Econometric Theory, **26**, 2010, 564–597.
doi:10.1017/S0266466609100099

A SPATIAL DYNAMIC PANEL DATA MODEL WITH BOTH TIME AND INDIVIDUAL FIXED EFFECTS

LUNG-FEI LEE
Ohio State University
JIHAI YU
University of Kentucky

This paper establishes asymptotic properties of quasi-maximum likelihood estimators for spatial dynamic panel data with both time and individual fixed effects when the number of individuals n and the number of time periods T can be large. We propose a data transformation approach to eliminate the time effects. When $n/T \to 0$, the estimators are $\sqrt{nT}$ consistent and asymptotically centered normal; when n is asymptotically proportional to T, they are $\sqrt{nT}$ consistent and asymptotically normal, but the limit distribution is not centered around 0; when $n/T \to \infty$, the estimators are consistent with rate T and have a degenerate limit distribution. We also propose a bias correction for our estimators. When $n^{1/3}/T \to 0$, the correction will asymptotically eliminate the bias and yield a centered confidence interval. The estimates from the transformation approach can be consistent when n is a fixed finite number.

1. INTRODUCTION

Recently, there has been a growing literature on the estimation of dynamic panel data models when both n and T are large (see, among others, Phillips and Moon, 1999; Hahn and Kuersteiner, 2002; Alvarez and Arellano, 2003; Hahn and Newey, 2004). For panel data with spatial interaction, we have Baltagi, Song, and Koh (2003), Kapoor, Kelejian, and Prucha (2007), and Yu, de Jong, and Lee (2007, 2008). Those papers focus on models with only individual effects. However, in the panel data literature, there are also two-way error component regression models where we have not only unobserved individual effects but also unobserved time effects (see, among others, Wallace and Hussain, 1969; Amemiya, 1971; Nerlove, 1971; Hahn and Moon, 2006). Hence, it is natural to study the spatial dynamic panel data (SDPD) models when both n and T are large with both individual

We thank two anonymous referees and the co-editor Jinyong Hahn for their comments and suggestions for improving this paper. Lee acknowledges financial support from NSF under grant SES-0519204. Address correspondence to Lung-fei Lee, Department of Economics, Ohio State University, Columbus, OH 43210, USA; e-mail: lflee@econ.ohio-state.edu.

 © Cambridge University Press, 2009 0266-4666/10 $15.00

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

effects and time effects. This would be useful for empirical applications where the time dummy effects might be important, such as in growth theory and regional economics (for recent empirical applications of panel data models with both time dummy effects and spatial effects, see Ertur and Koch, 2007; Foote, 2007).

In Yu et al. (2008), the consistency and asymptotic distribution of the maximum likelihood (ML) estimator and quasi–maximum likelihood (QML) estimator are established for the SDPD model when individual effects are included. Also, a bias correction procedure for the estimator is proposed. It is shown that as long as $n^{1/3}/T \to 0$, the correction will asymptotically eliminate the bias of the order $O(T^{-1})$ and yield a centered confidence interval. This approach can be directly extended to the estimation of models with both individual and time fixed effects, where "direct" means we also estimate the fixed time effects jointly with other parameters. When there are time effects included in the model, we might have additional bias of the order $O(n^{-1})$ in the estimates if we estimate the time effects directly (see Theorem 6 in Section 4.2). In this paper, we propose a transformation procedure to eliminate the time effects that can avoid the additional $O(n^{-1})$ order bias with the same asymptotic efficiency as the direct QML estimates when n is not relatively smaller than T. Our transformation procedure is particularly useful when $n/T \to 0$. For the latter, the estimates of the transformed approach have faster rates of convergence than those of the direct estimates. The direct estimates have a degenerate limit distribution, but the transformed estimates are properly centered and are asymptotically normal. The estimates from the transformed approach can even be consistent and asymptotically normal when n is a fixed finite number.

This paper is organized as follows. In Section 2, the model is introduced and the data transformation procedure is proposed. We then explain our method of estimation, which is a concentrated QML estimation. Section 3 establishes the consistency and asymptotic distribution of the QML estimator of the transformation approach and its bias corrected estimator. Section 4 presents the asymptotic properties of the direct approach estimator. Simulation results are reported to compare the two approaches. Section 5 concludes the paper. Some useful lemmas and proofs are collected in the Appendixes.

2. THE MODEL

2.1. Data Generating Process and Data Transformation

The model considered in this paper is

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

where $Y_{nt} = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ and $V_{nt} = (v_{1t}, v_{2t}, \ldots, v_{nt})'$ are $n \times 1$ column vectors and v_{it} is independent and identically distributed (i.i.d.) across i and t with zero mean and variance σ_0^2. Also, W_n is an $n \times n$ spatial weights matrix[1] that is

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

nonstochastic and generates the spatial dependence among cross sectional units y_{it}, X_{nt} is an $n \times k_x$ matrix of nonstochastic regressors, $\mathbf{c}_{n0}$ is an $n \times 1$ column vector of individual fixed effects, α_{t0} is a scalar of time effect and l_n is an $n \times 1$ column vector of ones.[2] The total number of parameters in this model is equal to the sum of the number of individuals n and the number of time periods T, plus the dimension of the common parameters $(\gamma, \rho, \beta', \lambda, \sigma^2)'$, which is $k_x + 4$. The matrix W_n is usually row normalized from a symmetric matrix, which ensures that all the weights are between 0 and 1 and weighting operations can be interpreted as an average of the neighboring values. Also, a spatial weights matrix row normalized has the property $W_n l_n = l_n$.

Define $S_n(\lambda) = I_n - \lambda W_n$ and $S_n \equiv S_n(\lambda_0) = I_n - \lambda_0 W_n$. Then, presuming S_n is invertible and denoting $A_n = S_n^{-1}(\gamma_0 I_n + \rho_0 W_n)$, (1) can be rewritten as $Y_{nt} = A_n Y_{n,t-1} + S_n^{-1} X_{nt}\beta_0 + S_n^{-1}\mathbf{c}_{n0} + \alpha_{t0} S_n^{-1} l_n + S_n^{-1} V_{nt}$. Assuming that the infinite sums are well defined, by continuous substitution,

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

$$= \mu_n + \mathcal{X}_{nt}\beta_0 + l_n \sum_{h=0}^{\infty} \frac{\alpha_{t-h,0}}{1-\lambda_0} \left(\frac{\gamma_0 + \rho_0}{1-\lambda_0}\right)^h + U_{nt}, \tag{2}$$

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

When both n and T go to infinity, one way to estimate (1) is to estimate all the parameters including both the time effects and individual effects in the model, which will yield a bias of the order $O(\max(n^{-1}, T^{-1}))$ for the common parameters (see Theorem 6). In this paper, we will introduce a data transformation approach (to eliminate the time effects) such that we can avoid the bias of the order $O(n^{-1})$, whereas the estimator may have the same asymptotic efficiency as the direct QML estimator. This transformation procedure is particularly useful when $n/T \to 0$, where the estimates of the transformed approach have faster rates of convergence than the rates of convergence of the direct estimates. Also, when $n/T \to 0$, the direct estimates have a degenerate limit distribution, but the transformed estimates are properly centered and are asymptotically normal. In the event that n is a fixed finite number, the estimates from the direct approach will be inconsistent whereas the estimates from the transformation approach can still be consistent and asymptotically normal.

Let $J_n = I_n - (1/n) l_n l_n'$ be the deviation from the group mean transformation, which is a projector. When $W_n l_n = l_n$, we have $J_n W_n = J_n W_n (J_n + (1/n) l_n l_n') = J_n W_n J_n$ because $J_n W_n l_n = J_n l_n = \mathbf{0}$. Hence,

$$(J_n Y_{nt}) = \lambda_0 (J_n W_n)(J_n Y_{nt}) + \gamma_0 (J_n Y_{n,t-1}) + \rho_0 (J_n W_n)(J_n Y_{n,t-1})$$

$$+ (J_n X_{nt})\beta_0 + (J_n \mathbf{c}_{n0}) + (J_n V_{nt}), \tag{3}$$

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

which does not involve the time effects, and $J_n\mathbf{c}_{n0}$ can be regarded as the transformed individual effects. Thus, we can estimate θ_0 and $J_n\mathbf{c}_{n0}$ on the basis of the transformed equation (3), where relevant variables are premultiplied by J_n.

A special feature of the transformed equation (3) is that the variance matrix of J_nV_{nt} is equal to $\sigma_0^2 J_n$ so that the elements of J_nV_{nt} are correlated. Also, J_n is singular with rank $(n-1)$ as J_n is an orthogonal projector with trace $(n-1)$. Hence, there is a linear dependence among the elements of J_nV_{nt}. An effective estimation method will eliminate the linear dependence in sample observations. This can be done with the eigenvalues and eigenvectors decomposition as in the theory of generalized inverses for the estimation of linear regression models (see, e.g., Theil, 1971, Ch. 6).

The eigenvalues of J_n are a single zero and $(n-1)$ ones. An eigenvector corresponding to the zero eigenvalue is proportional to l_n. Let $(F_{n,n-1}, l_n/\sqrt{n})$ be the orthonormal matrix of eigenvectors of J_n where $F_{n,n-1}$ corresponds to the eigenvalues of ones and $l_n/\sqrt{n}$ corresponds to the eigenvalue zero. As is derived in Section A.1 of Appendix A, the transformation of J_nY_{nt} to Y_{nt}^*, where $Y_{nt}^* = F'_{n,n-1}J_nY_{nt} = F'_{n,n-1}Y_{nt}$ is of dimension $(n-1)$, gives

$$Y_{nt}^* = \lambda_0 W_n^* Y_{nt}^* + \gamma_0 Y_{n,t-1}^* + \rho_0 W_n^* Y_{n,t-1}^* + X_{nt}^*\beta_0 + \mathbf{c}_{n0}^* + V_{nt}^*, \tag{4}$$

where $W_n^* = F'_{n,n-1}W_nF_{n,n-1}$, $X_{nt}^* = F'_{n,n-1}J_nX_{nt} = F'_{n,n-1}X_{nt}$, $\mathbf{c}_{n0}^* = F'_{n,n-1}J_n\mathbf{c}_{n0} = F'_{n,n-1}\mathbf{c}_{n0}$, $V_{nt}^* = F'_{n,n-1}J_nV_{nt} = F'_{n,n-1}V_{nt}$, and V_{nt}^* is an $(n-1)$-dimensional disturbance vector with zero mean and variance matrix $\sigma_0^2 I_{n-1}$. Equation (4) will provide the estimation of the structural parameters in the model. This equation is in the format of a typical spatial autoregressive model in panel data, where the number of observations is $T(n-1)$, reduced from the original sample observations by one for each period. Equation (4) is useful as it motivates the derivation of the likelihood function for Y_{nt}^* in our approach.

2.2. The Method of Maximum Likelihood Estimation

Denote $\theta = (\delta', \lambda, \sigma^2)'$ and $\zeta = (\delta', \lambda, \mathbf{c}_n')'$ where $\delta = (\gamma, \rho, \beta')'$. At the true value, $\theta_0 = (\delta_0', \lambda_0, \sigma_0^2)'$ and $\zeta_0 = (\delta_0', \lambda_0, \mathbf{c}_{n0}')'$ where $\delta_0 = (\gamma_0, \rho_0, \beta_0')'$. Suppose that V_{nt} is normally distributed $N(0, \sigma_0^2 I_n)$, the transformed V_{nt}^* will be $N(0, \sigma_0^2 I_{n-1})$, and (4) can be estimated by ML. The log likelihood function of (4) for Y_{nt}^* is

$$\ln L_{n,T}(\theta, \mathbf{c}_n^*) = -\frac{(n-1)T}{2}\ln 2\pi - \frac{(n-1)T}{2}\ln\sigma^2 + T\ln\left|I_{n-1} - \lambda W_n^*\right| - \frac{1}{2\sigma^2}\sum_{t=1}^{T} V_{nt}^{*\prime}(\theta)V_{nt}^*(\theta), \tag{5}$$

where $V_{nt}^*(\theta) = (I_{n-1} - \lambda W_n^*)Y_{nt}^* - Z_{nt}^*\delta - \mathbf{c}_n^*$ with $Z_{nt} = (Y_{n,t-1}, W_nY_{n,t-1}, X_{nt})$.[3] To use (4) for effective estimation, the determinant and inverse of $(I_{n-1} - \lambda W_n^*)$

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

are needed. As is derived in Section A.2 of Appendix A, $(I_{n-1}-\lambda W_n^*)=F'_{n,n-1}(I_n-\lambda W_n)F_{n,n-1}$, $|I_{n-1}-\lambda W_n^*|=1/(1-\lambda)\,|I_n-\lambda W_n|$, and $(I_{n-1}-\lambda W_n^*)^{-1}=F'_{n,n-1}(I_n-\lambda W_n)^{-1}F_{n,n-1}$. Thus, the tractability in computing the determinant of $(I_{n-1}-\lambda W_n^*)$ is exactly that of $(I_n-\lambda W_n)$. When W_n is constructed as a weights matrix row normalized from an original symmetric matrix, Ord (1975) has suggested a computationally tractable method for the evaluation of $|I_n-\lambda W_n|$ at various λ for the ML method. Thus, this will also be useful for evaluating the determinant of $(I_{n-1}-\lambda W_n^*)$ even though the row sums of the transformed spatial weights matrix W_n^* may not even be unity. Furthermore, a spatial autoregressive model is an equilibrium model in the sense that the observed outcomes are determined by the equation. That is, the matrix $(I_{n-1}-\lambda W_n^*)$ will be invertible. For the transformed equation (4), $(I_{n-1}-\lambda W_n^*)$ is invertible as long as the original matrix $(I_n-\lambda W_n)$ in (1) is invertible (see Section A.2 of Appendix A).

Also, as we have

$$
\begin{aligned}
&(I_{n-1}-\lambda W_n^*)Y_{nt}^*-Z_{nt}^*\delta-\mathbf{c}_n^*\\
&\quad=F'_{n,n-1}(I_n-\lambda W_n)F_{n,n-1}\cdot F'_{n,n-1}Y_{nt}-F'_{n,n-1}Z_{nt}\delta-F'_{n,n-1}\mathbf{c}_n\\
&\quad=F'_{n,n-1}(I_n-\lambda W_n)(I_n-\frac{1}{n}l_nl_n')Y_{nt}-F'_{n,n-1}Z_{nt}\delta-F'_{n,n-1}\mathbf{c}_n\\
&\quad=F'_{n,n-1}[(I_n-\lambda W_n)Y_{nt}-Z_{nt}\delta-\mathbf{c}_n],
\end{aligned}
$$

because $F'_{n,n-1}W_nl_n=F'_{n,n-1}l_n=\mathbf{0}$, it follows that

$$
\begin{aligned}
&V_{nt}^{*\prime}(\theta)V_{nt}^*(\theta)\\
&\quad=[(I_{n-1}-\lambda W_n^*)Y_{nt}^*-Z_{nt}^*\delta-\mathbf{c}_n^*]'[(I_{n-1}-\lambda W_n^*)Y_{nt}^*-Z_{nt}^*\delta-\mathbf{c}_n^*]\\
&\quad=[(I_n-\lambda W_n)Y_{nt}-Z_{nt}\delta-\mathbf{c}_n]'F_{n,n-1}F'_{n,n-1}[(I_n-\lambda W_n)Y_{nt}-Z_{nt}\delta-\mathbf{c}_n]\\
&\quad=[(I_n-\lambda W_n)Y_{nt}-Z_{nt}\delta-\mathbf{c}_n]'J_n[(I_n-\lambda W_n)Y_{nt}-Z_{nt}\delta-\mathbf{c}_n],
\end{aligned}
$$

by $F_{n,n-1}F'_{n,n-1}=J_n$. Hence, the log likelihood function (5) for Y_{nt}^* can be expressed in terms of Y_{nt} as

$$
\begin{aligned}
\ln L_{n,T}(\theta,\mathbf{c}_n)=&-\frac{(n-1)T}{2}\ln 2\pi-\frac{(n-1)T}{2}\ln\sigma^2-T\ln(1-\lambda)\\
&+T\ln|I_n-\lambda W_n|-\frac{1}{2\sigma^2}\sum_{t=1}^{T}V'_{nt}(\theta)J_nV_{nt}(\theta),
\end{aligned}
\tag{6}
$$

where $V_{nt}(\theta)=(I_n-\lambda W_n)Y_{nt}-Z_{nt}\delta-\mathbf{c}_n$ and J_n can be read as the generalized inverse of $\sigma_0^{-2}\mathrm{Var}(J_nV_{nt})$.

