $R_n$ is the corresponding eigenvector matrix. As $A_n = S_n^{-1}(\gamma_0 I_n + \rho_0 W_n)$, the eigenvalues matrix of $A_n$ is $D_n = (I_n - \lambda_0 \varpi_n)^{-1}(\gamma_0 I_n + \rho_0 \varpi_n)$ such that $A_n = R_n D_n R_n^{-1}$. When $W_n$ is row-normalized, all the eigenvalues are less than or equal to 1 in the absolute value, where it has definitely some eigenvalues being 1. Let $m_n$ be the number of unit eigenvalues of $W_n$ and let the first $m_n$ eigenvalues of $W_n$ be the unity. Hence, $D_n$ can be decomposed into two parts, one corresponding to the unit eigenvalues of $W_n$, and the other corresponding to the eigenvalues of $W_n$ which are smaller than 1. Define $J_n = \text{diag}\{\mathbf{1}'_{m_n}, 0, \cdots, 0\}$ with $\mathbf{1}_{m_n}$ being an $m_n \times 1$ vector of ones and $\tilde{D}_n = \text{diag}\{0, \cdots, 0, d_{n,m_n+1}, \cdots, d_{nn}\}$, where $|d_{ni}| < 1$, for $i = m_n + 1, \cdots, n$, are assumed.[2] As $J_n \cdot \tilde{D}_n = \mathbf{0}$, we have $A_n^h = (\frac{\gamma_0+\rho_0}{1-\lambda_0})^h R_n J_n R_n^{-1} + B_n^h$ where $B_n^h = R_n \tilde{D}_n^h R_n^{-1}$ for any $h = 1, 2, \cdots$. Hence, depending on the value of $\frac{\gamma_0+\rho_0}{1-\lambda_0}$, we have three cases. As $|\lambda_0| < 1$, which will be maintained under the Assumption 1 and 3 (see Section 14.3), we have the stable case when $\gamma_0 + \rho_0 + \lambda_0 < 1$; the spatial co-integration case when $\gamma_0 + \rho_0 + \lambda_0 = 1$ but $\gamma_0 \neq 1$; and the explosive case when $\gamma_0 + \rho_0 + \lambda_0 > 1$.

For the stable case, the rates of convergence of QMLEs are $\sqrt{nT}$, as shown in Yu, de Jong, and Lee (2008). For the spatial co-integration case where $Y_{nt}$ and

[1] Due to the presence of fixed individual and time effects, the $X_{nt}$ will not include time invariant or individual invariant regressors.

[2] We note that $d_{ni} = (\gamma_0 + \rho_0 \varpi_{ni})/(1 - \lambda_0 \varpi_{ni})$. Hence, if $\gamma_0 + \lambda_0 + \rho_0 < 1$, we have $d_{ni} < 1$ as $|\varpi_{ni}| \leq 1$. Some additional conditions are needed to ensure that $d_{ni} > -1$. See Appendix A.1.

$W_nY_{nt}$ are spatially co-integrated, Yu, de Jong, and Lee (2007) show that the QMLEs for such a model are $\sqrt{nT}$ consistent and asymptotically normal, but, the presence of the unstable components will make the estimators' asymptotic variance matrix singular. Consequently, a linear combination of the spatial and dynamic effects estimates can converge at a higher rate.[3] In addition to the above stable case and the spatial co-integration case, we may also have an explosive case in the event that some eigenvalues of $A_n$ are greater than unity in the absolute value.[4] In this chapter, we propose a unified transformation approach that can be used to estimate all three cases, namely, stable, spatial co-integrated, and explosive cases.

In earlier studies of the SDPD model, Yu, de Jong, and Lee (2007, 2008) consider the QMLE of the model with only the individual fixed effects. Subsequently, Lee and Yu (2010a) study the SDPD model with additional time effect when the process is stable. They propose a data transformation based on the deviation from cross-sectional mean, $I_n - \frac{1}{n}l_nl_n'$, to eliminate the time effects. That approach may be applied to study the unstable SDPD models with time effects but might not be able to eliminate unstable or explosive components. In this chapter, we report the use of a spatial difference operator, $I_n - W_n$, which may not only eliminate the time dummy effects, but also the possible unstable or explosive components, generated from the spatial co-integration or explosive roots. This implies that the spatial difference transformation can be applied to DGPs with stability, spatial co-integration, or explosive roots. The asymptotics of the resulting estimates can then be easily established for these DGPs. Thus, the transformation $I_n - W_n$ provides a unified estimation procedure for SDPD models.[5]

Denote $W_n^u = R_nJ_nR_n^{-1}$. Then, for $t \geq 0$, $Y_{nt}$ can be decomposed into a sum of a possible stable part, a possible unstable or explosive part, and a time effect part (see Appendix A.2 for proof)

$$Y_{nt} = Y_{nt}^u + Y_{nt}^s + Y_{nt}^{\alpha}, \tag{14.3}$$

[3] When $\gamma_0 + \lambda_0 + \rho_0 = 1$ and $\gamma_0 = 1$, the asymptotic properties of estimators are considered in Yu and Lee (2010). The QML estimate of the dynamic coefficient is $\sqrt{nT^3}$ consistent and the estimates of other parameters are $\sqrt{nT}$ consistent, and they are all asymptotically normal. Also, the sum of the contemporaneous and dynamic spatial effects will converge at $\sqrt{nT^3}$ rate.

[4] For the autoregressive AR(1) process in time series, asymptotic properties of the ordinary least square estimator have been investigated in White (1958, 1959), Anderson (1959), Nielsen (2001, 2005) and Phillips and Magdalinos (2007). For the SDPD due to its complexity, properties of a possible QMLE have not been investigated.

[5] We note that the spatial difference operator can be applied to cross-sectional units. However, its function is different from the time difference operator for a time series. The spatial difference operator does not eliminate pure time series unit root or explosive roots. Thus, the unified approach cannot be applied to the pure unit root SDPD models in Yu and Lee (2010).

where

$$Y_{nt}^s = \sum_{h=0}^{\infty} B_n^h S_n^{-1}(\mathbf{c}_{n0} + X_{n,t-h}\beta_0 + V_{n,t-h}),$$

$$Y_{nt}^u = W_n^u \left\{ \left( \frac{\gamma_0 + \rho_0}{1 - \lambda_0} \right)^{t+1} Y_{n,-1} + \frac{1}{(1-\lambda_0)} \left[ \sum_{h=0}^{t} \left( \frac{\gamma_0 + \rho_0}{1 - \lambda_0} \right)^h (\mathbf{c}_{n0} + X_{n,t-h}\beta_0 + V_{n,t-h}) \right] \right\},$$

$$Y_{nt}^{\alpha} = \frac{1}{(1-\lambda_0)} l_n \sum_{h=0}^{t} \alpha_{t-h,0} \left( \frac{\gamma_0 + \rho_0}{1 - \lambda_0} \right)^h.$$

The $Y_{nt}^u$ can be an unstable component when $\frac{\gamma_0+\rho_0}{1-\lambda_0} = 1$, which occurs when $\gamma_0 + \rho_0 + \lambda_0 = 1$ and $\lambda_0 \neq 1$. When $\gamma_0 + \rho_0 + \lambda_0 > 1$, it implies $\frac{\gamma_0+\rho_0}{1-\lambda_0} > 1$ and, hence, $Y_{nt}^u$ can be explosive. The $Y_{nt}^{\alpha}$ can be rather complicated as it depends on what exactly the time dummies represent. The $Y_{nt}$ can be explosive when $\alpha_{t0}$ represents some explosive functions of $t$, even when $\frac{\gamma_0+\rho_0}{1-\lambda_0}$ were smaller than 1. Without a specific time structure for $\alpha_{t0}$, it is desirable to eliminate this component for the estimation. The $Y_{nt}^s$ can be a stable component unless $\gamma_0 + \rho_0 + \lambda_0$ is much larger than 1. If the sum $\gamma_0 + \rho_0 + \lambda_0$ were too big, some of the eigenvalues $d_{ni}$ in $Y_{nt}^s$ might become larger than 1.

### 14.2.2 Data Transformation

Both the deviation from the cross-sectional mean $I_n - \frac{1}{n} l_n l_n'$ and the spatial difference operator $I_n - W_n$ can eliminate the $Y_{nt}^{\alpha}$ component in $Y_{nt}$. The transformation $I_n - W_n$ can be motivated via a feature of spatial co-integration below. Because $(I_n - W_n) l_n = 0$, $(I_n - W_n) Y_{nt}^{\alpha} = 0$. The $(I_n - W_n) Y_{nt}$ does not involve time dummies. In addition, because $W_n^u = R_n J_n R_n^{-1}$, it follows that $(I_n - W_n) W_n^u = R_n (I_n - D_n) J_n R_n^{-1} = 0$, and $(I_n - W_n) Y_{nt}^u = 0$. Therefore, $(I_n - W_n) Y_{nt} = (I_n - W_n) Y_{nt}^s$. That is, the transformation $I_n - W_n$ can eliminate not only time dummies but also the unstable component. Therefore, after the $(I_n - W_n)$ transformation, we will end up with the following equation:

$$(I_n - W_n) Y_{nt} = \lambda_0 W_n (I_n - W_n) Y_{nt} + \gamma_0 (I_n - W_n) Y_{n,t-1} + \rho_0 W_n (I_n - W_n) Y_{n,t-1} + (I_n - W_n) X_{nt}\beta_0 + (I_n - W_n)\mathbf{c}_{n0} + (I_n - W_n) V_{nt}. \quad (14.4)$$

The variance of $(I_n - W_n) V_{nt}$ is $\sigma_0^2 \Sigma_n$, where $\Sigma_n = (I_n - W_n)(I_n - W_n)'$. This transformed equation has less degrees of freedom than $n$. Denote the degree of freedom of Equation 14.4 as $n^*$. Then, $n^*$ is the rank of the variance matrix of $(I_n - W_n) V_{nt}$, which is the number of nonzero eigenvalues of $\Sigma_n$. Hence,

$n^* = n - m_n$ is also the number of non-unit eigenvalues[6] of $W_n$. Thus, the transformed variables do not have time effects and are all stable even when $\lambda_0 + \gamma_0 + \rho_0$ is equal to or greater than 1.

Let $[F_n, H_n]$ be the orthonormal matrix of eigenvectors and $\Lambda_n$ be the diagonal matrix of nonzero eigenvalues of $\Sigma_n$ such that $\Sigma_n F_n = F_n \Lambda_n$ and $\Sigma_n H_n = 0$. That is, the columns of $F_n$ consist of eigenvectors of nonzero eigenvalues and those of $H_n$ are for zero-eigenvalues of $\Sigma_n$. The $F_n$ is an $n \times n^*$ matrix and $\Lambda_n$ is an $n^* \times n^*$ diagonal matrix. Denote $W_n^* = \Lambda_n^{-1/2} F_n' W_n F_n \Lambda_n^{1/2}$ which is an $n^* \times n^*$ matrix. As is derived in Appendix A.3, we have

$$Y_{nt}^* = \lambda_0 W_n^* Y_{nt}^* + \gamma_0 Y_{n,t-1}^* + \rho_0 W_n^* Y_{n,t-1}^* + X_{nt}^* \beta_0 + \mathbf{c}_{n0}^* + V_{nt}^*, \tag{14.5}$$

where $Y_{nt}^* = \Lambda_n^{-1/2} F_n'(I_n - W_n) Y_{nt}$ and other variables are defined accordingly. Note that this transformed $Y_{nt}^*$ is an $n^*$ dimensional vector. Thus, at each $t$, after the removal of the time dummy variables as well as the unstable or explosive components in $Y_{nt}$, the remaining observations at period $t$ have only $n^*$ degrees of freedom. While the sum of the coefficients $\lambda_0 + \gamma_0 + \rho_0$ of this transformed equation can be equal to or greater than 1, the eigenvalues of $W_n^*$ are exactly those eigenvalues of $W_n$ not equal to the unity (see Appendix A.4) but less than 1 in the absolute value. It follows that 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 the absolute values even when $\lambda_0 + \gamma_0 + \rho_0 = 1$ with $|\lambda_0| < 1$ and $|\gamma_0| < 1$. For the explosive case with $\lambda_0 + \gamma_0 + \rho_0 > 1$, the eigenvalue of $A_n^*$ can be less than 1 only if $\frac{\rho_0 + \lambda_0}{1 - \gamma_0} < \frac{1}{\varpi_{\max 1}}$, where $\varpi_{\max 1}$ is the maximum positive eigenvalue of $W_n$ less than the unity (see Appendix A.1). Hence, the transformed model (Equation 14.5) is a stable one as long as $\lambda_0 + \gamma_0 + \rho_0$ is not much bigger than 1.[7]

The transformation $I_n - W_n$ for the case with $\gamma_0 + \rho_0 + \lambda_0 = 1$ but $\gamma_0 \neq 1$ has an interpretation as a spatial co-integrating matrix for elements of $Y_{nt}$. Denote time difference as $\Delta Y_{nt} = Y_{nt} - Y_{n,t-1}$. The reduced form Equation 14.2) implies that $\Delta Y_{nt} = (A_n - I_n) Y_{n,t-1} + S_n^{-1}(X_{nt}\beta_0 + \mathbf{c}_{n0} + V_{nt} + \alpha_{t0} l_n)$. For the case $\lambda_0 + \gamma_0 + \rho_0 = 1$ with $\gamma_0 \neq 1$, $A_n - I_n = (I_n - \lambda_0 W_n)^{-1}(\gamma_0 I_n + \rho_0 W_n) - I_n = (1 - \gamma_0)(I_n - \lambda_0 W_n)^{-1}(W_n - I_n)$. Hence, we have a vector error correction model (VECM) representation of Equation 14.2 as

$$\Delta Y_{nt} = (1 - \gamma_0)(I_n - \lambda_0 W_n)^{-1}(W_n - I_n) Y_{n,t-1} + S_n^{-1}(X_{nt}\beta_0 + \mathbf{c}_{n0} + V_{nt} + \alpha_{t0} l_n).$$

The matrix $I_n - W_n = R_n(I_n - \varpi_n) R_n^{-1}$ has its rank equal to the number of eigenvalues of $W_n$ different from 1. With the VECM representation, one may

---

6 This is so, because (1) the set $K_n$ of eigenvectors corresponding to the zero eigenvalues of $(I_n - W_n)(I_n - W_n)'$ is the same as that of $(I_n - W_n)'$; (2) the dimension of $K_n$ is the number of unit eigenvalues of $W_n'$; (3) $W_n = R_n \varpi_n R_n^{-1}$ if and only if $W_n' = R_n^{-1\prime} \varpi_n R_n'$, i.e., the eigenvalues of $W_n$ and $W_n'$ are the same.

7 Similar to Yu, de Jong, and Lee (2007) for the spatial co-integration case, we assume that the eigenvalues of $W_n$ with their absolute values less than 1 are bounded away from 1 for all $n$. Appendix A.1 provides sufficient conditions on the parameters of the model, which can imply this regularity condition.

regard $I_n - W_n$ as a co-integrating matrix with the co-integration rank as the number of non-unit eigenvalues of $W_n$. Hence, this transformation method has exploited the spatial co-integration of $Y_{nt}$'s for the estimation.

### 14.2.3 The Log-Likelihood Function

Suppose that $V_{nt}$ is normally distributed as $N(0, \sigma_0^2 I_n)$, the transformed $V_{nt}^*$ in Equation 14.5 will be $N(0, \sigma_0^2 I_{n^*})$. Denote $\delta = (\gamma, \rho, \beta')'$, $\theta = (\delta', \lambda)'$ and $S_n^*(\lambda) = I_{n^*} - \lambda W_n^*$. The log-likelihood function for $Y_{nt}^*$ in Equation 14.5 is

$$\ln L_{n,T}(\theta, \mathbf{c}_n^*) = -\frac{n^*T}{2}\ln 2\pi - \frac{n^*T}{2}\ln\sigma^2 + T\ln|S_n^*(\lambda)| - \frac{1}{2\sigma^2}\sum_{t=1}^{T} V_{nt}^{*\prime}(\theta, \mathbf{c}_n^*) V_{nt}^*(\theta, \mathbf{c}_n^*), \tag{14.6}$$

where $V_{nt}^*(\theta, \mathbf{c}_n^*) = S_n^*(\lambda) Y_{nt}^* - Z_{nt}^*\delta - \mathbf{c}_n^*$, $Z_{nt}^* = (Y_{n,t-1}^*, W_n^* Y_{n,t-1}^*, X_{nt}^*)$. In order to use Equation 14.6 for an effective estimation, the determinant and inverse of $S_n^*(\lambda)$ are needed. As is derived in Appendix A.4, using $S_n^*(\lambda) = \Lambda_n^{-1/2} F_n' S_n(\lambda) F_n \Lambda_n^{1/2}$, we have

$$|S_n^*(\lambda)| = \frac{1}{(1-\lambda)^{n-n^*}}|S_n(\lambda)|, \text{ and } S_n^{*-1}(\lambda) = \Lambda_n^{-1/2} F_n' S_n^{-1}(\lambda) F_n \Lambda_n^{1/2}. \tag{14.7}$$

Hence, the computation of the determinant of $S_n^*(\lambda)$ is not more complicated than $S_n(\lambda)$. Also,

$$\begin{aligned} V_{nt}^*(\theta, \mathbf{c}_n^*) &= S_n^*(\lambda) Y_{nt}^* - Z_{nt}^*\delta - \mathbf{c}_n^* \\ &= \Lambda_n^{-1/2} F_n' S_n(\lambda) F_n F_n'(I_n - W_n) Y_{nt} - \Lambda_n^{-1/2} F_n'(I_n - W_n) Z_{nt}\delta \\ &\quad - \Lambda_n^{-1/2} F_n'(I_n - W_n)\mathbf{c}_n \\ &= \Lambda_n^{-1/2} F_n'(I_n - W_n)[S_n(\lambda) Y_{nt} - Z_{nt}\delta - \mathbf{c}_n] \\ &= \Lambda_n^{-1/2} F_n'(I_n - W_n) V_{nt}(\theta, \mathbf{c}_n), \end{aligned}$$

by using $F_n F_n' + H_n H_n' = I_n$ and $H_n'(I_n - W_n) = \mathbf{0}$, where $Z_{nt} = (Y_{n,t-1}, W_n Y_{n,t-1}, X_{nt})$ and $V_{nt}(\theta, \mathbf{c}_n) = S_n(\lambda) Y_{nt} - Z_{nt}\delta - \mathbf{c}_n$. Hence,

$$V_{nt}^{*\prime}(\theta, \mathbf{c}_n^*) V_{nt}^*(\theta, \mathbf{c}_n^*) = V_{nt}'(\theta, \mathbf{c}_n)(I_n - W_n)'\Sigma_n^+(I_n - W_n) V_{nt}(\theta, \mathbf{c}_n), \tag{14.8}$$

where $\Sigma_n^+ = F_n \Lambda_n^{-1} F_n'$ is the generalized inverse of $\Sigma_n = (I_n - W_n)(I_n - W_n)'$. By using Equation 14.7 and 14.8, the log-likelihood function (Equation 14.6) for $Y^*_{nt}$ can be expressed in terms of $Y_{nt}$ as

$$\ln L_{n,T}(\theta, \mathbf{c}_n) = -\frac{n^*T}{2}\ln(2\pi\sigma^2) - (n-n^*)T\ln(1-\lambda) + T\ln|S_n(\lambda)| \\ -\frac{1}{2\sigma^2}\sum_{t=1}^{T}(S_n(\lambda)Y_{nt} - Z_{nt}\delta - \mathbf{c}_n)'(I_n - W_n)'\Sigma_n^{+}(I_n - W_n) \\ \times (S_n(\lambda)Y_{nt} - Z_{nt}\delta - \mathbf{c}_n). \tag{14.9}$$

Hence, after the transformation, the QML method is to estimate the SDPD model with only individual effects with $n^*$ cross-section units and $T$ time periods, where Equation 14.6 is the objective function. Alternatively, one may maximize Equation 14.9 expressed in terms of the original variables. However, although the components of $V_{nt}$ are i.i.d. in the model, the elements of $V^*_{nt}$ might not be independent (they are uncorrelated). The asymptotic analysis in Yu, de Jong, and Lee (2008) may not be directly carried over to the transformed model with the disturbances $V^*_{nt}$.[8] As Equation 14.6 is equivalent to Equation 14.9, we can analyze the asymptotic distribution of the estimator via Equation 14.9.

