Journal of Econometrics 154 (2010) 165–185

Contents lists available at ScienceDirect

Journal of Econometrics

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

Estimation of spatial autoregressive panel data models with fixed effects[☆]

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

[a] Department of Economics, Ohio State University, 43210 Columbus, OH, United States
[b] Department of Economics, University of Kentucky, United States

ARTICLE INFO

Article history:
Received 4 March 2008
Received in revised form
4 June 2009
Accepted 1 August 2009
Available online 12 August 2009

JEL classification:
C13
C23
R15

Keywords:
Spatial autoregression
Panel data
Fixed effects
Quasi-maximum likelihood estimation
Conditional likelihood

ABSTRACT

This paper establishes asymptotic properties of quasi-maximum likelihood estimators for SAR panel data models with fixed effects and SAR disturbances. A direct approach is to estimate all the parameters including the fixed effects. Because of the incidental parameter problem, some parameter estimators may be inconsistent or their distributions are not properly centered. We propose an alternative estimation method based on transformation which yields consistent estimators with properly centered distributions. For the model with individual effects only, the direct approach does not yield a consistent estimator of the variance parameter unless T is large, but the estimators for other common parameters are the same as those of the transformation approach. We also consider the estimation of the model with both individual and time effects.

1. Introduction

Spatial econometrics consists of econometric techniques dealing with the interactions of economic units in space, which can have physical or economic characteristic. The spatial autoregressive (SAR) model by Cliff and Ord (1973) has received the most attention in economics.[1] Panel data with spatial interaction is also of great interest, as it enables researchers to take into account the dynamics and control for the unobservable heterogeneity (e.g., Anselin, 1988; Baltagi et al., 2003, 2007; Elhorst, 2003; Kapoor et al., 2007; Yu et al., 2007, 2008; Yu and Lee, forthcoming).

For panel data models with fixed individual effects, when the time dimension T is fixed, we are likely to encounter the incidental parameter problem discussed in Neyman and Scott (1948). This is because the introduction of fixed effects increases the number of parameters. For the linear panel regression model with fixed effects, the direct maximum likelihood (ML) approach estimates jointly the common parameters and fixed effects. The corresponding ML estimates (MLEs) of the regression coefficients are known as the within estimates, which happen to be the likelihood estimates conditional on the time means of the dependent variables. However, the MLE of the variance parameter is inconsistent when T is finite. The inconsistency of the variance parameter is the one illustrated in Neyman and Scott (1948). For the SAR panel data models with fixed individual effects, similar findings of the direct ML approach are found in this paper. The direct approach will yield consistent estimates for the spatial and regression coefficients, except for the variance parameter when T is small (but the number of spatial units n is large).[2] For the SAR panel models with both fixed individual and time effects, the direct approach will be inconsistent for the estimation of the common parameters unless n is large. Even when both n and T are large, the distribution of the estimates of the common parameters would not be properly centered.

To eliminate the fixed effects, the method of conditional likelihood is used when sufficient statistics can be found for the

[☆] We would like to thank participants of the seminars at Syracuse University and the University of Kentucky, and two anonymous referees for helpful comments. Lee acknowledges financial support from NSF Grant No. SES-0519204.

* Corresponding author. Tel.: +1 614 292 5508; fax: +1 614 292 4192.
E-mail addresses: lflee@econ.ohio-state.edu (L.-f. Lee), jihai.yu@uky.edu (J. Yu).

[1] Early development in estimation and testing for cross sectional data can be found in Anselin (1988), Cressie (1993), Kelejian and Robinson (1993), and Anselin and Bera (1998), among others.

[2] When a dynamic effect is considered into the SAR panel data, we will have an "initial condition" problem which will cause the inconsistency of the direct likelihood estimates for all the parameters unless T is large (see Yu et al., 2007, 2008 and Yu and Lee, forthcoming). The initial value problem for the dynamic panel data model is well known (Nickell, 1981).

0304-4076/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2009.08.001

fixed effects. For the linear regression and logit panel models, the time average of the dependent variables for each cross sectional unit provides a sufficient statistic[3] (see Hsiao, 1986). For the normal panel regression model, the conditional likelihood can be constructed from some transformed data. In this paper, we investigate the use of similar transformations to the SAR panel model. By using a data transformation from $(I_T - \frac{1}{T} l_T l_T')$ to eliminate the individual effects where l_T is the $T \times 1$ vector of ones, the transformed equation can be estimated by the quasi-maximum likelihood (QML) approach. For the model with both individual and time fixed effects, one may combine the transformations from $(I_n - \frac{1}{n} l_n l_n')$ and $(I_T - \frac{1}{T} l_T l_T')$ to eliminate both effects.[4] The transformation approach for our models can either be justified as a conditional likelihood, a partial likelihood (Cox, 1975), or a modified likelihood based on a concentrated likelihood of the direct estimation (Kalbfleisch and Sprott, 1970; Cox and Reid, 1987; Lancaster, 2000; Arellano and Hahn, 2005).[5]

Panel regression models with SAR disturbances have recently been considered in the spatial econometrics literature. The model in Baltagi et al. (2003) is $Y_{nt} = X_{nt}\beta_0 + \mathbf{c}_{n0} + U_{nt}$, $U_{nt} = \lambda_0 W_n U_{nt} + V_{nt}$, $t = 1, 2, \ldots, T$, where elements of V_{nt} are *i.i.d.* $(0, \sigma_0^2)$, $\mathbf{c}_{n0}$ is an $n \times 1$ vector of individual error components, W_n is a spatial weights matrix, and the spatial correlation is in U_{nt}. A different specification in Kapoor et al. (2007) is $Y_{nt} = X_{nt}\beta_0 + U_{nt}^+$ and $U_{nt}^+ = \lambda_0 W_n U_{nt}^+ + \mathbf{d}_{n0} + V_{nt}$, $t = 1, 2, \ldots, T$, where $\mathbf{d}_{n0}$ is the vector of individual error components. Kapoor et al. (2007) propose a method of moment (MOM) procedure for the estimation of λ_0 along with the variances of $\mathbf{d}_{n0}$ and V_{nt}. The two models are different in terms of the variance matrices of the overall disturbances. The variance matrix in Baltagi et al. (2003) is more complicated and its inverse is computationally demanding; the one in Kapoor et al. (2007) has a special pattern and its inverse can be easier to compute. By the transformation $(I_n - \lambda_0 W_n)$, the data generating process (DGP) of Kapoor et al. (2007) becomes $Y_{nt} = X_{nt}\beta_0 + \mathbf{c}_{n0} + U_{nt}$ where $\mathbf{c}_{n0} = (I_n - \lambda_0 W_n)^{-1}\mathbf{d}_{n0}$ and $U_{nt} = U_{nt}^+ - (I_n - \lambda_0 W_n)^{-1}\mathbf{d}_{n0}$. The $U_{nt} = \lambda_0 W_n U_{nt} + V_{nt}$ forms a SAR process. This model implies spatial correlations in both the individual and disturbance components, $\mathbf{c}_{n0}$ and U_{nt}, having the same spatial effect parameter. Baltagi et al. (2007) formulate a model which allows for different spatial effects in both individual and disturbance components. Baltagi et al. (2003, 2007) have emphasized the test of spatial correlation in their models.

We note that, with the fixed effects specification, all these panel models have the same representation. By regarding $(I_n - \lambda_0 W_n)^{-1}\mathbf{d}_{n0}$ as a vector of unknown fixed effect parameters, the two equations are identical to a linear panel regression with fixed effects and SAR disturbances. In this paper, we consider the estimation of the SAR panel model with both spatial lag and spatial disturbances. For the model with individual effects, we consider the case where n is large but T can be finite or tends to infinity. For the model with both individual and time effects, we focus on the scenario with both n and T being large.[6]

This paper is organized as follows. In Section 2, the SAR panel model with individual fixed effects is introduced. We consider, first, the direct ML approach where the individual effects are also estimated. We find that when T is finite, the estimate of the variance parameter is inconsistent, but the estimates of the other common parameters are consistent and asymptotically normal. As an alternative estimation method, we propose a data transformation procedure, and establish the consistency and asymptotic distribution of the QML estimator of that approach. We demonstrate that the estimates (except the variance parameter) from the direct approach are identical to the corresponding estimates from the transformation approach. These results extend those of the within estimates and the conditional likelihood estimation of the linear panel regression model to the SAR panel model. Section 3 generalizes the model to include both individual and time effects. For the direct ML approach, even both n and T are large, the distribution of the estimates is not properly centered. The noncentrality can, however, be removed by some bias-correction procedure. On the contrary, the transformation approach will yield consistent estimates as long as either n or T are large, and their asymptotic distributions are normal and properly centered. For the model with both effects, the likelihood function from the transformation approach is not necessarily a conditional likelihood, but a partial likelihood function instead. It may also be justified as a certain modified likelihood function. Simulation results are reported in Section 4 to compare the two approaches. Section 5 concludes the paper. Proofs are collected in the Appendix.

2. The model with individual effects only

The SAR panel model with individual effects and SAR disturbances is

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

where $Y_{nt} = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ and $V_{nt} = (v_{1t}, v_{2t}, \ldots, v_{nt})'$ are $n \times 1$ vectors and v_{it} is *i.i.d.* across i and t with zero mean and variance σ_0^2. W_n is an $n \times n$ nonstochastic spatial weights matrix that generates the spatial dependence on y_{it} among cross sectional units. X_{nt} is an $n \times k$ matrix of nonstochastic time varying regressors, and $\mathbf{c}_{n0}$ is an $n \times 1$ vector of fixed effects. Similarly, M_n is an $n \times n$ spatial weights matrix for the disturbances. In practice, M_n may or may not be W_n.

In this paper, we consider, first, the estimation of the parameters including the fixed effects, and investigate the possible incidental parameter issue. We then consider the estimation after the elimination of the fixed effects. Define $S_n(\lambda) = I_n - \lambda W_n$ and $R_n(\rho) = I_n - \rho M_n$ for any λ and ρ. At the true parameter, $S_n = S_n(\lambda_0)$ and $R_n = R_n(\rho_0)$. Then, presuming S_n and R_n are invertible, (1) can be rewritten as

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

For notational purposes, we define $\tilde{Y}_{nt} = Y_{nt} - \bar{Y}_{nT}$ for $t = 1, 2, \ldots, T$, where $\bar{Y}_{nT} = \frac{1}{T}\sum_{t=1}^{T} Y_{nt}$. Similarly, $\tilde{X}_{nt} = X_{nt} - \bar{X}_{nT}$ and $\tilde{V}_{nt} = V_{nt} - \bar{V}_{nT}$. A list of frequently used notations is provided in Appendix A for easy reference. For our asymptotic analysis of the estimators, we make the following assumptions.

Assumption 1. *W_n and M_n are nonstochastic spatial weights matrices with zero diagonals.*

[3] Sufficient statistics might not be available for many other models. A well-known example is the probit panel regression model, even though probit and logit models are close substitutes (see Chamberlain, 1982). In addition to the conditional likelihood method, other methods to eliminate nuisance parameters have been discussed in Kalbfleisch and Sprott (1970), Cox and Reid (1987) and Lancaster (2000) among others.

[4] The use of the transformation from $(I_n - \frac{1}{n} l_n l_n')$ to eliminate time fixed effects has been considered in Lee and Yu (forthcoming) for a spatial dynamic panel model with large T. In a group setting with group fixed effects, a similar transformation can eliminate the group effects (Lee et al., 2008).

[5] However, our modified likelihood is not one of those which could be constructed from their formulas.

[6] We may point out, in some occasions, the implication of either n or T being finite. For a SAR model, because spatial interactions are highly parameterized, it is of interest only when n is large. Otherwise, a vector autoregression model would be preferable. For this reason, as suggested by a referee, we focus our attention on n being large.

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

Assumption 3. $S_n(\lambda)$ and $R_n(\rho)$ are invertible for all $\lambda \in \Lambda$ and $\rho \in \mathbb{P}$, where Λ and $\mathbb{P}$ are compact intervals. Furthermore, λ_0 is in the interior of Λ, and ρ_0 is in the interior of $\mathbb{P}$.[7]

Assumption 4. The elements of X_{nt} are nonstochastic and bounded,[8] uniformly in n and t. Also, under the asymptotic setting in Assumption 6, the limit of $\frac{1}{nT}\sum_{t=1}^{T}\tilde{X}'_{nt}R'_nR_n\tilde{X}_{nt}$ exists and is nonsingular.

Assumption 5. W_n and M_n are uniformly bounded in both row and column sums in absolute value (for short, UB).[9] Also $S_n^{-1}(\lambda)$ and $R_n^{-1}(\rho)$ are UB, [10] uniformly in $\lambda \in \Lambda$ and $\rho \in \mathbb{P}$.

Assumption 6. n is large, where T can be finite or large.[11]

The zero diagonal assumption helps the interpretation of the spatial effect, as self-influence shall be excluded in practice. In many empirical applications, each of the rows of W_n (and M_n) sums to 1, which ensures that all the weights are between 0 and 1. In this section, our estimation and analysis for the model do not require the feature of row-normalization. Assumption 2 provides *i.i.d.* regularity assumptions for v_{it}. We note that the disturbances in U_{nt} are allowed to be spatially correlated. It is the noise term in U_{nt} that are *i.i.d.* distributed. If there is unknown heteroskedasticity, the MLE (QMLE) will not be consistent. Methods such as the GMM in Lin and Lee (forthcoming) and that in Kelejian and Prucha (forthcoming) may be designed for that situation. Invertibility of $S_n(\lambda)$ and $R_n(\rho)$ in Assumption 3 guarantees that (2) is valid. Also, compactness is a condition for theoretical analysis on nonlinear functions. When W_n is row-normalized, a compact subset of $(-1, 1)$ has often been taken as the parameter space for λ in theory. So is the parameter space of ρ for a row-normalized M_n. When exogenous variables X_{nt} are included in the model, it is convenient to assume that they are uniformly bounded as in Assumption 4. Assumption 5 is originated by Kelejian and Prucha (1998, 2001) and also used in Lee (2004, 2007a). That W_n, M_n, $S_n^{-1}(\lambda)$ and $R_n^{-1}(\rho)$ are UB is a condition that limits the spatial correlation to a manageable degree. Assumption 6 allows two cases of interest: (i) both n and T are large; and (ii) n is large and T is fixed. For (ii), we are interested in the short panel data case in contrast to the case where T is large in other studies, e.g., Hahn and Kuersteiner (2002) and Yu et al. (2008).

2.1. The direct approach

Denote $\theta = (\beta', \lambda, \rho, \sigma^2)'$ and $\zeta = (\beta', \lambda, \rho)'$. At the true value, $\theta_0 = (\beta'_0, \lambda_0, \rho_0, \sigma_0^2)'$ and $\zeta_0 = (\beta'_0, \lambda_0, \rho_0)'$. The log likelihood function of (1), as if the disturbances were normally distributed, is

$$\ln L^d_{n,T}(\theta, \mathbf{c}_n) = -\frac{nT}{2}\ln(2\pi\sigma^2) + T\left[\ln|S_n(\lambda)| + \ln|R_n(\rho)|\right] - \frac{1}{2\sigma^2}\sum_{t=1}^{T}V'_{nt}(\zeta, \mathbf{c}_n)V_{nt}(\zeta, \mathbf{c}_n), \quad (3)$$

where $V_{nt}(\zeta, \mathbf{c}_n) = R_n(\rho)[S_n(\lambda)Y_{nt} - X_{nt}\beta - \mathbf{c}_n]$. We can estimate $\mathbf{c}_n$ directly and have the estimator of θ_0 via a concentrated log likelihood with $\mathbf{c}_n$ concentrated out:

$$\ln L^d_{n,T}(\theta) = -\frac{nT}{2}\ln(2\pi\sigma^2) + T\left[\ln|S_n(\lambda)| + \ln|R_n(\rho)|\right] - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta), \quad (4)$$

where $\tilde{V}_{nt}(\zeta) = R_n(\rho)[S_n(\lambda)\tilde{Y}_{nt} - \tilde{X}_{nt}\beta]$. The first and second order derivatives of (4) are (36) and (37) in Appendix B.

2.2. Transformation approach

To eliminate the individual effects, the deviation from the time mean operator, $J_T = (I_T - \frac{1}{T}l_Tl'_T)$, can be used. Because W_n is time invariant, the variables in the deviation form would still be a SAR model. Such a transformed model consists of $\tilde{Y}_{nt} = \lambda_0W_n\tilde{Y}_{nt} + \tilde{X}_{nt}\beta_0 + \tilde{U}_{nt}$ and $\tilde{U}_{nt} = \rho_0M_n\tilde{U}_{nt} + \tilde{V}_{nt}$. However, the resulting disturbances $\tilde{V}_{nt}$ would be linearly dependent over the time dimension. Without creating linear dependence in the resulting disturbances, a corresponding transformation can be based on the orthonormal eigenvector matrix of J_T. We use an orthogonal transformation which includes the Helmert transformation as a special case to eliminate the fixed effects. Let $[F_{T,T-1}, \frac{1}{\sqrt{T}}l_T]$ be the orthonormal eigenvector matrix of J_T, where $F_{T,T-1}$ is the $T \times (T-1)$ submatrix[12] corresponding to the eigenvalues of one. For any $n \times T$ matrix $[Z_{n1}, \ldots, Z_{nT}]$, define the transformed $n \times (T-1)$ matrix $[Z^*_{n1}, \ldots, Z^*_{n,T-1}] = [Z_{n1}, \ldots, Z_{nT}]F_{T,T-1}$. Similarly, $X^*_{nt} = [X^*_{nt,1}, X^*_{nt,2}, \ldots, X^*_{nt,k}]$. Then, (1) implies

$$Y^*_{nt} = \lambda_0W_nY^*_{nt} + X^*_{nt}\beta_0 + U^*_{nt},$$
$$U^*_{nt} = \rho_0M_nU^*_{nt} + V^*_{nt}, \; t = 1, \ldots, T-1. \quad (5)$$

Because $(V^{*\prime}_{n1}, \ldots, V^{*\prime}_{n,T-1})' = (F'_{T,T-1} \otimes I_n)(V'_{n1}, \ldots, V'_{nT})'$ and v_{it}'s are *i.i.d*,

$$E(V^{*\prime}_{n1}, \ldots, V^{*\prime}_{n,T-1})'(V^{*\prime}_{n1}, \ldots, V^{*\prime}_{n,T-1}) = \sigma_0^2(F'_{T,T-1} \otimes I_n)(F_{T,T-1} \otimes I_n) = \sigma_0^2I_{n(T-1)}.$$

Hence, v^*_{it}'s are uncorrelated for all i and t (and independent under normality), where v^*_{it} is the ith element of V^*_{nt}.

The log likelihood function of (5), as if the disturbances were normally distributed, is

$$\ln L_{n,T}(\theta) = -\frac{n(T-1)}{2}\ln(2\pi\sigma^2) + (T-1)[\ln|S_n(\lambda)| + \ln|R_n(\rho)|] - \frac{1}{2\sigma^2}\sum_{t=1}^{T-1}V^{*\prime}_{nt}(\zeta)V^*_{nt}(\zeta), \quad (6)$$

where $V^*_{nt}(\zeta) = R_n(\rho)[S_n(\lambda)Y^*_{nt} - X^*_{nt}\beta]$. For any n-dimensional column vectors p_{nt} and q_{nt}, as

---

7 Due to the nonlinearity of λ and ρ in the reduced form of the model, compactness of Λ and $\mathbb{P}$ is needed. However, the compactness of β and σ^2 is not necessary because the β and σ^2 estimates given λ and ρ are least squares type estimates.

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

9 We say a (sequence of $n \times n$) matrix P_n is uniformly bounded in row and column sums in absolute value if $\sup_{n\geq1}\|P_n\|_\infty < \infty$ and $\sup_{n\geq1}\|P_n\|_1 < \infty$, where $\|P_n\|_\infty = \sup_{1\leq i\leq n}\sum_{j=1}^{n}|p_{ij,n}|$ is the row sum norm and $\|P_n\|_1 = \sup_{1\leq j\leq n}\sum_{i=1}^{n}|p_{ij,n}|$ is the column sum norm.

10 This assumption has effectively ruled out some cases, and, hence, imposed limited dependence across spatial units. For example, if $\lambda_{0n} = 1 - 1/n$ under $n \to \infty$, it is a near unit root case for a cross sectional SAR model and S_n^{-1} will not be UB (see Lee and Yu, 2007).