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

2.3. First-Order Condition and Second-Order Condition of the ML Estimator

Therefore, if we transform the data from Y_{nt} to $Y_{nt}^* = F'_{n,n-1}Y_{nt}$, we will maximize (5) by searching over the parameter space. This is equivalent to the estimation of the spatial dynamic panel model with only individual effects with $(n-1)$ cross-section units and T time periods. Alternatively, one may maximize (6) instead. However, although the components of V_{nt} are i.i.d. in the model, the elements of V_{nt}^* might not be independent in general, even though they are uncorrelated. The original asymptotic analysis in Yu et al. (2008) does not directly carry over to the transformed model with the disturbances V_{nt}^*.[4] As (5) is equivalent to (6), it turns out that we can analyze the asymptotic distribution of the estimator just from (6).

Using first-order conditions, we concentrate out $\mathbf{c}_n$ in (6) and get[5]

$$\ln L_{n,T}(\theta) = -\frac{(n-1)T}{2}\ln 2\pi - \frac{(n-1)T}{2}\ln\sigma^2 - T\ln(1-\lambda) + T\ln|I_n - \lambda W_n| - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{V}'_{nt}(\theta)J_n\tilde{V}_{nt}(\theta), \tag{7}$$

where $\tilde{V}_{nt}(\theta) = (I_n - \lambda W_n)\tilde{Y}_{nt} - \tilde{Z}_{nt}\delta$ and $J_n\tilde{V}_{nt}(\theta) = J_n[(I_n - \lambda W_n)\tilde{Y}_{nt} - \tilde{Z}_{nt}\delta - \tilde{\alpha}_t l_n]$ because $J_n l_n = \mathbf{0}$. The concentrated likelihood function is different from that of the direct approach[6] (see equation (20) in Section 4) in that we have an attachment of $[\frac{T}{2}\ln 2\pi\sigma^2 - T\ln(1-\lambda)]$. For the concentrated likelihood function (7), the first-order derivatives and the second-order derivatives are (C.2) and (C.3) in Section C2 of Appendix C. Let $J_n\tilde{V}_{nt} = J_n[S_n\tilde{Y}_{nt} - \tilde{Z}_{nt}\delta_0 - \tilde{\alpha}_{t0}l_n]$. For the first-order derivative evaluated at θ_0, denote $G_n = W_n S_n^{-1}$; we have

$$\frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma_0^2}\frac{1}{\sqrt{(n-1)T}}\sum_{t=1}^{T}\tilde{Z}'_{nt}J_n\tilde{V}_{nt} \\ \frac{1}{\sigma_0^2}\frac{1}{\sqrt{(n-1)T}}\sum_{t=1}^{T}(G_n\tilde{Z}_{nt}\delta_0)'J_n\tilde{V}_{nt} + \frac{1}{\sigma_0^2}\frac{1}{\sqrt{(n-1)T}}\sum_{t=1}^{T}(\tilde{V}'_{nt}G'_nJ_n\tilde{V}_{nt} - \sigma_0^2\mathrm{tr}J_nG_n) \\ \frac{1}{2\sigma_0^4}\frac{1}{\sqrt{(n-1)T}}\sum_{t=1}^{T}(\tilde{V}'_{nt}J_n\tilde{V}_{nt} - (n-1)\sigma_0^2) \end{pmatrix}, \tag{8}$$

which is a linear and quadratic form of $\tilde{V}_{nt}$. For the information matrix, let $\Sigma_{\theta_0,nT} = -\mathrm{E}\left(\frac{1}{(n-1)T}\frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'}\right)$ and $\mathcal{H}_{nT} = 1/((n-1)T)\sum_{t=1}^{T}(\tilde{Z}_{nt}, G_n\tilde{Z}_{nt}\delta_0)'$

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

$J_n(\tilde{Z}_{nt}, G_n\tilde{Z}_{nt}\delta_0)$. We have

$$\Sigma_{\theta_0,nT} = \frac{1}{\sigma_0^2}\begin{pmatrix} \mathrm{E}\mathcal{H}_{nT} & * \\ \mathbf{0}_{1\times(k_x+3)} & 0 \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{(k_x+2)\times(k_x+2)} & * & * \\ \mathbf{0}_{1\times(k_x+2)} & \frac{1}{n-1}\left[\mathrm{tr}(G_n'J_nG_n)+\mathrm{tr}((J_nG_n)^2)\right] & * \\ \mathbf{0}_{1\times(k_x+2)} & \frac{1}{\sigma_0^2(n-1)}\mathrm{tr}(J_nG_n) & \frac{1}{2\sigma_0^4} \end{pmatrix}$$

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

3. QML ESTIMATORS

For our analysis of the asymptotic properties of the estimator, we make the following assumptions.

Assumption 1. W_n is a row normalized nonstochastic spatial weights matrix.

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

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

Assumption 4. The elements of X_{nt} are nonstochastic and bounded[7] uniformly in n and t, and $\lim_{T\to\infty}\frac{1}{nT}\sum_{t=1}^{T}\tilde{X}_{nt}'J_n\tilde{X}_{nt}$ exists and is nonsingular.

Assumption 5. W_n is uniformly bounded in row and column sums in absolute value (UB).[8] Also $S_n^{-1}(\lambda)$ is UB uniformly in $\lambda \in \Lambda$.

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

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

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

Assumption 1 is a standard normalization assumption in spatial econometrics. In many empirical applications, the rows of W_n sum to 1, which ensures that all the weights are between 0 and 1. If the row normalization in Assumption 1 does not hold, the transformation approach will not work, and we can use the direct approach presented in Section 4. Assumption 2 provides regularity assumptions for v_{it}, and our analysis is based on i.i.d. disturbances. If there is unknown heteroskedasticity, the ML estimator (or QML estimator) would not be consistent. Consistency methods such as the generalized method of moment (GMM) in Lin and Lee (2005) and that in Kelejian and Prucha (2007) may be designed for the model (which is for future research). If the disturbances have a parametric variance function or spatial correlation structure, the ML method may need to be generalized to take into account those structures. Invertibility of $S_n(\lambda)$ in Assumption 3 guarantees that (2) is valid. Also, compactness is a condition for theoretical analysis. In empirical application where W_n is row normalized, one just searches over a parameter space on $(-1, 1)$. For the case where W_n is not row normalized but its eigenvalues are real, Λ can be a closed interval contained in $(-1/|\omega_{n,\min}|, 1/\omega_{n,\max})$ where $\omega_{n,\min}$ and $\omega_{n,\max}$ are the minimum and maximum eigenvalues of W_n (Anselin, 1988). When exogenous variables X_{nt} are included in the model, it is convenient to assume that the exogenous regressors are uniformly bounded as in Assumption 4. Assumption 5 was originated by Kelejian and Prucha (1998, 2001) and is also used in Lee (2004, 2007). The uniform boundedness of W_n and $S_n^{-1}(\lambda)$ is a condition that limits the spatial correlation to a manageable degree. Assumption 6 is the absolute summability condition and row/column sum boundedness condition, which will play an important role in deriving asymptotic properties of the QML estimator. This assumption is essential for the paper because it limits the dependence between time series and between cross-sectional units.[9] Assumption 7 allows two cases: (a) $n \to \infty$ as $T \to \infty$; (b) n is fixed as $T \to \infty$. If n is fixed, the system (1), is a dynamic structural model, and its reduced form has a vector autoregressive representation with restricted coefficients and stationary properties (Hamilton, 1994, Ch. 11).

3.1. Consistency

For the concentrated log likelihood function (7) divided by the effective sample size $(n-1)T$, the corresponding expected value function is $Q_{n,T}(\theta) = \mathrm{E}\max_{\mathbf{c}_n} 1/((n-1)T)\ln L_{n,T}(\theta, \mathbf{c}_n)$, which is

$$Q_{n,T}(\theta) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 - \frac{1}{n-1}\ln(1-\lambda) + \frac{1}{n-1}\ln|S_n(\lambda)| - \frac{1}{2\sigma^2(n-1)T}\mathrm{E}\left(\sum_{t=1}^{T}\tilde{V}_{nt}'(\theta)J_n\tilde{V}_{nt}(\theta)\right). \quad (9)$$

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

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

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

For identification, if the information matrix $-\mathrm{E}\left(\frac{1}{(n-1)T}\frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'}\right)$ is nonsingular and $-\mathrm{E}\left(\frac{1}{(n-1)T}\frac{\partial^2 \ln L_{n,T}(\theta)}{\partial\theta\partial\theta'}\right)$ has full rank for any θ in some neighborhood $N(\theta_0)$ of θ_0, the parameters are locally identified (see Rothenberg, 1971). For the information matrix, using Lemma 16 in Yu et al. (2008), we have

$$\Sigma_{\theta_0,nT} = \frac{1}{\sigma_0^2}\begin{pmatrix} \mathrm{E}\mathcal{H}_{nT} & * \\ \mathbf{0}_{1\times(k_x+3)} & 0 \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{(k_x+2)\times(k_x+2)} & * & * \\ \mathbf{0}_{1\times(k_x+2)} & \frac{1}{n-1}\left[\mathrm{tr}(G_n'J_nG_n) + \mathrm{tr}((J_nG_n)^2)\right] & * \\ \mathbf{0}_{1\times(k_x+2)} & \frac{1}{\sigma_0^2(n-1)}\mathrm{tr}(J_nG_n) & \frac{1}{2\sigma_0^4} \end{pmatrix} + O\left(\frac{1}{T}\right), \tag{10}$$

which is nonsingular if $\mathrm{E}\mathcal{H}_{nT}$ is nonsingular or $1/(n-1)[\mathrm{tr}(G_n'J_nG_n) + \mathrm{tr}((J_nG_n)^2) - (2\mathrm{tr}^2(J_nG_n))/(n-1)]$ is nonzero (see Section D.2 of Appendix D for the proof). Also, its rank does not change in a small neighborhood of θ_0 (see Yu et al., 2008, eqn. (41)). When $\lim_{T\to\infty}\mathrm{E}\mathcal{H}_{nT}$ is nonsingular, the global identification of the parameters is shown in Theorem 1, which follows. When $\lim_{T\to\infty}\mathrm{E}\mathcal{H}_{nT}$ is singular, global identification can still be obtained from Theorem 2 (also in this section) via a condition on the variance structure of the model. Denote $\sigma_n^2(\lambda) = (\sigma_0^2/(n-1))\mathrm{tr}(S_n'^{-1}S_n'(\lambda)J_nS_n(\lambda)S_n^{-1})$.

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

THEOREM 2. *If $\lim_{n\to\infty}\left(\frac{1}{n}\ln\left|\sigma_0^2S_n'^{-1}S_n^{-1}\right| - \frac{1}{n}\ln\left|\sigma_n^2(\lambda)S_n'^{-1}(\lambda)S_n^{-1}(\lambda)\right|\right) \neq 0$ for $\lambda \neq \lambda_0$,[10] under Assumptions 1–7, θ_0 is globally identified and $\hat{\theta}_{nT} \xrightarrow{p} \theta_0$.*

3.2. Asymptotic Distribution

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

$$J_n\tilde{Z}_{nt} = J_n\tilde{Z}_{nt}^{(u)} - (J_n\bar{U}_{nT,-1},\ J_nW_n\bar{U}_{nT,-1},\ \mathbf{0}_{n\times k_x}), \tag{11}$$

where $\tilde{Z}_{nt}^{(u)} = ((\overset{\approx}{\mathcal{X}}_{n,t-1} + U_{n,t-1}), (W_n\overset{\approx}{\mathcal{X}}_{n,t-1} + W_nU_{n,t-1}), \tilde{X}_{nt})$ with $\overset{\approx}{\mathcal{X}}_{n,t-1} = \mathcal{X}_{n,t-1} - \bar{\mathcal{X}}_{nT,-1}$. Hence, $J_n\tilde{Z}_{nt}$ has two components: one is $J_n\tilde{Z}_{nt}^{(u)}$, uncorrelated

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

with V_{nt}; the other is $-(J_n\bar{U}_{nT,-1},\ J_nW_n\bar{U}_{nT,-1},\ \mathbf{0}_{n\times k_x})$, correlated with V_{nt} when $t \le T-1$. Therefore, from (8), the score can be decomposed into two parts such that $\frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta} = \frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L^{(u)}_{n,T}(\theta_0)}{\partial\theta} - \Delta_{nT}$ where

$$\frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L^{(u)}_{n,T}(\theta_0)}{\partial\theta}$$

$$= \begin{pmatrix} \frac{1}{\sigma_0^2}\frac{1}{\sqrt{(n-1)T}}\sum_{t=1}^{T}\tilde{Z}^{(u)\prime}_{nt}J_nV_{nt} \\ \frac{1}{\sigma_0^2}\frac{1}{\sqrt{(n-1)T}}\sum_{t=1}^{T}(G_n\tilde{Z}^{(u)}_{nt}\delta_0)'J_nV_{nt} + \frac{1}{\sigma_0^2}\frac{1}{\sqrt{(n-1)T}} \\ \qquad\qquad \sum_{t=1}^{T}(V'_{nt}G'_nJ_nV_{nt} - \sigma_0^2\mathrm{tr}J_nG_n) \\ \frac{1}{2\sigma_0^4}\frac{1}{\sqrt{(n-1)T}}\sum_{t=1}^{T}(V'_{nt}J_nV_{nt} - (n-1)\sigma_0^2) \end{pmatrix}, \tag{12}$$

Δ_{nT}

$$= \sqrt{\frac{n-1}{T}}\begin{pmatrix} \frac{1}{\sigma_0^2}\frac{T}{n-1}(J_n\bar{U}_{nT,-1},\ J_nW_n\bar{U}_{nT,-1}, \mathbf{0}_{n\times k_x})'\bar{V}_{nT} \\ \frac{1}{\sigma_0^2}\frac{T}{n-1}(J_nG_n(\bar{U}_{nT,-1}, W_n\bar{U}_{nT,-1},\ \mathbf{0}_{n\times k_x})\delta_0)'\bar{V}_{nT} \\ \qquad +\frac{1}{\sigma_0^2}\frac{T}{n-1}\bar{V}'_{nT}G'_nJ_n\bar{V}_{nT} \\ \frac{1}{2\sigma_0^4}\frac{T}{n-1}\bar{V}'_{nT}J_n\bar{V}_{nT} \end{pmatrix}. \tag{13}$$

