Journal of Econometrics 167 (2012) 16–37

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

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

Estimation for spatial dynamic panel data with fixed effects: The case of spatial cointegration[☆]

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

[a] Guanghua School of Management, Peking University, China

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

ARTICLE INFO

Article history:
Received 8 May 2008
Received in revised form
11 May 2011
Accepted 20 May 2011
Available online 4 November 2011

JEL classification:
C13
C23
R15

Keywords:
Spatial autoregression
Dynamic panels
Fixed effects
Quasi-maximum likelihood estimation
Bias correction
Generalized method of moments
Spatial cointegration
Unit root

ABSTRACT

Yu et al. (2008) establish asymptotic properties of quasi-maximum likelihood estimators for a stable spatial dynamic panel model with fixed effects when both the number of individuals n and the number of time periods T are large. This paper investigates unstable cases where there are unit roots generated by temporal and spatial correlations. We focus on the spatial cointegration model where some eigenvalues of the data generating process are equal to 1 and the outcomes of spatial units are cointegrated as in a vector autoregressive system. The asymptotics of the QML estimators are developed by reparameterization, and bias correction for the estimators is proposed. We also consider the 2SLS and GMM estimations when T could be small.

1. Introduction

Spatial econometrics deals with (spatial) interactions of economic units in cross section data, and it can be extended to panel data models. Baltagi et al. (2003, 2007) and Kapoor et al. (2007) investigate the estimation and testing of different spatial panel data models under random effects specifications. An alternative specification for a panel data model assumes fixed effects, which has the advantage of robustness in that the individual effects are allowed to correlate with included regressors in the model (Hausman, 1978). For static models, Lee and Yu (2010a) study the spatial panel data model with fixed effects, spatial lags and spatial autoregressive (SAR) disturbances. For dynamic models, we have the following spatial dynamic panel data (SDPD) model:

$$\begin{aligned} Y_{nt} &= \lambda_{10} W_n Y_{nt} + \gamma_0 Y_{n,t-1} + \rho_0 W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} \\ &\quad + \alpha_{t0} l_n + U_{nt} \quad \text{with } U_{nt} = \lambda_{20} M_n U_{nt} + V_{nt} \qquad (1) \end{aligned}$$

for $t = 1, 2, \ldots, T$. Here, $Y_{nt} = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ and $V_{nt} = (v_{1t}, v_{2t}, \ldots, v_{nt})'$ are $n \times 1$ column vectors, and v_{it}'s are i.i.d. across i and t with zero mean and variance σ_0^2. The W_n and M_n are $n \times n$ nonstochastic spatial weights matrices, X_{nt} is an $n \times k$ matrix of nonstochastic regressors,[1] and $\mathbf{c}_{n0}$ is an $n \times 1$ column vector of individual fixed effects. α_{t0} is a scalar representing a time effect, and l_n is an $n \times 1$ column vector of ones. The disturbances U_{nt} follow an SAR process with the spatial weights matrix M_n,

☆ We would like to thank participants of seminars at Ohio State University, University of Michgan, Yale University, University of Washington at Seattle, and the Third Symposium on Econometric Theory and Applications at HKUST; we would also like to thank Professor Peter Phillips and the co-editor, Professor Peter M. Robinson, of this journal for helpful comments as well as two anonymous referees. Yu acknowledges funding from National Science Foundation of China (Grant No. 71171005) and support from Center for Statistical Science of Peking University.

* Corresponding author. Tel.: +1 614 292 5508; fax: +1 614 292 4192.
E-mail addresses: jihai.yu@gmail.com (J. Yu), de-jong.8@osu.edu (R. de Jong), lee.1777@osu.edu (L.-f. Lee).

1 Due to the presence of fixed individual and time effects, X_{nt} does not include time invariant or individually invariant regressors.

0304-4076/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2011.05.014

which may or may not be the same as W_n.[2] The W_n and M_n are row-normalized with zero diagonals. A row-normalized W_n (resp. M_n) has the property $W_n l_n = l_n$ (resp. $M_n l_n = l_n$), which ensures that all the weights are between 0 and 1, and a weighting operation can be interpreted as an average of the neighboring values. The practical specification of Ord (1975) is constructed by the row-normalization of a symmetric matrix. Ord (1975) showed that such a row-normalized matrix is diagonalizable and all of its eigenvalues are real. We consider such a case that W_n is diagonalizable with all real eigenvalues.

Define $S_n(\lambda_1) = I_n - \lambda_1 W_n$ for an arbitrary value λ_1 and $S_n \equiv S_n(\lambda_{10})$ at the true parameter value. Presuming S_n is invertible,[3] by denoting $A_n = S_n^{-1}(\gamma_0 I_n + \rho_0 W_n)$, (1) can be rewritten as

$$Y_{nt} = A_n Y_{n,t-1} + S_n^{-1}(X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + U_{nt}). \quad (2)$$

Assumption 1. (i) W_n and M_n are row-normalized nonstochastic spatial weights matrices with zero diagonals; (ii) $(I_n - \lambda_{10} W_n)$ and $(I_n - \lambda_{20} M_n)$ are invertible with $|\lambda_{j0}| < 1$ for $j = 1$ and 2; (iii) W_n is diagonalizable, i.e., $W_n = \Gamma_n \varpi_n \Gamma_n^{-1}$ where Γ_n is the eigenvector matrix, ϖ_n is the diagonal eigenvalue matrix, and all the eigenvalues are real.

With ϖ_n in Assumption 1, the eigenvalue matrix of A_n is $D_n = (I_n - \lambda_{10}\varpi_n)^{-1}(\gamma_0 I_n + \rho_0 \varpi_n)$. Depending on the eigenvalues of A_n, we have three cases of SDPD models. When all the roots are less than 1 in absolute value, we call it a stable case. When all the roots are equal to 1, we term it a pure unit root case, which generalizes the unit root dynamic panel data model in the time series literature to include spatial elements. When some of the roots (but not all) are equal to 1, we define it as a spatial cointegration case, where the unit roots in the process are generated with mixed time and spatial dimensions. This terminology of spatial cointegration can be justified by regarding the spatial panels as a vector autorgressive (VAR) system and we can show the existence of cointegration relationships among spatial units (see Section 2.2).

For the SDPD model in this paper, the cointegrating space is completely known (determined by the spatial weights matrix) when cointegration occurs, while in the conventional cointegration time series literature it is the main object of inference (see Johansen, 1991 and Phillips, 1995, etc.). In the conventional cointegration, the dimension of a VAR process is fixed and relatively small; in the present paper, it can be large and asymptotically it tends to infinity. Such a system is of particular interest for the study of market integration across regions. Due to different eigenvalue matrices of the models, the asymptotics of the spatial cointegration case differs from those of the stable case and the pure unit root case. In Yu et al. (2008), the consistency and asymptotic distribution of the maximum likelihood (ML) and quasi-maximum likelihood (QML) estimators are established for the stable SDPD model. These estimators are $\sqrt{nT}$-consistent, and have biases of the order $O(1/T)$. Bias correction for the estimator is possible; when T grows faster than $n^{1/3}$, the bias corrected estimator yields a centered confidence interval.

When there are unit roots in the process, we find that the appropriate reparameterization motivated by Sims et al. (1990) provides a good device for our investigation.[4] The current paper's main finding is that, for the spatial cointegration case, all the estimates are $\sqrt{nT}$-consistent; but the variance matrix of the estimates is singular in the limit, and the sum of the spatial and dynamic effects estimates is superconsistent. For the pure unit root case, the dynamic effect estimate is superconsistent and other estimates are $\sqrt{nT}$-consistent, while the sum of spatial lag effect and spatial time lag effect estimates is also superconsistent. Bias corrections for spatial cointegration or pure unit root cases can also be constructed, but with different procedures as compared to the stable case.[5]

There is an interest in nonstationary panels for both independent panels and cross-sectionally correlated panels with common time factors. That literature is now large. The present paper differs from that literature in that it covers an unstable panel data model with the cross sectional dependence specified by local spatial interactions as well as (fixed) time effects. There are already extensive empirical applications for nonstationary panel data.[6] We expect that our model contributes to existing nonstationary panel data models of empirical interest. For example, Keller and Shiue (2007) investigate how the spatial feature influences the expansion of interregional trade by studying historical data on Chinese rice markets. The spatial effects are significant and the sum of the estimates of the spatial and dynamic effects is approximately 1. By applying the SDPD model to the rice price, we find that the regional prices follow a spatial cointegration process (Lee and Yu, 2010b).

The present paper is organized as follows. Section 2 further discusses the spatial cointegration model. Section 3 considers the QML estimation of the spatial cointegration model. We derive the consistency and asymptotic distribution of the parameter estimates via a reparameterization approach. The asymptotics of the estimator relies on both n and T tending to infinity. Also, a bias correction procedure is proposed. The asymptotics for the pure unit root case is summarized in Appendix D. Section 4 presents the GMM estimation for comparison. For the asymptotics of the GMM estimator, T is allowed to be finite and fixed. Monte Carlo results on the finite sample properties of these estimators are presented in Section 5. Conclusions are in Section 6. Some useful lemmas and proofs are collected in Appendices A–D.

2. The spatial cointegration model

The following subsections provide some different views on implied structures of the model (1).

2 The SAR disturbances and the addition of time effects in (1) are not covered in the earlier version of this paper, titled "Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large: a nonstationary case". These extensions become tractable with the reparameterization approach in the current version.

3 With S_n being invertible and W_n being row-normalized, the case of $\lambda_{10} = 1$ is ruled out.

4 The earlier version of this paper involves complicated algebraic manipulation when reparameterizations are not used, and so does Yu and Lee (2010) for the pure unit root case. Both papers have not considered time effects and spatial disturbances.

5 If the spatial cointegration feature were ignored and the model were treated as a stable one, then singularity of the limiting variance matrix of the (original) parameter vector would be ignored. In that case, asymptotic inference might be problematic if the limiting variance matrix were utilized as if it were a regular nonsingular one. In addition, due to different eigenvalue matrices, the bias correction procedure for the spatial cointegration case is different from those of stable and pure unit root cases. The bias correction would be important in order to have a properly centered limiting distribution for the estimator, which is useful for statistical inference. Thus, ignoring spatial cointegration (treating the model as a stable one or a pure unit root one) would result in an inappropriate bias correction procedure and the $O(1/T)$ bias might remain.

6 The applications include purchasing power parity, growth and convergence, money demand, exchange rate model, inflation rate convergence, interest rate, health care expenditure, hysteresis in unemployment, etc. See Choi (2006).

2.1. Decomposition of Y_{nt}

From Assumption 1 and the reduced form (2), we have $A_n = \Gamma_n D_n \Gamma_n^{-1}$. When W_n is row-normalized, all of its eigenvalues are less than or equal to 1 in absolute value by the spectral radius theorem, where it definitely has some eigenvalues being 1 because $W_n l_n = l_n$. Let m_n be the number of unit eigenvalues of W_n and, without loss of generality, let the first m_n eigenvalues of W_n be 1. Thus, the eigenvalue matrix of W_n is $\varpi_n = \text{diag}\{l'_{m_n}, \omega_{n,m_n+1}, \ldots, \omega_{nn}\}$ where l_{m_n} is an $m_n \times 1$ vector of ones and $|\omega_{nj}| < 1$ for $j = m_n + 1, \ldots, n$. By defining $\mathbb{J}_n = \text{diag}\{l'_{m_n}, 0, \ldots, 0\}$, ϖ_n can be decomposed into two parts: $\varpi_n = \mathbb{J}_n + \varpi_n^s$ with $\mathbb{J}_n$ corresponding to the unit eigenvalues, and the ϖ_n^s corresponding to eigenvalues smaller than 1 in absolute value. It follows that the eigenvalues D_n are $d_{nj} = \frac{\gamma_0+\rho_0\omega_{nj}}{1-\lambda_{10}\omega_{nj}}$ for $j = 1, \ldots, n$, which can be decomposed into two corresponding parts: $D_n = (\frac{\gamma_0+\rho_0}{1-\lambda_{10}})\mathbb{J}_n + \tilde{D}_n$, where $\tilde{D}_n = \text{diag}\{0, \ldots, 0, d_{n,m_n+1}, \ldots, d_{nn}\}$ with the nonzero diagonal elements corresponding to the eigenvalues of W_n less than 1. Denote $A_n^h = A_n \cdots A_n$ as the h-fold products of A_n and similarly for other matrices. As $\mathbb{J}_n \cdot \tilde{D}_n = \mathbf{0}$, we have $A_n^h = (\frac{\gamma_0+\rho_0}{1-\lambda_{10}})^h \Gamma_n \mathbb{J}_n \Gamma_n^{-1} + B_n^h$ where $B_n^h = \Gamma_n \tilde{D}_n^h \Gamma_n^{-1}$ for any $h = 1, 2, \ldots$.

Assumption 2. $\rho_0 + \gamma_0 + \lambda_{10} = 1$ with $\gamma_0 < 1$ and $\rho_0 < 1$.[7]

Assumption 2 together with λ_{10} in Assumption 1 implies that the largest eigenvalue of A_n is 1 and the remaining eigenvalues are smaller than 1 in absolute value. This is so because the derivative of $d_{nj} = \frac{\gamma_0+\rho_0\omega_{nj}}{1-\lambda_{10}\omega_{nj}}$ with respect to ω_{nj} is $\frac{\rho_0+\gamma_0\lambda_{10}}{(1-\lambda_{10}\omega_{nj})^2}$. Under $\gamma_0 + \rho_0 + \lambda_{10} = 1$, $\rho_0 + \gamma_0\lambda_{10} = (1-\gamma_0)(1-\lambda_{10})$. Thus, d_{nj} is an increasing function of ω_{nj}. With ω_{nj} in $[-1, 1]$, the corresponding range of d_{nj} is $[\frac{\gamma_0-\rho_0}{1+\lambda_{10}}, 1]$. Under $\gamma_0 + \rho_0 + \lambda_{10} = 1$, $\frac{\gamma_0-\rho_0}{1+\lambda_{10}} > -1$ if and only if $\rho_0 < 1$. Assumption 2 states explicitly that the spatial cointegration model is the main interest in this paper. When $\gamma_0 + \rho_0 + \lambda_{10} = 1$ and $\gamma_0 = 1$, we have the pure unit root case in that all the eigenvalues are equal to 1. When $\gamma_0 + \rho_0 + \lambda_{10} \neq 1$, we have either a stable or explosive case.

For $t \geq 0$, Y_{nt} can be decomposed into a sum of three possible different components. By denoting $W_n^u = \Gamma_n \mathbb{J}_n \Gamma_n^{-1}$, as is derived in Appendix C.1, we have

$$Y_{nt} = Y_{nt}^u + Y_{nt}^s + Y_{nt}^\alpha \tag{3}$$

under $\gamma_0 + \rho_0 + \lambda_{10} = 1$ where

$$Y_{nt}^u = W_n^u \left\{ Y_{n,-1} + \frac{1}{(1-\lambda_{10})} \sum_{h=1}^{t} \left(\mathbf{c}_{n0} + X_{n,t-h}\beta_0 + U_{n,t-h} \right) \right\},$$

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

$$Y_{nt}^\alpha = \frac{1}{(1-\lambda_{10})} l_n \sum_{h=0}^{t} \alpha_{t-h,0}.$$

When $\gamma_0 + \rho_0 + \lambda_{10} = 1$ but $\gamma_0 \neq 1$, the unit eigenvalues of A_n correspond exactly to those unit eigenvalues of W_n via the relation $D_n = (I_n - \lambda_{10}\varpi_n)^{-1}(\gamma_0 I_n + \rho_0 \varpi_n)$. W_n has some unit eigenvalues, but not all of them are equal to 1.[8] Hence, some eigenvalues of A_n, but not all, are equal to 1. The Y_{nt}^u can be an unstable component. If $\mathbf{c}_{n0}$ and/or the time mean of $X_{nt}\beta_0$ are nonzero, the $\sum_{h=1}^{t} \left(\mathbf{c}_{n0} + X_{n,t-h}\beta_0 \right)$ will generate a time trend. The $\sum_{h=1}^{t} U_{n,t-h}$ will generate a stochastic trend. These imply the unstability of Y_{nt}^u. The component Y_{nt}^α captures the time effect due to the time dummies. As absolute values of the elements in $\tilde{D}_n$ are less than 1, Y_{nt}^s will be a stable component.

Remark. The unit root case has all eigenvalues of A_n being 1. It occurs when $\gamma_0 + \rho_0 + \lambda_{10} = 1$ and $\gamma_0 = 1$, because $A_n = (I_n - \lambda_{10}W_n)^{-1}(\gamma_0 I_n + \rho_0 W_n) = (I_n - \lambda_{10}W_n)^{-1}(I_n - \lambda_{10}W_n) = I_n$. For this unit root case, the unit eigenvalues of A_n are not linked to the eigenvalues of W_n. Because W_n^u is defined from W_n, the decomposition in (3) is not revealing for the unit root case and, instead, one has

$$Y_{nt} = Y_{n,t-1} + S_n^{-1}(X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + U_{nt}). \tag{4}$$

2.2. Error Correction Model (ECM) representation

The model (1) has a VAR form. It is a large equation system as the number of spatial units n is usually large; however, it has restricted coefficient matrices. The model has a revealing ECM representation.

Denote $\Delta Y_{nt} = Y_{nt} - Y_{n,t-1}$ as a difference in time. From (2), we have $\Delta Y_{nt} = (A_n - I_n)Y_{n,t-1} + S_n^{-1}(X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + U_{nt})$. As $A_n - I_n = S_n^{-1}[(\rho_0 + \lambda_{10})W_n - (1-\gamma_0)I_n]$, (1) has the ECM representation

$$\Delta Y_{nt} = S_n^{-1}[(\rho_0 + \lambda_{10})W_n - (1-\gamma_0)I_n]Y_{n,t-1} + S_n^{-1}(X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + U_{nt}).$$

For the spatial cointegration case,

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

Because $W_n W_n^u = W_n^u$ and $W_n l_n = l_n$, the components Y_{nt}^u and Y_{nt}^α have the identities

$$W_n Y_{nt}^u = Y_{nt}^u, \qquad W_n Y_{nt}^\alpha = Y_{nt}^\alpha. \tag{5}$$

In this situation, $(I_n - W_n)$ is a cointegrating matrix for Y_{nt} because $(I_n - W_n)Y_{nt} = (I_n - W_n)Y_{nt}^s$, which depends only on the stable component. Thus, Y_{nt} is spatially cointegrated. The cointegration rank of $I_n - W_n = \Gamma_n(I_n - \varpi_n)\Gamma_n^{-1}$ is equal to $n - m_n$, which is the number of eigenvalues of W_n different from 1.

Remark. For the comparison, the pure unit root case has $\gamma_0 = 1$ and $A_n = I_n$. From (4), its ECM representation is

$$\Delta Y_{nt} = S_n^{-1}(X_{nt}\beta_0 + c_{n0} + \alpha_{t0} l_n + V_{nt}).$$

Each outcome of a spatial unit is unstable, and there is no cointegrating relation across spatial units.

2.3. Rotation and linear combinations of Y_{nt}

We can understand more on the structure of the model by taking rotation and linear combinations of the outcomes across spatial units.[9] As $W_n = \Gamma_n \varpi_n \Gamma_n^{-1}$ where all the eigenvalues of ϖ_n are real, the corresponding eigenvectors of Γ_n will be real. By premultiplying Γ_n^{-1} to (1),

$$\Gamma_n^{-1} Y_{nt} = \lambda_{10}\varpi_n \Gamma_n^{-1} Y_{nt} + \gamma_0 \Gamma_n^{-1} Y_{n,t-1} + \rho_0 \varpi_n \Gamma_n^{-1} Y_{n,t-1} + \Gamma_n^{-1} X_{nt}\beta_0 + \Gamma_n^{-1}\mathbf{c}_{n0} + \alpha_{t0}\sqrt{n} e_n + \Gamma_n^{-1} U_{nt},$$

[7] It is apparent that when $\gamma_0 + \rho_0 + \lambda_{10} = 1$, $\gamma_0 < 1$, $\rho_0 < 1$ and $\lambda_{10} < 1$, they imply $|\gamma_0| < 1$, $|\rho_0| < 1$ and $|\lambda_{10}| < 1$.

[8] As W_n has a zero diagonal in a general SAR model specification, the sum of all eigenvalues of W_n is zero.

[9] This view is suggested by a referee. We appreciate having his/her insight.

where $e_n = (1, 0, \ldots, 0)'$ is an $n \times 1$ vector, because $\Gamma_n^{-1} l_n = \sqrt{n} e_n$ as $\frac{1}{\sqrt{n}} l_n$ is the first eigenvector in Γ_n. By the partition $\varpi_n = [\mathrm{diag}(I_{m_n}, \varpi_{2n})]$ and, conformably, $\Gamma_n^{-1} = [\Gamma_{1n}^{\prime(1)}, \Gamma_{2n}^{\prime(2)}]'$, we have

$$\begin{aligned}\Gamma_{1n}^{(1)} Y_{nt} &= (1-\lambda_{10})^{-1}\Big[(\gamma_0+\rho_0)\Gamma_{1n}^{(1)} Y_{n,t-1} + \Gamma_{1n}^{(1)} X_{nt}\beta_0 \\ &\quad + \Gamma_{1n}^{(1)} \mathbf{c}_{n0} + \alpha_{t0}\sqrt{n} e_{n1} + \Gamma_{1n}^{(1)} U_{nt}\Big], \qquad (6)\end{aligned}$$

$$\begin{aligned}\Gamma_{2n}^{(2)} Y_{nt} &= (I_{2n}-\lambda_{10}\varpi_{2n})^{-1}\Big[(\gamma_0 I_{2n}+\rho_0\varpi_{2n})\Gamma_{2n}^{(2)} Y_{n,t-1} \\ &\quad + \Gamma_{2n}^{(2)} X_{nt}\beta_0 + \Gamma_{2n}^{(2)} \mathbf{c}_{n0} + \Gamma_{2n}^{(2)} U_{nt}\Big], \qquad (7)\end{aligned}$$

where $e_{n1} = (1, 0, \ldots, 0)'$ is an $m_n \times 1$ vector. For the unstable situation with $\gamma_0 + \rho_0 + \lambda_{10} = 1$, $(1-\lambda_{10})^{-1}(\gamma_0+\rho_0) = 1$. From (6) and (7), we can see that after Y_{nt} is transformed to $\Gamma_n^{-1} Y_{nt}$, the $m_n \times 1$ subvector $\Gamma_{1n}^{(1)} Y_{nt}$ is unstable and the remaining $(n-m_n)\times 1$ subvector $\Gamma_{2n}^{(2)} Y_{nt}$ is stable.

Denote $Y_{1nt} = \Gamma_{1n}^{(1)} Y_{nt}$ and define similar notations for other variables. By reparameterizing $\gamma_1 = (1-\lambda_{10})^{-1}(\gamma_0+\rho_0)$ and $\beta_1 = (1-\lambda_{10})^{-1}\beta_0$, etc., we can rewrite (6) as

$$Y_{1nt} = \gamma_1 Y_{1n,t-1} + X_{1nt}\beta_1 + \mathbf{c}_{1n} + \alpha_{1nt} + U_{1nt}. \qquad (8)$$

As each variable in Y_{1nt} has a unit root because $\gamma_1 = 1$ and Y_{1nt} is dominated by the linear trend[10] generated by the drift term $\mathbf{c}_{1n}$ and the exogenous variable X_{1nt}, we may expect that an estimate of γ_1 would be superconsistent. As γ_1 is a function of λ_{10}, γ_0 and ρ_0, we may expect that corresponding estimates of λ_{10}, γ_0 and ρ_0 would have some nonstandard features as compared with the stable case. Denote $Y_{2nt} = \Gamma_{2n}^{(2)} Y_{nt}$ and similarly for other variables. (7) can be written as

$$\begin{aligned}(I_{2n}-\lambda_{10}\varpi_{2n})Y_{2nt} &= (\gamma_0 I_{2n}+\rho_0\varpi_{2n})Y_{2n,t-1} \\ &\quad + X_{2nt}\beta_0 + \mathbf{c}_{2n} + U_{2nt}. \qquad (9)\end{aligned}$$

For the equations in (9), under Assumption 2, the coefficient matrix $(I_{2n}-\lambda_{10}\varpi_{2n})^{-1}(\gamma_0 I_{2n}+\rho_0\varpi_{2n})$ of $Y_{2n,t-1}$ in the autoregressive representation of Y_{2nt} is a diagonal matrix with all its elements less than 1 in absolute values. (9) provides the stable components of the transformed $\Gamma_n^{-1} Y_{nt}$. (8) and (9) together provide a system of univariate equations of autoregressive processes. Those in (8) are unit root processes and those in (9) are stable ones. The coefficients of each univariate process can be estimated with either linear or nonlinear regression methods.[11] There are constraints across the univariate processes, which suggest a pooling estimation.

While this pooling approach is revealing, we do not follow it in this paper as the QML approach with reparameterization in the next subsection is more direct without going through several steps of estimation. Possible spatial correlations in the disturbances in U_{nt} can also be easily taken into account in the estimation with the reparameterization approach.

