Estimation of spatial autoregressive panel data models with fixed effects*

Lung-fei Lee
Department of Economics
Ohio State University
lflee@econ.ohio-state.edu

Jihai Yu
Department of Economics
University of Kentucky
jihai.yu@uky.edu

First draft: March, 2008
This draft: October, 2008

Abstract

This paper establishes asymptotic properties of quasi-maximum likelihood estimators for SAR panel data models with fixed effects and SAR disturbances, where the time periods T can be finite or large (and the number of spatial units n tends to infinity). A direct approach is to estimate all the parameters including the fixed effects. In the presence of fixed effects, because of the incidental parameter problem, some parameter estimates may be inconsistent or have asymptotic bias. We propose alternative estimation methods based on transformation. For the model with only individual effects, the transformation approach yields consistent estimators for the common parameters even T is finite. 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. For the model with both individual and time effects, the transformation approach yields consistent estimators of all the common parameters regardless T is finite or large. When we estimate both individual and time effects directly, consistency of the variance parameter requires both n and T to be large and consistency of other common parameters requires n to be large.

JEL classification: C13; C23; R15
Keywords: Spatial autoregression, Panel data, Fixed effects, Quasi-maximum likelihood estimation, Conditional likelihood

*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.

1. Introduction

Spatial econometrics consists of econometric techniques dealing with the interactions of economic units in space, which can be physical or economic characteristic. A sample may consist of cross sectional observations, where 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 but also control for the unobservable heterogeneity (e.g., Anselin 1988, Baltagi et al. 2003, 2007, Kapoor et al. 2007, Yu et al. 2007, 2008 and Yu and Lee 2007). Baltagi et al. (2003) consider the specification test for spatial correlation in a panel regression with error component and SAR disturbances. Kapoor et al. (2007) provide a rigorous theoretical analysis of a panel model with SAR disturbances and error components. Baltagi et al. (2007) generalize Baltagi et al. (2003) by allowing for spatial correlations in both individual and error components such that they might have different spatial autoregressive parameters, which encompasses the spatial correlation specifications in Baltagi et al. (2003) and Kapoor et al. (2007). Instead of random effect error components, an alternative specification for panel data models assumes fixed effects. The fixed effects specification has the advantage of robustness in that the fixed effects are allowed to correlate with included regressors in the model (Hausman, 1978). Yu et al. (2008, 2007) and Yu and Lee (2007) consider, respectively, the time and spatial lags in a panel data setting with stationarity, spatial cointegration, and unit roots in the time dimension.

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 to be estimated. For the linear panel regression model with fixed effects, the direct maximum likelihood (ML) approach will estimate jointly the common parameters of interest and fixed effects. The corresponding ML estimates (MLEs) of the regression coefficients are known as the within estimates, which happen to be the conditional 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 this variance parameter is exactly the one illustrated in Neyman and Scott (1948). For the SAR panel data models with individual effects, similar findings of the direct ML approach are found in this paper. The direct estimation approach will yield consistent estimates for the spatial and regression coefficients except the variance of the disturbances when T is small (but n is large).[2] For the SAR panel models with both individual and time effects, the direct estimation approach will be inconsistent for the estimation of the common parameters unless n is large. Even when both n and T are large so that individual and time effects can be consistently estimated, the estimates of the common

[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 (2007)). The initial value problem for the dynamic panel data model is well known (Nickell, 1981).

parameters would still exhibit asymptotic biases.

To eliminate the fixed effects, the method of conditional likelihood is used when effective sufficient statistics can be found for each of the fixed effects. For the linear regression and logit panel models, the time average of the dependent variables provides the sufficient statistic[3] (see Hsiao, 1986). For the normal panel regression model, the conditional likelihood can be constructed as a likelihood function 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 vector of ones, the transformed disturbances are uncorrelated. The transformed equation can be estimated by the quasi-maximum likelihood (QML) approach. For the more general model with both individual and time fixed effects, one may combine a transformation from $(I_n - \frac{1}{n} l_n l_n')$ with another transformation from $(I_T - \frac{1}{T} l_T l_T')$ to eliminate both the individual and time fixed effects. By exploring the generalized inverse of the transformed equation, one may end up with a QML approach for the transformed model[4]. The transformation approach for our models can either be justified as a conditional likelihood approach or a modified likelihood approach 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 considered in Baltagi et al. (2003) is $Y_{nt} = X_{nt}\beta_0 + \mathbf{c}_{n0} + U_{nt}$, $U_{nt} = \lambda_0 W_n U_{nt} + V_{nt}$, $t = 1, 2, ..., T$, 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 and the spatial correlation is in U_{nt}. A different specification has been considered in Kapoor et al. (2007) with $Y_{nt} = X_{nt}\beta_0 + U_{nt}^+$ and $U_{nt}^+ = \lambda_0 W_n U_{nt}^+ + \mathbf{d}_{n0} + V_{nt}$, $t = 1, 2, ..., T$, 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 and the variance parameters of $\mathbf{d}_{n0}$ and V_{nt}. The two panel 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 variance matrix in Kapoor et al. (2007) has a special pattern and its inverse can be easier to compute. The model in Kapoor et al. (2007) implies spatial correlations in both the individual and error components having the same spatial effect parameter. Baltagi et al. (2007) formulate a model which allows for spatial correlations in both individual and error components having different spatial parameters. Both Baltagi et al. (2003) and Baltagi et al. (2007) have emphasized on the test of spatial correlation in their models. With the fixed effects specification, all these panel models have the

[3] Effective sufficient statistics might not be available for many other models. The well-known example is the probit panel regression model, where the time average of the dependent variables does not provide the sufficient statistic 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 from a model 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 (2007a) 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.

same representation. 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. By regarding $(I_n - \lambda_0 W_n)^{-1}\mathbf{d}_{n0}$ as a vector of unknown fixed effect parameters, these 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 investigate its consistency and possible asymptotic distribution. We find that, when T is finite, the estimate of the variance parameter is inconsistent but the estimates of the other common parameters, including the spatial effect parameters, can be consistent and asymptotically normal. These results correspond to the finding of the incidental parameter problem in Neyman and Scott (1948), where the model is a simple normal panel data model with different means as individual effects and a common variance parameter. 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 which estimates both the individual and time effects, even both n and T are large, we might encounter asymptotic biases for those estimates. The asymptotic biases 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 transformation approach is not necessarily a conditional likelihood approach but it corresponds to a modified concentrated 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 SAR disturbances and individual effects is

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

where $Y_{nt} = (y_{1t}, y_{2t}, ..., y_{nt})'$ and $V_{nt} = (v_{1t}, v_{2t}, ..., v_{nt})'$ are $n \times 1$ column vectors and v_{it} is *i.i.d.* across i

[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.

and t with zero mean and variance σ_0^2, W_n is an $n \times n$ spatial weights matrix, which is predetermined and 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 $n \times 1$ column vector of fixed effects. Similarly, M_n is an $n \times n$ spatial weights matrix for the disturbance process. In practice, M_n may or may not be W_n.

In this paper, we consider first the estimation of the model including the fixed effects and investigate the possible incidental parameter issue. We then consider the estimation after the elimination of the fixed effects. We use an orthogonal transformation which includes the Helmert transformation as a special case to eliminate the fixed effects[7]. Our asymptotic analysis will pay special attention to the case where n tends to infinity but T can be finite or large[8]. 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 $t = 1, 2, \cdots, T$ where $\bar{Y}_{nT} = \frac{1}{T}\sum_{t=1}^{T} Y_{nt}$. Similarly, we define $\tilde{X}_{nt} = X_{nt} - \bar{X}_{nT}$ and $\tilde{V}_{nt} = V_{nt} - \bar{V}_{nT}$. A list of frequently used notations is provided in the 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.

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

Assumption 3. The parameter space Θ of θ is a compact set, where $\theta = (\beta', \lambda, \rho, \sigma^2)'$. $S_n(\lambda)$ and $R_n(\rho)$ are invertible for all $\lambda \in \Lambda$ and $\rho \in \mathbb{P}$, where Λ and $\mathbb{P}$ are compact. Furthermore, λ_0 is in the interior of Λ and ρ_0 is in the interior of $\mathbb{P}$.[9]

Assumption 4. The elements of X_{nt} are nonstochastic and bounded[10], uniformly in n and t. Also, under the asymptotic setting in Assumption 6, the limit of $\frac{1}{nT}\sum_{t=1}^{T}\tilde{X}'_{nt}\tilde{X}_{nt}$ 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)[11]. Also $S_n^{-1}(\lambda)$ and $R_n^{-1}(\rho)$ are UB[12], uniformly in $\lambda \in \Lambda$ and $\rho \in \mathbb{P}$.

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

[7] In dynamic panel data, the first difference or Helmert transformation have often been used to eliminate the individual effects (see Anderson and Hsiao (1981) and Arellano and Bover (1995) among others).

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

[9] 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 can be relaxed. This is so if we investigate the relevant likelihood by concentrating out the β and σ^2 estimates given λ and ρ. For simplicity on the analysis via the likelihood with θ, we adopt the compactness of the whole space Θ.

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

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

[12] 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 (2007b)).

Assumption 1 is a standard normalization assumption in spatial econometrics. 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. Hence, we do not impose that feature in Assumption 1. The zero diagonal assumption helps the interpretation of the spatial effect, as self-influence shall be excluded in practice. 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 are unknown heteroskedasticity, the MLE (QMLE) would not be consistent. Consistent methods such as the GMM in Lin and Lee (2005) and that in Kelejian and Prucha (2007) 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 and M_n are row normalized, a compact subset of (-1,1) has often been taken as the parameter space in theory. When exogenous variables X_{nt} are included in the model, it is convenient to assume that the exogenous regressors are uniformly bounded as in Assumption 4. Assumption 5 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

For the estimation of the linear panel regression model with fixed individual effects, the ML approach which estimates the fixed effects directly provides consistent estimates of the common parameters except σ_0^2, which are known as the within estimates. For the SAR panel model with fixed individual effects, one may wonder whether or not the MLEs of spatial effects and regression coefficients will yield consistent estimates when T is small. As we will see below, this direct approach will yield consistent estimator for all the common parameters except σ_0^2; and the estimator of σ_0^2 is consistent only when T is large.

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

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

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 asymptotic analysis on the estimator of θ_0 via the concentrated log likelihood function. Using the first order condition for $\mathbf{c}_n$, the log likelihood function with $\mathbf{c}_n$ concentrated out is

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

where $\tilde{V}_{nt}(\zeta) = R_n(\rho)[S_n(\lambda)\tilde{Y}_{nt} - \tilde{X}_{nt}\beta]$.

Denote $\theta_T = \theta_0 - (\mathbf{0}_{1\times(k+2)}, \frac{1}{T}\sigma_0^2)'$. We show that, the extremum estimator $\hat{\theta}^d_{nT}$ derived from (4) may converge to θ_T in the sense that $\hat{\theta}^d_{nT} - \theta_T \xrightarrow{p} 0$ as n goes to infinity. Denote $G_n = W_n S_n^{-1}$ and

$$\mathcal{H}_{nT}(\rho) = \frac{1}{n(T-1)} \sum_{t=1}^{T} (\tilde{X}_{nt}, G_n \tilde{X}_{nt}\beta_0)' R_n'(\rho) R_n(\rho) (\tilde{X}_{nt}, G_n \tilde{X}_{nt}\beta_0), \tag{5}$$

$$\sigma_n^2(\rho) = \frac{\sigma_0^2}{n} tr[(R_n(\rho)R_n^{-1})'(R_n(\rho)R_n^{-1})],$$

$$\sigma_n^2(\lambda,\rho) = \frac{\sigma_0^2}{n} tr[(R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1})'(R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1})].$$

Assumption 7. Either (a) the limit of $\mathcal{H}_{nT}(\rho)$ is nonsingular for each possible ρ in $\mathbb{P}$ and the limit of $\left(\frac{1}{n}\ln\left|\sigma_0^2 R_n^{-1\prime} R_n^{-1}\right| - \frac{1}{n}\ln\left|\sigma_n^2(\rho) R_n^{-1}(\rho)' R_n^{-1}(\rho)\right|\right)$ is not zero for $\rho \neq \rho_0$; or (b) the limit of

$$\left(\frac{1}{n}\ln\left|\sigma_0^2 R_n^{-1\prime} S_n^{-1\prime} S_n^{-1} R_n^{-1}\right| - \frac{1}{n}\ln\left|\sigma_n^2(\lambda,\rho) R_n^{-1}(\rho)' S_n^{-1}(\lambda)' S_n^{-1}(\lambda) R_n^{-1}(\rho)\right|\right)$$

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

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 rely on either Assumption 7 (a) or extra information on the order of magnitudes of λ_0 and ρ_0. 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_0^2 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). The identification and consistency are shown in the following theorem.

Theorem 1 *Under Assumptions 1-7, θ_0 is identified and for the QMLE $\hat{\theta}^d_{nT}$ based on (4), $\hat{\theta}^d_{nT} - \theta_T \xrightarrow{p} 0$.*
Proof. *See Appendix B.3.* ■

For this theorem, we shall emphasize that $\hat{\sigma}^{2d}_{nT}$ does not converge to σ_0^2 when T is a fixed finite value as n tends to infinity. The component $(\hat{\beta}^{d\prime}_{nT}, \hat{\lambda}^d_{nT}, \hat{\rho}^d_{nT})'$ of $\hat{\theta}^d_{nT}$ will be consistent even when T is small. But for $\hat{\sigma}^{2d}_{nT}$, it will be consistent only when T is large.