Using first order conditions, we concentrate out $\mathbf{c}_n$ in Equation 14.9 to obtain the concentrated likelihood function in terms of $\theta$. For an $n \times 1$ vector at period $t$, $\Upsilon_{nt}$, we define the deviation from time means as $\tilde{\Upsilon}_{nt} = \Upsilon_{nt} - \bar{\Upsilon}_{nT}$ and $\tilde{\tilde{\Upsilon}}_{n,t-1} = \Upsilon_{n,t-1} - \bar{\Upsilon}_{nT,-1}$, where $\bar{\Upsilon}_{nT} = \frac{1}{T}\sum_{t=1}^{T}\Upsilon_{nt}$ and $\bar{\Upsilon}_{nT,-1} = \frac{1}{T}\sum_{t=1}^{T}\Upsilon_{n,t-1}$. The concentrated log-likelihood is

$$\ln L_{n,T}(\theta) = -\frac{n^*T}{2}\ln 2\pi - \frac{n^*T}{2}\ln\sigma^2 - (n-n^*)T\ln(1-\lambda) + T\ln|I_n - \lambda W_n| \\ -\frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{V}'_{nt}(\theta)(I_n - W_n)'\Sigma_n^{+}(I_n - W_n)\tilde{V}_{nt}(\theta), \tag{14.10}$$

where $\tilde{V}_{nt}(\theta) = S_n(\lambda)\tilde{Y}_{nt} - \tilde{Z}_{nt}\delta$ and $(I_n - W_n)\tilde{V}_{nt}(\theta) = (I_n - W_n)[S_n(\lambda)\tilde{Y}_{nt} - \tilde{Z}_{nt}\delta - \tilde{\alpha}_t l_n]$ because $(I_n - W_n)l_n = \mathbf{0}$. At $\theta_0$, $\tilde{V}_{nt} = S_n\tilde{Y}_{nt} - \tilde{Z}_{nt}\delta_0$. For Equation 14.10, its first- and second-order derivatives are Equation A.16 and A.17 in Appendix C.2.

## 14.3 Asymptotic Properties of QMLE

For our analysis of the asymptotic properties of estimators, we make the following assumptions. Denote $J^*_n = (I_n - W_n)'\Sigma_n^{+}(I_n - W_n)$. We note that $J^*_n$ is an orthonormal projector with rank $n^*$ (see Appendix A.5).

[8] One could not treat the components of $V^*_{nt}$ as if they were independent when the disturbances are not normally distributed. Furthermore, it is not clear whether $W^*_n$ and $A^*_n = (I_{n^*} - \lambda_0 W^*_n)^{-1}(\gamma_0 I_{n^*} + \rho_0 W^*_n)$ would be uniformly bounded in both row and column sums even though $W_n$ and $A_n$ are.

**Assumption 1** $W_n$ is a row-normalized nonstochastic spatial weights matrix with zero diagonals.

**Assumption 2** The disturbances $\{v_{it}\}$, $i = 1, 2, ..., n$ and $t = 1, 2, ..., T$, are i.i.d. across $i$ and $t$ with zero mean, variance $\sigma_0^2$ and $E|v_{it}|^{4+\eta} < \infty$ for some $\eta > 0$.

**Assumption 3** $S_n(\lambda)$ is invertible for all $\lambda \in \Lambda$. Furthermore, $\Lambda$ is compact and the true parameter $\lambda_0$ is in the interior of $\Lambda$.

**Assumption 4** The elements of $X_{nt}$ are nonstochastic and bounded, uniformly in $n$ and $t$, and the limit of $\frac{1}{nT}\sum_{t=1}^{T} \tilde{X}'_{nt} J_n^* \tilde{X}_{nt}$ exists and is nonsingular.

**Assumption 5** $W_n$ is uniformly bounded in row and column sums in the absolute value (for short, UB).[9] Also $S_n^{-1}(\lambda)$ is UB, uniformly in $\lambda \in \Lambda$.

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

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

Assumption 1 is a standard normalization assumption in spatial econometrics. In many empirical applications, the rows of $W_n$ sum to 1, which ensures that all the weights are between 0 and 1. Assumption 2 provides regularity assumptions for $v_{it}$. Assumption 3 guarantees that Equation 14.2 is valid. When exogenous variables $X_{nt}$ are included in the model, it is convenient to assume that their elements are uniformly bounded[10] as in Assumption 4. Assumption 5 is originated by Kelejian and Prucha (1998, 2001) and is also used in Lee (2004, 2007). The uniform boundedness of $W_n$ and $S_n^{-1}(\lambda)$ is a condition that limits the spatial correlation to a manageable degree. Assumption 6 is the absolute summability condition and row/column sum boundedness condition, which will play an important role for asymptotic properties of QML estimator. In order to justify the absolute summability of $B_n$, a sufficient condition is $\|B_n\| < 1$ for any matrix norm (see Horn and Johnson (1985), Corollary 5.6.16) that satisfies $\|B_n\| = \|abs(B_n)\|$. When $\|B_n\| < 1$, $\sum_{h=0}^{\infty} B_n^h$ exists and can be defined as $(I_n - B_n)^{-1}$. Assumption 7 allows two cases: (1) $n^* \to \infty$ as $T \to \infty$; (2) $n^*$ can remain finite as $T \to \infty$. Because (2) is similar to a vector autoregressive (VAR) model, our main interest is in (1). If Assumption 7 holds, then we say that $n^*, T \to \infty$ simultaneously. These assumptions are similar to those in Yu, de Jong, and Lee (2008).

### 14.3.1 Consistency

For the log-likelihood function Equation 14.10 divided by the effective sample size $n^*T$, we have the corresponding $Q_{n,T}(\theta) = \mathrm{E}\max_{\mathbf{c}_n} \frac{1}{n^*T} \ln L_{n,T}(\theta, \mathbf{c}_n)$.

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

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

Hence,[11]

$$\begin{aligned} Q_{n,T}(\theta) &= \frac{1}{n^*T}\mathrm{E}\ln L_{n,T}(\theta) \\ &= -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 - \frac{n-n^*}{n^*}\ln(1-\lambda) + \frac{1}{n^*}\ln|S_n(\lambda)| \\ &\quad - \frac{1}{2\sigma^2}\frac{1}{n^*T}\mathrm{E}\left(\sum_{t=1}^{T}\tilde{V}'_{nt}(\theta)J_n^*\tilde{V}_{nt}(\theta)\right). \end{aligned} \tag{14.11}$$

It is shown in Appendix D.2 that, under Assumptions 1–7, $\frac{1}{n^*T}\ln L_{n,T}(\theta) - Q_{n,T}(\theta) \xrightarrow{p} 0$ uniformly in $\theta \in \Theta$ and $Q_{n,T}(\theta)$ is uniformly equicontinuous for $\theta \in \Theta$. For the identification, denote the information matrix $\Sigma_{\theta_0,nT} = -\mathrm{E}(\frac{1}{n^*T}\frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'})$. If $\Sigma_{\theta_0,nT}$ is nonsingular and $-\mathrm{E}(\frac{1}{n^*T}\frac{\partial^2 \ln L_{n,T}(\theta)}{\partial\theta\partial\theta'})$ has full rank for $\theta$ in some neighborhood $N(\theta_0)$ of $\theta_0$, the parameters are locally identified (see Rothenberg 1971). Denote $\mathcal{H}_{nT} = \frac{1}{n^*T}\sum_{t=1}^{T}(\tilde{Z}_{nt}, G_n\tilde{Z}_{nt}\delta_0)'J_n^*(\tilde{Z}_{nt}, G_n\tilde{Z}_{nt}\delta_0)$ and $G_n^* = W_n^*S_n^{*-1}$. Using Lemma 15 in Yu, de Jong, and Lee (2008),

$$\begin{aligned} \Sigma_{\theta_0,nT} &= \frac{1}{\sigma_0^2}\begin{pmatrix} \mathrm{E}\mathcal{H}_{nT} & \mathbf{0}_{(k+3)\times 1} \\ \mathbf{0}_{1\times(k+3)} & 0 \end{pmatrix} \\ &\quad + \begin{pmatrix} \mathbf{0}_{(k+2)\times(k+2)} & \mathbf{0}_{(k+2)\times 1} & \mathbf{0}_{(k+2)\times 1} \\ \mathbf{0}_{1\times(k+2)} & \frac{1}{n^*}\left[tr(G_n^{*\prime}G_n^*) + tr(G_n^{*2})\right] & \frac{1}{\sigma_0^2 n^*}tr(G_n^*) \\ \mathbf{0}_{1\times(k+2)} & \frac{1}{\sigma_0^2 n^*}tr(G_n^*) & \frac{1}{2\sigma_0^4} \end{pmatrix} \\ &\quad + O\left(\frac{1}{T}\right), \end{aligned} \tag{14.12}$$

which is nonsingular if $\mathrm{E}\mathcal{H}_{nT}$ is nonsingular or $\frac{1}{n^*}[tr(G_n^{*\prime}G_n^*) + tr(G_n^{*2}) - \frac{2tr^2(G_n^*)}{n^*}]$ is positive (see Appendix D.1). Also, its rank does not change in a small neighborhood of $\theta_0$ (see Equation 14.49).

When $\lim_{T\to\infty}\mathrm{E}\mathcal{H}_{nT}$ is nonsingular, the parameters are identified.

**Theorem 14.1** *Under Assumptions 1–7, if* $\lim_{T\to\infty}\mathrm{E}\mathcal{H}_{nT}$ *is nonsingular,* $\theta_0$ *is identified and* $\hat{\theta}_{nT} \xrightarrow{p} \theta_0$.

**Proof** *See Appendix D.2.* ■

When $\lim_{T\to\infty}\mathrm{E}\mathcal{H}_{nT}$ is singular, identification can still be obtained from the following theorem. Denote $\sigma_n^2(\lambda) = \frac{\sigma_0^2}{n^*}tr(S_n'^{-1}S_n'(\lambda)J_n^*S_n(\lambda)S_n^{-1})$.

[11] Because $W_n = R_n\varpi_n R_n^{-1}$, $|S_n(\lambda)| = |I_n - \lambda\varpi_n| = (1-\lambda)^{m_n}\prod_{j=m_n+1}^{n}(1-\lambda\varpi_{nj})$. Therefore, $\frac{1}{n^*}\ln|S_n(\lambda)| - \frac{n-n^*}{n^*}\ln(1-\lambda) = \frac{1}{n^*}\sum_{j=m_n+1}^{n}(1-\lambda\varpi_{nj})$ shows that the division by $n^*$ is proper.

**Theorem 14.2** *Under Assumptions 1–7, if* $\lim_{n^*\to\infty}(\frac{1}{n^*}\ln|\sigma_0^2 S_n^{*-1\prime}S_n^{*-1}| - \frac{1}{n^*}\ln|\times \sigma_n^2(\lambda)S_n^{*\prime-1}(\lambda)S_n^{*-1}(\lambda)|) \neq 0$ *for* $\lambda \neq \lambda_0$, *then* $\theta_0$ *is identified*[12] *and* $\hat{\theta}_{nT} \xrightarrow{p} \theta_0$.

**Proof** See Appendix D.3. ■

### 14.3.2 Asymptotic Distribution

As $Z_{nt} = (Y_{n,t-1}, W_nY_{n,t-1}, X_{nt})$, we can decompose $(I_n - W_n)\tilde{Z}_{nt}$ such that

$$(I_n - W_n)\tilde{Z}_{nt} = (I_n - W_n)\tilde{Z}_{nt}^{(c)} - ((I_n - W_n)\bar{U}_{nT,-1},\ (I_n - W_n)W_n\bar{U}_{nT,-1},\ \mathbf{0}_{n\times k}), \tag{14.13}$$

where $\tilde{Z}_{nt}^{(c)} = ((\approx{\mathcal{X}}_{n,t-1} + U_{n,t-1}),\ (W_n\approx{\mathcal{X}}_{n,t-1} + W_nU_{n,t-1}),\ \tilde{X}_{nt})$ with $\approx{\mathcal{X}}_{n,t-1} = \mathcal{X}_{n,t-1} - \bar{\mathcal{X}}_{nT,-1}$, $\mathcal{X}_{nt} \equiv \sum_{h=0}^{\infty} B_n^h S_n^{-1} X_{n,t-h}$ and $U_{nt} \equiv \sum_{h=0}^{\infty} B_n^h S_n^{-1} V_{n,t-h}$. Hence, $(I_n - W_n)\tilde{Z}_{nt}$ has two components: one is $(I_n - W_n)\tilde{Z}_{nt}^{(c)}$, which is uncorrelated with $V_{nt}$; the remaining one can be correlated with $V_{nt}$ when $t \leq T-1$. Here, after the data transformation by $I_n - W_n$, the unstable or explosive components and time component in $\tilde{Z}_{nt}$ are all eliminated. Therefore, from Equation 14.45, the score can be decomposed into two parts such that

$$\frac{1}{\sqrt{n^*T}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta} = \frac{1}{\sqrt{n^*T}}\frac{\partial \ln L_{n,T}^{(c)}(\theta_0)}{\partial\theta} - \Delta_{nT}, \tag{14.14}$$

where

$$\frac{1}{\sqrt{n^*T}}\frac{\partial \ln L_{n,T}^{(c)}(\theta_0)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma_0^2}\frac{1}{\sqrt{n^*T}}\sum_{t=1}^{T}\tilde{Z}_{nt}^{(c)\prime}J_n^*V_{nt} \\ \frac{1}{\sigma_0^2}\frac{1}{\sqrt{n^*T}}\sum_{t=1}^{T}(G_n\tilde{Z}_{nt}^{(c)}\delta_0)'J_n^*V_{nt} + \frac{1}{\sigma_0^2}\frac{1}{\sqrt{n^*T}}\sum_{t=1}^{T}(V_{nt}'G_n'J_n^*V_{nt} - \sigma_0^2 tr\, G_n^*) \\ \frac{1}{2\sigma_0^4}\frac{1}{\sqrt{n^*T}}\sum_{t=1}^{T}(V_{nt}'J_n^*V_{nt} - n^*\sigma_0^2) \end{pmatrix}, \tag{14.15}$$

[12] For our asymptotic analysis, finite $n^*$ is allowed as long as $T$ is tending to infinity, even though that is not an interesting case for SAR models. When $n^*$ is finite, the condition is $\frac{1}{n^*}\ln|\sigma_0^2 S_n^{*-1\prime}S_n^{*-1}| - \frac{1}{n^*}\ln|\sigma_n^2(\lambda)S_n^{*\prime-1}(\lambda)S_n^{*-1}(\lambda)| \neq 0$ for $\lambda \neq \lambda_0$.

and

$$\Delta_{nT} = \sqrt{\frac{n^*}{T}} \begin{pmatrix} \frac{1}{\sigma_0^2}\frac{T}{n^*}(J_n^*\bar{U}_{nT,-1},\ J_n^*W_n\bar{U}_{nT,-1},\mathbf{0})'\bar{V}_{nT} \\ \frac{1}{\sigma_0^2}\frac{T}{n^*}(J_n^*G_n(\bar{U}_{nT,-1},W_n\bar{U}_{nT,-1},\mathbf{0})\delta_0)'\bar{V}_{nT} + \frac{1}{\sigma_0^2}\frac{T}{n^*}\bar{V}'_{nT}G'_nJ_n^*\bar{V}_{nT} \\ \frac{1}{2\sigma_0^4}\frac{T}{n^*}\bar{V}'_{nT}J_n^*\bar{V}_{nT} \end{pmatrix}. \tag{14.16}$$

Similarly to Yu, de Jong, and Lee (2008), the variance matrix of $\frac{1}{\sqrt{n^*T}}\frac{\partial \ln L_{n,T}^{(c)}(\theta_0)}{\partial\theta}$ is equal to

$$\mathrm{E}\left(\frac{1}{\sqrt{n^*T}}\frac{\partial \ln L_{n,T}^{(c)}(\theta_0)}{\partial\theta}\cdot\frac{1}{\sqrt{n^*T}}\frac{\partial \ln L_{n,T}^{(c)}(\theta_0)}{\partial\theta'}\right) = \Sigma_{\theta_0,nT} + \Omega_{\theta_0,n} + O(T^{-1}), \tag{14.17}$$

where $\Sigma_{\theta_0,nT}$ is in Equation 14.12 and

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

is a symmetric matrix with $\mu_4$ being the fourth moment of $v_{it}$. When $V_{nt}$ is normally distributed, $\Omega_{\theta_0,n} = 0$ because $\mu_4 - 3\sigma_0^4 = 0$. Denote $\Sigma_{\theta_0} = \lim_{T\to\infty}\Sigma_{\theta_0,nT}$ and $\Omega_{\theta_0} = \lim_{T\to\infty}\Omega_{\theta_0,n}$. The asymptotic distribution of $\frac{1}{\sqrt{n^*T}}\frac{\partial \ln L_{n,T}^{(c)}(\theta_0)}{\partial\theta}$ can be derived from the central limit theorem for martingale difference arrays (Lemma 14.3). For the term $\Delta_{nT}$, from Equation 14.36 in Lemma 14.1 and Equation 14.38 in Lemma 14.2, $\Delta_{nT} = \sqrt{\frac{n^*}{T}}a_{\theta_0,n} + O(\sqrt{\frac{n^*}{T^3}}) + O_p(\frac{1}{\sqrt{T}})$ where

$$a_{\theta_0,n} = \begin{pmatrix} \frac{1}{n^*}tr\left(\left(J_n^*\sum_{h=0}^{\infty}B_n^h\right)S_n^{-1}\right) \\ \frac{1}{n^*}tr\left(W_n\left(J_n^*\sum_{h=0}^{\infty}B_n^h\right)S_n^{-1}\right) \\ \mathbf{0}_{k\times 1} \\ \frac{1}{n^*}\gamma_0 tr(G_n\left(J_n^*\sum_{h=0}^{\infty}B_n^h\right)S_n^{-1}) + \frac{1}{n^*}\rho_0 tr(G_nW_n\left(J_n^*\sum_{h=0}^{\infty}B_n^h\right)S_n^{-1}) + \frac{1}{n^*}trG_n^* \\ \frac{1}{2\sigma_0^2} \end{pmatrix} \tag{14.18}$$

is $O(1)$. It is shown in Appendix D.4 that, under Assumptions 1–7, $\frac{1}{\sqrt{n^*T}} \frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta} + \Delta_{nT} \xrightarrow{d} N(0, \Sigma_{\theta_0} + \Omega_{\theta_0})$.