Similar to Yu et al. (2008), the variance matrix of $\frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L^{(u)}_{n,T}(\theta_0)}{\partial\theta}$ is equal to

$$\mathrm{E}\left(\frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L^{(u)}_{n,T}(\theta_0)}{\partial\theta}\cdot\frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L^{(u)}_{n,T}(\theta_0)}{\partial\theta'}\right)$$

$$= \Sigma_{\theta_0,nT} + \Omega_{\theta_0,n} + O\left(T^{-1}\right), \tag{14}$$

where $\Sigma_{\theta_0,nT}$ is in (10) and the following matrix is a symmetric matrix with μ_4 being the fourth moment of v_{it}:

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

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

When V_{nt} are normally distributed, $\Omega_{\theta_0,n}=0$ because $\mu_4-3\sigma_0^4=0$ for a normal distribution. Denote $\Sigma_{\theta_0}=\lim_{T\to\infty}\Sigma_{\theta_0,nT}$ and $\Omega_{\theta_0}=\lim_{T\to\infty}\Omega_{\theta_0,n}$; then

$$\lim_{T\to\infty}\mathrm{E}\left(\frac{1}{\sqrt{(n-1)T}}\frac{\partial\ln L_{n,T}^{(u)}(\theta_0)}{\partial\theta}\cdot\frac{1}{\sqrt{(n-1)T}}\frac{\partial\ln L_{n,T}^{(u)}(\theta_0)}{\partial\theta'}\right)=\Sigma_{\theta_0}+\Omega_{\theta_0}. \quad \textbf{(15)}$$

The asymptotic distribution of $\frac{1}{\sqrt{(n-1)T}}\frac{\partial\ln L_{n,T}^{(u)}(\theta_0)}{\partial\theta}$ can be derived from the central limit theorem for martingale difference arrays (Lemma 3 in Appendix B). For the term Δ_{nT}, from Lemmas 1 and 2 in Appendix B, $\Delta_{nT}=\sqrt{(n-1)/T}a_{\theta_0,n}+O(\sqrt{(n-1)/T^3})+O_p(\frac{1}{\sqrt{T}})$ where $a_{\theta_0,n}=O(1)$ and

$$a_{\theta,n}=\begin{pmatrix}\frac{1}{n-1}\mathrm{tr}\left(\left(J_n\sum_{h=0}^{\infty}A_n^h(\theta)\right)S_n^{-1}(\lambda)\right)\\ \frac{1}{n-1}\mathrm{tr}\left(W_n\left(J_n\sum_{h=0}^{\infty}A_n^h(\theta)\right)S_n^{-1}(\lambda)\right)\\ \mathbf{0}_{k_x\times 1}\\ \frac{1}{n-1}\gamma\,\mathrm{tr}(G_n(\lambda)\left(J_n\sum_{h=0}^{\infty}A_n^h(\theta)\right)S_n^{-1}(\lambda))\\ +\frac{1}{n-1}\rho\mathrm{tr}(G_n(\lambda)W_n\left(J_n\sum_{h=0}^{\infty}A_n^h(\theta)\right)S_n^{-1}(\lambda))+\frac{1}{n-1}\mathrm{tr}(J_nG_n(\lambda))\\ \frac{1}{2\sigma^2}\end{pmatrix}. \quad \textbf{(16)}$$

To obtain the asymptotic distribution of the score, we need the following additional assumption.

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

Assumption 8 is a condition for the nonsingularity of the information matrix Σ_{θ_0}. When $\lim_{T\to\infty}\mathrm{E}\mathcal{H}_{nT}$ is singular;[11] as long as we have $\lim_{n\to\infty}1/(n-1)\,\mathrm{tr}[(G_n'J_nG_n)+\mathrm{tr}((J_nG_n)^2)-2\mathrm{tr}^2(J_nG_n)/(n-1)]\neq 0$, the information matrix Σ_{θ_0} is still nonsingular (see Section D.2 of Appendix D).

CLAIM 2. *Under Assumptions 1–8,* $\frac{1}{\sqrt{(n-1)T}}\frac{\partial\ln L_{n,T}(\theta_0)}{\partial\theta}+\Delta_{nT}\xrightarrow{d}N(0,\Sigma_{\theta_0}+\Omega_{\theta_0})$. *When* $\{v_{it}\}$, $i=1,2,\ldots,n$ *and* $t=1,2,\ldots,T$, *are normal,* $\frac{1}{\sqrt{(n-1)T}}\frac{\partial\ln L_{n,T}(\theta_0)}{\partial\theta}+\Delta_{nT}\xrightarrow{d}N(0,\Sigma_{\theta_0})$.

Also, under Assumptions 1–8, we have $\frac{1}{(n-1)T}\frac{\partial^2\ln L_{n,T}(\theta)}{\partial\theta\partial\theta'}-\frac{1}{(n-1)T}\frac{\partial^2\ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'}=\|\theta-\theta_0\|\cdot O_p(1)$ and $\frac{1}{(n-1)T}\frac{\partial^2\ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'}-\frac{\partial^2 Q_{n,T}(\theta_0)}{\partial\theta\partial\theta'}=O_p\left(\frac{1}{\sqrt{(n-1)T}}\right)$ similar to (38) and (39) in Yu et al. (2008). Combined with Claim 2, we have the following theorem for the asymptotic distribution of $\hat{\theta}_{nT}$.

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

THEOREM 3. *Under Assumptions 1–8,*

$$\sqrt{(n-1)T}(\hat{\theta}_{nT}-\theta_0)+\sqrt{\frac{n-1}{T}}b_{\theta_0,nT}+O_p\left(\max\left(\sqrt{\frac{n-1}{T^3}},\sqrt{\frac{1}{T}}\right)\right)$$

$$\xrightarrow{d} N(0,\Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0}+\Omega_{\theta_0})\Sigma_{\theta_0}^{-1}), \tag{17}$$

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

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

From (17), the QML estimator has the bias $-(1/T)b_{\theta_0,nT}$, and the confidence interval is not properly centered when $\frac{n}{T}\to k$ where $0<k<\infty$. Furthermore, when $n/T\to\infty$, the presence of $b_{\theta_0,nT}$ causes $\hat{\theta}_{nT}$ to have a degenerate distribution. An analytical bias reduction procedure will correct the bias of the estimate. Define the bias corrected estimator as $\hat{\theta}_{nT}^1=\hat{\theta}_{nT}-\frac{\hat{B}_{nT}}{T}$ where $\hat{B}_{nT}=\left[-\Sigma_{\theta,nT}^{-1}\cdot a_{\theta,n}\right]\big|_{\theta=\hat{\theta}_{nT}}$. We will show that when $n/T^3\to 0$, $\hat{\theta}_{nT}^1$ is $\sqrt{nT}$ consistent and asymptotically centered normal even when $n/T\to\infty$. To show our result for the bias corrected estimator, we need the following additional assumption.

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

Assumption 9 can be justified through Lemma 5 in Appendix B. Our result for the bias corrected estimator is as follows.

THEOREM 4. *Under Assumptions 1–9, if* $\frac{n}{T^3}\to 0$, $\sqrt{(n-1)T}(\hat{\theta}_{nT}^1-\theta_0)\xrightarrow{d} N(0,\Sigma_{\theta_0}^{-1}+\Sigma_{\theta_0}^{-1}\Omega_{\theta_0}\Sigma_{\theta_0}^{-1})$.

3.3. Monte Carlo Results

We conduct a small Monte Carlo experiment to evaluate the performance of our ML estimator and the bias corrected estimator for the transformation approach. We generate samples from (1):

$$Y_{nt}=\lambda_0W_nY_{nt}+\gamma_0Y_{n,t-1}+\rho_0W_nY_{n,t-1}+X_{nt}\beta_0+\mathbf{c}_{n0}+\alpha_{t0}l_n+V_{nt},$$
$$t=1,2,\ldots,T,$$

using $\theta_0^a=(0.2,0.2,1,0.2,1)'$ and $\theta_0^b=(0.3,0.3,1,0.3,1)'$ where $\theta_0=(\gamma_0,\rho_0,\beta_0',\lambda_0,\sigma_0^2)'$ and X_{nt}, $\mathbf{c}_{n0}$, $\alpha_{T0}=(\alpha_1,\alpha_2,\ldots,\alpha_T)$, and V_{nt} are generated from independent standard normal distributions[12] and the spatial weights matrix we use

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

is a rook matrix.[13] We use $T = 10, 50$, and $n = 16, 49$. For each set of generated sample observations, we calculate the ML estimator $\hat{\theta}_{nT}$ and evaluate the bias $\hat{\theta}_{nT} - \theta_0$; we then construct the bias corrected estimator $\hat{\theta}^1_{nT}$ and evaluate the bias $\hat{\theta}^1_{nT} - \theta_0$. We do this 1,000 times to see if the bias is reduced on average by using the analytical bias correction procedure, i.e., to compare $\frac{1}{1,000}\sum_{i=1}^{1,000}(\hat{\theta}_{nT} - \theta_0)_i$ with $\frac{1}{1,000}\sum_{i=1}^{1,000}(\hat{\theta}^1_{nT} - \theta_0)_i$. We also compare not only the empirical standard error of the estimator but also the average of the estimated standard error of the estimator.[14] With two different values of θ_0 for each n and T, finite-sample properties of both estimators are summarized in Tables 1–3, where Table 1 is for the magnitude of biases, Table 2 is for the standard errors of the estimator, and Table 3 is for the size under 5% significance level for the hypothesis that a coefficient is at its true value.

We can see that both estimators have some biases but the bias corrected estimators reduce those biases. This is consistent with our asymptotic analysis, because the bias corrected estimator 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 estimator, as can be seen from Table 2. Additionally, the estimated variances and the empirical variances are close, and so the estimated variances constructed from the Hessian matrix may be reliable. From Table 3, we can see that the size distortion is much reduced after bias correction.

For different cases of n and T, we can see that for each given n, when T is larger, the biases of the two sets of estimators will be smaller and the variance will be smaller; for each given T, when n is larger, the biases of the two sets of

TABLE 1. Biases of transformation approach estimators

				Bias of $\hat{\theta}_{nT}$ (1st line) and $\hat{\theta}^1_{nT}$ (2nd line)				
Case	T	n	θ_0	γ	ρ	β	λ	σ^2
(1)	10	49	θ_0^a	−0.0627	−0.0045	−0.0101	−0.0013	−0.1159
				−0.0038	−0.0023	−0.0026	0.0004	−0.0276
(2)	10	49	θ_0^b	−0.0697	−0.0093	−0.0132	−0.0086	−0.1183
				−0.0037	−0.0001	−0.0023	−0.0029	−0.0307
(3)	50	49	θ_0^a	−0.0119	−0.0013	0.0001	0.0000	−0.0217
				0.0000	−0.0013	0.0004	0.0001	−0.0022
(4)	50	49	θ_0^b	−0.0131	−0.0017	−0.0001	−0.0006	−0.0218
				−0.0000	−0.0013	0.0004	−0.0004	−0.0023
(5)	50	16	θ_0^a	−0.0121	−0.0032	−0.0013	−0.0018	−0.0247
				−0.0002	−0.0033	−0.0009	−0.0018	−0.0052
(6)	50	16	θ_0^b	−0.0134	−0.0043	−0.0013	−0.0017	−0.0247
				−0.0004	−0.0042	−0.0008	−0.0015	−0.0052

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

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

TABLE 2. Standard errors of transformation approach estimators

				Empirical S.E. of $\hat{\theta}_{nT}$ (1st line) and $\hat{\theta}^1_{nT}$ (2nd line)				
Case	T	n	θ_0	γ	ρ	β	λ	σ^2
(1)	10	49	θ_0^a	0.0351	0.0666	0.0482	0.0539	0.0610
				0.0373	0.0707	0.0484	0.0540	0.0671
(2)	10	49	θ_0^b	0.0349	0.0643	0.0484	0.0514	0.0610
				0.0375	0.0711	0.0488	0.0523	0.0671
(3)	50	49	θ_0^a	0.0149	0.0270	0.0203	0.0228	0.0297
				0.0150	0.0273	0.0203	0.0228	0.0303
(4)	50	49	θ_0^b	0.0147	0.0252	0.0203	0.0215	0.0298
				0.0149	0.0256	0.0203	0.0215	0.0304
(5)	50	16	θ_0^a	0.0248	0.0504	0.0372	0.0412	0.0519
				0.0250	0.0510	0.0372	0.0412	0.0529
(6)	50	16	θ_0^b	0.0239	0.0484	0.0373	0.0403	0.0518
				0.0241	0.0492	0.0372	0.0403	0.0529
				Estimated S.E. of $\hat{\theta}_{nT}$ (1st line) and $\hat{\theta}^1_{nT}$ (2nd line)				
Case	T	n	θ_0	γ	ρ	β	λ	σ^2
(1)	10	49	θ_0^a	0.0324	0.0621	0.0456	0.0504	0.0572
				0.0340	0.0654	0.0478	0.0510	0.0629
(2)	10	49	θ_0^b	0.0323	0.0602	0.0457	0.0487	0.0571
				0.0339	0.0639	0.0479	0.0491	0.0628
(3)	50	49	θ_0^a	0.0142	0.0273	0.0204	0.0224	0.0283
				0.0143	0.0276	0.0206	0.0225	0.0288
(4)	50	49	θ_0^b	0.0139	0.0258	0.0204	0.0214	0.0283
				0.0141	0.0262	0.0206	0.0215	0.0289
(5)	50	16	θ_0^a	0.0254	0.0508	0.0366	0.0422	0.0504
				0.0256	0.0513	0.0369	0.0423	0.0514
(6)	50	16	θ_0^b	0.0247	0.0491	0.0366	0.0412	0.0504
				0.0249	0.0497	0.0369	0.0413	0.0514

estimators will not be smaller but the variance will be smaller. This is consistent with our theoretical prediction, because the bias is of the order $O(T^{-1})$ and the variance of the estimator is of the order $O((nT)^{-1})$. Also, the size distortion is much reduced when T is larger.

4. AN ALTERNATIVE APPROACH: DIRECT ESTIMATION

An alternative approach is to estimate both the individual effects and the time effects directly.[15] As is shown subsequently, the QML estimators from this approach will yield bias of the order $O(\max(n^{-1}, T^{-1}))$.

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

TABLE 3. Sizes of transformation approach estimators

				Sizes of $\hat{\theta}_{nT}$ (1st line) and $\hat{\theta}^1_{nT}$ (2nd line)				
Case	T	n	θ_0	γ	ρ	β	λ	σ^2
(1)	10	49	θ_0^a	0.4780	0.0710	0.0720	0.0640	0.5300
				0.0980	0.0780	0.0660	0.0680	0.1350
(2)	10	49	θ_0^b	0.5630	0.0670	0.0760	0.0610	0.5480
				0.0970	0.0890	0.0650	0.0710	0.1510
(3)	50	49	θ_0^a	0.1450	0.0520	0.0400	0.0550	0.1450
				0.0650	0.0540	0.0410	0.0540	0.0630
(4)	50	49	θ_0^b	0.1660	0.0420	0.0410	0.0550	0.1460
				0.0690	0.0490	0.0420	0.0560	0.0670
(5)	50	16	θ_0^a	0.0730	0.0410	0.0570	0.0410	0.1000
				0.0490	0.0460	0.0550	0.0410	0.0620
(6)	50	16	θ_0^b	0.0740	0.0470	0.0540	0.0480	0.1000
				0.0500	0.0530	0.0520	0.0470	0.0650

Note: $\theta_0^a = (0.2, 0.2, 1, 0.2, 1)$ and $\theta_0^b = (0.3, 0.3, 1, 0.3, 1)$. The sizes are obtained under 5% significance level.