The asymptotic distribution of $\hat{\theta}^d_{nT}$ can be derived from the Taylor expansion of $\frac{\partial \ln L^d_{n,T}(\hat{\theta}^d_{nT})}{\partial\theta}$ around θ_T. Denote $C_n = G_n - \frac{trG_n}{n} I_n$ and $D_n = H_n - \frac{trH_n}{n} I_n$.

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

Assumption 8 is a condition for the nonsingularity of the limit of the information matrix $\Sigma^d_{\theta_T,nT}$ (see (30)). When the limit of $\mathcal{H}_{nT}(\rho_0)$ is singular, as long as the limit of $\frac{1}{n^2}\left[tr(C_n^s C_n^s) tr(D_n^s D_n^s) - tr^2(C_n^s D_n^s)\right]$ is strictly positive, the limit of $\Sigma^d_{\theta_T,nT}$ remains nonsingular. Using the Lemmas in Appendix B.2, we have the following theorem.

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

$$\sqrt{nT}(\hat{\theta}^d_{nT} - \theta_T) \xrightarrow{d} N(0, \lim \frac{T}{T-1}(\Sigma^d_{\theta_T,nT})^{-1}(\Sigma^d_{\theta_T,nT} + \Omega^d_{\theta_T,nT})(\Sigma^d_{\theta_T,nT})^{-1}), \tag{6}$$

where the lim *is taken under Assumption 6, and* $\Sigma^d_{\theta_T,nT}$*,* $\Omega^d_{\theta_T,nT}$ *are in (30) and (32). Additionally, if* v_{it}*'s are normally distributed,* $\sqrt{nT}(\hat{\theta}^d_{nT} - \theta_T) \xrightarrow{d} N(0, \lim \frac{T}{T-1}(\Sigma^d_{\theta_T,nT})^{-1})$.

Proof. *See Appendix B.4.* ∎

From Theorem 2, we see that $\hat{\beta}^{d\prime}_{nT}, \hat{\lambda}^d_{nT}$ and $\hat{\rho}^d_{nT}$ are properly centered at their true parameter values, but $\sqrt{nT}(\hat{\sigma}^{2d}_{nT} - \sigma^2_0)$ may have an asymptotic bias even T also tends to infinity, unless $\frac{n}{T}$ goes to zero. However, from Theorem 1, it is straightforward to construct the bias corrected estimates as

$$\hat{\theta}^{d1}_{nT} = (\hat{\beta}^{d\prime}_{nT}, \hat{\lambda}^d_{nT}, \hat{\rho}^d_{nT}, \frac{T}{T-1}\hat{\sigma}^{2d}_{nT})'. \tag{7}$$

The bias corrected estimate of σ^2_0 becomes consistent and properly centered at σ^2_0 when n goes to infinity, regardless whether T goes to infinity or T is a fixed finite value.

2.2. Transformation Approach

2.2.1. Data Transformation and Conditional Likelihood

To eliminate the individual effects, which are invariant over time, the simplest transformation for the sample observations is the deviation from the time mean operator, $J_T = (I_T - \frac{1}{T}l_T l'_T)$, as in the panel regression model. Because W_n is also time invariant, the variables in the deviation from time means would still be a SAR model. Such a transformed model consists of $\tilde{Y}_{nt} = \lambda_0 W_n \tilde{Y}_{nt} + \tilde{X}_{nt}\beta_0 + \tilde{U}_{nt}$ and $\tilde{U}_{nt} = \rho_0 M_n \tilde{U}_{nt} + \tilde{V}_{nt}$. However, the resulted disturbances $\tilde{V}_{nt}$ would be linearly dependent over the time dimension. To eliminate the individual fixed effects but without creating linear dependence in the resulted disturbances, a transformation can be based on the orthonormal matrix of J_T. Let $[F_{T,T-1}, \frac{1}{\sqrt{T}}l_T]$ be the orthonormal matrix of the eigenvectors of J_T, where $F_{T,T-1}$ is the $T \times (T-1)$ eigenvector matrix[13] corresponding to the eigenvalues of one. For any $n \times T$ matrix $[Z_{n1}, \cdots, Z_{nT}]$, define the transformed $n \times (T-1)$ matrix $[Z^*_{n1}, \cdots, Z^*_{n,T-1}] = [Z_{n1}, \cdots, Z_{nT}]F_{T,T-1}$. Denote $X^*_{nt} = [X^*_{nt,1}, X^*_{nt,2}, \cdots, X^*_{nt,k}]$. Then, (1) implies

$$Y^*_{nt} = \lambda_0 W_n Y^*_{nt} + X^*_{nt}\beta_0 + U^*_{nt}, \quad U^*_{nt} = \rho_0 M_n U^*_{nt} + V^*_{nt}, \quad t = 1, \cdots, T-1. \tag{8}$$

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

$$E(V^{*\prime}_{n1}, \cdots, V^{*\prime}_{n,T-1})'(V^{*\prime}_{n1}, \cdots, V^{*\prime}_{n,T-1}) = \sigma^2_0(F'_{T,T-1} \otimes I_n)(F_{T,T-1} \otimes I_n) = \sigma^2_0 I_{n(T-1)}.$$

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}.

[13] 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.

The log likelihood function of (8), as if the disturbances were normally distributed, is

$$\ln L_{n,T}(\theta) = -\frac{n(T-1)}{2}\ln(2\pi\sigma^2) + (T-1)[\ln|S_n(\lambda)| + \ln|R_n(\rho)|] - \frac{1}{2\sigma^2}\sum_{t=1}^{T-1} V_{nt}^{*\prime}(\zeta)V_{nt}^*(\zeta), \quad (9)$$

where $V_{nt}^*(\zeta) = R_n(\rho)[S_n(\lambda)Y_{nt}^* - X_{nt}^*\beta]$. Thus, $V_{nt}^* = V_{nt}^*(\zeta_0)$. The QMLE $\hat{\theta}_{nT}$ is the extremum estimator derived from the maximization of (9). For any n-dimensional column vectors p_{nt} and q_{nt}, as

$$\begin{aligned} \sum_{t=1}^{T-1} p_{nt}^{*\prime} q_{nt}^* &= (p_{n1}', \cdots, p_{nT}')(F_{T,T-1}\otimes I_n)(F_{T,T-1}'\otimes I_n)(q_{n1}', \cdots, q_{nT}')' \\ &= (p_{n1}', \cdots, p_{nT}')(J_T\otimes I_n)(q_{n1}', \cdots, q_{nT}')' = \sum_{t=1}^{T}\tilde{p}_{nt}'\tilde{q}_{nt} \end{aligned}$$

by using $(\tilde{p}_{n1}, \cdots, \tilde{p}_{nT}) = (p_{n1}, \cdots, p_{nT})J_T$, (9) can be rewritten as

$$\ln L_{n,T}(\theta) = -\frac{n(T-1)}{2}\ln(2\pi\sigma^2) + (T-1)[\ln|S_n(\lambda)| + \ln|R_n(\rho)|] - \frac{1}{2\sigma^2}\sum_{t=1}^{T} \tilde{V}_{nt}'(\zeta)\tilde{V}_{nt}(\zeta), \quad (10)$$

where $\tilde{V}_{nt}(\zeta) = R_n(\rho)[S_n(\lambda)\tilde{Y}_{nt} - \tilde{X}_{nt}\beta]$. For (10), its first and second order derivatives are (33) and (34) in Appendix C.

We note that the likelihood function in (10) has a conditional likelihood interpretation. It is the conditional likelihood conditional on the time average $\bar{Y}_{nT}$, which is a sufficient statistic for $\mathbf{c}_{n0}$ under normality. This is so as follows. (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 $\tilde{V}_{nt}$, $t = 1, \cdots, T$, are independent of $\bar{V}_{nT}$ under normality, the likelihood in (10) is the product of conditional likelihoods of Y_{nt}, $t = 1, ..., T$ conditional on $\bar{Y}_{nT}$ (Hsiao 1986, Lancaster 2000).

2.2.2. Asymptotic Properties

Under the identification conditions from Assumption 7, we obtain the consistency of the transformation estimates.

Theorem 3 *Under Assumptions 1-7, θ_0 is identified and, for the QMLE $\hat{\theta}_{nT}$ based on (10), $\hat{\theta}_{nT} - \theta_0 \xrightarrow{p} 0$.*

Proof. *See Appendix C.3.* ∎

The asymptotic distribution of the QMLE $\hat{\theta}_{nT}$ can be derived from the Taylor expansion of $\frac{\partial \ln L_{n,T}(\hat{\theta}_{nT})}{\partial\theta}$ around θ_0. At θ_0, the first order derivative of the log likelihood function involves both linear and quadratic functions of $\tilde{V}_{nt}$ in (35). The asymptotic distribution of $\frac{1}{\sqrt{n(T-1)}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta}$ can be derived from the central limit theorem for martingale difference arrays (see Lemma C.3 in Appendix C). The variance matrix of $\frac{1}{\sqrt{n(T-1)}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta}$ is equal to $\Sigma_{\theta_0,nT} + \Omega_{\theta_0,nT}$, where $\Sigma_{\theta_0,nT} = -E\left(\frac{1}{n(T-1)}\frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'}\right)$ and $\Omega_{\theta_0,nT}$ are, respectively, in (36) and (37). When V_{nt} are normally distributed, $\Omega_{\theta_0,nT} = 0$ because $\mu_4 - 3\sigma_0^4 = 0$. Also, under Assumptions 1-7, we have $\frac{1}{n(T-1)}\frac{\partial^2 \ln L_{n,T}(\theta)}{\partial\theta\partial\theta'} - \frac{1}{n(T-1)}\frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'} = \|\theta - \theta_0\|\cdot O_p(1)$ and $\frac{1}{n(T-1)}\frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'} + \Sigma_{\theta_0,nT} = O_p(\frac{1}{\sqrt{n(T-1)}})$ from Lemma A.3 in Appendix A. Combined with Lemma C.3, we have the following asymptotic distribution of $\hat{\theta}_{nT}$.

Theorem 4 *Under Assumptions 1-6 and 7(a); or 1-6, 7(b) and 8,*

$$\sqrt{n(T-1)}(\hat{\theta}_{nT}-\theta_0) \xrightarrow{d} N(0, \lim \Sigma_{\theta_0,nT}^{-1}(\Sigma_{\theta_0,nT}+\Omega_{\theta_0,nT})\Sigma_{\theta_0,nT}^{-1}). \tag{11}$$

Additionally, if v_{it}'s are normally distributed, $\sqrt{n(T-1)}(\hat{\theta}_{nT}-\theta_0) \xrightarrow{d} N(0, \lim \Sigma_{\theta_0,nT}^{-1})$.

Proof. *See Appendix C.4.* ■

Hence, after the data transformation to eliminate the individual effects, the QMLE is consistent and asymptotically normal even when T is small[14].

2.3. Comparison of the Two Approaches

One may compare the concentrated log likelihood function in (4) of the direct approach with the log likelihood function in (10) from the transformation approach. We see that the difference is on the use of T in (4) but $(T-1)$ in (10). A closer comparison of the two log likelihood with a further concentration is revealing.

For the direct approach with (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

$$\hat{\beta}_{nT}^d(\lambda,\rho) = [\textstyle\sum_{t=1}^T \tilde{X}_{nt}' R_n'(\rho) R_n(\rho) \tilde{X}_{nt}]^{-1} [\sum_{t=1}^T \tilde{X}_{nt}' R_n'(\rho) R_n(\rho) S_n(\lambda) \tilde{Y}_{nt}],$$

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

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

$$\ln L_{n,T}^d(\lambda,\rho) = -\frac{nT}{2}(\ln(2\pi)+1) - \frac{nT}{2}\ln \hat{\sigma}_{nT}^{2d}(\lambda,\rho) + T[\ln|S_n(\lambda)| + \ln|R_n(\rho)|]. \tag{12}$$

For the transformed approach with (10), the corresponding estimates are

$$\hat{\beta}_{nT}(\lambda,\rho) = [\textstyle\sum_{t=1}^T \tilde{X}_{nt}' R_n'(\rho) R_n(\rho) \tilde{X}_{nt}]^{-1} [\sum_{t=1}^T \tilde{X}_{nt}' R_n'(\rho) R_n(\rho) S_n(\lambda) \tilde{Y}_{nt}],$$

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

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

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

Note that $\hat{\beta}_{nT}(\lambda,\rho) = \hat{\beta}_{nT}^d(\lambda,\rho)$, but $\hat{\sigma}_{nT}^{2d}(\lambda,\rho) = \frac{T-1}{T}\hat{\sigma}_{nT}^2(\lambda,\rho)$. Hence, (12) can be rewritten as

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

By comparing (13) and (14), we see that they yield the same maximizer $(\hat{\lambda}_{nT}, \hat{\rho}_{nT})$. As $\hat{\beta}_{nT}^d(\lambda,\rho)$ and $\hat{\beta}_{nT}(\lambda,\rho)$ are the same, we can conclude that the QMLE of $\zeta_0 = (\beta_0', \lambda_0, \rho_0)'$ from the direct approach will

[14] We have emphasized the asymptotics when n goes to infinity but T can be finite or infinity under Assumption 6. The result in Theorem 4 is valid even when n is finite but T tends to infinity.

yield the same consistent estimate as the transformation approach. However, the estimation of σ_0^2 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 $\hat{\sigma}^2_{nT}(\lambda, \rho)$. From the relation $\hat{\sigma}^{2d}_{nT}(\lambda, \rho) = \frac{T-1}{T}\hat{\sigma}^2_{nT}(\lambda, \rho)$, we see that the bias corrected estimate $\hat{\theta}^{d1}_{nT}$ is numerically equivalent to the estimate $\hat{\theta}_{nT}$. Hence, the ML estimation of the SAR panel model with fixed individual effects shares some common features with those of the ML estimation of the linear panel regression model with fixed effects.