To get the asymptotic distribution of the estimates, we need the following additional assumption.

**Assumption 8.** $\lim_{T\to\infty} \mathrm{E}\mathcal{H}_{nT}$ is nonsingular or $\lim_{n\to\infty} \frac{1}{n^*}[tr(G_n^{*\prime} G_n^*) + tr(G_n^{*2}) - \frac{2tr^2(G_n^*)}{n^*}] > 0$.

Assumption 8 is a condition for the nonsingularity of the limit of the information matrix $\Sigma_{\theta_0}$. When $\lim_{T\to\infty} \mathrm{E}\mathcal{H}_{nT}$ is singular,[13] as long as we have $\lim_{n\to\infty} \frac{1}{n^*}[tr(G_n^{*\prime} G_n^*) + tr((G_n^*)^2) - \frac{2tr^2(G_n^*)}{n^*}] > 0$, the information matrix $\Sigma_{\theta_0}$ is still nonsingular (see Appendix D.1). Hence, for the second order derivatives of the log-likelihood function, under Assumption 1–8, we have $\frac{1}{n^*T}\frac{\partial^2 \ln L_{n,T}(\theta)}{\partial\theta\partial\theta'} - \frac{1}{n^*T}\frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'} = \|\theta - \theta_0\| \cdot O_p(1)$, and $\frac{1}{n^*T}\frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'} - \frac{\partial^2 Q_{n,T}(\theta_0)}{\partial\theta\partial\theta'} = O_p(\frac{1}{\sqrt{n^*T}})$ from Appendix C.3. Thus, we have the following theorem for the asymptotic distribution of $\hat{\theta}_{nT}$.

**Theorem 14.3** *Under Assumptions 1–8,*

$$\sqrt{n^*T}(\hat{\theta}_{nT} - \theta_0) + \sqrt{\frac{n^*}{T}} b_{\theta_0,nT} + O_p\left(\max\left(\sqrt{\frac{n^*}{T^3}}, \sqrt{\frac{1}{T}}\right)\right)$$
$$\xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0} + \Omega_{\theta_0})\Sigma_{\theta_0}^{-1}), \tag{14.19}$$

*where* $b_{\theta_0,nT} = \Sigma_{\theta_0,nT}^{-1} a_{\theta_0,n}$ *is* $O(1)$.

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

*When* $\frac{n^*}{T} \to c < \infty$, $\sqrt{n^*T}(\hat{\theta}_{nT} - \theta_0) + \sqrt{c} b_{\theta_0,nT} \xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0} + \Omega_{\theta_0})\Sigma_{\theta_0}^{-1})$.

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

**Proof** *See Appendix D.4.* ■

### 14.3.3 Bias Correction

From Equation (14.19), the QML estimator has the leading bias $-\frac{1}{T} b_{\theta_0,nT}$ where $b_{\theta_0,nT} = \Sigma_{\theta_0,nT}^{-1} \cdot a_{\theta_0,n}$ and the confidence interval will not be centered when $\frac{n^*}{T} \to c < \infty$. Furthermore, when $T$ is relatively smaller than $n^*$, the presence of $b_{\theta_0,nT}$ causes $\hat{\theta}_{nT}$ to have a degenerate distribution. An analytical bias reduction procedure can be used to correct this bias of the estimate. Define the

[13] The $\lim_{T\to\infty} \mathrm{E}\mathcal{H}_{nT}$ can be singular if, for example, $\delta_0 = \mathbf{0}$.

bias corrected estimator as

$$\hat{\theta}^1_{nT} = \hat{\theta}_{nT} - \frac{\hat{B}_{nT}}{T}, \tag{14.20}$$

where, from Theorem 14.3, $\hat{B}_{nT} = [-\Sigma^{-1}_{\theta,nT} \cdot a_{\theta,n}]|_{\theta=\hat{\theta}_{nT}}$. We show that when $n^*/T^3 \to 0$, $\hat{\theta}^1_{nT}$ is $\sqrt{n^*T}$ consistent and asymptotically centered normal even when $n^*/T \to \infty$.

For the asymptotic properties of the bias corrected estimator, we need the following additional assumption.

**Assumption 9.** $\sum_{h=0}^{\infty} B_n^h(\theta)$ and $\sum_{h=1}^{\infty} h B_n^{h-1}(\theta)$ are uniformly bounded in either row sum or column sums, uniformly in a neighborhood of $\theta_0$.

Assumption 9 can be justified through Lemma 14.5 in Appendix B. Our result for the bias corrected estimator is as follows.

**Theorem 14.4** *If* $\frac{n^*}{T^3} \to 0$, *under Assumptions 1–9,* $\sqrt{n^*T}(\hat{\theta}^1_{nT} - \theta_0) \xrightarrow{d} N(0, \Sigma^{-1}_{\theta_0} + \Sigma^{-1}_{\theta_0}\Omega_{\theta_0}\Sigma^{-1}_{\theta_0})$.

**Proof** *See Appendix D.5.* ■

Hence, if $T$ grows faster than $n^{*1/3}$, the analytical bias adjusted estimator is asymptotically normal and centered properly around $\theta_0$. For the case $\frac{n}{T} \to c$, $\hat{\theta}^1_{nT}$ has removed the asymptotic bias $b_{\theta_0,nT}$. Note that $\frac{n}{T} \to c$ implies $T/n^{*1/3} \to \infty$. For the case $\frac{n^*}{T} \to \infty$, as long as $T/n^{*1/3} \to \infty$, $\hat{\theta}^1_{nT}$ is $\sqrt{n^*T}$ consistent, which is also an improvement upon the $T$ consistency of $\hat{\theta}_{nT}$. Thus, $\hat{\theta}^1_{nT}$ might have better performance, especially when $n^*$ is much larger than $T$.

### 14.3.4 Testing

For the unified transformation approach with a bias correction, we have $\sqrt{n^*T}(\hat{\theta}^1_{nT} - \theta_0) \xrightarrow{d} N(0, \Sigma^{-1}_{\theta_0} + \Sigma^{-1}_{\theta_0}\Omega_{\theta_0}\Sigma^{-1}_{\theta_0})$. Hence, we can use the bias corrected estimate $\hat{\theta}^1_{nT}$ for the statistical inference of $\gamma_0 + \rho_0 + \lambda_0$. Let $\Sigma_{\hat{\theta}^1_{nT},nT}$ and $\Omega_{\hat{\theta}^1_{nT},n}$ be consistent estimates for $\Sigma_{\theta_0,nT}$ and $\Omega_{\theta_0,n}$. We can construct $t$-statistic to test the null of spatial co-integration, i.e., $\gamma_0 + \rho_0 + \lambda_0 = 1$. Denote $r = (1, 1, 0_{1\times k_x}, 1, 0)'$. With $\theta = (\gamma, \rho, \beta', \lambda, \sigma^2)$, we are testing $r'\hat{\theta}^1_{nT} = 1$. The test statistic is

$$t = \frac{\sqrt{n^*T} \cdot (r'\hat{\theta}^1_{nT} - 1)}{\sqrt{r'(\Sigma^{-1}_{\hat{\theta}^1_{nT},nT} + \Sigma^{-1}_{\hat{\theta}^1_{nT},nT}\Omega_{\hat{\theta}^1_{nT},n}\Sigma^{-1}_{\hat{\theta}^1_{nT},nT})r}} \xrightarrow{d} N(0, 1), \tag{14.21}$$

because $\sqrt{n^*T} \cdot (r'\hat{\theta}^1_{nT} - 1) \xrightarrow{d} N(0, r'(\Sigma^{-1}_{\theta_0} + \Sigma^{-1}_{\theta_0}\Omega_{\theta_0}\Sigma^{-1}_{\theta_0})r)$ and $\Sigma^{-1}_{\hat{\theta}^1_{nT},nT} + \Sigma^{-1}_{\hat{\theta}^1_{nT},nT}\Omega_{\hat{\theta}^1_{nT},n}\Sigma^{-1}_{\hat{\theta}^1_{nT},nT} \xrightarrow{p} \Sigma^{-1}_{\theta_0} + \Sigma^{-1}_{\theta_0}\Omega_{\theta_0}\Sigma^{-1}_{\theta_0}$. We present a simulation in the next section to investigate the finite sample performance of the test statistic in terms of its significance level and power, under one-sided or two-sided tests.

## 14.4 Monte Carlo Results

We conduct a Monte Carlo experiment to evaluate the performance of the bias corrected MLE of this unified approach and compare them with other estimation methods under different DGPs:

$$\begin{aligned} \text{A}: \quad Y_{nt} &= \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} \\ &\quad + V_{nt}, \ \lambda_0 + \gamma_0 + \rho_0 < 1, \end{aligned} \tag{14.22}$$

$$\begin{aligned} \text{B}: \quad Y_{nt} &= \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n \\ &\quad + V_{nt}, \ \lambda_0 + \gamma_0 + \rho_0 < 1, \end{aligned} \tag{14.23}$$

$$\begin{aligned} \text{C}: \quad Y_{nt} &= \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} \\ &\quad + V_{nt}, \ \lambda_0 + \gamma_0 + \rho_0 = 1, \ \gamma_0 \neq 1, \end{aligned} \tag{14.24}$$

$$\begin{aligned} \text{D}: \quad Y_{nt} &= \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n \\ &\quad + V_{nt}, \ \lambda_0 + \gamma_0 + \rho_0 = 1, \ \gamma_0 \neq 1. \end{aligned} \tag{14.25}$$

The DGPs A and B are stable SDPD models with or without time dummy effects. The C and D are spatial co-integrated SDPD models with or without time dummy effects. We will also consider subsequently DGPs with explosive roots in E and F. We generate samples using $\theta_0 = (0.2, 0.2, 1, 0.2, 1)'$ for the stable cases and $\theta_0 = (0.4, 0.2, 1, 0.4, 1)'$ for the spatial co-integration cases where $\theta_0 = (\gamma_0, \rho_0, \beta_0', \lambda_0, \sigma_0^2)'$, and $X_{nt}$, $\mathbf{c}_{n0}$, $\alpha_{T0} = (\alpha_1, \alpha_2, \cdots, \alpha_T)'$ and $V_{nt}$ are generated from independent normal distributions.[14] The spatial weights matrix we use is a block diagonal matrix formed by a row-normalized queen matrix.[15] We use $T = 10, 50$, and $n = 18, 54$.

For each set of generated sample observations, we use two methods: one is the corresponding estimation method without any transformation when the model does not have time dummies, or using the deviation from group mean transformation when the model includes time dummies, and the other is the unified transformation method. We obtain the MLE $\hat{\theta}_{nT}$, construct the bias corrected estimator $\hat{\theta}^1_{nT}$ and evaluate the bias $\hat{\theta}^1_{nT} - \theta_0$. We do this 1000

[14] We generated the data with $20 + T$ periods and then take the last $T$ periods as our sample. And the initial value is generated as $N(0, I_n)$ in the simulation.

[15] We choose the spatial weights matrix such that it contains unit eigenvalues. We use the block diagonal matrix where each block uses the same weights matrix. By increasing the number of blocks, the number of unit eigenvalues of the block diagonal matrix will also increase, but the percentage remains a constant. In our simulation, when $n = 18$, $n^* = 16$; when $n = 54$, $n^* = 48$.

times. We also compare the empirical standard deviation (SD) and the empirical mean square error (RMSE) of these 1000 estimators. Also, a coverage probability (CP) is reported.[16] With different values of $n$ and $T$, finite sample properties of the bias corrected estimators[17] are summarized in Tables 14.1 and 14.2. Table 14.1 presents the results for the stable SDPD models; and Table 14.2 is for the spatial co-integration cases.

Because the unified transformation method will lose more degrees of freedom than the other methods, we expect less precision for the estimates from the unified transformation approach than the others. For our MC design with blocks of the queen matrix, the use of the unified transformation will result in more loss of degrees of freedom than that of the deviation from the group mean transformation for the models with time effects. It is of interest to see that the estimators by the unified transformation method perform well, and they are a little bit worse than the corresponding estimators in the loss of precision. All the estimates have small biases. The CPs are adequate except for some cases with small $T = 10$.

The unified transformation method would be of more interest for the explosive roots case. We conduct a simulation to check the performance of the unified estimator when the DGP is explosive:

$$\text{E}: \quad Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + V_{nt}, \ \lambda_0 + \gamma_0 + \rho_0 > 1, \tag{14.26}$$

$$\text{F}: \quad Y_{nt} = \lambda_0 W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + V_{nt}, \ \lambda_0 + \gamma_0 + \rho_0 > 1, \tag{14.27}$$

where $\theta_0 = (0.4, 0.4, 1, 0.4, 1)'$. Finite sample properties of both estimators are summarized in Table 14.3 for the bias corrected estimators. We can see that even though we have explosive roots in the DGP, the unified approach can still yield estimators with good finite sample performances, i.e., the biases are small and the CPs are adequate. However, if we use the QMLE without any transformation when the model does not have time dummies, or use the deviation from group mean transformation when the model includes time dummies, the estimates' Biases, SD, and RMSE become very large and the CPs are nearly zero when $T$ is large.

Finally, we present the simulation result of the size and power of the hypothesis testing of spatial co-integration, i.e., $H_0 : \lambda_0 + \gamma_0 + \rho_0 = 1$. We run 1000 repetitions to calculate the power for $n = 54$ and $T = 10$ or 50, where the power is obtained with a 1% or 5% significance levels. We first use the

[16] The coverage probability is obtained by using the estimated analytical standard errors of the estimators in each repitition.

[17] For the estimators before bias correction, they have a larger bias than the corresponding bias corrected estimators. As the comparison between the unified transformation estimators and the corresponding estimators are similar to the counterpart of bias corrected estimators, we do not report the tables of results to save space.

