INTERNATIONAL ECONOMIC REVIEW
Vol. 53, No. 4, November 2012

SPATIAL PANELS: RANDOM COMPONENTS VERSUS FIXED EFFECTS*

By Lung-fei Lee and Jihai Yu[1]

Ohio State University, U.S.A.; Peking University, China

This article investigates spatial panel data models with a space–time filter in disturbances. We consider their estimation by both fixed effects and random effects specifications. With a between equation properly defined, the difference of the random versus fixed effects models can be highlighted. We show that the random effects estimate is a pooling of the within and between estimates. A Hausman-type specification test and an Lagrangian multiplier test are proposed for the testing of the random components specification versus the fixed effects specification.

1. INTRODUCTION

Panel data with spatial interactions are of interest as they take into account dynamic and spatial dependence and also control for unobservable heterogeneity. Anselin (1988) provides a panel regression model with error components and spatial autoregressive (SAR) disturbances. Baltagi et al. (2003) consider specification tests for spatial correlation in that spatial panel regression model. Kapoor et al. (2007) propose a different specification with error components and an SAR structure in the overall disturbance and suggest a method of moments (MOM) estimation, and Fingleton (2008) adopts a similar approach to estimate a spatial panel model with SAR-dependent variables but with random components and a spatial moving average (SMA) structure in the overall disturbance. In an attempt to nest the Anselin (1988) and Kapoor et al. (2007) models, Baltagi et al. (2007a) suggest an extended model without restrictions on implied SAR structures in the error component and the remaining disturbance. As an alternative to the random effects specification, Lee and Yu (2010) investigate the estimation of spatial panel models under a fixed effects specification. The fixed effects model has the advantage of robustness in that fixed effects are allowed to depend on included regressors in the model. It also provides a unified model framework because different random effects models in Anselin (1988), Kapoor et al. (2007), and Baltagi et al. (2007a) reduce to the same fixed effects model.

In this article, we are interested in exploring the econometric content of random effects versus fixed effects spatial panels. As there are various random components specifications in spatial panels, we put forward a generalized specification in order to nest existing random effects spatial panels as special cases. Such a generalized model is motivated by Baltagi et al. (2007a) but goes beyond by including serial correlation and the SMA structure in disturbances. It incorporates different SAR structures on error components in Kapoor et al. (2007), generalized spatial error components in Baltagi et al. (2007a), serially correlated disturbances in Baltagi et al. (2007b), and also SMA disturbances in Fingleton (2008).[2] Although Kapoor et al. (2007) and Fingleton

*Manuscript received May 2010; revised November 2011.

[1] We would like to thank participants of Econometrics seminars in SUNY at Albany and Syracuse University, and also the audience in the 2010 North American Winter Meeting of the Econometric Society in Atlanta, Georgia; we would also like to thank Professors Badi Baltagi, Chihwa Kao, Kajal Lahiri, Kenneth Wolpin (the editor of this journal), and two anonymous referees for helpful comments. Yu acknowledges funding from the National Science Foundation of China (Grant No. 71171005) and support from the Center for Statistical Science of Peking University. Please address correspondence to: Jihai Yu, Guanghua School of Management, Peking University, Beijing, China 100871. Phone: +8610-62760702. Fax: +8610-62753624. E-mail: *jihai.yu@gmail.com*.

[2] This generalized spatial model provides a common specification for our analysis without restriction to a particular existing one. In addition, we can compare various special models within this generalized framework.

(2008) consider the MOM estimation and Baltagi et al. (2007a, 2007b) emphasize score tests of correlation, we investigate the maximum likelihood (ML) and quasi-maximum likelihood approaches for estimation and testing. Estimation under the fixed effects specification as well as that with random components will be considered.[3] The use of likelihood functions is revealing, as they can highlight the different econometric content of a random components specification versus its fixed effects counterpart; also, they can provide the conventional Hausman-type and Lagrangian multiplier (LM) tests of random versus fixed effects.