11 The case with a finite n and large T is of less interest as the incidental parameter problem does not occur in this model.

12 In dynamic panel data, the first difference and Helmert transformation have often been used to eliminate the individual effects (see Anderson and Hsiao, 1981; Arellano and Bover, 1995 among others). A special selection of $F_{T,T-1}$ gives rise to the Helmert transformation where V_{nt} is transformed to $(\frac{T-t}{T-t+1})^{1/2}[V_{nt} - \frac{1}{T-t}(V_{n,t+1} + \cdots + V_{nT})]$, which is of particular interest for dynamic panel data models.

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

by using $(\tilde{p}_{n1}, \ldots, \tilde{p}_{nT}) = (p_{n1}, \ldots, p_{nT})J_T$, (6) can be rewritten as

$$\ln L_{n,T}(\theta) = -\frac{n(T-1)}{2}\ln(2\pi\sigma^2) + (T-1)[\ln|S_n(\lambda)| + \ln|R_n(\rho)|] - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta), \tag{7}$$

where $\tilde{V}_{nt}(\zeta) = R_n(\rho)[S_n(\lambda)\tilde{Y}_{nt} - \tilde{X}_{nt}\beta]$ and its first and second order derivatives are (40) and (41) in Appendix B.

We note that the likelihood function in (7) has a conditional likelihood interpretation. It is the likelihood conditional on the time average $\bar{Y}_{nT}$, which is a sufficient statistic for $\mathbf{c}_{n0}$ under normality. This is because (1) implies that $\bar{Y}_{nT} = \lambda_0 W_n \bar{Y}_{nT} + \bar{X}_{nT}\beta_0 + \mathbf{c}_{n0} + \bar{U}_{nT}$ with $\bar{U}_{nT} = \rho_0 M_n \bar{U}_{nT} + \bar{V}_{nT}$, but $\mathbf{c}_{n0}$ does not appear in $\tilde{Y}_{nt} = \lambda_0 W_n \tilde{Y}_{nt} + \tilde{X}_{nt}\beta_0 + \tilde{U}_{nt}$ with $\tilde{U}_{nt} = \rho_0 M_n \tilde{U}_{nt} + \tilde{V}_{nt}$. As $\tilde{V}_{nt}, t = 1, \ldots, T$, are independent of $\bar{V}_{nT}$ under normality, the likelihood in (7) is the conditional likelihood of $Y_{nt}, t = 1, \ldots, T$ conditional on $\bar{Y}_{nT}$ (Hsiao, 1986; Lancaster, 2000).

2.3. Comparison of the two approaches

One may compare the concentrated log likelihood function in (4) of the direct approach with the one in (7) of the transformation approach. We see that the difference is on the use of T in (4) but $(T-1)$ in (7). A closer comparison of the two log likelihoods with a further concentration is revealing.

For (4), we can further concentrate out β and σ^2 and focus on (λ, ρ). The QMLEs of β and σ^2 given λ and ρ are

$$\hat{\beta}^d_{nT}(\lambda, \rho) = \left[\sum_{t=1}^{T}\tilde{X}'_{nt}R'_n(\rho)R_n(\rho)\tilde{X}_{nt}\right]^{-1} \times \left[\sum_{t=1}^{T}\tilde{X}'_{nt}R'_n(\rho)R_n(\rho)S_n(\lambda)\tilde{Y}_{nt}\right], \tag{8}$$

$$\hat{\sigma}^{2d}_{nT}(\lambda, \rho) = \frac{1}{nT}\sum_{t=1}^{T}[S_n(\lambda)\tilde{Y}_{nt} - \tilde{X}_{nt}\hat{\beta}^d_{nT}(\lambda, \rho)]' \times R'_n(\rho)R_n(\rho)[S_n(\lambda)\tilde{Y}_{nt} - \tilde{X}_{nt}\hat{\beta}^d_{nT}(\lambda, \rho)]. \tag{9}$$

The concentrated log likelihood function of (λ, ρ) of the direct approach is

$$\ln L^d_{n,T}(\lambda, \rho) = -\frac{nT}{2}(\ln(2\pi) + 1) - \frac{nT}{2}\ln\hat{\sigma}^{2d}_{nT}(\lambda, \rho) + T[\ln|S_n(\lambda)| + \ln|R_n(\rho)|]. \tag{10}$$

For (7), the corresponding estimates are

$$\hat{\beta}_{nT}(\lambda, \rho) = \left[\sum_{t=1}^{T}\tilde{X}'_{nt}R'_n(\rho)R_n(\rho)\tilde{X}_{nt}\right]^{-1} \times \left[\sum_{t=1}^{T}\tilde{X}'_{nt}R'_n(\rho)R_n(\rho)S_n(\lambda)\tilde{Y}_{nt}\right], \tag{11}$$

$$\hat{\sigma}^2_{nT}(\lambda, \rho) = \frac{1}{n(T-1)}\sum_{t=1}^{T}[S_n(\lambda)\tilde{Y}_{nt} - \tilde{X}_{nt}\hat{\beta}_{nT}(\lambda, \rho)]' \times R'_n(\rho)R_n(\rho)[S_n(\lambda)\tilde{Y}_{nt} - \tilde{X}_{nt}\hat{\beta}_{nT}(\lambda, \rho)], \tag{12}$$

and the concentrated log likelihood function of (λ, ρ) for the transformation approach is

$$\ln L_{n,T}(\lambda, \rho) = -\frac{n(T-1)}{2}(\ln(2\pi) + 1) - \frac{n(T-1)}{2}\ln\hat{\sigma}^2_{nT}(\lambda, \rho) + (T-1)[\ln|S_n(\lambda)| + \ln|R_n(\rho)|]. \tag{13}$$

Note that $\hat{\beta}_{nT}(\lambda, \rho) = \hat{\beta}^d_{nT}(\lambda, \rho)$, but $\hat{\sigma}^{2d}_{nT}(\lambda, \rho) = \frac{T-1}{T}\hat{\sigma}^2_{nT}(\lambda, \rho)$. Hence, (10) can be rewritten as

$$\ln L^d_{n,T}(\lambda, \rho) = -\frac{nT}{2}\left(\ln(2\pi) + \ln\frac{T-1}{T} + 1\right) - \frac{nT}{2}\ln\hat{\sigma}^2_{nT}(\lambda, \rho) + T[\ln|S_n(\lambda)| + \ln|R_n(\rho)|]. \tag{14}$$

By comparing (13) and (14), we see that they yield the same maximizer $(\hat{\lambda}_{nT}, \hat{\rho}_{nT})$. As $\hat{\beta}^d_{nT}(\lambda, \rho)$ and $\hat{\beta}_{nT}(\lambda, \rho)$ are the same, the QMLE of $\zeta_0 = (\beta'_0, \lambda_0, \rho_0)'$ from the direct approach is the same as that of the transformation approach. However, the estimate of σ^2_0 from the direct approach will not be consistent unless T is large, which can be seen from the difference of $\hat{\sigma}^{2d}_{nT}(\lambda, \rho)$ and $\hat{\sigma}^2_{nT}(\lambda, \rho)$. As $\hat{\sigma}^{2d}_{nT}(\lambda, \rho) = \frac{T-1}{T}\hat{\sigma}^2_{nT}(\lambda, \rho)$, we see that the bias corrected estimate $\frac{T}{T-1}\hat{\sigma}^{2d}_{nT}$ is numerically equivalent to $\hat{\sigma}^2_{nT}(\lambda, \rho)$. Hence, the ML estimation of the SAR panel model with fixed individual effects shares some common features with the ML estimation of the fixed effects linear panel regression model. The concentrated log likelihood function in (13) or (14) provides a common ground for the investigation of asymptotic properties of the two approaches. It also provides computational simplicity in terms of reduced dimension for optimization.

2.4. Consistency and asymptotic distributions of estimates

Denote $G_n = W_n S_n^{-1}$ and

$$\mathcal{H}_{nT}(\rho) = \frac{1}{n(T-1)}\sum_{t=1}^{T}(\tilde{X}_{nt}, G_n\tilde{X}_{nt}\beta_0)' \times R'_n(\rho)R_n(\rho)(\tilde{X}_{nt}, G_n\tilde{X}_{nt}\beta_0), \tag{15}$$

$$\sigma^2_n(\rho) = \frac{\sigma^2_0}{n}tr[(R_n(\rho)R_n^{-1})'(R_n(\rho)R_n^{-1})],$$

$$\sigma^2_n(\lambda, \rho) = \frac{\sigma^2_0}{n}tr[(R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1})'(R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1})].$$

Assumption 7. Either (a) the limit of $\mathcal{H}_{nT}(\rho)$ is nonsingular for each possible ρ in $\mathbb{P}$, and the limit of $(\frac{1}{n}\ln|\sigma^2_0 R_n^{-1\prime}R_n^{-1}| - \frac{1}{n}\ln|\sigma^2_n(\rho)R_n^{-1}(\rho)'R_n^{-1}(\rho)|)$ is not zero for $\rho \neq \rho_0$; or (b) the limit of

$$\left(\frac{1}{n}\ln|\sigma^2_0 R_n^{-1\prime}S_n^{-1\prime}S_n^{-1}R_n^{-1}| - \frac{1}{n}\ln|\sigma^2_n(\lambda, \rho)R_n^{-1}(\rho)'S_n^{-1}(\lambda)'S_n^{-1}(\lambda)R_n^{-1}(\rho)|\right)$$

is not zero for $(\lambda, \rho) \neq (\lambda_0, \rho_0)$, as n tends to infinity.

Assumption 7 states the identification conditions of the model, which generalizes those for a cross section SAR model in Lee and Liu (2006) to the panel case. The part (a) of Assumption 7 represents the possible identification of λ_0 and β_0 through the deterministic part of the reduced form equation of (1), and the identification of ρ_0 and σ^2_0 from the SAR process of U_{nt} in (1). The part (b) of Assumption 7 provides identification through the SAR process of the reduced form of disturbances of Y_{nt} in (2). When $M_n = W_n$ and $\lambda_0 \neq \rho_0$, the condition in 7(b) would not be satisfied as (λ_0, ρ_0) and (ρ_0, λ_0) could not be distinguished from each other. Identification will then rely on either Assumption 7(a) or extra information on the order of magnitudes of λ_0 and ρ_0. The identification and consistency are shown in the following theorem. The analysis follows from the concentrated likelihood (10) or (13) for λ_0 and ρ_0. Those of β_0 and σ^2_0 for the direct and transformation approaches follow, respectively, from (8)–(9) and (11)–(12).

Theorem 1. *Under Assumptions 1–7, θ_0 is identified. Furthermore,*

(1) *for the QMLE $\hat{\theta}^d_{nT}$ based on (8)–(10) of the direct approach, $\hat{\theta}^d_{nT} - \theta_T \xrightarrow{p} 0$ where $\theta_T = \theta_0 - (\mathbf{0}_{1\times(k+2)}, \frac{1}{T}\sigma_0^2)'$;*

(2) *for the QMLE $\hat{\theta}_{nT}$ based on (11) and (13) of the transformation approach, $\hat{\theta}_{nT} - \theta_0 \xrightarrow{p} 0$.*

Proof. See Appendix B.3. ■

For this theorem, $\hat{\sigma}^{2d}_{nT}$ does not converge to σ_0^2 when T is a fixed finite value as n tends to infinity. It will be consistent only when T is large. The $(\hat{\beta}^{d\prime}_{nT}, \hat{\lambda}^d_{nT}, \hat{\rho}^d_{nT})'$ of $\hat{\theta}^d_{nT}$ will be consistent even when T is small, because they are identical to those of the transformation approach. For the $\hat{\sigma}^2_{nT}$ of the transformation approach, it is a consistent estimate of σ_0^2 as long as n tends to infinity.

The asymptotic distribution of $\hat{\theta}^d_{nT}$ can be derived from the Taylor expansion of (38) around θ_T, and $\hat{\theta}_{nT}$ can be derived accordingly with (42).[13] Denote $C_n = \ddot{G}_n - \frac{tr\ddot{G}_n}{n}I_n$ and $D_n = H_n - \frac{trH_n}{n}I_n$ where $\ddot{G}_n = R_nG_nR_n^{-1}$ and $H_n = M_nR_n^{-1}$.

Assumption 8. The limit of $\frac{1}{n^2}\left[tr(C_n^sC_n^s)tr(D_n^sD_n^s) - tr^2(C_n^sD_n^s)\right]$ is strictly positive as n tends to infinity.

The first order derivative of the concentrated log likelihood function at the true parameters involves both linear and quadratic functions of $\tilde{V}_{nt}$. Its asymptotic distribution can be derived from a central limit theorem for martingale difference arrays (see Lemma A.1 in Appendix A). Assumption 8 is a condition for the nonsingularity of the limits of the information matrices of both approaches. When the limit of $\mathcal{H}_{nT}(\rho_0)$ is singular, as long as the limit of $\frac{1}{n^2}\left[tr(C_n^sC_n^s)tr(D_n^sD_n^s) - tr^2(C_n^sD_n^s)\right]$ is strictly positive, the limits of the information matrices remain nonsingular.

Theorem 2. *Under Assumptions 1–7 (a); or Assumptions 1–6, 7 (b) and 8,*

(1) *for the direct approach,*

$$\sqrt{nT}(\hat{\theta}^d_{nT} - \theta_T) \xrightarrow{d} N\Bigg(0, \lim \frac{T}{T-1}(\Sigma^d_{\theta_T,nT})^{-1} \times (\Sigma^d_{\theta_T,nT} + \Omega^d_{\theta_T,nT})(\Sigma^d_{\theta_T,nT})^{-1}\Bigg), \tag{16}$$

where the lim *is taken under Assumption 6, and $\Sigma^d_{\theta_T,nT}$, $\Omega^d_{\theta_T,nT}$ are in* (39) and (48)

(2) *for the transformation approach,*

$$\sqrt{n(T-1)}(\hat{\theta}_{nT} - \theta_0) \xrightarrow{d} N(0, \lim \Sigma^{-1}_{\theta_0,nT} \times (\Sigma_{\theta_0,nT} + \Omega_{\theta_0,nT})\Sigma^{-1}_{\theta_0,nT}), \tag{17}$$

where $\Sigma_{\theta_0,nT}$ and $\Omega_{\theta_0,nT}$ are in (43) and (49).

Proof. See Appendix B.4. ■

The $(\hat{\beta}^{d\prime}_{nT}, \hat{\lambda}^d_{nT}, \hat{\rho}^d_{nT})$ is identical to $(\hat{\beta}'_{nT}, \hat{\lambda}_{nT}, \hat{\rho}_{nT})$, and they are properly centered at their true parameter values. But for the estimate of σ_0^2 of the direct approach, $\sqrt{nT}(\hat{\sigma}^{2d}_{nT} - \sigma_0^2)$ may not be centered at 0, even though T also tends to infinity, unless $\frac{n}{T}$ goes to zero. On the other hand, for the transformed approach, $\sqrt{n(T-1)}(\hat{\sigma}^2_{nT} - \sigma_0^2)$ is properly centered at 0 even with a finite T. When V_{nt} are normally distributed, $\Omega^d_{\theta_T,nT} = 0$ and $\Omega_{\theta_0,nT} = 0$ because $\mu_4 - 3\sigma_0^4 = 0$. The difference between $\Sigma^d_{\theta_T,nT}$ and $\Sigma_{\theta_0,nT}$ (resp. $\Omega^d_{\theta_0,nT}$ and $\Omega_{\theta_T,nT}$) occurs at the corresponding elements associated with σ_T^2 and σ_0^2, as shown in (39) and (43) (resp. (48) and (49)). From Theorem 1, it is straightforward to construct the bias corrected estimates for the direct approach as

$$\hat{\theta}^{d1}_{nT} = \left(\hat{\beta}^{d\prime}_{nT}, \hat{\lambda}^d_{nT}, \hat{\rho}^d_{nT}, \frac{T}{T-1}\hat{\sigma}^{2d}_{nT}\right)'. \tag{18}$$

This bias corrected estimate is numerically the same as the estimate of the transformation approach.

In some social interaction models, if each unit has many neighbors, the QMLEs of some parameters in $(\beta_0', \lambda_0, \rho_0)'$ might have a lower rate of convergence. For the cross section SAR model with *i.i.d.* disturbances, Lee (2004) shows that when $G_nX_n\beta_0$ is asymptotically multicollinear with X_n, the information matrix is asymptotically singular and the MLEs of β_0 and λ_0 will have a lower rate of convergence. Only when $G_nX_n\beta_0$ is not multicollinear with X_n would the rate of convergence be regular $\sqrt{n}$ under the "many neighbors" setting. In the SAR panel data with SAR disturbances, we might have similar findings. Namely, when $\mathcal{H}_{nT}$ is singular, the estimates of β_0 and λ_0 will have a lower rate of convergence under the "many neighbors" setting. When $\mathcal{H}_{nT}$ is nonsingular, the QMLEs of (β_0', λ_0) have the regular rate. However, the rate of the MLE of ρ_0 would be lower under the "many neighbors" setting, regardless of the singularity of $\mathcal{H}_{nT}$ or not. Hence, the result in Lee (2004) would carry over to (β_0', λ_0), while the rate of the MLE of ρ_0 would always be lower.[14]

3. A general model with both individual and time effects

Both Baltagi et al. (2003) and Kapoor et al. (2007) focus on models with only individual effects. In the panel data literature, there are also two-way error component regression models where we have not only individual effects but also time effects (See Wallace and Hussain, 1969; Amemiya, 1971; Nerlove, 1971; Baltagi, 1995; Hahn and Moon, 2006, etc). The time effects might be important, for example, in growth theory and regional economics (see, e.g., Ertur and Koch, 2007 and Foote, 2007). Hence, we generalize (1) to

$$Y_{nt} = \lambda_0W_nY_{nt} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0}l_n + U_{nt},$$
$$U_{nt} = \rho_0M_nU_{nt} + V_{nt}, t = 1, 2, \ldots, T, \tag{19}$$

where α_{t0} is the fixed time effect. From a methodological point of view, the asymptotics are of interest only when both n and T tend to infinity.[15] When T tends to infinity, the time effects may cause the incidental parameter problem in addition to the individual effects. In the following sections, we consider the direct QML approach which estimates both the individual and time effects, and a transformation approach where both the individual and time effects are eliminated. For the transformation approach, we may first eliminate the individual effects in (19) by $F_{T,T-1}$ similar to (5), which yields

$$Y^*_{nt} = \lambda_0W_nY^*_{nt} + X^*_{nt}\beta_0 + \alpha^*_{t0}l_n + U^*_{nt},$$
$$U^*_{nt} = \rho_0M_nU^*_{nt} + V^*_{nt}, t = 1, 2, \ldots, T-1, \tag{20}$$

[13] Those can also be derived with the concentrated likelihood in (10) or (13) with the corresponding estimates of β_0 and σ_0^2 via (8)–(9) and (11)–(12). However, while these can be convenient for the marginal distributions for the estimates of β_0 and σ_0^2, it is algebraically tedious and is indirect to obtain the joint distributions. The derivations via the log likelihoods of (38) or (42) are more direct.

[14] We do not provide a rigorous analysis of this "many neighbors" case in this paper. However, by investigating the elements of information matrix of (10) or (13), we can infer the rates of convergence for the QMLEs of (λ_0, ρ_0), and hence the rates for the QMLEs of β_0 and σ_0^2. The "many neighbors" case is of special interest in social interaction models. One may have a deeper understanding of that model with the approach in Lee (2007b) via a group setting.

[15] If T is finite, the time effects can be regarded as a finite number of additional regression coefficients similar to the role of β.

where $[\alpha_{10}^* l_n, \alpha_{20}^* l_n, \ldots, \alpha_{T-1,0}^* l_n] = [\alpha_{10} l_n, \alpha_{20} l_n, \ldots, \alpha_{T0} l_n] F_{T,T-1}$ are transformed time effects. We can further transform (20) to eliminate the time effects.

3.1. Direct approach

The log likelihood function of (19) with both $\mathbf{c}_n$ and $(\alpha_1, \ldots, \alpha_T)'$ concentrated out is

$$\ln L_{n,T}^d(\theta) = -\frac{nT}{2}\ln(2\pi\sigma^2) + T[\ln|S_n(\lambda)| + \ln|R_n(\rho)|] - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{V}_{nt}'(\zeta)J_n\tilde{V}_{nt}(\zeta), \quad (21)$$

where $\tilde{V}_{nt}(\zeta) = R_n(\rho)[S_n(\lambda)\tilde{Y}_{nt} - \tilde{X}_{nt}\beta]$ and $J_n = I_n - \frac{1}{n}l_n l_n'$ is the deviation from the group mean transformation over spatial units. The first and second order derivatives of (21) are, respectively, (50) and (51) in Appendix C.