2.4. Reparameterization

Define $Z_{nt} = (Y_{n,t-1}, W_n Y_{n,t-1}, X_{nt})$, and $\theta = (\delta', \lambda_1, \lambda_2, \sigma^2)'$ where $\delta = (\gamma, \rho, \beta')'$. At true values, $\theta_0 = (\delta_0', \lambda_{10}, \lambda_{20}, \sigma_0^2)'$ where $\delta_0 = (\gamma_0, \rho_0, \beta_0')'$. From (5), Z_{nt} can be decomposed into:

$$\begin{aligned}Z_{nt} &= (Y^u_{n,t-1} + Y^\alpha_{n,t-1}, W_n(Y^u_{n,t-1} + Y^\alpha_{n,t-1}), \mathbf{0}_{n\times k}) \\ &\quad + (Y^s_{n,t-1}, W_n Y^s_{n,t-1}, X_{nt}) \\ &= (Y^u_{n,t-1} + Y^\alpha_{n,t-1})(1, 1, 0_{1\times k}) + Z^s_{nt},\end{aligned}$$

where $Z^s_{nt} = (Y^s_{n,t-1}, W_n Y^s_{n,t-1}, X_{nt})$ is the stable component. Denote $G_n = W_n S_n^{-1}$. The identities in (5) imply that

$$S_n^{-1} Y^u_{nt} = \frac{1}{1-\lambda_{10}} Y^u_{nt}, \qquad G_n Y^u_{nt} = \frac{1}{1-\lambda_{10}} Y^u_{nt},$$
$$S_n^{-1} Y^\alpha_{nt} = \frac{1}{1-\lambda_{10}} Y^\alpha_{nt}, \quad \text{and} \quad G_n Y^\alpha_{nt} = \frac{1}{1-\lambda_{10}} Y^\alpha_{nt}.$$

As the reduced form Eq. (2) is $Y_{nt} = S_n^{-1} Z_{nt}\delta_0 + S_n^{-1}(\mathbf{c}_{n0} + \alpha_{t0} l_n + U_{nt})$, we have $W_n Y_{nt} = G_n Z_{nt}\delta_0 + G_n(\mathbf{c}_{n0} + \alpha_{t0} l_n + U_{nt})$. Thus, when $\gamma_0 + \rho_0 + \lambda_{10} = 1$, we have

$$G_n Z_{nt}\delta_0 = (Y^u_{n,t-1} + Y^\alpha_{n,t-1}) + G_n Z^s_{nt}\delta_0.$$

The following implied identities are useful for the subsequent reparameterization approach:

$$W_n Y_{n,t-1} - Y_{n,t-1} = W_n Y^s_{n,t-1} - Y^s_{n,t-1}, \qquad (10)$$

$$W_n Y_{nt} - Y_{n,t-1} = (G_n Z^s_{nt}\delta_0 - Y^s_{n,t-1}) + G_n(\mathbf{c}_{n0} + \alpha_{t0} l_n + U_{nt}), \qquad (11)$$

and

$$\begin{aligned}W_n(Y_{nt} - Y_{n,t-1}) &= (G_n Z^s_{nt}\delta_0 - W_n Y^s_{n,t-1}) \\ &\quad + G_n(\mathbf{c}_{n0} + \alpha_{t0} l_n + U_{nt}). \qquad (12)\end{aligned}$$

In estimation, it is desirable to eliminate the time effect component Y^α_{nt}, especially when T is large. This can be achieved with the deviation from the cross sectional mean operator $J_n = I_n - \frac{1}{n} l_n l_n'$. Let $(F_{n,n-1}, \frac{1}{\sqrt{n}} l_n)$ be the orthonormal eigenvector matrix of J_n where $F_{n,n-1}$ is the $n \times (n-1)$ submatrix corresponding to the eigenvalues of ones. By using $F'_{n,n-1} l_n = 0$, the time effects could be eliminated. Denote $\Upsilon^*_{nt} = F'_{n,n-1}\Upsilon_{nt}$ for any $n \times 1$ vector Υ_{nt}. We have $Y^*_{nt} = Y^{*u}_{nt} + Y^{*s}_{nt}$ because $Y^{*\alpha}_{nt} = 0$ by $F'_{n,n-1} l_n = 0$. Furthermore, denote $W^*_n = F'_{n,n-1} W_n F_{n,n-1}$ and $G^*_n = W^*_n(I_{n-1} - \lambda_{10} W^*_n)^{-1}$. Corresponding to (10)–(12), because $W_n l_n = l_n$ and $M_n l_n = l_n$, one has

$$W^*_n Y^*_{n,t-1} - Y^*_{n,t-1} = W^*_n Y^{*s}_{n,t-1} - Y^{*s}_{n,t-1},$$
$$W^*_n Y^*_{nt} - Y^*_{n,t-1} = (G^*_n Z^{*s}_{nt}\delta_0 - Y^{*s}_{n,t-1}) + G^*_n(\mathbf{c}^*_{n0} + U^*_{nt})$$

and

$$W^*_n(Y^*_{nt} - Y^*_{n,t-1}) = (G^*_n Z^{*s}_{nt}\delta_0 - W^*_n Y^{*s}_{n,t-1}) + G^*_n(\mathbf{c}^*_{n0} + U^*_{nt}),$$

by using the relations $G^*_n = F'_{n,n-1} G_n F_{n,n-1}$ and $F'_{n,n-1} l_n = 0$. Also, $U^*_{nt} = \lambda_{20} M^*_n U^*_{nt} + V^*_{nt}$ where $M^*_n = F'_{n,n-1} M_n F_{n,n-1}$.

For the spatial cointegration case, (1) can be rearranged with reparameterization into

$$\begin{aligned}Y_{nt} &= \lambda_{10}[W_n Y_{nt} - Y_{n,t-1}] + \gamma^*_0 Y_{n,t-1} + \rho_0[W_n Y_{n,t-1} - Y_{n,t-1}] \\ &\quad + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + U_{nt},\end{aligned}$$

for $t = 1, 2, \ldots, T$, where $\gamma^*_0 = \rho_0 + \gamma_0 + \lambda_{10}$ is a reparameterized coefficient. With the time effects eliminated, it becomes

$$\begin{aligned}Y^*_{nt} &= \lambda_{10}[W^*_n Y^*_{nt} - Y^*_{n,t-1}] + \gamma^*_0 Y^*_{n,t-1} + \rho_0[W^*_n Y^*_{n,t-1} - Y^*_{n,t-1}] \\ &\quad + X^*_{nt}\beta_0 + \mathbf{c}^*_{n0} + U^*_{nt}. \qquad (13)\end{aligned}$$

Both the variables $W^*_n Y^*_{nt} - Y^*_{n,t-1}$ and $W^*_n Y^*_{n,t-1} - Y^*_{n,t-1}$ are stable. However, $Y^*_{n,t-1}$ is still unstable, which implies that an estimate of its coefficient γ^*_0 may be superconsistent.

---

10 As pointed by a referee, we note that the linear trend is present due to the autoregressive feature in the regression equation. If the process is $Y_{nt} = X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + F_{nt}$, where $F_{nt} = \lambda_{10} W_n F_{nt} + \gamma_0 F_{n,t-1} + \rho_0 W_n F_{n,t-1} + U_{nt}$, then there would be no linear trend in the process even if $\gamma_0 + \rho_0 + \lambda_{10} = 1$.

11 For (8), the time dummies shall be eliminated for estimation. If the number of unit eigenvalue m_n remains to be 1 for all n, (8) will vanish when time dummies are eliminated by transformation. The remaining equations will solely consist of those stable ones in (9). In this paper, we are interested in the asymptotic where m_n does not vanish relatively to n. The asymptotic for the case with m_n/n vanishing as n tends to infinity seems hard to handle and remains an open issue.

Remark. For the pure unit root SDPD case, i.e., $\rho_0 + \gamma_0 + \lambda_{10} = 1$ and $\gamma_0 = 1$, the desirable reparameterization will be different:

$$Y_{nt} = \lambda_{10} W_n [Y_{nt} - Y_{n,t-1}] + \gamma_0 Y_{n,t-1} + \rho_0^* W_n Y_{n,t-1} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + U_{nt},$$

where $\rho_0^* = \rho_0 + \lambda_{10}$ is a reparameterized coefficient. With the time effects eliminated, it is

$$Y_{nt}^* = \lambda_{10} W_n^* [Y_{nt}^* - Y_{n,t-1}^*] + \gamma_0 Y_{n,t-1}^* + \rho_0^* W_n^* Y_{n,t-1}^* + X_{nt}^*\beta_0 + \mathbf{c}_{n0}^* + U_{nt}^*. \quad (14)$$

The $W_n^*[Y_{nt}^* - Y_{n,t-1}^*]$ consists of stable elements only. However, $Y_{n,t-1}^*$ and $W_n^* Y_{n,t-1}^*$ are unstable, which may imply superconsistent estimates for γ_0 and ρ_0^*.[12]

To analyze the model and estimation, the following regularity conditions are assumed.

Assumption 3. The elements $\{v_{it}\}$, $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$ of V_{nt} in (1) are i.i.d. across i and t with zero mean, variance σ_0^2 and $E\,|v_{it}|^{4+\eta} < \infty$ for some $\eta > 0$.

Assumption 4. X_{nt} and $\mathbf{c}_{n0}$ are nonstochastic with $\sup_{n,T} \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} |x_{it,l}|^{2+\eta} < \infty$ for $l = 1, \ldots, k$ and $\sup_n \frac{1}{n}\sum_{i=1}^{n} |c_{ni}|^{2+\eta} < \infty$ for some $\eta > 0$, where $x_{it,l}$ is the (i, l) element of X_{nt} and c_{ni} is the i th element of $\mathbf{c}_{n0}$.

Assumption 5. $S_n(\lambda_1)$ is invertible for all $\lambda_1 \in \Lambda_1$, and $R_n(\lambda_2)$ is invertible for all $\lambda_2 \in \Lambda_2$. Furthermore, the parameter spaces Λ_1 and Λ_2 are compact[13]; the true parameters λ_{10} and λ_{20} are, respectively, in the interiors of Λ_1 and Λ_2.

Assumption 6. W_n and M_n are uniformly bounded in row and column sums in absolute value (for short, UB).[14] Also $S_n^{-1}(\lambda_1)$ and $R_n^{-1}(\lambda_2)$ are UB, uniformly in $\lambda_1 \in \Lambda_1$ and $\lambda_2 \in \Lambda_2$.

Assumption 7. $\sum_{h=1}^{\infty} \text{abs}(B_n^h)$ is UB where $\left[\text{abs}(B_n^h)\right]_{ij} = \left|(B_n^h)_{ij}\right|$.

Assumption 3 provides essential features of the disturbances of the model. When exogenous variables X_{nt} are included in the model, an empirical moment of a higher than second order restriction is imposed as in Assumption 4.[15] This is also true for the individual effects. The second empirical moment restrictions are useful for some sample statistics to be bounded in our asymptotic analysis. Higher than the second moment restrictions are used in a central limit theorem for a linear and quadratic form in Kelejian and Prucha (2001). Assumption 5 is standard for a regular SAR type model. Assumption 6 is originated by Kelejian and Prucha (1998, 2001) and also used by Lee (2004, 2007). The uniform boundedness condition is to limit the spatial correlation to a manageable degree. Assumption 7 limits the dependence among time series and cross sectional units for the stable component Y_{nt}^s.

3. QML estimation

In this section, we consider the quasi-maximum likelihood approach where individual effects are jointly estimated with common parameters in the model. In order for the QML estimates to be consistent, T is assumed to be large. Denote $\mathbb{X}_{nt} = \sum_{h=0}^{t-1} X_{nh}$ and $\tilde{t} = t - \frac{T+1}{2}$.

Assumption 8. T goes to infinity and n is a strictly increasing function of T, and $\lim_{T\to\infty} \frac{1}{nT^3}\sum_{t=1}^{T} (\mathbf{c}_{n0}\tilde{t} + \tilde{\mathbb{X}}_{nt}\beta_0)' W_n^{u\prime} R_n' J_n R_n W_n^u (\mathbf{c}_{n0}\tilde{t} + \tilde{\mathbb{X}}_{nt}\beta_0) \neq 0$.

Assumption 8 specifies that $n \to \infty$ as $T \to \infty$. We say that $n, T \to \infty$ simultaneously.[16] In Assumption 8, we specify the magnitude of the unstable component Y_{nt}^u. When either $\mathbf{c}_{n0}$ or β_0 is not zero, the deterministic trend generated by either $\mathbf{c}_{n0}$ or X_{nt} (with non-zero mean) provides the dominating factor.[17] As our focus is on the dynamic panel data model where the individual effects are usually present, the assumption of $\mathbf{c}_{n0} \neq 0$ is justifiable.

3.1. Likelihood and concentrated likelihood

Let

$$\mathbb{Z}_{nt}^* = [Y_{n,t-1}^*, W_n^* Y_{n,t-1}^* - Y_{n,t-1}^*, X_{nt}^*]$$

be the regressors after the reparameterization in (13). Denote $\tilde{\Upsilon}_{nt} = \Upsilon_{nt} - \bar{\Upsilon}_{nT}$ for any $n \times 1$ vector Υ_{nt} where $\bar{\Upsilon}_{nT} = \frac{1}{T}\sum_{t=1}^{T} \Upsilon_{nt}$. Furthermore, define $\gamma^* = \gamma + \rho + \lambda_1$, $\delta^* = (\gamma^*, \rho, \beta')'$ and $\theta^* = (\delta^{*\prime}, \lambda_1, \lambda_2, \sigma^2)'$. The concentrated likelihood function of (13) with $\mathbf{c}_{n0}$ concentrated out is

$$\ln L_{nT}(\theta^*) = -\frac{(n-1)T}{2}\ln(2\pi\sigma^2) + T\ln\left|S_n^*(\lambda_1)\right| + T\ln\left|R_n^*(\lambda_2)\right| - \frac{1}{2\sigma^2}\sum_{t=1}^{T} \tilde{V}_{nt}^{*\prime}(\theta^*)\tilde{V}_{nt}^*(\theta^*), \quad (15)$$

where $S_n^*(\lambda_1) = F_{n,n-1}' S_n(\lambda_1) F_{n,n-1}$, $R_n^*(\lambda_2) = F_{n,n-1}' R_n(\lambda_2) F_{n,n-1}$ and $\tilde{V}_{nt}^*(\theta^*) = R_n^*(\lambda_2)\{\tilde{Y}_{nt}^* - \lambda_1[W_n^*\tilde{Y}_{nt}^* - \tilde{Y}_{n,t-1}^*] - \tilde{\mathbb{Z}}_{nt}^*\delta^*\}$. From Lee and Yu (2010a), $\left|S_n^*(\lambda_1)\right| = \frac{1}{1-\lambda_1}\left|S_n(\lambda_1)\right|$, and similarly for $R_n^*(\lambda_2)$. Therefore, the tractability in computing the determinants of $S_n^*(\lambda_1)$ and $R_n^*(\lambda_2)$ is exactly those of $S_n(\lambda_1)$ and $R_n(\lambda_2)$. By using $\sum_{t=1}^{T} \tilde{V}_{nt}^{*\prime}(\theta^*)\tilde{V}_{nt}^*(\theta^*) = \sum_{t=1}^{T} \tilde{V}_{nt}'(\theta^*) J_n \tilde{V}_{nt}(\theta^*)$ where $\tilde{V}_{nt}(\theta^*) = R_n(\lambda_2)\{\tilde{Y}_{nt} - \lambda_1[W_n\tilde{Y}_{nt} - \tilde{Y}_{n,t-1}] - \tilde{\mathbb{Z}}_{nt}\delta^*\}$, (15) can be written as

$$\ln L_{nT}(\theta^*) = -\frac{(n-1)T}{2}\ln(2\pi\sigma^2) + T\ln\left|S_n^*(\lambda_1)\right| + T\ln\left|R_n^*(\lambda_2)\right| - \frac{1}{2\sigma^2}\sum_{t=1}^{T} \tilde{V}_{nt}'(\theta^*) J_n \tilde{V}_{nt}(\theta^*). \quad (16)$$

The first and second order derivatives of (16) are (48) and (49) in Appendix C.2. Define

$$\mathcal{H}_{nT}(\lambda_2) = \frac{1}{(n-1)T}\sum_{t=1}^{T} (\tilde{\mathbb{Z}}_{nt}, (G_n\tilde{Z}_{nt}^s\delta_0 - \tilde{Y}_{n,t-1}^s))' R_n'(\lambda_2) \times J_n R_n(\lambda_2)(\tilde{\mathbb{Z}}_{nt}, (G_n\tilde{Z}_{nt}^s\delta_0 - \tilde{Y}_{n,t-1}^s))$$

[12] For the spatial cointegration case, the (13) implies that only $\rho_0 + \gamma_0 + \lambda_{10}$ is superconsistent. However, for the pure unit root case, the (14) implies that both γ_0 and $\rho_0 + \lambda_{10}$ are superconsistent (so that their sum $\rho_0 + \gamma_0 + \lambda_{10}$ is also superconsistent); this feature cannot be revealed from (13).

[13] Note that in the literature for models with row-normalized spatial matrices, Λ_1 and Λ_2 are typically assumed to be compact subsets of $(-1, 1)$.

[14] 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.

[15] In the previous version, a stronger assumption of boundedness of elements in X_{nt} and $\mathbf{c}_{n0}$ is assumed for convenience. We relax it here to the empirical moment restriction for a more general setting as requested by a referee. Boundedness of relevant variables in this paper can be derived by Lemma 1 in Appendix B. For the CLT that requires $\sup_{n,T}\frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}|\tilde{x}_{it,l}|^{2+\eta} < \infty$ where $\tilde{x}_{it,l} = x_{it,l} - \frac{1}{T}\sum_{s=1}^{T} x_{is,l}$, it can be derived from $\sup_{n,T}\frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}|x_{it,l}|^{2+\eta} < \infty$ by the c_r inequality.

[16] Although the case with large n is the main interest in spatial econometrics, results in this paper also apply to the case with finite n and large T, where a VAR model would be preferable.

[17] When both β_0 and $\mathbf{c}_{n0}$ are zero, the dominant component of the process would be a random walk rather than a linear trend; thus, the inner product of the unstable component vector would then be of the order $O(nT^2)$ instead of $O(nT^3)$. However, the asymptotics of the estimates would be different and it would be more complicated to derive the asymptotic distribution of the estimates under such a setting. We shall leave this for a future pursuit.

$$\Sigma_{\theta_0^*,nT} = \frac{1}{\sigma_0^2}\begin{pmatrix} E\mathcal{H}_{nT} & * & * \\ \mathbf{0}_{1\times(k+3)} & 0 & * \\ \mathbf{0}_{1\times(k+3)} & 0 & 0 \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{(k+2)\times(k+2)} & * & * & * \\ \mathbf{0}_{1\times(k+2)} & \frac{1}{n-1}\mathrm{tr}((\ddot{G}_n' + \ddot{G}_n)J_n\ddot{G}_n) & * & * \\ \mathbf{0}_{1\times(k+2)} & \frac{1}{n-1}\mathrm{tr}((H_n' + H_n)J_n\ddot{G}_n) & \frac{1}{n-1}\mathrm{tr}((H_n'+H_n)J_nH_n) & * \\ \mathbf{0}_{1\times(k+2)} & \frac{1}{\sigma_0^2(n-1)}\mathrm{tr}(J_n\ddot{G}_n) & \frac{1}{\sigma_0^2(n-1)}\mathrm{tr}(J_nH_n) & \frac{1}{2\sigma_0^4} \end{pmatrix} \quad (17)$$

Box I.

and $\mathcal{H}_{nT} = \mathcal{H}_{nT}(\lambda_{20})$. The dominant part of the information matrix $-E\left(\frac{1}{(n-1)T}\frac{\partial^2 \ln L_{nT}(\theta_0^*)}{\partial\theta^*\partial\theta'^*}\right)$ is Eq. (17) given in Box I where $\ddot{G}_n = \ddot{W}_n(I_n - \lambda_{10}\ddot{W}_n)^{-1}$ with $\ddot{W}_n = R_nW_nR_n^{-1}$ and $H_n = M_nR_n^{-1}$. As the submatrix $\tilde{Y}_{n,t-1}$ in $\tilde{\mathbb{Z}}_{nt}$ after reparameterization has an unstable component, we need to rescale $\tilde{\mathbb{Z}}_{nt}$ such that

$$\tilde{\mathbb{Z}}_{nt}^{(o)} = \left[\frac{1}{T}\tilde{Y}_{n,t-1}, W_n\tilde{Y}_{n,t-1} - \tilde{Y}_{n,t-1}, \tilde{X}_{nt}\right]$$

becomes stable. Correspondingly, $\mathcal{H}_{nT}(\lambda_2)$ can be rescaled as

$$\mathcal{H}_{nT}^{(o)}(\lambda_2) = \begin{pmatrix} \mathcal{H}_{1,nT}^{(o)}(\lambda_2) & \mathcal{H}_{2,nT}^{(o)}(\lambda_2) \\ \mathcal{H}_{2,nT}^{(o)\prime}(\lambda_2) & \mathcal{H}_{3,nT}(\lambda_2) \end{pmatrix}, \quad (18)$$

where

$$\mathcal{H}_{1,nT}^{(o)}(\lambda_2) = \frac{1}{(n-1)T}\sum_{t=1}^T \tilde{\mathbb{Z}}_{nt}^{(o)\prime}R_n'(\lambda_2)J_nR_n(\lambda_2)\tilde{\mathbb{Z}}_{nt}^{(o)},$$

$$\mathcal{H}_{2,nT}^{(o)}(\lambda_2) = \frac{1}{(n-1)T}\sum_{t=1}^T \tilde{\mathbb{Z}}_{nt}^{(o)\prime}R_n'(\lambda_2)J_nR_n(\lambda_2) \times (G_n\tilde{Z}_{nt}^s\delta_0 - \tilde{Y}_{n,t-1}^s),$$

$$\mathcal{H}_{3,nT}(\lambda_2) = \frac{1}{(n-1)T}\sum_{t=1}^T (G_n\tilde{Z}_{nt}^s\delta_0 - \tilde{Y}_{n,t-1}^s)'R_n'(\lambda_2)J_nR_n \times (\lambda_2)(G_n\tilde{Z}_{nt}^s\delta_0 - \tilde{Y}_{n,t-1}^s).$$

Denote $\mathcal{H}_{nT}^{(o)} = \mathcal{H}_{nT}^{(o)}(\lambda_{20})$ and $\mathcal{H}_{j,nT}^{(o)} = \mathcal{H}_{j,nT}^{(o)}(\lambda_{20})$ for $j = 1, 2$. Similar to Appendix A.3 in Lee and Yu (2010a), the information matrix of (16) after rescaling will be nonsingular if

$$\frac{1}{\sigma_0^2(n-1)}\mathrm{tr}((\mathbb{D}_n' + \mathbb{D}_n)^2)\left(\mathcal{H}_{3,nT} - \mathcal{H}_{2,nT}^{(o)\prime}(\mathcal{H}_{1,nT}^{(o)})^{-1}\mathcal{H}_{2,nT}^{(o)}\right) + \Pi_n \neq 0, \quad (19)$$

where

$$\Pi_n = \frac{1}{4(n-1)^2}\left[\mathrm{tr}((\mathbb{C}_n' + \mathbb{C}_n)^2)\mathrm{tr}((\mathbb{D}_n' + \mathbb{D}_n)^2) - \mathrm{tr}^2((\mathbb{C}_n' + \mathbb{C}_n)(\mathbb{D}_n' + \mathbb{D}_n))\right] \quad (20)$$

with $\mathbb{C}_n = J_n\ddot{G}_n - \frac{\mathrm{tr}(J_n\ddot{G}_n)}{n-1}J_n$ and $\mathbb{D}_n = J_nH_n - \frac{\mathrm{tr}(J_nH_n)}{n-1}J_n$. Denote

$$\sigma_n^2(\lambda_1, \lambda_2) = \frac{\sigma_0^2}{n-1}\mathrm{tr}[(R_n(\lambda_2)S_n(\lambda_1) \times S_n^{-1}R_n^{-1})'J_n(R_n(\lambda_2)S_n(\lambda_1)S_n^{-1}R_n^{-1})].$$

Assumption 9. Either (a) $\lim_{T\to\infty}\mathcal{H}_{nT}^{(o)}(\lambda_2)$ exists and is nonsingular for each λ_2 in Λ_2, and the limit of $(\frac{1}{n-1}\ln|\sigma_0^2R_n^{-1\prime}J_nR_n^{-1}| - \frac{1}{n-1}\ln|\sigma_n^2(\lambda_{10}, \lambda_2)R_n^{-1}(\lambda_2)'J_nR_n^{-1}(\lambda_2)|)$ is not zero for $\lambda_2 \neq \lambda_{20}$; or (b) the limit of

$$\left(\frac{1}{n-1}\ln\left|\sigma_0^2R_n^{-1\prime}S_n^{-1\prime}J_nS_n^{-1}R_n^{-1}\right| - \frac{1}{n-1}\ln\left|\sigma_n^2(\lambda_1,\lambda_2) \times R_n^{-1}(\lambda_2)'S_n^{-1}(\lambda_1)'J_nS_n^{-1}(\lambda_1)R_n^{-1}(\lambda_2)\right|\right)$$

is not zero for $(\lambda_1, \lambda_2) \neq (\lambda_{10}, \lambda_{20})$, and $\lim_{T\to\infty}\frac{1}{nT}\sum_{t=1}^T\tilde{X}_{nt}'R_n'J_nR_n\tilde{X}_{nt}$ exists and is nonsingular.

Assumption 9 is for the identification of parameters $(\lambda_{10}, \lambda_{20})$. When Assumption 9(a) holds, the information matrix after rescaling is nonsingular because $\lim_{T\to\infty}\frac{1}{n-1}\mathrm{tr}((\mathbb{D}_n' + \mathbb{D}_n)^2)(\mathcal{H}_{3,nT} - \mathcal{H}_{2,nT}^{(o)\prime}(\mathcal{H}_{1,nT}^{(o)})^{-1}\mathcal{H}_{2,nT}^{(o)})$ in (19) would be positive. When $\lim_{T\to\infty}\frac{1}{n-1}\mathrm{tr}((\mathbb{D}_n' + \mathbb{D}_n)^2)(\mathcal{H}_{3,nT} - \mathcal{H}_{2,nT}^{(o)\prime}(\mathcal{H}_{1,nT}^{(o)})^{-1}\mathcal{H}_{2,nT}^{(o)}) = 0$, $(\lambda_{10}, \lambda_{20})$ can still be identified under Assumption 9(b). For the latter case, the subsequent Assumption 10 will be supplementary for the nonsingularity of the rescaled information matrix.

3.2. Asymptotic properties of QML estimators

Denote $\Phi_{1,T} = \mathrm{diag}(T, l_{k+4}')$ where l_{k+4} is the $(k+4)$ vector of ones.