For the random effects specification in a linear regression panel data model, it is shown in Maddala (1971) that generalized least squares (GLS) estimates of regression coefficients are weighted averages of within and between estimates. Hence, by pooling the within and between estimators, the GLS estimate will be more efficient relative to the within estimator under the random effects specification. In this article, we show that, with a properly defined between equation, the random effects estimate may also be interpreted as a pooling of estimates that explore spatial units' within and between sample information.

For the linear regression panel model, the within estimate will be consistent under both random effects and fixed effects specifications. The random effects estimator is consistent and can be more efficient than the within estimator under the random effects specification; however, it would be inconsistent if individual effects were correlated with regressors. Hausman (1978) has proposed a test of correlation of random effects with regressors for the panel regression model. For the Kapoor et al. (2007)-type spatial panel model, Mutl and Pfaffermayr (2011) have considered a Hausman-type test based on two-stage least square (2SLS) estimates. In the likelihood framework, it is natural to consider the Hausman-type test via ML estimates. Difference of the log-likelihood functions of the random components and fixed effects models, where the latter explores only the spatial units' within sample information (within equation), is highlighted by the likelihood function of the between equation. The difference of various random components specifications would also be revealed via the between equation.

We conduct some Monte Carlo experiments to investigate the performances of various estimates and testing procedures in this article.[4] We find that (i) estimates based on the random effects information are more efficient than the estimates based only on the spatial units' within sample information under the random effects model, and estimates of the random effects model have some bias under the fixed effects model where individual effects are correlated with regressors, although estimates of within equation are consistent under both specifications; (ii) omission of serial correlation causes some bias in estimates for either the random effects model or within equation; (iii) misspecification of different random effects models causes some bias; and (iv) omission of either SAR disturbances or SMA disturbances causes bias in the estimation of spatial structures of disturbances.

In empirical applications with spatial panel data, it seems that investigators tend to limit their focus on some spatial structures and ignore others, and in addition, no serial correlation is considered. For example, Moscone et al. (2007) study the spatial correlation of public health expenditures among British counties using a random components panel model without serial correlation, where either the SAR-dependent variable or spatial disturbances are included in the regression (but not both). Thus, it is of interest to investigate possible consequences of misspecifications due to omitting some spatial or serial correlation structures. From our limited Monte Carlo experiments, we see that incorporating proper spatial and serial correlation structures into the panel model is important for the estimation and testing.

This article is organized as follows. Section 2 presents the general model specification and discusses the estimation. We first consider the estimation of the fixed effects model, followed

[3] There are studies in the literature on dynamic panel models with spatial interactions. See Elhorst (2005), Su and Yang (2007), Yu et al. (2008, 2012), Korniotis (2010), and Yu and Lee (2010) among others. However, dynamic models involve an initial value problem, which needs special treatment when the panel is short. Panel data models without dynamics do not, in general, have such an issue. Hence, estimation methods and asymptotic properties of estimators can be different for static and dynamic models.

[4] Matlab codes for estimates and tests are available on request.

by the random effects one. We demonstrate that the random effects model can be decomposed into a within equation and a between equation, and the estimate of the random effects model can be regarded as the pooling of estimates of within and between equations. The distinctions of various random components specifications are captured by the between equation. We focus on the situation with T being finite as in the literature for the panel regression model.[5] Section 3 investigates the Hausman and LM tests for the random effects specification. The between equation provides the role to determine proper degrees of freedom for the Hausman test. Section 4 provides some Monte Carlo results on performances of estimates and test statistics. Also, consequences of misspecifications of omitting spatial and/or serial correlations are reported. Section 5 concludes. Some algebra and proofs are collected in the Appendix.