In some social interaction models (see, e.g., Case (1991)), if each unit has many neighbors, the QMLEs of the parameters in $\zeta_0 = (\beta_0', \lambda_0, \rho_0)'$ might have a lower rate of convergence. For the cross section SAR model without SAR disturbances, Lee (2004) shows that when the generated spatial regressor $G_nX_n\beta_0$ 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 "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 "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[15].

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) and Hahn and Moon (2006), etc). The model including both individual and time effects would be useful for empirical applications where the time effects might be important, for example, in growth theory and regional economics (see, e.g., Ertur and Koch (2007) and Foote (2007) for recent empirical applications of panel data models with both time dummies and spatial effects). Hence, we generalize (1) to

$$Y_{nt} = \lambda_0 W_n Y_{nt} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + U_{nt}, \quad U_{nt} = \rho_0 M_n U_{nt} + V_{nt}, \quad t = 1, 2, ..., T, \tag{15}$$

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[16]. 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 likelihood approach based on transformed data, where both the individual and time effects are eliminated. The direct approach

[15] We do not provide a rigorous analysis of this "many neighbors" case in this paper. However, by investigating the elements of information matrix of (12) 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.

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

is discussed in Section 3.1. For the transformation approach, we may first eliminate the individual effects in (15) by $F_{T,T-1}$ similar to (8), which yields

$$Y_{nt}^* = \lambda_0 W_n Y_{nt}^* + X_{nt}^* \beta_0 + \alpha_{t0}^* l_n + U_{nt}^*, \quad U_{nt}^* = \rho_0 M_n U_{nt}^* + V_{nt}^*, \quad t = 1, 2, ..., T-1, \tag{16}$$

where $[\alpha_{10}^* l_n, \alpha_{20}^* l_n, \cdots, \alpha_{T-1,0}^* l_n] = [\alpha_{10} l_n, \alpha_{20} l_n, \cdots, \alpha_{T0} l_n] F_{T,T-1}$ can be considered as the transformed time effects. We can make a further transformation to (16) to eliminate the time effects. This transformation approach is investigated in Section 3.2.

3.1. Direct Approach

With both time and individual effects, the log likelihood function of (15) is

$$\ln L_{n,T}^d(\theta, \mathbf{c}_n, \boldsymbol{\alpha}_T) = -\frac{nT}{2}\ln(2\pi\sigma^2) + T[\ln|S_n(\lambda)| + \ln|R_n(\rho)|] - \frac{1}{2\sigma^2}\sum_{t=1}^T V_{nt}'(\zeta, \mathbf{c}_n, \boldsymbol{\alpha}_T) V_{nt}(\zeta, \mathbf{c}_n, \boldsymbol{\alpha}_T), \tag{17}$$

where $V_{nt}(\zeta, \mathbf{c}_n, \boldsymbol{\alpha}_T) = R_n(\rho)[S_n(\lambda)Y_{nt} - X_{nt}\beta - \mathbf{c}_n - \alpha_t l_n]$ with $\boldsymbol{\alpha}_T = (\alpha_1, ..., \alpha_T)'$. Using the first order conditions for α_t and $\mathbf{c}_n$, the log likelihood function with both $\mathbf{c}_n$ and $\boldsymbol{\alpha}_T$ concentrated out is

$$\ln L_{n,T}^d(\theta) = -\frac{nT}{2}\ln(2\pi\sigma^2) + T[\ln|S_n(\lambda)| + \ln|R_n(\rho)|] - \frac{1}{2\sigma^2}\sum_{t=1}^T \tilde{V}_{nt}'(\zeta) J_n \tilde{V}_{nt}(\zeta), \tag{18}$$

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. For (18), the first and second order derivatives are, respectively, (38) and (39) in Appendix D.

The concentrated likelihood estimate of θ_0 from (18) can be derived from setting the first order derivatives in (38) to zero. Denote these estimates as $\hat{\theta}_{nT}^d$ and also denote[17]

$$\mathcal{H}_{nT}(\rho) = \frac{1}{(n-1)(T-1)}\sum_{t=1}^T (\tilde{X}_{nt}, G_n \tilde{X}_{nt}\beta_0)' R_n'(\rho) J_n R_n(\rho)(\tilde{X}_{nt}, G_n \tilde{X}_{nt}\beta_0), \tag{19}$$

$$\sigma_n^2(\rho) = \frac{\sigma_0^2}{n-1} tr[(R_n(\rho)R_n^{-1})' J_n (R_n(\rho)R_n^{-1})],$$

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

The following assumptions provide conditions for parameter identification. These assumptions modify those 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 $\frac{1}{nT}\sum_{t=1}^T \tilde{X}_{nt}' J_n \tilde{X}_{nt}$ exists and is nonsingular[18].

Assumption 7'. Either (a) the limit of $\mathcal{H}_{nT}(\rho)$ is nonsingular for each possible ρ in $\mathbb{P}$ and the limit of $\left(\frac{1}{n-1}\ln\left|\sigma_0^2 R_n^{-1\prime} J_n R_n^{-1}\right| - \frac{1}{n-1}\ln\left|\sigma_n^2(\rho) R_n^{-1}(\rho)' J_n R_n^{-1}(\rho)\right|\right)$ is not zero for $\rho \neq \rho_0$; or (b) the limit of $\left(\frac{1}{n-1}\ln\left|\sigma_0^2 R_n^{-1\prime} S_n^{-1\prime} J_n S_n^{-1} R_n^{-1}\right| - \frac{1}{n-1}\ln\left|\sigma_n^2(\lambda, \rho) R_n^{-1}(\rho)' S_n^{-1}(\lambda)' J_n S_n^{-1}(\lambda) R_n^{-1}(\rho)\right|\right)$ is not zero for $(\lambda, \rho) \neq (\lambda_0, \rho_0)$.

[17] $\mathcal{H}_{nT}(\rho)$, $\sigma_n^2(\rho)$ and $\sigma_n^2(\lambda, \rho)$ for Section 3 have different meanings from corresponding ones in Section 2 although they share the same notations. The difference is that, we have J_n matrix present in those in Section 3.

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

Assumption 8'. The limit of $\frac{1}{(n-1)^2}\left[tr(C_n^s C_n^s)tr(D_n^s D_n^s) - tr^2(C_n^s D_n^s)\right]$ is strictly positive, where $C_n = J_n G_n - \frac{trJ_nG_n}{n-1}J_n$ and $D_n = J_n H_n - \frac{trJ_nH_n}{n-1}J_n$.

When $M_n = W_n$ and $\lambda_0 \neq \rho_0$, the condition 7'(b) would not be satisfied as (λ_0, ρ_0) and (ρ_0, λ_0) could not be distinguished from each other, i.e., λ_0 and ρ_0 can interchange roles. Identification will rely on either Assumption 7'(a) or extra information on the order of magnitudes of λ_0 and ρ_0.

Theorem 5 *Under Assumptions 1-3,4',5,6 and 7', θ_0 is identified and, for the QMLE $\hat{\theta}_{nT}^d$ based on (18), $\hat{\theta}_{nT}^d - \theta_T \xrightarrow{p} 0$ where $\theta_T = \theta_0 - (\mathbf{0}_{1\times(k+2)}, \frac{1}{T}\sigma_0^2)'$.*

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

From Theorem 5, 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 in the main equation, which can be consistently estimated as n tends to infinity. However, the consistency of the variance parameter requires both n and T to be large.

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. However, for the score evaluated at θ_T, it will not be centered at zero when T is large, due to the incidental parameter problem from 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 (41) and[19]

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

Theorem 6 *Under Assumptions 1–3,4',5,6 and 7'(a); or 1–3,4',5,6, 7'(b) and 8',*

$$\sqrt{nT}(\hat{\theta}_{nT}^d - \theta_T) + \sqrt{\frac{T}{n}}b_{\theta_T,nT} + O_p\left(\sqrt{\frac{T}{n^3}}\right) \xrightarrow{d} N(0, \lim \frac{T}{T-1}(\Sigma_{\theta_T,nT}^d)^{-1}(\Sigma_{\theta_T,nT}^d + \Omega_{\theta_T,nT}^d)(\Sigma_{\theta_T,nT}^d)^{-1}), \quad (20)$$

where $\Sigma_{\theta_T,nT}^d$ and $\Omega_{\theta_T,nT}^d$ are in (41) and (42).
When $\frac{n}{T} \to c$, where $0 < c < \infty$,

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

When $\frac{n}{T} \to 0$, $n(\hat{\theta}_{nT}^d - \theta_T) + b_{\theta_T,nT} \xrightarrow{p} 0$.
When $\frac{n}{T} \to \infty$, $\sqrt{nT}(\hat{\theta}_{nT}^d - \theta_T) \xrightarrow{d} N(0, \lim \frac{T}{T-1}(\Sigma_{\theta_T,nT}^d)^{-1}(\Sigma_{\theta_T,nT}^d + \Omega_{\theta_T,nT}^d)(\Sigma_{\theta_T,nT}^d)^{-1})$.
Additionally, if v_{it}'s are normally distributed,

$$\sqrt{nT}(\hat{\theta}_{nT}^d - \theta_T) + \sqrt{\frac{T}{n}}b_{\theta_T,nT} + O_p\left(\sqrt{\frac{T}{n^3}}\right) \xrightarrow{d} N(0, \lim \frac{T}{T-1}(\Sigma_{\theta_T,nT}^d)^{-1}).$$

Proof. *See Appendix D.3.* ■

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

Hence, $\hat{\theta}_{nT}^d$ is $\sqrt{nT}$ consistent when n and T go to infinity but has the 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 bias is of the order $O(\max(\frac{1}{T}, \frac{1}{n}))$. The confidence interval for $\hat{\theta}_{nT}^d$ will not center properly around θ_0 when $\frac{n}{T} \to c$ for finite $c > 0$. The situation becomes worse when $\frac{n}{T}$ tends to zero. 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 the common parameter $\zeta_0 = (\beta_0', \lambda_0, \rho_0)'$ is still asymptotically centered, while only the estimate of σ_0^2 has the low T rate of convergence and its limiting distribution is degenerate[20]. In terms of consistency, when T is finite, the estimates of β_0, λ_0 and ρ_0 are consistent, but that of σ_0^2 is not, because the finite number of time effects only play the role of T additional regressors in the model, and the incidental parameter problem occurs due to the many individual effects. For ζ_0, there is no incidental parameter problem caused by a finite number of time dummies. The additional incidental parameter problem occurs when T goes to infinity at the same rate as or faster than n.[21] When T is relatively large, it generates the asymptotic biases in all the common parameters via $b_{\theta_T,nT}$.

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 compare analytically 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).[22]

The overall bias can be corrected in two steps – an additive correction and then 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

$$\hat{\theta}_{nT}^{d1} = \hat{\theta}_{nT}^d - \frac{\hat{B}_{nT}}{n}, \text{ and } \hat{\theta}_{nT}^{d2} = A_T \cdot \hat{\theta}_{nT}^{d1}, \tag{21}$$

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 D.5, $\hat{B}_{nT} + b_{\theta_T,nT} = O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{n}\right)\right)$. Hence, when $\frac{T}{n^3} \to 0$, $\hat{\theta}_{nT}^{d1}$ is $\sqrt{nT}$ consistent and asymptotically normal centered around θ_T. With rescaling[23] 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 centered around θ_0.

[20] 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 $\hat{\theta}_{nT}^d$.

[21] 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 asymptotic bias for 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).

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

[23] 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, asymptotic bias would be present unless $\frac{n}{T^3} \to 0$.

Theorem 7 *Under Assumptions 1–3,4',5, 6 and 7'(a); or 1–3,4',5,6, 7'(b) and 8', when* $\frac{T}{n^3} \to 0$,

$$\sqrt{nT}(\hat{\theta}_{nT}^{d2} - \theta_0) \xrightarrow{d} N(0, \lim \frac{T}{T-1}(\Sigma_{\theta_T,nT}^d)^{-1}(\Sigma_{\theta_T,nT}^d + \Omega_{\theta_T,nT}^d)(\Sigma_{\theta_T,nT}^d)^{-1}). \tag{22}$$

Additionally, if v_{it}*'s are normally distributed,* $\sqrt{nT}(\hat{\theta}_{nT}^{d2} - \theta_0) \xrightarrow{d} N(0, \frac{T}{T-1}\lim(\Sigma_{\theta_T,nT}^d)^{-1})$.

Proof. *See Appendix D.4.* ■

3.2. Transformation Approach

3.2.1. Data Transformation and the Likelihood Function

Our proposed method is motivated by the construction of the within estimator in the panel regression model. For a panel regression model with both individual and time effects, those 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 resulted variable is $y_{it} - y_{.t} - y_{i.} + y_{..}$. 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 (15), 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.

In order to eliminate the fixed effects but without creating linear dependence on the resulted disturbances, the transformations can be based on orthonormal matrices of eigenvectors of J_T and J_n. Let $(F_{n,n-1}, \frac{1}{\sqrt{n}}l_n)$ be the orthonormal matrix of J_n, where $F_{n,n-1}$ corresponds to the eigenvalues of ones and $\frac{1}{\sqrt{n}}l_n$ corresponds to the eigenvalue zero. Similar to Lee and Yu (2007a), we can further transform the n-dimensional vector Y_{nt}^* in (16) to an $(n-1)$-dimensional vector Y_{nt}^{**} such that $Y_{nt}^{**} = F'_{n,n-1}Y_{nt}^*$. For analytical purpose, we need W_n and M_n to be row normalized as argued above[24].