Theorem 1. *Under Assumptions 1–9, θ_0^* is identified and, for the QMLE $\hat{\theta}_{nT}^*$ that maximizes (16), $\Phi_{1,T}(\hat{\theta}_{nT}^* - \theta_0^*) \xrightarrow{p} 0$.*

Proof. See Appendix C.4. □

Thus, the estimates $\hat{\rho}_{nT}$ and $\hat{\lambda}_{nT}$ are consistent, and $\hat{\gamma}_{nT}^*$ is superconsistent. The asymptotic distribution of $\hat{\theta}_{nT}^*$ can be derived from the Taylor expansion of $\frac{\partial \ln L_{nT}(\hat{\theta}_{nT}^*)}{\partial\theta}$ around θ_0^*:

$$\Phi_{1,T}\sqrt{(n-1)T}(\hat{\theta}_{nT}^* - \theta_0^*) = \left(-\frac{1}{(n-1)T}\Phi_{1,T}^{-1}\frac{\partial^2 \ln L_{n,T}(\bar{\theta}_{nT}^*)}{\partial\theta\partial\theta'}\Phi_{1,T}^{-1}\right)^{-1} \times \frac{1}{\sqrt{(n-1)T}}\Phi_{1,T}^{-1}\frac{\partial \ln L_{n,T}(\theta_0^*)}{\partial\theta}, \quad (21)$$

where $\bar{\theta}_{nT}^*$ lies between $\hat{\theta}_{nT}^*$ and θ_0^*. Even though the score and the Hessian matrix have unstable components due to $\tilde{Y}_{n,t-1}^*$, they will be stable after the rescaling by $\Phi_{1,T}^{-1}$. We can rewrite $\frac{1}{\sqrt{(n-1)T}}\Phi_{1,T}^{-1}\frac{\partial \ln L_{n,T}(\theta_0^*)}{\partial\theta}$ as $\frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L_{nT}^{(o)}(\theta_0^*)}{\partial\theta^*} - \Delta_{\theta_0^*,nT}$ where

$$\frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L_{nT}^{(o)}(\theta_0^*)}{\partial\theta^*} = \begin{pmatrix} \frac{1}{\sigma_0^2\sqrt{(n-1)T}}\sum_{t=1}^T(\tilde{\mathbb{Z}}_{nt}^{(o)\prime}R_n'J_nV_{nt}) \\ \frac{1}{\sigma_0^2\sqrt{(n-1)T}}\sum_{t=1}^T\left((G_n\tilde{Z}_{nt}^s\delta_0 - \tilde{Y}_{n,t-1}^s)'R_n'J_nV_{nt}\right) + \frac{1}{\sigma_0^2\sqrt{(n-1)T}}\sum_{t=1}^T(V_{nt}'\ddot{G}_nJ_nV_{nt} - \sigma_0^2\mathrm{tr}(J_n\ddot{G}_n)) \\ \frac{1}{\sigma_0^2\sqrt{(n-1)T}}\sum_{t=1}^T(V_{nt}'H_nJ_nV_{nt} - \sigma_0^2\mathrm{tr}(J_nH_n)) \\ \frac{1}{2\sigma_0^4\sqrt{(n-1)T}}\sum_{t=1}^T(V_{nt}'J_nV_{nt} - (n-1)\sigma_0^2) \end{pmatrix},$$

and

$$\Delta_{\theta_0^*,nT} = \sqrt{\frac{T}{n-1}}\begin{pmatrix} \frac{1}{\sigma_0^2}\bar{\mathbb{Z}}_{nT}^{(o)\prime}R_n'J_n\bar{V}_{nT} \\ \frac{1}{\sigma_0^2}(G_n\bar{Z}_{nT}^s\delta_0 - \bar{Y}_{nT,-1}^s)'R_n'J_n\bar{V}_{nT} \\ +\frac{1}{\sigma_0^2}\bar{V}_{nT}'\ddot{G}_nJ_n\bar{V}_{nT} \\ \frac{1}{\sigma_0^2}\bar{V}_{nT}'H_nJ_n\bar{V}_{nT} \\ \frac{1}{2\sigma_0^4}\bar{V}_{nT}'J_n\bar{V}_{nT} \end{pmatrix}$$

$$= \sqrt{\frac{n-1}{T}}a_{\theta_0,n}^* + O_p\left(\max\left(\sqrt{\frac{n-1}{T^3}}, \frac{1}{\sqrt{T}}\right)\right), \qquad (22)$$

with $\bar{\mathbb{Z}}_{nT}^{(o)} = [\frac{1}{T}\bar{Y}_{n,T-1}^d, W_n\bar{Y}_{n,T-1} - \bar{Y}_{n,T-1}, \bar{X}_{nT}]$ and $\bar{Y}_{n,T-1}^d$ being the variable $\bar{Y}_{n,T-1}$ derived by subtracting $W_n^uY_{n,-1}$, i.e., $\bar{Y}_{n,T-1}^d$ is the time average of $\frac{1}{(1-\lambda_{10})}W_n^u\sum_{h=0}^t\left(\mathbf{c}_{n0} + X_{n,t-h}\beta_0 + U_{n,t-h}\right) + Y_{nt}^s + Y_{nt}^\alpha$ as seen from (3). Here, $\frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L_{nT}^{(o)}(\theta_0^*)}{\partial\theta^*}$ has zero mean and is normally distributed using the CLT in Yu et al. (2008) when both n and T are large, and its variance is $\Sigma_{\theta_0^*,nT}^{(o)} + \Omega_{\theta_0^*,nT} + O\left(\frac{1}{T}\right)$ where $\Sigma_{\theta_0^*,nT}^{(o)} = \Phi_{1,T}^{-1}\Sigma_{\theta_0^*,nT}\Phi_{1,T}^{-1}$ and $\Omega_{\theta_0^*,nT}$ is in Eq. (23) given in Box II. Also, $a_{\theta,n}^*$ is in Eq. (24) given in Box II with (see Appendix C.5 for the derivation)

$$a_{\lambda,n}^* = \frac{1}{n-1}\mathrm{tr}\Bigg((\gamma G_n(\lambda_1) + \rho G_n(\lambda_1)W_n - I_n) \times \left(\sum_{h=0}^\infty B_n^h(\theta)\right)S_n^{-1}(\lambda_1)\Bigg) + \frac{1}{n-1}\mathrm{tr}(J_n\ddot{G}_n(\lambda_1))$$
$$= \frac{1}{n-1}\sum_{j=m_n+1}^n\Bigg[\left(\frac{\gamma\omega_{nj}}{(1-\lambda_1\omega_{nj})} + \frac{\rho\omega_{nj}^2}{(1-\lambda_1\omega_{nj})} - 1\right) \times \frac{1}{(1-d_{nj}(\theta))}\frac{\omega_{nj}-1}{(1-\lambda_1\omega_{nj})} + \frac{\omega_{nj}}{(1-\lambda_1\omega_{nj})}\Bigg] - \left(1-\frac{m_n}{n-1}\right)\frac{1}{(1-\lambda_1)}.$$

where $d_{nj}(\theta) = \frac{\gamma+\rho\omega_{nj}}{1-\lambda_1\omega_{nj}}$.

Under Assumption 8, the deterministic trend is present in the model; hence, from (38)–(41) in Yu et al. (2008), we have

$$-\frac{1}{(n-1)T}\Phi_{1,T}^{-1}\frac{\partial^2 \ln L_{n,T}(\bar{\theta}_{nT}^*)}{\partial\theta\partial\theta'}\Phi_{1,T}^{-1} = \Sigma_{\theta_0^*,nT}^{(o)} + O_p\left(\max\left(\frac{1}{\sqrt{(n-1)T}}, \frac{1}{T}\right)\right).$$

Thus, the distribution of $\hat{\theta}_{nT}^*$ can be obtained from (21). Furthermore, by using the relation of θ and θ^*, we can obtain the distribution of the QMLE $\hat{\theta}_{nT}$ of the original parameters vector θ_0. Denote

$$P_1 = \begin{pmatrix} 1 & -1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & I_k & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & I_2 \end{pmatrix} \quad \text{and}$$

$$\Psi_{1,T} = P_1\Phi_{1,T}^{-1} = \begin{pmatrix} \frac{1}{T} & -1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & I_k & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & I_2 \end{pmatrix},$$

so that $\theta = P_1\theta^*$.

Assumption 10. $\lim_{n\to\infty}\Pi_n$ is strictly positive, where Π_n is defined in (20).

Assumption 10 combined with Assumption 9(b) is to assure the nonsingularity of the information matrix. It is useful if Assumption 9(a) fails.

Theorem 2. *Under Assumptions 1–8 and 9 (a) (Assumption 9 (a) can be replaced with 9 (b) and 10),*

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

where $b_{\theta_0,nT}^* \equiv [\Sigma_{\theta_0^*,nT}^{(o)}]^{-1}\cdot a_{\theta_0,n}^*$, *and*

$$\sqrt{(n-1)T}(\hat{\theta}_{nT} - \theta_0) + \Psi_{1,T}\sqrt{\frac{n-1}{T}}b_{\theta_0,nT}^* + O_p\left(\max\left(\sqrt{\frac{n-1}{T^3}}, \sqrt{\frac{1}{T}}\right)\right)$$
$$\xrightarrow{d} N(0, \lim_{T\to\infty}\Psi_{1,T}(\Sigma_{\theta_0^*,nT}^{(o)})^{-1}(\Sigma_{\theta_0^*,nT}^{(o)} + \Omega_{\theta_0,nT}) \times (\Sigma_{\theta_0^*,nT}^{(o)})^{-1}\Psi_{1,T}').$$

Proof. See Appendix C.6. □

From Theorem 2, we have

$$(\hat{\theta}_{nT} - \theta_0) = -\frac{1}{T}\Psi_{1,T}b_{\theta_0,nT}^* + \frac{\mathcal{N}}{\sqrt{nT}} + O_p\left(\max\left(\frac{1}{T^2}, \frac{1}{\sqrt{nT^2}}\right)\right)$$

where $\mathcal{N}$ is a normal r.v. with zero mean and a finite variance matrix. Thus, the leading term of the right hand stochastic expansion has the bias $-\frac{1}{T}\Psi_{1,T}b_{\theta_0,nT}^*$ of order $O(1/T)$. For the distribution of $\hat{\theta}_{nT}$, when T is relatively larger than n, the QMLEs are asymptotically properly centered normal; when n is asymptotically proportional to T, the estimators are asymptotically normal, but the limit distribution is not properly centered. For those cases, as $\sqrt{nT}(\hat{\theta}_{nT} - \theta_0) = O_p(1)$, $\hat{\theta}_{nT}$ is $\sqrt{nT}$-consistent. However, when T is relatively smaller than n such that $T/n \to 0$, as $T(\hat{\theta}_{nT} - \theta_0) = -\Psi_{1,T}b_{\theta_0,nT}^* + \sqrt{T/n}\mathcal{N} + O_p\left(\max\left(1/T, 1/\sqrt{n}\right)\right)$, $\hat{\theta}_{nT}$ is T-consistent but $T(\hat{\theta}_{nT} - \theta_0)$ has a degenerate limiting distribution with all its mass at the limiting value of $-\Psi_{1,T}b_{\theta_0,nT}^*$. As $\Psi_{1,T}$ is singular in the limit, it implies that some linear combination of $\hat{\theta}_{nT}$ might be superconsistent. This is exactly what we see from $\hat{\theta}_{nT}^*$ when $n/T \to c < \infty$, because

$$\sqrt{(n-1)T^3}(\hat{\gamma}_{nT} + \hat{\rho}_{nT} + \hat{\lambda}_{1,nT} - 1) = \sqrt{\frac{n}{T}}[b_{\theta_0,nT}^*]_1 + O_p(1)$$

where $[b_{\theta_0,nT}^*]_1$ is the first element of $b_{\theta_0,nT}^*$.

$$\Omega_{\theta_0^*,nT} = \frac{(\mu_4 - 3\sigma_0^4)}{\sigma_0^4} \begin{pmatrix} \mathbf{0}_{(k+2)\times(k+2)} & * & * & * \\ \mathbf{0}_{1\times(k+2)} & \frac{1}{n-1}\sum_{i=1}^n [(J_n\ddot{G}_n)_{ii}]^2 & * & * \\ \mathbf{0}_{1\times(k+2)} & \frac{1}{n-1}\sum_{i=1}^n [(J_n\ddot{G}_n)_{ii}(J_nH_n)_{ii}] & \frac{1}{n-1}\sum_{i=1}^n [(J_nH_n)_{ii}]^2 & * \\ \mathbf{0}_{1\times(k+2)} & \frac{1}{2\sigma_0^2(n-1)}\mathrm{tr}(J_n\ddot{G}_n) & \frac{1}{2\sigma_0^2(n-1)}\mathrm{tr}(J_nH_n) & \frac{1}{4\sigma_0^4} \end{pmatrix}. \tag{23}$$

Also, $a^*_{\theta_0,n}$ is $a^*_{\theta,n}$ evaluated at θ_0

$$a^*_{\theta,n} = \Phi_{1,T}^{-1} \begin{pmatrix} \frac{T}{2(1-\lambda_1)}\frac{m_n-1}{n-1} + \frac{1}{n-1}\mathrm{tr}\left(\sum_{h=0}^{\infty} B_n^h(\theta)S_n^{-1}(\lambda_1)\right) - \frac{1}{(1-\lambda_1)} \\ \frac{1}{n-1}\mathrm{tr}\left((W_n - I_n)\left(\sum_{h=0}^{\infty} B_n^h(\theta)\right)S_n^{-1}(\lambda_1)\right) \\ \mathbf{0}_{k\times 1} \\ a^*_{\lambda,n} \\ \frac{1}{n-1}\mathrm{tr}(J_nH_n(\lambda_2)) \\ \frac{1}{2\sigma^2} \end{pmatrix}$$

$$= \Phi_{1,T}^{-1} \begin{pmatrix} \frac{T}{2(1-\lambda_1)}\frac{m_n-1}{n-1} + \frac{1}{n-1}\sum_{j=m_n+1}^{n}\frac{1}{(1-d_{nj}(\theta))}\frac{1}{(1-\lambda_j\omega_{nj})} - \left(1 - \frac{m_n}{n-1}\right)\frac{1}{(1-\lambda_1)} \\ \frac{1}{n-1}\sum_{j=m_n+1}^{n}\frac{1}{(1-d_{nj}(\theta))}\frac{\omega_{nj}-1}{(1-\lambda_1\omega_{nj})} \\ \mathbf{0}_{k\times 1} \\ a^*_{\lambda,n} \\ \frac{1}{n-1}\mathrm{tr}(H_n(\lambda_2)) - \frac{1}{(1-\lambda_2)} \\ \frac{1}{2\sigma^2} \end{pmatrix} \tag{24}$$

Box II.

From Theorem 2, when $n/T \to c$ where $0 < c < \infty$, the QMLEs $\hat{\theta}_{nT}$ and $\hat{\theta}^*_{nT}$ have biases, and their confidence intervals are not centered. Furthermore, when T is small relative to n in the sense that $n/T \to \infty$, the presence of biases causes $\hat{\theta}_{nT}$ and $\Phi_{1,T}\hat{\theta}^*_{nT}$ to have the slower T-rate of convergence. An analytical bias reduction procedure is

$$\hat{\theta}^{*1}_{nT} = \hat{\theta}^*_{nT} - \frac{\hat{\varphi}^*_{nT}}{T} \quad \text{and} \quad \hat{\theta}^1_{nT} = \hat{\theta}_{nT} - \frac{\hat{\varphi}_{nT}}{T},$$

where

$$\hat{\varphi}^*_{nT} = \left[\left(\frac{1}{(n-1)T}\frac{\partial^2 \ln L_{n,T}(\theta^*)}{\partial\theta^*\partial\theta^{*\prime}}\right)^{-1}\Phi_{1,T}a^*_{\theta,n}\right]\Bigg|_{\theta^*=\hat{\theta}^*_{nT}} \quad \text{and}$$

$$\hat{\varphi}_{nT} = \left[\left(\frac{1}{(n-1)T}\frac{\partial^2 \ln L_{n,T}(\theta)}{\partial\theta\partial\theta'}\right)^{-1}a_n(\theta)\right]\Bigg|_{\theta=\hat{\theta}_{nT}}$$

with $a_n(\theta) = P_1'^{-1}\Phi_{1,T}a^*_{\theta,n}$. We show that, when $T/(n-1)^{1/3} \to \infty$, $\Phi_{1,T}\hat{\theta}^{*1}_{nT}$ and $\hat{\theta}^1_{nT}$ are $\sqrt{(n-1)T}$-consistent and asymptotically centered normal even when $n/T \to \infty$. For the bias corrected estimators, we need the following additional assumption. Denote

$$B_n(\theta) = \Gamma_n\tilde{D}_n(\theta)\Gamma_n^{-1} \quad \text{where } \tilde{D}_n(\theta) = \mathrm{diag}\left\{0, \ldots, 0, \frac{\gamma + \rho\omega_{n,m_n+1}}{1-\lambda_1\omega_{n,m_n+1}}, \ldots, \frac{\gamma+\rho\omega_{nn}}{1-\lambda_1\omega_{nn}}\right\}$$

so that $B_n = B_n(\theta_0)$.

Assumption 11. $\sum_{h=0}^{\infty} B_n^h(\theta)$ and $\sum_{h=1}^{\infty} hB_n^{h-1}(\theta)$ are uniformly bounded in either row sum or column sums, uniformly in a neighborhood of θ_0.

This assumption can be justified by more basic conditions as in Lemma 14 in Yu et al. (2008). Our result for the bias corrected estimators is as follows.

Theorem 3. *If $T/(n-1)^{1/3} \to \infty$, under Assumptions 1–8 and 9 (a) (Assumption 9 (a) can be replaced with 9 (b) and 10) and 11,*

$$\Phi_{1,T}\sqrt{(n-1)T}(\hat{\theta}^{*1}_{nT} - \theta_0) \xrightarrow{d} N(0, \lim_{T\to\infty}(\Sigma^{(o)}_{\theta_0^*,nT})^{-1}(\Sigma^{(o)}_{\theta_0^*,nT} + \Omega_{\theta_0,nT})(\Sigma^{(o)}_{\theta_0^*,nT})^{-1}),$$

and $\sqrt{(n-1)T}(\hat{\theta}^1_{nT} - \theta_0) \xrightarrow{d} N(0, \lim_{T\to\infty}\Psi_{1,T}(\Sigma^{(o)}_{\theta_0^*,nT})^{-1}(\Sigma^{(o)}_{\theta_0^*,nT} + \Omega_{\theta_0,nT})(\Sigma^{(o)}_{\theta_0^*,nT})^{-1}\Psi'_{1,T})$.

Proof. Similar to Theorem 4 in Yu et al. (2008). □

Hence, if T grows faster than $(n-1)^{1/3}$, the analytical bias correction will give us estimators that are asymptotically normal and properly centered. For the case $n/T \to c$, it implies $T/(n-1)^{1/3} \to \infty$. For the case $n/T \to \infty$, as long as $T/n^{1/3} \to \infty$, $\hat{\theta}^1_{nT}$ is $\sqrt{(n-1)T}$-consistent, which is also an improvement upon the T-consistency of $\hat{\theta}_{nT}$. Thus, $\hat{\theta}^1_{nT}$ might perform better in economic applications, especially when n is much larger than T.

4. GMM estimation

Instead of the QML, we may use GMM for estimation. For the SDPD models, the GMM uses lagged values and exogenous variables to construct linear moments. Contrary to the QML estimation where consistency requires T tending to infinity, the GMM is applicable even when T is small, because GMM estimation can be applied to an equation after the elimination of the fixed effects so that the GMME does not have the $O(1/T)$ bias as in MLE.[18]

Assumption 8′. n is large and T is finite.

4.1. The moment conditions

For (13) where the time effects are eliminated, the individual effects $\mathbf{c}_{n0}^*$ remain but they can be further eliminated by a data transformation over time. Let $[F_{T,T-1}, \frac{1}{\sqrt{T}}l_T]$ be the orthonormal matrix of the eigenvectors of $J_T = (I_T - \frac{1}{T}l_T l_T')$, where l_T is the T-dimensional vector of ones and $F_{T,T-1}$ is the $T\times(T-1)$ eigenvectors matrix corresponding to the eigenvalues of one. The $(n-1)\times T$ matrix $[Y_{n1}^*, Y_{n2}^*, \ldots, Y_{nT}^*]$ can be transformed into the $(n-1)\times(T-1)$ matrix $[Y_{n1}^{**}, Y_{n2}^{**}, \ldots, Y_{n,T-1}^{**}] = [Y_{n1}^*, Y_{n2}^*, \ldots, Y_{nT}^*]F_{T,T-1}$, and so is $[V_{n1}^{**}, \ldots, V_{n,T-1}^{**}] = [V_{n1}^*, \ldots, V_{nT}^*]F_{T,T-1}$. Similarly, such a transformation can be applied to each column of the exogenous regressors and we obtain X_{nt}^{**}. Also, $[Y_{n0}^{(**,-1)}, Y_{n1}^{(**,-1)}, \ldots, Y_{n,T-2}^{(**,-1)}] = [Y_{n0}^*, Y_{n1}^*, \ldots, Y_{n,T-1}^*]F_{T,T-1}$. We note that $Y_{n,t-1}^{(**,-1)}$ and $Y_{n,t-1}^{**}$ are not equal. As $l_T'F_{T,T-1} = 0$, it follows $[\mathbf{c}_{n0}^*, \ldots, \mathbf{c}_{n0}^*]F_{T,T-1} = 0$ so that individual effects are eliminated by the transformation. Thus,[19] for $t = 1, \ldots, T-1$,

$$\begin{aligned} Y_{nt}^{**} &= \lambda_{10}[W_n^* Y_{nt}^{**} - Y_{n,t-1}^{(**,-1)}] + \gamma_0^* Y_{n,t-1}^{(**,-1)} + \rho_0[W_n^* Y_{n,t-1}^{(**,-1)} \\ &\quad - Y_{n,t-1}^{(**,-1)}] + X_{nt}^{**}\beta_0 + R_n^{*-1}V_{nt}^{**} \\ &= \lambda_{10}[W_n^* Y_{nt}^{**} - Y_{n,t-1}^{(**,-1)}] + \mathbb{Z}_{nt}^{**}\delta_0^* + R_n^{*-1}V_{nt}^{**} \\ &= \mathbb{K}_{nt}^{**}\kappa_0^* + R_n^{*-1}V_{nt}^{**}, \end{aligned} \tag{25}$$

where

$$\mathbb{Z}_{nt}^{**} = [Y_{n,t-1}^{(**,-1)}, W_n^* Y_{n,t-1}^{(**,-1)} - Y_{n,t-1}^{(**,-1)}, X_{nt}^{**}],$$

$$\mathbb{K}_{nt}^{**} = [\mathbb{Z}_{nt}^{**}, W_n^* Y_{nt}^{**} - Y_{n,t-1}^{(**,-1)}],$$

$\kappa_0^* = (\delta_0^{*\prime}, \lambda_{10})'$ and $\delta_0^* = (\gamma_0^*, \rho_0, \beta_0')'$. Because $(V_{n-1,1}^{**\prime}, \ldots, V_{n-1,T-1}^{**\prime})' = (F_{T,T-1}' \otimes F_{n,n-1}')(V_{n1}', \ldots, V_{nT}')'$, we have

$$E(V_{n-1,1}^{**\prime}, \ldots, V_{n-1,T-1}^{**\prime})'(V_{n-1,1}^{**\prime}, \ldots, V_{n-1,T-1}^{**}) = \sigma_0^2 I_{(n-1)(T-1)}.$$

Hence, elements of $V_{n-1,t}^{**}$ are uncorrelated for all i and t. However, we note that $Y_{n,t-1}^{(**,-1)}$ is correlated with V_{nt}^{**}. For this reason, in order to estimate (25), where individual effects vanish, IVs are needed for $Y_{n,t-1}^{(**,-1)}$ and $W_n^* Y_{n,t-1}^{(**,-1)}$ for each t (and also for $W_n^* Y_{nt}^{**}$). For that purpose, a convenient and revealing $F_{T,T-1}$ is the Helmert transformation such that

$$V_{nt}^{**} = \left(\frac{T-t}{T-t+1}\right)^{\frac{1}{2}}\left[V_{nt}^* - \frac{1}{T-t}\sum_{h=t+1}^{T} V_{nh}^*\right] \quad \text{and}$$

$$Y_{n,t-1}^{(**,-1)} = \left(\frac{T-t}{T-t+1}\right)^{\frac{1}{2}}\left[Y_{n,t-1}^* - \frac{1}{T-t}\sum_{h=t}^{T-1} Y_{nh}^*\right].$$

The V_{nt}^{**} depends on current and future values, but not on past ones; thus, in addition to all strictly exogenous variables X_{ns}^{**} for $s = 1, \ldots, T-1$, the time lag variables $Y_{n0}^*, \ldots, Y_{n,t-1}^*$ can also be used to construct IVs for $Y_{n,t-1}^{(**,-1)}$ as in the literature of dynamic panel data models (Alvarez and Arellano, 2003, etc.).

For the linear moments, we stack up the data and construct moment conditions via the transformed Eq. (25). An IV matrix can take the form $\mathbf{Q}_{n,T-1}^* = (Q_{n,1}^{*\prime}, \ldots, Q_{n,T-1}^{*\prime})'$ where $Q_{n,t}^*$, with a fixed column dimension q greater than or equal to $k+3$, is designed from predetermined and strictly exogenous variables. For example, for $t = 1, \ldots, T-1$, we may use

$$Q_{nt}^* = [Y_{n,t-1}^*, W_n^* Y_{n,t-1}^*, W_n^{*2} Y_{n,t-1}^*, X_{nt}^{**}, W_n^* X_{nt}^{**}] \tag{26}$$

or $\ddot{Q}_{nt}^* = R_n^* Q_{nt}^*$ to form the IV matrix. Denote $\mathbf{V}_{n,T-1}^{**}(\kappa^*, \lambda_2) = (V_{n1}^{**\prime}(\kappa^*, \lambda_2), \ldots, V_{n,T-1}^{**\prime}(\kappa^*, \lambda_2))'$ where $V_{nt}^{**\prime}(\kappa^*, \lambda_2) = R_n^*(\lambda_2)(Y_{nt}^{**} - \mathbb{K}_{nt}^{**}\kappa^*)$. The linear moments are $\ddot{\mathbf{Q}}_{n,T-1}^{*\prime}\mathbf{V}_{n,T-1}^{**}(\kappa^*, \lambda_2)$, which can be used for a generalized two stage least squares estimator (G2SLS). For notational purposes, we define the following long vectors or large matrices of variables. Denote $\mathbf{Y}_{n,T-1}^{**} = [Y_{n1}^{**\prime}, \ldots, Y_{n,T-1}^{**\prime}]'$ as dependent regressors, $\mathbf{K}_{n,T-1}^{**} = (\mathbb{K}_{n1}^{**\prime}, \ldots, \mathbb{K}_{n,T-1}^{**\prime})$ as explanatory variables in the RHS of (25), $\mathbf{Y}_{n,T-1}^{(**,-1)} = [Y_{n0}^{(**,-1)\prime}, \ldots, Y_{n,T-2}^{(**,-1)\prime}]'$ as lagged values, $\mathbf{G}_{n,T-1}^* = I_{T-1} \otimes G_n^*$ and $\mathbf{R}_{n,T-1}^* = I_{T-1} \otimes R_n^*$. Also, $\ddot{\mathbf{Z}}_{n,T-1}^{**} = \mathbf{R}_{n,T-1}^*[\mathbb{Z}_{n1}^{**\prime}, \ldots, \mathbb{Z}_{n,T-1}^{**\prime}]'$, $\ddot{\mathbf{Y}}_{n,T-1}^{**} = \mathbf{R}_{n,T-1}^*\mathbf{Y}_{n,T-1}^{**}$, $\ddot{\mathbf{K}}_{n,T-1}^{**} = \mathbf{R}_{n,T-1}^*\mathbf{K}_{n,T-1}^{**}$, $\ddot{\mathbf{Y}}_{n,T-1}^{(**,-1)} = \mathbf{R}_{n,T-1}^*\mathbf{Y}_{n,T-1}^{(**,-1)}$ and $\ddot{\mathbf{G}}_{n,T-1}^* = I_{T-1} \otimes R_n^* G_n^* R_n^{*-1}$. The (infeasible) G2SLSE of κ_0^* is

$$\begin{aligned} \hat{\kappa}_{\text{g2sls}}^* &= (\ddot{\mathbf{K}}_{n,T-1}^{**\prime}\ddot{\mathbf{Q}}_{n,T-1}^*(\ddot{\mathbf{Q}}_{n,T-1}^{*\prime}\ddot{\mathbf{Q}}_{n,T-1}^*)^{-1}\ddot{\mathbf{Q}}_{n,T-1}^{*\prime}\ddot{\mathbf{K}}_{n,T-1}^{**})^{-1} \\ &\quad \times (\ddot{\mathbf{K}}_{n,T-1}^{**\prime}\ddot{\mathbf{Q}}_{n,T-1}^*(\ddot{\mathbf{Q}}_{n,T-1}^{*\prime}\ddot{\mathbf{Q}}_{n,T-1}^*)^{-1}\ddot{\mathbf{Q}}_{n,T-1}^{*\prime}\ddot{\mathbf{Y}}_{n,T-1}^{**}). \end{aligned}$$

As elements of $\ddot{\mathbf{K}}_{n,T-1}^{**}$ and $\ddot{\mathbf{Q}}_{n,T-1}^*$ could have large magnitude when T is large, they need to be rescaled. By denoting $\Phi_{k+3} = \mathrm{diag}(T, l_{k+2})$, identification of κ_0 requires that the following two matrices have full rank:

$$\mathrm{plim}_{T\to\infty}\frac{1}{(n-1)(T-1)}\Upsilon_q^{-1}\ddot{\mathbf{Q}}_{n,T-1}^{*\prime}\ddot{\mathbf{K}}_{n-1,T-1}^{**}\Phi_{k+3}^{-1} \quad \text{and}$$

$$\mathrm{plim}_{T\to\infty}\frac{1}{(n-1)(T-1)}\Upsilon_q^{-1}\ddot{\mathbf{Q}}_{n,T-1}^{*\prime}\ddot{\mathbf{Q}}_{n,T-1}^*\Upsilon_q^{-1},$$

where Υ_q is a diagonal matrix of which its diagonal elements specify the orders for the rescaling of $\ddot{\mathbf{Q}}_{n,T-1}^*$. With the identification condition of κ_0^*, we can obtain the asymptotic behavior of G2SLSE for κ_0^* from

$$\begin{aligned} \Phi_{k+3}\hat{\kappa}_{\text{g2sls}}^* &= \left(\Phi_{k+3}^{-1}\ddot{\mathbf{K}}_{n,T-1}^{**\prime}\ddot{\mathbf{Q}}_{n,T-1}^*\left(\ddot{\mathbf{Q}}_{n,T-1}^{*\prime}\ddot{\mathbf{Q}}_{n,T-1}^*\right)^{-1}\right. \\ &\quad \left.\times\ \ddot{\mathbf{Q}}_{n,T-1}^{*\prime}\ddot{\mathbf{K}}_{n,T-1}^{**}\Phi_{k+3}^{-1}\right)^{-1}\left(\Phi_{k+3}^{-1}\ddot{\mathbf{K}}_{n,T-1}^{**\prime}\ddot{\mathbf{Q}}_{n,T-1}^*\right. \\ &\quad \left.\times\ \left(\ddot{\mathbf{Q}}_{n,T-1}^{*\prime}\ddot{\mathbf{Q}}_{n,T-1}^*\right)^{-1}\ddot{\mathbf{Q}}_{n,T-1}^{*\prime}\ddot{\mathbf{Y}}_{n,T-1}^{**}\right). \end{aligned}$$

We see that while estimates for other regression parameters are consistent, the estimate for $\gamma_0^* \equiv \gamma_0 + \rho_0 + \lambda_{10}$ is superconsistent. For the feasible G2SLS estimation, an initial consistent estimate of λ_{20} is needed for R_n. For this propose, as an example, the method of moments (MOM) in Kelejian and Prucha (1998) can be used.