2. THE GENERAL SPATIAL PANEL MODEL

Consider the following model:

$$
\begin{aligned}
Y_{nt} &= \lambda_{10} W_{n1} Y_{nt} + z_n b_0 + X_{nt}\beta_0 + \mu_n + U_{nt}, \\
U_{nt} &= \lambda_{20} W_{n2} U_{nt} + (I_n + \delta_{20} M_{n2}) V_{nt} \text{ for } t = 1, \ldots, T, \\
\mu_n &= \lambda_{30} W_{n3} \mu_n + (I_n + \delta_{30} M_{n3}) \mathbf{c}_{n0}, \\
V_{nt} &= \rho_0 V_{n,t-1} + e_{nt} \text{ for } t = 2, \ldots, T,
\end{aligned}
\tag{1}
$$

where Y_{nt} is an $n \times 1$ column vector, X_{nt} is an $n \times k_x$ matrix of nonstochastic time varying regressors, and z_n is an $n \times k_z$ matrix captures nonstochastic time invariant regressors including the constant intercept.[6] The W_{nj} and M_{nj} are $n \times n$ nonstochastic spatial weights matrices that generate the spatial dependence, μ_n is an n-dimensional vector of individual effects[7] with spatial interactions, U_{nt} is the SAR error, which is also serially correlated, and $\mathbf{c}_{n0}$ and e_{nt} are independent with i.i.d. elements such that $c_{n0,i} \sim \text{i.i.d.}(0, \sigma_{c0}^2)$ and $e_{nt,i} \sim \text{i.i.d.}(0, \sigma_{e0}^2)$. As in the literature, we assume stationarity so that $V_{n1} \sim (0, (\sigma_{e0}^2/(1-\rho_0^2))I_n)$ and is independent with e_{nt} for $t = 2, \ldots, T$. In addition, U_{nt} and μ_n are allowed to incorporate the SMA feature. This is a generalized spatial panel model that incorporates spatial correlation, heterogeneity, and serial correlation in disturbances. It nests various spatial panels existing in the literature (where the abbreviation of models is from Baltagi et al., 2012):

(i) Kapoor et al. (2007) SAR-RE: $\lambda_{20} = \lambda_{30}$, $\delta_{20} = \delta_{30} = 0$, $\rho_0 = 0$, and $W_{n2} = W_{n3}$.
(ii) Fingleton (2008) SMA-RE: $\lambda_{20} = \lambda_{30} = 0$, $\delta_{20} = \delta_{30}$, $\rho_0 = 0$, and $M_{n2} = M_{n3}$.
(iii) Anselin (1988) RE-SAR: $\lambda_{30} = 0$, $\delta_{20} = 0$, $\delta_{30} = 0$, and $\rho_0 = 0$.
(iv) Anselin et al. (2008) RE-SMA: $\lambda_{20} = 0$, $\lambda_{30} = 0$, $\delta_{30} = 0$, and $\rho_0 = 0$.
(v) Baltagi et al. (2007a) Generalized RE-SAR: $\delta_{20} = 0$, $\delta_{30} = 0$ and $\rho_0 = 0$.
(vi) Baltagi et al. (2007b) RE-SAR with serial correlation: $\lambda_{30} = 0$, $\delta_{20} = 0$, and $\delta_{30} = 0$.
In addition, it also includes
(vii) SARMA-RE, where $\mu_n + U_{nt} = (I_n - \lambda_{20} W_{n2})^{-1}(I_n + \delta_{20} M_{n2})(\mathbf{c}_{n0} + V_{nt})$ with $\lambda_{20} = \lambda_{30}$, $\delta_{20} = \delta_{30}$, and $\rho_0 = 0$.

[5] In footnote 17, we make some brief comments on the situation with T tending to infinity.

[6] The exogenous variables X_{nt} (and also z_n) can include spatial features such that the regressor function includes $X_{nt}\beta_0$ and $M_{n1}X_{nt}\delta_{10}$, where $M_{n1}X_{nt}$ can capture the so-called spatial Durbin regressors (LeSage and Pace, 2009). Although the spatial Durbin regressors may be of interest in empirical applications, it does not introduce additional complication in theoretical analysis. Thus, for simplicity, we only use $X_{nt}\beta_0$ (and $z_n b_0$) for the regressor function.

[7] We may also have time effects in the model. In short panels, time effects will not cause the incidental parameter problem and they can be treated as regressors. However, for long panels, to avoid the incidental parameter problem, we may eliminate them before the estimation or allow them in the regression by the random effects specification.