4.1. Model, Likelihood Function, First-Order Condition, and Second-Order Condition

Denote $\boldsymbol{\alpha}_T = (\alpha_1, \alpha_2, \ldots, \alpha_T)$ as the time effects; the likelihood function of (1) is

$$\ln L^d_{n,T}(\theta, \mathbf{c}_n, \boldsymbol{\alpha}_T) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln\sigma^2 + T\ln|S_n(\lambda)| - \frac{1}{2\sigma^2}\sum_{t=1}^{T} V'_{nt}(\theta, \mathbf{c}_n, \boldsymbol{\alpha}_T)V_{nt}(\theta, \mathbf{c}_n, \boldsymbol{\alpha}_T), \quad \textbf{(18)}$$

where $V_{nt}(\theta, \mathbf{c}_n, \boldsymbol{\alpha}_T) = S_n(\lambda)Y_{nt} - Z_{nt}\delta - \mathbf{c}_n - \alpha_t l_n$. For (18), the first-order condition for α_t is $\frac{\partial \ln L^d(\theta, \mathbf{c}_n, \alpha_T)}{\partial \alpha_t} = (1/\sigma^2) l'_n V_{nt}(\theta, \mathbf{c}_n, \boldsymbol{\alpha}_T)$, and it implies $\hat{\alpha}_t(\theta, \mathbf{c}_n) = (1/n) l'_n (S_n(\lambda)Y_{nt} - Z_{nt}\delta - \mathbf{c}_n)$. It follows that $V_{nt}(\theta, \mathbf{c}_n, \hat{\boldsymbol{\alpha}}_T(\theta, \mathbf{c}_n)) = J_n[S_n(\lambda)Y_{nt} - Z_{nt}\delta - \mathbf{c}_n]$. Hence, the likelihood function with $\boldsymbol{\alpha}_T$ being concentrated out is

$$\ln L^d_{n,T}(\theta, \mathbf{c}_n) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln\sigma^2 + T\ln|S_n(\lambda)| - \frac{1}{2\sigma^2}\sum_{t=1}^{T}[V_{nt}(\theta, \mathbf{c}_n, \hat{\boldsymbol{\alpha}}_T(\theta, \mathbf{c}_n))]'[V_{nt}(\theta, \mathbf{c}_n, \hat{\boldsymbol{\alpha}}_T(\theta, \mathbf{c}_n))]. \quad \textbf{(19)}$$

Compared to (6), (19) is the concentrated likelihood function for the true data generating process where the time effects are concentrated out, whereas (6) is the

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

likelihood function for the transformed data. For (19), the first-order derivative of $\mathbf{c}_n$ is $\frac{\partial \ln L_{nT}^d(\theta,\mathbf{c}_n)}{\partial \mathbf{c}_n} = \frac{1}{\sigma^2}\sum_{t=1}^T V_{nt}(\theta,\mathbf{c}_n,\hat{\boldsymbol{\alpha}}_T(\theta,\mathbf{c}_n))$, and it implies $J_n\hat{\mathbf{c}}_n(\theta) = J_n\left[S_n(\lambda)\bar{Y}_{nT} - \bar{Z}_{nT}\delta\right]$. The likelihood function with both $\mathbf{c}_n$ and $\boldsymbol{\alpha}_T$ concentrated out is

$$\ln L_{n,T}^d(\theta) = -\frac{nT}{2}\ln 2\pi - \frac{nT}{2}\ln\sigma^2 + T\ln|S_n(\lambda)| - \frac{1}{2\sigma^2}\sum_{t=1}^T \tilde{V}_{nt}'(\theta)J_n\tilde{V}_{nt}(\theta). \tag{20}$$

For (20), the first-order derivatives are (E.1) and (E.2) in Section E.1 of Appendix E. Hence,

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^d(\theta_0)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^T \tilde{Z}_{nt}'J_n\tilde{V}_{nt} \\ \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^T (G_n\tilde{Z}_{nt}\delta_0)'J_n\tilde{V}_{nt} + \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^T \left(\tilde{V}_{nt}'G_n'J_n\tilde{V}_{nt} - \sigma_0^2\mathrm{tr}G_n\right) \\ \frac{1}{2\sigma_0^4}\frac{1}{\sqrt{nT}}\sum_{t=1}^T (\tilde{V}_{nt}'J_n\tilde{V}_{nt} - n\sigma_0^2) \end{pmatrix}, \tag{21}$$

which is a linear and quadratic form of $\tilde{V}_{nt}$. For the information matrix of (20), denote $\Sigma_{\theta_0,nT}^d = -\mathrm{E}\left(\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^d(\theta_0)}{\partial\theta\partial\theta'}\right)$ and $\mathcal{H}_{nT}^d = \frac{1}{nT}\sum_{t=1}^T(\tilde{Z}_{nt}, G_n\tilde{Z}_{nt}\delta_0)'J_n(\tilde{Z}_{nt}, G_n\tilde{Z}_{nt}\delta_0)$; we have[16]

$$\Sigma_{\theta_0,nT}^d = \frac{1}{\sigma_0^2}\begin{pmatrix} \mathrm{E}\mathcal{H}_{nT}^d & * \\ \mathbf{0}_{1\times(k_x+3)} & 0 \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{(k_x+2)\times(k_x+2)} & * & * \\ \mathbf{0}_{1\times(k_x+2)} & \frac{1}{n}\left[\mathrm{tr}(G_n'J_nG_n) + \mathrm{tr}(G_n^2)\right] & * \\ \mathbf{0}_{1\times(k_x+2)} & \frac{1}{\sigma_0^2 n}\mathrm{tr}(J_nG_n) & \frac{1}{2\sigma_0^4} \end{pmatrix}$$

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

Here, $\mathcal{H}_{nT}^d$ is the covariance matrix of regressors of the reduced form (C.1) after demeaning from both the time dimension and the cross-sectional dimension (see Section C.1 of Appendix C for the reduced form).

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

4.2. Asymptotic Properties

Similarly to the transformation approach, we can obtain the asymptotic properties of the QML estimator. To do that, we need to strengthen Assumption 7 such that both n and T are large. This is so that the time dummies can be consistently estimated. However, in contrast with the transformation approach, we do not need the row normalization of the spatial weights matrix.

Assumption 1′. W_n is a nonstochastic spatial weights matrix.

Assumption 7′. n is an increasing function of T, and T goes to infinity.

THEOREM 5. *Under Assumptions 1′, 2–6, and 7′, if* $\lim_{T\to\infty} \mathrm{E}\mathcal{H}^d_{nT}$ *is nonsingular or*

$$\lim_{n\to\infty}\left(\frac{1}{n}\ln\left|\sigma_0^2 S_n'^{-1}S_n^{-1}\right| - \frac{1}{n}\ln\left|\sigma_n^2(\lambda)S_n'^{-1}(\lambda)S_n^{-1}(\lambda)\right|\right)\neq 0 \quad \textit{for } \lambda\neq\lambda_0,$$

θ_0 is globally identified and $\hat{\theta}^d_{nT}\xrightarrow{p}\theta_0$.

From (11) and (21), we have $\frac{1}{\sqrt{nT}}\frac{\partial \ln L^d_{n,T}(\theta_0)}{\partial\theta} = \frac{1}{\sqrt{nT}}\frac{\partial \ln L^{d(u)}_{n,T}(\theta_0)}{\partial\theta} - \Delta_{1,nT} - \Delta_{2,nT}$ where

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L^{d(u)}_{n,T}(\theta_0)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\tilde{Z}^{(u)\prime}_{nt}J_nV_{nt} \\ \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}(G_n\tilde{Z}^{(u)}_{nt}\delta_0)'J_nV_{nt} + \frac{1}{\sigma_0^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}(V'_{nt}G'_nJ_nV_{nt} - \sigma_0^2\mathrm{tr}J_nG_n) \\ \frac{1}{2\sigma_0^4}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}(V'_{nt}J_nV_{nt} - (n-1)\sigma_0^2) \end{pmatrix}, \tag{22}$$

$$\Delta_{1,nT} = \sqrt{\frac{n}{T}}\begin{pmatrix} \frac{1}{\sigma_0^2}\frac{T}{n}(J_n\bar{U}_{nT,-1},\ J_nW_n\bar{U}_{nT,-1},\mathbf{0}_{n\times k_x})'\bar{V}_{nT} \\ \frac{1}{\sigma_0^2}\frac{T}{n}(J_nG_n(\bar{U}_{nT,-1},W_n\bar{U}_{nT,-1},\ \mathbf{0}_{n\times k_x})\delta_0)'\bar{V}_{nT} \\ \quad +\frac{1}{\sigma_0^2}\frac{T}{n}\bar{V}'_{nT}G'_nJ_n\bar{V}_{nT} \\ \frac{1}{2\sigma_0^4}\frac{T}{n}\bar{V}'_{nT}J_n\bar{V}_{nT} \end{pmatrix}, \tag{23}$$

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

$$\Delta_{2,nT} = \begin{pmatrix} \mathbf{0}_{(k_x+2)\times 1} \\ \sqrt{\frac{T}{n}}(\mathrm{tr}G_n - \mathrm{tr}J_nG_n) \\ \sqrt{\frac{T}{n}}\frac{1}{2\sigma_0^2} \end{pmatrix} = \sqrt{\frac{T}{n}} \cdot a_{\theta_0,2}, \quad \text{and} \tag{24}$$

$$a_{\theta_0,2} = \left(\mathbf{0}_{1\times(k_x+2)}, \frac{1}{n} l_n' G_n l_n, \frac{1}{2\sigma_0^2} \right)'. \tag{25}$$

The asymptotic distribution of $\frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}^{d(u)}(\theta_0)}{\partial \theta}$ can be derived from Lemma 3. For the term $\Delta_{1,nT}$, from Lemma 2, $\Delta_{1,nT} = \sqrt{\frac{n}{T}} a_{\theta_0,n,1} + O(\sqrt{\frac{n}{T^3}}) + O_p(\frac{1}{\sqrt{T}})$ where

$$a_{\theta_0,n,1} = \begin{pmatrix} \frac{1}{n}\mathrm{tr}\left(\left(J_n \Sigma_{h=0}^{\infty} A_n^h\right) S_n^{-1}\right) \\ \frac{1}{n}\mathrm{tr}\left(W_n \left(J_n \Sigma_{h=0}^{\infty} A_n^h\right) S_n^{-1}\right) \\ \mathbf{0}_{k_x\times 1} \\ \frac{1}{n}\gamma_0 \mathrm{tr}\left(G_n \left(J_n \Sigma_{h=0}^{\infty} A_n^h\right) S_n^{-1}\right) + \frac{1}{n}\rho_0 \mathrm{tr}\left(G_n W_n \left(J_n \Sigma_{h=0}^{\infty} A_n^h\right) S_n^{-1}\right) \\ + \frac{1}{n}\mathrm{tr}(J_n G_n) \\ \frac{n-1}{n}\frac{1}{2\sigma_0^2} \end{pmatrix}. \tag{26}$$

The asymptotic distribution of the QML estimator is as follows.

THEOREM 6. *Under Assumptions 1′, 2–6, 7′, and 8,*

$$\sqrt{nT}(\hat{\theta}_{nT}^d - \theta_0) + \sqrt{\frac{n}{T}} b_{\theta_0,nT,1} + \sqrt{\frac{T}{n}} b_{\theta_0,nT,2} + O_p\left(\max\left(\sqrt{\frac{n}{T^3}}, \sqrt{\frac{T}{n^3}}, \sqrt{\frac{1}{T}}\right)\right) \xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0} + \Omega_{\theta_0})\Sigma_{\theta_0}^{-1}), \tag{27}$$

where $b_{\theta_0,nT,1} = (\Sigma_{\theta_0,nT}^d)^{-1} a_{\theta_0,n,1}$ *and* $b_{\theta_0,nT,2} = (\Sigma_{\theta_0,nT}^d)^{-1} a_{\theta_0,2}$ *are* $O(1)$ *and* $a_{\theta_0,n,1}$, $a_{\theta_0,2}$ *are defined in (26) and (25).*

When $\frac{n}{T} \to 0$, $n(\hat{\theta}_{nT}^d - \theta_0) + b_{\theta_0,nT,2} \xrightarrow{p} 0$.

When $\frac{n}{T} \to k < \infty$, $\sqrt{nT}(\hat{\theta}_{nT}^d - \theta_0) + \sqrt{k} b_{\theta_0,nT,1} + \sqrt{k^{-1}} b_{\theta_0,nT,2} \xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0} + \Omega_{\theta_0})\Sigma_{\theta_0}^{-1})$.

When $\frac{n}{T} \to \infty$, $T(\hat{\theta}_{nT}^d - \theta_0) + b_{\theta_0,nT,1} \xrightarrow{p} 0$.

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

From (27), the QML estimator has the bias $-(1/T)b_{\theta_0,nT,1}$ and $-(1/n)b_{\theta_0,nT,2}$, and the confidence interval is not centered when $\frac{n}{T} \to k$ where $0 < k < \infty$.[17] Furthermore, when T and n do not have the same order, the presence of $b_{\theta_0,nT,1}$ and $b_{\theta_0,nT,2}$ causes $\hat{\theta}^d_{nT}$ to have a degenerate distribution. An analytical bias reduction procedure will correct the bias $B_{1,nT} = -b_{\theta_0,nT,1}$ and $B_{2,nT} = -b_{\theta_0,nT,2}$ by constructing estimates $\hat{B}_{1,nT}$, $\hat{B}_{2,nT}$ and defining the bias corrected estimator as $\hat{\theta}^{d1}_{nT} = \hat{\theta}^d_{nT} - \frac{\hat{B}_{1,nT}}{T} - \frac{\hat{B}_{2,nT}}{n}$ where $\hat{B}_{1,nT} = \left[-(\Sigma^d_{\theta,nT})^{-1} \cdot a_{\theta,n,1}\right]\big|_{\theta=\hat{\theta}^d_{nT}}$ and $\hat{B}_{2,nT} = \left[-(\Sigma^d_{\theta,nT})^{-1} \cdot a_{\theta,2}\right]\big|_{\theta=\hat{\theta}^d_{nT}}$. We show that when $n/T^3 \to 0$ and $T/n^3 \to 0$, $\hat{\theta}^{d1}_{nT}$ is $\sqrt{nT}$ consistent and asymptotically centered normal.

THEOREM 7. *Under Assumptions 1′, 2–6, 7′, 8, and 9, if* $\frac{n}{T^3} \to 0$ *and* $\frac{T}{n^3} \to 0$, $\sqrt{nT}\left(\hat{\theta}^{d1}_{nT} - \theta_0\right) \xrightarrow{d} N\left(0, \Sigma^{-1}_{\theta_0} + \Sigma^{-1}_{\theta_0}\Omega_{\theta_0}\Sigma^{-1}_{\theta_0}\right)$.