3.2. Transformation approach

In a panel regression model with both individual and time effects, these effects can be eliminated by taking deviations from time and cross section means. For example, for y_{it}, denote $y_{i.} = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $y_{.t} = \frac{1}{n}\sum_{i=1}^{n} y_{it}$ and $y_{..} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} y_{it}$. The within estimator of β in the panel regression model is to regress $y_{it} - y_{.t} - y_{i.} + y_{..}$ on $x_{it} - x_{.t} - x_{i.} + x_{..}$ (see, e.g., Wallace and Hussain, 1969; Baltagi, 1995). The within estimator is a conditional MLE of y_{it}'s conditional on all $y_{i.}$ and $y_{.t}$. In terms of matrices, these transformations correspond to J_T and J_n. With W_n and M_n being row normalized, $J_nW_nJ_n = J_nW_n$ and $J_nM_nJ_n = J_nM_n$. Using these transformations for (19), we have $J_n\tilde{Y}_{nt} = \lambda_0 J_nW_nJ_n\tilde{Y}_{nt} + J_n\tilde{X}_{nt}\beta_0 + J_n\tilde{U}_{nt}$ with $J_n\tilde{U}_{nt} = \rho_0 J_nM_nJ_n\tilde{U}_{nt} + J_n\tilde{V}_{nt}$. The elements of $J_n\tilde{Y}_{nt}$, etc., are in the deviation form from both individual and time means. This transformed equation is in the form of a SAR model without individual or time effects. The parameters can then be estimated from this equation.

Without creating linear dependence on the resulting disturbances, the transformations can be based on the orthonormal eigenvector matrices of J_T and J_n. Let $(F_{n,n-1}, \frac{1}{\sqrt{n}}l_n)$ be the orthonormal eigenvector matrix of J_n, where $F_{n,n-1}$ is the $n \times (n-1)$ submatrix corresponding to the eigenvalues of one. Similar to Lee and Yu (forthcoming), we can further transform the n-dimensional vector Y_{nt}^* in (20) to an $(n-1)$-dimensional vector Y_{nt}^{**} such that $Y_{nt}^{**} = F_{n,n-1}'Y_{nt}^*$. For this transformation approach and a likelihood estimation, we need W_n and M_n to be row normalized.[16]

Assumption 1′. W_n and M_n are row normalized nonstochastic spatial weights matrices with zero diagonals.

With W_n and M_n being row normalized, (20) can be transformed into

$$Y_{nt}^{**} = \lambda_0(F_{n,n-1}'W_nF_{n,n-1})Y_{nt}^{**} + X_{nt}^{**}\beta_0 + U_{nt}^{**},$$
$$U_{nt}^{**} = \rho_0(F_{n,n-1}'M_nF_{n,n-1})U_{nt}^{**} + V_{nt}^{**}, \quad (22)$$

for $t = 1, \ldots, T-1$ where $X_{nt}^{**} = F_{n,n-1}'X_{nt}^*$ and $V_{nt}^{**} = F_{n,n-1}'V_{nt}^*$. After the transformations, the effective sample size is now $(n-1)(T-1)$. Because $(V_{n1}^{**\prime}, \ldots, V_{n,T-1}^{**\prime})' = (I_{T-1} \otimes F_{n,n-1}')(V_{n1}^{*\prime}, \ldots, V_{n,T-1}^{*\prime})' = (F_{T,T-1}' \otimes F_{n,n-1}')(V_{n1}', \ldots, V_{nT}')'$, we have

$$E(V_{n1}^{**\prime}, \ldots, V_{n,T-1}^{**\prime})'(V_{n1}^{**\prime}, \ldots, V_{n,T-1}^{**}) = \sigma_0^2(F_{T,T-1}' \otimes F_{n,n-1}')(F_{T,T-1} \otimes F_{n,n-1}) = \sigma_0^2(I_{T-1} \otimes I_{n-1}).$$

Hence, the elements v_{it}^{**}'s of V_{nt}^{**} are uncorrelated for all i and t.

The log likelihood function for (22) is

$$\ln L_{n,T}(\theta) = -\frac{(n-1)(T-1)}{2}\ln(2\pi\sigma^2) + (T-1)\ln\left|I_{n-1} - \lambda F_{n,n-1}'W_nF_{n,n-1}\right| + (T-1)\ln\left|I_{n-1} - \rho F_{n,n-1}'M_nF_{n,n-1}\right| - \frac{1}{2\sigma^2}\sum_{t=1}^{T-1}V_{nt}^{**\prime}(\zeta)V_{nt}^{**}(\zeta), \quad (23)$$

where $V_{nt}^{**}(\zeta) = R_n^*(\rho)[(I_{n-1} - \lambda F_{n,n-1}'W_nF_{n,n-1})Y_{nt}^{**} - X_{nt}^{**}\beta]$ with $R_n^*(\rho) = I_{n-1} - \rho F_{n,n-1}'M_nF_{n,n-1}$. From Lemma A.2 in Appendix A, the determinant and inverse of $(I_{n-1} - \lambda F_{n,n-1}'W_nF_{n,n-1})$ are

$$\left|I_{n-1} - \lambda F_{n,n-1}'W_nF_{n,n-1}\right| = \frac{1}{1-\lambda}\left|I_n - \lambda W_n\right|,$$
$$(I_{n-1} - \lambda F_{n,n-1}'W_nF_{n,n-1})^{-1} = F_{n,n-1}'(I_n - \lambda W_n)^{-1}F_{n,n-1},$$

and similarly for $(I_{n-1} - \rho F_{n,n-1}'M_nF_{n,n-1})$. For any n-dimensional column vectors p_{nt} and q_{nt},

$$\sum_{t=1}^{T-1} p_{nt}^{**\prime}q_{nt}^{**} = (p_{n1}', \ldots, p_{nT}')(F_{T,T-1} \otimes F_{n,n-1}) \times (F_{T,T-1}' \otimes F_{n,n-1}')(q_{n1}', \ldots, q_{nT}')' = (p_{n1}', \ldots, p_{nT}')(J_T \otimes J_n)(q_{n1}', \ldots, q_{nT}')' = \sum_{t=1}^{T}\tilde{p}_{nt}'J_n\tilde{q}_{nt}.$$

This implies that (23) is equal to

$$\ln L_{n,T}(\theta) = -\frac{(n-1)(T-1)}{2}\ln(2\pi\sigma^2) - (T-1)[\ln(1-\lambda) + \ln(1-\rho)] + (T-1)[\ln|S_n(\lambda)| + \ln|R_n(\rho)|] - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{V}_{nt}'(\zeta)J_n\tilde{V}_{nt}(\zeta), \quad (24)$$

and its first and second order derivatives are in (54) and (55) of Appendix C.

We note that this likelihood function is, in general, not necessarily a conditional likelihood for the spatial model, because the sample average over spatial units at each t might not be a sufficient statistic for the time dummy. The cross section average $\frac{1}{n}l_n'W_nY_{nt}$ might not equal $c \cdot y_{.t}$ for some scalar c, unless the column sums of W_n are all equal to 1. In practice, W_n is usually row-normalized but not column-normalized. Conversely, the likelihood in (24) is a partial likelihood function.[17] It may also be regarded as a modification of the concentrated likelihood in (21) as $L_{n,T}(\theta) = L_{n,T}^d(\theta)A_{nT}(\theta)$ where $A_{nT}(\theta) = (2\pi\sigma^2)^{\frac{n+T-1}{2}}\{[(1-\lambda)(1-\rho)]^{T-1}|S_n(\lambda)| \cdot |R_n(\rho)|\}^{-1}$. The factor $A_{nT}(\theta)$ modifies the

[16] When W_n and M_n are not row normalized, we can still eliminate the transformed time effects; however, we will not have the presentation of (22). In that case, a likelihood formulation would not be feasible, and alternative estimation methods, such as the generalized method of moment, would be possible. Such an estimation approach is beyond the scope of this paper.

[17] The density function of $(Y_{n1}, \ldots, Y_{nT})$ of (19) can be decomposed as $f(Y_{n1}, \ldots, Y_{nT}|\theta, \mathbf{c}_n, \alpha_1, \ldots, \alpha_T) = f(Y_{n1}, \ldots, Y_{nT}|Y_{n1}^{**}, \ldots, Y_{n,T-1}^{**}, \theta, \mathbf{c}_n, \alpha_1, \ldots, \alpha_T) \times f(Y_{n1}^{**}, \ldots, Y_{n,T-1}^{**}|\theta)$, where $f(Y_{n1}^{**}, \ldots, Y_{n,T-1}^{**}|\theta)$ is the density of (22).

concentrated likelihood of the direct approach so that the modified likelihood can improve upon the concentrated likelihood function. Various ways to construct a modified likelihood function from the concentrated likelihood of a direct approach are in Cox and Reid (1987), Lancaster (2000), and Arellano and Hahn (2005). The Cox and Reid (1987) and Lancaster (2000) approach involves orthogonal parameterization, which is model specific. The ones in Arellano and Hahn (2005) involve approximations and are related to bias corrected estimation. For our model, the likelihood modification does not seem to relate to theirs, as our intention is not to make a bias correction on the direct estimate. Our approach is motivated by the estimation of the within equation, which relies on the linearity feature of the specified model.[18]

3.3. Comparison of the two approaches

For the direct approach, from (21), we have

$$\hat{\beta}^d_{nT}(\lambda,\rho) = \left[\sum_{t=1}^T \tilde{X}'_{nt}R'_n(\rho)J_nR_n(\rho)\tilde{X}_{nt}\right]^{-1} \times \left[\sum_{t=1}^T \tilde{X}'_{nt}R'_n(\rho)J_nR_n(\rho)S_n(\lambda)\tilde{Y}_{nt}\right], \tag{25}$$

$$\hat{\sigma}^{2d}_{nT}(\lambda,\rho) = \frac{1}{nT}\sum_{t=1}^T[S_n(\lambda)\tilde{Y}_{nt} - \tilde{X}_{nt}\hat{\beta}^d_{nT}(\lambda,\rho)]' \times R'_n(\rho)J_nR_n(\rho)[S_n(\lambda)\tilde{Y}_{nt} - \tilde{X}_{nt}\hat{\beta}^d_{nT}(\lambda,\rho)], \tag{26}$$

and hence, the concentrated log likelihood function of (λ, ρ) is

$$\ln L^d_{n,T}(\lambda,\rho) = -\frac{nT}{2}(\ln(2\pi)+1) - \frac{nT}{2}\ln\hat{\sigma}^{2d}_{nT}(\lambda,\rho) + T[\ln|S_n(\lambda)| + \ln|R_n(\rho)|]. \tag{27}$$

For the transformation approach, from (24), the corresponding estimates are

$$\hat{\beta}_{nT}(\lambda,\rho) = \left[\sum_{t=1}^T \tilde{X}'_{nt}R'_n(\rho)J_nR_n(\rho)\tilde{X}_{nt}\right]^{-1} \times \left[\sum_{t=1}^T \tilde{X}'_{nt}R'_n(\rho)J_nR_n(\rho)S_n(\lambda)\tilde{Y}_{nt}\right], \tag{28}$$

$$\hat{\sigma}^2_{nT}(\lambda,\rho) = \frac{1}{(n-1)(T-1)}\sum_{t=1}^T[S_n(\lambda)\tilde{Y}_{nt} - \tilde{X}_{nt}\hat{\beta}_{nT}(\lambda,\rho)]' \times R'_n(\rho)J_nR_n(\rho)[S_n(\lambda)\tilde{Y}_{nt} - \tilde{X}_{nt}\hat{\beta}_{nT}(\lambda,\rho)], \tag{29}$$

and the concentrated log likelihood function of (λ, ρ) is

$$\ln L_{n,T}(\lambda,\rho) = -\frac{(n-1)(T-1)}{2}(\ln(2\pi)+1) - \frac{(n-1)(T-1)}{2}\ln\hat{\sigma}^2_{nT}(\lambda,\rho) - (T-1)[\ln(1-\lambda)+\ln(1-\rho)] + (T-1) \times[\ln|S_n(\lambda)| + \ln|R_n(\rho)|]. \tag{30}$$

Note that $\hat{\beta}_{nT}(\lambda,\rho) = \hat{\beta}^d_{nT}(\lambda,\rho)$, but

$$\hat{\sigma}^{2d}_{nT}(\lambda,\rho) = \frac{(n-1)(T-1)}{nT}\hat{\sigma}^2_{nT}(\lambda,\rho).$$

Hence, (27) can be rewritten as

$$\ln L^d_{n,T}(\lambda,\rho) = -\frac{nT}{2}\left(\ln(2\pi) + \ln\frac{(n-1)(T-1)}{nT} + 1\right) - \frac{nT}{2}\ln\hat{\sigma}^2_{nT}(\lambda,\rho) + T[\ln|S_n(\lambda)| + \ln|R_n(\rho)|]. \tag{31}$$

By comparing (30) and (31), we can see that $\ln L^d_{n,T}(\lambda,\rho)$ and $\ln L_{n,T}(\lambda,\rho)$ do not yield the same estimates of λ_0 and ρ_0. By ignoring irrelevant constant terms, the difference between $\frac{1}{(n-1)(T-1)}\ln L_{n,T}(\lambda,\rho)$ and $\frac{1}{nT}\ln L^d_{n,T}(\lambda,\rho)$ is $\frac{1}{n(n-1)}[\ln|S_n(\lambda)| + \ln|R_n(\rho)|] - \frac{1}{n-1}[\ln(1-\lambda)+\ln(1-\rho)]$. The direct and transformation approaches will yield asymptotically similar sample average objective functions when n is large, because their difference will vanish when n tends to infinity. Hence, under large n case, the estimate $\hat{\zeta}^d_{nT}$ of the direct approach and $\hat{\zeta}_{nT}$ of the transformation approach are both consistent. However, they are not numerically identical.

3.4. Consistency and asymptotic distributions of estimates

Denote[19]

$$\mathcal{H}_{nT}(\rho) = \frac{1}{(n-1)(T-1)}\sum_{t=1}^T(\tilde{X}_{nt}, G_n\tilde{X}_{nt}\beta_0)'R'_n(\rho) \times J_nR_n(\rho)(\tilde{X}_{nt}, G_n\tilde{X}_{nt}\beta_0), \tag{32}$$

$$\sigma^2_n(\rho) = \frac{\sigma^2_0}{n-1}tr[(R_n(\rho)R_n^{-1})'J_n(R_n(\rho)R_n^{-1})],$$

$$\sigma^2_n(\lambda,\rho) = \frac{\sigma^2_0}{n-1}tr[(R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1})'J_n(R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1})].$$

The following assumptions provide conditions for parameter identification. These assumptions modify the assumptions in Section 2 in that J_n will be involved.

Assumption 4′. The elements of X_{nt} are nonstochastic and bounded, uniformly in n and t. Under the setting in Assumption 6, the limit of $\frac{1}{nT}\sum_{t=1}^T \tilde{X}'_{nt}R'_nJ_nR_n\tilde{X}_{nt}$ exists and is nonsingular.[20]

Assumption 7′. Either (a) the limit of $\mathcal{H}_{nT}(\rho)$ is nonsingular for each ρ in $\mathbb{P}$ and the limit of $(\frac{1}{n-1}\ln|\sigma^2_0R_n^{-1\prime}J_nR_n^{-1}| - \frac{1}{n-1}\ln|\sigma^2_n(\rho)R_n^{-1}(\rho)'J_nR_n^{-1}(\rho)|)$ is not zero for $\rho \neq \rho_0$; or (b) the limit of $(\frac{1}{n-1}\ln|\sigma^2_0R_n^{-1\prime}S_n^{-1\prime}J_nS_n^{-1}R_n^{-1}| - \frac{1}{n-1}\ln|\sigma^2_n(\lambda,\rho)R_n^{-1}(\rho)'S_n^{-1}(\lambda)'J_nS_n^{-1}(\lambda)R_n^{-1}(\rho)|)$ is not zero for $(\lambda,\rho) \neq (\lambda_0,\rho_0)$.

Assumption 8′. The limit of $\frac{1}{(n-1)^2}\left[tr(C^s_nC^s_n)tr(D^s_nD^s_n) - tr^2(C^s_nD^s_n)\right]$ is strictly positive, where $C_n = J_n\ddot{G}_n - \frac{trJ_n\ddot{G}_n}{n-1}J_n$ and $D_n = J_nH_n - \frac{trJ_nH_n}{n-1}J_n$.

Theorem 3. (1) *For the QMLE $\hat{\theta}^d_{nT}$ based on (21), under Assumptions 1–3, 4′, 5, 6 and 7′, $\hat{\theta}^d_{nT} - \theta_T \xrightarrow{p} 0$ where $\theta_T = \theta_0 - (\mathbf{0}_{1\times(k+2)}, \frac{1}{T}\sigma^2_0)'$.*
(2) *For the QMLE $\hat{\theta}_{nT}$ based on (24), under Assumptions 1′, 2, 3, 4′, 5, 6 and 7′, $\hat{\theta}_{nT} - \theta_0 \xrightarrow{p} 0$.*

Proof. The arguments will be similar to those in the proof of Theorem 1. ■

[18] The approaches in Cox and Reid (1987) and Arellano and Hahn (2005) may be applied to nonlinear models.

[19] $\mathcal{H}_{nT}(\rho)$, $\sigma^2_n(\rho)$ and $\sigma^2_n(\lambda,\rho)$ for Section 3 are different from those in Section 2 although they share the same notations. The difference is that we have degrees of freedom adjustment and J_n matrix present in those in Section 3.

[20] This assumption rules out regressors which are either time or cross section invariant.

For the direct approach, the consistency of the QMLE of $\zeta_0 = (\beta_0', \lambda_0, \rho_0)'$ requires only n to be large. If T were finite, the time dummies would introduce an additional finite number of regression coefficients, which can be consistently estimated as n tends to infinity. However, the consistency of the estimate of the variance parameter requires both n and T to be large. For the transformation approach, all the estimates in $\hat{\theta}_{nT}$ will be consistent even when T is small.

Similar to the previous sections, the asymptotic properties of $\hat{\theta}_{nT}^d$ can be obtained by the Taylor expansion of $\frac{\partial \ln L_{n,T}^d(\theta)}{\partial \theta}$ around θ_T, and that of $\hat{\theta}_{nT}$ from $\frac{\partial \ln L_{n,T}(\theta)}{\partial \theta}$ around θ_0. For the direct approach, however, the score evaluated at θ_T will not be centered at zero when T is large, due to the incidental parameter problem induced by time effects. Denote $b_{\theta_T,nT} = (\Sigma_{\theta_T,nT}^d)^{-1} a_{\theta_T,n}$ where $\Sigma_{\theta_T,nT}^d$ is in (53) and[21]

$$a_{\theta_T,n} = \left(\mathbf{0}_{1\times k}, \frac{1}{n} l_n' R_n G_n R_n^{-1} l_n, \frac{1}{n} l_n' H_n l_n, \frac{1}{2\sigma_T^2} \right)'.$$

Theorem 4. (1) *For the direct approach, under Assumptions 1–3, 4′, 5, 6 and 7′ (a); or Assumptions 1–3, 4′, 5, 6 and 7′ (b) and 8′*

(i) *when $\frac{n}{T} \to c$, where $0 < c < \infty$,*

$$\sqrt{nT}(\hat{\theta}_{nT}^d - \theta_T) + \frac{1}{\sqrt{c}} b_{\theta_T,nT} \xrightarrow{d} N(0, \lim (\Sigma_{\theta_T,nT}^d)^{-1} \times (\Sigma_{\theta_T,nT}^d + \Omega_{\theta_T,nT}^d)(\Sigma_{\theta_T,nT}^d)^{-1});$$

(ii) *when $\frac{n}{T} \to 0$, $n(\hat{\theta}_{nT}^d - \theta_T) + b_{\theta_T,nT} \xrightarrow{p} 0$; but*

(iii) *when $\frac{n}{T} \to \infty$, $\sqrt{nT}(\hat{\theta}_{nT}^d - \theta_T) \xrightarrow{d} N(0, \lim \frac{T}{T-1} (\Sigma_{\theta_T,nT}^d)^{-1} (\Sigma_{\theta_T,nT}^d + \Omega_{\theta_T,nT}^d)(\Sigma_{\theta_T,nT}^d)^{-1})$, where $\Sigma_{\theta_T,nT}^d$ and $\Omega_{\theta_T,nT}^d$ are in (53) and (58).*

(2) *For the transformation approach, under 1′ 2, 3, 4′, 5, 6 and 7′ (a); or 1′ 2, 3, 4′, 5, 6 and 7′ (b) and 8′,*

$$\sqrt{(n-1)(T-1)}(\hat{\theta}_{nT} - \theta_0) \xrightarrow{d} N(0, \lim \Sigma_{\theta_0,nT}^{-1} \times (\Sigma_{\theta_0,nT} + \Omega_{\theta_0,nT}) \Sigma_{\theta_0,nT}^{-1}). \tag{33}$$

where $\Sigma_{\theta_T,nT}$ and $\Omega_{\theta_T,nT}$ are in (57) and (59).