Parent and LeSage (2008) apply the Markov Chain Monte Carlo method to a linear panel regression model where spatially and serially correlated disturbances are present. The product of the quasi-difference over time and the spatial transformation is called the space-time filter in Parent and LeSage (2008). In this article, the serial correlation in V_{nt} incorporates the space-time filter in U_{nt}. We consider the ML estimation under both fixed effects and random effects specifications on μ_n. The parameter subvector $\theta_{10} = (\beta_0', \lambda_{10}, \lambda_{20}, \delta_{20}, \rho_0, \sigma_{e0}^2)'$ can be estimated from both fixed and random effects models. The remaining parameters in $\theta_{20} = (b_0', \lambda_{30}, \delta_{30}, \sigma_{c0}^2)'$ can only be estimated under the random effects specification.

2.1. *Fixed Individual Effects.* With individual effects μ_n being fixed parameters, the model is

$$
\begin{aligned}
(2) \qquad & Y_{nt} = \lambda_{10} W_{n1} Y_{nt} + X_{nt}\beta_0 + \mu_n + U_{nt}, \\
& U_{nt} = \lambda_{20} W_{n2} U_{nt} + (I_n + \delta_{20} M_{n2}) V_{nt} \quad \text{for} \quad t = 1, \ldots, T, \\
& V_{nt} = \rho_0 V_{n,t-1} + e_{nt} \quad \text{for} \quad t = 2, \ldots, T,
\end{aligned}
$$

where the time invariant regression function $z_n b_0$ has been implicitly absorbed by μ_n. Because elements of μ_n are fixed parameters, their spatial structure would be irrelevant. Thus, the specifications in Anselin (1988) and Kapoor et al. (2007) would yield the same fixed effects model once the random effects are conditioned upon. For the linear panel regression model with or without serial correlation, relevant model specifications and estimation methods are summarized in Hsiao (2003) and Baltagi (2008). For the panel regression model with serially correlated disturbances, Kiefer (1980) and Bhargava et al. (1982) investigate the fixed effects specification and recognize possible inconsistency of estimates when T is finite, in particular, the serial correlation parameters.[8]

This section will consider the estimation method by eliminating fixed effects before estimation. To eliminate individual effects, we suggest the use of first difference rather than the deviation from time mean. This is so because serial correlation in time series can be better dealt with in a recursive fashion.[9] For any $n \times 1$ vector Z_{nt}, denote $\Delta Z_{nt} = Z_{nt} - Z_{n,t-1}$ as the first difference. With the first difference on (2), it gives

$$
(3) \qquad
\begin{aligned}
& \Delta Y_{nt} = \lambda_{10} W_{n1} \Delta Y_{nt} + \Delta X_{nt}\beta_0 + \Delta U_{nt}, \\
& \Delta U_{nt} = \lambda_{20} W_{n2} \Delta U_{nt} + (I_n + \delta_{20} M_{n2}) \Delta V_{nt} \quad \text{for} \quad t = 2, \ldots, T,
\end{aligned}
$$

where

$$
\begin{aligned}
& \Delta V_{nt} = \rho_0 \Delta V_{n,t-1} + \Delta e_{nt} \quad \text{for} \quad t = 3, \ldots, T, \\
& \Delta V_{n2} = e_{n2} - (1 - \rho_0) V_{n1},
\end{aligned}
$$

and Δe_{nt} becomes a moving average (MA) process. Define $(T-1) \times (T-1)$ matrices $Q_{T-1} = Q_{T-1}(\rho_0)$ in (A.5). We see that Q_{T-1} is the quasi-difference transformation matrix. Denote $\mathbf{Y}_{nT} = (Y_{n1}', Y_{n2}', \ldots, Y_{nT}')'$, $\mathbf{Y}_{n,T-1}^d = (\Delta Y_{n2}', \ldots, \Delta Y_{nT}')'$, and other variables accordingly. Let $S_{nj}(\lambda_j) = I_n - \lambda_j W_{nj}$ for any possible λ_j for $j = 1, 2, 3$ and $B_{nj}(\delta_j) = (I_n + \delta_j M_{nj})$ for any possible

[8] The incidental parameter problem is also well known for dynamic panel models with fixed effects (Nickell, 1981).