For a more sophisticated GMM estimation, additional quadratic moments can also be incorporated (Lee, 2007), which can increase the efficiency of estimates. Let $\mathbf{P}_{n,T-1,j}^* = I_{T-1} \otimes F_{n,n-1}'P_{nj}F_{n,n-1}$ for some nonstochastic $n \times n$ UB matrix P_{nj} with the property $\mathrm{tr}(P_{nj}J_n) = 0$, which is equivalent to $\mathrm{tr}(P_{nj}^*) = 0$ where $P_{nj}^* =$

[18] A likelihood function could not be derived for the resulting equation after the elimination of fixed effects.

[19] Because $Y_{n,t-1}^{(**,-1)}$ is not $Y_{n,t-1}^{**}$, the transformed Eq. (25) does not form an SDPD process by itself. For this reason, an ML or QML approach for (25) is not feasible.

$F'_{n,n-1}P_{nj}F_{n,n-1}$. The rescaled moment conditions are

$$g_{nT}^{(o)}(\kappa^{*\prime},\lambda_2)=\begin{pmatrix}\Upsilon_q^{-1}\ddot{\mathbf{Q}}^{*\prime}_{n,T-1}\mathbf{V}^{**}_{n,T-1}(\kappa^{*\prime},\lambda_2)\\ \mathbf{V}^{**\prime}_{n,T-1}(\kappa^{*\prime},\lambda_2)\mathbf{P}^{*}_{n,T-1,1}\mathbf{V}^{**}_{n,T-1}(\kappa^{*\prime},\lambda_2)\\ \vdots\\ \mathbf{V}^{**\prime}_{n,T-1}(\kappa^{*\prime},\lambda_2)\mathbf{P}^{*}_{n,T-1,m}\mathbf{V}^{**}_{n,T-1}(\kappa^{*\prime},\lambda_2)\end{pmatrix}, \quad (27)$$

where $\mathbf{V}^{**}_{n,T-1}(\kappa^{*\prime},\lambda_2)=\mathbf{R}^{*}_{n,T-1}(\lambda_2)\mathbf{U}^{**}_{n,T-1}(\kappa^{*})$ and $\mathbf{U}^{**}_{n,T-1}(\kappa^{*})$ is the long vector stacked with $Y^{**}_{nt}-\mathbb{K}^{**}_{nt}\kappa^{*}$ for $t=1,\ldots,T-1$. Denote $\ddot{\mathbf{Q}}^{*}_{n,T-1}(\lambda_2)=\mathbf{R}^{*}_{n,T-1}(\lambda_2)\mathbf{Q}^{*}_{n,T-1}$.

Assumption 9′. The $(n-1)(T-1)\times q$ IV matrix $\mathbf{Q}^{*}_{n,T-1}$ is constructed as in (26).[20] For each possible λ_2, $\text{plim}_{n\to\infty}\frac{1}{(n-1)(T-1)}\Upsilon_q^{-1}\ddot{\mathbf{Q}}^{*}_{n,T-1}(\lambda_2)'[\ddot{\mathbf{Z}}^{**}_{n,T-1},\ddot{\mathbf{G}}^{*}_{n,T-1}\ddot{\mathbf{Z}}^{**}_{n,T-1}\delta_0-\ddot{\mathbf{Y}}^{(**,-1)}_{n,T-1}]\Phi_{k+3}^{-1}$ is of full rank $k+3$ and $\text{plim}_{n\to\infty}\frac{1}{(n-1)(T-1)}\Upsilon_q^{-1}\ddot{\mathbf{Q}}^{*}_{n,T-1}(\lambda_2)'\ddot{\mathbf{Q}}^{*}_{n,T-1}(\lambda_2)\Upsilon_q^{-1}$ is of full rank q. Also, $\text{tr}(R_n^{*\prime}(\lambda_2)^{-1}R_n^{*\prime}P^{*}_{nj}R^{*}_nR^{*}_n(\lambda_2)^{-1})=0$ for all $j=1,\ldots,m$ have a unique solution at λ_{20}.

The first part in Assumption 9′ is the identification condition for κ_0^{*} and the second part is the identification condition for λ_{20}.

4.2. Consistency and asymptotic distribution of GMME

Denote D_{nT} and R_{nT} as in Box III where b_{lz} and $b_{l\lambda}$ for $l=1,\ldots,m$ are defined in (63) and (64) in Appendix C.7. We see that $D^{(o)}_{nT}=\Upsilon_{q+2}^{-1}D_{nT}\Phi_{k+4}^{-1}$ is scaled to have $O_p(1)$ elements where $\Upsilon_{q+2}=\begin{pmatrix}\Upsilon_q & \mathbf{0}_{q\times 2}\\ \mathbf{0}_{2\times q} & I_2\end{pmatrix}$and $\Phi_{k+4}=\text{diag}(T,l'_{k+3})$ as before. Also, $R^{(o)}_{nT}=\Upsilon_{q+2}^{-1}R_{nT}\Phi_{k+4}^{-1}$ is scaled to have $O(\frac{1}{T})$. Thus, we have

$$\frac{1}{(n-1)(T-1)}\frac{\partial g^{(o)}_{nT}(\hat{\theta}_{nT})}{\partial\theta'}=D^{(o)}_{nT}+R^{(o)}_{nT}+O_p\left(\frac{1}{\sqrt{nT}}\right),$$

where $D^{(o)}_{nT}$ is $O(1)$ and $R^{(o)}_{nT}$ is $O\left(\frac{1}{T}\right)$. Let $\mathcal{D}^{(o)}_{nT}=D^{(o)}_{nT}+R^{(o)}_{nT}$. When T is large, the component $R^{(o)}_{nT}$ will disappear and $\mathcal{D}^{(o)}_{nT}$ will be reduced to $D^{(o)}_{nT}$ asymptotically. Let

$$\Delta_{mn,T}=[\text{vec}(\mathbf{P}^{*\prime}_{n,T-1,1}),\ldots,\text{vec}(\mathbf{P}^{*\prime}_{n,T-1,m})]'[\text{vec}(\mathbf{P}^{*\prime}_{n,T-1,1}+\mathbf{P}^{*}_{n,T-1,1}),\ldots,\text{vec}(\mathbf{P}^{*\prime}_{n,T-1,m}+\mathbf{P}^{*}_{n,T-1,m})], \quad (29)$$

and $\omega_{nm,T}=[\text{vec}_D(\mathbf{P}^{*}_{n,T-1,1}),\ldots,\text{vec}_D(\mathbf{P}^{*}_{n,T-1,m})]$, where $\text{vec}_D(\mathbf{P}^{*}_{n,T-1,j})$ denotes a column vector formed by the diagonal elements of $\mathbf{P}^{*}_{n,T-1,j}$. The variance matrix of quadratic and linear moments in (27) can be approximated by Eq. (30) is given in Box IV. When v_{it} is normally distributed, the second component of Σ_{nT} will be zero because $\mu_4-3\sigma_0^4=0$.

By analyzing $g^{(o)\prime}_{nT}(\kappa^{*\prime},\lambda_2)\Sigma^{(o)-1}_{nT}g^{(o)}_{nT}(\kappa^{*\prime},\lambda_2)$, we can obtain the asymptotic properties of GMM estimates for the reparameterized coefficients. We denote Π^{*} as the parameter space of $\pi^{*}=(\kappa^{*\prime},\lambda_2)'$, which is assumed to be compact. Denote

$$P_2=\begin{pmatrix}1&-1&0&-1&0\\0&1&0&0&0\\0&0&I_k&0&0\\0&0&0&1&0\\0&0&0&0&1\end{pmatrix}\quad\text{and}$$

$$\Psi_{2,T}=P_2\Phi_{k+4}^{-1}=\begin{pmatrix}\frac{1}{T}&-1&0&-1&0\\0&1&0&0&0\\0&0&I_k&0&0\\0&0&0&1&0\\0&0&0&0&1\end{pmatrix},$$

so that $(\kappa',\lambda_2)'=P_2(\kappa^{*\prime},\lambda_2)'$, where $\kappa=(\gamma,\rho,\beta',\lambda_1)'$, provides the original parameter vector. Correspondingly, $\pi=(\kappa',\lambda_2)'$ has its compact parameter space Π.

Theorem 4. *Under Assumptions 1–7, 8 (or 8′) and 9′, suppose we use the moment conditions in (27), the optimum GMME* $\hat{\pi}^{*}_{o,nT}\equiv(\hat{\kappa}^{*\prime}_{o,nT},\hat{\lambda}_{2,o,nT})'$ *derived from* $\min_{\pi^{*}\in\Pi^{*}}g^{(o)\prime}_{nT}(\pi^{*})\Sigma^{(o)-1}_{nT}g^{(o)}_{nT}(\pi^{*})$ *has*

$$\sqrt{(n-1)(T-1)}\Phi_{k+4}(\hat{\pi}^{*}_{o,nT}-\pi_0^{*})\xrightarrow{d}N(0,\text{plim}_{n\to\infty}(\mathcal{D}^{(o)\prime}_{nT}\Sigma^{(o)-1}_{nT}\mathcal{D}^{(o)}_{nT})^{-1}). \quad (31)$$

Suppose that $\hat{\Sigma}^{(o)-1}_{nT}-\Sigma^{(o)-1}_{nT}=o_p(1)$. *The feasible optimum GMME derived from* $\min_{\pi^{*}\in\Pi^{*}}g^{(o)\prime}_{nT}(\pi^{*})\hat{\Sigma}^{(o)-1}_{nT}g^{(o)}_{nT}(\pi^{*})$ *has the same asymptotic distribution in (31).*

Also, from $\pi=P_2\pi^{*}$, *we have*

$$\sqrt{(n-1)(T-1)}(\hat{\pi}_{o,nT}-\pi_0)\xrightarrow{d}N(0,\text{plim}_{n\to\infty}\Psi_{2,T}(\mathcal{D}^{(o)\prime}_{nT}\Sigma^{(o)-1}_{nT}\mathcal{D}^{(o)}_{nT})^{-1}\Psi'_{2,T}). \quad (32)$$

Proof. See Appendix C.7. □

Thus, the GMM estimates are consistent and asymptotically normal. From $\hat{\kappa}_{o,nT}$ in (32), the singularity of $\Psi_{2,T}$ in the limit reveals that some linear combination of estimates of the original parameters is super-consistent, i.e., the estimate for γ_0^{*} is $\sqrt{nT^3}$-consistent. Compared with MLE, the GMME does not have an asymptotic bias. However, when T is large so that the bias of MLE is not an issue, the asymptotic variance matrix of MLE is smaller than that of GMME from (27) (see Appendix C.8 for details).

5. Monte Carlo

We conduct a small Monte Carlo experiment to evaluate the performance of our QML estimators, their bias corrected estimators, 2SLSE, G2SLSE and GMME for the spatial cointegration model. Samples are generated from (1) using $\theta_0=(0.4,0.2,1,0.4,0.2,1)'$ where $\theta_0=(\gamma_0,\rho_0,\beta_0',\lambda_{10},\lambda_{20},\sigma_0^2)'$, and X_{nt}, $\mathbf{c}_{n0}$, α_{t0} and V_{nt} are generated from independent normal distributions.[21] The spatial weights matrix we use is a block diagonal matrix formed by a row-normalized rook matrix.[22] We use $T=5$, 10, 20, 50 and $n=96$. With 1000 repetitions for each case, finite sample properties of those estimators are summarized in Tables 1 and 2, where Table 1 is for $T=5$ and 10 and Table 2 is for $T=20$ and $T=50$. For each case, we report the bias (Bias), empirical standard deviation (E-SD), root mean square error (RMSE), coverage probability (CP) and theoretical standard deviation (T-SD).

For the QML estimators, we see that they have some biases, but the bias corrected estimators reduce those biases on average. This is consistent with our asymptotic analysis because the bias

20 Some additional spatial expansion terms can also be included as IVs.

21 We generated the spatial panel 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. We have also experimented with different true values for the DGP. The results are similar in their general features and are not reported here to save space. These results are available upon request.

22 We use the rook matrix based on an $r=4$ board (so that $n=r^2$). The rook matrix represents a square tessellation with a connectivity of four for the inner fields on the chessboard and two and three for the corner and border fields, respectively. Most regional structures in spatial econometrics are made up of regions with connectivity close to the range of the rook tessellation. 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 increase, but the percentage remains a constant. In our simulation for $n=96$, we have $m_n=6$. Also, for the rook matrix in the simulation, the second largest eigenvalue is 0.796.

$$D_{nT} = -\frac{1}{(n-1)(T-1)} \times \begin{pmatrix} \ddot{\mathbf{Q}}^{*\prime}_{n,T-1}\ddot{\mathbf{Z}}^{**}_{n,T-1} & \ddot{\mathbf{Q}}^{*\prime}_{n,T-1}(\ddot{\mathbf{G}}^{*}_{n,T-1}\ddot{\mathbf{Z}}^{**}_{n,T-1}\delta_0 - \ddot{\mathbf{Y}}^{(**,-1)}_{n,T-1}) & \mathbf{0}_{q\times 1} \\ \mathbf{0}_{1\times(k+2)} & \sigma_0^2 \mathrm{tr}(\ddot{\mathbf{G}}^{*\prime}_{n,T-1}(\mathbf{P}^{*\prime}_{n,T-1,1}+\mathbf{P}^{*}_{n,T-1,1})) & \sigma_0^2 \mathrm{tr}(\mathbf{H}^{*\prime}_{n,T-1}(\mathbf{P}^{*\prime}_{n,T-1,1}+\mathbf{P}^{*}_{n,T-1,1})) \\ \vdots & \vdots & \vdots \\ \mathbf{0}_{1\times(k+2)} & \sigma_0^2 \mathrm{tr}(\ddot{\mathbf{G}}^{*\prime}_{n,T-1}(\mathbf{P}^{*\prime}_{n,T-1,m}+\mathbf{P}^{*}_{n,T-1,m})) & \sigma_0^2 \mathrm{tr}(\mathbf{H}^{*\prime}_{n,T-1}(\mathbf{P}^{*\prime}_{n,T-1,m}+\mathbf{P}^{*}_{n,T-1,m})) \end{pmatrix} \quad \text{and}$$

$$R_{nT} = \frac{1}{T-1}\begin{pmatrix} \mathbf{0}_{q\times k_z} & \mathbf{0}_{q\times 1} & \mathbf{0}_{q\times 1} \\ b_{1z} & b_{1\lambda} & 0 \\ \vdots & \vdots & \vdots \\ b_{mz} & b_{m\lambda} & 0 \end{pmatrix} \qquad (28)$$

Box III.

$$\Sigma^{(o)}_{nT} = \sigma_0^4 \begin{pmatrix} \frac{1}{\sigma_0^2}\frac{1}{(n-1)(T-1)}\Upsilon_q^{-1}\ddot{\mathbf{Q}}^{*\prime}_{n,T-1}\ddot{\mathbf{Q}}^{*}_{n,T-1}\Upsilon_q^{-1} & \mathbf{0}_{q\times m} \\ \mathbf{0}_{m\times q} & \frac{1}{(n-1)(T-1)}\Delta_{nm,T} \end{pmatrix} + \frac{1}{(n-1)(T-1)}\begin{pmatrix} \mathbf{0}_{q\times q} & * \\ \mathbf{0}_{m\times q} & (\mu_4 - 3\sigma_0^4)\omega'_{nm,T}\omega_{nm,T} \end{pmatrix} \qquad (30)$$

Box IV.

corrected estimators will eliminate the bias of order $O(T^{-1})$. Also, the bias reduction is achieved while there is no significant increase in the variance of the estimators. Before the bias correction, the CPs of the estimators under 95% confidence level have lower values due to the bias, especially when T is relatively small. After the bias correction, the CPs are close to the specified 95% confidence level. T-SDs are similar to E-SDs, which means that the negative inverse of the Hessian matrix provides proper estimates for the variances of estimators. For the sum $\hat{\gamma}_{nT} + \hat{\rho}_{nT} + \hat{\lambda}_{1,nT}$, when T is small, the superconsistency might not appear; however, when T is large, the sum is much closer to 1, and the E-SD is much smaller than that of the individual estimates. For different cases of T, we see that when T is larger, the biases of estimators will be smaller and the variances will be smaller.

For the 2SLS, the linear IV is $Q^*_{nt} = [Y^*_{n,t-1}, W^*_n Y^*_{n,t-1}, \ldots, W^{*5}_n Y^*_{n,t-1}, X^{**}_{nt}, W^*_n X^{**}_{nt}]$. After we obtain the 2SLSE, we utilize the residuals and construct quadratic moments[23] to obtain estimates of λ_{20} and σ_0^2. For the G2SLS, the linear IV is $\ddot{Q}^*_{nt} = \hat{R}^*_n Q^*_{nt}$, where $\hat{R}^*_n$ is estimated from above. After we obtain the G2SLS, the residuals are updated to compute the estimate of σ_0^2, which is reported in the tables. For GMM, we use the same linear moments as the G2SLS, and we use two quadratic moments with $P^*_{n1} = W^*_n - \frac{\mathrm{tr}(W^*_n)}{n-1} I_{n-1}$ and $P^*_{n2} = W^{*2}_n - \frac{\mathrm{tr}(W^{*2}_n)}{n-1} I_{n-1}$. For the GMM search, the starting value for the parameters is obtained from the above G2SLS. From Tables 1–2, we see that the biases of 2SLS, G2SLS and GMM estimates are small even when T is small. Their biases are smaller than those of the bias corrected MLE for $T = 5$ and 10. However, their E-SDs are larger than those of MLEs and so are the RMSEs. When T is larger, the biases of the 2SLS, G2SLS and GMM estimates are of the similar magnitude as the MLE, but with larger E-SDs. This implies that MLE is more appropriate for the estimation of long panels relative to those IV and GMM estimates.

6. Conclusion

This paper investigates unstable SDPD models where there are unit roots generated by temporal and spatial correlations in the DGP. The spatial cointegration model refers to the case where some (but not all) eigenvalues of the DGP are equal to 1; the pure unit root model refers to the case where all the eigenvalues are equal to 1. We develop the consistency, superconsistency, and asymptotics of the QML and GMM estimators by reparameterization and data transformation, and also propose bias correction for QML estimates. For the spatial cointegration case, compared to the stable case in Yu et al. (2008), the individual coefficient estimates can be $\sqrt{nT}$-consistent as usual, but the sum of the spatial and dynamic effects estimates is superconsistent.

The SDPD model analyzed in this paper does not incorporate cross section dependence due to unobserved individual specific macroeconomic variables or shocks. For future research, it may be of interest to model common and persistent shocks in a random component or factor structural framework jointly with the spatial setting. This extension is of interest as those features can be important in many macroeconomic applications.

Appendix A. Notations

The following list summarizes some frequently used notations in the text:

$S_n(\lambda_1) = I_n - \lambda_1 W_n$ for an arbitrary value λ_1 and $S_n = I_n - \lambda_{10} W_n$.
$R_n(\lambda_2) = I_n - \lambda_2 M_n$ for an arbitrary value λ_2 and $R_n = I_n - \lambda_{20} M_n$.
$G_n = W_n S_n^{-1}$ and $H_n = M_n R_n^{-1}$.
$W^*_n = F'_{n,n-1} W_n F_{n,n-1}$ and $M^*_n = F'_{n,n-1} M_n F_{n,n-1}$.
$G^*_n = W^*_n (I_n - \lambda_{10} W^*_n)^{-1} = F'_{n,n-1} G_n F_{n,n-1}$.
$A_n = S_n^{-1}(\gamma_0 I_n + \rho_0 W_n)$ and $A_n = \Gamma_n D_n \Gamma_n^{-1}$ where Γ_n is the eigenvector matrix of W_n and D_n is the diagonal eigenvalue matrix of A_n.
$W^u_n = \Gamma_n \mathbb{J}_n \Gamma_n^{-1}$ where $\mathbb{J}_n = \mathrm{Diag}\{l'_{m_n}, 0, \ldots, 0\}$ and l_{m_n} is an $m_n \times 1$ vector of ones.
$F_{n,n-1}$ is the $n \times (n-1)$ eigenvectors matrix of $J_n = I_n - \frac{1}{n} l_n l'_n$ corresponding to the eigenvalues of one.
$\tilde{\Upsilon}_{nt} = \Upsilon_{nt} - \bar{\Upsilon}_{nT}$, $\Upsilon^*_{nt} = F'_{n,n-1}\Upsilon_{nt}$ and $\tilde{\Upsilon}^*_{nt} = F'_{n,n-1}\tilde{\Upsilon}_{nt}$ where $\bar{\Upsilon}_{nT} = \frac{1}{T}\sum_{t=1}^{T}\Upsilon_{nt}$.
$\mathbb{Z}^*_{nt} = [Y^*_{n,t-1}, W^*_n Y^*_{n,t-1} - Y^*_{n,t-1}, X^*_{nt}]$ and $\tilde{\mathbb{Z}}^{(o)}_{nt} = [\frac{1}{T}\tilde{Y}_{n,t-1}, W_n \tilde{Y}_{n,t-1} - \tilde{Y}_{n,t-1}, \tilde{X}_{nt}]$.

[23] We use two quadratic moments $P^*_{n1} = W^*_n - \frac{\mathrm{tr}(W^*_n)}{n-1} I_{n-1}$ and $P^*_{n2} = W^{*2}_n - \frac{\mathrm{tr}(W^{*2}_n)}{n-1} I_{n-1}$ with an optimum GMM estimation for this purpose. This slightly extends the MOM estimation in Kelejian and Prucha (1999).

Table 1
Comparison of estimates under $T = 5$, 10 and $n = 96$.

	T	n		γ	ρ	β	λ_1	λ_2	σ^2	*sum*
(1) MLE	5	96	Bias	−0.1549	0.0215	−0.0471	−0.0251	0.0018	−0.2492	−0.1584
			E-SD	0.0388	0.0625	0.0494	0.0887	0.1152	0.0582	0.0896
			RMSE	0.1597	0.0661	0.0683	0.0922	0.1152	0.2559	0.1820
			CP	0.0140	0.9200	0.8000	0.8990	0.8910	0.0160	0.3230
			T-SD	0.0341	0.0575	0.0457	0.0742	0.0976	0.0497	0.0647
(2) MLE1	5	96	Bias	−0.0106	0.0394	−0.0006	0.0327	−0.0599	−0.1037	0.0615
			E-SD	0.0467	0.0885	0.0517	0.1027	0.1281	0.0701	0.1368
			RMSE	0.0479	0.0969	0.0517	0.1078	0.1414	0.1251	0.1500
			CP	0.8250	0.7650	0.9170	0.8150	0.8240	0.4500	0.6900
			T-SD	0.0378	0.0688	0.0500	0.0688	0.0961	0.0598	0.0560
(3) 2SLS	5	96	Bias	−0.0077	0.0020	−0.0046	0.0061	−0.0033	−0.0075	0.0004
			E-SD	0.0803	0.1160	0.0543	0.0942	0.1158	0.0945	0.1371
			RMSE	0.0807	0.1161	0.0544	0.0944	0.1158	0.0948	0.1371
			CP	0.9390	0.9490	0.9540	0.9500	–	–	0.9590
			T-SD	0.0766	0.1131	0.0566	0.0963	–	–	0.1306
(4) G2SLS	5	96	Bias	−0.0067	0.0015	−0.0029	0.0026	−0.0033	−0.0067	−0.0025
			E-SD	0.0802	0.1156	0.0540	0.0940	0.1158	0.0945	0.1354
			RMSE	0.0805	0.1156	0.0541	0.0940	0.1158	0.0948	0.1354
			CP	0.9390	0.9410	0.9600	0.9410	–	–	0.9590
			T-SD	0.0762	0.1115	0.0564	0.0935	–	–	0.1277
(5) GMM	5	96	Bias	−0.0084	0.0070	−0.0040	−0.0055	0.0141	−0.0082	−0.0069
			E-SD	0.0799	0.1199	0.0542	0.0963	0.1232	0.0919	0.1464
			RMSE	0.0803	0.1201	0.0544	0.0965	0.1240	0.0922	0.1466
			CP	0.9380	0.9300	0.9550	0.9210	0.9170	–	0.9430
			T-SD	0.0759	0.1112	0.0563	0.0907	0.1167	–	0.1281
(1′) MLE	10	96	Bias	−0.0730	0.0176	−0.0132	0.0001	−0.0094	−0.1208	−0.0552
			E-SD	0.0252	0.0434	0.0341	0.0529	0.0721	0.0450	0.0402
			RMSE	0.0772	0.0468	0.0366	0.0529	0.0727	0.1289	0.0683
			CP	0.1250	0.9190	0.9120	0.9290	0.9230	0.2020	0.5440
			T-SD	0.0228	0.0419	0.0322	0.0480	0.0666	0.0410	0.0279
(2′) MLE1	10	96	Bias	−0.0017	0.0144	0.0008	0.0137	−0.0242	−0.0342	0.0264
			E-SD	0.0271	0.0497	0.0342	0.0558	0.0734	0.0494	0.0469
			RMSE	0.0272	0.0517	0.0342	0.0574	0.0773	0.0601	0.0538
			CP	0.9080	0.8940	0.9350	0.8980	0.9010	0.8000	0.7270
			T-SD	0.0241	0.0465	0.0338	0.0478	0.0670	0.0451	0.0263
(3′) 2SLS	10	96	Bias	−0.0000	−0.0003	−0.0010	0.0008	0.0001	−0.0069	0.0005
			E-SD	0.0409	0.0629	0.0346	0.0582	0.0760	0.0524	0.0538
			RMSE	0.0409	0.0629	0.0346	0.0582	0.0760	0.0528	0.0539
			CP	0.9540	0.9660	0.9550	0.9550	–	–	0.9410
			T-SD	0.0422	0.0645	0.0349	0.0577	–	–	0.0507
(4′) G2SLS	10	96	Bias	−0.0000	0.0002	−0.0004	−0.0002	0.0001	−0.0067	−0.0000
			E-SD	0.0407	0.0632	0.0345	0.0588	0.0760	0.0525	0.0532
			RMSE	0.0407	0.0632	0.0345	0.0588	0.0760	0.0529	0.0532
			CP	0.9510	0.9630	0.9530	0.9460	–	–	0.9420
			T-SD	0.0418	0.0643	0.0348	0.0576	–	–	0.0499
(5′) GMM	10	96	Bias	−0.0002	0.0030	−0.0006	−0.0031	0.0065	−0.0067	−0.0003
			E-SD	0.0408	0.0640	0.0347	0.0603	0.0779	0.0525	0.0571
			RMSE	0.0408	0.0641	0.0347	0.0604	0.0782	0.0529	0.0571
			CP	0.9530	0.9530	0.9530	0.9330	0.9290	–	0.9320
			T-SD	0.0418	0.0639	0.0348	0.0564	0.0748	–	0.0501

Note: $\theta_0 = (0.4, 0.2, 1, 0.4, 0.2, 1)'$ and the sum is $\rho + \gamma + \lambda_1$.