Proof. See Appendix C.3. ■

Hence, $\hat{\theta}_{nT}^d$ is $\sqrt{nT}$ consistent when n and T go to infinity, but it has a leading bias which is the sum of $-(\mathbf{0}_{1\times(k+2)}, \frac{1}{T}\sigma_0^2)'$ and $-\frac{1}{n} b_{\theta_T,nT}$. The confidence interval for $\hat{\theta}_{nT}^d$ will not properly center around θ_0 when $\frac{n}{T} \to c$ for finite $c > 0$. When $\frac{n}{T} \to 0$, i.e., T is large relative to n, the bias component with $b_{\theta_T,nT}$ is the dominating one, and $\hat{\theta}_{nT}^d$ has the low n rate of convergence and its limiting distribution is degenerate. On the other hand, when $\frac{n}{T} \to \infty$, i.e., T is small relative to n, the estimate of $\zeta_0 = (\beta_0', \lambda_0, \rho_0)'$ is asymptotically centered; while only the estimate of σ_0^2 has the low T rate of convergence, and its limiting distribution is degenerate.[22] When T is finite, there is no additional incidental parameter problem caused by a finite number of time dummies. The additional incidental parameter problem occurs only when T goes to infinity at the same rate as n or faster than n.[23] For the transformation approach, the QMLE $\hat{\theta}_{nT}$ is consistent and asymptotically normal, and it is properly centered. When both n and T are large, estimates of the parameters based on the two approaches will be consistent and have the same asymptotic variance matrix.

In general, analytical bias reduction procedures are possible. Arellano and Hahn (2005) review various bias-correction methods for nonlinear panel data models with fixed individual effects. To correct for the bias due to the presence of incidental parameter problem, they analytically compare the bias correction of (i) estimators; (ii) moment equation (the score); and (iii) concentrated likelihood. A restricted case that relies on parameters orthogonality (Cox and Reid, 1987; Lancaster, 2000) is also discussed. For our SAR panel data model with fixed effects, we develop a bias correction procedure corresponding to (i).[24]

The overall bias can be corrected in two steps — an additive correction followed by a scalar adjustment in the σ^2 component. The first step is to correct for the bias of $\frac{1}{n} b_{\theta_T,nT}$ and the second step is to correct the bias of $(\mathbf{0}_{1\times(k+2)}, \frac{1}{T}\sigma_0^2)'$. Denote

$$\hat{\theta}_{nT}^{d1} = \hat{\theta}_{nT}^d - \frac{\hat{B}_{nT}}{n}, \quad \text{and} \quad \hat{\theta}_{nT}^{d2} = A_T \cdot \hat{\theta}_{nT}^{d1}, \tag{34}$$

where $\hat{B}_{nT} = [-(\Sigma_{\theta,nT}^d)^{-1} \cdot a_{\theta,n}]|_{\theta=\hat{\theta}_{nT}^d}$ and $A_T = \begin{pmatrix} I_{k+2} & \mathbf{0}_{(k+2)\times 1} \\ \mathbf{0}_{1\times(k+2)} & \frac{T}{T-1} \end{pmatrix}$. As is shown in Appendix C.4, $\hat{B}_{nT} + b_{\theta_T,nT} = O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{n}\right)\right)$. With rescaling[25] of the estimator of σ^2 in $\hat{\theta}_{nT}^{d1}$ by $\frac{T}{T-1}$, the bias corrected $\hat{\theta}_{nT}^{d2}$ is asymptotically normal and centered around θ_0.

Theorem 5. *Under Assumptions 1–3, 4′, 5, 6 and 7′ (a); or Assumptions 1–3, 4′, 5, 6 and 7′ (b) and 8′, when $\frac{T}{n^3} \to 0$,*

$$\sqrt{nT}(\hat{\theta}_{nT}^{d2} - \theta_0) \xrightarrow{d} N\left(0, \lim \frac{T}{T-1} (\Sigma_{\theta_T,nT}^d)^{-1} \times (\Sigma_{\theta_T,nT}^d + \Omega_{\theta_T,nT}^d)(\Sigma_{\theta_T,nT}^d)^{-1}\right). \tag{35}$$

Proof. See Appendix C.4. ■

4. Monte Carlo

We conduct a small Monte Carlo experiment to evaluate the performance of estimates under different settings. We first look into the model (1) with individual effects but no time effects, and compare the performance of the transformation approach with the direct approach. Then, we investigate the model (19) with both individual and time effects.

We first generate samples from (1):

$$Y_{nt} = \lambda_0 W_n Y_{nt} + X_{nt}\beta_0 + \mathbf{c}_{n0} + U_{nt},$$
$$U_{nt} = \rho_0 M_n U_{nt} + V_{nt}, t = 1, 2, \ldots, T,$$

using $\theta_0^a = (1.0, 0.2, 0.5, 1)'$ and $\theta_0^b = (1, 0.5, 0.2, 1)'$ where $\theta_0 = (\beta_0', \lambda_0, \rho_0, \sigma_0^2)'$. X_{nt}, $\mathbf{c}_{n0}$ and V_{nt} are generated from independent standard normal distributions, and both the spatial

[21] When W_n and M_n are row normalized, $a_{\theta_T,n}$ will be reduced to $(\mathbf{0}_{1\times k_x}, \frac{1}{1-\lambda_0}, \frac{1}{1-\rho_0}, \frac{1}{2\sigma_T^2})'$.

[22] It is $T(\hat{\sigma}_{nT}^{2d} - \sigma_0^2) + \sigma_0^2 \xrightarrow{p} 0$ where $\hat{\sigma}_{nT}^{2d}$ is the last entry of $\hat{\theta}_{nT}^d$.

[23] It is of interest to see that, for the panel model without time dynamics, the finite or relatively short T (relative to n) do not cause noncentrality in the distribution for estimates of most of the common parameters except the variance parameter of the disturbances. This feature differs from those of the dynamic panel data models in Hahn and Kuersteiner (2002) and Hahn and Moon (2006), and spatial dynamic panel models in Yu et al. (2008).

[24] Other bias correction methods for SAR panel data via (ii) and (iii) might be possible and would be of interest in future research.

[25] An alternative is to correct that entry in an additive fashion as $\hat{\sigma}_{nT}^2 + \frac{1}{T}\hat{\sigma}_{nT}^2$. However, for a finite T, such bias correction would not yield a consistent estimate for σ_0^2; and for T being large, the distribution will not be properly centered unless $\frac{n}{T^3} \to 0$.

Table 1
Transformation and direct approaches: model with individual effects only.

	T	n	θ_0		β	λ	ρ	σ_1^2	σ_2^2
(1)	5	49	θ_0^a	Bias	−0.0027	0.0096	−0.0279	−0.0216	−0.2173
				E-SD	0.0766	0.1377	0.1459	0.1067	0.0854
				RMSE	0.0766	0.1380	0.1485	0.1089	0.2334
				T-SD	0.0743	0.1355	0.1371	0.1043	0.0746
(2)	5	49	θ_0^b	Bias	−0.0039	−0.0173	0.0021	−0.0027	−0.2182
				E-SD	0.0736	0.1150	0.1590	0.1044	0.0835
				RMSE	0.0737	0.1163	0.1590	0.1068	0.2336
				T-SD	0.0718	0.1134	0.1574	0.1024	0.0733
(3)	10	49	θ_0^a	Bias	−0.0005	0.0040	−0.0110	−0.0116	−0.1104
				E-SD	0.0492	0.0948	0.0939	0.0704	0.0633
				RMSE	0.0492	0.0949	0.0945	0.0713	0.1273
				T-SD	0.0496	0.0925	0.0921	0.0701	0.0599
(4)	10	49	θ_0^b	Bias	−0.0011	−0.0066	0.0007	−0.0120	−0.1108
				E-SD	0.0466	0.0759	0.1053	0.0691	0.0622
				RMSE	0.0466	0.0762	0.1053	0.0702	0.1271
				T-SD	0.0475	0.0755	0.1069	0.0687	0.0586
(5)	50	9	θ_0^a	Bias	0.0003	0.0072	−0.0126	−0.0082	−0.0280
				E-SD	0.0501	0.0844	0.0810	0.0713	0.0699
				RMSE	0.0501	0.0847	0.0820	0.0718	0.0753
				T-SD	0.0499	0.0842	0.0787	0.0704	0.0682
(6)	50	9	θ_0^b	Bias	−0.0010	−0.0065	0.0018	−0.0093	−0.0291
				E-SD	0.0481	0.0664	0.0961	0.0708	0.0694
				RMSE	0.0482	0.0668	0.0962	0.0714	0.0752
				T-SD	0.0475	0.0645	0.0967	0.0689	0.0669
(7)	50	16	θ_0^a	Bias	−0.0010	0.0021	−0.0050	−0.0079	−0.0278
				E-SD	0.0380	0.0692	0.0660	0.0536	0.0525
				RMSE	0.0380	0.0692	0.0662	0.0542	0.0594
				T-SD	0.0374	0.0663	0.0641	0.0528	0.0512
(8)	50	16	θ_0^b	Bias	−0.0015	−0.0037	0.0016	−0.0082	−0.0280
				E-SD	0.0367	0.0549	0.0792	0.0526	0.0516
				RMSE	0.0367	0.0550	0.0793	0.0532	0.0587
				T-SD	0.0356	0.0524	0.0762	0.0516	0.0501
(9)	50	49	θ_0^a	Bias	−0.0009	−0.0011	−0.0004	−0.0025	−0.0224
				E-SD	0.0220	0.0405	0.0401	0.0305	0.0298
				RMSE	0.0220	0.0405	0.0401	0.0306	0.0373
				T-SD	0.0214	0.0404	0.0396	0.0303	0.0294
(10)	50	49	θ_0^b	Bias	−0.0007	−0.0031	0.0026	−0.0019	−0.0219
				E-SD	0.0212	0.0321	0.0465	0.0297	0.0291
				RMSE	0.0212	0.0323	0.0466	0.0298	0.0365
				T-SD	0.0203	0.0324	0.0464	0.0296	0.0287

Note: 1. $\theta_0^a = (1, 0.2, 0.5, 1)$ and $\theta_0^b = (1, 0.5, 0.2, 1)$.
2. The column of σ_1^2 is from the transformation approach; and the column of σ_2^2 is from the direct approach.
3. The transformation approach and the direct approach yield the same estimate of $\zeta_0 = (\beta_0', \lambda_0, \rho_0)'$.

weights matrices W_n and M_n are the same rook matrices.[26] We use some combinations of $T = 5, 10, 50$, and $n = 9, 16, 49$. For each set of generated sample observations, we calculate the ML estimator and evaluate the bias. We do this 1000 times. With two different values of θ_0 for various combinations of n and T, finite sample properties of estimators are summarized in Table 1. For each case, we report the empirical bias (Bias), the empirical standard deviation (E-SD), the empirical root mean square error (RMSE) and the theoretical standard deviation (T-SD).[27]

We see that both approaches provide the same estimate of $\zeta_0 = (\beta_0', \lambda_0, \rho_0)'$, and they have small biases when either n or T are large. From the last two columns in Table 1, the transformation approach yields a consistent estimator of σ_0^2 when either n or T is large; however, the estimator of σ_0^2 by the direct approach has a small bias only when T is large. For E-SDs, RMSEs and T-SDs for the estimators of $\zeta_0 = (\beta_0', \lambda_0, \rho_0)'$, they are small when either n or T are large. Also, T-SDs are similar to E-SDs, which means that the negative inverse of the Hessian matrix provides proper estimates for the variances of estimators.

We then generate samples from (19):

$$Y_{nt} = \lambda_0 W_n Y_{nt} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_t l_n + U_{nt},$$
$$U_{nt} = \rho_0 M_n U_{nt} + V_{nt}, t = 1, 2, \ldots, T,$$

using the same n, T, θ_0^a, θ_0^b, W_n and M_n. The X_{nt}, $\mathbf{c}_{n0}$, $\boldsymbol{\alpha}_{T0} = (\alpha_1, \alpha_2, \ldots, \alpha_T)$ and V_{nt} are generated from independent standard normal distributions. The finite sample properties of the estimators are summarized in Tables 2–4. Table 2 is for the direct approach, and Table 3 is for those estimators after bias correction; Table 4 is for the transformation approach.

We see that the biases of estimates of ζ_0 and σ_0^2, based on the transformation approach, are small when either n or T are large. For the direct approach, the bias of the estimate of ζ_0 is small when n is large, and the bias is large when n is small and T might be large; the bias for the estimate of σ_0^2 is small only when both n and T are large. After the bias correction for the direct approach, by comparing Tables 2 and 3, we see that the bias correction reduces the biases of

26 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.

27 The T-SD is obtained from diagonal elements of the negative inverse of the estimated Hessian matrix.

Table 2
Direct approach: model with both time and individual effects.

	T	n	θ_0		β	λ	ρ	σ^2
(1)	5	49	θ_0^a	Bias	0.0021	0.0271	−0.0904	−0.2207
				E-SD	0.0749	0.1213	0.1342	0.0843
				RMSE	0.0749	0.1243	0.1618	0.2362
				T-SD	0.0662	0.1254	0.1338	0.1026
(2)	5	49	θ_0^b	Bias	−0.0017	−0.0382	0.0183	−0.2267
				E-SD	0.0733	0.1063	0.1443	0.0831
				RMSE	0.0733	0.1129	0.1455	0.2415
				T-SD	0.0642	0.1090	0.1478	0.1013
(3)	10	49	θ_0^a	Bias	0.0038	0.0241	−0.0779	−0.1151
				E-SD	0.0488	0.0856	0.0910	0.0623
				RMSE	0.0489	0.0889	0.1198	0.1308
				T-SD	0.0468	0.0900	0.0952	0.0688
(4)	10	49	θ_0^b	Bias	0.0001	−0.0305	−0.0178	−0.1216
				E-SD	0.0471	0.0733	0.0980	0.0622
				RMSE	0.0471	0.0794	0.0996	0.1366
				T-SD	0.0450	0.0771	0.1060	0.0679
(5)	50	9	θ_0^a	Bias	−0.0014	−0.0179	−0.3438	−0.1260
				E-SD	0.0519	0.0541	0.0566	0.0649
				RMSE	0.0520	0.0570	0.3484	0.1417
				T-SD	0.0488	0.0983	0.1140	0.0605
(6)	50	9	θ_0^b	Bias	−0.0091	−0.1959	−0.1330	−0.1258
				E-SD	0.0526	0.0528	0.0571	0.0651
				RMSE	0.0534	0.2029	0.1447	0.1416
				T-SD	0.0479	0.0965	0.1192	0.0619
(7)	50	16	θ_0^a	Bias	0.0038	0.0262	−0.1964	−0.0608
				E-SD	0.0377	0.0496	0.0551	0.0498
				RMSE	0.0379	0.0561	0.2040	0.0786
				T-SD	0.0365	0.0713	0.0803	0.0493
(8)	50	16	θ_0^b	Bias	−0.0021	−0.0948	−0.0539	−0.0692
				E-SD	0.0375	0.0461	0.0578	0.0500
				RMSE	0.0376	0.1054	0.0791	0.0854
				T-SD	0.0354	0.0660	0.0862	0.0494
(9)	50	49	θ_0^a	Bias	0.0030	0.0195	−0.0671	−0.0272
				E-SD	0.0217	0.0365	0.0385	0.0291
				RMSE	0.0219	0.0413	0.0774	0.0398
				T-SD	0.0210	0.0409	0.0428	0.0297
(10)	50	49	θ_0^b	Bias	−0.0002	−0.0286	−0.0132	−0.0335
				E-SD	0.0213	0.0314	0.0428	0.0288
				RMSE	0.0213	0.0425	0.0448	0.0442
				T-SD	0.0201	0.0347	0.0479	0.0293

Note: $\theta_0^a = (1, 0.2, 0.5, 1)$ and $\theta_0^b = (1, 0.5, 0.2, 1)$.

the direct approach estimates, without significant increase in the variance (S-TD).[28] This is consistent with the theoretical prediction.

We also run the simulation when V_{nt} is generated from the exponential distribution with unit variance (demeaned by the population mean). The disturbances are skewed. To save space, the Monte Carlo simulation is reported only for the $T = 10$ and $n = 49$ case. By comparing with the corresponding estimates in Tables 1–4 under the normal disturbances, we see that the biases and SDs are similar except that the SDs of the estimates for σ_0^2 in Table 5 are relatively larger.

5. Conclusion

In this paper, we consider the estimation of SAR panel models with fixed effects and SAR disturbances.

We first consider the model with individual effects only where the time periods T can be finite (or infinite), while the number of spatial units n is large. If T is finite, the direct ML approach by estimating all the parameters including the fixed effects will yield consistent estimators for the common parameters except the variance parameter. These features are similar to the direct ML estimation of the linear panel regression model with fixed individual effects. As an alternative estimation approach, we suggest the use of a transformation approach, which eliminates the individual fixed effects and can provide consistent estimates for all the common parameters, including the variance. In the transformation approach, the individual effects are eliminated by taking deviation from time average for each spatial unit. A likelihood function, which takes into account the generalized inverse of the resulting disturbances, can be constructed from the transformed data. The transformation approach is shown to be a conditional likelihood approach if the disturbances were normally distributed.

We consider, next, the model with both individual and time effects. We show that the direct approach will yield consistent estimates when both n and T are large. However, the distribution of the estimates is not properly centered. Bias correction procedures are useful to remove the noncentrality. For the practical case where the spatial weights matrices are row-normalized, likelihood type estimation based on transformed data is also available where both the individual and time effects can be eliminated. The common parameter estimates from the transformed approach are consistent when either n or T is large, and the asymptotic distributions are properly centered. Monte Carlo results are provided to illustrate finite sample properties of the various estimators.

While Baltagi et al. (2003, 2007) and Kapoor et al. (2007) consider spatial models with random effects, the SAR models in this paper have fixed effects. The proposed estimation methods

28 For the T-SD of the bias corrected estimates, its values are also similar to those of the estimates before bias correction.

Table 3
Bias corrected direct approach: model with both time and individual effects.