[9] If the time deviation transformation is used, the disturbances in the estimation equation would be

$$
(*) \qquad V_{nt} - \bar{V}_{n.} = \rho_0 \big(V_{n,t-1} - \bar{V}_{n.}^{(-1)}\big) + e_{nt} - \bar{e}_{n.},
$$

for $t = 2, \ldots, T$, where $\bar{V}_{n.} = \frac{1}{T}\sum_{t=1}^{T} V_{nt}$ and $\bar{V}_{n.}^{(-1)} = \frac{1}{T}\sum_{t=0}^{T-1} V_{nt}$. The process in (*) is no longer a Markov process and does not process a recursive structure. In particular, we note that $V_{nt} - \bar{V}_{n.}$ and $V_{n,t-1} - \bar{V}_{n.}^{(-1)}$ for $t = 2, \ldots, T$ are not the same random variables. A quasi-difference to eliminate the serial correlation in (*) would not be feasible. Also, the variance matrix of such random variables over time would be complicated, and so would be its inverse.

δ_j for $j = 2, 3$; correspondingly, let $S_{nj} = I_n - \lambda_{j0} W_{nj}$ for $j = 1, 2, 3$ and $B_{nj} = (I_n + \delta_{j0} M_{nj})$ for $j = 2, 3$ at true parameter values. Then (3) can be rewritten as

$$(I_{T-1} \otimes S_{n1})\mathbf{Y}^d_{n,T-1} = \mathbf{X}^d_{n,T-1}\beta_0 + \big(I_{T-1} \otimes S^{-1}_{n2} B_{n2}\big)\mathbf{V}^d_{n,T-1},$$

where

$$(Q_{T-1} \otimes I_n)\mathbf{V}^d_{n,T-1} = \mathbf{e}^d_{n,T-1} = (\Delta V'_{n2}, \Delta e'_{n3}, \ldots, \Delta e'_{nT})'.$$

Thus, one has

$$(Q_{T-1} \otimes S_{n1})\mathbf{Y}^d_{n,T-1} = (Q_{T-1} \otimes I_n)\mathbf{X}^d_{n,T-1}\beta_0 + \big(I_{T-1} \otimes S^{-1}_{n2} B_{n2}\big)\mathbf{e}^d_{n,T-1}.$$

As $\Delta V_{n2} = e_{n2} - (1 - \rho_0)V_{n1}$, we have $\mathrm{Var}(\Delta V_{n2}) = (2\sigma^2_{e0}/(1 + \rho_0))I_n$, $\mathrm{Cov}(\Delta V_{n2}, \Delta e_{n3}) = -\sigma^2_{e0} I_n$, and $\mathrm{Cov}(\Delta V_{n2}, \Delta e_{nt}) = 0$ for $t = 4, \ldots, T$.[10] Define $H_{T-1} = H_{T-1}(\rho_0)$ in (A.6). The $\sigma^2_{e0} H_{T-1} \otimes I_n$ is the variance matrix[11] of $\mathbf{e}^d_{n,T-1}$. The likelihood function (when the disturbances are normal; otherwise, a quasi-likelihood) of θ_1 in terms of $\mathbf{Y}^d_{n,T-1}$ is therefore