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, from Lemma A.2, (16) can be transformed into

$$Y_{nt}^{**} = \lambda_0(F'_{n,n-1}W_nF_{n,n-1})Y_{nt}^{**} + X_{nt}^{**}\beta_0 + U_{nt}^{**}, \quad U_{nt}^{**} = \rho_0(F'_{n,n-1}M_nF_{n,n-1})U_{nt}^{**} + V_{nt}^{**}, \tag{23}$$

for $t = 1, ..., T-1$ 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 $(V_{n1}^{**\prime}, \cdots, V_{n,T-1}^{**\prime})' = (I_{T-1} \otimes F'_{n,n-1})(V_{n1}^{*\prime}, \cdots, V_{n,T-1}^{*\prime})' = (I_{T-1} \otimes F'_{n,n-1})(F'_{T,T-1} \otimes I_n)(V'_{n1}, \cdots, V'_{n,T-1})' = (F'_{T,T-1} \otimes F'_{n,n-1})(V'_{n1}, \cdots, V'_{n,T-1})'$, we have

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

[24] 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 (23). In that case, the likelihood function 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.

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

The log likelihood function for (23) is

$$\begin{aligned}\ln L_{n,T}(\theta) &= -\frac{(n-1)(T-1)}{2}\ln(2\pi\sigma^2) + (T-1)\ln\left|I_{n-1} - \lambda F_{n,n-1}' W_n F_{n,n-1}\right| \\ &\quad + (T-1)\ln\left|I_{n-1} - \rho F_{n,n-1}' M_n F_{n,n-1}\right| - \frac{1}{2\sigma^2}\sum_{t=1}^{T-1} V_{nt}^{**\prime}(\zeta) V_{nt}^{**}(\zeta), \qquad (24)\end{aligned}$$

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

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

and, similarly, for $(I_{n-1} - \rho F_{n,n-1}' M_n F_{n,n-1})$. For any n-dimensional column vector p_{nt} and q_{nt}, as $J_n(p_{n1}, \cdots, p_{nT})J_T = J_n(\tilde{p}_{n1}, \cdots, \tilde{p}_{nT})$, it follows that

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

This implies that (24) is numerically identical to

$$\begin{aligned}\ln L_{n,T}(\theta) &= -\frac{(n-1)(T-1)}{2}\ln(2\pi\sigma^2) - (T-1)[\ln(1-\lambda) + \ln(1-\rho)] \\ &\quad + (T-1)[\ln|S_n(\lambda)| + \ln|R_n(\rho)|] - \frac{1}{2\sigma^2}\sum_{t=1}^{T} \tilde{V}_{nt}'(\zeta) J_n \tilde{V}_{nt}(\zeta). \qquad (25)\end{aligned}$$

We note that this likelihood function is, in general, not necessarily a conditional likelihood as the sample average over spatial units at each t might not be a sufficient statistic for the time dummy. This is because the cross section average $\frac{1}{n} l_n' W_n Y_{nt}$ might not equal to $c \cdot y_{\cdot t}$ for some scalar c, unless the column sums of W_n are all equal to unity. In practice, W_n is usually row-normalized but not column-normalized. On the other hand, the likelihood in (25) may be regarded as a modification of the concentrated likelihood in (17) 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 concentrated likelihood of the direct approach so that the modified likelihood can improve upon the concentrated likelihood function. There are various ways to construct a modified likelihood function from the concentrated likelihood of a direct approach in Cox and Reid (1987), Lancaster (2000), and Arellano and Hahn (2005), which involve approximations and related to bias corrected estimation. The Cox and Reid (1987) and Lancaster (2000) approach involves orthogonal parameterization, which is model specific. For our model, our 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[25].

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

3.2.2. Asymptotic Properties

The first and second order derivatives of (25) are (43) and (44), and the score is in (45) in Appendix E. With the identification conditions in Assumption 7', we can obtain the consistency of the transformation estimates.

Theorem 8 *Under Assumptions 1',2,3,4',5,6 and 7', θ_0 is identified and, for the QMLE $\hat{\theta}_{nT}$ based on (25), $\hat{\theta}_{nT} - \theta_0 \xrightarrow{p} 0$.*

Proof. *The steps of the proof are similar to that of Theorem 3.* ■

The estimates from the transformation approach will be consistent even when T is small[26]. This is different from the direct approach, where the consistency of the estimates requires both n and T to be large.

The variance matrix of $\frac{1}{\sqrt{(n-1)(T-1)}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial \theta}$ is equal to $\Sigma_{\theta_0,nT} + \Omega_{\theta_0,nT}$, where $\Sigma_{\theta_0,nT}$ and $\Omega_{\theta_0,nT}$ are, respectively, in (46) and (47). The asymptotics of the estimates from the transformation approach can be obtained similarly to that of Theorem 4.

Theorem 9 *Under Assumptions 1',2,3,4',5,6 and 7'(a); or 1',2,3,4',5,6,7'(b) and 8',*

$$\sqrt{(n-1)(T-1)}(\hat{\theta}_{nT} - \theta_0) \xrightarrow{d} N(0, \lim \Sigma^{-1}_{\theta_0,nT}(\Sigma_{\theta_0,nT} + \Omega_{\theta_0,nT})\Sigma^{-1}_{\theta_0,nT}). \tag{26}$$

Additionally, if v_{it}'s are normally distributed, $\sqrt{(n-1)(T-1)}(\hat{\theta}_{nT} - \theta_0) \xrightarrow{d} N(0, \lim \Sigma^{-1}_{\theta_0,nT})$.

Proof. *See Appendix E.3.* ■

Hence, after the data transformation to eliminate both the individual and time effects, the QMLE is consistent and asymptotically normal when either n or T are large, and it is properly centered. The direct approach needs both n and T to be large for consistency; when T is large, it requires n to be large enough to have a properly centered distribution. When both n and T are large, estimates of the two approaches will be consistent and have the same asymptotic variance matrix.

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, where we compare the performance of the transformation approach in Section 2.2 with the direct approach in Section 2.1. Then, we investigate the model (15) with both individual and time effects, where we compare the transformation approach in Section 3.2 with the direct approach in Section 3.1.

We first generate samples from (1):

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

[26] We have emphasized on the asymptotic when n goes to infinity but T can be finite or infinity (under Assumption 6). However, the result in Theorem 8 is valid even when n is finite but T tends to infinity.

using $\theta_0^a = (1.0, 0.2, 0.5, 1)'$ and $\theta_0^b = (1, 0.5, 0.2, 1)'$ where $\theta_0 = (\beta_0', \lambda_0, \rho_0, \sigma_0^2)'$, and X_{nt}, $\mathbf{c}_{n0}$ and V_{nt} are generated from independent standard normal distributions, and both the spatial weights matrices W_n and M_n are the same rook matrices[27]. We use $T = 5$, 10, 50, and $n = 9$, 16, 49. For each set of generated sample observations, we calculate the ML estimator $\hat{\theta}_{nT}$ and evaluate the bias $\hat{\theta}_{nT} - \theta_0$. We do this for 1000 times to get $\frac{1}{1000}\sum_{i=1}^{1000}(\hat{\theta}_{nT} - \theta_0)_i$. With two different values of θ_0 for each n and T, finite sample properties of both estimators are summarized in Table 1. For each case, we report the bias (Bias), empirical standard deviation (E-SD), root mean square error (RMSE) and theoretical standard deviation (T-SD)[28]. Both approaches provide the same estimate of $\zeta_0 = (\beta_0', \lambda_0, \rho_0)'$ while the estimator of σ_0^2 by the direct approach has a larger bias. The transformation approach yields a consistent estimator of σ_0^2 and the direct approach does not, which can be seen from the last two columns in Table 1 when T is small. The Biases, E-SDs, RMSEs and T-SDs for the estimators of $\zeta_0 = (\beta_0', \lambda_0, \rho_0)'$ are small when either n or T are large. 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. Also, when T is larger, the bias of the estimator of σ_0^2 by the direct approach decreases.

We then generate samples from (15):

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

using the same n, T, θ_0^a, θ_0^b, W_n and M_n. The X_{nt}, $\mathbf{c}_{n0}$, $\boldsymbol{\alpha}_{T0} = (\alpha_1, \alpha_2, \cdots, \alpha_T)$ 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 estimators using the direct approach in Section 3.1, and Table 3 is for those estimators after bias correction. Table 4 is for the estimators using the transformation approach in Section 3.2. We see that the bias of the transformation approach is small when either n or T are large. For the direct approach, the bias is small when n is large, and the bias is large when n is small and T might be large, while the bias for the estimate of σ_0^2 is small only when both n and T are large. By comparing Table 2 and Table 3, the bias correction reduces the biases of the direct approach estimates, without significant increase in the variance (S-TD).[29] This is consistent with the theoretical prediction.

Tables 1-4 here

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 only individual effects where the time periods T can be finite (or infinity) while the number of spatial units n is large. If T is finite, the direct ML estimation by estimating jointly all

[27] 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.

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

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

the parameters including the fixed effects will yield consistent estimators for the common parameters except the variance of disturbances. 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 transformation approach, which eliminates the individual fixed effects and can provide consistent estimates for all the common parameters including the variance of disturbances. 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 resulted 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 fixed effects. We show that the direct approach will yield consistent estimates when both n and T are large. However, asymptotic biases may still be presented. Bias correction procedures are useful to remove those asymptotic biases. 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 by proper transformations. 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), Baltagi et al. (2007) and Kapoor et al. (2007) consider spatial models with random effects, the SAR model in this paper considers a fixed effects specification. The proposed estimation methods are robust regardless of the 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 may be investigated in future research.

Appendix A. Notations and Some Lemmas

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

$S_n(\lambda) = I_n - \lambda W_n$ 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 $t = 1, 2, \cdots, T$ where $\bar{Y}_{nT} = \frac{1}{T}\sum_{t=1}^{T} Y_{nt}$.

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

$\theta = (\beta', \lambda, \rho, \sigma^2)'$ 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 $Q_{nT} = \sum_{t=1}^{T} \left(D_{nt}' V_{nt} + V_{nt}' B_n V_{nt} - \sigma_0^2 tr B_n\right)$ 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 (where there is no moving averages of past disturbances in $Q_{n,T}$) 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,

$$\begin{array}{lll} J_n F_{n,n-1} = F_{n,n-1}, & F_{n,n-1}' F_{n,n-1} = I_{n-1}, & J_n l_n = \mathbf{0}, \\ F_{n,n-1}' l_n = \mathbf{0}, & F_{n,n-1} F_{n,n-1}' + \frac{1}{n} l_n l_n' = I_n, & F_{n,n-1} F_{n,n-1}' = J_n. \end{array}$$

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

Proof: The derivation of these results can be found in Appendix A.2 of Lee and Yu (2007a). ■

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 for Sections 3, the corresponding Hessian matrix of $\ln L_{n,T}(\theta)$ of the transformation approaches with $\theta_1 = \theta_0$, has the following properties:*

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

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

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

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 $i,j = 1,2,\cdots,k+4$.

Similarly, for $\ln L_{n,T}^d(\theta)$ of the direct approaches 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,*

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

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 $\mathbf{V}_{nT} = (V'_{n1}, V'_{n2}, \cdots, V'_{nT})'$. With $J_T = I_T - \frac{1}{T} l_T l'_T$, 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 for cross moments of quadratic forms,

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

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,

$$\begin{aligned} & E(\textstyle\sum_{t=1}^T \tilde{V}'_{nt} A_n \tilde{V}_{nt})(\sum_{t=1}^T \tilde{V}'_{nt} B_n \tilde{V}_{nt}) \\ = \; & (\mu_4 - 3\sigma_0^4)\dfrac{(T-1)^2}{T} vec'_D(A_n) vec_D(B_n) + \sigma_0^4[(T-1)^2 tr(A_n) tr(B_n) + (T-1) tr(A_n B_n^s). \end{aligned}$$

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,

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

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

$$cov[(\textstyle\sum_{t=1}^T \tilde{V}'_{nt} A_n \tilde{V}_{nt}), \sum_{t=1}^T D'_{nt} \tilde{V}_{nt}] = E\mathbf{V}'_{nT}(J_T \otimes A_n)\mathbf{V}_{nT}\tilde{\mathbf{D}}'_{nT}\mathbf{V}_{nT} = (1 - \dfrac{1}{T})\mu_3 \sum_{t=1}^T \sum_{i=1}^n a_{n,ii} \cdot \tilde{d}_{nt,i} = 0,$$