	T	n	θ_0		β	λ	ρ	σ^2
(1)	5	49	θ_0^a	Bias	−0.0015	0.0131	−0.0371	−0.0202
				E-SD	0.0761	0.1368	0.1487	0.1073
				RMSE	0.0761	0.1375	0.1533	0.1092
				T-SD	0.0747	0.1373	0.1356	0.0938
(2)	5	49	θ_0^b	Bias	−0.0033	−0.0192	−0.0013	−0.0236
				E-SD	0.0735	0.1197	0.1623	0.1051
				RMSE	0.0736	0.1213	0.1623	0.1078
				T-SD	0.0722	0.1189	0.1572	0.0925
(3)	10	49	θ_0^a	Bias	0.0005	0.0082	−0.0216	−0.0106
				E-SD	0.0498	0.0971	0.1012	0.0705
				RMSE	0.0498	0.0975	0.1035	0.0713
				T-SD	0.0500	0.0941	0.0929	0.0666
(4)	10	49	θ_0^b	Bias	−0.0008	−0.0094	−0.0022	−0.0132
				E-SD	0.0470	0.0822	0.1107	0.0699
				RMSE	0.0471	0.0827	0.1107	0.0712
				T-SD	0.0477	0.0802	0.1095	0.0655
(5)	50	9	θ_0^a	Bias	−0.0007	−0.0083	−0.1737	−0.0274
				E-SD	0.0538	0.0801	0.0861	0.0714
				RMSE	0.0538	0.0805	0.1939	0.0765
				T-SD	0.0523	0.0998	0.1052	0.0668
(6)	50	9	θ_0^b	Bias	−0.0018	−0.1080	−0.0545	−0.0301
				E-SD	0.0523	0.0763	0.0881	0.0719
				RMSE	0.0523	0.1322	0.1036	0.0780
				T-SD	0.0503	0.0969	0.1234	0.0675
(7)	50	16	θ_0^a	Bias	0.0006	0.0096	−0.0645	−0.0052
				E-SD	0.0387	0.0673	0.0732	0.0532
				RMSE	0.0387	0.0680	0.0976	0.0534
				T-SD	0.0384	0.0722	0.0727	0.0520
(8)	50	16	θ_0^b	Bias	−0.0003	−0.0349	−0.0106	−0.0112
				E-SD	0.0373	0.0624	0.0800	0.0534
				RMSE	0.0373	0.0715	0.0807	0.0545
				T-SD	0.0364	0.0655	0.0876	0.0514
(9)	50	49	θ_0^a	Bias	−0.0003	0.0017	−0.0079	−0.0010
				E-SD	0.0222	0.0414	0.0425	0.0304
				RMSE	0.0222	0.0414	0.0433	0.0304
				T-SD	0.0216	0.0413	0.0405	0.0300
(10)	50	49	θ_0^b	Bias	−0.0005	−0.0061	0.0015	−0.0022
				E-SD	0.0213	0.0353	0.0485	0.0298
				RMSE	0.0213	0.0358	0.0486	0.0298
				T-SD	0.0204	0.0347	0.0484	0.0294

Note: 1. $\theta_0^a = (1, 0.2, 0.5, 1)$ and $\theta_0^b = (1, 0.5, 0.2, 1)$.
2. The T-SD uses the bias corrected estimates.

are robust to different specifications in Baltagi et al. (2003) and Kapoor et al. (2007), and are computationally simpler than the ML approach for the estimation of the generalized random effects model in Baltagi et al. (2007). However, when the individual effects are random in the true DGP, proper methods which take into account the random effects' variance structure can improve the efficiency of the estimates. Hausman's type specification test of fixed effects versus random effects may be constructed. These will be investigated in future research.

Appendix A. Notations and some lemmas

The following list summarizes some frequently used notations in either the text or the Appendices A–C:

$S_n(\lambda) = I_n - \lambda W_n$ for any possible λ and $S_n = I_n - \lambda_0 W_n$.

$R_n(\rho) = I_n - \rho M_n$ for any possible ρ and $R_n = I_n - \rho_0 M_n$.

$G_n = W_n S_n^{-1}$ and $H_n = M_n R_n^{-1}$.

$\tilde{Y}_{nt} = Y_{nt} - \bar{Y}_{nT}$ for $t = 1, 2, \ldots, T$ where $\bar{Y}_{nT} = \frac{1}{T}\sum_{t=1}^T Y_{nt}$.

$\ddot{W}_n = R_n W_n R_n^{-1}$, $\ddot{G}_n = \ddot{W}_n(I_n - \lambda_0 \ddot{W}_n)^{-1} = R_n G_n R_n^{-1}$, $\ddot{X}_{nt} = R_n \tilde{X}_{nt}$.

$\theta = (\beta', \lambda, \rho, \sigma^2)'$ and $\zeta = (\beta', \lambda, \rho)'$.

$A_n^s = A_n' + A_n$ for any $n \times n$ matrix A_n.

In Section 2, $\mathcal{H}_{nT}(\rho) = \frac{1}{n(T-1)}\sum_{t=1}^T (\tilde{X}_{nt}, G_n\tilde{X}_{nt}\beta_0)' R_n'(\rho) R_n(\rho) (\tilde{X}_{nt}, G_n\tilde{X}_{nt}\beta_0)$.

In Section 3, $\mathcal{H}_{nT}(\rho) = \frac{1}{(n-1)(T-1)}\sum_{t=1}^T (\tilde{X}_{nt}, G_n\tilde{X}_{nt}\beta_0)' R_n'(\rho) J_n R_n(\rho) (\tilde{X}_{nt}, G_n\tilde{X}_{nt}\beta_0)$.

Lemma A.1. *Suppose that $\{B_n\}$ is a sequence of symmetric UB matrix with elements $b_{n,ij}$, and D_{nt} is a sequence of constant vectors with its elements $d_{nt,i}$ uniformly bounded. The moment $E(|v_{it}|^{4+2\delta})$ for some $\delta > 0$ of v_{it} exists. Let $\sigma_{Q_{nT}}^2$ be the variance of Q_{nT} where $Q_{nT} = \sum_{t=1}^T \left(D_{nt}' V_{nt} + V_{nt}' B_n V_{nt} - \sigma_0^2 trB_n\right)$ such that $\sigma_{Q_{nT}}^2 = \sigma_0^2 \sum_{t=1}^T D_{nt}' D_{nt} + T[\left(\mu_4 - 3\sigma_0^4\right)\sum_{i=1}^n b_{n,ii}^2 + 2\sigma_0^4 tr(B_n^2)] + 2\mu_3 \sum_{t=1}^T \sum_{i=1}^n d_{nt,i} b_{n,ii}$. Assume that the variance $\sigma_{Q_{nT}}^2$ is $O(nT)$ with $\{\frac{1}{nT}\sigma_{Q_{nT}}^2\}$ bounded away from zero. If either n or T are large, then $\frac{Q_{nT}}{\sigma_{Q_{nT}}} \xrightarrow{d} N(0, 1)$.*

Proof. When T is fixed and n is large, this is Lemma A.13 in Lee (2004), which is essentially the CLT in Kelejian and Prucha (2001). When T is large and n is either fixed or large, it is a special case (there is no moving averages of past disturbances in Q_{nT}) of the CLT in Yu et al. (2008). ■

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

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

Table 4
Transformation approach: model with both time and individual effects.

	T	n	θ_0		β	λ	ρ	σ^2
(1)	5	49	θ_0^a	Bias	−0.0020	0.0121	−0.0300	−0.0223
				E-SD	0.0764	0.1403	0.1529	0.1078
				RMSE	0.0764	0.1408	0.1558	0.1100
				T-SD	0.0751	0.1406	0.1481	0.1045
(2)	5	49	θ_0^b	Bias	−0.0042	−0.0167	0.0017	−0.0242
				E-SD	0.0737	0.1227	0.1658	0.1052
				RMSE	0.0738	0.1238	0.1658	0.1079
				T-SD	0.0723	0.1223	0.1654	0.1031
(3)	10	49	θ_0^a	Bias	−0.0001	0.0056	−0.0137	−0.0124
				E-SD	0.0500	0.0986	0.1031	0.0706
				RMSE	0.0500	0.0988	0.1040	0.0717
				T-SD	0.0502	0.0955	0.0994	0.0702
(4)	10	49	θ_0^b	Bias	−0.0013	−0.0064	−0.0005	−0.0133
				E-SD	0.0471	0.0836	0.1126	0.0700
				RMSE	0.0471	0.0839	0.1126	0.0712
				T-SD	0.0478	0.0816	0.1122	0.0691
(5)	50	9	θ_0^a	Bias	0.0010	0.0098	−0.0102	−0.0110
				E-SD	0.0546	0.1038	0.1260	0.0729
				RMSE	0.0546	0.1042	0.1264	0.0738
				T-SD	0.0540	0.1021	0.1276	0.0721
(6)	50	9	θ_0^b	Bias	−0.0017	−0.0010	0.0028	−0.0121
				E-SD	0.0512	0.1094	0.1306	0.0745
				RMSE	0.0512	0.1094	0.1306	0.0755
				T-SD	0.0507	0.1066	0.1314	0.0731
(7)	50	16	θ_0^a	Bias	−0.0011	0.0019	−0.0046	−0.0093
				E-SD	0.0393	0.0755	0.0845	0.0540
				RMSE	0.0393	0.0755	0.0846	0.0548
				T-SD	0.0390	0.0737	0.0830	0.0532
(8)	50	16	θ_0^b	Bias	−0.0019	−0.0031	0.0013	−0.0095
				E-SD	0.0373	0.0709	0.0915	0.0537
				RMSE	0.0373	0.0710	0.0915	0.0546
				T-SD	0.0365	0.0684	0.0894	0.0529
(9)	50	49	θ_0^a	Bias	−0.0009	−0.0011	−0.0002	−0.0026
				E-SD	0.0222	0.0422	0.0434	0.0305
				RMSE	0.0222	0.0423	0.0434	0.0306
				T-SD	0.0216	0.0417	0.0428	0.0304
(10)	50	49	θ_0^b	Bias	−0.0008	−0.0030	0.0025	−0.0021
				E-SD	0.0213	0.0358	0.0494	0.0298
				RMSE	0.0213	0.0360	0.0494	0.0299
				T-SD	0.0204	0.0351	0.0487	0.0298

Note: $\theta_0^a = (1, 0.2, 0.5, 1)$ and $\theta_0^b = (1, 0.5, 0.2, 1)$.

Lemma A.2. *For $W_n^* = F'_{n,n-1} W_n F_{n,n-1}$, when W_n is row normalized, $\left|I_{n-1} - \lambda W_n^*\right| = \frac{1}{1-\lambda} \left|I_n - \lambda W_n\right|$ and $(I_{n-1} - \lambda W_n^*)^{-1} = F'_{n,n-1}(I_n - \lambda W_n)^{-1} F_{n,n-1}$.*

Proof. The derivation of these results can be found in Appendix A.2 of Lee and Yu (forthcoming). ■

The following Lemma is applicable to both the models, under relevant assumptions in the corresponding Sections 2 and 3.

Lemma A.3. *Let $\|\theta - \theta_1\|$ be the Euclidean norm of $\theta - \theta_1$, and Θ_1 be a neighborhood of θ_1. Under the assumptions in the relevant section, the corresponding Hessian matrix of $\ln L_{n,T}(\theta)$ of the transformation approach with $\theta_1 = \theta_0$, has the following properties:*

$$-\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}(\theta)}{\partial\theta\partial\theta'} - \left(-\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}(\theta_1)}{\partial\theta\partial\theta'}\right) = \|\theta - \theta_1\| \cdot O_p(1),$$

$$\left(-\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}(\theta_1)}{\partial\theta\partial\theta'}\right) - \Sigma_{\theta_1,nT} = O_p\left(\frac{1}{\sqrt{nT}}\right),$$

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

and

$$\sup_{\theta\in\Theta_1}\left|-\frac{1}{nT}E\frac{\partial^2 \ln L_{n,T}(\theta)}{\partial\theta\partial\theta'} - \Sigma_{\theta_1,nT}\right|_{ij} = \sup_{\theta\in\Theta_1}\|\theta - \theta_1\| \cdot O(1),$$

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

Similarly, for $\ln L^d_{n,T}(\theta)$ of the direct approach with $\theta_1 = \theta_T$, the corresponding properties above hold.

Proof. When n is large and T is fixed, the derivation is similar to Lee (2004) for the cross sectional SAR model. When T is large and n could be finite and large, the derivation is similar to (38)–(41) in Yu et al. (2008). ■

Lemma A.4. *Suppose that $\{A_n\}$ and $\{B_n\}$ are sequences of matrices with elements $a_{n,ij}$ and $b_{n,ij}$, and $\{D_{nt}\}$ is a sequence of constant column vectors with its elements $d_{nt,i}$. Then,*

$$cov\left[\left(\sum_{t=1}^T \tilde{V}'_{nt}A_n\tilde{V}_{nt}\right), \left(\sum_{t=1}^T \tilde{V}'_{nt}B_n\tilde{V}_{nt}\right)\right]$$
$$= (\mu_4 - 3\sigma_0^4)\frac{(T-1)^2}{T}vec'_D(A_n)vec_D(B_n) + \sigma_0^4(T-1)tr(A_nB_n^s),$$

and $cov[(\sum_{t=1}^T \tilde{V}'_{nt}A_n\tilde{V}_{nt}), \sum_{t=1}^T D'_{nt}\tilde{V}_{nt}] = 0$, where $vec_D(A_n)$ is a column vector formed by the diagonal elements of A_n, and $B_n^s = B_n + B'_n$.

Proof. Denote $\mathbf{V}_{nT} = (V'_{n1}, V'_{n2}, \ldots, V'_{nT})'$. With $J_T = I_T - \frac{1}{T}l_Tl'_T$, we have $\sum_{t=1}^T \tilde{V}'_{nt}A_n\tilde{V}_{nt} = \mathbf{V}'_{nT}(J_T \otimes A_n)\mathbf{V}_{nT}$. Hence, using the formulas

Table 5
Estimates with non-normal disturbances: both approaches under different DGPs.

	T	n	θ_0		β	λ	ρ	σ_1^2	σ_2^2
Transformation and direct approach, model with individual effects only									
(1)	10	49	θ_0^a	Bias	−0.0006	0.0007	−0.0091	−0.0047	−0.1042
				E-SD	0.0501	0.0922	0.0914	0.1401	0.1261
				RMSE	0.0501	0.0922	0.0918	0.1402	0.1636
				T-SD	0.0497	0.0926	0.0920	0.1001	0.1045
(2)	10	49	θ_0^b	Bias	0.0005	−0.0112	0.0056	−0.0042	−0.1038
				E-SD	0.0476	0.0754	0.1045	0.1387	0.1248
				RMSE	0.0476	0.0762	0.1047	0.1368	0.1623
				T-SD	0.0476	0.0761	0.1071	0.0991	0.1038
Direct approach, model with both effects									
(3)	10	49	θ_0^a	Bias	0.0048	0.0199	−0.0727	–	−0.1083
				E-SD	0.0492	0.0825	0.0870	–	0.1252
				RMSE	0.0495	0.0849	0.1134	–	0.1656
				T-SD	0.0469	0.0902	0.0950	–	0.1021
(4)	10	49	θ_0^b	Bias	0.0016	−0.0329	−0.0135	–	−0.1140
				E-SD	0.0478	0.0703	0.0943	–	0.1241
				RMSE	0.0478	0.0776	0.0953	–	0.1685
				T-SD	0.0451	0.0776	0.1063	–	0.1012
Direct approach after bias correction, model with both effects									
(5)	10	49	θ_0^a	Bias	0.0013	0.0034	−0.0158	–	−0.0033
				E-SD	0.0502	0.0934	0.0961	–	0.1404
				RMSE	0.0503	0.0935	0.0974	–	0.1404
				T-SD	0.0501	0.0943	0.0925	–	0.1010
(6)	10	49	θ_0^b	Bias	0.0008	−0.0119	0.0024	–	−0.0064
				E-SD	0.0478	0.0794	0.1070	–	0.1391
				RMSE	0.0478	0.0803	0.1071	–	0.1391
				T-SD	0.0479	0.0806	0.1097	–	0.0996
Transformation approach, model with both effects									
(7)	10	49	θ_0^a	Bias	0.0005	0.0004	−0.0076	−0.0053	–
				E-SD	0.0504	0.0959	0.0988	0.1400	–
				RMSE	0.0504	0.0959	0.0991	0.1401	–
				T-SD	0.0503	0.0957	0.0994	0.0755	–
(8)	10	49	θ_0^b	Bias	0.0003	−0.0093	0.0043	−0.0048	–
				E-SD	0.0478	0.0816	0.1098	0.1389	–
				RMSE	0.0478	0.0821	0.1099	0.1390	–
				T-SD	0.0479	0.0821	0.1127	0.0744	–

Note: 1. $\theta_0^a = (1, 0.2, 0.5, 1)$ and $\theta_0^b = (1, 0.5, 0.2, 1)$.
2. The column of σ_1^2 is from the transformation approach;
and the column of σ_2^2 is from the direct approach.
For the T-SD under non-normal disturbances, $\Omega^d_{\theta_T,nT}$ and $\Omega_{\theta_0,nT}$ are not zero.

for cross moments of quadratic forms,

$$\begin{aligned}
&E\left(\sum_{t=1}^T \tilde{V}'_{nt}A_n\tilde{V}_{nt}\right)\left(\sum_{t=1}^T \tilde{V}'_{nt}B_n\tilde{V}_{nt}\right)\\
&\quad = E\mathbf{V}'_{nT}(J_T\otimes A_n)\mathbf{V}_{nT}\mathbf{V}'_{nT}(J_T\otimes B_n)\mathbf{V}_{nT}\\
&\quad = (\mu_4 - 3\sigma_0^4)vec'_D(J_T\otimes A_n)vec_D(J_T\otimes B_n)\\
&\qquad + \sigma_0^4[tr(J_T\otimes A_n)tr(J_T\otimes B_n) + tr(J_T\otimes A_n)(J_T\otimes B_n^s)].
\end{aligned}$$

Using the fact that $tr(J_T\otimes A_n) = tr(J_T)tr(A_n) = (T-1)tr(A_n)$ and $vec_D(J_T\otimes A_n) = (1-\frac{1}{T})l_T\otimes vec_D(A_n)$, we have $vec'_D(J_T\otimes A_n)vec_D(J_T\otimes B_n) = \frac{(T-1)^2}{T}vec'_D(A_n)vec_D(B_n)$. Hence,

$$\begin{aligned}
&E\left(\sum_{t=1}^T \tilde{V}'_{nt}A_n\tilde{V}_{nt}\right)\left(\sum_{t=1}^T \tilde{V}'_{nt}B_n\tilde{V}_{nt}\right)\\
&\quad = (\mu_4 - 3\sigma_0^4)\frac{(T-1)^2}{T}vec'_D(A_n)vec_D(B_n)\\
&\qquad + \sigma_0^4[(T-1)^2tr(A_n)tr(B_n) + (T-1)tr(A_nB_n^s)].
\end{aligned}$$

Also, we have $E(\sum_{t=1}^T \tilde{V}'_{nt}A_n\tilde{V}_{nt}) = \sigma_0^2(T-1)tr(A_n)$ and $E(\sum_{t=1}^T \tilde{V}'_{nt}B_n\tilde{V}_{nt}) = \sigma_0^2(T-1)tr(B_n)$. Therefore,

$$\begin{aligned}
&cov\left[\left(\sum_{t=1}^T \tilde{V}'_{nt}A_n\tilde{V}_{nt}\right), \left(\sum_{t=1}^T \tilde{V}'_{nt}B_n\tilde{V}_{nt}\right)\right]\\
&\quad = (\mu_4 - 3\sigma_0^4)\frac{(T-1)^2}{T}vec'_D(A_n)vec_D(B_n)\\
&\qquad + \sigma_0^4(T-1)tr(A_nB_n^s).
\end{aligned}$$

For the covariance between the quadratic form and the linear form, $cov[(\sum_{t=1}^T \tilde{V}'_{nt}A_n\tilde{V}_{nt}), \sum_{t=1}^T D'_{nt}\tilde{V}_{nt}] = E[(\sum_{t=1}^T \tilde{V}'_{nt}A_n\tilde{V}_{nt})\times\sum_{t=1}^T D'_{nt}\tilde{V}_{nt}]$. Denote $\tilde{\mathbf{D}}_{nT} = (\tilde{D}'_{n1},\ldots,\tilde{D}'_{nT})'$ where $\tilde{D}_{nt} = D_{nt} - \frac{1}{T}\sum_{s=1}^T D_{ns}$ with elements $\tilde{d}_{nt,i}$. It follows that

$$\begin{aligned}
&cov\left[\left(\sum_{t=1}^T \tilde{V}'_{nt}A_n\tilde{V}_{nt}\right), \sum_{t=1}^T D'_{nt}\tilde{V}_{nt}\right]\\
&\quad = E\mathbf{V}'_{nT}(J_T\otimes A_n)\mathbf{V}_{nT}\tilde{\mathbf{D}}'_{nT}\mathbf{V}_{nT}\\
&\quad = \left(1-\frac{1}{T}\right)\mu_3\sum_{t=1}^T\sum_{i=1}^n a_{n,ii}\cdot\tilde{d}_{nt,i} = 0,
\end{aligned}$$

where μ_3 is the third moment of v_{it}, and the last equality holds because $\sum_{t=1}^T \tilde{d}_{nt,i} = 0$. ■

Appendix B. The direct and transformation approaches in Section 2

B.1. The first and second order derivatives of (4) for the direct approach

For the concentrated log likelihood function (4), the first and second order derivatives are given in Box I. The score of the log likelihood function evaluated at θ_T is

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L^d_{n,T}(\theta_T)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma_T^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\ddot{X}'_{nt}\tilde{V}_{nt} \\ \frac{1}{\sigma_T^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}(\ddot{G}_n\ddot{X}_{nt}\beta_0)'\tilde{V}_{nt} + \frac{1}{\sigma_T^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}(\tilde{V}'_{nt}\ddot{G}'_n\tilde{V}_{nt} - \sigma_T^2 tr\ddot{G}_n) \\ \frac{1}{\sigma_T^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}(\tilde{V}'_{nt}H'_n\tilde{V}_{nt} - \sigma_T^2 trH_n) \\ \frac{1}{2\sigma_T^4}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}(\tilde{V}'_{nt}\tilde{V}_{nt} - n\sigma_T^2) \end{pmatrix}. \quad (38)$$

From the second order condition in (37), we have $\Sigma^d_{\theta_T,nT}$ given in Box II.