$F_{T,T-1}$ is the $T \times (T-1)$ matrix of Helmert transformation.
For QMLE, $\theta = (\delta', \lambda_1, \lambda_2, \sigma^2)'$ where $\delta = (\gamma, \rho, \beta')'$. After reparameterization, $\gamma^* = \gamma + \rho + \lambda_1$, $\delta^* = (\gamma^*, \rho, \beta')'$ and $\theta^* = (\delta^{*\prime}, \lambda_1, \lambda_2, \sigma^2)'$.
For GMM, $\pi = (\kappa', \lambda_2)'$ where $\kappa = (\gamma, \rho, \beta', \lambda_1)'$. After reparameterization, $\gamma^* = \gamma + \rho + \lambda_1$, $\kappa^{*\prime} = (\gamma^*, \rho, \beta', \lambda_1)'$ and $\pi^* = (\kappa^{*\prime}, \lambda_2)'$.

Appendix B. Some lemmas

Denote $\mathbb{U}_{nt} = \sum_{h=1}^{\infty} P_{nt,h} V_{n,t+1-h}$ and $\mathbb{W}_{nt} = \sum_{h=1}^{\infty} Q_{nt,h} V_{n,t+1-h}$, where $P_{nt,h}$ and $Q_{nt,h}$ are $n \times n$ nonstochastic matrices in the form of $P_{nt,h} = B_{1n} P_{nt}^h$ and $Q_{nt,h} = B_{2n} Q_{nt}^h$ for some nonstochastic UB matrices P_{nt}, Q_{nt}, B_{n1} and B_{n2}, where P_{nt}^h and Q_{nt}^h are the P_{nt} and Q_{nt} to the power of h. Furthermore, $\sum_{h=1}^{\infty} \mathrm{abs}(P_{nt}^h)$ and $\sum_{h=1}^{\infty} \mathrm{abs}(Q_{nt}^h)$ are UB uniformly in t, where $[\mathrm{abs}(P_{nt})]_{ij} = \left|P_{nt,ij}\right|$. Also, in the following Lemmas 1 and 2, let $\mathbf{C}_{nT} = (C_{n1}', \ldots, C_{nT}')'$ be a nonstochastic $nT \times 1$ vector with its elements satisfying $\sup_{n,T} \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} c_{nt,i}^2 < \infty$. Also, let $\mathcal{B}_n$ be an $n \times n$ nonstochastic UB matrix and denote $\xi_{nt} = \sum_{h=0}^{t-1} V_{nh}$ where V_{nh} is the vector of disturbances in (1). We note that $\bar{\xi}_{nT} = \frac{1}{T} \sum_{t=1}^{T} \xi_{nt} = \sum_{h=1}^{T} \frac{h}{T} V_{n,T-h}$. Also, as $\bar{\mathbb{U}}_{nT} = \left(\sum_{t=1}^{T} \mathbb{U}_{nt}\right)/T$, we have $\bar{\mathbb{U}}_{nT} = \sum_{h=1}^{\infty} \bar{P}_{nT,h} V_{n,T+1-h}$, where

$$\bar{P}_{nT,h} = \begin{cases} \dfrac{1}{T} \displaystyle\sum_{g=1}^{h} P_{n,T-h+g,g} & \text{for } h \le T \\ \dfrac{1}{T} \displaystyle\sum_{g=h-T+1}^{h} P_{n,T-h+g,g} & \text{for } h > T. \end{cases} \quad (33)$$

Lemma 1. *If $\mathbf{B}_{nT}$ is an $nT \times nT$ nonstochastic UB matrix and $\mathbf{F}_{nT}$ is a nonstochastic $nT \times 1$ vector with its elements satisfying $\sup_{n,T} \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} f_{nt,i}^2 < \infty$, then $\sup_{n,T} \left|\frac{1}{nT} \mathbf{C}_{nT}' \mathbf{B}_{nT} \mathbf{F}_{nT}\right| < \infty$.*

Table 2
Comparison of estimates under $T = 20, 50$ and $n = 96$.

	T	n		γ	ρ	β	λ_1	λ_2	σ^2	*sum*
(1) MLE	20	96	Bias	−0.0366	0.0099	−0.0045	0.0084	−0.0136	−0.0569	−0.0183
			E-SD	0.0163	0.0304	0.0242	0.0337	0.0478	0.0321	0.0160
			RMSE	0.0401	0.0320	0.0247	0.0347	0.0498	0.0653	0.0243
			CP	0.3650	0.9350	0.9380	0.9180	0.9260	0.5400	0.6960
			T-SD	0.0157	0.0304	0.0228	0.0325	0.0461	0.0310	0.0119
(2) MLE1	20	96	Bias	−0.0018	0.0006	−0.0010	0.0040	−0.0087	−0.0091	0.0028
			E-SD	0.0168	0.0324	0.0243	0.0340	0.0480	0.0337	0.0122
			RMSE	0.0169	0.0324	0.0243	0.0343	0.0488	0.0349	0.0125
			CP	0.9350	0.9330	0.9360	0.9260	0.9370	0.9060	0.9570
			T-SD	0.0161	0.0320	0.0233	0.0333	0.0467	0.0326	0.0119
(3) 2SLS	20	96	Bias	−0.0010	0.0002	−0.0015	0.0018	−0.0023	−0.0015	0.0009
			E-SD	0.0261	0.0402	0.0244	0.0380	0.0507	0.0344	0.0233
			RMSE	0.0261	0.0402	0.0245	0.0380	0.0508	0.0344	0.0233
			CP	0.9580	0.9550	0.9420	0.9520	–	–	0.9500
			T-SD	0.0262	0.0409	0.0236	0.0381	–	–	0.0208
(4) G2SLS	20	96	Bias	−0.0011	0.0006	−0.0012	0.0013	−0.0023	−0.0013	0.0007
			E-SD	0.0260	0.0409	0.0244	0.0385	0.0507	0.0345	0.0231
			RMSE	0.0260	0.0409	0.0244	0.0385	0.0508	0.0345	0.0231
			CP	0.9570	0.9450	0.9400	0.9470	–	–	0.9510
			T-SD	0.0260	0.0410	0.0235	0.0386	–	–	0.0206
(5) GMM	20	96	Bias	−0.0011	0.0012	−0.0012	0.0005	−0.0002	−0.0015	0.0006
			E-SD	0.0259	0.0409	0.0244	0.0387	0.0517	0.0343	0.0232
			RMSE	0.0259	0.0409	0.0244	0.0387	0.0517	0.0343	0.0232
			CP	0.9540	0.9440	0.9420	0.9470	0.9460	–	0.9470
			T-SD	0.0259	0.0407	0.0235	0.0379	0.0510	–	0.0206
(1′) MLE	50	96	Bias	−0.0142	0.0049	−0.0005	0.0059	−0.0064	−0.0226	−0.0034
			E-SD	0.0095	0.0196	0.0144	0.0206	−0.0299	0.0204	0.0045
			RMSE	0.0171	0.0202	0.0144	0.0214	0.0306	0.0304	0.0056
			CP	0.7050	0.9380	0.9540	0.9320	0.9330	0.7910	0.8530
			T-SD	0.0097	0.0196	0.0144	0.0201	0.0288	0.0203	0.0035
(2′) MLE1	50	96	Bias	−0.0004	−0.0001	0.0001	0.0006	−0.0005	−0.0023	0.0001
			E-SD	0.0096	0.0202	0.0144	0.0209	0.0301	0.0208	0.0037
			RMSE	0.0096	0.0202	0.0144	0.0209	0.0301	0.0210	0.0037
			CP	0.9480	0.9340	0.9530	0.9440	0.9260	0.9420	0.9410
			T-SD	0.0098	0.0200	0.0145	0.0205	0.0290	0.0207	0.0035
(3′) 2SLS	50	96	Bias	0.0001	−0.0004	−0.0000	0.0006	0.0005	−0.0010	0.0003
			E-SD	0.0151	0.0241	0.0144	0.0236	0.0323	0.0209	0.0061
			RMSE	0.0151	0.0241	0.0144	0.0236	0.0323	0.0210	0.0061
			CP	0.9480	0.9510	0.9560	0.9460	–	–	0.9550
			T-SD	0.0150	0.0242	0.0146	0.0233	–	–	0.0061
(4′) G2SLS	50	96	Bias	0.0000	−0.0001	0.0000	0.0004	0.0005	−0.0009	0.0003
			E-SD	0.0150	0.0248	0.0144	0.0244	0.0323	0.0210	0.0062
			RMSE	0.0150	0.0248	0.0144	0.0244	0.0323	0.0210	0.0062
			CP	0.9420	0.9440	0.9540	0.9410	–	–	0.9550
			T-SD	0.0149	0.0244	0.0146	0.0237	–	–	0.0061
(5′) GMM	50	96	Bias	0.0000	0.0002	0.0000	0.0000	0.0014	−0.0010	0.0003
			E-SD	0.0150	0.0248	0.0144	0.0243	0.0331	0.0209	0.0063
			RMSE	0.0150	0.0248	0.0144	0.0243	0.0331	0.0209	0.0063
			CP	0.9410	0.9390	0.9540	0.9380	0.9410	–	0.9550
			T-SD	0.0148	0.0241	0.0146	0.0233	0.0315	–	0.0061

Note: $\theta_0 = (0.4, 0.2, 1, 0.4, 0.2, 1)'$ and the sum is $\rho + \gamma + \lambda_1$.

Lemma 2. *Under Assumptions 1, 3 and 4 and large n and T, for ξ_{nt}, $\mathbf{c}_{n0}$, $\mathbb{X}_{nt}$, $\mathbb{U}_{nt}$, C_{nt} and their cross products,*

$$\frac{1}{nT}\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t} + \tilde{\mathbb{X}}_{nt}\beta_0)'\mathcal{B}_n(\mathbf{c}_{n0}\tilde{t} + \tilde{\mathbb{X}}_{nt}\beta_0) = O\left(T^2\right); \tag{34}$$

$$\frac{1}{nT}\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t} + \tilde{\mathbb{X}}_{nt}\beta_0)'\mathcal{B}_n\tilde{\xi}_{nt} = O_p\left(\sqrt{\frac{T^3}{n}}\right) \quad \textit{with zero mean;} \tag{35}$$

$$\frac{1}{nT}\sum_{t=1}^{T}\tilde{\xi}_{nt}'\mathcal{B}_n\tilde{\xi}_{nt} - E\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{\xi}_{nt}'\mathcal{B}_n\tilde{\xi}_{nt}\right) = O_p\left(\frac{T}{\sqrt{n}}\right) \tag{36}$$

where $E(\frac{1}{nT}\sum_{t=1}^{T}\tilde{\xi}_{nt}'\mathcal{B}_n\tilde{\xi}_{nt}) = O(T)$;

$$\frac{1}{nT}\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t} + \tilde{\mathbb{X}}_{nt}\beta_0)'\mathcal{B}_n C_{nt} = O(T); \tag{37}$$

$$\frac{1}{nT}\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t} + \tilde{\mathbb{X}}_{nt}\beta_0)'\tilde{\mathbb{U}}_{nt} = O_p\left(\sqrt{\frac{T}{n}}\right) \quad \textit{with mean zero;} \tag{38}$$

$$\frac{1}{nT}\sum_{t=1}^{T}\tilde{\xi}_{nt}'\mathcal{B}_n C_{nt} = O_p\left(\sqrt{\frac{T}{n}}\right) \quad \textit{with mean zero;} \tag{39}$$

$$\frac{1}{nT}\sum_{t=1}^{T}\tilde{\xi}_{nt}'\tilde{\mathbb{U}}_{nt} - E\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{\xi}_{nt}'\tilde{\mathbb{U}}_{nt}\right) = O_p\left(\sqrt{\frac{T}{n}}\right), \tag{40}$$

where $E(\frac{1}{nT}\sum_{t=1}^{T}\tilde{\xi}_{nt}'\tilde{\mathbb{U}}_{nt}) = O(1)$ *and*

$$\frac{1}{nT}\sum_{t=1}^{T}\tilde{\mathbb{U}}_{nt}'\tilde{\mathbb{W}}_{nt} - E\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{\mathbb{U}}_{nt}'\tilde{\mathbb{W}}_{nt}\right) = O\left(\frac{1}{\sqrt{nT}}\right), \tag{41}$$

where $E(\frac{1}{nT}\sum_{t=1}^{T}\tilde{\mathbb{U}}_{nt}'\tilde{\mathbb{W}}_{nt}) = O(1)$.

Lemma 3. *Under Assumptions 1, 3 and 4 and large n and T, for* $\mathbf{c}_{n0}, \mathbb{X}_{nt}, \xi_{nt}, V_{nt}, \mathbb{U}_{nt}$ *and their cross products,*

$$\frac{1}{nT}\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)'\mathcal{B}_n V_{nt} = O_p\left(\sqrt{\frac{T}{n}}\right); \quad (42)$$

$$\frac{1}{nT}\sum_{t=1}^{T}\xi_{nt}'\mathcal{B}_n V_{nt} = O_p\left(\frac{1}{\sqrt{n}}\right); \quad (43)$$

$$\frac{1}{n}\bar{\xi}_{nT}'\mathcal{B}_n\bar{V}_{nT} - E\frac{1}{n}\bar{\xi}_{nT}'\mathcal{B}_n\bar{V}_{nT} = O_p\left(\frac{1}{\sqrt{n}}\right), \quad (44)$$

where $E\frac{1}{n}\bar{\xi}_{nT}'\mathcal{B}_n\bar{V}_{nT} = \sigma_0^2\frac{1}{n}\frac{(T-1)}{2T}\mathrm{tr}(\mathcal{B}_n) = O(1)$;

$$\frac{1}{n}\bar{\mathbb{U}}_{n,T-1}'\bar{V}_{nT} - E\frac{1}{n}\bar{\mathbb{U}}_{n,T-1}'\bar{V}_{nT} = O_p\left(\frac{1}{\sqrt{nT}}\right), \quad (45)$$

where $E\frac{1}{n}\bar{\mathbb{U}}_{n,T-1}'\bar{V}_{nT} = \frac{1}{nT^2}\sigma_0^2\mathrm{tr}\left(\sum_{t=1}^{T-1}\sum_{h=1}^{t}P_{nt,h}\right)+O\left(\frac{1}{T^2}\right)$; *and*

$$\frac{1}{nT}\sum_{t=1}^{T}\tilde{V}_{nt}'\mathcal{B}_n\tilde{V}_{nt} = \left(1-\frac{1}{T}\right)\sigma_0^2\frac{1}{n}\mathrm{tr}(\mathcal{B}_n)+O_p\left(\frac{1}{\sqrt{nT}}\right). \quad (46)$$

Proof for Lemma 1. By Cauchy's inequality, $|\frac{1}{nT}\mathbf{C}_{nT}'\mathbf{B}_{nT}\mathbf{F}_{nT}| \le \frac{1}{nT}|\mathbf{C}_{nT}'\mathbf{C}_{nT}|^{1/2}|\mathbf{F}_{nT}'\mathbf{B}_{nT}'\mathbf{B}_{nT}\mathbf{F}_{nT}|^{1/2}$. Let $\rho(\mathbf{B}_{nT}'\mathbf{B}_{nT})$ be the spectral radius of $\mathbf{B}_{nT}'\mathbf{B}_{nT}$ (the largest eigenvalue in absolute value). For any matrix norm $\|\cdot\|$, it is known that $\rho(\mathbf{B}_{nT}'\mathbf{B}_{nT}) \le \|\mathbf{B}_{nT}'\mathbf{B}_{nT}\|$ (see Horn and Johnson, 1985). Taking $\|\mathbf{B}_{nT}'\mathbf{B}_{nT}\|$ to be either $\|\mathbf{B}_{nT}'\mathbf{B}_{nT}\|_\infty$ or $\|\mathbf{B}_{nT}'\mathbf{B}_{nT}\|_1$, it follows that $\left\|\mathbf{B}_{nT}'\mathbf{B}_{nT}\right\|$ is bounded because $\mathbf{B}_{nT}$ is UB. Thus, as $\mathbf{B}_{nT}'\mathbf{B}_{nT}$ is non-negative definite, we have $\mathbf{B}_{nT}'\mathbf{B}_{nT} \le \rho(\mathbf{B}_{nT}'\mathbf{B}_{nT})I_n \le \|\mathbf{B}_{nT}'\mathbf{B}_{nT}\|I_n$. Hence, $\mathbf{F}_{nT}'\mathbf{B}_{nT}'\mathbf{B}_{nT}\mathbf{F}_{nT} \le \|\mathbf{B}_{nT}'\mathbf{B}_{nT}\|\mathbf{F}_{nT}'\mathbf{F}_{nT}$. It follows that

$$\left|\frac{1}{nT}\mathbf{C}_{nT}'\mathbf{B}_{nT}\mathbf{F}_{nT}\right| \le \|\mathbf{B}_{nT}'\mathbf{B}_{nT}\|\cdot\left|\frac{1}{nT}\mathbf{C}_{nT}'\mathbf{C}_{nT}\right|^{1/2}\cdot\left|\frac{1}{nT}\mathbf{F}_{nT}'\mathbf{F}_{nT}\right|^{1/2}.$$

With the empirical moments of $\mathbf{C}_{nT}$ and $\mathbf{F}_{nT}$, $\sup_{n,T}\left|\frac{1}{nT}\mathbf{C}_{nT}'\mathbf{B}_{nT}\mathbf{F}_{nT}\right| < \infty$. □

Proof for Lemma 2. Eq. (34): Let $\rho(\frac{\mathcal{B}_n+\mathcal{B}_n'}{2})$ be the spectral radius of $\frac{\mathcal{B}_n+\mathcal{B}_n'}{2}$. Hence, $\left|\frac{1}{nT}\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)'\mathcal{B}_n(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)\right| \le \rho\left(\frac{\mathcal{B}_n+\mathcal{B}_n'}{2}\right)\cdot\frac{1}{nT}\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)'(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0) \le \|\mathcal{B}_n\|\cdot\frac{1}{nT}\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)'(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)$. The $\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)'(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)$ can be written as $(\mathbf{C}_{nT}\tilde{t}+\mathbf{F}_{nT}T)'(\mathbf{C}_{nT}\tilde{t}+\mathbf{F}_{nT}T)$ where $\mathbf{C}_{nT} = l_T\otimes\mathbf{c}_{n0}$ and $\mathbf{F}_{nT} = (\frac{1}{T}L_TJ_T\otimes I_n)\mathbf{X}_{nT}^{(-1)}\beta_0$ with L_T being a lower diagonal matrix with unit elements and $\mathbf{X}_{nT}^{(-1)} = (X_{n0}', X_{n1}', \ldots, X_{n,T-1}')'$. By using Lemma 1, $\frac{1}{nT}(\mathbf{C}_{nT}+\mathbf{F}_{nT})'(\mathbf{C}_{nT}+\mathbf{F}_{nT}) = O(1)$ under Assumption 4. Thus, we have $|\frac{1}{nT}\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)'\mathcal{B}_n(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)| = O(T^2)$.
Eq. (35): as $E\xi_{nt}\xi_{ns}' = \sigma_0^2\min\{t,s\}I_n$ and $\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)'\mathcal{B}_n\tilde{\xi}_{nt} = \sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)'\mathcal{B}_n\xi_{nt}$, we have

$$\begin{aligned}
&\mathrm{Var}\left(\frac{1}{nT}\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)'\mathcal{B}_n\tilde{\xi}_{nt}\right)\\
&=\frac{1}{n^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)'\mathcal{B}_n(E\xi_{nt}\xi_{ns}')\mathcal{B}_n'(\mathbf{c}_{n0}\tilde{s}+\tilde{\mathbb{X}}_{ns}\beta_0)\\
&=\frac{\sigma_0^2}{n^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\min\{s,t\}(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)'\mathcal{B}_n\mathcal{B}_n'(\mathbf{c}_{n0}\tilde{s}+\tilde{\mathbb{X}}_{ns}\beta_0)\\
&\le\frac{\sigma_0^2T^3}{n}\frac{1}{n}\left(\frac{1}{T^3}\sum_{t=1}^{T}t(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)'\right)\mathcal{B}_n\mathcal{B}_n'\\
&\quad\times\left(\frac{1}{T^2}\sum_{s=1}^{T}\left(\mathbf{c}_{n0}\tilde{s}+\tilde{\mathbb{X}}_{ns}\beta_0\right)\right) = O\left(\frac{T^3}{n}\right),
\end{aligned}$$

where the last equality follows from the empirical moments of $\mathbf{c}_{n0}$ and X_{nt} similar to the proof for (34).
Eq. (36): we have $\frac{1}{nT}\sum_{t=1}^{T}\tilde{\xi}_{nt}'\mathcal{B}_n\tilde{\xi}_{nt} = \frac{1}{nT}\sum_{t=1}^{T}(\xi_{nt}'\mathcal{B}_n\xi_{nt}) - \frac{1}{n}\bar{\xi}_{nT}'\mathcal{B}_n\bar{\xi}_{nT}$.

For the first part, $E(\frac{1}{nT}\sum_{t=1}^{T}(\xi_{nt}'\mathcal{B}_n\xi_{nt})) = \sigma_0^2\mathrm{tr}(\mathcal{B}_n)(\frac{1}{nT}\sum_{t=1}^{T}t) = O(T)$ and $\mathrm{Var}(\frac{1}{nT}\sum_{t=1}^{T}\xi_{nt}'\mathcal{B}_n\xi_{nt}) = \frac{1}{n^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}(\xi_{nt}'\mathcal{B}_n\xi_{nt}, \xi_{ns}'\mathcal{B}_n\xi_{ns})$. Using Lemma 4 in Yu et al. (2008) for the variance of $\mathbb{U}_{nt}'\mathbb{W}_{nt}$ (in our case here, $\mathbb{U}_{nt} = \sum_{h=1}^{\infty}P_{nt,h}V_{n,t+1-h}$ and $\mathbb{W}_{nt} = \sum_{h=1}^{\infty}Q_{nt,h}V_{n,t+1-h}$, where $P_{nt,h} = I_n$ and $Q_{nt,h} = \mathcal{B}_n$ for $h \le t$, and $P_{nt,h} = Q_{nt,h} = 0$ for $h > t$), we have $\mathrm{Var}(\frac{1}{nT}\sum_{t=1}^{T}\xi_{nt}'\mathcal{B}_n\xi_{nt}) = O(\frac{T^2}{n})$.

For the second part, $E(\frac{1}{n}\bar{\xi}_{nT}'\mathcal{B}_n\bar{\xi}_{nT}) = E(\frac{1}{n}\mathbb{U}_{nT}^{+\prime}\mathbb{W}_{nT}^{+})$ where $\mathbb{U}_{nT}^{+} = \sum_{h=1}^{\infty}P_{nT,h}^{+}V_{n,T-h}$ and $\mathbb{W}_{nT}^{+} = \sum_{h=1}^{\infty}Q_{nT,h}^{+}V_{n,T-h}$ with

$$P_{nT,h}^{+} = \begin{cases} I_n\dfrac{h}{T} & \text{for } h \le T\\ 0 & \text{for } h > T\end{cases} \quad\text{and}\quad Q_{nT,h}^{+} = \begin{cases} \mathcal{B}_n\dfrac{h}{T} & \text{for } h \le T\\ 0 & \text{for } h > T.\end{cases} \quad (47)$$

Then, using Lemmas 2 and 4 in Yu et al. (2008), $E(\frac{1}{n}\bar{\xi}_{nT}'\mathcal{B}_n\bar{\xi}_{nT}) = O(T)$ and $\mathrm{Var}(\frac{1}{n}\bar{\xi}_{nT}'\mathcal{B}_n\bar{\xi}_{nT}) = \frac{1}{n^2}\mathrm{Var}(\mathbb{U}_{nT}^{+\prime}\mathbb{W}_{nT}^{+}) = O(\frac{T^2}{n})$ because $\sum_{h=1}^{\infty}P_{nT,h}^{+}P_{nT,h}^{+\prime} = \sum_{h=1}^{T}(\frac{h}{T})^2I_n$, $\sum_{h=1}^{\infty}Q_{nT,h}^{+\prime}P_{nT,h}^{+} = \sum_{h=1}^{T}(\frac{h}{T})^2\mathcal{B}_n'$, $\sum_{h=1}^{\infty}Q_{nT,h}^{+}Q_{nT,h}^{+\prime} = \sum_{h=1}^{T}(\frac{h}{T})^2\mathcal{B}_n\mathcal{B}_n'$ and $\sum_{h=1}^{T}h^2 = O(T^3)$.
Eq. (37): we have $\left|\frac{1}{nT}\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)'\mathcal{B}_n\tilde{C}_{nt}\right| \le \frac{1}{T}\sum_{t=1}^{T}\left|\frac{\mathbf{c}_{n0}'\mathcal{B}_n\tilde{C}_{nt}}{n}\right|\cdot|\tilde{t}|+\frac{1}{T}\sum_{t=1}^{T}\left|\frac{(\tilde{\mathbb{X}}_{nt}\beta_0)'\mathcal{B}_n\tilde{C}_{nt}}{n}\right|$. By Lemma 1, $\left|\frac{1}{nT}\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)'\mathcal{B}_n\tilde{C}_{nt}\right| = O(T)$ similar to the proof for (34).
Eq. (38): the $\frac{1}{nT}\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)'\tilde{\mathbb{U}}_{nt} = \frac{1}{nT}\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)'\mathbb{U}_{nt}$ has zero mean and its variance matrix is $\frac{1}{n^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\mathbf{c}_{n0}\tilde{s}+\tilde{\mathbb{X}}_{ns}\beta_0)'E(\mathbb{U}_{nt}\mathbb{U}_{ns}')(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)$ where $E(\mathbb{U}_{nt}\mathbb{U}_{ns}') = \sigma_0^2(\sum_{h=1}^{\infty}P_{nt,t-s+h}P_{ns,h}')$ for $t \ge s$ from Lemma 2 in Yu et al. (2008). This variance matrix can be written as $\frac{1}{n^2T^2}(\mathbf{C}_{nT}\tilde{t}+\mathbf{F}_{nT}T)'\mathbf{P}_{nT}(\mathbf{C}_{nT}\tilde{t}+\mathbf{F}_{nT}T)$ where

$$\mathbf{P}_{nT} = \begin{pmatrix} \sum_{h=1}^{\infty}P_{n1,h}P_{n1,h}' & \cdots & \sum_{h=1}^{\infty}P_{nT,T-1+h}P_{n1,h}'\\ \vdots & \ddots & \vdots\\ \sum_{h=1}^{\infty}P_{n1,T-1+h}P_{nT,h}' & \cdots & \sum_{h=1}^{\infty}P_{nT,h}P_{nT,h}'\end{pmatrix}$$

has bounded row and column sums uniformly in n and T due to the form of $P_{nt,h} = B_{1n}P_{nt}^{h}$, and $\mathbf{C}_{nT}$ and $\mathbf{F}_{nT}$ are defined in the proof for (34). Thus, the order of the elements of this variance matrix is $O(\frac{T}{n})$, which implies that $\frac{1}{nT}\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)'\tilde{\mathbb{U}}_{nt} = O_p(\sqrt{\frac{T}{n}})$.
Eq. (39): as $E\xi_{nt}\xi_{ns}' = \sigma_0^2\min\{t,s\}I_n$, we have

$$\mathrm{Var}\left(\frac{1}{nT}\sum_{t=1}^{T}\tilde{\xi}_{nt}'\mathcal{B}_n\tilde{C}_{nt}\right) = \frac{1}{n^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}(\tilde{C}_{nt}'\mathcal{B}_n'(E(\xi_{nt}\xi_{ns}'))\mathcal{B}_n\tilde{C}_{ns})$$

$$= \frac{\sigma_0^2}{n^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\min\{s,t\}(\tilde{C}_{nt}'\mathcal{B}_n'\mathcal{B}_n\tilde{C}_{ns}) = O\left(\frac{T}{n}\right).$$

Eq. (40): we have $\frac{1}{nT}\sum_{t=1}^{T}\tilde{\xi}_{nt}'\tilde{\mathbb{U}}_{nt} = \frac{1}{nT}\sum_{t=1}^{T}\tilde{\xi}_{nt}'\tilde{\mathbb{U}}_{nt} - \frac{1}{n}\bar{\xi}_{nT}'\bar{\mathbb{U}}_{nT}$.