**TABLE 14.1**

Performance of Estimators When the DGP Is Stable

| | $T$ | $n$ | Estimator | | $\gamma$ | $\rho$ | $\beta$ | $\lambda$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|---|---|
| **No Time Dummy in the DGP (Equation 14.22):** | | | | | | | | | |
| (1) | 10 | 54 | A | Bias | −0.0010 | −0.0015 | 0.0016 | −0.0086 | −0.0288 |
| | | | | SD | 0.0320 | 0.0659 | 0.0451 | 0.0517 | 0.0592 |
| | | | | RMSE | 0.0439 | 0.0926 | 0.0619 | 0.0721 | 0.0860 |
| | | | | CP | 0.9400 | 0.9040 | 0.9290 | 0.9410 | 0.8500 |
| | 10 | 54 | Unified | Bias | −0.0010 | −0.0002 | 0.0001 | −0.0108 | −0.0305 |
| | | | | SD | 0.0375 | 0.1409 | 0.0494 | 0.1139 | 0.0674 |
| | | | | RMSE | 0.0515 | 0.1908 | 0.0676 | 0.1535 | 0.0981 |
| | | | | CP | 0.9250 | 0.9360 | 0.9300 | 0.9780 | 0.8470 |
| (2) | 50 | 18 | A | Bias | −0.0007 | −0.0011 | −0.0009 | −0.0025 | −0.0043 |
| | | | | SD | 0.0235 | 0.0476 | 0.0337 | 0.0393 | 0.0470 |
| | | | | RMSE | 0.0317 | 0.0652 | 0.0458 | 0.0536 | 0.0647 |
| | | | | CP | 0.9570 | 0.9410 | 0.9480 | 0.9510 | 0.9230 |
| | 50 | 18 | Unified | Bias | −0.0006 | 0.0004 | −0.0025 | −0.0077 | −0.0062 |
| | | | | SD | 0.0274 | 0.1034 | 0.0370 | 0.0873 | 0.0534 |
| | | | | RMSE | 0.0371 | 0.1414 | 0.0505 | 0.1176 | 0.0736 |
| | | | | CP | 0.9450 | 0.9400 | 0.9440 | 0.9470 | 0.9250 |
| (3) | 50 | 54 | A | Bias | −0.0002 | −0.0009 | 0.0000 | −0.0007 | −0.0015 |
| | | | | SD | 0.0136 | 0.0275 | 0.0195 | 0.0227 | 0.0272 |
| | | | | RMSE | 0.0182 | 0.0370 | 0.0259 | 0.0311 | 0.0377 |
| | | | | CP | 0.9570 | 0.9500 | 0.9620 | 0.9320 | 0.9320 |
| | 50 | 54 | Unified | Bias | 0.0002 | 0.0018 | 0.0001 | −0.0015 | −0.0017 |
| | | | | SD | 0.0158 | 0.0596 | 0.0214 | 0.0504 | 0.0309 |
| | | | | RMSE | 0.0211 | 0.0800 | 0.0284 | 0.0684 | 0.0424 |
| | | | | CP | 0.9480 | 0.9520 | 0.9600 | 0.9420 | 0.9320 |
| **Time Dummy in the DGP (Equation 14.23):** | | | | | | | | | |
| (1) | 10 | 54 | B | Bias | −0.0036 | −0.0000 | 0.0016 | −0.0066 | −0.0283 |
| | | | | SD | 0.0323 | 0.0700 | 0.0455 | 0.0550 | 0.0597 |
| | | | | RMSE | 0.0452 | 0.0987 | 0.0632 | 0.0769 | 0.0871 |
| | | | | CP | 0.9190 | 0.9160 | 0.9260 | 0.9280 | 0.8620 |
| | 10 | 54 | Unified | Bias | −0.0043 | −0.0010 | 0.0008 | −0.0086 | −0.0312 |
| | | | | SD | 0.0375 | 0.1405 | 0.0495 | 0.1140 | 0.0673 |
| | | | | RMSE | 0.0519 | 0.1950 | 0.0690 | 0.1532 | 0.0968 |
| | | | | CP | 0.9270 | 0.9170 | 0.9200 | 0.9760 | 0.8600 |
| (2) | 50 | 18 | B | Bias | −0.0009 | −0.0025 | −0.0024 | −0.0013 | −0.0046 |
| | | | | SD | 0.0242 | 0.0595 | 0.0347 | 0.0498 | 0.0484 |
| | | | | RMSE | 0.0330 | 0.0804 | 0.0477 | 0.0683 | 0.0662 |
| | | | | CP | 0.9510 | 0.9460 | 0.9430 | 0.9310 | 0.9380 |
| | 50 | 18 | Unified | Bias | 0.0005 | 0.0074 | −0.0030 | −0.0023 | −0.0253 |
| | | | | SD | 0.0273 | 0.1033 | 0.0370 | 0.0873 | 0.0523 |
| | | | | RMSE | 0.0375 | 0.1411 | 0.0507 | 0.1178 | 0.0732 |
| | | | | CP | 0.9440 | 0.9420 | 0.9410 | 0.9480 | 0.9330 |
| (3) | 50 | 54 | B | Bias | 0.0002 | −0.0008 | 0.0002 | 0.0007 | −0.0015 |
| | | | | SD | 0.0136 | 0.0290 | 0.0196 | 0.0241 | 0.0269 |
| | | | | RMSE | 0.0185 | 0.0395 | 0.0268 | 0.0334 | 0.0373 |
| | | | | CP | 0.9470 | 0.9410 | 0.9360 | 0.9270 | 0.9400 |
| | 50 | 54 | Unified | Bias | 0.0004 | 0.0003 | −0.0001 | −0.0033 | −0.0031 |
| | | | | SD | 0.0158 | 0.0596 | 0.0214 | 0.0504 | 0.0309 |
| | | | | RMSE | 0.0215 | 0.0811 | 0.0291 | 0.0693 | 0.0421 |
| | | | | CP | 0.9500 | 0.9410 | 0.9430 | 0.9430 | 0.9400 |

*Note:* $\theta_0 = (0.2, 0.2, 1, 0.2, 1)'$

**TABLE 14.2**

Performance of Estimators When the DGP Is Spatial Co-integrated

| | *T* | *n* | **Estimator** | | $\gamma$ | $\rho$ | $\beta$ | $\lambda$ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|---|---|
| **No Time Dummy in the DGP (Equation 14.24):** | | | | | | | | | |
| (1) | 10 | 54 | C | Bias | 0.0065 | 0.0518 | 0.0073 | 0.0007 | −0.0330 |
| | | | | SD | 0.0314 | 0.0531 | 0.0452 | 0.0405 | 0.0594 |
| | | | | RMSE | 0.0447 | 0.0899 | 0.0626 | 0.0589 | 0.0876 |
| | | | | CP | 0.9090 | 0.7640 | 0.9190 | 0.9160 | 0.8330 |
| | 10 | 54 | Unified | Bias | −0.0023 | 0.0023 | −0.0005 | −0.0173 | −0.0354 |
| | | | | SD | 0.0354 | 0.1353 | 0.0490 | 0.1113 | 0.0659 |
| | | | | RMSE | 0.0494 | 0.1843 | 0.0674 | 0.1560 | 0.0982 |
| | | | | CP | 0.9180 | 0.9300 | 0.9240 | 0.9230 | 0.8350 |
| (2) | 50 | 18 | C | Bias | 0.0001 | 0.0046 | −0.0006 | −0.0046 | −0.0039 |
| | | | | SD | 0.0224 | 0.0365 | 0.0338 | 0.0314 | 0.0473 |
| | | | | RMSE | 0.0303 | 0.0495 | 0.0460 | 0.0425 | 0.0651 |
| | | | | CP | 0.9550 | 0.9450 | 0.9470 | 0.9530 | 0.9280 |
| | 50 | 18 | Unified | Bias | −0.0013 | −0.0005 | −0.0024 | −0.0078 | −0.0062 |
| | | | | SD | 0.0249 | 0.0969 | 0.0367 | 0.0851 | 0.0525 |
| | | | | RMSE | 0.0337 | 0.1306 | 0.0501 | 0.1171 | 0.0725 |
| | | | | CP | 0.9460 | 0.9470 | 0.9420 | 0.9520 | 0.9280 |
| (3) | 50 | 54 | C | Bias | 0.0003 | 0.0033 | 0.0003 | −0.0035 | −0.0010 |
| | | | | SD | 0.0129 | 0.0210 | 0.0195 | 0.0181 | 0.0274 |
| | | | | RMSE | 0.0175 | 0.0287 | 0.0260 | 0.0250 | 0.0380 |
| | | | | CP | 0.9510 | 0.9380 | 0.9620 | 0.9360 | 0.9340 |
| | 50 | 54 | Unified | Bias | −0.0002 | 0.0002 | 0.0002 | −0.0002 | −0.0016 |
| | | | | SD | 0.0143 | 0.0558 | 0.0212 | 0.0491 | 0.0304 |
| | | | | RMSE | 0.0193 | 0.0757 | 0.0282 | 0.0675 | 0.0418 |
| | | | | CP | 0.9470 | 0.9490 | 0.9600 | 0.9430 | 0.9310 |
| **Time Dummy in the DGP (Equation 14.25):** | | | | | | | | | |
| (1) | 10 | 54 | D | Bias | 0.0030 | 0.0483 | 0.0059 | 0.0006 | −0.0323 |
| | | | | SD | 0.0316 | 0.0557 | 0.0456 | 0.0435 | 0.0597 |
| | | | | RMSE | 0.0450 | 0.0917 | 0.0638 | 0.0631 | 0.0882 |
| | | | | CP | 0.8550 | 0.6790 | 0.8880 | 0.8600 | 0.8500 |
| | 10 | 54 | Unified | Bias | −0.0054 | 0.0020 | −0.0000 | −0.0160 | −0.0364 |
| | | | | SD | 0.0354 | 0.1349 | 0.0490 | 0.1114 | 0.0658 |
| | | | | RMSE | 0.0501 | 0.1871 | 0.0686 | 0.1559 | 0.0969 |
| | | | | CP | 0.9160 | 0.9110 | 0.9240 | 0.9220 | 0.8510 |
| (2) | 50 | 18 | D | Bias | −0.0004 | 0.0017 | −0.0024 | −0.0039 | −0.0050 |
| | | | | SD | 0.0226 | 0.0441 | 0.0347 | 0.0404 | 0.0483 |
| | | | | RMSE | 0.0308 | 0.0594 | 0.0477 | 0.0560 | 0.0661 |
| | | | | CP | 0.9560 | 0.9480 | 0.9400 | 0.9310 | 0.9340 |
| | 50 | 18 | Unified | Bias | −0.0007 | 0.0032 | −0.0029 | −0.0024 | −0.0059 |
| | | | | SD | 0.0248 | 0.0967 | 0.0367 | 0.0851 | 0.0525 |
| | | | | RMSE | 0.0339 | 0.1323 | 0.0503 | 0.1171 | 0.0721 |
| | | | | CP | 0.9530 | 0.9450 | 0.9410 | 0.9500 | 0.9290 |
| (3) | 50 | 54 | D | Bias | 0.0004 | 0.0015 | 0.0003 | −0.0017 | −0.0014 |
| | | | | SD | 0.0130 | 0.0219 | 0.0197 | 0.0193 | 0.0276 |
| | | | | RMSE | 0.0175 | 0.0297 | 0.0269 | 0.0266 | 0.0375 |
| | | | | CP | 0.9360 | 0.8810 | 0.9240 | 0.9180 | 0.9430 |
| | 50 | 54 | Unified | Bias | −0.0001 | −0.0004 | −0.0000 | −0.0027 | −0.0030 |
| | | | | SD | 0.0143 | 0.0557 | 0.0212 | 0.0491 | 0.0304 |
| | | | | RMSE | 0.0195 | 0.0758 | 0.0288 | 0.0685 | 0.0415 |
| | | | | CP | 0.9420 | 0.9430 | 0.9370 | 0.9350 | 0.9400 |

*Note:* $\theta_0 = (0.4, 0.2, 1, 0.4, 1)'$

**TABLE 14.3**

Performance of Estimators When the DGP Is Explosive

| | *T* | *n* | **Estimator** | | γ | ρ | β | λ | $\sigma^2$ |
|---|---|---|---|---|---|---|---|---|---|
| **No Time Dummy in the DGP (Equation 14.26):** | | | | | | | | | |
| (1) | 10 | 54 | A | Bias | 0.0053 | 0.0395 | 0.0049 | −0.0336 | −0.0241 |
| | | | | SD | 0.0336 | 0.0584 | 0.0465 | 0.0422 | 0.0626 |
| | | | | RMSE | 0.0340 | 0.0705 | 0.0467 | 0.0540 | 0.0670 |
| | | | | CP | 0.9200 | 0.8890 | 0.9270 | 0.8630 | 0.9230 |
| | 10 | 54 | Unified | Bias | −0.0018 | 0.0031 | −0.0007 | −0.0196 | −0.0360 |
| | | | | SD | 0.0379 | 0.1382 | 0.0504 | 0.1201 | 0.0716 |
| | | | | RMSE | 0.0380 | 0.1382 | 0.0504 | 0.1217 | 0.0801 |
| | | | | CP | 0.9170 | 0.9310 | 0.9270 | 0.9100 | 0.8070 |
| (2) | 50 | 18 | A | Bias | ****** | ****** | 2.4973 | −0.0624 | ****** |
| | | | | SD | ****** | ****** | 264.78 | 0.2958 | ****** |
| | | | | RMSE | ****** | ****** | 264.79 | 0.3023 | ****** |
| | | | | CP | 0.0150 | 0.0090 | 0.0140 | 0.0130 | 0.0110 |
| | 50 | 18 | Unified | Bias | −0.0013 | −0.0013 | −0.0025 | −0.0088 | −0.0065 |
| | | | | SD | 0.0246 | 0.0931 | 0.0373 | 0.0878 | 0.0543 |
| | | | | RMSE | 0.0246 | 0.0931 | 0.0374 | 0.0882 | 0.0547 |
| | | | | CP | 0.9480 | 0.9440 | 0.9420 | 0.9260 | 0.9050 |
| (3) | 50 | 54 | A | Bias | ****** | ****** | −4.1263 | −0.0668 | ****** |
| | | | | SD | ****** | ****** | 724.64 | 0.3096 | ****** |
| | | | | RMSE | ****** | ****** | 724.66 | 0.3167 | ****** |
| | | | | CP | 0.0010 | 0.0000 | 0.0000 | 0.0010 | 0.0000 |
| | 50 | 54 | Unified | Bias | −0.0004 | −0.0006 | 0.0002 | −0.0005 | −0.0016 |
| | | | | SD | 0.0139 | 0.0557 | 0.0203 | 0.0510 | 0.0315 |
| | | | | RMSE | 0.0139 | 0.0557 | 0.0203 | 0.0510 | 0.0315 |
| | | | | CP | 0.9450 | 0.9380 | 0.9600 | 0.9250 | 0.9130 |
| **Time dummy in the DGP (Equation 14.27):** | | | | | | | | | |
| (1) | 10 | 54 | B | Bias | 0.0021 | 0.0386 | 0.0037 | −0.0305 | −0.0257 |
| | | | | SD | 0.0346 | 0.0635 | 0.0482 | 0.0462 | 0.0639 |
| | | | | RMSE | 0.0347 | 0.0743 | 0.0483 | 0.0554 | 0.0689 |
| | | | | CP | 0.9190 | 0.8870 | 0.9240 | 0.8880 | 0.9100 |
| | 10 | 54 | Unified | Bias | −0.0049 | 0.0029 | −0.0003 | −0.0191 | −0.0371 |
| | | | | SD | 0.0390 | 0.1435 | 0.0529 | 0.1200 | 0.0688 |
| | | | | RMSE | 0.0394 | 0.1435 | 0.0529 | 0.1216 | 0.0782 |
| | | | | CP | 0.9120 | 0.9060 | 0.9230 | 0.9090 | 0.8090 |
| (2) | 50 | 18 | B | Bias | ****** | ****** | −4.0205 | −0.0478 | ****** |
| | | | | SD | ****** | ****** | 105.34 | 0.2891 | ****** |
| | | | | RMSE | ****** | ****** | 105.41 | 0.2931 | ****** |
| | | | | CP | 0.1030 | 0.0640 | 0.0960 | 0.0790 | 0.0660 |
| | 50 | 18 | Unified | Bias | −0.0011 | 0.0014 | −0.0030 | −0.0033 | −0.0061 |
| | | | | SD | 0.0248 | 0.0972 | 0.0378 | 0.0885 | 0.0536 |
| | | | | RMSE | 0.0248 | 0.0972 | 0.0379 | 0.0885 | 0.0540 |
| | | | | CP | 0.9520 | 0.9390 | 0.9430 | 0.9260 | 0.9110 |
| (3) | 50 | 54 | B | Bias | ****** | ****** | −35.49 | −0.0596 | ****** |
| | | | | SD | ****** | ****** | 835.56 | 0.3128 | ****** |
| | | | | RMSE | ****** | ****** | 836.31 | 0.3184 | ****** |
| | | | | CP | 0.0020 | 0.0000 | 0.0010 | 0.0040 | 0.0000 |
| | 50 | 54 | Unified | Bias | −0.0001 | −0.0009 | −0.0001 | −0.0030 | −0.0031 |
| | | | | SD | 0.0143 | 0.0553 | 0.0215 | 0.0521 | 0.0308 |
| | | | | RMSE | 0.0143 | 0.0553 | 0.0215 | 0.0522 | 0.0310 |
| | | | | CP | 0.9410 | 0.9370 | 0.9380 | 0.9220 | 0.9270 |

*Note:* 1. $\theta_0$= (0.4, 0.4, 1, 0.4, 1)′.
2. ****** denotes an explosive number, which is of the order $10^{11}$ for the column of $\sigma^2$, and $10^5$ for other columns.

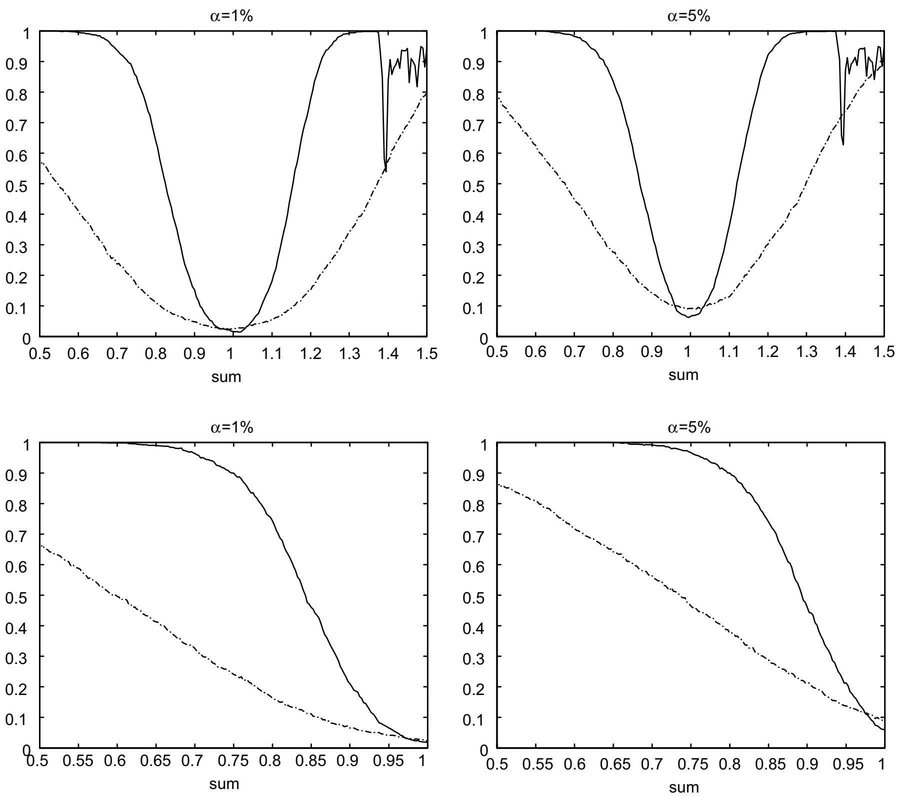


**FIGURE 14.1**
Power curves under the unified approach for $H_0 : \gamma_0 + \rho_0 + \lambda_0 = 1$.

unified approach to get the power curves. The results are in Figure 14.1. For the two-sided tests, the sum $\lambda_0 + \gamma_0 + \rho_0$ under the alternative hypothesis ranges from 0.65 to 1.35 with a $\frac{0.7}{200}$ increment; for the one-sided test with $H_1 : \lambda_0 + \gamma_0 + \rho_0 < 1$, the sum $\lambda_0 + \gamma_0 + \rho_0$ ranges from 0.65 to 1.0 with a $\frac{0.35}{200}$ increment. From Figure 14.2, we can see that the empirical sizes[18] are close to the theoretical ones and the tests are more powerful when $T = 50$ than those for the small $T = 10$. The power seems reasonable for the large $T = 50$. We run additional simulations where we use the corresponding estimation method without any transformation. Figure 14.2 is the counterparts[19] of Table 14.1.