B.2. The first and second order derivatives of (7) for the transformation approach

For the first and second order derivatives of (7),

$$\frac{\partial \ln L_{n,T}(\theta)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'\tilde{V}_{nt}(\zeta) \\ \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\zeta) - (T-1)trG_n(\lambda) \\ \frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\zeta))'\tilde{V}_{nt}(\zeta) - (T-1)trH_n(\rho) \\ \frac{1}{2\sigma^4}\sum_{t=1}^{T}\left(\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta) - n\frac{T-1}{T}\sigma^2\right) \end{pmatrix}, \quad (40)$$

and see Box III. At true θ_0, we have the equation in Box IV and the information matrix $\Sigma_{\theta_0,nT} = -E\left(\frac{1}{n(T-1)}\frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'}\right) =$

$$\frac{1}{\sigma_0^2}\begin{pmatrix} \mathcal{H}_{nT} & * & * \\ \mathbf{0}_{1\times(k+1)} & 0 & * \\ \mathbf{0}_{1\times(k+1)} & 0 & 0 \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{k\times k} & * & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n}tr(\ddot{G}^s_n\ddot{G}_n) & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n}tr(H^s_n\ddot{G}_n) & \frac{1}{n}tr(H^s_nH_n) & * \\ \mathbf{0}_{1\times k} & \frac{1}{\sigma_0^2 n}tr(\ddot{G}_n) & \frac{1}{\sigma_0^2 n}tr(H_n) & \frac{1}{2\sigma_0^4} \end{pmatrix}, \quad (43)$$

where $\mathcal{H}_{nT} = \frac{1}{n(T-1)}\sum_{t=1}^{T}(\ddot{X}_{nt}, \ddot{G}_n\ddot{X}_{nt}\beta_0)'(\ddot{X}_{nt}, \ddot{G}_n\ddot{X}_{nt}\beta_0)$.

B.3. Proof for Theorem 1 (Consistency)

Without loss of generality, we will present the analysis under the asymptotic setting that n tends to infinity with a fixed finite T. The extension to the case with infinity T is immediate. We first prove the consistency of the estimates of (λ_0, ρ_0) via the concentrated likelihood, which are the same (up to a constant proportion) for the direct and transformation approaches. The probability limits of the estimates of other parameters for both approaches can then be derived.

Global identification of (λ_0, ρ_0):

Corresponding to $\frac{1}{n(T-1)}\ln L_{n,T}(\lambda, \rho)$ in (13), define $Q_{n,T}(\lambda, \rho) = \max_{\beta,\sigma^2} E\frac{1}{n(T-1)}\ln L_{n,T}(\theta)$. Denote

$$\mathcal{H}_{\lambda_0,nT}(\rho) = \mathcal{H}_{3,nT}(\rho) - \mathcal{H}'_{2,nT}(\rho)\mathcal{H}^{-1}_{1,nT}(\rho)\mathcal{H}_{2,nT}(\rho),$$

where $\mathcal{H}_{i,nT}(\rho)$ for $i = 1, 2, 3$ are the corresponding components of $\mathcal{H}_{nT}(\rho)$ in (15). We have

$$Q_{n,T}(\lambda, \rho) = -(\ln(2\pi) + 1) - \frac{1}{2}\ln\sigma^{*2}_{nT}(\lambda, \rho) + \frac{1}{n}[\ln|S_n(\lambda)| + \ln|R_n(\rho)|], \quad (44)$$

where

$$\sigma^{*2}_{nT}(\lambda, \rho) = (\lambda_0 - \lambda)^2\mathcal{H}_{\lambda_0,nT}(\rho) + \sigma_0^2\frac{1}{n}tr(R_n'^{-1}S_n'^{-1}S'_n(\lambda)R'_n(\rho)R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}). \quad (45)$$

At the true parameters, $Q_{n,T}(\lambda_0, \rho_0) = -\frac{1}{2}(\ln 2\pi + 1) - \frac{1}{2}\ln\sigma_0^2 + \frac{1}{n}\ln|S_n(\lambda_0)| + \frac{1}{n}\ln|R_n(\rho_0)|$. We are going to prove that $\lim Q_{n,T}(\lambda, \rho) < \lim Q_{n,T}(\lambda_0, \rho_0)$ for any $(\lambda, \rho) \neq (\lambda_0, \rho_0)$. We have

$$\begin{aligned} Q_{n,T}(\lambda, \rho) - Q_{n,T}(\lambda_0, \rho_0) &= -\frac{1}{2}[\ln\sigma^{*2}_{nT}(\lambda, \rho) - \ln\sigma_0^2] \\ &\quad + \frac{1}{n}\ln|S_n(\lambda)| - \frac{1}{n}\ln|S_n(\lambda_0)| + \frac{1}{n}\ln|R_n(\rho)| - \frac{1}{n}\ln|R_n(\rho_0)| \\ &= T_{1,n}(\lambda, \rho) - T_{2,nT}(\lambda, \rho), \end{aligned}$$

where

$$\begin{aligned} T_{1,n}(\lambda, \rho) &= -\frac{1}{2}[\ln\sigma_n^2(\lambda, \rho) - \ln\sigma_0^2] \\ &\quad + \frac{1}{n}\ln|S_n(\lambda)| - \frac{1}{n}\ln|S_n(\lambda_0)| \\ &\quad + \frac{1}{n}\ln|R_n(\rho)| - \frac{1}{n}\ln|R_n(\rho_0)|, \end{aligned}$$

and $T_{2,nT}(\lambda, \rho) = \ln\left(1 + \frac{(\lambda_0-\lambda)^2(\mathcal{H}_{3,nT}(\rho) - \mathcal{H}'_{2,nT}(\rho)\mathcal{H}^{-1}_{1,nT}(\rho)\mathcal{H}_{2,nT}(\rho))}{\sigma_n^2(\lambda,\rho)}\right)$.

Consider the pure spatial process $Y_{nt} = \lambda_0 W_n Y_{nt} + U_{nt}$ with $U_{nt} = \rho_0 M_n U_{nt} + V_{nt}$ for a period t. The log likelihood function of this transformed process is

$$\ln L_{p,n}(\lambda, \rho, \sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 + \ln|S_n(\lambda)| + \ln|R_n(\rho)| - \frac{1}{2\sigma^2}V'_{nt}(\lambda, \rho)J_nV'_{nt}(\lambda, \rho),$$

where $V_{nt}(\lambda, \rho) = R_n(\rho)S_n(\lambda)Y_{nt}$. Let $Q_{p,n}(\lambda, \rho) = \max_{\sigma^2} E\frac{1}{n}\ln L_{p,n}(\lambda, \rho, \sigma^2)$ and $Q_{p,n}(\lambda_0, \rho_0)$ be $Q_{p,n}(\lambda, \rho)$ evaluated at (λ_0, ρ_0). It follows that $Q_{p,n}(\lambda, \rho) - Q_{p,n}(\lambda_0, \rho_0) = T_{1,n}(\lambda, \rho)$. By the information inequality, $Q_{p,n}(\lambda, \rho) - Q_{p,n}(\lambda_0, \rho_0) \leq 0$. Thus, $T_{1,n}(\lambda, \rho) \leq 0$ for any (λ, ρ). Also, as $(\lambda_0 - \lambda)^2(\mathcal{H}_{3,nT}(\rho) - \mathcal{H}'_{2,nT}(\rho)\mathcal{H}^{-1}_{1,nT}(\rho)\mathcal{H}_{2,nT}(\rho))$ is a quadratic function of λ given ρ and $\sigma_n^2(\lambda, \rho)$ is bounded away from zero,[29] $T_{2,nT}(\lambda, \rho) \geq 0$.

29 The $Q_{p,n}(\lambda, \rho) - Q_{p,n}(\lambda_0, \rho_0) \leq 0$ for any (λ, ρ) implies that $-\frac{1}{2}\ln\sigma_n^2(\lambda, \rho) \leq -\frac{1}{2}\ln\sigma_0^2 + \frac{1}{n}\ln|S_n(\lambda)| - \frac{1}{n}\ln|S_n(\lambda_0)| + \frac{1}{n}\ln|R_n(\rho)| - \frac{1}{n}\ln|R_n(\rho_0)|$. As $\frac{1}{n}\ln|S_n(\lambda)| - \frac{1}{n}\ln|S_n(\lambda_0)|$ and $\frac{1}{n}\ln|R_n(\rho)| - \frac{1}{n}\ln|R_n(\rho_0)|$ are $O(1)$ uniformly in (λ, ρ), $-\ln\sigma_n^2(\lambda, \rho)$ is bounded from above as σ_0^2 is bounded away from 0. Hence, $\sigma_n^2(\lambda, \rho)$ is bounded away from 0.

$$\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^{d}(\theta)}{\partial\theta}=\begin{pmatrix}\frac{1}{\sigma^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'\tilde{V}_{nt}(\zeta)\\ \frac{1}{\sigma^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\left((R_n(\rho)W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\zeta)-\sigma^2 trG_n(\lambda)\right)\\ \frac{1}{\sigma^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}\left((H_n(\rho)\tilde{V}_{nt}(\zeta))'\tilde{V}_{nt}(\zeta)-\sigma^2 trH_n(\rho)\right)\\ \frac{1}{2\sigma^4}\frac{1}{\sqrt{nT}}\sum_{t=1}^{T}(\tilde{V}_{nt}'(\zeta)\tilde{V}_{nt}(\zeta)-n\sigma^2)\end{pmatrix}, \tag{36}$$

$$\begin{aligned}&\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^{d}(\theta)}{\partial\theta\partial\theta'}\\ &=-\frac{1}{nT}\begin{pmatrix}\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'R_n(\rho)\tilde{X}_{nt} & * & * & *\\ \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'R_n(\rho)\tilde{X}_{nt} & \left[\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'R_n(\rho)W_n\tilde{Y}_{nt}+Ttr(G_n^2(\lambda))\right] & * & *\\ \left[\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'H_n(\rho)\tilde{V}_{nt}(\zeta)+\frac{1}{\sigma^2}\sum_{t=1}^{T}(M_n\tilde{X}_{nt})'\tilde{V}_{nt}(\zeta)\right] & \left[\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})_n'H_n(\rho)\tilde{V}_{nt}(\zeta)+\frac{1}{\sigma^2}\sum_{t=1}^{T}(M_nW_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\zeta)\right] & 0 & 0\\ \frac{1}{\sigma^4}\sum_{t=1}^{T}\tilde{V}_{nt}'(\zeta)R_n(\rho)\tilde{X}_{nt} & \frac{1}{\sigma^4}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\zeta) & 0 & 0\end{pmatrix}\\ &\quad-\frac{1}{nT}\begin{pmatrix}\mathbf{0}_{k\times k} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times 1}\\ \mathbf{0}_{1\times k} & 0 & 0 & 0\\ \mathbf{0}_{1\times k} & 0 & \left[\frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\zeta))'H_n(\rho)\tilde{V}_{nt}(\zeta)+T\operatorname{tr}(H_n^2(\rho))\right] & *\\ \mathbf{0}_{1\times k} & 0 & \frac{1}{\sigma^4}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\zeta))'\tilde{V}_{nt}(\zeta) & -\frac{nT}{2\sigma^4}+\frac{1}{\sigma^6}\sum_{t=1}^{T}(\tilde{V}_{nt}'(\zeta)\tilde{V}_{nt}(\zeta))\end{pmatrix}\end{aligned} \tag{37}$$

Box I.

$$\begin{aligned}\Sigma_{\theta_T,nT}^{d}&=-E\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^{d}(\theta_T)}{\partial\theta\partial\theta'}\\ &=\begin{pmatrix}\frac{1}{\sigma_T^2 nT}\sum_{t=1}^{T}\ddot{X}_{nt}'\ddot{X}_{nt} & * & * & *\\ \frac{1}{\sigma_T^2 nT}\sum_{t=1}^{T}(\ddot{G}_n\ddot{X}_{nt}\beta_0)'\ddot{X}_{nt} & \frac{1}{\sigma_T^2 nT}\sum_{t=1}^{T}(\ddot{G}_n\ddot{X}_{nt}\beta_0)'\ddot{G}_n\ddot{X}_{nt}\beta_0+\frac{1}{n}tr\ddot{G}_n^s\ddot{G}_n & * & *\\ \mathbf{0}_{1\times k} & \frac{1}{n}tr(H_n^s\ddot{G}_n) & \frac{1}{n}tr(H_n^sH_n) & *\\ \mathbf{0}_{1\times k} & \frac{1}{\sigma_T^2 n}tr(\ddot{G}_n) & \frac{1}{\sigma_T^2 n}tr(H_n) & \frac{1}{2\sigma_T^4}\end{pmatrix}\end{aligned} \tag{39}$$

Box II.

$$\begin{aligned}&-\frac{\partial^2 \ln L_{n,T}(\theta)}{\partial\theta\partial\theta'}\\ &=\begin{pmatrix}\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'R_n(\rho)\tilde{X}_{nt} & * & * & *\\ \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'R_n(\rho)\tilde{X}_{nt} & \left[\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'R_n(\rho)W_n\tilde{Y}_{nt}+(T-1)tr(G_n^2(\lambda))\right] & * & *\\ \left[\frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\zeta))'R_n(\rho)\tilde{X}_{nt}+\frac{1}{\sigma^2}\sum_{t=1}^{T}\tilde{V}_{nt}'(\zeta)M_n\tilde{X}_{nt}\right] & \left[\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})_n'H_n(\rho)\tilde{V}_{nt}(\zeta)+\frac{1}{\sigma^2}\sum_{t=1}^{T}(M_nW_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\zeta)\right] & 0 & 0\\ \frac{1}{\sigma^4}\sum_{t=1}^{T}\tilde{V}_{nt}'(\zeta)R_n(\rho)\tilde{X}_{nt} & \frac{1}{\sigma^4}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\zeta) & 0 & 0\end{pmatrix}\\ &\quad+\begin{pmatrix}\mathbf{0}_{k\times k} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times 1}\\ \mathbf{0}_{1\times k} & 0 & 0 & 0\\ \mathbf{0}_{1\times k} & 0 & \left[\frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\zeta))'H_n(\rho)\tilde{V}_{nt}(\zeta)+(T-1)tr(H_n^2(\rho))\right] & *\\ \mathbf{0}_{1\times k} & 0 & \frac{1}{\sigma^4}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\zeta))'\tilde{V}_{nt}(\zeta) & -\frac{n(T-1)}{2\sigma^4}+\frac{1}{\sigma^6}\sum_{t=1}^{T}(\tilde{V}_{nt}'(\zeta)\tilde{V}_{nt}(\zeta))\end{pmatrix}\end{aligned} \tag{41}$$

Box III.

$$\frac{1}{\sqrt{n(T-1)}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma_0^2}\frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^T \ddot{X}_{nt}'\tilde{V}_{nt} \\ \frac{1}{\sigma_0^2}\frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^T (\ddot{G}_n\ddot{X}_{nt}\beta_0)'\tilde{V}_{nt} + \frac{1}{\sigma_0^2}\frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^T\left(\tilde{V}_{nt}'\ddot{G}_n'\tilde{V}_{nt} - \frac{T-1}{T}\sigma_0^2 tr\ddot{G}_n\right) \\ \frac{1}{\sigma_0^2}\frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^T\left(\tilde{V}_{nt}'H_n'\tilde{V}_{nt} - \frac{T-1}{T}\sigma_0^2 trH_n\right) \\ \frac{1}{2\sigma_0^4}\frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^T\left(\tilde{V}_{nt}'\tilde{V}_{nt} - n\frac{T-1}{T}\sigma_0^2\right)\end{pmatrix} \quad (42)$$

Box IV.

Under Assumption 7(a), $\mathcal{H}_{3,nT}(\rho) - \mathcal{H}_{2,nT}'(\rho)\mathcal{H}_{1,nT}^{-1}(\rho)\mathcal{H}_{2,nT}(\rho)$ is positive so that $T_{2,nT}(\lambda,\rho) > 0$ for $\lambda \neq \lambda_0$ given any ρ. Given λ_0, ρ_0 is the unique maximizer of $\lim T_{1,n}(\lambda,\rho)$ under

$$\lim\left(\frac{1}{n}\ln\left|\sigma_0^2 R_n^{-1\prime}R_n^{-1}\right| - \frac{1}{n}\ln\left|\sigma_n^2(\rho)R_n^{-1}(\rho)'R_n^{-1}(\rho)\right|\right) \neq 0$$

for $\rho \neq \rho_0$.

Hence, (λ_0,ρ_0) are identified. When Assumption 7(a) fails, identification requires that $T_{1,n}(\lambda,\rho)$ is strictly less than zero. Under Assumption 7(b), we will have $T_{1,n}(\lambda,\rho) < 0$ whenever $(\lambda,\rho) \neq (\lambda_0,\rho_0)$. Hence, $\lim[Q_{n,T}(\lambda_0,\rho_0) - Q_{n,T}(\lambda,\rho)] > 0$ if $(\lambda,\rho) \neq (\lambda_0,\rho_0)$. This proves the global identification.

Uniform convergence of $\frac{1}{n(T-1)}\ln L_{n,T}(\lambda,\rho) - Q_{n,T}(\lambda,\rho)$:

As $\frac{1}{nT}\ln L_{n,T}^d(\lambda,\rho) - Q_{n,T}^d(\lambda,\rho) = \frac{1}{n(T-1)}\ln L_{n,T}(\lambda,\rho) - Q_{n,T}(\lambda,\rho)$ where $Q_{n,T}^d(\lambda,\rho) = \max_{\beta,\sigma^2} E\frac{1}{nT}\ln L_{n,T}^d(\theta)$, we prove the latter for simplicity. Denote

$$\tilde{\mathcal{X}}_{x,nt}(\rho) = R_n(\rho)(G_n\tilde{X}_{nt}\beta_0 - \tilde{X}_{nt}'\mathcal{H}_{1,nT}^{-1}(\rho)\mathcal{H}_{2,nT}(\rho)),$$

and

$$\mathcal{V}_{x,nT}(\rho) = \frac{1}{n(T-1)}\sum_{t=1}^T \tilde{X}_{nt}'R_n'(\rho)R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}\tilde{V}_{nt}.$$

We have

$$\begin{aligned}\hat{\sigma}_{nT}^2(\lambda,\rho) &= (\lambda_0-\lambda)^2\mathcal{H}_{\lambda_0,nT}(\rho) \\ &+ \frac{1}{n(T-1)}\sum_{t=1}^T \tilde{V}_{nt}'R_n'^{-1}S_n'^{-1}S_n'(\lambda)R_n'(\rho)R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}\tilde{V}_{nt} \\ &+ 2(\lambda_0-\lambda)\frac{1}{n(T-1)}\sum_{t=1}^T \tilde{\mathcal{X}}_{x,nt}'(\rho)R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}\tilde{V}_{nt} \\ &- \mathcal{V}_{x,nT}'(\rho)\mathcal{H}_{1,nT}^{-1}(\rho)\mathcal{V}_{x,nT}(\rho). \end{aligned} \quad (46)$$

Hence,

$$\hat{\sigma}_{nT}^2(\lambda_0,\rho_0) = \sigma_0^2 + O_p\left(\frac{1}{\sqrt{nT}}\right) \quad \text{and} \quad \sigma_{nT}^{*2}(\lambda_0,\rho_0) = \sigma_0^2. \quad (47)$$

From (13) and (44), $\frac{1}{n(T-1)}\ln L_{n,T}(\lambda,\rho) - Q_{n,T}(\lambda,\rho) = \frac{1}{2}\ln\sigma_{nT}^{*2}(\lambda,\rho) - \frac{1}{2}\ln\hat{\sigma}_{nT}^2(\lambda,\rho)$. By the mean value theorem, $\frac{1}{n(T-1)}\ln L_{n,T}(\lambda,\rho) - Q_{n,T}(\lambda,\rho) = -\frac{1}{2}\frac{1}{\tilde{\sigma}_{nT}^2(\lambda,\rho)}(\hat{\sigma}_{nT}^2(\lambda,\rho) - \sigma_{nT}^{*2}(\lambda,\rho))$ where $\tilde{\sigma}_{nT}^2(\lambda,\rho)$ lies between $\hat{\sigma}_{nT}^2(\lambda,\rho)$ and $\sigma_{nT}^{*2}(\lambda,\rho)$. We need to show that (1) $\hat{\sigma}_{nT}^2(\lambda,\rho) - \sigma_{nT}^{*2}(\lambda,\rho) \xrightarrow{p} 0$ uniformly in λ and ρ and (2) $\tilde{\sigma}_{nT}^2(\lambda,\rho)$ is bounded away from zero uniformly in λ and ρ in probability.