For the first part, using Lemma 2 in Yu et al. (2008), $E(\frac{1}{nT}\sum_{t=1}^{T}\xi_{nt}'\mathbb{U}_{nt}) = \sigma_0^2\frac{1}{nT}\mathrm{tr}(\sum_{t=1}^{T}\sum_{h=1}^{t}P_{nt,h}) = O(1)$ because $\sum_{h=1}^{\infty}\mathrm{abs}(P_{nt,h})$ is UB. Also,

$$\mathrm{Var}\left(\frac{1}{nT}\sum_{t=1}^{T}\xi_{nt}'\mathbb{U}_{nt}\right)$$

$$= \frac{1}{n^2T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}(\xi_{nt}'\mathbb{U}_{nt},\xi_{ns}'\mathbb{U}_{ns}) = O\left(\frac{T}{n}\right).$$

This is so as follows. As $\xi_{nt} = \sum_{h=1}^{\infty}Q_{nt,h}V_{n,t-h}$ where $Q_{nt,h} = \begin{cases} I_n & \text{for } h=1,2,\ldots,t \\ 0 & \text{for } h\geq t+1\end{cases}$, we have $\mathrm{Var}(\frac{1}{nT}\sum_{t=1}^{T}\xi_{nt}'\mathbb{U}_{nt}) = O(\frac{T}{n})$ using Lemma 4 in Yu et al. (2008) because the leading factor $\sum_{h=1}^{\infty}Q_{ns,h}Q_{ns,h}' = \sum_{h=1}^{s}I_n = s\cdot I_n$ and $\sum_{h=1}^{T}\sum_{s=1}^{t}s = O(T^3)$.

For the second part, $E(\frac{1}{n}\bar{\xi}_{nT}'\bar{\mathbb{U}}_{nT}) = \frac{\sigma_0^2}{nT}\mathrm{tr}(\sum_{h=1}^{T}h\cdot\bar{P}_{nT,h})$ where $\bar{P}_{nT,h}$ is specified in (33). So, $E(\frac{1}{n}\bar{\xi}_{nT}'\bar{\mathbb{U}}_{nT}) = O(1)$. Also, $\mathrm{Var}(\frac{1}{n}\bar{\xi}_{nT}'\bar{\mathbb{U}}_{nT}) = \frac{1}{n^2}\mathrm{Var}(\mathbb{W}_{nT}^{+\prime}\bar{\mathbb{U}}_{nT})$ where $\bar{\mathbb{U}}_{nT} = \sum_{h=1}^{\infty}\bar{P}_{nT,h}V_{n,t+1-h}$ and $\mathbb{W}_{nT}^{+} = \sum_{h=1}^{\infty}P_{nT,h}^{+}V_{n,t+1-h}$ with $\bar{P}_{nT,h}$ specified in (33) and $P_{nT,h}^{+}$ specified in (47). Then, using Lemma 4 in Yu et al. (2008), we have $\mathrm{Var}(\frac{1}{n}\bar{\xi}_{nT}'\bar{\mathbb{U}}_{nT}) = O(\frac{T}{n})$.

Eq. (41): this is Lemma 7 in Yu et al. (2008). □

Proof for Lemma 3. Eq. (42): denote $\rho(\mathcal{B}_n\mathcal{B}_n')$ as the spectral radius of $\mathcal{B}_n\mathcal{B}_n'$. Then,

$$\mathrm{Var}\left(\frac{1}{nT}\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)'\mathcal{B}_nV_{nt}\right)$$

$$\leq \frac{\sigma_0^2}{n^2T^2}\cdot\rho(\mathcal{B}_n\mathcal{B}_n')\cdot\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)'(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)$$

$$\leq \frac{\sigma_0^2}{n^2T^2}\cdot\left\|\mathcal{B}_n\mathcal{B}_n'\right\|\cdot\sum_{t=1}^{T}(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)'(\mathbf{c}_{n0}\tilde{t}+\tilde{\mathbb{X}}_{nt}\beta_0)$$

$$= O\left(\frac{T}{n}\right).$$

Eq. (43): $\mathrm{Var}\left(\frac{1}{nT}\sum_{t=1}^{T}\xi_{nt}'\mathcal{B}_nV_{nt}\right) = \frac{\sigma_0^2}{n^2T^2}\sum_{t=1}^{T}E(\xi_{nt}'\mathcal{B}_n\mathcal{B}_n'\xi_{nt}) = \frac{\sigma_0^4}{n^2T^2}\cdot\mathrm{tr}(\mathcal{B}_n\mathcal{B}_n')\cdot\sum_{t=1}^{T}(t-1) = O\left(\frac{1}{n}\right)$. Then, $\frac{1}{nT}\sum_{t=1}^{T}\xi_{nt}'\mathcal{B}_nV_{nt} = O_p\left(\frac{1}{\sqrt{n}}\right)$ with zero mean.

Eq. (44): as $\bar{\xi}_{nT} = \frac{1}{T}\sum_{t=1}^{T}\xi_{nt} = \frac{1}{T}\sum_{t=1}^{T}\sum_{h=0}^{t-1}V_{nh} = \frac{1}{T}\sum_{t=1}^{T}(T-t+1)V_{n,t-1}$, we have

$$E(\bar{\xi}_{nT}'\mathcal{B}_n\bar{V}_{nT}) = \frac{1}{T^2}E\left[\left(\sum_{t=1}^{T}(T-t+1)V_{n,t-1}\right)'\mathcal{B}_n\sum_{t=1}^{T}V_{nt}\right]$$

$$= \sigma_0^2\frac{(T-1)}{2T}\mathrm{tr}(\mathcal{B}_n).$$

Also, $\mathrm{Var}(\bar{\xi}_{nT}'\mathcal{B}_n\bar{V}_{nT}) = \mathrm{Var}(\mathbb{U}_{nT}^{+\prime}\mathbb{W}_{nT}^{+})$, where $\mathbb{U}_{nT}^{+} = \sum_{h=1}^{\infty}P_{nT,h}^{+}V_{n,T-h}$ with $P_{nT,h}^{+}$ in (47) and $\mathbb{W}_{nT}^{+} = \sum_{h=1}^{\infty}Q_{nT,h}^{+}V_{n,T-h}$ with $Q_{nT,h}^{+} = \begin{cases}\mathcal{B}_n\cdot\frac{1}{T} & \text{for } h\leq T\\ 0 & \text{for } h>T\end{cases}$. From Lemma 4 in Yu et al. (2008), $\mathrm{Var}(\bar{\xi}_{nT}'\mathcal{B}_n\bar{V}_{nT}) = O(n)$. Hence, $E(\bar{\xi}_{nT}'\mathcal{B}_n\bar{V}_{nT}) = O(n)$ and $\mathrm{Var}(\bar{\xi}_{nT}'\mathcal{B}_n\bar{V}_{nT}) = O(n)$.

Eq. (45): this is implied by Lemma 11 in Yu et al. (2008). □

Appendix C. Some algebra and proof for Theorems 1, 2 and 4

C.1. Deriving (3)

From (2), by iterative substitution, we have

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

For the term associated with time effects, as $S_n^{-1}l_n = \frac{1}{1-\lambda_{10}}l_n$ and $A_n = S_n^{-1}(\gamma_0I_n+\rho_0W_n) = (\gamma_0I_n+\rho_0W_n)S_n^{-1}$, using $W_nl_n = l_n$, we have $A_n^hS_n^{-1}l_n = \frac{1}{1-\lambda_{10}}(\frac{\gamma_0+\rho_0}{1-\lambda_{10}})^hl_n$. By $A_n^h = (\frac{\gamma_0+\rho_0}{1-\lambda_{10}})^h\Gamma_n\mathbb{J}_n\Gamma_n^{-1}+B_n^h$ for $h=1,2,\ldots$, and $\Gamma_n\mathbb{J}_n\Gamma_n^{-1}S_n^{-1} = S_n^{-1}\Gamma_n\mathbb{J}_n\Gamma_n^{-1} = \frac{1}{1-\lambda_{10}}\Gamma_n\mathbb{J}_n\Gamma_n^{-1}$,[24] the above equation can be written as

$$Y_{nt} = A_n^{t+1}Y_{n,-1} + \sum_{h=0}^{t}B_n^hS_n^{-1}(\mathbf{c}_{n0}+X_{n,t-h}\beta_0+U_{n,t-h}) + \frac{1}{1-\lambda_0}\sum_{h=0}^{t}\left(\frac{\gamma_0+\rho_0}{1-\lambda_{10}}\right)^h\alpha_{t-h,0}l_n + \frac{1}{1-\lambda_{10}}\times\sum_{h=1}^{t}\left(\frac{\gamma_0+\rho_0}{1-\lambda_{10}}\right)^h\Gamma_n\mathbb{J}_n\Gamma_n^{-1}(\mathbf{c}_{n0}+X_{n,t-h}\beta_0+U_{n,t-h}).$$

For $A_n^{t+1}Y_{n,-1}$, we have $A_n^{t+1}Y_{n,-1} = (\frac{\gamma_0+\rho_0}{1-\lambda_{10}})^{t+1}\Gamma_n\mathbb{J}_n\Gamma_n^{-1}Y_{n,-1}+B_n^{t+1}Y_{n,-1}$, where

$$B_n^{t+1}Y_{n,-1} = \sum_{h=t+1}^{\infty}B_n^hS_n^{-1}(\mathbf{c}_{n0}+X_{n,t-h}\beta_0+U_{n,t-h}) + \frac{1}{1-\lambda_{10}}\sum_{h=t+1}^{\infty}\alpha_{t-h,0}B_n^hl_n,$$

using $B_nA_n = B_n^2$ and $B_nS_n^{-1} = S_n^{-1}B_n$. Because Γ_n is the eigenvectors matrix of W_n and its first column is l_n, we have $\Gamma_n^{-1}l_n = \sqrt{n}e_n$ where $e_n = (1,0,\ldots,0)'$. As $\tilde{D}_ne_n = 0$, it follows that $B_nl_n = 0$. Hence, we can decompose Y_{nt} as $Y_{nt} = Y_{nt}^u+Y_{nt}^s+Y_{nt}^\alpha$ in (3).

C.2. FOC and SOC for (16)

Using $\mathrm{tr}(G_n(\lambda_1)) - \mathrm{tr}(J_nG_n(\lambda_1)) = \frac{1}{1-\lambda_1}$ and $\mathrm{tr}(G_n^2(\lambda_1)) - \mathrm{tr}((J_nG_n(\lambda_1))^2) = \frac{1}{(1-\lambda_1)^2}$ (see Lee and Yu, 2010a), the first and second order derivatives for the log likelihood function (16) are given in Box V. As $\mathrm{tr}(J_nG_n) = \mathrm{tr}(J_n\ddot{G}_n) = \mathrm{tr}(G_n) - \frac{1}{1-\lambda_{10}}$ because W_n and M_n are row-normalized, the score vector is Eq. (50) is given in Box VI and the dominant part of the information matrix is in (17).

C.3. Concentrated likelihood on (λ_1,λ_2)

As we might have a higher rate of convergence for $\hat{\gamma}_{nT}^*$ when $T\to\infty$, it is more convenient to further concentrate out the log

24 As $S_n = \Gamma_n(I_n-\lambda_{10}\varpi_n)\Gamma_n^{-1}$ where $\varpi_n = \mathbb{J}_n+\tilde{D}_n$, we have $S_n^{-1}\Gamma_n\mathbb{J}_n\Gamma_n^{-1} = \Gamma_n(I_n-\lambda_{10}\varpi_n)^{-1}\mathbb{J}_n\Gamma_n^{-1} = \frac{1}{1-\lambda_{10}}\Gamma_n\mathbb{J}_n\Gamma_n^{-1}$ and also $\Gamma_n\mathbb{J}_n\Gamma_n^{-1}S_n^{-1} = \Gamma_n\mathbb{J}_n(I_n-\lambda_{10}\varpi_n)^{-1}\Gamma_n^{-1} = \frac{1}{1-\lambda_{10}}\Gamma_n\mathbb{J}_n\Gamma_n^{-1}$ because $(I_n-\lambda_{10}\varpi_n)^{-1}\mathbb{J}_n = \mathbb{J}_n(I_n-\lambda_{10}\varpi_n)^{-1} = \frac{1}{1-\lambda_{10}}\mathbb{J}_n$.

$$\frac{\partial \ln L_{nT}(\theta^*)}{\partial \theta^*} = \begin{pmatrix} \frac{\partial \ln L_{nT}(\theta^*)}{\partial \delta^*} \\ \frac{\partial \ln L_{nT}(\theta^*)}{\partial \lambda_1} \\ \frac{\partial \ln L_{nT}(\theta^*)}{\partial \lambda_2} \\ \frac{\partial \ln L_{nT}(\theta^*)}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\lambda_2)\tilde{\mathbb{Z}}_{nt})'J_n\tilde{V}_{nt}(\theta^*) \\ \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\lambda_2)(W_n\tilde{Y}_{nt}-\tilde{Y}_{n,t-1}))'J_n\tilde{V}_{nt}(\theta^*) - T\operatorname{tr}(J_nG_n(\lambda_1)) \\ \frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n(\lambda_2)\tilde{V}_{nt}(\theta^*))'J_n\tilde{V}_{nt}(\theta^*) - T\operatorname{tr}(J_nH_n(\lambda_2)) \\ \frac{1}{2\sigma^4}\sum_{t=1}^{T}(\tilde{V}'_{nt}(\theta^*)J_n\tilde{V}_{nt}(\theta^*) - (n-1)\sigma^2) \end{pmatrix}, \tag{48}$$

and

$$\frac{\partial^2 \ln L_{nT}(\theta^*)}{\partial \theta^* \partial \theta^{*\prime}} = \begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\lambda_2)\tilde{\mathbb{Z}}_{nt})'J_nR_n(\lambda_2)\tilde{\mathbb{Z}}_{nt} & * & * & * \\ \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\lambda_2)(W_n\tilde{Y}_{nt}-\tilde{Y}_{n,t-1}))'J_nR_n(\lambda_2)\tilde{\mathbb{Z}}_{nt} & 0 & 0 & 0 \\ \left[\frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n(\lambda_2)\tilde{V}_{nt}(\theta^*))'J_nR_n(\lambda_2)\tilde{\mathbb{Z}}_{nt} + \frac{1}{\sigma^2}\sum_{t=1}^{T}\tilde{V}'_{nt}(\theta^*)J_nM_n\tilde{\mathbb{Z}}_{nt}\right] & 0 & 0 & 0 \\ \frac{1}{\sigma^4}\sum_{t=1}^{T}\tilde{V}'_{nt}(\theta^*)J_nR_n(\lambda_2)\tilde{\mathbb{Z}}_{nt} & 0 & 0 & 0 \end{pmatrix}$$

$$+ \begin{pmatrix} \mathbf{0}_{k\times k} & * & * & * \\ \mathbf{0}_{1\times k} & \left[\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\lambda_2)(W_n\tilde{Y}_{nt}-\tilde{Y}_{n,t-1}))'J_nR_n(\lambda_2)W_n\tilde{Y}_{nt} + T\operatorname{tr}(J_nG_n^2(\lambda_1))\right] & * & * \\ \mathbf{0}_{1\times k} & \left[\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\lambda_2)(W_n\tilde{Y}_{nt}-\tilde{Y}_{n,t-1}))'J_nH_n(\lambda_2)\tilde{V}_{nt}(\theta^*) + \frac{1}{\sigma^2}\sum_{t=1}^{T}(M_n(W_n\tilde{Y}_{nt}-\tilde{Y}_{n,t-1}))'J_n\tilde{V}_{nt}(\theta^*)\right] & 0 & 0 \\ \mathbf{0}_{1\times k} & \frac{1}{\sigma^4}\sum_{t=1}^{T}(R_n(\lambda_2)(W_n\tilde{Y}_{nt}-\tilde{Y}_{n,t-1}))'J_n\tilde{V}_{nt}(\theta^*) & 0 & 0 \end{pmatrix}$$

$$+ \begin{pmatrix} \mathbf{0}_{k\times k} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times 1} \\ \mathbf{0}_{1\times k} & 0 & 0 & 0 \\ \mathbf{0}_{1\times k} & 0 & \left[\begin{matrix}\frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n(\lambda_2)\tilde{V}_{nt}(\theta^*))'J_nH_n(\lambda_2)\tilde{V}_{nt}(\theta^*) \\ +T\operatorname{tr}(J_nH_n^2(\lambda_2))\end{matrix}\right] & * \\ \mathbf{0}_{1\times k} & 0 & \frac{1}{\sigma^4}\sum_{t=1}^{T}(H_n(\lambda_2)\tilde{V}_{nt}(\theta^*))'J_n\tilde{V}_{nt}(\theta^*) & \left[\begin{matrix}-\frac{(n-1)T}{2\sigma^4} \\ +\frac{1}{\sigma^6}\sum_{t=1}^{T}(\tilde{V}'_{nt}(\theta^*)J_n\tilde{V}_{nt}(\theta^*))\end{matrix}\right] \end{pmatrix} \tag{49}$$

Box V.

likelihood function. From (48), the concentrated estimates are

$$\hat{\delta}^*_{nT}(\lambda_1,\lambda_2) = \left[\sum_{t=1}^{T}\tilde{\mathbb{Z}}'_{nt}R'_n(\lambda_2)J_nR_n(\lambda_2)\tilde{\mathbb{Z}}_{nt}\right]^{-1} \times \left[\sum_{t=1}^{T}\tilde{\mathbb{Z}}'_{nt}R'_n(\lambda_2)J_nR_n(\lambda_2)(\tilde{Y}_{nt}-\lambda_1(W_n\tilde{Y}_{nt}-\tilde{Y}_{n,t-1}))\right], \tag{51}$$

$$\hat{\sigma}^2_{nT}(\lambda_1,\lambda_2) = \frac{1}{(n-1)T}\sum_{t=1}^{T}\left[\tilde{Y}_{nt}-\lambda_1(W_n\tilde{Y}_{nt}-\tilde{Y}_{n,t-1}) - \tilde{\mathbb{Z}}_{nt}\hat{\delta}^*_{nT}(\lambda_1,\lambda_2)\right]' R'_n(\lambda_2)J_nR_n(\lambda_2) \times \left[\tilde{Y}_{nt}-\lambda_1(W_n\tilde{Y}_{nt}-\tilde{Y}_{n,t-1}) - \tilde{\mathbb{Z}}_{nt}\hat{\delta}^*_{nT}(\lambda_1,\lambda_2)\right].$$

Therefore, the concentrated likelihood on λ_1 and λ_2 from (16) is

$$\ln L_{nT}(\lambda_1,\lambda_2) = -\frac{(n-1)T}{2}(\ln(2\pi)+1) - \frac{(n-1)T}{2}\ln\hat{\sigma}^2_{nT}(\lambda_1,\lambda_2) + T[\ln|S^*_n(\lambda_1)| + \ln|R^*_n(\lambda_2)|]. \tag{52}$$

Denote

$$\mathcal{H}_{\lambda_{10},nT}(\lambda_2) = \mathcal{H}_{3,nT}(\lambda_2) - \mathcal{H}^{(o)\prime}_{2,nT}(\lambda_2)\mathcal{H}^{(o)-1}_{1,nT}(\lambda_2)\mathcal{H}^{(o)}_{2,nT}(\lambda_2), \tag{53}$$

$$\begin{aligned}\tilde{\mathcal{X}}_{z,nt}(\lambda_2) &= R_n(\lambda_2)(G_n\tilde{Z}^s_{nt}\delta_0 - \tilde{Y}^s_{n,t-1} - \tilde{\mathbb{Z}}_{nt}\mathcal{H}^{-1}_{1,nT}(\lambda_2)\mathcal{H}_{2,nT}(\lambda_2)) \\ &= R_n(\lambda_2)(G_n\tilde{Z}^s_{nt}\delta_0 - \tilde{Y}^s_{n,t-1} - \tilde{\mathbb{Z}}^{(o)\prime}_{nt}\mathcal{H}^{(o)-1}_{1,nT}(\lambda_2)\mathcal{H}^{(o)}_{2,nT}(\lambda_2)),\end{aligned} \tag{54}$$

and

$$\mathcal{V}_{x,nT}(\lambda_2) = \frac{1}{(n-1)T}\sum_{t=1}^{T}\tilde{\mathbb{Z}}'_{nt}R'_n(\lambda_2)J_nR_n(\lambda_2)S_n(\lambda_1)S_n^{-1}R_n^{-1}\tilde{V}_{nt}, \tag{55}$$

$$\mathcal{V}^{(o)}_{x,nT}(\lambda_2) = \frac{1}{(n-1)T}\sum_{t=1}^{T}\tilde{\mathbb{Z}}^{(o)\prime}_{nt}R'_n(\lambda_2)J_nR_n(\lambda_2)S_n(\lambda_1)S_n^{-1}R_n^{-1}\tilde{V}_{nt},$$

for notational simplicity. From (51), we have

$$\hat{\sigma}_{nT}^2(\lambda_1,\lambda_2) = (\lambda_{10}-\lambda_1)^2\mathcal{H}_{\lambda_{10},nT}(\lambda_2) + \frac{1}{(n-1)T} \times \sum_{t=1}^{T} \tilde{V}'_{nt}R_n'^{-1}S_n'^{-1}S_n'(\lambda_1)R_n'(\lambda_2)J_nR_n(\lambda_2) \times S_n(\lambda_1)S_n^{-1}R_n^{-1}\tilde{V}_{nt} + 2(\lambda_{10}-\lambda_1)\frac{1}{(n-1)T} \times \sum_{t=1}^{T}\tilde{\mathcal{X}}'_{z,nt}(\lambda_2)J_nR_n(\lambda_2)S_n(\lambda_1)S_n^{-1}R_n^{-1}\tilde{V}_{nt} - \mathcal{V}'_{x,nT}(\lambda_2)\mathcal{H}_{1,nT}^{-1}(\lambda_2)\mathcal{V}_{x,nT}(\lambda_2), \quad (56)$$

where $\mathcal{V}'_{x,nT}(\lambda_2)\mathcal{H}_{1,nT}^{-1}(\lambda_2)\mathcal{V}_{x,nT}(\lambda_2) = \mathcal{V}_{x,nT}^{(0)\prime}(\lambda_2)\mathcal{H}_{1,nT}^{(0)-1}(\lambda_2)\mathcal{V}_{x,nT}^{(0)}(\lambda_2)$. Hence, elements in $\hat{\sigma}_{nT}^2(\lambda_1,\lambda_2)$ can be considered as stable. Corresponding to (52), we have

$$Q_{nT}(\lambda_1,\lambda_2) = \max_{\delta^*,\sigma^2}\frac{1}{n(T-1)}\ln L_{nT}(\theta) = -(\ln(2\pi)+1) - \ln\sigma_{nT}^{*2}(\lambda_1,\lambda_2) + [\ln|S_n^*(\lambda_1)| + \ln|R_n^*(\lambda_2)|], \quad (57)$$

where

$$\sigma_{nT}^{*2}(\lambda_1,\lambda_2) = (\lambda_{10}-\lambda_1)^2E\mathcal{H}_{\lambda_{10},nT}(\lambda_2) + \sigma_0^2\frac{1}{n-1}\mathrm{tr}(R_n'^{-1}S_n'^{-1}S_n'(\lambda_1) \times R_n'(\lambda_2)J_nR_n(\lambda_2)S_n(\lambda_1)S_n^{-1}R_n^{-1}) + O\left(\frac{1}{T}\right). \quad (58)$$

By Lemmas 2 and 3, $\hat{\sigma}_{nT}^2(\lambda_{10},\lambda_{20}) = \sigma_0^2 + O_p\left(\max\left(\frac{1}{\sqrt{nT}},\frac{1}{T}\right)\right)$, and $\sigma_{nT}^{*2}(\lambda_{10},\lambda_{20}) = \sigma_0^2 + O_p\left(\frac{1}{T}\right)$.

C.4. Proof for Theorem 1

We shall scratch the essential arguments here only, because the remaining details are similar to those in Yu et al. (2008).

(i) $\lim_{T\to\infty}[Q_{nT}(\lambda_{10},\lambda_{20}) - Q_{nT}(\lambda_1,\lambda_2)] > 0$ *for any* $(\lambda_1,\lambda_2) \neq (\lambda_{10},\lambda_{20})$.

As $\sigma_n^2(\lambda_1,\lambda_2) = \frac{\sigma_0^2}{n-1}\mathrm{tr}[(R_n(\lambda_2)S_n(\lambda_1)S_n^{-1}R_n^{-1})'J_n(R_n(\lambda_2)S_n(\lambda_1)S_n^{-1}R_n^{-1})]$, we have

$$Q_{nT}(\lambda_1,\lambda_2) - Q_{nT}(\lambda_{10},\lambda_{20}) = T_{1,n}(\lambda_1,\lambda_2) - T_{2,nT}(\lambda_1,\lambda_2)$$

where

$$T_{1,n}(\lambda_1,\lambda_2) = -\frac{1}{2}[\ln\sigma_n^2(\lambda_1,\lambda_2) - \ln\sigma_{nT}^{*2}(\lambda_{10},\lambda_{20})] + \frac{1}{n}\ln|S_n(\lambda_1)| - \frac{1}{n}\ln|S_n(\lambda_{10})| + \frac{1}{n}\ln|R_n(\lambda_2)| - \frac{1}{n}\ln|R_n(\lambda_{20})| - \frac{1}{n-1}(\ln(1-\lambda_1) - \ln(1-\lambda_{10})) - \frac{1}{n-1}(\ln(1-\lambda_2) - \ln(1-\lambda_{20})),$$

and $T_{2,nT}(\lambda_1,\lambda_2) = \ln\left(1 + \frac{(\lambda_{10}-\lambda_1)^2\mathcal{H}_{\lambda_{10},nT}(\lambda_2)}{\sigma_n^2(\lambda_1,\lambda_2)}\right)$ with $\mathcal{H}_{\lambda_{10},nT}(\lambda_2)$ in (53).