[18] For the empirical size, the $T = 10$ case has 2.4%, 2.2%, 9.1%, and 8.8% from the first row to the second row, and the $T = 50$ case has 1.6%, 1.7%, 6.5%, and 5.8%. As the significance level are 1%, 1%, 5%, and 5% correspondingly, a larger $T$ will yield empirical sizes closer to the theoretical values.

[19] For the first row in Table 14.2, when the sum $\lambda_0 + \gamma_0 + \rho_0$ is much larger than 1 (i.e., the process is explosive), the estimates might not be available due to overflow without the unified transformation. Hence, for the two-sided power curves, we allow the sum only up to 1.3.

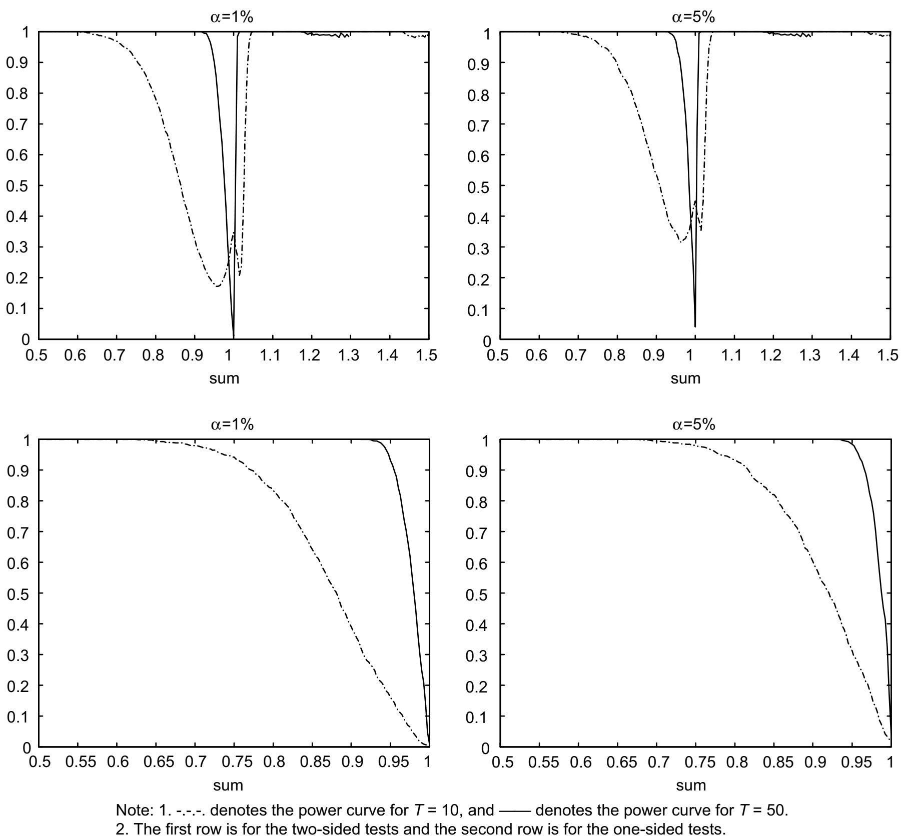


**FIGURE 14.2**
Power curves under Yu, de Jong, and Lee (2007) for $H_0 : \gamma_0 + \rho_0 + \lambda_0 = 1$.

We can see that, when $\lambda_0 + \gamma_0 + \rho_0 < 1$, the test is more powerful by using the corresponding method without any transformation; when $\lambda_0 + \gamma_0 + \rho_0 > 1$, the power curves are irregular and we need to rely on the unified approach for the inferences.[20]

## 14.5 Conclusion

This chapter establishes asymptotic properties of QMLEs for SDPD models with both time and individual fixed effects when both the number of individuals $n$ and the number of time periods $T$ can be large. Instead of using different

[20] For the empirical size, the $T = 10$ case has 34.8%, 0.3%, 44.9%, and 1.5% from the first row to the second row in Table 14.2, and the $T = 50$ case has 1.1%, 0.8%, 4%, and 4%. Hence, when $T$ is small, the empirical sizes could be far away from the theoretical values.

estimation methods depending on whether the DGP has time effects or not and whether the DGP is stable or not, we propose a data transformation approach to eliminate both the time effects and the possible unstable or explosive effects. The transformation is motivated by the possible co-integration relationship in the SDPD model, which is implied by the unit eigenvalues in the spatial weights matrix $W_n$. Unlike the co-integration in the multi-variate time series, the co-integrating vector is known and does not need to be estimated. With the proposed data transformation, the possible unstable or explosive components and time effects can be eliminated.

The transformation uses the co-integrating matrix. The effective sample size $n^*$ after transformation corresponds to the co-integration rank, which is the number of eigenvalues not equal to the unity. This transformation is of particular value when the process may contain explosive roots, as usual estimation methods can be poorly performed under such a situation. For the unified approach, when $T$ is relatively larger than $n^*$, the estimators are $\sqrt{n^*T}$ consistent and asymptotically centered normal; when $n^*$ is asymptotically proportional to $T$, the estimators are $\sqrt{n^*T}$ consistent and asymptotically normal, but the limit distribution is not centered around 0; when $T$ is relatively smaller than $n^*$, the estimators are consistent with rate $T$ and have a degenerate limit distribution. We also propose a bias correction for our estimators. We show that when $T$ grows faster than $n^{*1/3}$, the correction will asymptotically eliminate the bias and yield a centered confidence interval. Monte Carlo experiments have demonstrated a desirable finite sample performance of the estimator. A test statistic for testing possible spatial co-integration is also considered. In Lee and Yu (2010b), this unified estimation approach is applied to study the market integration in Keller and Shiue (2007) with the SDPD model and test for the spatial co-integration.

## Appendices

### A Some Notes

#### *A.1 The Eigenvalues of $A_n$: Three Cases of the DGP*

From Subsection 14.2.1, the eigenvalues matrix of $A_n$ can be decomposed as $D_n = \frac{\gamma_0+\rho_0}{1-\lambda_0} J_n + \tilde{D}_n$, where $J_n = \text{diag}\{1_{m_n}, 0, \cdots, 0\}$ and $\tilde{D}_n = \text{diag}\{0, \cdots, 0, d_{n,m_n+1}, \cdots, d_{nn}\}$ with $|d_{ni}| < 1$. Hence, $A_n^h = (\frac{\gamma_0+\rho_0}{1-\lambda_0})^h R_n J_n R_n^{-1} + B_n^h$ with $B_n^h = R_n \tilde{D}_n^h R_n^{-1}$. As $d_{ni} = \frac{\gamma_0+\rho_0\omega_{ni}}{1-\lambda_0\omega_{ni}}$, the derivative of $d_{ni} = \frac{\gamma_0+\rho_0\omega_{ni}}{1-\lambda_0\omega_{ni}}$ as a function of $\omega_{ni}$ is $\frac{\partial(\frac{\gamma_0+\rho_0\omega_{ni}}{1-\lambda_0\omega_{ni}})}{\partial\omega_{ni}} = \frac{\rho_0+\gamma_0\lambda_0}{(1-\lambda_0\omega_{ni})^2}$. Thus, $d_{ni}$ is a monotonic function of $\omega_{ni}$. Our setting assumes that $|d_{ni}| < 1$ whenever $d_{ni} \neq 1$. This requirement can be satisfied with appropriate restriction on the parameter space of $\rho_0$, $\gamma_0$ and $\lambda_0$ as shown below.

The case with $\rho_0 + \gamma_0\lambda_0 = 0$ implies that $d_{ni}$ is a constant function of $\omega_{ni}$. As $|\lambda_0| < 1$ (implied by Assumptions 1 and 3), the derivative is zero if and only if $\rho_0 + \gamma_0\lambda_0 = 0$, i.e., $\rho_0 = -\lambda_0\gamma_0$. In this situation, $d_{ni} = \frac{\gamma_0+\rho_0\omega_{ni}}{1-\lambda_0\omega_{ni}} = \gamma_0$, and all $|d_{ni}| < 1$ if $|\gamma_0| < 1$.[21] The $d_{ni}$ is a strictly increasing function of $\omega_{ni}$ if and only if $\rho_0 + \lambda_0\gamma_0 > 0$; otherwise it is a strictly decreasing function of $\omega_{ni}$ when $\rho_0 + \lambda_0\gamma_0 < 0$. Let $\gamma_0 + \rho_0 + \lambda_0 = 1 + a$, where $a$ is a constant. We have the stable case when $\gamma_0 + \rho_0 + \lambda_0 < 1$; the spatial cointegration case when $\gamma_0 + \rho_0 + \lambda_0 = 1$ but $\gamma_0 \neq 1$; and the explosive case when $\gamma_0 + \rho_0 + \lambda_0 > 1$. The condition $\rho_0 + \gamma_0\lambda_0 > 0$ $(< 0)$ is equivalent to $(1 - \gamma_0)(1 - \lambda_0) > -a$ $(< -a)$ because $(1 - \gamma_0)(1 - \lambda_0) = \rho_0 + \gamma_0\lambda_0 - a$.

Assume that $d_{ni}$ is an increasing function of $\omega_{ni}$. As $W_n$ is row-normalized, $-1 \leq \omega_{ni} \leq 1$ for all $i$. With the relation $d_{ni} = \frac{\gamma_0+\rho_0\omega_{ni}}{1-\lambda_0\omega_{ni}}$ on $[-1, 1]$, $d_{ni} = \frac{\gamma_0-\rho_0}{1+\lambda_0}$ at $\omega_{ni} = -1$, and $d_{ni} = \frac{\gamma_0+\rho_0}{1-\lambda_0}$ at $\omega_{ni} = 1$. Hence, the smallest eigenvalue of $A_n$ will be greater than or equal to $\frac{\gamma_0-\rho_0}{1+\lambda_0}$, and the largest eigenvalue will occur at $\omega_{ni} = 1$. Hence, the possible range of $d_{ni}$ with $\omega_{ni}$ in $[-1, 1]$ is $[\frac{\gamma_0-\rho_0}{1+\lambda_0}, \frac{\gamma_0+\rho_0}{1-\lambda_0}]$. The smallest eigenvalue of $A_n$ will be greater than $-1$ if

$$\frac{\gamma_0 - \rho_0}{1 + \lambda_0} > -1 \Leftrightarrow 1 + \gamma_0 + \lambda_0 > \rho_0 \Leftrightarrow 1 - \rho_0 > -\frac{a}{2}.$$

Also, whenever $\omega_{ni} < \frac{1-\gamma_0}{\rho_0+\lambda_0}$, the corresponding $d_{ni} < 1$. This is so, because the critical value $\omega^*$ such that $\frac{\gamma_0+\rho_0\omega^*}{1-\lambda_0\omega^*} = 1$ is at $\omega^* = \frac{1-\gamma_0}{\rho_0+\lambda_0} = 1 - \frac{a}{(\rho_0+\lambda_0)}$.

In summary, for any eigenvalue $\omega_{ni}$ of $W_n$ (with $|\omega_{ni}| \leq 1$), the corresponding eigenvalue of $A_n$ is $d_{ni} = \frac{\gamma_0+\rho_0\omega_{ni}}{1-\lambda_0\omega_{ni}}$. Under the situation $(1-\gamma_0)(1-\lambda_0) > -a$, we have $d_{ni} < 1$ if $\omega_{ni} < 1 - \frac{a}{\rho_0+\lambda_0}$; and $d_{ni} > -1$ if $1 - \rho_0 > -\frac{a}{2}$.

Hence, we have the following sufficient conditions for three cases in our studies. Assume that $|\lambda_0| < 1$ and $(1 - \gamma_0)(1 - \lambda_0) > -a$.

1. Stable case: $a < 0$. If $\rho_0 + \lambda_0 > 0$, all $d_{ni} \leq 1$ (because $\omega_{ni} < 1 - \frac{a}{\rho_0+\lambda_0}$); if $1 - \rho_0 > -\frac{a}{2}$, $-1 < d_{ni}$.
2. Spatial co-integration case: $a = 0$. When $\omega_{ni} = 1$, $d_{ni} = 1$; when $\omega_{ni} < 1$ and $1 - \rho_0 > 0$, then $|d_{ni}| < 1$.
3. Explosive case: $a > 0$. When $\omega_{ni} = 1$, $d_{ni} > 1$; when $\omega_{ni} < 1 - \frac{a}{\rho_0+\lambda_0} = \frac{1-\gamma_0}{\rho_0+\lambda_0}$, $|d_{ni}| < 1$; furthermore, with $1 - \rho_0 > -\frac{a}{2}$, $|d_{ni}| < 1$.

### *A.2 Decomposition*

From Equation 14.2, by iterative substitution, we have

$$Y_{nt} = A_n^{t+1} Y_{n,-1} + \sum_{h=0}^{t} A_n^h S_n^{-1}(\mathbf{c}_{n0} + X_{n,t-h}\beta_0 + V_{n,t-h} + \alpha_{t-h,0} l_n).$$

[21] For this special case, the model becomes $Y_{nt} = \gamma_0 Y_{n,t-1} + S_n^{-1}(X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + V_{nt})$. Hence, this case is $Y_{nt} = \gamma_0 Y_{n,t-1} + S_n^{-1} X_{nt}\beta_0 + \frac{\alpha_{t0}}{1-\lambda_0} l_n + \epsilon_{nt}$, where $\epsilon_{nt} = \lambda_0 W_n \epsilon_{nt} + \mathbf{c}_{n0} + V_{nt}$ has the panel disturbance structure in Kapoor, Kelejian, and Prucha (2007). This model is close to the one considered in Su and Yang (2007) except for the resulting regressor term.

As $S_n^{-1} l_n = \frac{1}{1-\lambda_0} l_n$ and $A_n = S_n^{-1}(\gamma_0 I_n + \rho_0 W_n) = (\gamma_0 I_n + \rho_0 W_n) S_n^{-1}$, using $W_n l_n = l_n$, we have $A_n^h S_n^{-1} l_n = \frac{1}{1-\lambda_0}(\frac{\gamma_0+\rho_0}{1-\lambda_0})^h l_n$. By $A_n^h = (\frac{\gamma_0+\rho_0}{1-\lambda_0})^h R_n J_n R_n^{-1} + B_n^h$ and $R_n J_n R_n^{-1} S_n^{-1} = S_n^{-1} R_n J_n R_n^{-1} = \frac{1}{1-\lambda_0} R_n J_n R_n^{-1}$ (see Proposition B.4 in Yu, de Jong, and Lee 2007), the above equation can be written as

$$Y_{nt} = A_n^{t+1} Y_{n,-1} + \sum_{h=0}^{t} B_n^h S_n^{-1}(\mathbf{c}_{n0} + X_{n,t-h}\beta_0 + V_{n,t-h}) + \frac{1}{1-\lambda_0}\sum_{h=0}^{t}\left(\frac{\gamma_0+\rho_0}{1-\lambda_0}\right)^h \times \alpha_{t-h,0} l_n + \frac{1}{1-\lambda_0}\sum_{h=0}^{t}\left(\frac{\gamma_0+\rho_0}{1-\lambda_0}\right)^h R_n J_n R_n^{-1}(\mathbf{c}_{n0} + X_{n,t-h}\beta_0 + V_{n,t-h}).$$

For $A_n^{t+1} Y_{n,-1}$, we have $A_n^{t+1} Y_{n,-1} = (\frac{\gamma_0+\rho_0}{1-\lambda_0})^{t+1} R_n J_n R_n^{-1} Y_{n,-1} + B_n^{t+1} Y_{n,-1}$, where

$$B_n^{t+1} Y_{n,-1} = \sum_{h=t+1}^{\infty} B_n^h S_n^{-1}(\mathbf{c}_{n0} + X_{n,t-h}\beta_0 + V_{n,t-h}) + \frac{1}{1-\lambda_0}\sum_{h=t+1}^{\infty} \alpha_{t-h,0} B_n^h l_n,$$

using $B_n A_n = B_n^2$ and $B_n S_n^{-1} = S_n^{-1} B_n$. The item with $B_n^h l_n$ is zero. Because $R_n$ is the eigenvectors matrix of $W_n$ and its first column is $l_n$, we have $R_n^{-1} l_n = e_{n1}$ which is the first unit vector. As $\tilde{D}_n e_{n1} = 0$, it follows that $B_n l_n = 0$. Hence, we can decompose $Y_{nt}$ as $Y_{nt} = Y_{nt}^u + Y_{nt}^s + Y_{nt}^\alpha$, which is Equation 14.3.

The $Y_{nt}^s$ represents a stable component as the eigenvalues of $B_n$ can be less than unity in absolute value for many parameter values (see Appendix A.1). The $Y_{nt}^\alpha$ captures the component due to time dummies. As $|\frac{\gamma_0+\rho_0}{1-\lambda_0}| < 1$ if and only if $-1 < \gamma_0 + \rho_0 + \lambda_0 < 1$ because $\lambda_0 < 1$, $Y_{nt}^u$ is also stable when $\gamma_0 + \rho_0 + \lambda_0 < 1$. But when $\gamma_0 + \rho_0 + \lambda_0 = 1$ ($> 1$), then $\frac{\gamma_0+\rho_0}{1-\lambda_0} = 1$ ($> 1$) and $Y_{nt}^u$ may represent the unstable or explosive components.

### *A.3 Data Transformation*

We can transform Equation 14.1 by $I_n - W_n$ into Equation 14.4, where the remaining $(I_n - W_n)\mathbf{c}_{n0}$ can be regarded as the individual effects. A special feature of the transformed Equation 14.4 is that the variance matrix of $(I_n - W_n)V_{nt}$ is equal to $\sigma_0^2 \Sigma_n \equiv \sigma_0^2 (I_n - W_n)(I_n - W_n)'$, which is singular. Hence, there is a linear dependence among the elements of $(I_n - W_n)V_{nt}$. An effective estimation method shall eliminate the linear dependence. This can be done with the eigenvalues and eigenvectors decomposition (see, e.g., Theil 1971, Chapter 6).