By comparing Theorems 6 and 7 with Theorems 3 and 4, we see that both estimators are consistent and have the same limiting variances when the limiting distributions are not degenerate. For the transformed approach, the estimator has the bias of the order $O(T^{-1})$, and the bias correction requires $n/T^3 \to 0$. For the direct approach, the estimator has the bias of the order $O(\max(n^{-1}, T^{-1}))$, and the bias correction requires not only $n/T^3 \to 0$ but also $T/n^3 \to 0$. Hence, the transformed approach has an advantage over the direct approach especially when n is relatively small. But the direct approach also has its own merit in that we

TABLE 4. Biases of direct approach estimators

				Bias of $\hat{\theta}^d_{nT}$ (1st line) and $\hat{\theta}^{d1}_{nT}$ (2nd line)				
Case	T	n	θ_0	γ	ρ	β	λ	σ^2
(1)	10	49	θ^a_0	−0.0610	−0.0014	−0.0092	−0.0312	−0.1318
				−0.0035	0.0026	−0.0022	−0.0009	−0.0298
(2)	10	49	θ^b_0	−0.0669	−0.0014	−0.0117	−0.0391	−0.1327
				−0.0032	0.0115	−0.0015	−0.0008	−0.0333
(3)	50	49	θ^a_0	−0.0100	0.0051	0.0010	−0.0296	−0.0394
				0.0002	0.0001	0.0007	−0.0019	−0.0031
(4)	50	49	θ^b_0	−0.0099	0.0108	0.0016	−0.0297	−0.0379
				0.0002	0.0014	0.0007	−0.0007	−0.0033
(5)	50	16	θ^a_0	−0.0067	0.0115	−0.0030	−0.0973	−0.0842
				0.0012	0.0013	0.0006	−0.0199	−0.0130
(6)	50	16	θ^b_0	−0.0032	0.0272	−0.0008	−0.1038	−0.0785
				0.0026	0.0082	0.0014	−0.0213	−0.0117

Note: $\theta^a_0 = (0.2, 0.2, 1, 0.2, 1)$ and $\theta^b_0 = (0.3, 0.3, 1, 0.3, 1)$.

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

TABLE 5. Standard errors of direct approach estimators

				Empirical S.E. of $\hat{\theta}^d_{nT}$ (1st line) and $\hat{\theta}^{d1}_{nT}$ (2nd line)				
Case	T	n	θ_0	γ	ρ	β	λ	σ^2
(1)	10	49	θ^a_0	0.0350	0.0668	0.0482	0.0529	0.0600
				0.0373	0.0712	0.0484	0.0539	0.0670
(2)	10	49	θ^b_0	0.0347	0.0646	0.0484	0.0509	0.0601
				0.0376	0.0801	0.0490	0.0533	0.0669
(3)	50	49	θ^a_0	0.0149	0.0270	0.0203	0.0225	0.0292
				0.0150	0.0273	0.0203	0.0228	0.0303
(4)	50	49	θ^b_0	0.0147	0.0252	0.0203	0.0209	0.0293
				0.0149	0.0258	0.0203	0.0212	0.0304
(5)	50	16	θ^a_0	0.0249	0.0499	0.0375	0.0376	0.0489
				0.0251	0.0508	0.0373	0.0394	0.0525
(6)	50	16	θ^b_0	0.0238	0.0483	0.0375	0.0381	0.0492
				0.0241	0.0504	0.0373	0.0399	0.0525
				Estimated S.E. of $\hat{\theta}^d_{nT}$ (1st line) and $\hat{\theta}^{d1}_{nT}$ (2nd line)				
Case	T	n	θ_0	γ	ρ	β	λ	σ^2
(1)	10	49	θ^a_0	0.0321	0.0615	0.0452	0.0488	0.0556
				0.0340	0.0653	0.0478	0.0491	0.0622
(2)	10	49	θ^b_0	0.0320	0.0596	0.0453	0.0471	0.0558
				0.0338	0.0637	0.0478	0.0469	0.0623
(3)	50	49	θ^a_0	0.0141	0.0270	0.0202	0.0217	0.0275
				0.0143	0.0276	0.0206	0.0217	0.0286
(4)	50	49	θ^b_0	0.0138	0.0255	0.0203	0.0208	0.0277
				0.0141	0.0260	0.0206	0.0206	0.0287
(5)	50	16	θ^a_0	0.0246	0.0488	0.0354	0.0376	0.0458
				0.0255	0.0507	0.0368	0.0375	0.0495
(6)	50	16	θ^b_0	0.0239	0.0471	0.0355	0.0366	0.0463
				0.0248	0.0489	0.0368	0.0360	0.0498

do not need the weights matrix to be row normalized. We conduct a small Monte Carlo experiment to compare the two estimators where we use the same simulated data. Table 4 is the counterpart of Table 1, Table 5 is the counterpart of Table 2, and Table 6 is the counterpart of Table 3. From Tables 4 and 6, for biases and sizes, the performance of the bias adjusted estimates can be much better than the performance of the unadjusted estimates. By comparing Tables 4–6 with Tables 1–3, we can see that the biases of estimates from the transformation approach are smaller than those of the direct approach; also, for the sizes of the estimates, the direct approach has distorted sizes compared to the transformation approach when

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

TABLE 6. Sizes of direct approach estimators

				Sizes of $\hat{\theta}^d_{nT}$ (1st line) and $\hat{\theta}^{d1}_{nT}$ (2nd line)				
Case	T	n	θ_0	γ	ρ	β	λ	σ^2
(1)	10	49	θ_0^a	0.4670	0.0760	0.0700	0.1060	0.6440
				0.1010	0.0850	0.0680	0.0810	0.1600
(2)	10	49	θ_0^b	0.5420	0.0680	0.0770	0.1400	0.6410
				0.0990	0.1160	0.0690	0.0860	0.1650
(3)	50	49	θ_0^a	0.1280	0.0540	0.0450	0.2740	0.3210
				0.0670	0.0550	0.0410	0.0670	0.0750
(4)	50	49	θ_0^b	0.1220	0.0790	0.0460	0.2960	0.3020
				0.0710	0.0490	0.0430	0.0660	0.0760
(5)	50	16	θ_0^a	0.0630	0.0680	0.0670	0.7370	0.4640
				0.0530	0.0530	0.0600	0.0990	0.1080
(6)	50	16	θ_0^b	0.0520	0.0980	0.0630	0.7950	0.4140
				0.0540	0.0720	0.0610	0.1040	0.1030

Note: $\theta_0^a = (0.2, 0.2, 1, 0.2, 1)$ and $\theta_0^b = (0.3, 0.3, 1, 0.3, 1)$. The sizes are obtained under 5% significance level.

n is relatively small. There is not much difference in the corresponding variances of the estimates of the two approaches.

5. CONCLUSION

This paper establishes asymptotic properties of quasi–maximum likelihood estimators for spatial dynamic panel data with both individual effects and time effects when both the number of individuals n and the number of time periods T can be large. A possible approach is to estimate both effects directly. But instead of direct estimation, we propose a data transformation approach so that the magnitude of the bias is only of the order $O(T^{-1})$ rather than $O(\max(T^{-1}, n^{-1}))$ in the direct estimation approach. When n is asymptotically proportional to T, the estimator is $\sqrt{nT}$ consistent and asymptotically normal, but the limit distribution is not centered around 0; when $n/T \to 0$, the estimator is $\sqrt{nT}$ consistent and asymptotically centered normal; when $n/T \to \infty$, the estimator is consistent with rate T and has a degenerate limit distribution. We also propose a bias correction for our estimator. We show that when $n^{1/3}/T \to 0$, the correction will asymptotically eliminate the bias and yield a centered confidence interval. This transformation approach has an advantage over the direct approach especially when n is relatively small. When n is relatively small, the estimates of the transformed approach have faster rates of convergence than those of the direct estimates. The direct estimates have a degenerate limit distribution, but the transformed estimates are properly centered and are asymptotically normal.

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

NOTES

1. Several notable weights matrices are (a) the contiguity matrix in regional science constructed and by boundary of neighboring regions (Anselin, 1988); (b) the input-output matrix in industrial organization (Conley and Dupor, 2003); (c) the social interaction matrix (Case, 1991), among others.

2. Because of the presence of fixed individual and time effects, the X_{nt} will not include time invariant or individual invariant regressors.

3. We note that the likelihood function for Y_{nt}^* does not necessarily have a conditional likelihood function interpretation. One may suspect that $\bar{y}_{.t} = (1/n)l_n' Y_{nt}$ would be a sufficient statistic for α_{t0}. But that is not the case in general when $\lambda_0 \neq 0$. The reason is that, by $(1/n)l_n' Y_{nt} = \lambda_0 (1/n) l_n' W_n Y_{nt} + (1/n) l_n' Z_{nt}\delta_0 + (1/n) l_n' \mathbf{c}_{n0} + \alpha_{t0} + (1/n) l_n' V_{nt}$, $(1/n) l_n' W_n Y_{nt}$ is not a function of $(1/n) l_n' Y_{nt}$ in general as $l_n' W_n$ is not proportional to l_n'.

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

5. For notational purposes, we define for any $n \times 1$ vector at period t, Υ_{nt}; we have $\ddot{\Upsilon}_{nt} = \Upsilon_{nt} - \bar{\Upsilon}_{nT}$ and $\overset{\approx}{\Upsilon}_{n,t-1} = \Upsilon_{n,t-1} - \bar{\Upsilon}_{nT,-1}$ for $t = 1, 2, \ldots, T$ where $\bar{\Upsilon}_{nT} = \frac{1}{T}\sum_{t=1}^{T} \Upsilon_{nt}$ and $\bar{\Upsilon}_{nT,-1} = \frac{1}{T}\sum_{t=1}^{T} \Upsilon_{n,t-1}$.

6. The attached term can be regarded as an adjustment by the transformation approach to eliminate the bias of the order $O(n^{-1})$ that will appear in the direct approach.

7. If X_{nt} is allowed to be stochastic and unbounded, appropriate moment conditions can be imposed instead. Also, X_{nt} may not include any time effects such as time trend, because the time dummies will absorb any time specific variables.

8. We say a (sequence of $n \times n$) matrix P_n is UB 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.

9. Assumptions 5 and 6 have effectively ruled out some cases and, hence, imposed limited dependence across units or time series. For example, if $\lambda_{0n} = 1 - 1/n$ under $n \to \infty$, it is a near unit root case for a cross-sectional spatial autoregressive model, and S_n^{-1} will not be UB (see Lee and Yu, 2007). For a spatial dynamic panel model, if $\lambda_0 + \rho_0 + \gamma_0 = 1$, Y_{nt} might have deterministic trends and also a nonstationary stochastic component (for details, see Yu et al., 2007; Yu and Lee, 2007). Assumption 5 rules out the near unit root case, and Assumption 6 rules out the nonstationary case.

10. When n is finite, the condition is $\frac{1}{n}\ln|\sigma_0^2 S_n'^{-1} S_n^{-1}| - \frac{1}{n}\ln|\sigma_n^2(\lambda) S_n'^{-1}(\lambda) S_n^{-1}(\lambda)| + \left(\frac{1}{n}\ln\frac{\sigma_0^2}{(1-\lambda_0)^2} - \frac{1}{n}\ln\frac{\sigma_n^2(\lambda)}{(1-\lambda)^2}\right) \neq 0$ for $\lambda \neq \lambda_0$. See Section D.4 of Appendix D.

11. The $\lim_{T\to\infty} \mathrm{E}\mathcal{H}_{nT}$ can be singular if, e.g., $\delta_0 = \mathbf{0}$.

12. We generate 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 generated the data with a much longer history $1,000 + T$, and the results are similar.

13. We use the rook matrix based on an r 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 empirically observed regional structures in spatial econometrics are made up of regions with connectivity close to the range of the rook tessellation.

14. The empirical variance is the diagonal elements of $\frac{1}{1,000}\sum_{i=1}^{1,000}[(\hat{\theta}_{nT} - mean(\hat{\theta}_{nT}))(\hat{\theta}_{nT} - mean(\hat{\theta}_{nT}))']_i$, and the estimated variance is the diagonal elements of $\frac{1}{1,000}\sum_{i=1}^{1,000}\left(\Sigma^{-1}_{\hat{\theta}_{nT},nT}\right)_i$.

15. Because components of $\mathbf{c}_n$ and $\boldsymbol{\alpha}_T$ appear additively, a convenient location normalization is to set the sum of the components of $\mathbf{c}_n$ to zero, i.e., the elements of $\mathbf{c}_n$ have a zero (empirical) mean.

16. The $\mathcal{H}_{nT}^d$ differs from $\mathcal{H}_{nT}$ in the transformation approach only in the division by n in $\mathcal{H}_{nT}^d$ instead of $(n-1)$ in $\mathcal{H}_{nT}$.

17. In Hahn and Moon (2006) with a simple AR(1) model where the direct approach is used, the inclusion of time effects will not change the bias of the estimate of the dynamic coefficient, and the

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

bias correction procedure is the same as that of Hahn and Kuersteiner (2002) where only individual effects are included. As Hahn and Moon (2006) stated, their result is limited only to a linear model. From Theorem 6, we can see that the time effects will introduce additional $O(n^{-1})$ bias when we have spatial effects in the model.

REFERENCES

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

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

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

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

Case, A.C. (1991) Spatial patterns in household demand. *Econometrica* 59, 953–965.

Conley, T.G. & W. Dupor (2003) A spatial analysis of sectoral complementarity. *Journal of Political Economy* 111, 311–352.

Ertur, C. & W. Koch (2007) Growth, technological interdependence and spatial externalities: Theory and evidence. *Journal of Applied Econometrics* 22, 1033–1062.

Foote, C.L. (2007) Space and Time in Macroeconomic Panel Data: Young Workers and State-Level Unemployment Revisited. Working paper 07–10, Federal Reserve Bank of Boston.

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

Hahn, J. & H.R. Moon (2006) Reducing bias of MLE in a dynamic panel model. *Econometric Theory* 22, 499–512.

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

Hamilton, J.D. (1994) *Time Series Analysis*. Princeton University Press.

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

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

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

Kelejian, H.H. & I.R. Prucha (2007) Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances. Working paper, University of Maryland.

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. & J. Yu (2007) Near Unit Root in the Spatial Autoregressive Model. Working paper, Ohio State University.

Lin, X. & L.F. Lee (2005) GMM Estimation of Spatial Autoregressive Models with Unknown Heteroskedasticity. Working paper, Ohio State University.

Nerlove, M. (1971) A note on error components models. *Econometrica* 39, 383–396.

Ord, J.K. (1975) Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70, 120–126.

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

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

Theil, H. (1971) *Principles of Econometrics*. Wiley.

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

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

Yu, J., R. de Jong, & L.F. Lee (2007) Quasi-Maximum Likelihood Estimators for Spatial Dynamic Panel Data with Fixed Effects when both n and T Are Large: A Nonstationary Case. Working paper, Ohio State University.

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

Yu, J. & L.F. Lee (2007) Estimation of Unit Root Spatial Dynamic Panel Data Models. Working paper, Ohio State University.