Consider the pure spatial process $Y_{nt} = \lambda_{10}W_nY_{nt} + \alpha_{t0}l_n + U_{nt}$ with $U_{nt} = \lambda_{20}M_nU_{nt} + V_{nt}$ for a period t. By using the $F_{n,n-1}$ transformation to eliminate the time effect, the resulting log likelihood function of the transformed process is $\ln L_{p,n}(\lambda_1,\lambda_2,\sigma^2) = -\frac{n-1}{2}\ln 2\pi - \frac{n-1}{2}\ln\sigma^2 - \ln(1-\lambda_1) - \ln(1-\lambda_2) + \ln|S_n(\lambda_1)| + \ln|R_n(\lambda_2)| - \frac{1}{2\sigma^2}V'_{nt}(\lambda_1,\lambda_2)J_nV'_{nt}(\lambda_1,\lambda_2)$, where $V_{nt}(\lambda_1,\lambda_2) = R_n(\lambda_2)S_n(\lambda_1)Y_{nt}$. Let $Q_{p,n}(\lambda_1,\lambda_2) = \max_{\sigma^2}E\frac{1}{n}\ln L_{p,n}(\lambda_1,\lambda_2,\sigma^2)$. It follows that $Q_{p,n}(\lambda_1,\lambda_2) - Q_{p,n}(\lambda_{10},\lambda_{20}) = T_{1,n}(\lambda_1,\lambda_2)$. By the information inequality, $Q_{p,n}(\lambda_1,\lambda_2) - Q_{p,n}(\lambda_{10},\lambda_{20}) \le 0$. Thus, $T_{1,n}(\lambda_1,\lambda_2) \le 0$ for any (λ_1,λ_2).

Under Assumption 9(a), $\mathcal{H}_{\lambda_{10},nT}(\lambda_2)$ is positive so that $T_{2,nT}(\lambda_1,\lambda_2) > 0$ for $\lambda_1 \neq \lambda_{10}$ given any λ_2. Given λ_{10}, λ_{20} is the unique maximizer of $\lim T_{1,n}(\lambda_1,\lambda_2)$ under

$$\lim\left(\frac{1}{n-1}\ln\left|\sigma_0^2R_n^{-1\prime}J_nR_n^{-1}\right| - \frac{1}{n-1} \times \ln\left|\sigma_n^2(\lambda_1,\lambda_2)R_n^{-1}(\lambda_2)'J_nR_n^{-1}(\lambda_2)\right|\right) \neq 0 \quad \text{for } \lambda_2 \neq \lambda_{20}.$$

When Assumption 9(a) fails, identification requires that $T_{1,n}(\lambda_1,\lambda_2)$ is strictly less than zero. Under Assumption 9(b), $T_{1,n}(\lambda_1,\lambda_2) < 0$ whenever $(\lambda_1,\lambda_2) \neq (\lambda_{10},\lambda_{20})$. This proves the global identification.

(ii) $\frac{1}{(n-1)T}\ln L_{nT}(\lambda_1,\lambda_2) - Q_{nT}(\lambda_1,\lambda_2) \xrightarrow{p} 0$ *uniformly in* $\lambda_1 \in \Lambda_1$ *and* $\lambda_2 \in \Lambda_2$.

From (52) and (57), $\frac{1}{(n-1)T}\ln L_{nT}(\lambda_1,\lambda_2) - Q_{nT}(\lambda_1,\lambda_2) = -\frac{1}{2}\frac{1}{\tilde{\sigma}_{nT}^2(\lambda_1,\lambda_2)}(\hat{\sigma}_{nT}^2(\lambda_1,\lambda_2) - \sigma_{nT}^{*2}(\lambda_1,\lambda_2))$ by the mean value theorem where $\tilde{\sigma}_{nT}^2(\lambda_1,\lambda_2)$ lies between $\hat{\sigma}_{nT}^2(\lambda_1,\lambda_2)$ and $\sigma_{nT}^{*2}(\lambda_1,\lambda_2)$. From (56) and (58), as $\mathcal{V}_{x,nT}^{(0)}(\lambda_2) = o_p(1)$ from (22) and (24), we have $[\hat{\sigma}_{nT}^2(\lambda_1,\lambda_2) - \hat{\sigma}_{nT}^{*2}(\lambda_1,\lambda_2)] \xrightarrow{p} 0$ uniformly in λ_1 and λ_2. Also, as $\tilde{\sigma}_{nT}^2(\lambda_1,\lambda_2)$ lies between $\hat{\sigma}_{nT}^2(\lambda_1,\lambda_2)$ and $\sigma_{nT}^{*2}(\lambda_1,\lambda_2)$, we have $\frac{1}{\tilde{\sigma}_{nT}^2(\lambda_1,\lambda_2)} \le \max\{\frac{1}{\hat{\sigma}_{nT}^2(\lambda_1,\lambda_2)},\frac{1}{\sigma_{nT}^{*2}(\lambda_1,\lambda_2)}\}$. Because $\mathcal{H}_{\lambda_{10},nT}(\lambda_2)$ is nonnegative and $\sigma_n^2(\lambda_1,\lambda_2)$ is uniformly bounded away from zero, $\hat{\sigma}_{nT}^2(\lambda_1,\lambda_2)$ and $\sigma_{nT}^{*2}(\lambda_1,\lambda_2)$ are uniformly bounded away from zero. Hence, $\frac{1}{\tilde{\sigma}_{nT}^2(\lambda_1)}$ is uniformly bounded. Thus, $\frac{1}{(n-1)T}\ln L_{nT}(\lambda_1,\lambda_2) - Q_{nT}(\lambda_1,\lambda_2) \xrightarrow{p} 0$ uniformly in λ_1 and λ_2.

(iii) $Q_{nT}(\lambda_1,\lambda_2)$ *is uniformly equicontinuous in* λ_1 *and* λ_2 *in any compact parameter space* Λ_1 *and* Λ_2:

This follows because $\frac{1}{n-1}\ln|S_n(\lambda_1)|$, $\frac{1}{n-1}\ln|R_n(\lambda_2)|$, $(\lambda_1-\lambda_{10})^2\mathcal{H}_{\lambda_{10},nT}(\lambda_2)$ and $\sigma_n^2(\lambda_1,\lambda_2)$ are uniformly equicontinuous in λ_1 and λ_2.

The consistency of $(\hat{\lambda}_{1,nT},\hat{\lambda}_{2,nT})$ then follows from the above global identification, uniform convergence and uniform equicontinuity. Denote $\Upsilon_{\delta,T} = \mathrm{diag}(T, l'_{k+2})$. Using $\tilde{Y}_{nt} - \lambda_{10}(W_n\tilde{Y}_{nt} - \tilde{Y}_{n,t-1}) = \tilde{\mathbb{Z}}_{nt}\delta_0 + \tilde{U}_{nt}$ from (13), we have

$$\Upsilon_{\delta,T}\left(\hat{\delta}_{nT}^*(\lambda_{10},\lambda_{20}) - \delta_0\right) = \left[\frac{1}{(n-1)T}\sum_{t=1}^{T}\tilde{\mathbb{Z}}_{nt}^{(0)\prime}R_n'J_nR_n\tilde{\mathbb{Z}}_{nt}^{(0)}\right]^{-1} \times \left[\frac{1}{(n-1)T}\sum_{t=1}^{T}\tilde{\mathbb{Z}}_{nt}^{(0)\prime}R_n'J_nR_n\tilde{U}_{nt}\right].$$

The $\frac{1}{(n-1)T}\sum_{t=1}^{T}\tilde{\mathbb{Z}}_{nt}^{(0)\prime}R_n'J_nR_n\tilde{U}_{nt}$ can be decomposed into two parts $\frac{1}{(n-1)T}\sum_{t=1}^{T}\mathbb{Z}_{nt}^{(0)\prime}R_n'J_nR_nU_{nt}$ and $\frac{1}{(n-1)}\bar{\mathbb{Z}}_{nT}^{(0)\prime}R_n'J_nR_n\bar{U}_{nT}$, where the former is $O_p\left(\frac{1}{\sqrt{(n-1)T}}\right)$. From (22) and (24), $\frac{1}{(n-1)}\bar{\mathbb{Z}}_{nT}^{(0)\prime}R_n'J_nR_n\bar{U}_{nT}$ is $o_p(1)$. Hence, $\Upsilon_{\delta,T}\hat{\delta}_{nT}^*(\lambda_{10},\lambda_{20})$ is consistent. By the consistency of $\hat{\lambda}_{1,nT}$, $\hat{\lambda}_{2,nT}$, $\Upsilon_{\delta,T}\hat{\delta}_{nT}^*(\hat{\lambda}_{1,nT},\hat{\lambda}_{2,nT})$ converges in probability to $\Upsilon_{\delta,T}\hat{\delta}_{nT}^*(\lambda_{10},\lambda_{20})$. Thus, with the consistency of $\hat{\lambda}_{1,nT}$, $\hat{\lambda}_{2,nT}$ and $\Upsilon_{\delta,T}\hat{\delta}_{nT}^*(\hat{\lambda}_{1,nT},\hat{\lambda}_{2,nT})$, the consistency of $\hat{\sigma}_{nT}^2(\hat{\lambda}_{1,nT},\hat{\lambda}_{2,nT})$ in (56) follows. □

$$\frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L_{nT}(\theta_0^*)}{\partial\theta^*}=\begin{pmatrix}\frac{1}{\sigma_0^2\sqrt{(n-1)T}}\sum_{t=1}^{T}\tilde{\mathbb{Z}}_{nt}'R_n'J_n\tilde{V}_{nt}\\ \frac{1}{\sigma_0^2\sqrt{(n-1)T}}\sum_{t=1}^{T}(G_n\tilde{Z}_{nt}^s\delta_0-\tilde{Y}_{n,t-1}^s)'R_n'J_n\tilde{V}_{nt}+\frac{1}{\sigma_0^2\sqrt{(n-1)T}}\sum_{t=1}^{T}(\tilde{V}_{nt}'\ddot{G}_nJ_n\tilde{V}_{nt}-\sigma_0^2\mathrm{tr}(J_n\ddot{G}_n))\\ \frac{1}{\sigma_0^2\sqrt{(n-1)T}}\sum_{t=1}^{T}(\tilde{V}_{nt}'H_nJ_n\tilde{V}_{nt}-\sigma_0^2\mathrm{tr}(J_nH_n))\\ \frac{1}{2\sigma_0^4\sqrt{(n-1)T}}\sum_{t=1}^{T}(\tilde{V}_{nt}'J_n\tilde{V}_{nt}-(n-1)\sigma_0^2)\end{pmatrix} \tag{50}$$

Box VI.

$$a_{\theta,n}^*=\Phi_{1,T}^{-1}\begin{pmatrix}\frac{T}{2(1-\lambda_1)}\frac{1}{n-1}\mathrm{tr}(J_nR_n(\lambda_2)W_n^uR_n^{-1}(\lambda_2))+\frac{1}{n-1}\mathrm{tr}\left(J_nR_n(\lambda_2)\sum_{h=0}^{\infty}B_n^h(\theta)S_n^{-1}(\lambda_1)R_n^{-1}(\lambda_2)\right)\\ \frac{1}{n-1}\mathrm{tr}\left(J_nR_n(\lambda_2)(W_n-I_n)\left(\sum_{h=0}^{\infty}B_n^h(\theta)\right)S_n^{-1}(\lambda_1)R_n^{-1}(\lambda_2)\right)\\ \mathbf{0}_{k\times 1}\\ a_{\lambda,n}^*\\ \frac{1}{n-1}\mathrm{tr}(J_nH_n(\lambda_2))\\ \frac{1}{2\sigma^2}\end{pmatrix} \tag{59}$$

Box VII.

C.5. Bias terms

From $\Delta_{\theta_0^*,nT}$ in (22), by Lemma 3, the bias term is Eq. (59) given in Box VII where

$$a_{\lambda,n}^*=\frac{1}{n-1}\mathrm{tr}\Bigg(J_nR_n(\lambda_2)(\gamma G_n(\lambda_1)+\rho G_n(\lambda_1)W_n-I_n)\times\left(\sum_{h=0}^{\infty}B_n^h(\theta)\right)S_n^{-1}(\lambda_1)R_n^{-1}(\lambda_2)\Bigg)+\frac{1}{n-1}\mathrm{tr}(J_n\ddot{G}_n(\lambda_1)).$$

The first element in $a_{\theta,n}^*$ is obtained from $\frac{1}{\sigma_0^2}\frac{1}{n-1}(\bar{Y}_{n,T-1}^{ud})'R_n'J_n\bar{V}_{nT}=\frac{1}{2(1-\lambda_{10})}\frac{(T-1)}{T}\frac{1}{n-1}\mathrm{tr}(J_nR_nW_n^uR_n^{-1})+O_p(\frac{1}{\sqrt{n}})$ where $\bar{Y}_{n,T-1}^{ud}$ is the variable $\bar{Y}_{n,T-1}^{u}$ derived by subtracting $W_n^uY_{n,-1}$, i.e., $\bar{Y}_{n,T-1}^{ud}$ is the time average of $\frac{1}{(1-\lambda_{10})}W_n^u\sum_{h=0}^{t}\left(\mathbf{c}_{n0}+X_{n,t-h}\beta_0+U_{n,t-h}\right)$ from (3), and $\frac{1}{\sigma_0^2}\frac{1}{n-1}\bar{Y}_{n,T-1}^{s\prime}R_n'J_n\bar{V}_{nT}=\frac{1}{T}\frac{1}{n-1}\mathrm{tr}\left(J_nR_n\sum_{h=0}^{\infty}B_n^hS_n^{-1}\right)+O(\frac{1}{T^2})+O_p(\frac{1}{\sqrt{nT^2}})$ from Lemma 3. The (59) can be simplified to (24) as follows.

Let us investigate $\mathrm{tr}\left(J_nR_n(\lambda_2)\sum_{h=0}^{\infty}B_n^h(\theta)S_n^{-1}(\lambda_1)R_n^{-1}(\lambda_2)\right)$ first. As $J_n=I_n-\frac{1}{n}l_nl_n'$ and W_n and M_n are row-normalized,

$$\begin{aligned}&\mathrm{tr}\left(J_nR_n(\lambda_2)\sum_{h=0}^{\infty}B_n^h(\theta)S_n^{-1}(\lambda_1)R_n^{-1}(\lambda_2)\right)\\&=\mathrm{tr}\left(\sum_{h=0}^{\infty}B_n^h(\theta)S_n^{-1}(\lambda_1)\right)-\frac{1}{n(1-\lambda_1)(1-\lambda_2)}\times\mathrm{tr}\left(R_n(\lambda_2)\sum_{h=0}^{\infty}B_n^h(\theta)l_nl_n'\right)\\&=\mathrm{tr}\left(\sum_{h=0}^{\infty}B_n^h(\theta)S_n^{-1}(\lambda_1)\right)-\frac{1}{(1-\lambda_1)}.\end{aligned}$$

The last equality follows from $B_n(\theta)l_n=0$ because $B_n(\theta)=\Gamma\tilde{D}_n(\theta)\Gamma^{-1}$, $\tilde{D}_n(\theta)=\mathrm{diag}\{0,\ldots,0,d_{n,m_n+1},\ldots,d_{nn}\}$ and $\Gamma^{-1}l_n=\sqrt{n}e_n$ where $e_n=(1,0,\ldots,0)'$. Similar arguments and $(W_n-I_n)l_n=0$ imply

$$\mathrm{tr}\left(J_nR_n(\lambda_2)(W_n-I_n)\left(\sum_{h=0}^{\infty}B_n^h(\theta)\right)S_n^{-1}(\lambda_1)R_n^{-1}(\lambda_2)\right)=\mathrm{tr}\left((W_n-I_n)\left(\sum_{h=0}^{\infty}B_n^h(\theta)\right)S_n^{-1}(\lambda_1)\right),$$

and

$$a_{\lambda,n}^*=\frac{1}{n-1}\mathrm{tr}\Bigg((\gamma G_n(\lambda_1)+\rho G_n(\lambda_1)W_n-I_n)\times\left(\sum_{h=0}^{\infty}B_n^h(\theta)\right)S_n^{-1}(\lambda_1)\Bigg)+\frac{1}{n-1}\mathrm{tr}(J_n\ddot{G}_n(\lambda_1)),$$

when $\gamma+\rho+\lambda_1=1$. For the bias item related with W_n^u, we have

$$\mathrm{tr}(J_nR_n(\lambda_2)W_n^uR_n^{-1}(\lambda_2))=\mathrm{tr}(W_n^u)-\frac{1}{n(1-\lambda_2)}\mathrm{tr}(R_n(\lambda_2)W_n^ul_nl_n')=m_n-1,$$

by using $W_n^u=\Gamma_n\mathbb{J}_n\Gamma_n^{-1}$. Thus, the bias term (59) is simplified to (24).

For a further simplification, we note that $B_n(\theta)=\Gamma\tilde{D}_n(\theta)\Gamma^{-1}$ and $S_n^{-1}(\lambda_1)=\Gamma(I_n-\lambda_1\varpi_n)^{-1}\Gamma^{-1}$ implies $B_n(\theta)S_n^{-1}(\lambda_1)=\Gamma\tilde{D}_n(\theta)(I_n-\lambda_1\varpi_n)^{-1}\Gamma^{-1}$. Thus,

$$\begin{aligned}\mathrm{tr}\left(\sum_{h=0}^{\infty}B_n^h(\theta)S_n^{-1}(\lambda_1)\right)&=\mathrm{tr}\left(S_n^{-1}(\lambda_1)\right)+\mathrm{tr}\left(\sum_{h=1}^{\infty}B_n^h(\theta)S_n^{-1}(\lambda_1)\right)\\&=\left(\frac{m_n}{(1-\lambda_1)}+\sum_{j=m_n+1}^{n}\frac{1}{(1-\lambda_1\omega_{nj})}\right)\\&\quad+\sum_{j=m_n+1}^{n}\frac{d_{nj}(\theta)}{(1-d_{nj}(\theta))}\frac{1}{(1-\lambda_1\omega_{nj})}\end{aligned}$$

$$= \frac{m_n}{(1-\lambda_1)} + \sum_{j=m_n+1}^{n} \frac{1}{(1-d_{nj}(\theta))(1-\lambda_1\omega_{nj})}.$$

With $\mathrm{tr}(G_n) = \sum_{j=1}^{n} \frac{\omega_{nj}}{(1-\lambda_1\omega_{nj})}$, the bias term (24) can also be written in terms of the eigenvalues of W_n.

C.6. Proof for Theorem 2

From the reparameterization, $\frac{1}{(n-1)T}\sum_{t=1}^{T}(W_n\tilde{Y}_{n,t-1}-\tilde{Y}_{n,t-1})' R_n' J_n R_n (W_n\tilde{Y}_{n,t-1}-\tilde{Y}_{n,t-1}) = O_p(1)$, which behaves like the stable components. Hence, $\frac{1}{(n-1)T}\sum_{t=1}^{T}(W_n\tilde{Y}_{n,t-1}-\tilde{Y}_{n,t-1})'R_n'J_nR_n\tilde{Y}^u_{n,t-1} = O_p(T)$ by the Cauchy–Schwarz inequality. This implies that the rescaling with $\Sigma^{(o)}_{\theta_0^*,nT} = \Phi_{1,T}^{-1}\Sigma_{\theta_0^*,nT}\Phi_{1,T}^{-1}$ is appropriate.

For the term $-\frac{1}{(n-1)T}\Phi_{1,T}^{-1}\frac{\partial^2 \ln L_{n,T}(\bar{\theta}^*_{nT})}{\partial\theta\partial\theta'}\Phi_{1,T}^{-1}$ in the Taylor expansion in (21), by using Lemmas 2 and 3 and (38)–(41) in Yu et al. (2008),

$$-\frac{1}{(n-1)T}\Phi_{1,T}^{-1}\frac{\partial^2 \ln L_{n,T}(\bar{\theta}^*_{nT})}{\partial\theta\partial\theta'}\Phi_{1,T}^{-1} = \Sigma^{(o)}_{\theta_0^*,nT} + O_p\left(\max\left(\frac{1}{\sqrt{(n-1)T}},\frac{1}{T}\right)\right). \tag{60}$$

For the term $\frac{1}{\sqrt{(n-1)T}}\Phi_{1,T}^{-1}\frac{\partial \ln L_{n,T}(\theta_0^*)}{\partial\theta}$ in (21), from (22) and (24),

$$\frac{1}{\sqrt{(n-1)T}}\Phi_{1,T}^{-1}\frac{\partial \ln L_{n,T}(\theta_0^*)}{\partial\theta} = \frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L^{(o)}_{nT}(\theta_0^*)}{\partial\theta^*} - \sqrt{\frac{n-1}{T}}a_{\theta_0^*,n} + O_p\left(\max\left(\sqrt{\frac{n-1}{T^3}},\frac{1}{\sqrt{T}}\right)\right). \tag{61}$$

By combining (60) and (61) together with (21) where $\frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L^{(o)}_{nT}(\theta_0^*)}{\partial\theta^*}$ is asymptotically normal under both n and T are large, we have the results in Theorem 2. □

C.7. Proof for Theorem 4

Consistency: For the objective function $g^{(o)\prime}_{nT}(\pi^*)\Sigma^{(o)-1}_{nT}g^{(o)}_{nT}(\pi^*)$ where $\pi^* = (\kappa^{*\prime},\lambda_2)$, we first derive the uniform convergence of $\frac{1}{(n-1)(T-1)}a_{nT}g^{(o)}_{nT}(\kappa^{*\prime},\lambda_2)$ where $a_{nT} = (\Sigma^{(o)}_{nT})^{-1/2}$. Combined with the identification in Assumption 9′, the consistency of optimum GMME $\hat{\pi}^*_{o,nT} \equiv (\hat{\kappa}^{*\prime}_{o,nT},\hat{\lambda}_{2,o,nT})'$ will follow. Partition the $(m+q)\times(m+q)$ matrix a_{nT} as $(a^{(1)}_{nT},\ldots,a^{(m)}_{nT},a^{(Q)}_{nT})$. Then,

$$\frac{1}{(n-1)(T-1)}a_{nT}g^{(o)}_{nT}(\pi^*) = \frac{1}{(n-1)(T-1)} \times a^{(Q)}_{nT}\Upsilon_q^{-1}\ddot{\mathbf{Q}}^{*\prime}_{n,T-1}\mathbf{V}^{**}_{n,T-1}(\pi^*) + \frac{1}{(n-1)(T-1)} \times \mathbf{V}^{**\prime}_{n,T-1}(\pi^*)\left(\sum_{j=1}^{m}a^{(j)}_{nT}\mathbf{P}_{n,T-1,l}\right)\mathbf{V}^{**}_{n,T-1}(\pi^*), \tag{62}$$

where, by expansion, $\mathbf{V}^{**}_{n,T-1}(\pi^*) = \mathbf{d}^*_{n,T-1}(\pi^*) + \mathbf{R}^*_{n,T-1}(\lambda_2)\mathbf{S}^*_{n,T-1}(\lambda_1)\mathbf{S}^{*-1}_{n,T-1}\mathbf{R}^{*-1}_{n,T-1}\mathbf{V}^{**}_{n,T-1}$ with $\mathbf{d}^*_{n,T-1}(\pi^*) = \mathbf{R}^*_{n,T-1}(\lambda_2)[(\lambda_{10}-\lambda_1)\mathbf{G}^*_{n,T-1}\mathbf{Z}^*_{n,T-1}\delta_0 + \mathbf{Z}^*_{n,T-1}(\delta_0-\delta)]$.

For the first part in (62), $\frac{1}{(n-1)(T-1)}a^{(Q)}_{nT}\Upsilon_q^{-1}\ddot{\mathbf{Q}}^{*\prime}_{n,T-1}\mathbf{V}^{**}_{n,T-1}(\pi^*) = \frac{1}{(n-1)(T-1)}a^{(Q)}_{nT}\Upsilon_q^{-1}\ddot{\mathbf{Q}}^{*\prime}_{n,T-1}\mathbf{d}^*_{n,T-1}(\pi^*) + \frac{1}{(n-1)(T-1)}a^{(Q)}_{nT}\Upsilon_q^{-1}\ddot{\mathbf{Q}}^{*\prime}_{n,T-1}\mathbf{R}^*_{n,T-1}(\lambda_2)\mathbf{S}^*_{n,T-1}(\lambda_1)\mathbf{S}^{*-1}_{n,T-1}\mathbf{R}^{*-1}_{n,T-1}\mathbf{V}^*_{n,T-1}$, where its second term is $o_p(1)$ uniformly in $\pi^*\in\Pi^*$ because Q^*_{nt} is predetermined and its elements are $O_p(1)$ uniformly in n and t. For the second part in (62), we have

$$\mathbf{V}^{**\prime}_{n,T-1}(\pi^*)\left(\sum_{j=1}^{m}a^{(j)}_{nT}\mathbf{P}_{n,T-1,l}\right)\mathbf{V}^{**}_{n,T-1}(\pi^*) = \mathbf{d}^{*\prime}_{n,T-1}(\pi^*)\left(\sum_{j=1}^{m}a^{(j)}_{nT}\mathbf{P}_{n,T-1,l}\right)\mathbf{d}^*_{n,T-1}(\pi^*) + \mathbf{l}_{n,T-1}(\pi^*) + \mathbf{q}_{n,T-1}(\pi^*)$$

where $\mathbf{l}_{n,T-1}(\pi^*) = \mathbf{d}^{*\prime}_{n,T-1}(\pi^*)(\sum_{j=1}^{m}a^{(j)}_{nT}\mathbf{P}_{n,T-1,l})\mathbf{R}^*_{n,T-1}(\lambda_2)\mathbf{S}^*_{n,T-1}(\lambda_1)\mathbf{S}^{*-1}_{n,T-1}\mathbf{R}^{*-1}_{n,T-1}\mathbf{V}^{**}_{n,T-1}$ and

$$\mathbf{q}_{n,T-1}(\pi^*) = (\mathbf{R}^*_{n,T-1}(\lambda_2)\mathbf{S}^*_{n,T-1}(\lambda_1)\mathbf{S}^{*-1}_{n,T-1}\mathbf{R}^{*-1}_{n,T-1}\mathbf{V}^{**}_{n,T-1})'\left(\sum_{j=1}^{m}a^{(j)}_{nT}\mathbf{P}_{n,T-1,l}\right) \times \mathbf{R}^*_{n,T-1}(\lambda_2)\mathbf{S}^*_{n,T-1}(\lambda_1)\mathbf{S}^{*-1}_{n,T-1}\mathbf{R}^{*-1}_{n,T-1}\mathbf{V}^{**}_{n,T-1}.$$

For any n-dimensional column vectors p_{nt} and q_{nt}, we have

$$(p'_{n1},\ldots,p'_{nT})(F_{T,T-1}\otimes I_n)(F'_{T,T-1}\otimes I_n)(q'_{n1},\ldots,q'_{nT})' = (p'_{n1},\ldots,p'_{nT})(J_T\otimes I_n)(q'_{n1},\ldots,q'_{nT})' = \sum_{t=1}^{T}\tilde{p}'_{nt}\tilde{q}_{nt}$$

by using $(\tilde{p}_{n1},\ldots,\tilde{p}_{nT}) = (p_{n1},\ldots,p_{nT})J_T$. Thus, Lemma 3 can be applied so that $\frac{1}{(n-1)(T-1)}\mathbf{l}_{n,T-1}(\theta)$ and $\frac{1}{(n-1)(T-1)}\mathbf{q}_{n,T-1}(\theta)$ converge uniformly to well defined limits.