Let $[F_n, H_n]$ be the orthonormal matrix of eigenvectors and $\Lambda_n$ be the diagonal matrix of nonzero eigenvalues of $\Sigma_n$ such that $\Sigma_n F_n = F_n \Lambda_n$ and $\Sigma_n H_n = 0$. That is, the columns of $F_n$ consist of eigenvectors of nonzero eigenvalues and those of $H_n$ are for zero-eigenvalues of $\Sigma_n$. Let $n^*$ be the number of nonzero eigenvalues. The $F_n$ is an $n \times n^*$ matrix and $\Lambda_n$ is an $n^* \times n^*$ diagonal matrix. Thus,

$$\Sigma_n F_n = F_n \Lambda_n, \quad F_n' F_n = I_{n^*}, \quad \Sigma_n H_n = \mathbf{0}, \quad H_n' H_n = I_{n-n^*}, \\ F_n' H_n = \mathbf{0}, \quad F_n F_n' + H_n H_n' = I_n, \quad F_n \Lambda_n F_n' = \Sigma_n. \tag{14.28}$$

Because $\Sigma_n H_n = 0$, it implies that $(I_n - W_n)' H_n = 0$. In turn, $W_n(I_n - W_n) = W_n(F_n F_n' + H_n H_n')(I_n - W_n) = W_n F_n F_n'(I_n - W_n)$. Denote $W_n^* = \Lambda_n^{-1/2} F_n' W_n F_n \Lambda_n^{1/2}$ which is a $n^* \times n^*$ matrix. This matrix can be regarded as a spatial weights matrix for the following transformed equation:

$$Y_{nt}^* = \lambda_0 W_n^* Y_{nt}^* + \gamma_0 Y_{n,t-1}^* + \rho_0 W_n^* Y_{n,t-1}^* + X_{nt}^* \beta_0 + \boldsymbol{c}_{n0}^* + V_{nt}^*, \tag{14.29}$$

where $Y_{nt}^* = \Lambda_n^{-1/2} F_n'(I_n - W_n) Y_{nt}$ and other variables are defined correspondingly. Note that this transformed $Y_{nt}^*$ is an $n^*$ dimensional vector. Hence, after the transformation, the observations at time period $t$ have only $n^*$ degrees of freedom. Equation 14.29 shall provide the structural parameters for estimation. This equation is in the format of a typical SAR model in panel data, where the number of observations is $n^* T$.

### A.4 Determinant and Inverse of $S_n^*(\lambda) \equiv I_{n^*} - \lambda W_n^*$

We note that $S_n^* = \Lambda_n^{-1/2} F_n' S_n F_n \Lambda_n^{1/2}$. Let $\mu$ be a scalar. Because $(I_n - W_n) \cdot H_n = 0$,

$$\begin{aligned}
&[F_n, H_n]'(\mu I_n - W_n)[F_n, H_n] \\
&= \begin{pmatrix} \mu I_{n^*} - F_n' W_n F_n & -F_n' W_n H_n \\ -H_n' W_n F_n & \mu I_{n-n^*} - H_n' W_n H_n \end{pmatrix} = \begin{pmatrix} \mu I_{n^*} - F_n' W_n F_n & -F_n' W_n H_n \\ \mathbf{0} & (\mu - 1) I_{n-n^*} \end{pmatrix}.
\end{aligned}$$

Hence, $|\mu I_n - W_n| = (\mu - 1)^{n-n^*} |\mu I_{n^*} - F_n' W_n F_n|$. Because $|\mu I_{n^*} - W_n^*| = |\mu I_{n^*} - \Lambda_n^{-1/2} F_n' W_n F_n \Lambda_n^{1/2}| = |\mu I_{n^*} - F_n' W_n F_n|$, $|\mu I_n - W_n| = (\mu - 1)^{n-n^*} |\mu I_{n^*} - W_n^*|$. As $W_n$ has $(n - n^*)$ unit eigenvalues, the eigenvalues of $W_n^*$ are exactly the remaining eigenvalues of $W_n$, which are less than unity in the absolute value. Furthermore,

$$|S_n^*(\lambda)| = \frac{1}{(1-\lambda)^{n-n^*}} |S_n(\lambda)|. \tag{14.30}$$

Thus, the tractability in computing the determinant of $S_n^*(\lambda)$ is exactly that of $S_n(\lambda)$. When $W_n$ is constructed as a weights matrix that is row-normalized from an original symmetric matrix, Ord (1975) has suggested a computationally tractable method for the evaluation of $|S_n(\lambda)|$ at various $\lambda$ for the ML method. This is useful for evaluating the determinant of $S_n^*(\lambda)$ even though the row sums of $W_n^*$ may not even be unity.

Furthermore, a SAR model is an equilibrium model in the sense that the observed outcomes are determined by the equation. That is, the matrix $S_n^*(\lambda)$ shall be invertible. For the transformed equation (Equation 14.29), $S_n^*(\lambda)$ is invertible as long as the original matrices $S_n(\lambda)$ in Equation 14.1 is invertible. We can see that

$$S_n^{*-1}(\lambda) = \Lambda_n^{-1/2} F_n' S_n^{-1}(\lambda) F_n \Lambda_n^{1/2}, \tag{14.31}$$

because

$$\begin{aligned} S_n^*(\lambda)\cdot\Lambda_n^{-1/2}F_n'S_n^{-1}(\lambda)F_n\Lambda_n^{1/2} &= \Lambda_n^{-1/2}F_n'S_n(\lambda)F_nF_n'S_n^{-1}(\lambda)F_n\Lambda_n^{1/2}\\ &= \Lambda_n^{-1/2}F_n'S_n(\lambda)(I_n - H_nH_n')S_n^{-1}(\lambda)F_n\Lambda_n^{1/2}\\ &= I_{n^*} - \Lambda_n^{-1/2}F_n'S_n(\lambda)H_nH_n'S_n^{-1}(\lambda)F_n\Lambda_n^{1/2} = I_{n^*}, \end{aligned}$$

as $H_n'W_n = H_n'$, $H_n'S_n^{-1}(\lambda) = \frac{1}{1-\lambda}H_n'$ and $H_n'F_n = \mathbf{0}$.

### A.5 About $tr(G_n^*(\lambda))$

We have $G_n^*(\lambda) = \Lambda_n^{-1/2}F_n'G_n(\lambda)F_n\Lambda_n^{1/2}$. This is so because, from Equation 14.31,

$$\begin{aligned} G_n^*(\lambda) &= W_n^*S_n^{-1*}(\lambda) = \Lambda_n^{-1/2}F_n'W_nF_nF_n'S_n^{-1}(\lambda)F_n\Lambda_n^{1/2}\\ &= \Lambda_n^{-1/2}F_n'W_n(I_n - H_nH_n')S_n^{-1}(\lambda)F_n\Lambda_n^{1/2}\\ &= \Lambda_n^{-1/2}F_n'W_nS_n^{-1}(\lambda)F_n\Lambda_n^{1/2} - \Lambda_n^{-1/2}F_n'W_nH_nH_n'S_n^{-1}(\lambda)F_n\Lambda_n^{1/2}\\ &= \Lambda_n^{-1/2}F_n'W_nS_n^{-1}(\lambda)F_n\Lambda_n^{1/2} = \Lambda_n^{-1/2}F_n'G_n(\lambda)F_n\Lambda_n^{1/2}, \end{aligned}$$

because $H_n'S_n^{-1}(\lambda)F_n = \frac{1}{1-\lambda}H_n'F_n = \mathbf{0}$. Hence,

$$tr(G_n^*(\lambda)) = tr(F_n'G_n(\lambda)F_n) = tr[G_n(\lambda)(I_n - H_nH_n')] = tr(G_n(\lambda)) - \frac{n-n^*}{1-\lambda}, \tag{14.32}$$

where the last equality holds because $H_n'W_n = H_n'$ and $H_n'S_n^{-1}(\lambda) = \frac{1}{1-\lambda}H_n'$ implies that

$$\begin{aligned} tr(G_n(\lambda)H_nH_n') &= tr(H_n'G_n(\lambda)H_n) = tr(H_n'W_nS_n^{-1}(\lambda)H_n) = \frac{1}{1-\lambda}tr(H_n'H_n)\\ &= \frac{n-n^*}{1-\lambda}. \end{aligned}$$

As $G_n^{*2}(\lambda) = \Lambda_n^{-1/2}F_n'G_n(\lambda)F_nF_n'G_n(\lambda)F_n\Lambda_n^{1/2}$, we have

$$\begin{aligned} tr(G_n^{*2}(\lambda)) &= tr(F_n'G_n(\lambda)F_nF_n'G_n(\lambda)F_n) = tr(G_n(\lambda)F_nF_n'G_n(\lambda)F_nF_n')\\ &= tr(G_n(\lambda)(I_n - H_nH_n')G_n(\lambda)(I_n - H_nH_n')). \end{aligned}$$

Using $H_n'G_n(\lambda) = \frac{1}{(1-\lambda)}H_n'$ and $H_n'H_n = I_{n-n^*}$, we have $[G_n(\lambda)(I_n - H_nH_n')]^2 = [G_n(\lambda)]^2[I_n - H_nH_n']$ and

$$tr(G_n^{*2}(\lambda)) = tr(G_n^2(\lambda)) - \frac{n-n^*}{(1-\lambda)^2}, \tag{14.33}$$

because $H_n'G_n^2(\lambda)H_n = \frac{1}{(1-\lambda)^2}H_n'H_n = \frac{1}{(1-\lambda)^2}I_{n-n^*}$. In terms of the eigenvalues of $W_n$, as $W_n = R_n\varpi R_n^{-1}$, $tr(G_n^*(\lambda)) = \sum_{j=m_n+1}^{n}\frac{\varpi_{nj}}{1-\lambda\varpi_{nj}}$ and $tr(G_n^{*2}(\lambda)) = \sum_{j=m_n+1}^{n}\frac{\varpi_{nj}^2}{(1-\lambda\varpi_{nj})^2}$.

Also, as $J_n^* = (I_n - W_n)'\Sigma_n^+(I_n - W_n)$ and $(I_n - W_n)G_n(\lambda) = G_n(\lambda)(I_n - W_n)$, Equation 14.32 implies that

$$\begin{aligned} tr(J_n^* G_n(\lambda)) &= tr(G_n(\lambda)(I_n - W_n)(I_n - W_n)' F_n \Lambda_n^{-1} F_n') \\ &= tr(G_n(\lambda) F_n F_n') = tr(G_n(\lambda)(I_n - H_n H_n')) \\ &= tr(G_n^*(\lambda)). \end{aligned} \tag{14.34}$$

For $J_n^*$, we have $tr(J_n^*) = tr((I_n - W_n)' F_n \Lambda_n^{-1} F_n'(I_n - W_n)) = tr(\Lambda_n^{-1}\Lambda_n) = n^*$ by using Equation 14.28. The $J_n^*$ is an orthogonal projector. This is so, because $J_n^*$ is symmetric and $J_n^* J_n^* = (I_n - W_n)'\Sigma_n^+(I_n - W_n) \cdot (I_n - W_n)'\Sigma_n^+(I_n - W_n) = (I_n - W_n)'\Sigma_n^+\Sigma_n\Sigma_n^+(I_n - W_n) = (I_n - W_n)'\Sigma_n^+(I_n - W_n) = J_n^*$.

## B Lemmas for Some Statistics in the Model

The following lemmas can be found in Yu, de Jong, and Lee (2008). These lemmas provide orders for relevant terms in the score and the Hessian matrix of the log-likelihood function. They include also a CLT for linear and quadratic forms of disturbances. Denote $\mathbb{U}_{nt} = \sum_{h=1}^{\infty} P_{nh} V_{n,t+1-h}$, where $\{P_{nh}\}_{h=1}^{\infty}$ is a sequence of $n \times n$ nonstochastic square matrices.

**Assumption A1** The disturbances $\{v_{it}\}$, $i = 1, 2, ..., n$ and $t = 1, 2, ..., T$, are i.i.d. across $i$ and $t$ with zero mean, variance $\sigma_0^2$ and $\mathrm{E}|v_{it}|^{4+\eta} < \infty$ for some $\eta > 0$.
**Assumption A2** $\sum_{h=1}^{\infty} abs(P_{nh})$ is UB.
**Assumption A3** The elements of $n \times 1$ vector $D_{nt}$ are nonstochastic and bounded, uniformly in $n$ and $t$.
**Assumption A4** $n$ is a nondecreasing function of $T$ and $T$ goes to infinity.

**Lemma 14.1** *Under Assumptions A1 and A4, for an $n \times n$ nonstochastic matrix $\mathcal{B}_n$, uniformly bounded in row and column sums,*

$$\frac{1}{nT}\sum_{t=1}^{T} V_{nt}'\mathcal{B}_n V_{nt} - \mathrm{E}(\frac{1}{nT}\sum_{t=1}^{T} V_{nt}'\mathcal{B}_n V_{nt}) = O_p\left(\frac{1}{\sqrt{nT}}\right), \tag{14.35}$$

$$\frac{1}{n}\bar{V}_{nT}'\mathcal{B}_n\bar{V}_{nT} - \mathrm{E}(\frac{1}{n}\bar{V}_{nT}'\mathcal{B}_n\bar{V}_{nT}) = O_p\left(\frac{1}{\sqrt{nT^2}}\right), \tag{14.36}$$

*and*

$$\frac{1}{nT}\sum_{t=1}^{T} \tilde{V}_{nt}'\mathcal{B}_n \tilde{V}_{nt} - \mathrm{E}(\frac{1}{nT}\sum_{t=1}^{T} \tilde{V}_{nt}'\mathcal{B}_n \tilde{V}_{nt}) = O_p\left(\frac{1}{\sqrt{nT}}\right), \tag{14.37}$$

*where* $\mathrm{E}(\frac{1}{nT}\sum_{t=1}^{T} V_{nt}'\mathcal{B}_n V_{nt}) = O(1)$, $\mathrm{E}(\frac{1}{n}\bar{V}_{nT}'\mathcal{B}_n\bar{V}_{nT}) = O(T^{-1})$ *and* $\mathrm{E}(\frac{1}{nT}\sum_{t=1}^{T} \tilde{V}_{nt}'\mathcal{B}_n\tilde{V}_{nt}) = O(1)$.

**Lemma 14.2** *Under Assumptions A1, A2, and A4,*

$$\sqrt{\frac{T}{n}}(\bar{\mathbb{U}}'_{nT,-1}\bar{V}_{nT} - \mathrm{E}(\bar{\mathbb{U}}'_{nT,-1}\bar{V}_{nT})) = O_p\left(\frac{1}{\sqrt{T}}\right), \tag{14.38}$$

*where* $\sqrt{\frac{T}{n}}\mathrm{E}(\bar{\mathbb{U}}'_{nT,-1}\bar{V}_{nT}) = \sqrt{\frac{n}{T}}\frac{1}{n}\sigma_0^2 tr\left(\sum_{h=1}^{\infty} P_{nh}\right) + O\left(\sqrt{\frac{n}{T^3}}\right)$.

For the lemma that follows, we will consider the following form:

$$Q_{nT} = \sum_{t=1}^{T}(\mathbb{U}'_{n,t-1}V_{nt} + D'_{nt}V_{nt} + V'_{nt}\mathcal{B}_n V_{nt} - \sigma_0^2 tr(\mathcal{B}_n)) = \sum_{t=1}^{T}\sum_{i=1}^{n} z_{nt,i},$$

where $\mathcal{B}_n$ is a $n \times n$ nonstochastic symmetric matrix which is UB, and $z_{nt,i} = (u_{i,t-1} + d_{nti})v_{it} + b_{n,ii}(v_{it}^2 - \sigma_0^2) + 2(\sum_{j=1}^{i-1} b_{n,ij}v_{jt})v_{it}$, where $b_{n,ij}$ is the $(i, j)$ element of $\mathcal{B}_n$ and $d_{nti}$ is the $i$th element of $D_{nt}$. Then, for the mean and variance of $Q_{nT}$, $\mu_{Q_{nT}} = 0$ and

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

where $\mu_s = Ev_{it}^s$ for $s = 3, 4$.

**Lemma 14.3** *Under Assumptions A1, A2, A3, A4, and that $\mathcal{B}_n$ is UB, if the sequence $\frac{1}{nT}\sigma^2_{Q_{nT}}$ is bounded away from zero, then, $\frac{Q_{nT}}{\sigma_{Q_{nT}}} \xrightarrow{d} N(0, 1)$.*

Denote $Z_{nt} = (Y_{n,t-1}, W_n Y_{n,t-1}, X_{nt})$, we are going to provide some lemmas related to $(I_n - W_n)\tilde{Z}_{nt}$, $(I_n - W_n)\bar{Z}_{nT}$ and $\tilde{V}_{nt}$, $\bar{V}_{nT}$ of the model Equation 14.1.

**Lemma 14.4** *Under Assumptions 1–7, for an $n \times n$ nonstochastic UB matrix $\mathcal{B}_n$,*

$$\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}'_{nt}(I_n - W_n)'\mathcal{B}_n(I_n - W_n)\tilde{Z}_{nt} - \mathrm{E}\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}'_{nt}(I_n - W_n)'\mathcal{B}_n(I_n - W_n)\tilde{Z}_{nt}$$
$$= O_p\left(\frac{1}{\sqrt{nT}}\right), \tag{14.39}$$

*and*

$$\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}'_{nt}(I_n - W_n)'\mathcal{B}_n(I_n - W_n)\tilde{V}_{nt} - \mathrm{E}\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}'_{nt}(I_n - W_n)'\mathcal{B}_n(I_n - W_n)\tilde{V}_{nt}$$
$$= O_p\left(\frac{1}{\sqrt{nT}}\right), \quad (14.40)$$

*where* $\mathrm{E}\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}'_{nt}(I_n - W_n)'\mathcal{B}_n(I_n - W_n)\tilde{Z}_{nt}$ *is* $O(1)$ *and* $\mathrm{E}\frac{1}{nT}\sum_{t=1}^{T}\tilde{Z}'_{nt}(I_n - W_n)'\mathcal{B}_n(I_n - W_n)\tilde{V}_{nt}$ *is* $O\left(\frac{1}{T}\right)$.