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 Approach in Section 2.1

B.1. The First and Second Order Derivatives

For the concentrated log likelihood function (4), the first and second order derivatives are

$$\frac{1}{\sqrt{nT}} \frac{\partial \ln L^d_{n,T}(\theta)}{\partial \theta} = \begin{pmatrix} \frac{1}{\sigma^2} \frac{1}{\sqrt{nT}} \sum_{t=1}^T (R_n(\rho)\tilde{X}_{nt})' \tilde{V}_{nt}(\zeta) \\ \frac{1}{\sigma^2} \frac{1}{\sqrt{nT}} \sum_{t=1}^T \Big((R_n(\rho) W_n \tilde{Y}_{nt})' \tilde{V}_{nt}(\zeta) - \sigma^2 tr G_n(\lambda) \Big) \\ \frac{1}{\sigma^2} \frac{1}{\sqrt{nT}} \sum_{t=1}^T \Big((H_n(\rho)\tilde{V}_{nt}(\zeta))' \tilde{V}_{nt}(\zeta) - \sigma^2 tr H_n(\rho) \Big) \\ \frac{1}{2\sigma^4} \frac{1}{\sqrt{nT}} \sum_{t=1}^T (\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta) - n\sigma^2) \end{pmatrix}, \tag{27}$$

$$
\begin{aligned}
&\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^d(\theta)}{\partial\theta\partial\theta'} \qquad (28)\\
= \; & -\frac{1}{nT}\begin{pmatrix}
\frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)\tilde{X}_{nt})'R_n(\rho)\tilde{X}_{nt} & * & * & * \\
\frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)W_n\tilde{Y}_{nt})'R_n(\rho)\tilde{X}_{nt} & \begin{array}{c}\frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)W_n\tilde{Y}_{nt})'R_n(\rho)W_n\tilde{Y}_{nt} \\ +Ttr(G_n^2(\lambda))\end{array} & * & * \\
\begin{array}{c}\frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)\tilde{X}_{nt})'H_n(\rho)\tilde{V}_{nt}(\zeta) \\ +\frac{1}{\sigma^2}\sum_{t=1}^T (M_n\tilde{X}_{nt})'\tilde{V}_{nt}(\zeta)\end{array} & \begin{array}{c}\frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)W_n\tilde{Y}_{nt})'_n H_n(\rho)\tilde{V}_{nt}(\zeta) \\ +\frac{1}{\sigma^2}\sum_{t=1}^T (M_nW_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\zeta)\end{array} & 0 & 0 \\
\frac{1}{\sigma^4}\sum_{t=1}^T \tilde{V}'_{nt}(\zeta)R_n(\rho)\tilde{X}_{nt} & \frac{1}{\sigma^4}\sum_{t=1}^T (R_n(\rho)W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\zeta) & 0 & 0
\end{pmatrix} \\
& -\frac{1}{nT}\begin{pmatrix}
\mathbf{0}_{k\times k} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times 1} \\
\mathbf{0}_{1\times k} & 0 & 0 & 0 \\
\mathbf{0}_{1\times k} & 0 & \begin{array}{c}\frac{1}{\sigma^2}\sum_{t=1}^T (H_n(\rho)\tilde{V}_{nt}(\zeta))'H_n(\rho)\tilde{V}_{nt}(\zeta) \\ +T\mathrm{tr}(H_n^2(\rho))\end{array} & * \\
\mathbf{0}_{1\times k} & 0 & \frac{1}{\sigma^4}\sum_{t=1}^T (H_n(\rho)\tilde{V}_{nt}(\zeta))'\tilde{V}_{nt}(\zeta) & -\frac{nT}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{t=1}^T (\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta))
\end{pmatrix}.
\end{aligned}
$$

The score of the log likelihood function evaluated at θ_T is

$$
\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^d(\theta_T)}{\partial\theta} = \begin{pmatrix}
\frac{1}{\sigma_T^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^T \ddot{X}'_{nt}\tilde{V}_{nt} \\
\frac{1}{\sigma_T^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^T (\ddot{G}_n\ddot{X}_{nt}\beta_0)'\tilde{V}_{nt} + \frac{1}{\sigma_T^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^T (\tilde{V}'_{nt}\ddot{G}'_n\tilde{V}_{nt} - \sigma_T^2 tr\ddot{G}_n) \\
\frac{1}{\sigma_T^2}\frac{1}{\sqrt{nT}}\sum_{t=1}^T (\tilde{V}'_{nt}H'_n\tilde{V}_{nt} - \sigma_T^2 trH_n) \\
\frac{1}{2\sigma_T^4}\frac{1}{\sqrt{nT}}\sum_{t=1}^T (\tilde{V}'_{nt}\tilde{V}_{nt} - n\sigma_T^2)
\end{pmatrix}. \qquad (29)
$$

From the second order condition in (28), we have $\Sigma_{\theta_T,nT}^d = -E\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^d(\theta_T)}{\partial\theta\partial\theta'} =$

$$
\begin{pmatrix}
\frac{1}{\sigma_T^2 nT}\sum_{t=1}^T \ddot{X}'_{nt}\ddot{X}_{nt} & * & * & * \\
\frac{1}{\sigma_T^2 nT}\sum_{t=1}^T (\ddot{G}_n\ddot{X}_{nt}\beta_0)'\ddot{X}_{nt} & \frac{1}{\sigma_T^2 nT}\sum_{t=1}^T (\ddot{G}_n\ddot{X}_{nt}\beta_0)'\ddot{G}_n\ddot{X}_{nt}\beta_0 + \frac{1}{n}tr\ddot{G}_n^s\ddot{G}_n & * & * \\
\mathbf{0}_{1\times k} & \frac{1}{n}tr(H_n^s\ddot{G}_n) & \frac{1}{n}tr(H_n^sH_n) & * \\
\mathbf{0}_{1\times k} & \frac{1}{\sigma_T^2 n}tr(\ddot{G}_n) & \frac{1}{\sigma_T^2 n}tr(H_n) & \frac{1}{2\sigma_T^4}
\end{pmatrix}. \qquad (30)
$$

B.2. Some Lemmas

The following lemmas provide some basic elements for consistency and asymptotic analysis of the direct estimator.

Lemma B.1 *Under Assumptions 1-6,* $\frac{1}{nT}\ln L_{n,T}^d(\theta) - Q_{n,T}^d(\theta) \xrightarrow{p} 0$ *uniformly in* $\theta \in \Theta$ *and* $Q_{n,T}^d(\theta)$ *is uniformly equicontinuous for* $\theta \in \Theta$*, where* $Q_{n,T}^d(\theta) = E\frac{1}{nT}\ln L_{n,T}^d(\theta)$.

Lemma B.2 *Under Assumption 7,* $\Sigma_{\theta_T,nT}^d$ *in (30) is nonsingular.*

Lemma B.3 *Under Assumptions 1–6 and 7(a); or 1–6,7(b) and 8,* $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^d(\theta_T)}{\partial\theta} \xrightarrow{d} N(0, \lim \frac{T}{T-1}(\Sigma_{\theta_T,nT}^d + \Omega_{\theta_T,n}^d))$ *where* $\Sigma_{\theta_T,nT}^d$ *and* $\Omega_{\theta_T,n}^d$ *are in (30) and (32). When* v_{it}*'s are normally distributed,*

$$
\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^d(\theta_T)}{\partial\theta} \xrightarrow{d} N(0, \lim \frac{T}{T-1}\Sigma_{\theta_T,nT}^d).
$$

Proof of Lemma B.1

Uniform convergence:

We have $Q_{n,T}^d(\theta) = E\frac{1}{nT}\ln L_{n,T}^d(\theta)$ and $\frac{1}{nT}\ln L_{n,T}^d(\theta) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 + \frac{1}{n}\ln|S_n(\lambda)| + \frac{1}{n}\ln|R_n(\rho)| - \frac{1}{2\sigma^2 nT}\sum_{t=1}^T \tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta)$. From the DGP, $\tilde{V}_{nt}(\zeta) = R_n(\rho)[S_n(\lambda)S_n^{-1}\tilde{X}_{nt}\beta_0 - \tilde{X}_{nt}\beta + S_n(\lambda)S_n^{-1}R_n^{-1}\tilde{V}_{nt}]$ and hence,

$$E\frac{1}{nT}\sum_{t=1}^{T}\tilde{V}_{nt}'(\zeta)\tilde{V}_{nt}(\zeta)=\frac{1}{nT}\sum_{t=1}^{T}(S_n(\lambda)S_n^{-1}\tilde{X}_{nt}\beta_0-\tilde{X}_{nt}\beta)'R_n'(\rho)R_n(\rho)(S_n(\lambda)S_n^{-1}\tilde{X}_{nt}\beta_0-\tilde{X}_{nt}\beta)$$
$$+\frac{1}{n}\sigma_T^2 tr(R_n^{-1\prime}S_n^{-1\prime}S_n'(\lambda)R_n'(\rho)R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}). \quad (31)$$

where $\sigma_T^2=\frac{T-1}{T}\sigma_0^2$. When T is large, from Lemma 15 in Yu et al. (2008), we have $\frac{1}{nT}\sum_{t=1}^{T}\tilde{V}_{nt}'B_n\tilde{V}_{nt}-E\frac{1}{nT}\sum_{t=1}^{T}\tilde{V}_{nt}'B_n\tilde{V}_{nt}\xrightarrow{p}0$ 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, as Θ is a bounded set, G_n and $R_n(\rho)$ are UB, and σ^2 is bounded away from zero in Θ,

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

uniformly in θ in Θ.

Uniform equicontinuity:

We have $Q_{nT}^d(\theta)=-\frac{1}{2}\ln 2\pi-\frac{1}{2}\ln\sigma^2+\frac{1}{n}\ln|S_n(\lambda)|+\frac{1}{n}\ln|R_n(\rho)|-\frac{1}{2\sigma^2 nT}E\sum_{t=1}^{T}\tilde{V}_{nt}'(\zeta)\tilde{V}_{nt}(\zeta)$. For $\frac{1}{nT}E\sum_{t=1}^{T}\tilde{V}_{nt}'(\zeta)\tilde{V}_{nt}(\zeta)$ in (31), the first term is equal to $(\beta'-\beta_0',\lambda-\lambda_0)\mathcal{H}_{nT}^d(\rho)(\beta'-\beta_0',\lambda-\lambda_0)'$ where $\mathcal{H}_{nT}^d(\rho)=\frac{1}{nT}\sum_{t=1}^{T}(\tilde{X}_{nt},G_n\tilde{X}_{nt}\beta_0)'R_n'(\rho)R_n(\rho)(\tilde{X}_{nt},G_n\tilde{X}_{nt}\beta_0)$, by using $S_n(\lambda)S_n^{-1}=I_n-(\lambda-\lambda_0)G_n$; the second term is equal to $\sigma_{nT}^2(\lambda,\rho)$ where $\sigma_{nT}^2(\lambda,\rho)=\frac{\sigma_T^2}{n}tr[(R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1})'(R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1})]$. There terms are all polynomial functions of θ. To prove $Q_{n,T}^d(\theta)$ is uniformly equicontinuous in θ, the following are sufficient: (i) $\ln\sigma^2$ is uniformly continuous; (ii) $\frac{1}{n}\ln|S_n(\lambda)|$ and $\frac{1}{n}\ln|R_n(\rho)|$ are uniformly equicontinuous; (iii) $(\beta'-\beta_0',\lambda-\lambda_0)\mathcal{H}_{nT}^d(\rho)(\beta'-\beta_0',\lambda-\lambda_0)'$ is uniformly equicontinuous; (iv) $\sigma_{nT}^2(\lambda,\rho)$ is uniformly equicontinuous.

The (i) is obvious because σ^2 is bounded away from zero in Θ. For (ii), $\frac{1}{n}\ln|S_n(\lambda_2)|-\frac{1}{n}\ln|S_n(\lambda_1)|=\frac{1}{n}tr\left(W_nS_n^{-1}\left(\bar{\lambda}\right)\right)(\lambda_2-\lambda_1)$ where $\bar{\lambda}$ lies between λ_2 and λ_1. As $S_n^{-1}(\lambda)$ is UB, uniformly in $\theta\in\Theta$, $\frac{1}{n}tr\left(W_nS_n^{-1}\left(\bar{\lambda}\right)\right)$ is bounded, and, hence, $\frac{1}{n}\ln|S_n(\lambda)|$ is uniformly equicontinuous. Similarly, $\frac{1}{n}\ln|R_n(\rho)|$ is uniformly equicontinuous. For (iii) and (iv), the functions are simply polynomial functions of β, λ and ρ. The uniform equicontinuity of those functions on any bounded set are apparent. ■

Proof of Lemma B.2

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. Denote $C_n=G_n-\frac{trG_n}{n}I_n$ and $D_n=H_n-\frac{trH_n}{n}I_n$ so that $\frac{1}{n}tr(\ddot{G}_n^s\ddot{G}_n)-2\left(\frac{tr\ddot{G}_n}{n}\right)^2=\frac{1}{2n}tr(C_n^sC_n^s)$, $\frac{1}{n}tr(H_n^sH_n)-2\left(\frac{trH_n}{n}\right)^2=\frac{1}{2n}tr(D_n^sD_n^s)$ 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 $\mathcal{H}_{\beta,nT}=\frac{1}{nT}\sum_{t=1}^{T}\ddot{X}_{nt}'\ddot{X}_{nt}$, $\mathcal{H}_{\beta\lambda,nT}=\frac{1}{nT}\sum_{t=1}^{T}\ddot{X}_{nt}'\ddot{G}_n\ddot{X}_{nt}\beta_0$,

[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 $\mathbf{V}_{nT}=(V_{n1}',\cdots,V_{nT})'$ 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}.