APPENDIX A: Some Notes

A.1. Data Transformation. Let $(F_{n,n-1}, l_n/\sqrt{n})$ be the orthonormal matrix of $J_n = I_n - (1/n)l_n l_n'$ where $F_{n,n-1}$ corresponds to the eigenvalues of ones and $l_n/\sqrt{n}$ corresponds to the eigenvalue zero. Thus,

$$J_n F_{n,n-1} = F_{n,n-1}, \qquad F_{n,n-1}' F_{n,n-1} = I_{n-1}, \qquad J_n l_n = \mathbf{0},$$

$$F_{n,n-1}' l_n = \mathbf{0}, \quad F_{n,n-1} F_{n,n-1}' + \frac{1}{n} l_n l_n' = I_n, \qquad F_{n,n-1} F_{n,n-1}' = J_n. \tag{A.1}$$

The transformation of Y_{nt} to Y_{nt}^*, where $Y_{nt}^* = F_{n,n-1}' Y_{nt}$ is a vector with dimension $n-1$, gives

$$Y_{nt}^* = \lambda_0 F_{n,n-1}' W_n Y_{nt} + \gamma_0 Y_{n,t-1}^* + \rho_0 F_{n,n-1}' W_n Y_{n,t-1} + X_{nt}^* \beta_0 + \mathbf{c}_{n0}^* + V_{nt}^*,$$

where $Y_{nt}^* = F_{n,n-1}' Y_{nt}$, $X_{nt}^* = F_{n,n-1}' X_{nt}$, $\mathbf{c}_{n0}^* = F_{n,n-1}' \mathbf{c}_{n0}$, and $V_{nt}^* = F_{n,n-1}' V_{nt}$. We observe that $F_{n,n-1}' W_n = F_{n,n-1}' W_n (F_{n,n-1} F_{n,n-1}' + (1/n) l_n l_n') = F_{n,n-1}' W_n F_{n,n-1} F_{n,n-1}'$ because $W_n l_n = l_n$ and $F_{n,n-1}' l_n = \mathbf{0}$. Denoting $W_n^* = F_{n,n-1}' W_n F_{n,n-1}$, one arrives at

$$Y_{nt}^* = \lambda_0 W_n^* Y_{nt}^* + \gamma_0 Y_{n,t-1}^* + \rho_0 W_n^* Y_{n,t-1}^* + X_{nt}^* \beta_0 + \mathbf{c}_{n0}^* + V_{nt}^*, \tag{A.2}$$

where V_{nt}^* is an $(n-1)$-dimensional disturbance vector with zero mean and variance matrix $\sigma_0^2 I_{n-1}$.

A.2. Determinant and Inverse of $I_{n-1} - \lambda W_n^*$. We note that $(I_{n-1} - \lambda W_n^*) = F_{n,n-1}' (I_n - \lambda W_n) F_{n,n-1}$. For the determinant, we have

$$\begin{aligned}
&\left[F_{n,n-1}, l_n/\sqrt{n}\right]' (I_n - \lambda W_n) \left[F_{n,n-1}, l_n/\sqrt{n}\right] \\
&\quad = \begin{pmatrix} F_{n,n-1}'(I_n - \lambda W_n) F_{n,n-1} & F_{n,n-1}'(I_n - \lambda W_n)(l_n/\sqrt{n}) \\ (l_n'/\sqrt{n})(I_n - \lambda W_n) F_{n,n-1} & (l_n'/\sqrt{n})(I_n - \lambda W_n)(l_n/\sqrt{n}) \end{pmatrix} \\
&\quad = \begin{pmatrix} F_{n,n-1}'(I_n - \lambda W_n) F_{n,n-1} & \mathbf{0} \\ -\lambda l_n' W_n F_{n,n-1}/\sqrt{n} & 1 - \lambda \end{pmatrix},
\end{aligned}$$

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

because $F'_{n,n-1}W_n l_n = F'_{n,n-1}l_n = \mathbf{0}$ and $(1/n)l'_n W_n l_n = 1$. Hence, $\left|I_{n-1} - \lambda W_n^*\right| = \left|F'_{n,n-1}(I_n - \lambda W_n)F_{n,n-1}\right| = (1/1-\lambda)|I_n - \lambda W_n|$. The inverse of $I_{n-1} - \lambda W_n^*$ is $(I_{n-1} - \lambda W_n^*)^{-1} = F'_{n,n-1}(I_n - \lambda W_n)^{-1}F_{n,n-1}$, because we have

$$
\begin{aligned}
&(I_{n-1} - \lambda W_n^*) \cdot F'_{n,n-1}(I_n - \lambda W_n)^{-1}F_{n,n-1} \\
&\quad = F'_{n,n-1}(I_n - \lambda W_n)F_{n,n-1}F'_{n,n-1}(I_n - \lambda W_n)^{-1}F_{n,n-1} \\
&\quad = F'_{n,n-1}(I_n - \lambda W_n)J_n(I_n - \lambda W_n)^{-1}F_{n,n-1} \\
&\quad = F'_{n,n-1}F_{n,n-1} - F'_{n,n-1}(I_n - \lambda W_n)\frac{l_n l'_n}{n}(I_n - \lambda W_n)^{-1}F_{n,n-1} = I_{n-1},
\end{aligned}
$$

as $F'_{n,n-1}l_n = \mathbf{0}$ and $F'_{n,n-1}W_n l_n = \mathbf{0}$. We also have

$$
\begin{aligned}
W_n^*(I_{n-1} - \lambda W_n^*)^{-1} &= F'_{n,n-1}W_n F_{n,n-1}F'_{n,n-1}(I_n - \lambda W_n)^{-1}F_{n,n-1} \\
&= F'_{n,n-1}W_n J_n(I_n - \lambda W_n)^{-1}F_{n,n-1} \\
&= F'_{n,n-1}W_n(I_n - \lambda W_n)^{-1}F_{n,n-1},
\end{aligned}
$$

and $(I_{n-1} - \lambda W_n^*)^{-1}W_n^* = W_n^*(I_{n-1} - \lambda W_n^*)^{-1} = F'_{n,n-1}W_n(I_n - \lambda W_n)^{-1}F_{n,n-1}$.

A.3. About $\mathbf{tr}[(\boldsymbol{J_n G_n(\lambda)})^{\boldsymbol{m}}]$. As is shown in Section A.2 that $S_n^{*-1}(\lambda) = (I_{n-1} - \lambda W_n^*)^{-1} = F'_{n,n-1}S_n^{-1}(\lambda)F_{n,n-1}$, it follows that $G_n^*(\lambda) = W_n^* S_n^{*-1}(\lambda) = F'_{n,n-1}W_n F_{n,n-1} \cdot F'_{n,n-1}S_n^{-1}(\lambda)F_{n,n-1} = F'_{n,n-1}W_n J_n S_n^{-1}(\lambda)F_{n,n-1} = F'_{n,n-1}G_n(\lambda)F_{n,n-1}$ because $F'_{n,n-1}W_n J_n = F'_{n,n-1}W_n$. Furthermore, as $G_n(\lambda)l_n = \frac{1}{(1-\lambda)}l_n$ and $F'_{n,n-1}G_n(\lambda)l_n = \mathbf{0}$, by induction, one has $G_n^{*m}(\lambda) = F'_{n,n-1}G_n^m(\lambda)F_{n,n-1}$, $m = 0, 1, 2, \ldots$. From this, $\mathrm{tr}(G_n^{*m}(\lambda)) = \mathrm{tr}(J_n G_n^m(\lambda))$ because $F_{n,n-1}F'_{n,n-1} = J_n$. Furthermore, because $G_n^m(\lambda)l_n = \frac{1}{(1-\lambda)^m}l_n$, $\mathrm{tr}(G_n^{*m}(\lambda)) - \mathrm{tr}(G_n^m(\lambda)) = -(1/n)\mathrm{tr}(G_n^m(\lambda)l_n l'_n) = -(1/n(1-\lambda)^m)$ $\mathrm{tr}(l_n l'_n) = -(1/(1-\lambda)^m)$, $m = 1, 2, \cdots$.

APPENDIX B: Some Lemmas for the Model

The following lemmas can be found in Yu et al. (2008). Denote $\mathbb{U}_{nt} = \sum_{h=1}^{\infty} P_{nh}V_{n,t+1-h}$ where $\{P_{nh}\}_{h=1}^{\infty}$ is a sequence of $n \times n$ nonstochastic square matrices.

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

Assumption A2. $\sum_{h=1}^{\infty} abs(P_{hn})$ is UB.

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

Assumption A4. n is a nondecreasing function of T, and T goes to infinity.

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

LEMMA 1. *Under Assumptions A1 and A4, for an $n \times n$ nonstochastic matrix B_n that is UB,* $\frac{1}{nT}\sum_{t=1}^{T} V_{nt}' B_n V_{nt} - \mathrm{E}\left(\frac{1}{nT}\sum_{t=1}^{T} V_{nt}' B_n V_{nt}\right) = O_p\left(\frac{1}{\sqrt{nT}}\right)$, $(1/n)\bar{V}_{nT}' B_n \bar{V}_{nT} - \mathrm{E}\left((1/n)\bar{V}_{nT}' B_n \bar{V}_{nT}\right) = O_p\left(\frac{1}{\sqrt{nT^2}}\right)$ *and* $\frac{1}{nT}\sum_{t=1}^{T} \tilde{V}_{nt}' B_n \tilde{V}_{nt} - \mathrm{E}\left(\frac{1}{nT}\sum_{t=1}^{T} \tilde{V}_{nt}' B_n \tilde{V}_{nt}\right) = O_p\left(\frac{1}{\sqrt{nT}}\right)$, *where* $\mathrm{E}\left(\frac{1}{nT}\sum_{t=1}^{T} V_{nt}' B_n V_{nt}\right) = (1/n)\sigma_0^2 \mathrm{tr}(B_n) = O(1)$ *and* $\mathrm{E}\left((1/n)\bar{V}_{nT}' B_n \bar{V}_{nT}\right) = (1/nT)\sigma_0^2 \mathrm{tr}(B_n) = O\left(T^{-1}\right)$.

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

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

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

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

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

where $\mu_s = \mathrm{E}v_{it}^s$ for $s = 3, 4$.

LEMMA 3. *Under Assumptions A1, A2, A3, and A4 and where B_n is UB, if the sequence $(1/nT)\sigma_{Q_{nT}}^2$ is bounded away from zero, then* $\frac{Q_{nT}}{\sigma_{Q_{nT}}} \xrightarrow{d} N(0,1)$.

LEMMA 4. *Let $Z_{nt} = (Y_{n,t-1}, W_n Y_{n,t-1}, X_n)$, under Assumptions 1–7, for an $n \times n$ nonstochastic matrix B_n that is UB,*

$$\frac{1}{nT}\sum_{t=1}^{T}(J_n \tilde{Z}_{nt})' B_n (J_n \tilde{Z}_{nt}) - \mathrm{E}\frac{1}{nT}\sum_{t=1}^{T}(J_n \tilde{Z}_{nt})' B_n (J_n \tilde{Z}_{nt}) = O_p\left(\frac{1}{\sqrt{nT}}\right), \qquad \textbf{(B.1)}$$

$$\frac{1}{nT}\sum_{t=1}^{T}(J_n \tilde{Z}_{nt})' B_n \tilde{V}_{nt} - \mathrm{E}\frac{1}{nT}\sum_{t=1}^{T}(J_n \tilde{Z}_{nt})' B_n \tilde{V}_{nt} = O_p\left(\frac{1}{\sqrt{nT}}\right), \qquad \textbf{(B.2)}$$

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

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

LEMMA 5. *If* $\|A_n(\theta_0)\|_\infty < 1$ *(resp:* $\|A_n(\theta_0)\|_1 < 1$*), then the row sum (resp: column sum) of* $\sum_{h=0}^{\infty} A_n^h(\theta)$ *and* $\sum_{h=1}^{\infty} h A_n^{h-1}(\theta)$ *is bounded uniformly in n and in a neighborhood of* θ_0.

APPENDIX C: Concentrated QML of Transformation Approach

C.1. Reduced Form of (1). From (1), we have $Y_{nt} = S_n^{-1}(Z_{nt}\delta_0 + \mathbf{c}_{n0} + \alpha_t l_n + V_{nt})$ and $W_n Y_{nt} = G_n Z_{nt}\delta_0 + G_n \mathbf{c}_{n0} + \alpha_t G_n l_n + G_n V_{nt}$. Hence, using $S_n^{-1} = I_n + \lambda_0 G_n$, we have $Y_{nt} = Z_{nt}\delta_0 + \lambda_0 G_n Z_{nt}\delta_0 + S_n^{-1}\mathbf{c}_{n0} + \alpha_t S_n^{-1} l_n + S_n^{-1} V_{nt}$. Using $S_n^{-1} l_n = (1/1-\lambda_0) l_n$ and $J_n l_n = \mathbf{0}$, we have

$$J_n \tilde{Y}_{nt} = J_n \tilde{Z}_{nt}\delta_0 + \lambda_0 J_n G_n \tilde{Z}_{nt}\delta_0 + J_n S_n^{-1}\tilde{V}_{nt}. \tag{C.1}$$

Similarly, because $J_n G_n l_n = J_n W_n S_n^{-1} l_n = J_n W_n J_n S_n^{-1} l_n = \mathbf{0}$, we can obtain $J_n W_n \tilde{Y}_{nt} = J_n G_n \tilde{Z}_{nt}\delta_0 + J_n G_n \tilde{V}_{nt}$.

C.2. First-Order Condition and Second-Order Condition of the ML Estimator. For the concentrated likelihood function (7), using $\mathrm{tr}G_n(\lambda) - \mathrm{tr}(J_n G_n(\lambda)) = (1/1-\lambda)$ and $\mathrm{tr}(G_n^2(\lambda)) - \mathrm{tr}((J_n G_n(\lambda))^2) = \frac{1}{(1-\lambda)^2}$ (see Section A.3 of Appendix A), the first-order derivatives are

$$\frac{\partial \ln L_{n,T}(\theta)}{\partial \theta} = \begin{pmatrix} \frac{\partial \ln L_{n,T}(\theta)}{\partial \delta} \\ \frac{\partial \ln L_{n,T}(\theta)}{\partial \lambda} \\ \frac{\partial \ln L_{n,T}(\theta)}{\partial \sigma^2} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^{T} (J_n \tilde{Z}_{nt})' \tilde{V}_{nt}(\theta) \\ \frac{1}{\sigma^2}\sum_{t=1}^{T} \left((J_n W_n \tilde{Y}_{nt})' \tilde{V}_{nt}(\theta)\right) - T\,\mathrm{tr}G_n(\lambda) + \frac{T}{1-\lambda} \\ \frac{1}{2\sigma^4}\sum_{t=1}^{T} (\tilde{V}_{nt}'(\theta) J_n \tilde{V}_{nt}(\theta) - (n-1)\sigma^2) \end{pmatrix}, \tag{C.2}$$

and the second-order derivatives are

$$\frac{\partial^2 \ln L_{n,T}(\theta)}{\partial \theta \partial \theta'} \tag{C.3}$$

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

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

APPENDIX D: Proofs for Claims and Theorems

D.1. Proof of Claim 1.