**Lemma 14.5** *If* $\|B_n(\theta_0))\|_\infty < 1$ *(resp:* $\|B_n(\theta_0))\|_1 < 1$*), then the row sum (resp: column sum) of* $\sum_{h=0}^{\infty} B_n^h(\theta)$ *and* $\sum_{h=1}^{\infty} hB_n^{h-1}(\theta)$ *are bounded uniformly in n and in a neighborhood of* $\theta_0$*.*

## C Concentrated QML of the Transformation Approach

### *C.1 Reduced Form of Equation 14.1*

From Equation 14.1, we have $Y_{nt} = S_n^{-1}(Z_{nt}\delta_0 + \mathbf{c}_{n0} + \alpha_t l_n + V_{nt})$ and $W_nY_{nt} = G_nZ_{nt}\delta_0 + G_n\mathbf{c}_{n0} + \alpha_t G_n l_n + G_n V_{nt}$. By using $S_n^{-1} = I_n + \lambda_0 G_n$, $Y_{nt} = Z_{nt}\delta_0 + \lambda_0 G_n Z_{nt}\delta_0 + S_n^{-1}\mathbf{c}_{n0} + \alpha_t S_n^{-1} l_n + S_n^{-1}V_{nt}$. With $S_n^{-1}l_n = \frac{1}{1-\lambda_0}l_n$ and $(I_n - W_n)l_n = \mathbf{0}$,

$$\tilde{Y}_{nt} = \tilde{Z}_{nt}\delta_0 + \lambda_0 G_n\tilde{Z}_{nt}\delta_0 + \frac{\tilde{\alpha}_t}{1-\lambda_0}l_n + S_n^{-1}\tilde{V}_{nt},$$

and

$$(I_n - W_n)\tilde{Y}_{nt} = (I_n - W_n)\tilde{Z}_{nt}\delta_0 + \lambda_0(I_n - W_n)G_n\tilde{Z}_{nt}\delta_0 + (I_n - W_n)S_n^{-1}\tilde{V}_{nt}. \quad (14.41)$$

Similarly, as $W_n\tilde{Y}_{nt} = G_n\tilde{Z}_{nt}\delta_0 + \tilde{\alpha}_t G_n l_n + G_n\tilde{V}_{nt}$,

$$(I_n - W_n)W_n\tilde{Y}_{nt} = (I_n - W_n)G_n\tilde{Z}_{nt}\delta_0 + (I_n - W_n)G_n\tilde{V}_{nt}, \quad (14.42)$$

because $(I_n - W_n)G_n l_n = \frac{1}{1-\lambda_0}(I_n - W_n)l_n = \mathbf{0}$.

### C.2 FOC and SOC of the Concentrated Log-Likelihood

Denote $J_n^* = (I_n - W_n)'\Sigma_n^+(I_n - W_n)$ and $G_n^* = W_n^* S_n^{*-1}$. By using $trG_n(\lambda) - tr(G_n^*(\lambda)) = \frac{n-n^*}{1-\lambda}$ and $tr(G_n^2(\lambda)) - tr(G_n^{*2}(\lambda)) = \frac{n-n^*}{(1-\lambda)^2}$ (see Appendix A.5), the first-order derivatives of Equation 14.10 are

$$\frac{\partial \ln L_{n,T}(\theta)}{\partial \theta} = \begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^{T}(J_n^*\tilde{Z}_{nt})'\tilde{V}_{nt}(\theta) \\ \frac{1}{\sigma^2}\sum_{t=1}^{T}((J_n^*W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\theta)) - T\,trG_n^*(\lambda) \\ \frac{1}{2\sigma^4}\sum_{t=1}^{T}(\tilde{V}_{nt}'(\theta)J_n\tilde{V}_{nt}(\theta) - n^*\sigma^2) \end{pmatrix}, \tag{14.43}$$

and the second order derivatives are

$$\frac{\partial^2 \ln L_{n,T}(\theta)}{\partial\theta\partial\theta'} = -\begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^{T}\tilde{Z}_{nt}'J_n^*\tilde{Z}_{nt} & \frac{1}{\sigma^2}\sum_{t=1}^{T}\tilde{Z}_{nt}'J_n^*W_n\tilde{Y}_{nt} & \frac{1}{\sigma^4}\sum_{t=1}^{T}\tilde{Z}_{nt}'J_n^*\tilde{V}_{nt}(\theta) \\ * & \frac{1}{\sigma^2}\sum_{t=1}^{T}((W_n\tilde{Y}_{nt})'J_n^*W_n\tilde{Y}_{nt}) + T\,tr((G_n^*(\lambda))^2) & \frac{1}{\sigma^4}\sum_{t=1}^{T}(W_n\tilde{Y}_{nt})'J_n^*\tilde{V}_{nt}(\theta) \\ * & * & -\frac{n^*T}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{t=1}^{T}\tilde{V}_{nt}'(\theta)J_n^*\tilde{V}_{nt}(\theta) \end{pmatrix}. \tag{14.44}$$

At $\theta_0$,

$$\frac{1}{\sqrt{n^*T}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma_0^2}\frac{1}{\sqrt{n^*T}}\sum_{t=1}^{T}\tilde{Z}_{nt}'J_n^*\tilde{V}_{nt} \\ \frac{1}{\sigma_0^2}\frac{1}{\sqrt{n^*T}}\sum_{t=1}^{T}(G_n\tilde{Z}_{nt}\delta_0)'J_n^*\tilde{V}_{nt} + \frac{1}{\sigma_0^2}\frac{1}{\sqrt{n^*T}}\sum_{t=1}^{T}(\tilde{V}_{nt}'G_n'J_n^*\tilde{V}_{nt} - \sigma_0^2 trG_n^*) \\ \frac{1}{2\sigma_0^4}\frac{1}{\sqrt{n^*T}}\sum_{t=1}^{T}(\tilde{V}_{nt}'J_n^*\tilde{V}_{nt} - n^*\sigma_0^2) \end{pmatrix}, \tag{14.45}$$

which is a linear and quadratic form of $\tilde{V}_{nt}$. For the information matrix,

$$\Sigma_{\theta_0,nT} = \frac{1}{\sigma_0^2}\begin{pmatrix} \mathrm{E}\mathcal{H}_{nT} & \mathbf{0}_{(k+3)\times 1} \\ \mathbf{0}_{1\times(k+3)} & 0 \end{pmatrix}$$
$$+\begin{pmatrix} \mathbf{0}_{(k+2)\times(k+2)} & \mathbf{0}_{(k+2)\times 1} & \mathbf{0}_{(k+2)\times 1} \\ \mathbf{0}_{1\times(k+2)} & \frac{1}{n^*}\left[tr(G_n' J_n^* G_n) + tr((G_n^*)^2)\right] & \frac{1}{\sigma_0^2 n^*} tr(J_n^* G_n) \\ \mathbf{0}_{1\times(k+2)} & \frac{1}{\sigma_0^2 n^*} tr(J_n^* G_n) & \frac{1}{2\sigma_0^4} \end{pmatrix}$$
$$-\begin{pmatrix} \mathbf{0}_{(k+2)\times(k+2)} & * & * \\ \frac{1}{\sigma_0^2 n^*}\mathrm{E}(G_n \bar{V}_{nT})' J_n^* \bar{Z}_{nT} & \frac{2}{\sigma_0^2 n^*}\mathrm{E}[(G_n \bar{Z}_{nT}\delta_0)' J_n^* G_n \bar{V}_{nT}] + \frac{1}{n^* T} tr(G_n' J_n^* G_n) & * \\ \frac{1}{\sigma_0^4 n^*}\mathrm{E}(\bar{Z}_{nT}' J_n^* \bar{V}_{nT})' & \frac{1}{\sigma_0^4 n^*}\mathrm{E}[(G_n \bar{Z}_{nT}\delta_0)' J_n^* \bar{V}_{nT}]' + \frac{1}{\sigma_0^2 n^* T} tr(J_n^* G_n) & \frac{1}{T}\frac{1}{\sigma_0^4} \end{pmatrix}.$$

### *C.3 About* $-\frac{1}{n^* T}\frac{\partial^2 \ln L_{nT}(\theta)}{\partial\theta\partial\theta'}$

Denote $\|\theta - \theta_0\|$ as the Euclidean norm of $\theta - \theta_0$, and $\Theta_1$ as a neighborhood of $\theta_0$, then, we have

$$\frac{1}{n^* T}\frac{\partial^2 \ln L_{nT}(\theta)}{\partial\theta\partial\theta'} - \frac{1}{n^* T}\frac{\partial^2 \ln L_{nT}(\theta_0)}{\partial\theta\partial\theta'} = \|\theta - \theta_0\| \cdot O_p(1), \tag{14.46}$$

$$\frac{1}{n^* T}\frac{\partial^2 \ln L_{nT}(\theta_0)}{\partial\theta\partial\theta'} + \Sigma_{\theta_0,nT} = O_p\left(\frac{1}{\sqrt{n^* T}}\right), \tag{14.47}$$

$$\sup_{\theta\in\Theta}\left|\frac{1}{n^* T}\frac{\partial^2 \ln L_{nT}(\theta)}{\partial\theta\partial\theta'} - \frac{1}{n^* T}\mathrm{E}\frac{\partial^2 \ln L_{nT}(\theta)}{\partial\theta\partial\theta'}\right|_{ij} = O_p\left(\frac{1}{\sqrt{n^* T}}\right), \tag{14.48}$$

and

$$\sup_{\theta\in\Theta_1}\left|\frac{1}{n^* T}\mathrm{E}\frac{\partial^2 \ln L_{nT}(\theta)}{\partial\theta\partial\theta'} + \Sigma_{\theta_0,nT}\right|_{ij} = \sup_{\theta\in\Theta_1}\|\theta - \theta_0\| \cdot O(1) \tag{14.49}$$

for all $i, j = 1, 2, \cdots, k+4$. These are Equation A.11 to Equation A.14 in Yu, de Jong, and Lee (2008).

## D Proofs for Claims and Theorems

### *D.1 Proof of nonsingularity of the information matrix*

The result can be proved by using an argument by contradiction. For $\Sigma_{\theta_0} \equiv \lim_{T\to\infty}\Sigma_{\theta_0,nT}$, where $\Sigma_{\theta_0,nT}$ is Equation 14.12, we shall prove that $\Sigma_{\theta_0}\alpha = 0$

implies $\alpha = 0$, where $\alpha = (\alpha_1', \alpha_2, \alpha_3)'$, $\alpha_2, \alpha_3$ are scalars and $\alpha_1$ is $(k+2)\times 1$ vector. If this is true, then, columns of $\Sigma_{\theta_0}$ would be linear independent so that $\Sigma_{\theta_0}$ would be nonsingular. Denote $\mathcal{H}_\delta = \text{plim}_{T\to\infty}\frac{1}{n^*T}\sum_{t=1}^T \tilde{Z}_{nt}' J_n^* \tilde{Z}_{nt}$, $\mathcal{H}_{\delta\lambda} = \text{plim}_{T\to\infty}\frac{1}{n^*T}\sum_{t=1}^T \tilde{Z}_{nt}' J_n^* G_n \tilde{Z}_{nt}\delta_0$, $\mathcal{H}_{\lambda\delta} = \mathcal{H}_{\delta\lambda}'$ and $\mathcal{H}_\lambda = \text{plim}_{T\to\infty}\frac{1}{n^*T}\sum_{t=1}^T (G_n\tilde{Z}_{nt}\delta_0)' J_n^* G_n \tilde{Z}_{nt}\delta_0$. Then

$$\Sigma_{\theta_0} = \frac{1}{\sigma_0^2}\begin{pmatrix} \mathcal{H}_\delta & \mathcal{H}_{\delta\lambda} & \mathbf{0}_{(k+2)\times 1} \\ \mathcal{H}_{\lambda\delta} & \text{E}\mathcal{H}_\lambda + \lim_{n\to\infty}\frac{\sigma_0^2}{n^*}[tr(G_n'J_n^*G_n) + tr((G_n^*)^2)] & \lim_{n\to\infty}\frac{1}{n^*}tr(J_n^*G_n) \\ \mathbf{0}_{1\times(k+2)} & \lim_{n\to\infty}\frac{1}{n^*}tr(J_n^*G_n) & \frac{1}{2\sigma_0^2} \end{pmatrix}.$$

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

$$\begin{aligned} &\mathcal{H}_\delta \times \alpha_1 + \mathcal{H}_{\delta\lambda}\times\alpha_2 = 0, \\ &\frac{1}{\sigma_0^2}\mathcal{H}_{\lambda\delta}\times\alpha_1 + \left(\frac{1}{\sigma_0^2}\mathcal{H}_\lambda + \lim_{n\to\infty}\frac{1}{n^*}\left[tr(G_n'J_n^*G_n) + tr((G_n^*)^2)\right]\right) \\ &\quad\times\alpha_2 + \lim_{n\to\infty}\frac{1}{\sigma_0^2 n^*}tr(J_n^*G_n)\times\alpha_3 = 0, \\ &\lim_{n\to\infty}\frac{1}{n^*}tr(J_n^*G_n)\times\alpha_2 + \frac{1}{2\sigma_0^2}\times\alpha_3 = 0. \end{aligned}$$

From the first equation, $\alpha_1 = -(\mathcal{H}_\delta)^{-1}\mathcal{H}_{\delta\lambda}\times\alpha_2$; from the third equation, $\alpha_3 = -2\lim_{n\to\infty}\frac{\sigma_0^2}{n^*}tr(J_n^*G_n)\times\alpha_2$. By eliminating $\alpha_1$ and $\alpha_3$, the remaining equation becomes

$$\left\{\left(\frac{1}{\sigma_0^2}(\mathcal{H}_\lambda - \mathcal{H}_{\lambda\delta}\mathcal{H}_\delta^{-1}\mathcal{H}_{\delta\lambda})\right) + \lim_{n\to\infty}\frac{1}{n^*}\left[tr(G_n'J_n^*G_n) + tr((G_n^*)^2) - 2\frac{tr^2(J_n^*G_n)}{n^*}\right]\right\}\times\alpha_2 = 0.$$

Using Equation 14.34 and that $J_n^*$ is idempotent, denote $C_n = G_n^* - \frac{tr(G_n^*)}{n^*}$, we have

$$\begin{aligned} tr(G_n'J_n^*G_n) + tr((G_n^*)^2) - 2\frac{tr^2(J_n^*G_n)}{n^*} &= tr(G_n^{*\prime}G_n^*) + tr((G_n^*)^2) - 2\frac{tr^2(G_n^*)}{n^*} \\ &= \frac{1}{2}tr(C_n' + C_n)(C_n' + C_n)', \end{aligned}$$

which is nonnegative. Hence, if the limit of $\text{E}\mathcal{H}_{nT}$ is nonsingular or the limit of $\frac{1}{n^*}(tr(G_n^{*\prime}G_n^*) + tr((G_n^*)^2) - 2\frac{tr^2(G_n^*)}{n^*})$ is nonzero, we have $\alpha_2 = 0$ and hence $\alpha = 0$. This proves the nonsingularity of $\Sigma_{\theta_0}$. ■

### *D.2 Proof of Theorem 14.1*

*To prove* $\frac{1}{n^*T}\ln L_{n,T}(\theta) - Q_{n,T}(\theta) \xrightarrow{p} 0$ *uniformly in* $\theta$ *in any compact parameter space* $\Theta$:

From $\tilde{V}_{nt}(\theta) \equiv S_n(\lambda)\tilde{Y}_{nt} - \tilde{Z}_{nt}\delta - \tilde{\alpha}_t l_n$ and $\tilde{V}_{nt} = S_n\tilde{Y}_{nt} - \tilde{Z}_{nt}\delta_0 - \tilde{\alpha}_{t0}l_n$, using $J_n^* l_n = \mathbf{0}$, we have $J_n^*\tilde{V}_{nt}(\theta) = J_n^*\tilde{V}_{nt} - (\lambda-\lambda_0)J_n^* W_n\tilde{Y}_{nt} - J_n^*\tilde{Z}_{nt}(\delta-\delta_0)$. As $\Theta$ is compact and $\sigma^2$ is bounded away from zero in $\Theta$, by Lemma 14.1 and 14.4,

$$\frac{1}{n^*T}\ln L_{n,T}(\theta) - Q_{n,T}(\theta)$$
$$= -\frac{1}{2\sigma^2}\left(\frac{1}{n^*T}\sum_{t=1}^{T}\tilde{V}'_{nt}(\theta)J_n^*\tilde{V}_{nt}(\theta) - \frac{1}{n^*T}\mathrm{E}\sum_{t=1}^{T}\tilde{V}'_{nt}(\theta)J_n^*\tilde{V}_{nt}(\theta)\right) \xrightarrow{p} 0$$

uniformly in $\theta$ in $\Theta$.

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

For $Q_{n,T}(\theta)$ in Equation 14.11, as $J_n^*\tilde{V}_{nt}(\theta) = J_n^*[S_n(\lambda)\tilde{Y}_{nt} - \tilde{Z}_{nt}\delta]$ and $\tilde{Y}_{nt} = S_n^{-1}\tilde{Z}_{nt}\delta_0 + S_n^{-1}\tilde{V}_{nt} + \frac{\tilde{\alpha}_{t0}}{1-\lambda_0}l_n$,

$$J_n^*\tilde{V}_{nt}(\theta) = J_n^*[S_n(\lambda)S_n^{-1}\tilde{Z}_{nt}\delta_0 - \tilde{Z}_{nt}\delta + S_n(\lambda)S_n^{-1}\tilde{V}_{nt}]$$

because $J_n^* l_n = \mathbf{0}$. Hence,

$$\begin{aligned}\mathrm{E}\frac{1}{n^*T}\sum_{t=1}^{T}\tilde{V}'_{nt}(\theta)J_n^*\tilde{V}_{nt}(\theta) &= \frac{1}{n^*T}\mathrm{E}\sum_{t=1}^{T}(S_n(\lambda)S_n^{-1}\tilde{Z}_{nt}\delta_0 - \tilde{Z}_{nt}\delta)'J_n^*(S_n(\lambda)\\ &\quad\times S_n^{-1}\tilde{Z}_{nt}\delta_0 - \tilde{Z}_{nt}\delta) + \frac{1}{n^*}\frac{T-1}{T}\sigma_0^2 tr\\ &\quad\times(S_n^{-1\prime}S_n'(\lambda)J_n^*S_n(\lambda)S_n^{-1}) + \frac{2}{n^*T}\mathrm{E}\sum_{t=1}^{T}\\ &\quad\times(S_n(\lambda)S_n^{-1}\tilde{Z}_{nt}\delta_0 - \tilde{Z}_{nt}\delta)'J_n^*S_n(\lambda)S_n^{-1}\tilde{V}_{nt}. \quad (14.50)\end{aligned}$$

With these terms, similar to Lee and Yu (2010a), it can be shown that $Q_{n,T}(\theta)$ is uniformly equicontinuous in $\theta$ in any compact parameter space $\Theta$.