To prove (1): We have $\hat{\sigma}_{nT}^2(\lambda,\rho)$ and $\sigma_{nT}^{*2}(\lambda,\rho)$ in (45) and (46). When T is large, from Lemma 15 in Yu et al. (2008), $\frac{1}{nT}\sum_{t=1}^T\tilde{V}_{nt}'B_n\tilde{V}_{nt} - E\frac{1}{nT}\sum_{t=1}^T\tilde{V}_{nt}'B_n\tilde{V}_{nt} \xrightarrow{p} 0$ and $\frac{1}{nT}\sum_{t=1}^T\tilde{X}_{nt}'B_n\tilde{V}_{nt} \xrightarrow{p} 0$ for any UB matrix B_n. When n is large and T is finite, the results still hold by using Lemma A.12 in Lee (2004).[30] Hence, $[\hat{\sigma}_{nT}^2(\lambda,\rho) - \sigma_{nT}^{*2}(\lambda,\rho)] \xrightarrow{p} 0$ uniformly in λ and ρ. To prove (2): As $\tilde{\sigma}_{nT}^2(\lambda,\rho)$ lies between $\hat{\sigma}_{nT}^2(\lambda,\rho)$ and $\sigma_{nT}^{*2}(\lambda,\rho)$, we have $\frac{1}{\tilde{\sigma}_{nT}^2(\lambda,\rho)} \leq \max\{\frac{1}{\hat{\sigma}_{nT}^2(\lambda,\rho)}, \frac{1}{\sigma_{nT}^{*2}(\lambda,\rho)}\}$. As $\mathcal{H}_{3,nT}(\rho) - \mathcal{H}_{2,nT}'(\rho)\mathcal{H}_{1,nT}^{-1}(\rho)\mathcal{H}_{2,nT}(\rho)$ is nonnegative definite by the Cauchy–Schwarz inequality, and $\sigma_n^2(\lambda,\rho)$ is uniformly bounded away from zero, $\hat{\sigma}_{nT}^2(\lambda,\rho)$ and $\sigma_{nT}^{*2}(\lambda,\rho)$ are uniformly bounded away from zero. Hence, $\frac{1}{\tilde{\sigma}_{nT}^2(\lambda,\rho)}$ is uniformly bounded. Combining (1) $\hat{\sigma}_{nT}^2(\lambda,\rho) - \sigma_{nT}^{*2}(\lambda,\rho) \xrightarrow{p} 0$ uniformly in λ and ρ, and (2) $\frac{1}{\tilde{\sigma}_{nT}^2(\lambda,\rho)}$ is $O_p(1)$ uniformly in λ and ρ, we have $\frac{1}{n(T-1)}\ln L_{n,T}(\lambda,\rho) - Q_{n,T}(\lambda,\rho) \xrightarrow{p} 0$ uniformly in λ and ρ.

Uniform equicontinuity of $Q_{n,T}(\lambda,\rho)$:

From (44) and (45), $Q_{n,T}(\lambda,\rho)$ *is uniformly equicontinuous in* λ *and* ρ due to the facts: (1) $\frac{1}{n}\ln|S_n(\lambda)|$ and $\frac{1}{n}\ln|R_n(\rho)|$ are uniformly equicontinuous in λ and ρ; (2) $(\lambda-\lambda_0)^2\mathcal{H}_{\lambda_0,nT}(\rho)$ is uniformly equicontinuous in λ and ρ; (3) $\sigma_n^2(\lambda,\rho)$ is uniformly equicontinuous in λ and ρ.

Combining the global identification, uniform convergence and equicontinuity, the consistency of $(\hat{\lambda}_{nT},\hat{\rho}_{nT})$ and, equivalently, $(\hat{\lambda}_{nT}^d,\hat{\rho}_{nT}^d)$ follows.

Estimates for other parameters:

From (8), the consistency of $\hat{\beta}_{nT}^d(\hat{\lambda}_{nT}^d,\hat{\rho}_{nT}^d)$ can be easily obtained, where $(\hat{\beta}_{nT}^d,\hat{\lambda}_{nT}^d,\hat{\rho}_{nT}^d)$ is numerically the same as $(\hat{\beta}_{nT},\hat{\lambda}_{nT},\hat{\rho}_{nT})$ from Section 2.3. From (9) and (12), we can see that $\frac{T}{T-1}\hat{\sigma}_{nT}^{2d}(\hat{\lambda}_{nT}^d,\hat{\rho}_{nT}^d) - \sigma_0^2 \xrightarrow{p} 0$ and $\hat{\sigma}_{nT}^2(\hat{\lambda}_{nT}^d,\hat{\rho}_{nT}^d) - \sigma_0^2 \xrightarrow{p} 0$. Hence, the results follow. ■

B.4. Proof for Theorem 2 (Asymptotic Distribution)

For the direct approach, according to the Taylor expansion,

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

where $\bar{\theta}_{nT}^d$ lies between θ_T and $\hat{\theta}_{nT}^d$. As we have

$$-\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^d(\bar{\theta}_{nT}^d)}{\partial\theta\partial\theta'}$$

30 $\frac{1}{nT}\sum_{t=1}^T\tilde{V}_{nt}'B_n\tilde{V}_{nt} = \frac{1}{nT}\mathbf{V}_{nT}'A_{nT}\mathbf{V}_{nT}$ where $\mathbf{V}_{nT} = (V_{n1}',\ldots,V_{nT}')'$ and $A_{nT} = J_T\otimes B_n$. As A_{nT} is UB due to the special pattern of J_T and B_n being UB, $\frac{1}{nT}\mathbf{V}_{nT}'A_{nT}\mathbf{V}_{nT}$ is just a quadratic form of $\mathbf{V}_{nT}$ with a UB matrix A_{nT}.

$$= \left(\left(-\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^d(\bar{\theta}_{nT}^d)}{\partial\theta\partial\theta'} - \left(-\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^d(\theta_T)}{\partial\theta\partial\theta'}\right)\right) + \left(-\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^d(\theta_T)}{\partial\theta\partial\theta'} - \Sigma_{\theta_T,nT}^d\right) + \Sigma_{\theta_T,nT}^d \right.$$

where the first term is $\|\bar{\theta}_{nT}^d - \theta_T\| \cdot O_p(1)$ and the second term is $O_p(\frac{1}{\sqrt{nT}})$ (see Lemma A.3), $-\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^d(\bar{\theta}_{nT}^d)}{\partial\theta\partial\theta'} = \left\|\bar{\theta}_{nT}^d - \theta_T\right\| \cdot O_p(1) + O_p\left(\frac{1}{\sqrt{nT}}\right) + \Sigma_{\theta_T,nT}^d$.

Under Assumptions 7 and 8, $\Sigma_{\theta_T,nT}^d$ in (39) is nonsingular. We can prove the nonsingularity of the limiting information matrix by using an argument by contradiction (similar to Lee, 2004). We need to prove that $\lim \Sigma_{\theta_T,nT}^d c = 0$ implies $c = 0$ where $c = (c_1', c_2, c_3, c_4)'$, c_2, c_3, c_4 are scalars and c_1 is $k \times 1$ vector. With C_n and D_n defined in Assumption 8, $\frac{1}{n}tr(\ddot{G}_n^s\ddot{G}_n) - 2\left(\frac{tr\ddot{G}_n}{n}\right)^2 = \frac{1}{2n}tr(C_n^sC_n^s)$, $\frac{1}{n}tr(H_n^sH_n) - 2\left(\frac{trH_n}{n}\right)^2 = \frac{1}{2n}tr(D_n^sD_n^s)$ and $\frac{1}{n}tr(H_n^s\ddot{G}_n) - 2\frac{trH_n}{n}\frac{tr\ddot{G}_n}{n} = \frac{1}{2n}tr(C_n^sD_n^s)$. Also, denote $\mathcal{H}_{\beta,nT} = \frac{1}{nT}\sum_{t=1}^T \ddot{X}_{nt}'\ddot{X}_{nt}$, $\mathcal{H}_{\beta\lambda,nT} = \frac{1}{nT}\sum_{t=1}^T \ddot{X}_{nt}'\ddot{G}_n\ddot{X}_{nt}\beta_0$, $\mathcal{H}_{\lambda\beta,nT} = \mathcal{H}_{\beta\lambda,nT}'$ and $\mathcal{H}_{\lambda,nT} = \frac{1}{nT}\sum_{t=1}^T(\ddot{G}_n\ddot{X}_{nt}\beta_0)'\ddot{G}_n\ddot{X}_{nt}\beta_0$. By the method of substitution and elimination, $\lim \Sigma_{\theta_T,nT}^d c = 0$ will imply

$$\left\{\lim\left(\frac{1}{\sigma_T^2}\frac{1}{n}tr(D_n^sD_n^s)\left(\mathcal{H}_{\lambda,nT} - \mathcal{H}_{\lambda\beta,nT}(\mathcal{H}_{\beta,nT})^{-1}\mathcal{H}_{\beta\lambda,nT}\right) + \Phi_n\right)\right\} \times c_2 = 0$$

where $\Phi_n = \frac{1}{4n^2}\left[tr(C_n^sC_n^s)tr(D_n^sD_n^s) - tr^2(C_n^sD_n^s)\right]$ and $\mathcal{H}_{\lambda,nT} - \mathcal{H}_{\lambda\beta,nT}(\mathcal{H}_{\beta,nT})^{-1}\mathcal{H}_{\beta\lambda,nT}$ are nonnegative by the Cauchy–Schwarz inequality. Hence, the nonsingularity of $\lim \Sigma_{\theta_T,nT}^d$ follows from Assumption 7.

For $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^d(\theta_T)}{\partial\theta}$, it is a linear and quadratic form of $\tilde{V}_{nt}$ with zero mean because $E\tilde{V}_{nt}'\tilde{V}_{nt} = \frac{T-1}{T}n\sigma_0^2 = n\sigma_T^2$. For its variance, as $\ddot{X}_{nt}$ is uncorrelated with V_{nt}, using Lemma A.4, we have

$$E\left(\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^d(\theta_T)}{\partial\theta}\cdot\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^d(\theta_T)}{\partial\theta'}\right) = \frac{T}{T-1}(\Sigma_{\theta_T,nT}^d + \Omega_{\theta_T,n}^d)$$

where $\Sigma_{\theta_T,nT}^d$ is in (39) and

$$\Omega_{\theta_T,nT}^d = \frac{(T-1)}{T}\frac{(\mu_4 - 3\sigma_0^4)}{\sigma_0^4} \times \begin{pmatrix} \mathbf{0}_{k\times k} & * & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n}\sum_{i=1}^n \ddot{G}_{n,ii}^2 & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n}\sum_{i=1}^n \ddot{G}_{n,ii}H_{n,ii} & \frac{1}{n}\sum_{i=1}^n H_{n,ii}^2 & * \\ \mathbf{0}_{1\times k} & \frac{1}{2\sigma_T^2 n}tr\ddot{G}_n & \frac{1}{2\sigma_T^2 n}trH_n & \frac{1}{4\sigma_T^4} \end{pmatrix}. \quad (48)$$

When V_{nt} are normally distributed, $\Omega_{\theta_T,nT}^d = \mathbf{0}_{(k+3)\times(k+3)}$ because $\mu_4 - 3\sigma_0^4 = 0$. By using the central limit theorem in Lemma A.1, $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^d(\theta_T)}{\partial\theta} \xrightarrow{d} N(0, \lim \frac{T}{T-1}(\Sigma_{\theta_T,nT}^d + \Omega_{\theta_T,n}^d))$.

Because $\left\|\bar{\theta}_{nT}^d - \theta_T\right\| = o_p(1)$ and $\Sigma_{\theta_T,nT}^d$ is nonsingular in the limit, $\left(-\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^d(\bar{\theta}_{nT}^d)}{\partial\theta\partial\theta'}\right)^{-1}$ is $O_p(1)$. It follows that $\hat{\theta}_{nT}^d - \theta_T = O_p\left(\frac{1}{\sqrt{nT}}\right)$. Hence,

$$\sqrt{nT}(\hat{\theta}_{nT}^d - \theta_T) = \left(\Sigma_{\theta_T,nT}^d + O_p\left(\frac{1}{\sqrt{nT}}\right)\right)^{-1} \times \left(\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^d(\theta_T)}{\partial\theta}\right).$$

Using the fact that

$$\left(\Sigma_{\theta_T,nT}^d + O_p\left(\frac{1}{\sqrt{nT}}\right)\right)^{-1} = (\Sigma_{\theta_T,nT}^d)^{-1} + O_p\left(\frac{1}{\sqrt{nT}}\right),$$

we have

$$\sqrt{nT}(\hat{\theta}_{nT}^d - \theta_T) \xrightarrow{d} N(0, \lim \frac{T}{T-1}(\Sigma_{\theta_T,nT}^d)^{-1} \times (\Sigma_{\theta_T,nT}^d + \Omega_{\theta_T,nT}^d)(\Sigma_{\theta_T,nT}^d)^{-1}).$$

For the transformation approach, the proof is similar. For the variance matrix of the estimates $\hat{\theta}_{nT}$, $\Sigma_{\theta_0,nT}$ is in (43) and

$$\Omega_{\theta_0,nT} = \frac{(T-1)}{T}\frac{(\mu_4 - 3\sigma_0^4)}{\sigma_0^4} \times \begin{pmatrix} \mathbf{0}_{k\times k} & * & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n}\sum_{i=1}^n \ddot{G}_{n,ii}^2 & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n}\sum_{i=1}^n \ddot{G}_{n,ii}H_{n,ii} & \frac{1}{n}\sum_{i=1}^n H_{n,ii}^2 & * \\ \mathbf{0}_{1\times k} & \frac{1}{2\sigma_0^2 n}tr\ddot{G}_n & \frac{1}{2\sigma_0^2 n}trH_n & \frac{1}{4\sigma_0^4} \end{pmatrix}. \quad (49)$$

As the log likelihood function in the transformation approach has a proper degree of freedom adjustment (from nT to $n(T-1)$), the location of $\hat{\theta}_{nT}$ is properly centered at θ_0; while for the direct approach, θ_T provides the convenient location for analysis. ■

Appendix C. The direct and transformation approaches in Section 3

C.1. The first and second order derivatives of (21) for the Direct Approach

The first and second order derivatives of the concentrated log likelihood in (21) are

$$\frac{\partial \ln L_{n,T}^d(\theta)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^T(R_n(\rho)\tilde{X}_{nt})'J_n\tilde{V}_{nt}(\zeta) \\ \frac{1}{\sigma^2}\sum_{t=1}^T(R_n(\rho)W_n\tilde{Y}_{nt})'J_n\tilde{V}_{nt}(\zeta) - T\mathrm{tr}G_n(\lambda) \\ \frac{1}{\sigma^2}\sum_{t=1}^T(H_n(\rho)\tilde{V}_{nt}(\zeta))'J_n\tilde{V}_{nt}(\zeta) - T\mathrm{tr}H_n(\rho) \\ \frac{1}{2\sigma^4}\sum_{t=1}^T(\tilde{V}_{nt}'(\zeta)J_n\tilde{V}_{nt}(\zeta) - n\sigma^2) \end{pmatrix}, \quad (50)$$

and see the equation in Box V. For the first order derivative evaluated at θ_T, it has two components

$$\frac{\partial \ln L_{n,T}^d(\theta_T)}{\partial\theta} = \frac{\partial \ln L_{n,T}^{d,u}(\theta_T)}{\partial\theta} - T\cdot a_{\theta_T,n} \quad (52)$$

$$-\frac{\partial^2 \ln L_{n,T}^d(\theta)}{\partial\theta\partial\theta'}$$
$$= \begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)\tilde{X}_{nt})'J_nR_n(\rho)\tilde{X}_{nt} & * & * & * \\ \frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)W_n\tilde{Y}_{nt})'J_nR_n(\rho)\tilde{X}_{nt} & \left[\frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)W_n\tilde{Y}_{nt})'J_nR_n(\rho)W_n\tilde{Y}_{nt} + T\mathrm{tr}(G_n^2(\lambda))\right] & * & * \\ \left[\frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)\tilde{X}_{nt})'J_nH_n(\rho)\tilde{V}_{nt}(\zeta) + \frac{1}{\sigma^2}\sum_{t=1}^T (M_n\tilde{X}_{nt})'J_n\tilde{V}_{nt}(\zeta)\right] & \left[\frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)W_n\tilde{Y}_{nt})'_nJ_nH_n(\rho)\tilde{V}_{nt}(\zeta) + \frac{1}{\sigma^2}\sum_{t=1}^T (M_nW_n\tilde{Y}_{nt})'J_n\tilde{V}_{nt}(\zeta)\right] & 0 & 0 \\ \frac{1}{\sigma^4}\sum_{t=1}^T \tilde{V}'_{nt}(\zeta)J_nR_n(\rho)\tilde{X}_{nt} & \frac{1}{\sigma^4}\sum_{t=1}^T (R_n(\rho)W_n\tilde{Y}_{nt})'J_n\tilde{V}_{nt}(\zeta) & 0 & 0 \end{pmatrix}$$
$$+ \begin{pmatrix} \mathbf{0}_{k\times k} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times 1} \\ \mathbf{0}_{1\times k} & 0 & 0 & 0 \\ \mathbf{0}_{1\times k} & 0 & \left[\frac{1}{\sigma^2}\sum_{t=1}^T (H_n(\rho)\tilde{V}_{nt}(\zeta))'J_nH_n(\rho)\tilde{V}_{nt}(\zeta) + T\mathrm{tr}(H_n^2(\rho))\right] & * \\ \mathbf{0}_{1\times k} & 0 & \frac{1}{\sigma^4}\sum_{t=1}^T (H_n(\rho)\tilde{V}_{nt}(\zeta))'J_n\tilde{V}_{nt}(\zeta) & -\frac{nT}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{t=1}^T (\tilde{V}'_{nt}(\zeta)J_n\tilde{V}_{nt}(\zeta)) \end{pmatrix} \quad (51)$$

Box V.

$$\Sigma_{\theta_T,nT}^d = -E\frac{1}{nT}\frac{\partial \ln L_{n,T}^{2d}(\theta_T)}{\partial\theta\partial\theta'} = \begin{pmatrix} \frac{1}{\sigma_T^2 nT}\sum_{t=1}^T \ddot{X}'_{nt}J_n\ddot{X}_{nt} & * & * & * \\ \frac{1}{\sigma_T^2 nT}\sum_{t=1}^T (\ddot{G}_n\ddot{X}_{nt}\beta_0)'J_n\ddot{X}_{nt} & \frac{1}{\sigma_T^2 nT}\sum_{t=1}^T (\ddot{G}_n\ddot{X}_{nt}\beta_0)'J_n\ddot{G}_n\ddot{X}_{nt}\beta_0 + \frac{1}{n}tr\ddot{G}_n^sJ_n\ddot{G}_n & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n}tr(H_n^sJ_n\ddot{G}_n) & \frac{1}{n}tr(H_n^sH_n) & * \\ \mathbf{0}_{1\times k} & \frac{1}{\sigma_T^2 n}tr(\ddot{G}_n) & \frac{1}{\sigma_T^2 n}tr(H_n) & \frac{1}{2\sigma_T^4} \end{pmatrix} \quad (53)$$

Box VI.

where

$$\frac{\partial \ln L_{n,T}^{d,u}(\theta_T)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma_T^2}\sum_{t=1}^T \ddot{X}'_{nt}J_n\tilde{V}_{nt} \\ \frac{1}{\sigma_T^2}\sum_{t=1}^T (\ddot{G}_n\ddot{X}_{nt}\beta_0)'J_n\tilde{V}_{nt} + \frac{1}{\sigma_T^2}\sum_{t=1}^T (\tilde{V}'_{nt}\ddot{G}'_nJ_n\tilde{V}_{nt} - \sigma_T^2 tr\ddot{G}'_nJ_n) \\ \frac{1}{\sigma_T^2}\sum_{t=1}^T (\tilde{V}'_{nt}H'_nJ_n\tilde{V}_{nt} - \sigma_T^2 trH'_nJ_n) \\ \frac{1}{2\sigma_T^4}\sum_{t=1}^T (\tilde{V}'_{nt}J_n\tilde{V}_{nt} - (n-1)\sigma_T^2) \end{pmatrix},$$

and

$$a_{\theta_T,n} = \left(\mathbf{0}_{1\times k}, \frac{1}{n}l'_nR_nG_nR_n^{-1}l_n, \frac{1}{n}l'_nH_nl_n, \frac{1}{2\sigma_T^2}\right)'.$$

For the second order derivative evaluated at θ_T, see Box VI.