$\mathcal{H}_{\lambda\beta,nT} = \mathcal{H}'_{\beta\lambda,nT}$ and $\mathcal{H}_{\lambda,nT} = \frac{1}{nT}\sum_{t=1}^T(\ddot{G}_n\tilde{X}_{nt}\beta_0)'\ddot{G}_n\tilde{X}_{nt}\beta_0$. By the method of substitution and elimination, $\lim \Sigma^d_{\theta_T,nT}c = 0$ 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 $\lim \Sigma^d_{\theta_T,nT}$ follows from Assumption 7. ■

Proof of Lemma B.3

For (29), 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 $E(\frac{1}{\sqrt{nT}}\frac{\partial \ln L^d_{n,T}(\theta_T)}{\partial\theta} \cdot \frac{1}{\sqrt{nT}}\frac{\partial \ln L^d_{n,T}(\theta_T)}{\partial\theta'}) = \frac{T}{T-1}(\Sigma^d_{\theta_T,nT} + \Omega^d_{\theta_T,n})$ where $\Sigma^d_{\theta_T,nT}$ is in (30) and

$$\Omega^d_{\theta_T,nT} = \frac{(T-1)}{T}\frac{(\mu_4 - 3\sigma_0^4)}{\sigma_0^4}\begin{pmatrix} \mathbf{0}_{k\times k} & * & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n}\sum_{i=1}^n \ddot{G}^2_{n,ii} & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n}\sum_{i=1}^n \ddot{G}_{n,ii}H_{n,ii} & \frac{1}{n}\sum_{i=1}^n H^2_{n,ii} & * \\ \mathbf{0}_{1\times k} & \frac{1}{2\sigma_T^2 n}tr\ddot{G}_n & \frac{1}{2\sigma_T^2 n}trH_n & \frac{1}{4\sigma_T^4}\end{pmatrix}. \tag{32}$$

When V_{nt} are normally distributed, $\Omega^d_{\theta_T,nT} = \mathbf{0}_{(k+3)\times(k+3)}$ because $\mu_4 - 3\sigma_0^4 = 0$. By using the central limit theorem in Lemma A.1, (29) will be asymptotically normally distributed. ■

B.3. Proof for Theorem 1

Without loss of generality, we are presenting 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 with a slight modification of assumptions, because the arguments below follow the consistency framework in White (1994) which allows a sequence of true parameters to depend on sample sizes.

As $E\frac{1}{nT}\sum_{t=1}^T\tilde{V}'_{nt}\tilde{V}_{nt} = \frac{T-1}{T}\sigma_0^2 = \sigma_T^2$, at θ_T, we have $\frac{1}{nT}E\ln L^d_{n,T}(\theta_T) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma_T^2 + \frac{1}{n}[\ln|S_n| + \ln|R_n|] - \frac{1}{2}$. As $\tilde{Y}_{nt} = S_n^{-1}(\tilde{X}_{nt}\beta_0 + R_n^{-1}\tilde{V}_{nt})$ and $S_n(\lambda)S_n^{-1} = I_n + (\lambda_0 - \lambda)G_n$, we have

$$\tilde{V}_{nt}(\zeta) = R_n(\rho)[S_n(\lambda)S_n^{-1}R_n^{-1}\tilde{V}_{nt} + (\lambda_0 - \lambda)G_n\tilde{X}_{nt}\beta_0 + \tilde{X}_{nt}(\beta_0 - \beta)].$$

Denote

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

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

It follows that

$\frac{1}{nT}E\ln L^d_{n,T}(\theta) - \frac{1}{nT}E\ln L^d_{n,T}(\theta_T)$
$= -\frac{1}{2}(\ln\sigma^2 - \ln\sigma_T^2) + \frac{1}{n}\ln|S_n(\lambda)| - \frac{1}{n}\ln|S_n| + \frac{1}{n}\ln|R_n(\rho)| - \frac{1}{n}\ln|R_n| - \left(\frac{1}{2\sigma^2}\frac{1}{nT}\sum_{t=1}^T E\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta) - \frac{1}{2}\right)$
$= T_{1,n}(\lambda,\rho,\sigma^2) - \frac{1}{2\sigma^2}T_{2,n,T}(\beta,\lambda,\rho)$

where, using $\frac{1}{nT}\sum_{t=1}^{T} E\tilde{V}_{nt}'(\zeta)\tilde{V}_{nt}(\zeta)$ in (31),

$$\begin{aligned} T_{1,n}(\lambda,\rho,\sigma^2) &= -\frac{1}{2}(\ln\sigma^2 - \ln\sigma_T^2) + \frac{1}{n}[\ln|S_n(\lambda)| - \ln|S_n| + \ln|R_n(\rho)| - \ln|R_n|] - \frac{1}{2\sigma^2}(\sigma_{nT}^2(\lambda,\rho) - \sigma^2), \\ T_{2,n,T}(\beta,\lambda,\rho) &= \frac{1}{nT}\sum_{t=1}^{T}\left\{ \begin{array}{c} (\tilde{X}_{nt}(\beta_0-\beta) + (\lambda_0-\lambda)G_n\tilde{X}_{nt}\beta_0)'R_n'(\rho) \\ \times R_n(\rho)(\tilde{X}_{nt}(\beta_0-\beta) + (\lambda_0-\lambda)G_n\tilde{X}_{nt}\beta_0) \end{array}\right\}. \end{aligned}$$

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, where the variance of the disturbances is σ_T^2. Using the information inequality from this pure spatial process, $T_{1,n}(\lambda,\rho,\sigma^2) \leq 0$ for any (λ,ρ,σ^2). Also, $T_{2,n,T}(\beta,\lambda,\rho)$ is a quadratic function of β and λ with a positive semidefinite matrix given ρ.

Under the condition in Assumption 7 (a) that $\lim_{n\to\infty}\mathcal{H}_{nT}^d(\rho)$ is nonsingular[31], $T_{2,n,T}(\beta,\lambda,\rho) > 0$ given any ρ whenever $(\beta,\lambda) \neq (\beta_0,\lambda_0)$. Hence, (β,λ) is identified. Given λ_0, ρ_0 and σ_T^2 are the unique maximizer of $\lim_{n\to\infty} T_{1,n}(\lambda,\rho,\sigma^2)$ under the condition that $\lim_{n\to\infty}(\frac{1}{n}\ln\left|\sigma_T^2 R_n^{-1\prime}R_n^{-1}\right| - \frac{1}{n}\ln\left|\sigma_{nT}^2(\rho)R_n^{-1}(\rho)'R_n^{-1}(\rho)\right|)$ is not zero for $\rho \neq \rho_0$.[32] Hence, $(\beta,\lambda,\rho,\sigma^2)$ is identified.

When $\lim_{n\to\infty}\mathcal{H}_{nT}^d(\rho)$ is singular, β_0 and λ_0 cannot be identified from $T_{2,n,T}(\beta,\lambda,\rho)$. Identification requires that $\lim_{n\to\infty} T_{1,n}(\lambda,\rho,\sigma^2)$ is strictly less than zero. As $T_{1,n}(\lambda,\rho,\sigma^2) \leq 0$ by the information inequality for the pure spatial process, $\lim_{n\to\infty} T_{1,n}(\lambda,\rho,\sigma^2) \neq 0$ is equivalent to $\lim_{n\to\infty}(\frac{1}{n}\ln\left|\sigma_T^2 R_n^{-1\prime}S_n^{-1\prime}S_n^{-1}R_n^{-1}\right| - \frac{1}{n}\ln\left|\sigma_{nT}^2(\lambda,\rho)R_n^{-1}(\rho)'S_n^{-1}(\lambda)'S_n^{-1}(\lambda)R_n^{-1}(\rho)\right|) \neq 0$ for any $(\lambda,\rho,\sigma^2) \neq (\lambda_0,\rho_0,\sigma_T^2)$ (this is similar to Lee (2004), Proof of Theorem 4.1). Given λ_0 is identified, β_0 can be identified from $\lim_{n\to\infty} T_{2,n,T}(\beta,\lambda,\rho)$.

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

B.4. Proof for Theorem 2

According to the Taylor expansion,

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

where $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^d(\theta_T)}{\partial\theta} \xrightarrow{d} N(0, \lim\frac{T}{T-1}(\Sigma_{\theta_T,nT}^d + \Omega_{\theta_T,nT}^d))$ and $\bar{\theta}_{nT}^d$ lies between θ_T and $\hat{\theta}_{nT}^d$. As we have $-\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^d(\bar{\theta}_{nT}^d)}{\partial\theta\partial\theta'} = \left(-\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^d(\bar{\theta}_{nT}^d)}{\partial\theta\partial\theta'} - \left(-\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^d(\theta_T)}{\partial\theta\partial\theta'}\right)\right) + \left(-\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^d(\theta_T)}{\partial\theta\partial\theta'} - \Sigma_{\theta_T,nT}^d\right) + \Sigma_{\theta_T,nT}^d$ where the first term is $\left\|\bar{\theta}_{nT}^d - \theta_T\right\| \cdot O_p(1)$ and the second term is $O_p\left(\frac{1}{\sqrt{nT}}\right)$ (see Lemma A.3), $-\frac{1}{nT}\frac{\partial^2 \ln L_{n,T}^d(\bar{\theta}_{nT}^d)}{\partial\theta\partial\theta'} = \left\|\bar{\theta}_{nT}^d - \theta_T\right\| \cdot O_p(1) + O_p\left(\frac{1}{\sqrt{nT}}\right) + \Sigma_{\theta_T,nT}^d$. 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, $\sqrt{nT}(\hat{\theta}_{nT}^d - \theta_T) = \left(\Sigma_{\theta_T,nT}^d + O_p\left(\frac{1}{\sqrt{nT}}\right)\right)^{-1} \cdot \left(\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{n,T}^d(\theta_T)}{\partial\theta}\right)$. 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 $\sqrt{nT}(\hat{\theta}_{nT}^d - \theta_T) \xrightarrow{d} N(0, \lim\frac{T}{T-1}(\Sigma_{\theta_T,nT}^d)^{-1}(\Sigma_{\theta_T,nT}^d + \Omega_{\theta_T,nT}^d)(\Sigma_{\theta_T,nT}^d)^{-1})$. ■

Appendix C. The Transformation Approach in Section 2.2

[31] We note that $\mathcal{H}_{nT}^d(\rho) = \frac{T-1}{T}\mathcal{H}_{nT}(\rho)$.

[32] This is equivalent to the identification of a pure SAR model in Lee (2004) as the common coefficient $\frac{T-1}{T}$ in both σ_T^2 and $\sigma_{nT}^2(\rho)$ will be cancelled out.

C.1. The First and Second Order Derivatives

For the first and second order derivatives of (10),

$$\frac{\partial \ln L_{n,T}(\theta)}{\partial \theta} = \begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)\tilde{X}_{nt})'\tilde{V}_{nt}(\zeta) \\ \frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\zeta) - (T-1)trG_n(\lambda) \\ \frac{1}{\sigma^2}\sum_{t=1}^T (H_n(\rho)\tilde{V}_{nt}(\zeta))'\tilde{V}_{nt}(\zeta) - (T-1)trH_n(\rho) \\ \frac{1}{2\sigma^4}\sum_{t=1}^T (\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta) - n\frac{T-1}{T}\sigma^2) \end{pmatrix}, \tag{33}$$

and

$$-\frac{\partial^2 \ln L_{n,T}(\theta)}{\partial \theta \partial \theta'}$$

$$= \begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)\tilde{X}_{nt})'R_n(\rho)\tilde{X}_{nt} & * & * & * \\ \frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)W_n\tilde{Y}_{nt})'R_n(\rho)\tilde{X}_{nt} & \begin{matrix}\frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)W_n\tilde{Y}_{nt})'R_n(\rho)W_n\tilde{Y}_{nt} \\ +(T-1)tr(G_n^2(\lambda))\end{matrix} & * & * \\ \begin{matrix}\frac{1}{\sigma^2}\sum_{t=1}^T (H_n(\rho)\tilde{V}_{nt}(\zeta))'R_n(\rho)\tilde{X}_{nt}) \\ +\frac{1}{\sigma^2}\sum_{t=1}^T \tilde{V}'_{nt}(\zeta)M_n\tilde{X}_{nt}\end{matrix} & \begin{matrix}\frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)W_n\tilde{Y}_{nt})'_n H_n(\rho)\tilde{V}_{nt}(\zeta) \\ +\frac{1}{\sigma^2}\sum_{t=1}^T (M_nW_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\zeta)\end{matrix} & 0 & 0 \\ \frac{1}{\sigma^4}\sum_{t=1}^T \tilde{V}'_{nt}(\zeta)R_n(\rho)\tilde{X}_{nt} & \frac{1}{\sigma^4}\sum_{t=1}^T (R_n(\rho)W_n\tilde{Y}_{nt})'\tilde{V}_{nt}(\zeta) & 0 & 0 \end{pmatrix} \tag{34}$$

$$+\begin{pmatrix} \mathbf{0}_{k\times k} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times 1} \\ \mathbf{0}_{1\times k} & 0 & 0 & 0 \\ \mathbf{0}_{1\times k} & 0 & \begin{matrix}\frac{1}{\sigma^2}\sum_{t=1}^T (H_n(\rho)\tilde{V}_{nt}(\zeta))'H_n(\rho)\tilde{V}_{nt}(\zeta) \\ +(T-1)tr(H_n^2(\rho))\end{matrix} & * \\ \mathbf{0}_{1\times k} & 0 & \frac{1}{\sigma^4}\sum_{t=1}^T (H_n(\rho)\tilde{V}_{nt}(\zeta))'\tilde{V}_{nt}(\zeta) & -\frac{n(T-1)}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{t=1}^T (\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta)) \end{pmatrix}.$$

At true θ_0, we have

$$\frac{1}{\sqrt{n(T-1)}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial \theta} = \begin{pmatrix} \frac{1}{\sigma_0^2}\frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^T \ddot{X}'_{nt}\tilde{V}_{nt} \\ \frac{1}{\sigma_0^2}\frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^T (\ddot{G}_n\ddot{X}_{nt}\beta_0)'\tilde{V}_{nt} + \frac{1}{\sigma_0^2}\frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^T (\tilde{V}'_{nt}\ddot{G}'_n\tilde{V}_{nt} - \frac{T-1}{T}\sigma_0^2 tr\ddot{G}_n) \\ \frac{1}{\sigma_0^2}\frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^T (\tilde{V}'_{nt}H'_n\tilde{V}_{nt} - \frac{T-1}{T}\sigma_0^2 trH_n) \\ \frac{1}{2\sigma_0^4}\frac{1}{\sqrt{n(T-1)}}\sum_{t=1}^T (\tilde{V}'_{nt}\tilde{V}_{nt} - n\frac{T-1}{T}\sigma_0^2) \end{pmatrix}, \tag{35}$$

and the information matrix $\Sigma_{\theta_0,nT} = -E\left(\frac{1}{n(T-1)}\frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'}\right) =$

$$\frac{1}{\sigma_0^2}\begin{pmatrix} \mathcal{H}_{nT} & * & * \\ \mathbf{0}_{1\times(k+1)} & 0 & * \\ \mathbf{0}_{1\times(k+1)} & 0 & 0 \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{k\times k} & * & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n}tr\ddot{G}_n^s\ddot{G}_n) & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n}tr(H_n^s\ddot{G}_n) & \frac{1}{n}tr(H_n^sH_n) & * \\ \mathbf{0}_{1\times k} & \frac{1}{\sigma_0^2 n}tr(\ddot{G}_n) & \frac{1}{\sigma_0^2 n}tr(H_n) & \frac{1}{2\sigma_0^4} \end{pmatrix}, \tag{36}$$

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

C.2. Some Lemmas

The following basic properties of the likelihood function, its score, and information matrix are for the transformation approach. As the proofs of these lemmas are similar to those in Appendix B.2, they are omitted.