*To prove the identification:*

As $tr J_n^* = n^*$, $\mathrm{E}\sum_{t=1}^{T}\tilde{V}'_{nt}J_n^*\tilde{V}_{nt} = n^*(T-1)\sigma_0^2$ from Lemma 14.1. Hence, $\frac{1}{n^*T}\mathrm{E}\ln L_{n,T}(\theta) - \frac{1}{n^*T}\mathrm{E}\ln L_{n,T}(\theta_0) = -\frac{1}{2}(\ln\sigma^2 - \ln\sigma_0^2) + \frac{1}{n^*}\ln|S_n(\lambda)| - \frac{1}{n^*}\ln|S_n| - \frac{n-n^*}{n^*}(\ln(1-\lambda) - \ln(1-\lambda_0)) - (\frac{1}{2\sigma^2}\frac{1}{n^*T}\sum_{t=1}^{T}E\tilde{V}'_{nt}(\theta)J_n^*\tilde{V}_{nt}(\theta) - \frac{T-1}{2T})$. By using $S_n(\lambda)S_n^{-1} = I_n + (\lambda_0-\lambda)G_n$, from Equation 14.50, $\frac{1}{n^*T}\mathrm{E}\ln L_{n,T}(\theta) - \frac{1}{n^*T}\mathrm{E}\ln L_{n,T}(\theta_0) = T_{1,n}(\lambda,\sigma^2) - \frac{1}{2\sigma^2}T_{2,n,T}(\delta,\lambda) + O(T^{-1})$, where

$$\begin{aligned}T_{1,n}(\lambda,\sigma^2) &= -\frac{1}{2}(\ln\sigma^2 - \ln\sigma_0^2) + \frac{1}{n^*}\ln|S_n(\lambda)| - \frac{1}{n^*}\ln|S_n| - \frac{n-n^*}{n^*}\\ &\quad\times(\ln(1-\lambda) - \ln(1-\lambda_0)) - \frac{1}{2\sigma^2}(\sigma_n^2(\lambda) - \sigma^2),\end{aligned}$$

and

$$T_{2,n,T}(\delta,\lambda) = \frac{1}{n^*T}\sum_{t=1}^{T} \mathrm{E}\{[\tilde{Z}_{nt}(\delta_0-\delta)+(\lambda_0-\lambda)G_n\tilde{Z}_{nt}\delta_0]' J_n^* \times [\tilde{Z}_{nt}(\delta_0-\delta)+(\lambda_0-\lambda)G_n\tilde{Z}_{nt}\delta_0]\},$$

where $\sigma_n^2(\lambda) = \frac{\sigma_0^2}{n^*} tr(S_n^{-1\prime} S_n'(\lambda) J_n^* S_n(\lambda) S_n^{-1})$. Consider the pure spatial process $Y_{nt} = \lambda_0 W_n Y_{nt} + \alpha_t l_n + V_{nt}$ for a single period $t$. With similar data transformation as in Equation 14.5, the log-likelihood function of this process is

$$\ln L_{p,n}(\lambda,\sigma^2) = -\frac{n^*}{2}\ln 2\pi - \frac{n^*}{2}\ln\sigma^2 - (n-n^*)\ln(1-\lambda) + \ln|S_n(\lambda)| - \frac{1}{2\sigma^2} V_{nt}'(\lambda) J_n^* V_{nt}'(\lambda), \tag{14.51}$$

where $V_{nt}(\lambda) = S_n(\lambda) Y_{nt}$. Let $\mathrm{E}_p(\cdot)$ be the expectation operator for $Y_{nt}$ based on this pure spatial autoregressive process. It follows that

$$\mathrm{E}_p(\frac{1}{n^*}\ln L_{p,n}(\lambda,\sigma^2)) - \mathrm{E}_p(\frac{1}{n^*}\ln L_{p,n}(\lambda_0,\sigma_0^2)) = -\frac{1}{2}(\ln\sigma^2 - \ln\sigma_0^2) + \frac{1}{n^*}\ln|S_n(\lambda)| - \frac{1}{n^*}\ln|S_n| - \frac{n-n^*}{n^*}(\ln(1-\lambda) - \ln(1-\lambda_0)) - \frac{1}{2\sigma^2}(\sigma_n^2(\lambda)-\sigma^2),$$

which equals to $T_{1,n}(\lambda,\sigma^2)$. By the information inequality, $\ln L_{p,n}(\lambda,\sigma^2) - \ln L_{p,n}(\lambda_0,\sigma_0^2) \leq 0$. Thus, $T_{1,n}(\lambda,\sigma^2) \leq 0$ for any $(\lambda,\sigma^2)$.

For $T_{2,n,T}(\delta,\lambda)$, it is a quadratic function of $\delta$ and $\lambda$. Under the assumed condition that $\lim_{T\to\infty} \mathrm{E}\mathcal{H}_{nT}$ is nonsingular, $\lim_{T\to\infty} T_{2,n,T}(\delta,\lambda) > 0$ whenever $(\delta,\lambda) \neq (\delta_0,\lambda_0)$. So, $(\delta,\lambda)$ is globally identified. Given $\lambda_0$, $\sigma_0^2$ is also the unique maximizer of $T_{1,n}(\lambda_0,\sigma^2)$ for any given $n^*$. In the event that $n^* \to \infty$, $\sigma_0^2$ is the unique maximizer of $\lim_{T\to\infty} T_{1,n}(\lambda_0,\sigma^2)$. Hence, $(\delta,\lambda,\sigma^2)$ is globally identified.

By combining the results above together, the consistency follows. ■

### *D.3 Proof of Theorem 14.2*

When the limit of $\mathrm{E}\mathcal{H}_{nT}$ is singular, $\delta_0$ and $\lambda_0$ cannot be identified from $T_{2,n,T}(\delta,\lambda)$ in Appendix D.2. Identification requires that the limit of $T_{1,n}(\lambda,\sigma^2)$ is strictly less than zero whenever $(\lambda,\sigma^2) \neq (\lambda_0,\sigma_0^2)$. Thus, the identification will just be from the likelihood function Equation 14.51. By concentrating out

$\sigma^2$ in Equation 14.51, we have the concentrated log-likelihood function

$$\begin{aligned}\ln L_{p,n}(\lambda) &= -\frac{n^*}{2}(\ln(2\pi)+1) - \frac{n^*}{2}\ln\hat{\sigma}^2_{nt}(\lambda) - (n-n^*)\ln(1-\lambda) + \ln|S_n(\lambda)| \\ &= -\frac{n^*}{2}(\ln(2\pi)+1) - \frac{n^*}{2}\ln\hat{\sigma}^2_{nt}(\lambda) + \ln|S^*_n(\lambda)|\end{aligned}$$

from Equation 14.30, where $\hat{\sigma}^2_{nt}(\lambda) = \frac{1}{n^*}V'_{nt}(\lambda)J^*_n V_{nt}(\lambda)$. Also, we have the corresponding $Q_n(\lambda) = \max_{\sigma^2} \mathrm{E}(\ln L_{p,n}(\lambda,\sigma^2)) = -\frac{n^*}{2}(\ln(2\pi)+1) - \frac{n^*}{2}\ln\sigma^2_n(\lambda) + \ln|S^*_n(\lambda)|$. Identification of $\lambda_0$ requires that $\lim_{n\to\infty}\frac{1}{n^*}[Q_n(\lambda) - Q_n(\lambda_0)] \neq 0$ whenever $\lambda \neq \lambda_0$, which is equivalent to

$$\frac{1}{n^*}\ln\left|\sigma^2_0 S^{*\prime -1}_n S^{*-1}_n\right| - \frac{1}{n^*}\ln\left|\sigma^2_n(\lambda)S^{*\prime -1}_n(\lambda)S^{*-1}_n(\lambda)\right| \neq 0 \text{ for } \lambda \neq \lambda_0.$$

After $\lambda_0$ is identified, $\sigma^2_0$ is then identified. Also, given $\lambda_0$, $\delta_0$ can then be identified from $\lim_{T\to\infty} T_{2,n,T}(\delta,\lambda)$. Combined with uniform convergence and equicontinuity, the consistency follows. ■

### D.4 Proof of Theorem 14.3

From Equation 14.13,

$$J^*_n\tilde{Z}_{nt} = J^*_n\tilde{Z}^{(c)}_{nt} - (J^*_n\bar{U}_{nT,-1},\ J^*_n W_n\bar{U}_{nT,-1},\ \mathbf{0}_{n\times k}), \tag{14.52}$$

where $J^*_n\tilde{Z}^{(c)}_{nt}$ is uncorrelated with $V_{nt}$ and the remaining term is correlated with $V_{nt}$ when $t \leq T-1$. For the score decomposition $\frac{1}{\sqrt{n^*T}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta} = \frac{1}{\sqrt{n^*T}}\frac{\partial \ln L^{(c)}_{n,T}(\theta_0)}{\partial\theta} - \Delta_{nT}$ in Equation 14.14, the first term is a linear and quadratic form of $V_{nt}$, and the asymptotic distribution can be derived from the CLT for martingale difference arrays (Lemma 14.3). Hence,

$$\frac{1}{\sqrt{n^*T}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta} + \Delta_{nT} \xrightarrow{d} N(0, \Sigma_{\theta_0} + \Omega_{\theta_0}).$$

For $\Delta_{nT}$, from Equation 14.36 in Lemma 14.1 and Equation 14.38 in Lemma 14.2, we have $\Delta_{nT} = \sqrt{\frac{n^*}{T}}a_{\theta_0,n} + O(\sqrt{\frac{n^*}{T^3}}) + O_p(\frac{1}{\sqrt{T}})$ where $a_{\theta_0,n}$ specified in Equation 14.18 is $O(1)$.

The Taylor expansion gives $\sqrt{n^*T}(\hat{\theta}_{nT} - \theta_0) = (-\frac{1}{n^*T}\frac{\partial^2 \ln L_{n,T}(\bar{\theta}_{nT})}{\partial\theta\partial\theta'})^{-1}\frac{1}{\sqrt{n^*T}} \times \frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta}$, where $\theta_{nT}$ lies between $\theta_0$ and $\hat{\theta}_{nT}$. Similar to Lee and Yu (2010a), we have $\hat{\theta}_{nT} - \theta_0 = O_p(\max(\frac{1}{\sqrt{n^*T}}, \frac{1}{T}))$. Using the fact that $(-\frac{1}{n^*T}\frac{\partial^2 \ln L_{n,T}(\bar{\theta}_{nT})}{\partial\theta\partial\theta'})^{-1} = \Sigma^{-1}_{\theta_0,nT} + O_p(\max(\frac{1}{\sqrt{n^*T}}, \frac{1}{T}))$, given that $\Sigma_{\theta_0,nT}$ is nonsingular and its inverse is of order $O(1)$, we have

$$\sqrt{n^*T}(\hat{\theta}_{nT}-\theta_0)=\left(-\frac{1}{n^*T}\frac{\partial^2\ln L_{n,T}(\bar{\theta}_{nT})}{\partial\theta\partial\theta'}\right)\cdot\left(\frac{1}{\sqrt{n^*T}}\frac{\partial\ln L^{(c)}_{n,T}(\theta_0)}{\partial\theta}-\Delta_{nT}\right)$$

$$=\Sigma^{-1}_{\theta_0,nT}\cdot\frac{1}{\sqrt{n^*T}}\frac{\partial\ln L^{(c)}_{n,T}(\theta_0)}{\partial\theta}+O_p\left(\max\left(\frac{1}{\sqrt{n^*T}},\frac{1}{T}\right)\right)$$

$$\cdot\frac{1}{\sqrt{n^*T}}\frac{\partial\ln L^{(c)}_{n,T}(\theta_0)}{\partial\theta}-\Sigma^{-1}_{\theta_0,nT}\cdot\Delta_{nT}$$

$$-O_p\left(\max\left(\frac{1}{\sqrt{n^*T}},\frac{1}{T}\right)\right)\cdot\Delta_{nT},$$

which implies that

$$\sqrt{n^*T}(\hat{\theta}_{nT}-\theta_0)+\Sigma^{-1}_{\theta_0,nT}\cdot\Delta_{nT}+O_p\left(\max\left(\frac{1}{\sqrt{n^*T}},\frac{1}{T}\right)\right)\cdot\Delta_{nT}$$

$$=(\Sigma^{-1}_{\theta_0,nT}+o_p(1))\cdot\frac{1}{\sqrt{n^*T}}\frac{\partial\ln L^{(c)}_{n,T}(\theta_0)}{\partial\theta}. \quad (14.53)$$

As $\Sigma_{\theta_0}=\lim_{T\to\infty}\Sigma_{\theta_0,nT}$ exists, then using $\Delta_{nT}=\sqrt{\frac{n^*}{T}}a_{\theta_0,n}+O(\sqrt{\frac{n^*}{T^3}})+O_p(\frac{1}{\sqrt{T}})$ with $a_{\theta_0,n}=O(1)$ and $\frac{1}{\sqrt{n^*T}}\frac{\partial\ln L^{(c)}_{n,T}(\theta_0)}{\partial\theta}\xrightarrow{d}N(0,\Sigma_{\theta_0}+\Omega_{\theta_0})$, the result in the theorem follows. ■

### *D.5 Proof for Theorem 14.4*

Theorem 14.3 states that $\sqrt{n^*T}(\theta_{nT}-\theta_0)+\sqrt{\frac{n^*}{T}}b_{\theta_0,nT}+O_p(\max(\sqrt{\frac{n^*}{T^3}},\frac{1}{\sqrt{T}}))\xrightarrow{d}N(0,\Sigma^{-1}_{\theta_0}(\Sigma_{\theta_0}+\Omega_{\theta_0})\Sigma^{-1}_{\theta_0})$. As the bias corrected estimator is $\hat{\theta}^1_{nT}=\hat{\theta}_{nT}+\frac{1}{T}(-\frac{1}{n^*T}\mathrm{E}\frac{\partial^2\ln L_{nT}(\hat{\theta}_{nT})}{\partial\theta\partial\theta'})^{-1}\cdot a_n(\hat{\theta}_{nT})$ where $a_n(\theta)=a_{\theta,n}$, we have $\sqrt{n^*T}(\hat{\theta}^1_{nT}-\theta_0)\xrightarrow{d}N(0,\Sigma^{-1}_{\theta_0}(\Sigma_{\theta_0}+\Omega_{\theta_0})\Sigma^{-1}_{\theta_0})$ if

$$\sqrt{\frac{n^*}{T}}\left(\left(-\frac{1}{n^*T}\mathrm{E}\frac{\partial^2\ln L_{nT}(\hat{\theta}_{nT})}{\partial\theta\partial\theta'}\right)^{-1}a_n(\hat{\theta}_{nT})-\Sigma^{-1}_{\theta_0,nT}a_n(\theta_0)\right)\xrightarrow{p}0 \quad (14.54)$$

and $\frac{n^*}{T^3}\to 0$. Similar to Lee and Yu (2010a), Equation 14.54 can be proved under the assumed regularity conditions. ■

## References

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

Amemiya, T. 1971. The estimation of the variances in a variance-components model. *International Economic Review* 12:1–13.

Anderson, T. W. 1959. On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics* 30:676–687.

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

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

Hahn, J., and 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.

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

Hamilton, J. 1994. *Times Series Analysis.* Princeton, NJ: Princeton University Press.

Horn, R., and C. Johnson. 1985. *Matrix Algebra.* New York: Cambridge University Press.

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

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

Keller, W., and 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. Manuscript, Department of Finance, University of Notre Dame, South Bend, IN.

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.

Lee, L. F., and J. Yu. 2010a. A spatial dynamic panel data model with both time and individual fixed effects. *Econometric Theory* 26:564–597.

Lee, L. F., and J. Yu. 2010b. Some recent developments in spatial panel data models. *Regional Science and Urban Economics* 40:255–271.

Nielsen, B. 2001. The asymptotic distribution of unit root tests of unstable autoregressive processes. *Econometrica* 69:211–219.

Nielsen, B. 2005. Strong consistency results for least squares estimators in general vector autoregressions with deterministic terms. *Econometric Theory* 21:534–561.

Ord, J. K. 1975. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70:120–297.

Phillips, P. C. B., and T. Magdalinos. 2007. Limit theory for moderate deviations from a unit root. *Journal of Econometrics* 136:115–130.

Phillips, P. C. B., and H. R. Moon. 1999. Linear regression limit theory for nonstationary panel data. *Econometrica* 67:1057–1111.

Rothenberg, T. J. 1971. Identification in parametric models. *Econometrica* 39:577–591.

Sims, C. A., J. H. Stock, and M. W. Watson. 1990. Inference in linear time series models with some unit roots. *Econometrica* 58:113–144.

Su, L., and Z. Yang. 2007. QML estimation of dynamic panel data models with spatial errors. Manuscript, Singapore Management University.

Theil, H. 1971. *Principles of Econometrics*. New York: John Wiley & Sons.

Wallace, T. D., and A. Hussain. 1969. The use of error components models in combining cross-section and time-series data. *Econometrica* 37:55–72.

White, J. S. 1958. The limiting distribution of the serial correlation coefficient in the explosive case I. *Annals of Mathematical Statistics* 29:1188–1197.

White, J. S. 1959. The limiting distribution of the serial correlation coefficient in the explosive case II. *Annals of Mathematical Statistics* 30:831–834.

White, H. 1994. *Estimation, Inference and Specification Analysis.* New York: Cambridge University Press.

Yu, J., R. de Jong, and 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. Manuscript, The Ohio State University, Columbus, OH.

Yu, J., R. de Jong, and 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., and L. F. Lee. 2010. Estimation of unit root spatial dynamic panel data models. *Econometric Theory*, forthcoming. doi:10.1017/s0266466609990600