C.2. The first and second order derivatives of (24) for the transformation approach

Using $trG_n(\lambda) - tr(J_nG_n(\lambda)) = \frac{1}{1-\lambda}$ and $tr(G_n^2(\lambda)) - tr((J_nG_n(\lambda))^2) = \frac{1}{(1-\lambda)^2}$ (see Lee and Yu, forthcoming), the first and second order derivatives of the concentrated log likelihood function (24) are

$$\frac{\partial \ln L_{n,T}(\theta)}{\partial\theta} = \begin{pmatrix} \frac{\partial \ln L_{n,T}(\theta)}{\partial\beta} \\ \frac{\partial \ln L_{n,T}(\theta)}{\partial\lambda} \\ \frac{\partial \ln L_{n,T}(\theta)}{\partial\rho} \\ \frac{\partial \ln L_{n,T}(\theta)}{\partial\sigma^2} \end{pmatrix}$$
$$= \begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)\tilde{X}_{nt})'J_n\tilde{V}_{nt}(\zeta) \\ \frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)W_n\tilde{Y}_{nt})'J_n\tilde{V}_{nt}(\zeta) - (T-1)trJ_nG_n(\lambda) \\ \frac{1}{\sigma^2}\sum_{t=1}^T (H_n(\rho)\tilde{V}_{nt}(\zeta))'J_n\tilde{V}_{nt}(\zeta) - (T-1)trJ_nH_n(\rho) \\ \frac{1}{2\sigma^4}\sum_{t=1}^T \left(\tilde{V}'_{nt}(\zeta)J_n\tilde{V}_{nt}(\zeta) - (n-1)\frac{T-1}{T}\sigma^2\right) \end{pmatrix}, \quad (54)$$

and the equation in Box VII. The score vector and the information matrix are given in Box VIII and

$$\Sigma_{\theta_0,nT} = -E\left(\frac{1}{(n-1)(T-1)}\frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'}\right)$$
$$= \frac{1}{\sigma_0^2}\begin{pmatrix} \mathcal{H}_{nT} & * & * \\ \mathbf{0}_{1\times(k+1)} & 0 & * \\ \mathbf{0}_{1\times(k+1)} & 0 & 0 \end{pmatrix}$$

$$\frac{\partial^2 \ln L_{n,T}(\theta)}{\partial\theta\partial\theta'} = \begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'J_nR_n(\rho)\tilde{X}_{nt} & * & * & * \\ \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'J_nR_n(\rho)\tilde{X}_{nt} & \left[\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'J_nR_n(\rho)W_n\tilde{Y}_{nt} + (T-1)tr(J_nG_n^2(\lambda))\right] & * & * \\ \left[\frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\zeta))'J_nR_n(\rho)\tilde{X}_{nt} + \frac{1}{\sigma^2}\sum_{t=1}^{T}\tilde{V}'_{nt}(\zeta)J_nM_n\tilde{X}_{nt}\right] & \left[\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'_nJ_nH_n(\rho)\tilde{V}_{nt}(\zeta) + \frac{1}{\sigma^2}\sum_{t=1}^{T}(M_nW_n\tilde{Y}_{nt})'J_n\tilde{V}_{nt}(\zeta)\right] & 0 & 0 \\ \frac{1}{\sigma^4}\sum_{t=1}^{T}\tilde{V}'_{nt}(\zeta)J_nR_n(\rho)\tilde{X}_{nt} & \frac{1}{\sigma^4}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'J_n\tilde{V}_{nt}(\zeta) & 0 & 0 \end{pmatrix}$$
$$+ \begin{pmatrix} \mathbf{0}_{k\times k} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times 1} \\ \mathbf{0}_{1\times k} & 0 & 0 & 0 \\ \mathbf{0}_{1\times k} & 0 & \left[\frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\zeta))'J_nH_n(\rho)\tilde{V}_{nt}(\zeta) + (T-1)tr(J_nH_n^2(\rho))\right] & * \\ \mathbf{0}_{1\times k} & 0 & \frac{1}{\sigma^4}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\zeta))'J_n\tilde{V}_{nt}(\zeta) & \left[-\frac{(n-1)(T-1)}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{t=1}^{T}(\tilde{V}'_{nt}(\zeta)J_n\tilde{V}_{nt}(\zeta))\right] \end{pmatrix} \quad (55)$$

Box VII.

$$\frac{1}{\sqrt{(n-1)(T-1)}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma_0^2\sqrt{(n-1)(T-1)}}\sum_{t=1}^{T}(\ddot{X}'_{nt}J_n\tilde{V}_{nt}) \\ \frac{1}{\sigma_0^2\sqrt{(n-1)(T-1)}}\sum_{t=1}^{T}\left((\ddot{G}_n\ddot{X}_{nt}\beta_0)'J_n\tilde{V}_{nt}\right) + \frac{1}{\sigma_0^2\sqrt{(n-1)(T-1)}}\sum_{t=1}^{T}\left(\tilde{V}'_{nt}\ddot{G}_nJ_n\tilde{V}_{nt} - \frac{T-1}{T}\sigma_0^2 trJ_n\ddot{G}_n\right) \\ \frac{1}{\sigma_0^2\sqrt{(n-1)(T-1)}}\sum_{t=1}^{T}\left(\tilde{V}'_{nt}H_nJ_n\tilde{V}_{nt} - \frac{T-1}{T}\sigma_0^2 trJ_nH_n\right) \\ \frac{1}{2\sigma_0^4\sqrt{(n-1)(T-1)}}\sum_{t=1}^{T}\left(\tilde{V}'_{nt}J_n\tilde{V}_{nt} - \frac{T-1}{T}(n-1)\sigma_0^2\right) \end{pmatrix} \quad (56)$$

Box VIII.

$$+ \begin{pmatrix} \mathbf{0}_{k\times k} & * & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n-1}tr(\ddot{G}_n^sJ_n\ddot{G}_n) & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n-1}tr(H_n^sJ_n\ddot{G}_n) & \frac{1}{n-1}tr(H_n^sJ_nH_n) & * \\ \mathbf{0}_{1\times k} & \frac{1}{\sigma_0^2(n-1)}tr(J_n\ddot{G}_n) & \frac{1}{\sigma_0^2(n-1)}tr(J_nH_n) & \frac{1}{2\sigma_0^4} \end{pmatrix}, \quad (57)$$

where $\mathcal{H}_{nT} = \frac{1}{(n-1)(T-1)}\sum_{t=1}^{T}(\ddot{X}_{nt}, G_n\ddot{X}_{nt}\beta_0)'J_n(\ddot{X}_{nt}, G_n\ddot{X}_{nt}\beta_0)$.

C.3. *Proof for* Theorem 4 *(Asymptotic Distribution)*

For the direct approach, according to the Taylor expansion,

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

Here, $\bar{\theta}_{nT}^d$ lies between θ_T and $\hat{\theta}_{nT}^d$, and $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^d(\theta_T)}{\partial\theta} + \sqrt{\frac{T}{n}}a_{\theta_T,n} = \frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^{d(u)}(\theta_T)}{\partial\theta} \xrightarrow{d} N(0, \lim \frac{T}{T-1}(\Sigma_{\theta_T,nT}^d + \Omega_{\theta_T,nT}^d))$ where $\Sigma_{\theta_T,nT}^d$ is in (53) and

$$\Omega_{\theta_T,nT}^d = \frac{(T-1)}{T}\frac{(\mu_4 - 3\sigma_0^4)}{\sigma_0^4}$$

$$\times \begin{pmatrix} \mathbf{0}_{k\times k} & * & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n}\sum_{i=1}^{n}[(J_n\ddot{G}_n)_{ii}]^2 & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n}\sum_{i=1}^{n}[J_n\ddot{G}_n]_{ii}[J_nH_n]_{ii} & \frac{1}{n}\sum_{i=1}^{n}[(J_nH_n)_{ii}]^2 & * \\ \mathbf{0}_{1\times k} & \frac{1}{2\sigma_T^2 n}trJ_n\ddot{G}_n & \frac{1}{2\sigma_T^2 n}trJ_nH_n & \frac{1}{4\sigma_T^4} \end{pmatrix}. \quad (58)$$

Because $\left\|\bar{\theta}_{nT}^d - \theta_T\right\| = o_p(1)$ and $\Sigma_{\theta_T,nT}^d$ is nonsingular in the limit, $\left(-\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^d(\bar{\theta}_{nT}^d)}{\partial\theta\partial\theta'}\right)^{-1}$ is $O_p(1)$. Hence, $\sqrt{nT}(\hat{\theta}_{nT}^d - \theta_T) = O_p(1)(O_p(1) + O(\sqrt{\frac{T}{n}}))$, which implies that $\hat{\theta}_{nT}^d - \theta_T = O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{n}\right)\right)$. In turn, $\left(-\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^d(\bar{\theta}_{nT}^d)}{\partial\theta\partial\theta'}\right)^{-1} = (\Sigma_{\theta_T,nT}^d)^{-1} + O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{n}\right)\right)$. It follows that

$$\begin{aligned}\sqrt{nT}(\hat{\theta}_{nT}^d - \theta_T) &= \left(-\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^d(\bar{\theta}_{nT}^d)}{\partial\theta\partial\theta'}\right)\cdot\left(\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^{d(u)}(\theta_T)}{\partial\theta} - \sqrt{\frac{T}{n}}a_{\theta_T,n}\right) \\ &= (\Sigma_{\theta_T,nT}^d)^{-1}\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^{d(u)}(\theta_T)}{\partial\theta} + O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{n}\right)\right)\end{aligned}$$

$$\times \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}^{d(u)}(\theta_T)}{\partial \theta} - (\Sigma_{\theta_T,nT}^d)^{-1} \cdot \sqrt{\frac{T}{n}} a_{\theta_T,n} - O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{n}\right)\right) \cdot \sqrt{\frac{T}{n}} a_{\theta_T,n},$$

and, hence,

$$\sqrt{nT}(\hat{\theta}_{nT}^d - \theta_T) + (\Sigma_{\theta_T,nT}^d)^{-1} \cdot \sqrt{\frac{T}{n}} a_{\theta_T,n} + O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{n}\right)\right)\sqrt{\frac{T}{n}} a_{\theta_T,n} = ((\Sigma_{\theta_T,nT}^d)^{-1} + o_p(1)) \cdot \frac{1}{\sqrt{nT}} \frac{\partial \ln L_{n,T}^{d(u)}(\theta_T)}{\partial \theta}.$$

Therefore, we have the results in Theorem 4 for the direct approach.

For the transformation approach, the proof is similar. For the variance matrix of $\hat{\theta}_{nT}$, the information matrix $\Sigma_{\theta_0,nT}$ is in (57) and

$$\Omega_{\theta_0,nT} = \frac{(T-1)}{T} \frac{(\mu_4 - 3\sigma_0^4)}{\sigma_0^4} \times \begin{pmatrix} \mathbf{0}_{k\times k} & * & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n-1}\sum_{i=1}^n [(J_n\ddot{G}_n)_{ii}]^2 & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n-1}\sum_{i=1}^n [(J_n\ddot{G}_n)_{ii}(J_nH_n)_{ii}] & \frac{1}{n-1}\sum_{i=1}^n [(J_nH_n)_{ii}]^2 & * \\ \mathbf{0}_{1\times k} & \frac{1}{2\sigma_0^2(n-1)} tr(J_n\ddot{G}_n) & \frac{1}{2\sigma_0^2(n-1)} tr(J_nH_n) & \frac{1}{4\sigma_0^4} \end{pmatrix}. \quad (59)$$

Because a degree of freedom has been properly adjusted (from nT to $(n-1)(T-1)$) for the likelihood function in the transformation approach, the score has zero mean and the resulting asymptotic distribution is properly centered at the true parameter vector. ■

C.4. Proof for Theorem 5 (Bias Correction)

We have $\sqrt{nT}(\hat{\theta}_{nT}^d - \theta_T) + \sqrt{\frac{T}{n}}(\Sigma_{\theta_T,nT}^d)^{-1} a_{\theta_T,n} + O_p\left(\sqrt{\frac{T}{n^3}}\right) \xrightarrow{d} N(0, \lim \frac{T}{T-1}(\Sigma_{\theta_T,nT}^d)^{-1}(\Sigma_{\theta_T,nT}^d + \Omega_{\theta_T,nT}^d)(\Sigma_{\theta_T,nT}^d)^{-1})$ from Theorem 4. As the first step bias corrected estimator is $\hat{\theta}_{nT}^{d1} = \hat{\theta}_{nT}^d + \frac{1}{n}(\Sigma_{\hat{\theta}_{nT}^d,nT}^d)^{-1} a_n(\hat{\theta}_{nT}^d)$ where $a_n(\theta) = a_{\theta,n}$, we will have

$$\sqrt{nT}(\hat{\theta}_{nT}^{d1} - \theta_T) \xrightarrow{d} N(0, \lim \frac{T}{T-1}(\Sigma_{\theta_T,nT}^d)^{-1} \times (\Sigma_{\theta_T,nT}^d + \Omega_{\theta_T,nT}^d)(\Sigma_{\theta_T,nT}^d)^{-1})$$

if $\frac{T}{n^3} \to 0$ and

$$\sqrt{\frac{T}{n}}\left(\left(-\frac{1}{nT} E \frac{\partial^2 \ln L_{nT}^d(\hat{\theta}_{nT}^d)}{\partial\theta\partial\theta'}\right)^{-1} a_n(\hat{\theta}_{nT}^d) - (\Sigma_{\theta_T,nT}^d)^{-1} a_n(\theta_T)\right) \xrightarrow{p} 0,$$

where $-\frac{1}{nT} E \frac{\partial^2 \ln L_{nT}^d(\hat{\theta}_{nT}^d)}{\partial\theta\partial\theta'} = \Sigma_{\hat{\theta}_{nT}^d,nT}^d$ is the information matrix evaluated at $\hat{\theta}_{nT}^d$. The first condition is assumed in the theorem. For the second condition, as

$$\hat{\theta}_{nT}^d - \theta_T = O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{n}\right)\right) \quad \text{and}$$

$$-\frac{1}{nT} E \frac{\partial^2 \ln L_{nT}^d(\hat{\theta}_{nT})}{\partial\theta\partial\theta'} = (\Sigma_{\theta_T,nT}^d)^{-1} + O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{n}\right)\right),$$

we have

$$\sqrt{\frac{T}{n}}\left\{\left(-\frac{1}{nT} E \frac{\partial^2 \ln L_{nT}^d(\hat{\theta}_{nT})}{\partial\theta\partial\theta'}\right)^{-1} a_n(\hat{\theta}_{nT}^d) - (\Sigma_{\theta_T,nT}^d)^{-1} a_n(\theta_T)\right\}$$
$$= (\Sigma_{\theta_T,nT}^d)^{-1}\sqrt{\frac{T}{n}}\left(a_n(\hat{\theta}_{nT}^d) - a_n(\theta_T)\right) + a_n(\hat{\theta}_{nT}^d) \times O_p\left(\max\left(\frac{1}{n}, \sqrt{\frac{T}{n^3}}\right)\right)$$
$$= (\Sigma_{\theta_T,nT}^d)^{-1}\frac{\partial a_n(\theta_{nT}^*)}{\partial\theta'}\sqrt{\frac{T}{n}}\left(\hat{\theta}_{nT}^d - \theta_T\right) + a_n(\hat{\theta}_{nT}^d) \times O_p\left(\max\left(\frac{1}{n}, \sqrt{\frac{T}{n^3}}\right)\right)$$

where θ_{nT}^* lies between $\hat{\theta}_{nT}^d$ and θ_T. From the explicit form of $a_n(\theta)$, $\frac{\partial a_n(\theta_{nT}^*)}{\partial\theta'}$ is bounded in probability. Thus, as $\hat{\theta}_{nT}^d - \theta_T = O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{n}\right)\right)$, the second condition is satisfied under $\frac{n}{T^3} \to 0$. Consequently,

$$\sqrt{nT}(\hat{\theta}_{nT}^{d1} - \theta_T) \xrightarrow{d} N\left(0, \lim \frac{T}{T-1}(\Sigma_{\theta_T,nT}^d)^{-1} \times (\Sigma_{\theta_T,nT}^d + \Omega_{\theta_T,nT}^d)(\Sigma_{\theta_T,nT}^d)^{-1}\right).$$

The remaining bias in the variance parameter is adjusted in $\hat{\theta}_{nT}^{d2} = A_T \cdot \hat{\theta}_{nT}^{d1}$, where $A_T = \begin{pmatrix} I_{k+2} & \mathbf{0}_{(k+2)\times 1} \\ \mathbf{0}_{1\times(k+2)} & \frac{T}{T-1} \end{pmatrix}$. After this adjustment, (35) follows. ■

References

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

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

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

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

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

Arellano, M., Hahn, J., 2005. Understanding bias in nonlinear panel models: Some recent developments. In: Blundell, R., Newey, W.K., Persson, T. (Eds.), Advances in Econometric and Economics: Theory and Applications. In: Ninth World Congress, vol. II. Econometric Society Monographs.

Baltagi, B., 1995. Econometric Analysis of Panel Data. John Wiley and Sons, New York.

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

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

Chamberlain, G., 1982. Multivariate regression models for panel data. Journal of Econometrics 18, 5–46.

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

Cox, D.R., 1975. Partial likelihood. Biometrika 62, 269–276.

Cox, D.R., Reid, N., 1987. Parameter orthogonality and approximate conditional inference. Journal of the Royal Statistical Society. Series B (Methodological) 49, 1–39.

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

Elhorst, J.P., 2003. Specification and estimation of spatial panel data models. International Regional Science Review 26, 244–268.

Ertur, C., Koch, W., 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 No. 07-10, Federal Reserve Bank of Boston.

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

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

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

Kalbfleisch, J.D., Sprott, D.A., 1970. Application of likelihood methods to models involving large numbers of parameters. Journal of the Royal Statistical Society. Series B (Methodological) 32, 175–208.

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

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

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

Kelejian, H.H., Prucha, I.R., 2007. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics (forthcoming).

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

Lancaster, T., 2000. The incidental parameter problem since 1948. Journal of Econometrics 95, 391–413.

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

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

Lee, L.F., 2007b. Identification and estimation of econometrics models with group interactions, contextual factors and fixed effects. Journal of Econometrics 140, 333–374.

Lee, L.F., Liu, X., 2006. Efficient GMM estimation of a spatial autoregressive model with autoregressive disturbances. Manuscript, Ohio State University.

Lee, L.F., Liu, X., Lin, X., 2008. Specification and estimation of social interaction models with network structure, contextual factors, correlation and fixed effects. Manuscript, Ohio State University.

Lee, L.F., Yu, J., 2007. A spatial dynamic panel data model with both time and individual fixed effects. Econometric Theory (forthcoming).

Lee, L.F., Yu, J., 2007. Near unit root in the spatial autoregressive model. Manuscript, Ohio State University.

Lin, X., Lee, L.F., 2005. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. Journal of Econometrics (forthcoming).

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

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

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

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

Yu, J., de Jong, R., Lee, L.F., 2007. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large: a nonstationary case. Manuscript, Ohio State University.

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

Yu, J., Lee, L.F., 2007. Estimation of unit root spatial dynamic panel data models. Econometric Theory (forthcoming).