Lemma C.1 *Under Assumptions 1-6,* $\frac{1}{n(T-1)}\ln L_{n,T}(\theta) - Q_{n,T}(\theta) \xrightarrow{p} 0$ *uniformly in* $\theta \in \Theta$ *and* $Q_{n,T}(\theta)$ *is uniformly equicontinuous for* $\theta \in \Theta$*, where* $Q_{n,T}(\theta) = E\frac{1}{n(T-1)}\ln L_{n,T}(\theta)$.

Lemma C.2 *Under Assumption 7, $\Sigma_{\theta_0,nT}$ in (36) is nonsingular.*

Lemma C.3 *Under Assumptions 1-6 and 7(a); or 1-6, 7(b) and 8, $\frac{1}{\sqrt{n(T-1)}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta} \xrightarrow{d} N(0, \lim(\Sigma_{\theta_0,nT}+\Omega_{\theta_0,nT}))$ where $\Sigma_{\theta_0,nT}$ is in (36)*

$$\Omega_{\theta_0,nT} = \frac{(T-1)}{T}\frac{(\mu_4 - 3\sigma_0^4)}{\sigma_0^4}\begin{pmatrix} \mathbf{0}_{k\times k} & * & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n}\sum_{i=1}^n \ddot{G}_{n,ii}^2 & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n}\sum_{i=1}^n \ddot{G}_{n,ii}H_{n,ii} & \frac{1}{n}\sum_{i=1}^n H_{n,ii}^2 & * \\ \mathbf{0}_{1\times k} & \frac{1}{2\sigma_0^2 n}tr\ddot{G}_n & \frac{1}{2\sigma_0^2 n}trH_n & \frac{1}{4\sigma_0^4} \end{pmatrix}. \tag{37}$$

C.3. Proof of Theorem 3

Compared with the direct approach in Appendix B, the transformation approach has a degree of freedom adjustment (from nT to $n(T-1)$) for the log likelihood function. For the direct approach, the maximum of the limiting average log likelihood function is at $(\lambda_0, \rho_0, \beta_0,' \sigma_T^2)$ with a finite T while it is $(\lambda_0, \rho_0, \beta_0,' \sigma_0^2)$ for that of the transformation approach. Taking into these differences, the proof arguments are similar. ■

C.4. Proof of Theorem 4

The proof is similar to the direct approach, where $\sqrt{nT}(\hat{\theta}_{nT}^d - \theta_T)$ is considered. Due to a degree of freedom adjustment (from nT to $n(T-1)$) for the log likelihood function in the transformation approach, 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 D. The Direct Approach in Section 3.1

D.1. The First and Second Order Derivatives of (18)

The first and second order derivatives of the concentrated log likelihood in (18) are

$$\frac{\partial \ln L_{n,T}^d(\theta)}{\partial\theta} = \begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)\tilde{X}_{nt})'J_n\tilde{V}_{nt}(\zeta) \\ \frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)W_n\tilde{Y}_{nt})'J_n\tilde{V}_{nt}(\zeta) - T\mathrm{tr}G_n(\lambda) \\ \frac{1}{\sigma^2}\sum_{t=1}^T (H_n(\rho)\tilde{V}_{nt}(\zeta))'J_n\tilde{V}_{nt}(\zeta) - T\mathrm{tr}H_n(\rho) \\ \frac{1}{2\sigma^4}\sum_{t=1}^T (\tilde{V}_{nt}'(\zeta)J_n\tilde{V}_{nt}(\zeta) - n\sigma^2) \end{pmatrix}, \tag{38}$$

and

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

$$= \begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)\tilde{X}_{nt})'J_nR_n(\rho)\tilde{X}_{nt} & * & * & * \\ \frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)W_n\tilde{Y}_{nt})'J_nR_n(\rho)\tilde{X}_{nt} & \frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)W_n\tilde{Y}_{nt})'J_nR_n(\rho)W_n\tilde{Y}_{nt} + T\mathrm{tr}(G_n^2(\lambda)) & * & * \\ \frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)\tilde{X}_{nt})'J_nH_n(\rho)\tilde{V}_{nt}(\zeta) + \frac{1}{\sigma^2}\sum_{t=1}^T (M_n\tilde{X}_{nt})'J_n\tilde{V}_{nt}(\zeta) & \frac{1}{\sigma^2}\sum_{t=1}^T (R_n(\rho)W_n\tilde{Y}_{nt})_n'J_nH_n(\rho)\tilde{V}_{nt}(\zeta) + \frac{1}{\sigma^2}\sum_{t=1}^T (M_nW_n\tilde{Y}_{nt})'J_n\tilde{V}_{nt}(\zeta) & 0 & 0 \\ \frac{1}{\sigma^4}\sum_{t=1}^T \tilde{V}_{nt}'(\zeta)J_nR_n(\rho)\tilde{X}_{nt} & \frac{1}{\sigma^4}\sum_{t=1}^T (R_n(\rho)W_n\tilde{Y}_{nt})'J_n\tilde{V}_{nt}(\zeta) & 0 & 0 \end{pmatrix} \tag{39}$$

$$+ \begin{pmatrix} \mathbf{0}_{k\times k} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times 1} \\ \mathbf{0}_{1\times k} & 0 & 0 & 0 \\ \mathbf{0}_{1\times k} & 0 & \frac{1}{\sigma^2}\sum_{t=1}^T (H_n(\rho)\tilde{V}_{nt}(\zeta))'J_nH_n(\rho)\tilde{V}_{nt}(\zeta) + T\mathrm{tr}(H_n^2(\rho)) & * \\ \mathbf{0}_{1\times k} & 0 & \frac{1}{\sigma^4}\sum_{t=1}^T (H_n(\rho)\tilde{V}_{nt}(\zeta))'J_n\tilde{V}_{nt}(\zeta) & -\frac{nT}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{t=1}^T (\tilde{V}_{nt}'(\zeta)J_n\tilde{V}_{nt}(\zeta)) \end{pmatrix}.$$

For the first order derivative evaluated at θ_T, it has two components such that

$$\frac{\partial \ln L^d_{n,T}(\theta_T)}{\partial \theta} = \frac{\partial \ln L^{d,u}_{n,T}(\theta_T)}{\partial \theta} - T \cdot a_{\theta_T,n} \tag{40}$$

where

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

and

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

For the second order derivative evaluated at θ_T, we have $\Sigma^d_{\theta_T,nT} = -E\frac{1}{nT}\frac{\partial \ln L^{2d}_{n,T}(\theta_T)}{\partial \theta \partial \theta'} =$

$$\begin{pmatrix} \frac{1}{\sigma_T^2 nT}\sum_{t=1}^T \ddot{X}'_{nt} J_n \ddot{X}_{nt} & * & * & * \\ \frac{1}{\sigma_T^2 nT}\sum_{t=1}^T (\ddot{G}_n \ddot{X}_{nt}\beta_0)' J_n \ddot{X}_{nt} & \frac{1}{\sigma_T^2 nT}\sum_{t=1}^T (\ddot{G}_n \ddot{X}_{nt}\beta_0)' J_n \ddot{G}_n \ddot{X}_{nt}\beta_0 + \frac{1}{n} tr \ddot{G}^s_n J_n \ddot{G}_n & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n} tr(H^s_n J_n \ddot{G}_n) & \frac{1}{n} tr(H^s_n H_n) & * \\ \mathbf{0}_{1\times k} & \frac{1}{\sigma_T^2 n} tr(\ddot{G}_n) & \frac{1}{\sigma_T^2 n} tr(H_n) & \frac{1}{2\sigma_T^4} \end{pmatrix}. \tag{41}$$

D.2. Some Lemmas

The following lemmas are for the direct approach of the model with both individual and time fixed effects. The proofs are similar to those for the model with only individual effects, and are omitted. Here, we note that there may be an asymptotic bias in the score in Lemma D.3 due to the additional time dummies.

Lemma D.1 *Under Assumptions 1-3,4',5,6 and 7',* $\frac{1}{nT}\ln L^d_{n,T}(\theta) - Q^d_{n,T}(\theta) \xrightarrow{p} 0$ *uniformly in* $\theta \in \Theta$ *and* $Q^d_{n,T}(\theta)$ *is uniformly equicontinuous for* $\theta \in \Theta$*, where* $Q^d_{n,T}(\theta) = E\frac{1}{nT}\ln L^d_{n,T}(\theta)$.

Lemma D.2 *Under Assumption 7',* $\Sigma^d_{\theta_T,nT}$ *in (41) is nonsingular.*

Lemma D.3 *Under Assumptions 1-3,4',5,6 and 7'(a); or 1-3,4',5,6,7'(b) and 8',* $\frac{1}{\sqrt{nT}}\frac{\partial \ln L^d_{n,T}(\theta_T)}{\partial \theta} + \sqrt{\frac{T}{n}} a_{\theta_T,n} \xrightarrow{d} N(0, \lim \frac{T}{T-1}(\Sigma^d_{\theta_T,nT} + \Omega^d_{\theta_T,nT}))$ *where* $\Sigma^d_{\theta_T,nT}$ *is in (41) and*

$$\Omega^d_{\theta_T,nT} = \frac{(T-1)}{T}\frac{(\mu_4 - 3\sigma_0^4)}{\sigma_0^4}\begin{pmatrix} \mathbf{0}_{k\times k} & * & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n}\sum_{i=1}^n [(J_n \ddot{G}_n)_{ii}]^2 & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n}\sum_{i=1}^n [J_n \ddot{G}_n]_{ii}[J_n H_n]_{ii} & \frac{1}{n}\sum_{i=1}^n [(J_n H_n)_{ii}]^2 & * \\ \mathbf{0}_{1\times k} & \frac{1}{2\sigma_T^2 n} tr J_n \ddot{G}_n & \frac{1}{2\sigma_T^2 n} tr J_n H_n & \frac{1}{4\sigma_T^4} \end{pmatrix}. \tag{42}$$

D.3. Proof for Theorem 6

According to the Taylor expansion,

$$\sqrt{nT}(\hat{\theta}^d_{nT} - \theta_T) = \left(-\frac{1}{nT}\frac{\partial^2 \ln L^d_{n,T}(\bar{\theta}^d_{nT})}{\partial \theta \partial \theta'}\right)^{-1} \cdot \left(\frac{1}{\sqrt{nT}}\frac{\partial \ln L^d_{n,T}(\theta_T)}{\partial \theta}\right)$$

where $\frac{1}{\sqrt{nT}}\frac{\partial \ln L^d_{n,T}(\theta_T)}{\partial\theta} + \sqrt{\frac{T}{n}}a_{\theta_T,n} = \frac{1}{\sqrt{nT}}\frac{\partial \ln L^{d(u)}_{n,T}(\theta_T)}{\partial\theta} \xrightarrow{d} N(0, \lim \frac{T}{T-1}(\Sigma^d_{\theta_T,nT} + \Omega^d_{\theta_T,nT}))$ by Lemma D.3 and $\bar{\theta}^d_{nT}$ lies between θ_T and $\hat{\theta}^d_{nT}$. Because $\left\|\bar{\theta}^d_{nT} - \theta_T\right\| = o_p(1)$ and $\Sigma^d_{\theta_T,nT}$ is nonsingular in the limit, $\left(-\frac{1}{nT}\frac{\partial^2 \ln L^d_{n,T}(\bar{\theta}^d_{nT})}{\partial\theta\partial\theta'}\right)^{-1}$ is $O_p(1)$. Hence, $\sqrt{nT}(\hat{\theta}^d_{nT} - \theta_T) = O_p(1)(O_p(1) + O(\sqrt{\frac{T}{n}}))$, which implies that $\hat{\theta}^d_{nT} - \theta_T = O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{n}\right)\right)$. In turn, $\left(-\frac{1}{nT}\frac{\partial^2 \ln L^d_{n,T}(\bar{\theta}^d_{nT})}{\partial\theta\partial\theta'}\right)^{-1} = (\Sigma^d_{\theta_T,nT})^{-1} + O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{n}\right)\right)$. It follows that