The above uniform convergence and the identification condition in Assumption 9′ imply the consistency of the estimator.

Distribution: By the Taylor expansion,

$$\sqrt{(n-1)(T-1)}\Phi_{k+4}(\hat{\pi}^*_{o,nT}-\pi_0^*) = -\left[\frac{\partial g^{(o)\prime}_{nT}(\hat{\pi}^*_{o,nT})/\partial\pi^*}{(n-1)(T-1)}(\Sigma^{(o)}_{nT})^{-1}\frac{\partial g^{(o)}_{nT}(\bar{\pi}^*_{o,nT})/\partial\pi^{*\prime}}{(n-1)(T-1)}\right]^{-1} \times \frac{\partial g^{(o)\prime}_{nT}(\hat{\pi}^*_{o,nT})/\partial\pi^*}{(n-1)(T-1)}(\Sigma^{(o)}_{nT})^{-1}\frac{g^{(o)}_{nT}(\pi_0^*)}{\sqrt{(n-1)(T-1)}},$$

where $\bar{\pi}^*_{o,nT}$ lies between $\hat{\pi}^*_{o,nT}$ and π_0^*. By Lemma 3 and $\hat{\pi}^*_{o,nT}-\pi_0^* = o_p(1)$, we have $\frac{1}{(n-1)(T-1)}\frac{\partial g^{(o)}_{nT}(\hat{\pi}^*_{o,nT})}{\partial\pi^{*\prime}} = D^{(o)}_{nT} + R^{(o)}_{nT} + o_p(1)$ where $D^{(o)}_{nT}$ is $O(1)$ and $R^{(o)}_{nT}$ is $O(1/T)$. Here, for the elements of $R^{(o)}_{nT}$ in (28), we have Eqs. (63) and (64) given in Box VIII where $P^s_{nl} = P_{nl} + P'_{nl}$ for $l = 1,\ldots,m$. By denoting $\mathcal{D}^{(o)}_{nT} = D^{(o)}_{nT} + R^{(o)}_{nT}$, we have $\frac{1}{(n-1)(T-1)}\frac{\partial g^{(o)}_{nT}(\hat{\pi}^*_{o,nT})}{\partial\pi^{*\prime}} = \mathcal{D}^{(o)}_{nT} + o_p(1)$. Thus, $\frac{\partial g^{(o)\prime}_{nT}(\hat{\pi}^*_{o,nT})/\partial\pi^*}{(n-1)(T-1)}(\Sigma^{(o)}_{nT})^{-1}\frac{\partial g^{(o)}_{nT}(\bar{\pi}^*_{o,nT})/\partial\pi^{*\prime}}{(n-1)(T-1)} = \mathcal{D}^{(o)\prime}_{nT}(\Sigma^{(o)}_{nT})^{-1}\mathcal{D}^{(o)}_{nT} + o_p(1)$. Also, by using the CLT, we have $\frac{1}{\sqrt{(n-1)(T-1)}}(\Sigma^{(o)}_{nT})^{-1/2}g^{(o)}_{nT}(\theta_0) \xrightarrow{d} N(0,\mathrm{plim}_{n\to\infty}I_{(m+q)})$. For large T, the CLT in Yu et al. (2008) can be applied. For the case of a finite T, the CLT in Kelejian and Prucha (2001) can be extended here.[25] Thus,

$$\sqrt{(n-1)(T-1)}\Phi_{k+4}(\hat{\pi}^*_{o,nT}-\pi_0^*) \xrightarrow{d} N\left(0,\mathrm{plim}_{n\to\infty}\left(\mathcal{D}^{(o)\prime}_{nT}(\Sigma^{(o)}_{nT})^{-1}\mathcal{D}^{(o)}_{nT}\right)^{-1}\right).$$

[25] The linear moments in Kelejian and Prucha (2001) are based on the product of nonstochastic variables and disturbances. However, as the proof of the CLT in Kelejian and Prucha (2001) is based on the martingale CLT, it can be extended to cover the predetermined variables situation without additional complication.

$$b_{lz} = \sigma_0^2 \begin{pmatrix} \frac{T}{2(1-\lambda_1)} \frac{1}{n-1} \mathrm{tr}(J_n P_n^s J_n R_n W_n^u R_n^{-1}) + \frac{1}{n-1} \mathrm{tr}\left(J_n P_n^s J_n R_n \frac{1}{T-1} \sum_{h=1}^{T-1} (T-h) B_n^{h-1} S_n^{-1} R_n^{-1} \right) \\ \frac{1}{n-1} \mathrm{tr}\left(J_n P_n^s J_n R_n (W_n - I_n) \frac{1}{T-1} \sum_{h=1}^{T-1} (T-h) B_n^{h-1} S_n^{-1} R_n^{-1} \right) \\ \mathbf{0}_{k\times 1} \end{pmatrix}' \tag{63}$$

and

$$b_{l\lambda} = \frac{\sigma_0^2}{n-1} \mathrm{tr}\left(J_n P_n^s J_n R_n (\gamma_0 G_n + \rho_0 G_n W_n - I_n) \frac{1}{T-1} \sum_{h=1}^{T-1} (T-h) B_n^{h-1} S_n^{-1} R_n^{-1} \right) \tag{64}$$

Box VIII.

When $\Sigma_{nT}^{(o)}$ is replaced by $\hat{\Sigma}_{nT}^{(o)}$ so that $\hat{\Sigma}_{nT}^{(o)} = \Sigma_{nT}^{(o)} + o_p(1)$, we will have the same asymptotic distribution by similar arguments as Proposition 2 in Lee (2007). □

C.8. Comparing variance matrices of MLE and GMME

For the (rescaled) variance matrix of MLE for π^*, by using the inverse of the partitioned matrix of (17), it is $[\Sigma_{\pi_0^*,nT}]^{-1}$ where

$$\Sigma_{\pi_0^*,nT} = \frac{1}{\sigma_0^2} \begin{pmatrix} E\mathcal{H}_{nT}^{(o)} & * \\ \mathbf{0}_{1\times(k+3)} & 0 \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{(k+2)\times(k+2)} & * \\ \mathbf{0}_{2\times(k+2)} & \boldsymbol{\Sigma}_{2,mle,n} \end{pmatrix} \tag{65}$$

where $\boldsymbol{\Sigma}_{2,mle,n} = \frac{1}{n-1}\begin{pmatrix} \mathrm{tr}(\mathbb{C}_n'\mathbb{C}_n) & \mathrm{tr}(\mathbb{C}_n'\mathbb{D}_n) \\ \mathrm{tr}(\mathbb{D}_n'\mathbb{C}_n) & \mathrm{tr}(\mathbb{D}_n'\mathbb{D}_n) \end{pmatrix}$ with $\mathbb{C}_n$ and $\mathbb{D}_n$ defined after (19). For the (rescaled) variance matrix of GMME, it is the inverse of $D_{nT}^{(o)\prime}\Sigma_{nT}^{(o)-1}D_{nT}^{(o)}$ and

$$D_{nT}^{(o)\prime}\Sigma_{nT}^{(o)-1}D_{nT}^{(o)} = \frac{1}{\sigma_0^2} \begin{pmatrix} E\mathcal{G}_{nT}^{(o)} & * \\ \mathbf{0}_{1\times(k+3)} & 0 \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{(k+2)\times(k+2)} & * \\ \mathbf{0}_{2\times(k+2)} & \boldsymbol{\Sigma}_{2,gmm,nT} \end{pmatrix} \tag{66}$$

where

$$\mathcal{G}_{nT}^{(o)} = \frac{1}{(n-1)(T-1)} (\ddot{\mathbf{Z}}_{n,T-1}^{**}\Phi_{k+3}^{-1}, \ddot{\mathbf{G}}_{n,T-1}^{*}\ddot{\mathbf{Z}}_{n,T-1}^{**}\delta_0 - \ddot{\mathbf{Y}}_{n,T-1}^{(**,-1)})' \times \mathbf{M}_{Q,nT}(\ddot{\mathbf{Z}}_{n,T-1}^{**}\Phi_{k+3}^{-1}, \ddot{\mathbf{G}}_{n,T-1}^{*}\ddot{\mathbf{Z}}_{n,T-1}^{**}\delta_0 - \ddot{\mathbf{Y}}_{n,T-1}^{(**,-1)}),$$

$$\boldsymbol{\Sigma}_{2,gmm,nT} = \frac{1}{(n-1)(T-1)} C_{mn,T}\Delta_{mn,T}^{-1}C_{mn,T}',$$

with

$$C_{mn,T} = \begin{bmatrix} \mathrm{tr}((\mathbf{P}_{n,T-1,1}^{*\prime} + \mathbf{P}_{n,T-1,1}^{*})\ddot{\mathbf{G}}_{n,T-1}^{*}) & \cdots & \mathrm{tr}((\mathbf{P}_{n,T-1,m}^{*\prime} + \mathbf{P}_{n,T-1,m}^{*})\ddot{\mathbf{G}}_{n,T-1}^{*}) \\ \mathrm{tr}((\mathbf{P}_{n,T-1,1}^{*\prime} + \mathbf{P}_{n,T-1,1}^{*})\mathbf{H}_{n,T-1}^{*}) & \cdots & \mathrm{tr}((\mathbf{P}_{n,T-1,m}^{*\prime} + \mathbf{P}_{n,T-1,m}^{*})\mathbf{H}_{n,T-1}^{*}) \end{bmatrix}$$

and $\Delta_{mn,T}$ in (29).

To compare (65) and (66), we can investigate two components separately. For $\mathcal{H}_{nT}^{(o)}$ and $\mathcal{G}_{nT}^{(o)}$, we have

$$\begin{aligned} &(\ddot{\mathbf{Z}}_{n,T-1}^{**}\Phi_{k+3}^{-1}, \ddot{\mathbf{G}}_{n,T-1}^{*}\ddot{\mathbf{Z}}_{n,T-1}^{**}\delta_0 - \ddot{\mathbf{Y}}_{n,T-1}^{(**,-1)})'\mathbf{M}_{Q,nT}(\ddot{\mathbf{Z}}_{n,T-1}^{**}\Phi_{k+3}^{-1}, \ddot{\mathbf{G}}_{n,T-1}^{*}\ddot{\mathbf{Z}}_{n,T-1}^{**}\delta_0 - \ddot{\mathbf{Y}}_{n,T-1}^{(**,-1)}) \\ &\le (\ddot{\mathbf{Z}}_{n,T-1}^{**}\Phi_{k+3}^{-1}, \ddot{\mathbf{G}}_{n,T-1}^{*}\ddot{\mathbf{Z}}_{n,T-1}^{**}\delta_0 - \ddot{\mathbf{Y}}_{n,T-1}^{(**,-1)})'(\ddot{\mathbf{Z}}_{n,T-1}^{**}\Phi_{k+3}^{-1}, \ddot{\mathbf{G}}_{n,T-1}^{*}\ddot{\mathbf{Z}}_{n,T-1}^{**}\delta_0 - \ddot{\mathbf{Y}}_{n,T-1}^{(**,-1)}) \\ &= \sum_{t=1}^{T} (\tilde{\mathbb{Z}}_{nt}^{(o)}, (G_n\tilde{Z}_{nt}^{s}\delta_0 - \tilde{Y}_{n,t-1}^{s}))'R_n'J_nR_n(\tilde{\mathbb{Z}}_{nt}^{(o)}, (G_n\tilde{Z}_{nt}^{s}\delta_0 - \tilde{Y}_{n,t-1}^{s})), \end{aligned}$$

which implies that $\lim \mathcal{G}_{nT}^{(o)} \le \lim \mathcal{H}_{nT}^{(o)}$. For $\boldsymbol{\Sigma}_{2,mle,n}$ and $\boldsymbol{\Sigma}_{2,gmm,nT}$, by the Cauchy–Schwarz inequality, we have

$$C_{mn,T}\Delta_{mn,T}^{-1}C_{mn,T}' \le (T-1)\begin{pmatrix} \mathrm{tr}(\mathbb{C}_n'\mathbb{C}_n) & \mathrm{tr}(\mathbb{C}_n'\mathbb{D}_n) \\ \mathrm{tr}(\mathbb{D}_n'\mathbb{C}_n) & \mathrm{tr}(\mathbb{D}_n'\mathbb{D}_n) \end{pmatrix},$$

which implies that $\lim(\boldsymbol{\Sigma}_{2,gmm,nT} - \boldsymbol{\Sigma}_{2,mle,n}) \le 0$. Hence, the variance matrix of MLE would be smaller than that of GMME asymptotically.[26]

Appendix D. Unit root SDPD[27]

This section investigates the unit root case where all the eigenvalues of (1) are equal to 1.

Assumption 2′. $\rho_0 + \gamma_0 + \lambda_{10} = 1$ with $\gamma_0 = 1$.

Under Assumption 2′, we have $A_n = (I_n - \lambda_{10}W_n)^{-1}(I_n - \lambda_{10}W_n) = I_n$. Hence, all the eigenvalues of A_n will be 1. This holds even if the weights matrix W_n is not row-normalized. Also, there is no stable component Y_{nt}^s in (3).[28]

With the re-parameterization in (14), define $\mathbb{Z}_{nt}^* = [Y_{n,t-1}^*, W_n^*Y_{n,t-1}^*, X_{nt}^*]$ and $\theta^* = (\delta^{*\prime}, \lambda_1, \lambda_2, \sigma^2)'$ with $\delta^* = (\gamma, \rho^*, \beta)'$ and $\rho^* = \rho + \lambda_1$. Thus, the elements of $Y_{n,t-1}^*$ and $W_n^*Y_{n,t-1}^*$ have a higher order so that $\hat{\gamma}_{nT}$ and $\hat{\rho}_{nT}^*$ might be superconsistent, while other parameters will have regular $\sqrt{nT}$-consistency. For the reparameterized regressor associated with λ_{10}, from (12), we have $W_n^*\tilde{Y}_{nt}^* - W_n^*\tilde{Y}_{n,t-1}^* = G_n^*\tilde{Z}_{nt}^{*s}\delta_0 - W_n^*\tilde{Y}_{n,t-1}^{*s} + G_n^*\tilde{U}_{nt}^*$.

D.1. Concentrated likelihood

The concentrated likelihood function of (14) is

$$\ln L_{nT}(\theta^*) = -\frac{(n-1)T}{2}\ln(2\pi\sigma^2) + T\ln\left|S_n^*(\lambda_1)\right| + T\ln\left|R_n^*(\lambda_2)\right| - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{V}_{nt}'(\theta^*)J_n\tilde{V}_{nt}(\theta^*). \tag{67}$$

For (67), the dominant part of the information matrix $-E\Big(\frac{1}{(n-1)T}\frac{\partial^2 \ln L_{nT}(\theta_0^*)}{\partial\theta^*\partial\theta^{\prime *}}\Big)$ is Eq. (68) given in Box IX, where $\mathcal{H}_{nT} = \frac{1}{(n-1)T}$

[26] When $\mu_4 - 3\sigma_0^4 \neq 0$, the two variance matrices cannot be compared directly.

[27] In this section, the notations of $\mathbb{Z}_{nt}$ are different from that in Section 3. In Section 3, $\mathbb{Z}_{nt}^* = [Y_{n,t-1}^*, W_n^*Y_{n,t-1}^* - Y_{n,t-1}^*, X_{nt}^*]$ so that the first element in $\mathbb{Z}_{nt}^*$ is unstable. Here, $\mathbb{Z}_{nt}^* = [Y_{n,t-1}^*, W_n^*Y_{n,t-1}^*, X_{nt}^*]$ so that both the first and second elements are unstable.

[28] Thus, Assumption 7 and the second part in Assumption 8 are irrelevant in this section.

$$\Sigma_{\theta_0^*,nT} = \frac{1}{\sigma_0^2}\begin{pmatrix} E\mathcal{H}_{nT} & * & * \\ \mathbf{0}_{1\times(k+1)} & 0 & * \\ \mathbf{0}_{1\times(k+1)} & 0 & 0 \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{k\times k} & * & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n-1}\mathrm{tr}((\ddot{G}_n' + \ddot{G}_n)J_n\ddot{G}_n) & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n-1}\mathrm{tr}((H_n' + H_n)J_n\ddot{G}_n) & \frac{1}{n-1}\mathrm{tr}((H_n'+H_n)J_nH_n) & * \\ \mathbf{0}_{1\times k} & \frac{1}{\sigma_0^2(n-1)}\mathrm{tr}(J_n\ddot{G}_n) & \frac{1}{\sigma_0^2(n-1)}\mathrm{tr}(J_nH_n) & \frac{1}{2\sigma_0^4} \end{pmatrix} \quad (68)$$

Box IX.

$\sum_{t=1}^{T}(\tilde{\mathbb{Z}}_{nt}, (\tilde{G}_n\tilde{Z}_{nt}^s\delta_0 - W_n\tilde{Y}_{n,t-1}^s))'R_n'J_nR_n(\tilde{\mathbb{Z}}_{nt}, (\tilde{G}_n\tilde{Z}_{nt}^s\delta_0 - W_n\tilde{Y}_{n,t-1}^s))$. For this unit root model, we need to have a different rescaling matrix $\Phi_{2,T} = \mathrm{diag}(T, T, l_{k+3}')$ so that $\Sigma_{\theta_0^*,nT}^{(o)} \equiv \Phi_{2,T}^{-1}\Sigma_{\theta_0^*,nT}\Phi_{2,T}^{-1}$ is $O(1)$.

D.2. Asymptotic properties of ML estimators

Under Assumption 9 (where $\mathcal{H}_{nT}(\lambda_2)$ has different elements from Section 3 due to the different reparameterization in this section), we have the following consistency theorem.

Theorem 5. *Under Assumptions 1, 2′, 3–6 and 9, when both n and T are large, θ_0^* is identified and $\Phi_{2,T}(\hat{\theta}_{nT}^* - \theta_0^*) \xrightarrow{p} 0$ for the QMLE $\hat{\theta}_{nT}^*$ that maximizes (67).*

Proof. Similar to Theorem 1. □

Thus, the estimate $\hat{\lambda}_{nT}$ is consistent, and $\hat{\gamma}_{nT}$ and $\hat{\rho}_{nT}^*$ are superconsistent. Similar to Section 3, the asymptotic distribution of $\hat{\theta}_{nT}^*$ can be derived from the Taylor expansion of $\frac{\partial \ln L_{nT}(\hat{\theta}_{nT}^*)}{\partial\theta}$ around θ_0^*. Different from Section 3, the bias term here is[29]

$$a_{\theta,n}^* = \left(\frac{1}{2(n-1)}\mathrm{tr}\left(J_nS_n^{-1}(\lambda_1)\right), \frac{1}{2(n-1)}\mathrm{tr}\left(J_nG(\lambda_1)\right), \mathbf{0}_{1\times k}, \frac{1}{n-1}\mathrm{tr}(J_nG_n(\lambda_1)), \frac{1}{n-1}\mathrm{tr}(J_nH_n(\lambda_2)), \frac{1}{2\sigma^2}\right)'.$$

For the unit root SDPD case, denote

$$P_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & I_k & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & I_2 \end{pmatrix} \quad \text{and}$$

$$\Psi_{3,T} = \Phi_{1,T}P_3\Phi_{2,T}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & T^{-1} & 0 & -1 & 0 \\ 0 & 0 & I_k & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & I_2 \end{pmatrix},$$

so that $\theta = P_3\theta^*$.

Theorem 6. *Under Assumptions 1, 2′, 3–6, and 9 (a) (Assumption 9 (a) can be replaced by 9 (b) and 10), when both n and T are large,*[30]

$$\Phi_{2,T}\sqrt{(n-1)T}(\hat{\theta}_{nT}^* - \theta_0^*) + \sqrt{\frac{n-1}{T}}b_{\theta_0,nT}^* + O_p\left(\max\left(\sqrt{\frac{n-1}{T^3}}, \sqrt{\frac{1}{T}}\right)\right) \xrightarrow{d} N\left(0, \lim_{T\to\infty}(\Sigma_{\theta_0^*,nT}^{(o)})^{-1}(\Sigma_{\theta_0^*,nT}^{(o)} + \Omega_{\theta_0,nT})(\Sigma_{\theta_0^*,nT}^{(o)})^{-1}\right),$$

where $b_{\theta_0,nT}^* \equiv [\Sigma_{\theta_0^*,nT}^{(o)}]^{-1}\cdot a_{\theta_0,n}^*$, and

$$\Phi_{1,T}\sqrt{(n-1)T}(\hat{\theta}_{nT} - \theta_0) + \Psi_{3,T}\sqrt{\frac{n-1}{T}}b_{\theta_0,nT}^* + O_p\left(\max\left(\sqrt{\frac{n-1}{T^3}}, \sqrt{\frac{1}{T}}\right)\right) \xrightarrow{d} N\Big(0, \lim_{T\to\infty}\Psi_{3,T}(\Sigma_{\theta_0^*,nT}^{(o)})^{-1}(\Sigma_{\theta_0^*,nT}^{(o)} + \Omega_{\theta_0,nT})(\Sigma_{\theta_0^*,nT}^{(o)})^{-1}\Psi_{3,T}'\Big).$$

Proof. Similar to Theorem 2. □

Hence, $(\hat{\theta}_{nT} - \theta_0) = \Phi_{1,T}^{-1}\big[-\Psi_{3,T}\frac{1}{T}b_{\theta_0,nT}^* + O_p\big(\max(\frac{1}{\sqrt{nT}}, \frac{1}{T^2})\big)\big]$. As $\Phi_{1,T} = \mathrm{diag}(T, l_{k+4}')$, $\hat{\gamma}_{nT}$ is superconsistent. Because $\Psi_{3,T}$ is singular in the limit, it implies that some linear combination of $\hat{\theta}_{nT}$ will also be superconsistent. This is exactly what we have seen from $\hat{\rho}_{nT}^*$. Similar to Section 3, an analytical bias reduction procedure is $\hat{\theta}_{nT}^{*1} = \hat{\theta}_{nT}^* - \frac{\hat{\varphi}_{nT}^*}{T}$ and $\hat{\theta}_{nT}^1 = \hat{\theta}_{nT} - \frac{\hat{\varphi}_{nT}}{T}$, where we may have $\hat{\varphi}_{nT}^* = \left[\left(E\left(\frac{1}{(n-1)T}\frac{\partial^2 \ln L_{n,T}(\theta^*)}{\partial\theta^*\partial\theta^{*\prime}}\right)\right)^{-1}a_{\theta,n}^*\right]\Big|_{\theta^*=\hat{\theta}_{nT}^*}$ and $\hat{\varphi}_{nT} = \left[\left(E\left(\frac{1}{(n-1)T}\frac{\partial^2 \ln L_{n,T}(\theta)}{\partial\theta\partial\theta'}\right)\right)^{-1}a_n(\theta)\right]\Big|_{\theta=\hat{\theta}_{nT}}$ with $a_n(\theta) = P_3'^{-1}\Phi_{2,T}a_{\theta,n}^*$. We can show that under $T/(n-1)^{1/3} \to \infty$, $\hat{\theta}_{nT}^{*1}$ and $\hat{\theta}_{nT}^1$ are $\sqrt{(n-1)T}$ consistent and asymptotically centered normal even when $n/T \to \infty$.

Theorem 7. *If both n and T are large and $T/(n-1)^{1/3} \to \infty$, under Assumptions 1, 2′, 3–6, and 9 (a) (Assumption 9 (a) can be replaced by 9 (b) and 10),*

$$\Phi_{2,T}\sqrt{(n-1)T}(\hat{\theta}_{nT}^{*1} - \theta_0) \xrightarrow{d} N(0, \lim_{T\to\infty}(\Sigma_{\theta_0^*,nT}^{(o)})^{-1}(\Sigma_{\theta_0^*,nT}^{(o)} + \Omega_{\theta_0,nT})(\Sigma_{\theta_0^*,nT}^{(o)})^{-1}),$$

and $\Phi_{1,T}\sqrt{(n-1)T}(\hat{\theta}_{nT}^1 - \theta_0) \xrightarrow{d} N(0, \lim_{T\to\infty}\Psi_{3,T}(\Sigma_{\theta_0^*,nT}^{(o)})^{-1}(\Sigma_{\theta_0^*,nT}^{(o)} + \Omega_{\theta_0,nT})(\Sigma_{\theta_0^*,nT}^{(o)})^{-1}\Psi_{3,T}')$.

Proof. Similar to Theorem 4 in Yu et al. (2008). □

[29] Because $\mathrm{tr}(J_nR_nS_n^{-1}(\lambda_1)R_n^{-1})$ is equal to $\mathrm{tr}(J_nS_n^{-1}(\lambda_1))$, the R_n matrix does not appear.

[30] For the unit root SDPD model here, the row-normalization of W_n and M_n in Assumption 1 can be relaxed. However, the parameter spaces of λ_j's need to be adjusted accordingly. Some expressions of relevant statistics, such as biases, would adjust accordingly. We keep Assumption 1 for simplicity.

References

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

Baltagi, B., Egger, P., Pfaffermayr, M., 2007. A generalized spatial panel data model with random effects. Working Paper. Syracuse University.

Baltagi, B., Song, S.H., Koh, W., 2003. Testing panel data regression models with spatial error correlation. Journal of Econometrics 117, 123–150.
Choi, I., 2006. Nonstationary panels. In: Mills, T., Patterson, K. (Eds.), Palgrave Handbooks of Econometrics, vol. 1, Econometric Theory. Palgrave Macmillan, pp. 511–539.
Hausman, J.A., 1978. Specification tests in econometrics. Econometrica 46, 1251–1271.
Horn, R., Johnson, C., 1985. Matrix Algebra. Cambridge University Press.
Johansen, S., 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. Econometrica 59, 1551–1580.
Kapoor, M., Kelejian, H.H., Prucha, I.R., 2007. Panel data models with spatially correlated error components. Journal of Econometrics 140, 97–130.
Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance. Journal of Real Estate Finance and Economics 17 (1), 99–121.
Kelejian, H.H., Prucha, I.R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40, 509–533.
Kelejian, H.H., Prucha, I.R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. Journal of Econometrics 104, 219–257.
Keller, W., Shiue, C.H., 2007. The origin of spatial interaction. Journal of Econometrics 140, 304–332.
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., Yu, J., 2010a. Estimation of spatial autoregressive panel data models with fixed effects. Journal of Econometrics 154, 165–185.
Lee, L.F., Yu, J., 2010b. Some recent developments in spatial panel data models. Regional Science and Urban Economics 40, 255–271.
Ord, J.K., 1975. Estimation methods for models of spatial interaction. Journal of the American Statistical Association 70, 120–126.
Phillips, P.C.B., 1995. Fully modified least squares and vector autoregression. Econometrica 63, 1023–1078.
Sims, C.A., Stock, J.H., Watson, M.W., 1990. Inference in linear time series models with some unit roots. Econometrica 58, 113–144.
Yu, J., de Jong, R., Lee, L.-F., 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., Lee, L.F., 2010. Estimation of unit root spatial dynamic panel data models. Econometric Theory 26, 1332–1362.