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

From $\tilde{V}_{nt}(\theta) \equiv S_n(\lambda)\tilde{Y}_{nt} - \tilde{Z}_{nt}\delta - \tilde{\alpha}_t l_n$ and $\tilde{V}_{nt} = S_n\tilde{Y}_{nt} - \tilde{Z}_{nt}\delta_0 - \tilde{\alpha}_{t0}l_n$, using $J_n l_n = \mathbf{0}$, we have $J_n\tilde{V}_{nt}(\theta) = J_n\tilde{V}_{nt} - (\lambda-\lambda_0)J_nW_n\tilde{Y}_{nt} - J_n\tilde{Z}_{nt}(\delta-\delta_0)$. As Θ is compact so that λ, δ are bounded in Θ, using Lemmas 1 and 4, $\frac{1}{(n-1)T}\sum_{t=1}^{T}\tilde{V}'_{nt}(\theta)J_n\tilde{V}_{nt}(\theta) - \frac{1}{(n-1)T}\mathrm{E}\sum_{t=1}^{T}\tilde{V}'_{nt}(\theta)J_n\tilde{V}_{nt}(\theta) \xrightarrow{p} 0$ uniformly in θ in Θ. Also, from (9) and by using the fact that σ^2 is bounded away from zero in Θ,

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

uniformly in θ in Θ.

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

We have $Q_{nT}(\theta) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 - (1/n-1)\ln(1-\lambda) + (1/n-1)\ln|S_n(\lambda)| - \frac{1}{2\sigma^2(n-1)T}\mathrm{E}\sum_{t=1}^{T}\tilde{V}'_{nt}(\theta)J_n\tilde{V}_{nt}(\theta)$. As $J_n\tilde{V}_{nt}(\theta) = J_n\left[S_n(\lambda)\tilde{Y}_{nt} - \tilde{Z}_{nt}\delta\right]$ and $\tilde{Y}_{nt} = S_n^{-1}\tilde{Z}_{nt}\delta_0 + S_n^{-1}\tilde{V}_{nt} + (\tilde{\alpha}_{t0}/1-\lambda_0)l_n$, we have $J_n\tilde{V}_{nt}(\theta) = J_n[S_n(\lambda)S_n^{-1}\tilde{Z}_{nt}\delta_0 - \tilde{Z}_{nt}\delta + S_n(\lambda)S_n^{-1}\tilde{V}_{nt}]$ because $J_n l_n = \mathbf{0}$. Hence, $\mathrm{E}\frac{1}{(n-1)T}\sum_{t=1}^{T}\tilde{V}'_{nt}(\theta)J_n\tilde{V}_{nt}(\theta) = \frac{1}{(n-1)T}\mathrm{E}\sum_{t=1}^{T}(S_n(\lambda)S_n^{-1}\tilde{Z}_{nt}\delta_0 - \tilde{Z}_{nt}\delta)'J_n(S_n(\lambda)S_n^{-1}\tilde{Z}_{nt}\delta_0 - \tilde{Z}_{nt}\delta) + (1/n-1)\frac{T-1}{T}\sigma_0^2\mathrm{tr}(S_n^{-1'}S'_n(\lambda)J_nS_n(\lambda)S_n^{-1}) + \frac{2}{(n-1)T}\mathrm{E}\sum_{t=1}^{T}(S_n(\lambda)S_n^{-1}\tilde{Z}_{nt}\delta_0 - \tilde{Z}_{nt}\delta)'J_nS_n(\lambda)S_n^{-1}\tilde{V}_{nt}$. Similar to Yu et al. (2008), $Q_{n,T}(\theta)$ is uniformly equicontinuous in θ in the compact parameter space Θ. ■

D.2. Proof of Nonsingularity of the Information Matrix. We can prove the result by using an argument by contradiction. For $\Sigma_{\theta_0} \equiv \lim_{T\to\infty}\Sigma_{\theta_0,nT}$, where $\Sigma_{\theta_0,nT}$ is defined in (10), we need to prove that $\Sigma_{\theta_0}c = 0$ implies $c = 0$ where $c = (c_1', c_2, c_3)'$ and c_2, c_3 are scalars and c_1 is a $(k_x+2)\times 1$ vector. If this is true, then the columns of Σ_{θ_0} would be linearly independent so that Σ_{θ_0} would be nonsingular. Denote $\mathcal{H}_\delta = \mathrm{plim}_{T\to\infty}\frac{1}{(n-1)T}\sum_{t=1}^{T}\tilde{Z}'_{nt}J_n\tilde{Z}_{nt}$, $\mathcal{H}_{\delta\lambda} = \mathrm{plim}_{T\to\infty}\frac{1}{(n-1)T}\sum_{t=1}^{T}\tilde{Z}'_{nt}J_nG_n\tilde{Z}_{nt}\delta_0$, $\mathcal{H}_{\lambda\delta} = \mathcal{H}'_{\delta\lambda}$, and $\mathcal{H}_\lambda = \mathrm{plim}_{T\to\infty}\frac{1}{(n-1)T}\sum_{t=1}^{T}(G_n\tilde{Z}_{nt}\delta_0)'J_nG_n\tilde{Z}_{nt}\delta_0$. Then

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

Hence, $\Sigma_{\theta_0}c = 0$ implies $\mathcal{H}_\delta \times c_1 + \mathcal{H}_{\delta\lambda} \times c_2 = 0$,

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

$$\frac{1}{\sigma_0^2}\mathcal{H}_{\lambda\delta}\times c_1+\left(\frac{1}{\sigma_0^2}\mathcal{H}_{\lambda}+\lim_{n\to\infty}\frac{1}{n-1}\left[\mathrm{tr}(G_n'J_nG_n)+\mathrm{tr}((J_nG_n)^2)\right]\right)$$
$$\times c_2+\lim_{n\to\infty}\frac{1}{\sigma_0^2(n-1)}\mathrm{tr}(J_nG_n)\times c_3=0,$$

and $\lim_{n\to\infty}\frac{1}{(n-1)}\mathrm{tr}(J_nG_n)\times c_2+\frac{1}{2\sigma_0^2}\times c_3=0$. From the first equation, $c_1=-(\mathcal{H}_\delta)^{-1}\mathcal{H}_{\delta\lambda}\times c_2$; from the third equation, $c_3=-2\lim_{n\to\infty}(\sigma_0^2/n-1)\mathrm{tr}(J_nG_n)\times c_2$. By eliminating c_1 and c_3, the remaining equation becomes

$$\left\{\left(\frac{1}{\sigma_0^2}\left(\mathcal{H}_\lambda-\mathcal{H}_{\lambda\delta}\mathcal{H}_\delta^{-1}\mathcal{H}_{\delta\lambda}\right)\right)\right.$$
$$\left.+\lim_{n\to\infty}\frac{1}{n-1}\left[\mathrm{tr}(G_n'J_nG_n)+\mathrm{tr}((J_nG_n)^2)-2\frac{\mathrm{tr}^2(J_nG_n)}{n-1}\right]\right\}\times c_2=0.$$

Denote $C_n=J_nG_n-\mathrm{tr}(J_nG_n)/(n-1)J_n$; then $\mathrm{tr}(G_n'J_nG_n)+\mathrm{tr}((J_nG_n)^2)-2\mathrm{tr}^2(J_nG_n)(n-1)=\frac{1}{2}\mathrm{tr}(C_n'+C_n)(C_n'+C_n)'$, which is nonnegative. Hence, if the limit of $\mathrm{E}\mathcal{H}_{nT}$ is nonsingular or the limit of $1/(n-1)(\mathrm{tr}(G_n'J_nG_n)+\mathrm{tr}((J_nG_n)^2)-2\mathrm{tr}^2(J_nG_n)/(n-1))$ is nonzero, we have $c_2=0$, and hence $c=0$. This proves the nonsingularity of Σ_{θ_0}. ■

D.3. Proof of Theorem 1. As $\mathrm{E}\sum_{t=1}^{T}\tilde{V}_{nt}'J_n\tilde{V}_{nt}=(n-1)(T-1)\sigma_0^2$ according to Lemma 1, at θ_0, (9) implies $\mathrm{E}\ln L_{n,T}(\theta_0)=-\frac{(n-1)T}{2}\ln 2\pi-\frac{(n-1)T}{2}\ln\sigma_0^2-T\ln(1-\lambda_0)+T\ln|S_n|-\frac{(n-1)(T-1)}{2}$. Denote $\sigma_n^2(\lambda)=\sigma_0^2/(n-1)\mathrm{tr}(S_n^{-1\prime}S_n'(\lambda)J_nS_n(\lambda)S_n^{-1})$. By using $S_n(\lambda)S_n^{-1}=I_n+(\lambda_0-\lambda)G_n$, it follows that

$$\frac{1}{(n-1)T}\mathrm{E}\ln L_{n,T}(\theta)-\frac{1}{(n-1)T}\mathrm{E}\ln L_{n,T}(\theta_0)$$
$$=-\frac{1}{2}\left(\ln\sigma^2-\ln\sigma_0^2\right)+\frac{1}{n-1}\ln|S_n(\lambda)|-\frac{1}{n-1}\ln|S_n|$$
$$-\left(\frac{1}{2\sigma^2}\frac{1}{(n-1)T}\sum_{t=1}^{T}\mathrm{E}\tilde{V}_{nt}'(\theta)J_n\tilde{V}_{nt}(\theta)-\frac{T-1}{2T}\right)$$
$$-\frac{1}{n-1}(\ln(1-\lambda)-\ln(1-\lambda_0))$$
$$=T_{1,n}\left(\lambda,\sigma^2\right)-\frac{1}{2\sigma^2}T_{2,n,T}(\delta,\lambda)+O\left(T^{-1}\right),$$

where

$$T_{1,n}(\lambda,\sigma^2)=-\frac{1}{2}(\ln\sigma^2-\ln\sigma_0^2)+\frac{1}{n-1}\ln|S_n(\lambda)|-\frac{1}{n-1}\ln|S_n|-\frac{1}{2\sigma^2}(\sigma_n^2(\lambda)-\sigma^2)$$
$$-\frac{1}{n-1}(\ln(1-\lambda)-\ln(1-\lambda_0)),$$

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

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

Consider the pure spatial process $Y_{nt}=\lambda_0 W_n Y_{nt}+\alpha_t l_n+V_{nt}$ for a period t. Similarly as (6) after data transformation, we get the log likelihood function of this process as

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

where $V_{nt}(\lambda)=S_n(\lambda)Y_{nt}$. Let $\mathrm{E}_p(\cdot)$ be the expectation operator for Y_{nt} based on this pure spatial autoregressive process. It follows that $\mathrm{E}_p(1/(n-1)\ln L_{p,n}(\lambda,\sigma^2))-\mathrm{E}_p(1/(n-1)\ln L_{p,n}(\lambda_0,\sigma_0^2))=-\frac{1}{2}(\ln\sigma^2-\ln\sigma_0^2)+1/(n-1)\ln|S_n(\lambda)|-1/(n-1)\ln|S_n(\lambda_0)|-\frac{1}{2\sigma^2}(\sigma_n^2(\lambda)-\sigma^2)-1/(n-1)(\ln(1-\lambda)-\ln(1-\lambda_0))$, which equals to $T_{1,n}(\lambda,\sigma^2)$. By the information inequality, $\ln L_{p,n}(\lambda,\sigma^2)-\ln L_{p,n}(\lambda_0,\sigma_0^2)\le 0$. Thus, $T_{1,n}(\lambda,\sigma^2)\le 0$ for any (λ,σ^2). For $T_{2,n,T}(\delta,\lambda)$, it is a quadratic function of δ and λ. Under the assumed condition that $\lim_{T\to\infty}\mathrm{E}\mathcal{H}_{nT}$ is nonsingular, $\lim_{T\to\infty}T_{2,n,T}(\delta,\lambda)>0$ whenever $(\delta,\lambda)\neq(\delta_0,\lambda_0)$. So, (δ,λ) is globally identified. Given λ_0, σ_0^2 is the unique maximizer of $T_{1,n}(\lambda_0,\sigma^2)$ for any given n. In the event that $n\to\infty$, σ_0^2 is the unique maximizer of $\lim_{T\to\infty}T_{1,n}(\lambda_0,\sigma^2)$. Hence, $(\delta,\lambda,\sigma^2)$ is globally identified.

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

D.4. Proof of Theorem 2. From the proof of Theorem 1 (see Section D.3),

$$\frac{1}{(n-1)T}\mathrm{E}\ln L_{n,T}(\theta)-\frac{1}{(n-1)T}\mathrm{E}\ln L_{n,T}(\theta_0)=T_{1,n}(\lambda,\sigma^2)-\frac{1}{2\sigma^2}T_{2,n,T}(\delta,\lambda)+O(T^{-1}).$$

When the limit of $\mathrm{E}\mathcal{H}_{nT}$ is singular, δ_0 and λ_0 cannot be identified from $T_{2,n,T}(\delta,\lambda)$. Global identification requires that the limit of $T_{1,n}(\lambda,\sigma^2)$ is strictly less than zero. Thus, we require global identification just from the pure spatial model with the likelihood function (D.1). Using the concentrating approach by concentrating out σ^2 in (D.1), we have the concentrated likelihood function

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

where $\hat{\sigma}_{nt}^2(\lambda)=1/(n-1)V_{nt}'(\lambda)J_nV_{nt}(\lambda)$. Also, we have corresponding $Q_n(\lambda)=\max_{\sigma^2}\mathrm{E}(\ln L_{p,n}(\lambda,\sigma^2))=-(n-1)/2(\ln(2\pi)+1)-(n-1)/2\ln\sigma_n^2(\lambda)-\ln(1-\lambda)+\ln|S_n(\lambda)|$. Global identification of λ_0 requires that $\lim_{n\to\infty}\frac{1}{n}[Q_n(\lambda)-Q_{nt}(\lambda_0)]\neq 0$ whenever $\lambda\neq\lambda_0$, which is equivalent to $\lim_{n\to\infty}\frac{1}{n}[\ln|S_n(\lambda)|-\ln|S_n|-(n-1)/2(\ln\sigma_n^2(\lambda)-\ln\sigma_n^2(\lambda_0))-(\ln(1-\lambda)-\ln(1-\lambda_0))]\neq 0$ whenever $\lambda\neq\lambda_0$. By rearranging the terms, $Q_n(\lambda)-Q_{nt}(\lambda_0)\neq 0$ whenever $\lambda\neq\lambda_0$ is just equivalent to

$$\frac{1}{n}\ln\left|\sigma_0^2S_n^{-1\prime}S_n^{-1}\right|-\frac{1}{n}\ln\left|\sigma_n^2(\lambda)S_n^{-1}(\lambda)'S_n^{-1}(\lambda)\right|+\left(\frac{1}{n}\ln\frac{\sigma_0^2}{(1-\lambda_0)^2}-\frac{1}{n}\ln\frac{\sigma_n^2(\lambda)}{(1-\lambda)^2}\right)\neq 0$$

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

for $\lambda \neq \lambda_0$. When $n \to \infty$, it becomes $\lim_{n\to\infty}\left(\frac{1}{n}\ln\left|\sigma_0^2 S_n^{-1\prime} S_n^{-1}\right| - \frac{1}{n}\ln\left|\sigma_n^2(\lambda) S_n^{-1}(\lambda)'\right.\right.$ $\left.\left.S_n^{-1}(\lambda)\right|\right) \neq 0$ for $\lambda \neq \lambda_0$. After λ_0 is identified, σ_0^2 is then identified. Also, given λ_0, δ_0 can be identified from $\lim_{T\to\infty} T_{2,n,T}(\delta, \lambda)$. Combined with uniform convergence and equicontinuity in Claim 1, the consistency follows. ■