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

and, hence,

$$\begin{aligned}&\sqrt{nT}(\hat{\theta}^d_{nT} - \theta_T) + (\Sigma^d_{\theta_T,nT})^{-1}\cdot\sqrt{\frac{T}{n}}a_{\theta_T,n} + O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{n}\right)\right)\sqrt{\frac{T}{n}}a_{\theta_T,n} \\ = &\ ((\Sigma^d_{\theta_T,nT})^{-1} + o_p(1))\cdot\frac{1}{\sqrt{nT}}\frac{\partial \ln L^{d(u)}_{n,T}(\theta_T)}{\partial\theta}.\end{aligned}$$

Using Lemma D.3, we have the result (20) in Theorem 6. ■

D.4. Proof for Theorem 7

We have $\sqrt{nT}(\hat{\theta}^d_{nT} - \theta_T) + \sqrt{\frac{T}{n}}(\Sigma^d_{\theta_T,nT})^{-1}a_{\theta_T,n} + O_p\left(\sqrt{\frac{T}{n^3}}\right) \xrightarrow{d} N(0, \lim \frac{T}{T-1}(\Sigma^d_{\theta_T,nT})^{-1}(\Sigma^d_{\theta_T,nT} + \Omega^d_{\theta_T,nT})(\Sigma^d_{\theta_T,nT})^{-1})$ from Theorem 6. As the first step bias corrected estimator is $\hat{\theta}^{d1}_{nT} = \hat{\theta}^d_{nT} + \frac{1}{n}(\Sigma^d_{\hat{\theta}^d_{nT},nT})^{-1}a_n(\hat{\theta}^d_{nT})$ where $a_n(\theta) = a_{\theta,n}$, we will have $\sqrt{nT}(\hat{\theta}^{d1}_{nT} - \theta_T) \xrightarrow{d} N(0, \lim \frac{T}{T-1}(\Sigma^d_{\theta_T,nT})^{-1}(\Sigma^d_{\theta_T,nT} + \Omega^d_{\theta_T,nT})(\Sigma^d_{\theta_T,nT})^{-1})$ if $\frac{T}{n^3} \to 0$ and $\sqrt{\frac{T}{n}}\left(\left(-\frac{1}{nT}E\frac{\partial^2 \ln L^d_{nT}(\hat{\theta}^d_{nT})}{\partial\theta\partial\theta'}\right)^{-1}a_n(\hat{\theta}^d_{nT}) - (\Sigma^d_{\theta_T,nT})^{-1}a_n(\theta_T)\right) \xrightarrow{p} 0$, where $-\frac{1}{nT}E\frac{\partial^2 \ln L^d_{nT}(\hat{\theta}^d_{nT})}{\partial\theta\partial\theta'} = \Sigma^d_{\hat{\theta}^d_{nT},nT}$ is the information matrix evaluated at $\hat{\theta}^d_{nT}$. The first condition is assumed in the Theorem. For the second condition, as $\hat{\theta}^d_{nT} - \theta_T = O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{n}\right)\right)$ and $-\frac{1}{nT}E\frac{\partial^2 \ln L^d_{nT}(\hat{\theta}_{nT})}{\partial\theta\partial\theta'} = (\Sigma^d_{\theta_T,nT})^{-1} + O_p\left(\max\left(\frac{1}{\sqrt{nT}}, \frac{1}{n}\right)\right)$, we have

$$\begin{aligned}&\sqrt{\frac{T}{n}}\left\{\left(-\frac{1}{nT}E\frac{\partial^2 \ln L^d_{nT}(\hat{\theta}_{nT})}{\partial\theta\partial\theta'}\right)^{-1}a_n(\hat{\theta}^d_{nT}) - (\Sigma^d_{\theta_T,nT})^{-1}a_n(\theta_T)\right\} \\ = &\ (\Sigma^d_{\theta_T,nT})^{-1}\sqrt{\frac{T}{n}}\left(a_n(\hat{\theta}^d_{nT}) - a_n(\theta_T)\right) + a_n(\hat{\theta}^d_{nT}) \times O_p\left(\max\left(\frac{1}{n}, \sqrt{\frac{T}{n^3}}\right)\right) \\ = &\ (\Sigma^d_{\theta_T,nT})^{-1}\frac{\partial a_n(\theta^*_{nT})}{\partial\theta'}\sqrt{\frac{T}{n}}\left(\hat{\theta}^d_{nT} - \theta_T\right) + a_n(\hat{\theta}^d_{nT}) \times O_p\left(\max\left(\frac{1}{n}, \sqrt{\frac{T}{n^3}}\right)\right)\end{aligned}$$

where θ^*_{nT} lies between $\hat{\theta}^d_{nT}$ 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,

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

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, (22) follows. ■

Appendix E. The Transformation Approach in Section 3.2

E.1. The First and Second Order Derivatives of (25)

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 (2007a)), the first and second order derivatives of the concentrated log likelihood function (25) are

$$\frac{\partial \ln L_{n,T}(\theta)}{\partial \theta} = \begin{pmatrix} \frac{\partial \ln L_{n,T}(\theta)}{\partial \beta} \\ \frac{\partial \ln L_{n,T}(\theta)}{\partial \lambda} \\ \frac{\partial \ln L_{n,T}(\theta)}{\partial \rho} \\ \frac{\partial \ln L_{n,T}(\theta)}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'J_n\tilde{V}_{nt}(\zeta) \\ \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'J_n\tilde{V}_{nt}(\zeta) - (T-1)trJ_nG_n(\lambda) \\ \frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\zeta))'J_n\tilde{V}_{nt}(\zeta) - (T-1)trJ_nH_n(\rho) \\ \frac{1}{2\sigma^4}\sum_{t=1}^{T}(\tilde{V}_{nt}'(\zeta)J_n\tilde{V}_{nt}(\zeta) - (n-1)\frac{T-1}{T}\sigma^2) \end{pmatrix}, \quad (43)$$

and

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

$$= \begin{pmatrix} \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)\tilde{X}_{nt})'J_nR_n(\rho)\tilde{X}_{nt} & * & * & * \\ \frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'J_nR_n(\rho)\tilde{X}_{nt} & \begin{matrix}\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'J_nR_n(\rho)W_n\tilde{Y}_{nt} \\ +(T-1)tr(J_nG_n^2(\lambda))\end{matrix} & * & * \\ \begin{matrix}\frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\zeta))'J_nR_n(\rho)\tilde{X}_{nt}) \\ +\frac{1}{\sigma^2}\sum_{t=1}^{T}\tilde{V}_{nt}'(\zeta)J_nM_n\tilde{X}_{nt}\end{matrix} & \begin{matrix}\frac{1}{\sigma^2}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})_n'J_nH_n(\rho)\tilde{V}_{nt}(\zeta) \\ +\frac{1}{\sigma^2}\sum_{t=1}^{T}(M_nW_n\tilde{Y}_{nt})'J_n\tilde{V}_{nt}(\zeta)\end{matrix} & 0 & 0 \\ \frac{1}{\sigma^4}\sum_{t=1}^{T}\tilde{V}_{nt}'(\zeta)J_nR_n(\rho)\tilde{X}_{nt} & \frac{1}{\sigma^4}\sum_{t=1}^{T}(R_n(\rho)W_n\tilde{Y}_{nt})'J_n\tilde{V}_{nt}(\zeta) & 0 & 0 \end{pmatrix} \quad (44)$$

$$+ \begin{pmatrix} \mathbf{0}_{k\times k} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times 1} \\ \mathbf{0}_{1\times k} & 0 & 0 & 0 \\ \mathbf{0}_{1\times k} & 0 & \begin{bmatrix} \frac{1}{\sigma^2}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\zeta))'J_nH_n(\rho)\tilde{V}_{nt}(\zeta) \\ +(T-1)tr(J_nH_n^2(\rho)) \end{bmatrix} & * \\ \mathbf{0}_{1\times k} & 0 & \frac{1}{\sigma^4}\sum_{t=1}^{T}(H_n(\rho)\tilde{V}_{nt}(\zeta))'J_n\tilde{V}_{nt}(\zeta) & \begin{bmatrix} -\frac{(n-1)(T-1)}{2\sigma^4} \\ +\frac{1}{\sigma^6}\sum_{t=1}^{T}(\tilde{V}_{nt}'(\zeta)J_n\tilde{V}_{nt}(\zeta)) \end{bmatrix} \end{pmatrix}.$$

From (43), the score vector and the information matrix are

$$\frac{1}{\sqrt{(n-1)(T-1)}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial \theta} \quad (45)$$

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

and

$$\Sigma_{\theta_0,nT} = -E\left(\frac{1}{(n-1)(T-1)}\frac{\partial^2 \ln L_{n,T}(\theta_0)}{\partial\theta\partial\theta'}\right) \tag{46}$$

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

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

E.2. Some Lemmas

Lemma E.1 *Under Assumptions 1',2,3,4',5,6 and 7',* $\frac{1}{(n-1)(T-1)}\ln L_{n,T}(\theta) - Q_{n,T}(\theta) \xrightarrow{p} 0$ *uniformly in* $\theta\in\Theta$ *and* $Q_{n,T}(\theta)$ *is uniformly equicontinuous for* $\theta\in\Theta$*, where* $Q_{n,T}(\theta) = E\frac{1}{(n-1)(T-1)}\ln L_{n,T}(\theta)$.

Lemma E.2 *Under Assumption 7',* $\Sigma_{\theta_0,nT}$ *in (46) is nonsingular.*

Lemma E.3 *Under Assumptions 1',2,3,4',5,6 and 7'(a); or 1',2,3,4',5,6,7'(b) and 8',* $\frac{1}{\sqrt{(n-1)(T-1)}}\frac{\partial \ln L_{n,T}(\theta_0)}{\partial\theta} \xrightarrow{d} N(0, \lim(\Sigma_{\theta_0,nT} + \Omega_{\theta_0,nT}))$ *where* $\Sigma_{\theta_0,nT}$ *is in (46) and*

$$\Omega_{\theta_0,nT} = \frac{(T-1)}{T}\frac{(\mu_4 - 3\sigma_0^4)}{\sigma_0^4}\begin{pmatrix} \mathbf{0}_{k\times k} & * & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n-1}\sum_{i=1}^n[(J_n\ddot{G}_n)_{ii}]^2 & * & * \\ \mathbf{0}_{1\times k} & \frac{1}{n-1}\sum_{i=1}^n[(J_n\ddot{G}_n)_{ii}(J_nH_n)_{ii}] & \frac{1}{n-1}\sum_{i=1}^n[(J_nH_n)_{ii}]^2 & * \\ \mathbf{0}_{1\times k} & \frac{1}{2\sigma_0^2(n-1)}tr(J_n\ddot{G}_n) & \frac{1}{2\sigma_0^2(n-1)}tr(J_nH_n) & \frac{1}{4\sigma_0^4} \end{pmatrix}. \tag{47}$$

E.3. Proof for Theorem 9

The proof is similar to the direct approach in Appendix D. 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 resulted asymptotic distribution is properly centered at the true parameter vector. ■

References

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

Amemiya, T., 1985. *Advanced Econometrics.* Harvard University Press, Cambridge, MA.

Anderson, T.W. and C. Hsiao, 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. and A.K. Bera, 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics, A. Ullah and D.E.A. Giles (eds.). *Handbook of Applied Economics Statistics*, Marcel Dekker, New York.

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

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

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

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

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

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

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

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

Cox, D.R. and N. Reid, 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.

Ertur C. and W. Koch, 2007. Growth, technological interdependence and spatial externalities: theory and evidence. *Journal of Applied Econometrics* 22, 1033-1062.

Foote, C.L., 2007. Space and time in macroeconomic panel data: young workers and state-level unemployment revisited. Working Paper No. 07-10, Federal Reserve Bank of Boston.

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

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

Hausman, J.A., 1978. Specification tests in econometrics. *Econometrica* 46, 1251-1271.

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

Kalbfleisch, J.D. and D.A. Sprott, 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. and I.R. Prucha, 2007. Panel data models with spatially correlated error components. *Journal of Econometrics* 140, 97-130.

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

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

Kelejian H.H. and I.R. Prucha, 2007. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Forthcoming in *Journal of Econometrics.*

Kelejian, H.H. and D. Robinson, 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. and X. Liu, 2006. Efficient GMM estimation of a spatial autoregressive model with autoregressive disturbances. Manuscript, Ohio State University.

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

Lee, L.F. and J. Yu, 2007a. A spatial dynamic panel data model with both time and individual fixed effects. Manuscript, Ohio State University.

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

Lin, X. and L.F. Lee, 2005. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. Forthcoming in *Journal of Econometrics*.

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

Neyman, J. and E.L. Scott, 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. & A. Hussain, 1969. The use of error components models in combining cross-section and time-series data. *Econometrica* 37, 55-72.

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

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

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

Yu, J. and L.F. Lee, 2007. Estimation of unit root spatial dynamic panel data models. Manuscript, Ohio State University.

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	0.0683
(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 $\zeta_0 = (\beta_0', \lambda_0, \rho_0)'$.

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 $\theta_0^b = (1, 0.5, 0.2, 1)$.

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. θ_0^a= $(1, 0.2, 0.5, 1)$ and θ_0^b= $(1, 0.5, 0.2, 1)$.
2. The T-SD uses the bias corrected estimates.

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 $\theta_0^b = (1, 0.5, 0.2, 1)$.