D.5. Proof of Claim 2. The score can be decomposed into two parts such that $\frac{1}{\sqrt{(n-1)T}}$ $\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta} = \frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial\theta} - \Delta_{nT}$, where $\frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial\theta}$ is defined in (12) and Δ_{nT} is defined in (13). For the first part, it is a linear and quadratic form of V_{nt}, and its asymptotic distribution can be derived by Lemma 3. For the term Δ_{nT}, from Lemmas 1 and 2, we have $\Delta_{nT} = \sqrt{(n-1)/T}\, a_{\theta_0,n} + O(\sqrt{(n-1)/T^3}) + O_p(\frac{1}{\sqrt{T}})$ where $a_{\theta_0,n}$ is $O(1)$ and specified in (16). Combined, $\frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta} + \Delta_{nT} \xrightarrow{d} N(0, \Sigma_{\theta_0} + \Omega_{\theta_0})$. ■

D.6. Proof of Theorem 3. Taylor expansion gives

$$\sqrt{(n-1)T}(\hat\theta_{nT} - \theta_0) = \left(-(1)/(n-1)T\,\frac{\partial^2 \ln L_{n,T}(\bar\theta_{nT})}{\partial\theta\partial\theta'}\right)^{-1}\left(\frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial\theta} - \Delta_{nT}\right),$$

where $\bar\theta_{nT}$ lies between θ_0 and $\hat\theta_{nT}$. Similarly as in Appendix D in Yu et al. (2008), we have $\hat\theta_{nT} - \theta_0 = O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{T}\right)\right)$. As $\Sigma_{\theta_0,nT}$ is nonsingular and its inverse is of order $O(1)$, using $\left(-(1)/(n-1)T\,\frac{\partial^2 \ln L_{n,T}(\bar\theta_{nT})}{\partial\theta\partial\theta'}\right)^{-1} = \Sigma_{\theta_0,nT}^{-1} + O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{T}\right)\right)$, we have $\sqrt{(n-1)T}(\hat\theta_{nT} - \theta_0) + \Sigma_{\theta_0,nT}^{-1}\cdot\Delta_{nT} + O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{T}\right)\right)\cdot\Delta_{nT} = (\Sigma_{\theta_0,nT}^{-1} +$ $o_p(1))\cdot\frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial\theta}$. As $\Sigma_{\theta_0} = \lim_{T\to\infty}\Sigma_{\theta_0,nT}$ exists, then using Claim 2 and that $\Delta_{nT} = \sqrt{(n-1)/T}\, a_{\theta_0,n} + O(\sqrt{(n-1)/T^3}) + O_p(\frac{1}{\sqrt{T}})$ with $a_{\theta_0,n} = O(1)$,

$$\sqrt{(n-1)T}(\hat\theta_{nT} - \theta_0) + \sqrt{\frac{n-1}{T}}\,\Sigma_{\theta_0,nT}^{-1} a_{\theta_0,n} + O_p\left(\max\left(\sqrt{\frac{n}{T^3}}, \sqrt{\frac{1}{T}}\right)\right)$$

$$\xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0} + \Omega_{\theta_0})\Sigma_{\theta_0}^{-1}). \qquad \textbf{(D.3)}$$

■

D.7. Proof of Theorem 4. As $\hat\theta_{nT}^1 = \hat\theta_{nT} + \frac{1}{T}\left(-\frac{1}{(n-1)T}\mathrm{E}\frac{\partial^2 \ln L_{nT}(\hat\theta_{nT})}{\partial\theta\partial\theta'}\right)^{-1}\cdot a_n(\hat\theta_{nT})$ where $a_n(\theta) = a_{\theta,n}$, (D.3) will imply that $\sqrt{(n-1)T}(\hat\theta_{nT}^1 - \theta_0) \xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0} +$ $\Omega_{\theta_0})\Sigma_{\theta_0}^{-1})$ if

$$\sqrt{\frac{n-1}{T}}\left(\left(-\frac{1}{(n-1)T}\mathrm{E}\frac{\partial^2 \ln L_{nT}(\hat\theta_{nT})}{\partial\theta\partial\theta'}\right)^{-1} a_n(\hat\theta_{nT}) - \Sigma_{\theta_0,nT}^{-1} a_n(\theta_0)\right) \xrightarrow{p} 0 \qquad \textbf{(D.4)}$$

and $\frac{n}{T^3} \to 0$. Assuming that $\frac{n}{T^3} \to 0$, we are going to prove (D.4). This can be proved similarly as in Theorem 4 in Yu et al. (2008). ■

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

APPENDIX E: Direct Approach

E.1. Concentrated Likelihood Function and its First-Order Condition and Second-Order Condition. The first- and second-order derivatives of the concentrated log likelihood in (20) are

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^{d}(\theta)}{\partial \theta}=\frac{1}{\sqrt{nT}}\begin{pmatrix}\frac{\partial \ln L_{n,T}^{d}(\theta)}{\partial \delta}\\ \frac{\partial \ln L_{n,T}^{d}(\theta)}{\partial \lambda}\\ \frac{\partial \ln L_{n,T}^{d}(\theta)}{\partial \sigma^{2}}\end{pmatrix}$$

$$=\begin{pmatrix}\frac{1}{\sigma^{2}}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}(J_n\tilde{Z}_{nt})'\tilde{V}_{nt}(\theta)\\ \frac{1}{\sigma^{2}}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\left((J_nW_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\theta)-\sigma^{2}\mathrm{tr}G_n(\lambda)\right)\\ \frac{1}{2\sigma^{4}}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}(\tilde{V}_{nt}'(\theta)J_n\tilde{V}_{nt}(\theta)-n\sigma^{2})\end{pmatrix}, \tag{E.1}$$

$$\frac{1}{nT}\frac{\partial^{2}\ln L_{n,T}^{d}(\theta)}{\partial\theta\partial\theta'} \tag{E.2}$$

$$=-\frac{1}{nT}\begin{pmatrix}\frac{1}{\sigma^{2}}\sum_{t=1}^{T}\tilde{Z}_{nt}'J_n\tilde{Z}_{nt} & \frac{1}{\sigma^{2}}\sum_{t=1}^{T}\tilde{Z}_{nt}'J_nW_n\tilde{Y}_{nt} & \frac{1}{\sigma^{4}}\sum_{t=1}^{T}\tilde{Z}_{nt}'J_n\tilde{V}_{nt}(\theta)\\ * & \frac{1}{\sigma^{2}}\sum_{t=1}^{T}((W_n\tilde{Y}_{nt})'J_nW_n\tilde{Y}_{nt} & \frac{1}{\sigma^{4}}\sum_{t=1}^{T}(W_n\tilde{Y}_{nt})'J_n\tilde{V}_{nt}(\theta)\\ & +\mathrm{tr}(G_n^{2}(\lambda))) & \\ * & * & -\frac{nT}{2\sigma^{4}}+\frac{1}{\sigma^{6}}\sum_{t=1}^{T}\tilde{V}_{nt}'(\theta)\\ & & \times J_n\tilde{V}_{nt}(\theta)\end{pmatrix}.$$

Similar to the variance matrix of $\frac{1}{\sqrt{(n-1)T}}\frac{\partial \ln L_{n,T}^{(u)}(\theta_0)}{\partial\theta}$ in (12), that of $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^{d(u)}(\theta_0)}{\partial\theta}$ in (22) is equal to

$$\mathrm{E}\left(\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^{d(u)}(\theta_0)}{\partial\theta}\cdot\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^{d(u)}(\theta_0)}{\partial\theta'}\right)=\Sigma_{\theta_0,nT}^{d}+\Omega_{\theta_0,n}^{d}+O\left(T^{-1}\right), \tag{E.3}$$

and $\Omega_{\theta_0,n}^{d}=(\mu_4-3\sigma_0^{4})/\sigma_0^{4}\begin{pmatrix}0&0&0\\0&\frac{1}{n}\sum_{i=1}^{n}G_{n,ii}^{2}&\frac{1}{2\sigma_0^{2}n}\mathrm{tr}G_n\\0&\frac{1}{2\sigma_0^{2}n}\mathrm{tr}G_n&\frac{1}{4\sigma_0^{4}}\end{pmatrix}$ is a symmetric matrix with μ_4 being the fourth moment of v_{it}, where $G_{n,ii}$ is the (i,i) entry of G_n. Denote $\Sigma_{\theta_0}=\lim_{T\to\infty}\Sigma_{\theta_0,nT}^{d}$ and $\Omega_{\theta_0}=\lim_{T\to\infty}\Omega_{\theta_0,n}^{d}$ (note that $\lim_{T\to\infty}\Sigma_{\theta_0,nT}^{d}=\lim_{T\to\infty}\Sigma_{\theta_0,nT}$ and $\lim_{T\to\infty}\Omega_{\theta_0,n}^{d}=\lim_{T\to\infty}\Omega_{\theta_0,n}$); then

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

$$\lim_{T\to\infty} \mathrm{E}\left(\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^{d(u)}(\theta_0)}{\partial\theta}\cdot\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^{d(u)}(\theta_0)}{\partial\theta'}\right) = \Sigma_{\theta_0}+\Omega_{\theta_0}. \tag{E.4}$$

CLAIM 3. *Under Assumptions 1′, 2–6, 7′, and 8,* $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^{d}(\theta_0)}{\partial\theta}+\Delta_{1,nT}+\Delta_{2,nT}\xrightarrow{d} N(0,\Sigma_{\theta_0}+\Omega_{\theta_0})$ *where* $\Delta_{1,nT}$ *and* $\Delta_{2,nT}$ *are, respectively, in (23) and (24). When* $\{v_{it}\}$, $i=1,2,\ldots,n$ *and* $t=1,2,\ldots,T$, *are normal,* $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^{d}(\theta_0)}{\partial\theta}+\Delta_{1,nT}+\Delta_{2,nT}\xrightarrow{d} N(0,\Sigma_{\theta_0})$.

Proof. This is similar to Claim 2. ■

CLAIM 4. *Under Assumptions 1′, 2–6, 7′, and 8,* $\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^{d}(\theta)}{\partial\theta\partial\theta'}-\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^{d}(\theta_0)}{\partial\theta\partial\theta'}=\|\theta-\theta_0\|\cdot O_p(1)$ *and* $\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^{d}(\theta_0)}{\partial\theta\partial\theta'}-\frac{1}{nT}\mathrm{E}\frac{\partial^2 \ln L_{n,T}^{d}(\theta_0)}{\partial\theta\partial\theta'}=O_p\left(\frac{1}{\sqrt{nT}}\right)$.

Proof. These are similar to (38) and (39) in Yu et al. (2008). ■

E.2. Proof of Theorem 5. For the log likelihood function (18) divided by the sample size nT, we have corresponding $Q_{n,T}^{d}(\theta)=\mathrm{E}\max_{\mathbf{c}_n,\alpha_T}\frac{1}{nT}\ln L_{n,T}(\theta,\mathbf{c}_n,\alpha_T)$. Hence,

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

Similarly to the proof of Claim 1 in Section D.1 of Appendix D, we can show the uniform convergence of $\frac{1}{nT}\ln L_{n,T}^{d}(\theta)-Q_{n,T}^{d}(\theta)$ to zero in probability and that $Q_{n,T}^{d}(\theta)$ is uniformly equicontinuous in θ in any compact parameter space Θ. When $\lim_{T\to\infty}\mathcal{H}_{nT}^{d}$ is nonsingular, we can show that $\hat{\theta}_{nT}^{d}$ is globally identified and consistent similarly to the proof of Theorem 1 in Section D.3 of Appendix D. When $\lim_{T\to\infty}\mathcal{H}_{nT}^{d}$ is singular, as long as

$$\lim_{n\to\infty}\left(\frac{1}{n}\ln\left|\sigma_0^2 S_n'^{-1}S_n^{-1}\right|-\frac{1}{n}\ln\left|\sigma_n^2(\lambda)S_n'^{-1}(\lambda)S_n^{-1}(\lambda)\right|\right)\neq 0 \quad \text{for } \lambda\neq\lambda_0,$$

θ_0 is still globally identified, and $\hat{\theta}_{nT}^{d}\xrightarrow{p}\theta_0$. The proof can be done similarly as in Section D.4 of Appendix D. ■

E.3. Proof of Theorem 6. Taylor expansion gives

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

where $\bar{\theta}_{nT}$ lies between θ_0 and $\hat{\theta}_{nT}^{d}$ and $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^{d}(\theta_0)}{\partial\theta}=\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^{d(u)}(\theta_0)}{\partial\theta}-\Delta_{1,nT}-\Delta_{2,nT}$. Similarly as in Section D.6 of Appendix D, we obtain the result. ■

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press

E.4. Proof of Theorem 7. The bias corrected estimator is

$$\hat{\theta}_{nT}^{d1} = \hat{\theta}_{nT}^{d} + \frac{1}{T}\left(-\frac{1}{nT}\mathrm{E}\frac{\partial^2 \ln L_{nT}^{d}(\hat{\theta}_{nT}^{d})}{\partial\theta\partial\theta'}\right)^{-1} a_{1,n}(\hat{\theta}_{nT}^{d}) + \frac{1}{n}\left(-\frac{1}{nT}\mathrm{E}\frac{\partial^2 \ln L_{nT}^{d}(\hat{\theta}_{nT}^{d})}{\partial\theta\partial\theta'}\right)^{-1} a_2(\hat{\theta}_{nT}^{d}),$$

where $a_{1,n}(\theta) = a_{\theta,n,1}$ and $a_2(\theta) = a_{\theta,2}$, we will have $\sqrt{nT}(\hat{\theta}_{nT}^{d1} - \theta_0) \xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1}(\Sigma_{\theta_0} + \Omega_{\theta_0})\Sigma_{\theta_0}^{-1})$ if

$$\sqrt{\frac{n}{T}}\left(\left(-\frac{1}{nT}\mathrm{E}\frac{\partial^2 \ln L_{nT}^{d}(\hat{\theta}_{nT}^{d})}{\partial\theta\partial\theta'}\right)^{-1} a_{1,n}(\hat{\theta}_{nT}^{d}) - (\Sigma_{\theta_0,nT}^{d})^{-1} a_{1,n}(\theta_0)\right) \xrightarrow{p} 0 \qquad \textbf{(E.6)}$$

and

$$\sqrt{\frac{T}{n}}\left(\left(-\frac{1}{nT}\mathrm{E}\frac{\partial^2 \ln L_{nT}^{d}(\hat{\theta}_{nT}^{d})}{\partial\theta\partial\theta'}\right)^{-1} a_2(\hat{\theta}_{nT}^{d}) - (\Sigma_{\theta_0,nT}^{d})^{-1} a_2(\theta_0)\right) \xrightarrow{p} 0, \qquad \textbf{(E.7)}$$

when $\frac{n}{T^3} \to 0$, $\frac{T}{n^3} \to 0$. Assuming that $\frac{n}{T^3} \to 0$ and $\frac{T}{n^3} \to 0$, (E.6) and (E.7) can be proved similarly as in Section D.7 of Appendix D. ■

https://doi.org/10.1017/S0266466609100099 Published online by Cambridge University Press