(4)
$$L_{w,nT}(\theta_1) = (2\pi)^{-n(T-1)/2} |S_{n1}(\lambda_1)|^{T-1} |S_{n2}(\lambda_2)|^{T-1} \left|B_{n2}(\delta_2)\right|^{-(T-1)} \times \left|\sigma^2_e H_{T-1}(\rho)\right|^{-n/2} |Q_{T-1}(\rho)|^n \exp\left(-\frac{1}{2\sigma^2_e}\mathbf{e}'^d_{n,T-1}(\theta_1)\big(H^{-1}_{T-1}(\rho) \otimes I_n\big)\mathbf{e}^d_{n,T-1}(\theta_1)\right),$$

where

$$\mathbf{e}^d_{n,T-1}(\theta_1) = \big(I_{T-1} \otimes B^{-1}_{n2}(\delta_2) S_{n2}(\lambda_2)\big)\big[(Q_{T-1}(\rho) \otimes S_{n1}(\lambda_1))\mathbf{Y}^d_{n,T-1} - (Q_{T-1}(\rho) \otimes I_n)\mathbf{X}^d_{n,T-1}\beta\big]$$

and $|Q_{T-1}(\rho)|^n = 1$.

Alternatively, denoting $\mathbf{V}_{nT}(\theta_1) = (I_T \otimes B^{-1}_{n2}(\delta_2) S_{n2}(\lambda_2))\,[(I_T \otimes S_{n1}(\lambda_1))\mathbf{Y}_{nT} - \mathbf{X}_{nT}\beta - \mathbf{c}_n]$ with $\mathbf{c}_n$ being an arbitrary vector for $\mathbf{c}_{n0}$, the likelihood function can be written based on $\mathbf{V}_{nT}(\theta_1)$. By defining $L_{T-1,T}$ in (A.4) as the first difference transformation matrix, we have the relation $\mathbf{e}^d_{n,T-1}(\theta_1) = (Q_{T-1}(\rho) L_{T-1,T} \otimes I_n)\mathbf{V}_{nT}(\theta_1)$.[12] This implies that $\mathbf{e}'^d_{n,T-1}(\theta_1)(H^{-1}_{T-1}(\rho) \otimes I_n)\mathbf{e}^d_{n,T-1}(\theta_1)$ in (4) can be written as $\mathbf{V}'_{nT}(\theta_1)(\mathbb{J}_T(\rho) \otimes I_n)\mathbf{V}_{nT}(\theta_1)$, where[13] $\mathbb{J}_T(\rho) \equiv L'_{T-1,T} Q'_{T-1}(\rho) H^{-1}_{T-1}(\rho) Q_{T-1}(\rho) L_{T-1,T}$ has the explicit expression in (A.8). The $\mathbb{J}_T(\rho)$ matrix is to eliminate the individual effects and serial correlation in the disturbances $\mathbf{V}_{nT}$. When $\rho = 0$, we have $\mathbb{J}_T(0) = L'_{T-1,T} H^{-1}_{T-1} L_{T-1,T} = I_T - l_T l'_T/T$ as the usual deviation from time mean operation, where l_T is a $T \times 1$ vector of ones. Thus, the log-likelihood (4) can be rewritten as

[10] If V_{n1} were assumed to be fixed, $\mathrm{Var}(\Delta V_{n2}) = \mathrm{Var}(e_{n2}) = \sigma^2_{e0} I_n$, $\mathrm{Cov}(\Delta V_{n2}, \Delta e_{n3}) = -\sigma^2_{e0} I_n$, and $\mathrm{Cov}(\Delta V_{n2}, \Delta e_{nt}) = 0$, for $t = 4, \ldots, T$. Also, when V_{n1} were random, it could be in general $(0, \frac{\sigma^2_{e0}}{\tau} I_n)$ in Baltagi and Li (1991). If these were the selected specifications, the (1,1) entry of $H_{T-1}(\rho)$ in the following would be modified accordingly.

[11] For $H_{T-1}(\rho)$, its inverse and determinant have closed form expressions (see Hsiao et al., 2002, or Appendix A in this article).

[12] Even though $\mathbf{c}_n$ appears in $\mathbf{V}_{nT}(\theta_1)$, it would be eliminated by the differencing operator $L_{T-1,T}$. Thus, one may ignore $\mathbf{c}_n$ in $\mathbf{V}_{nT}(\theta_1)$ for analysis.

[13] The inverse of $H_{T-1}(\rho)$ is needed to derive $\mathbb{J}_T(\rho)$. As $H^{-1}_{T-1}(\rho)$ is not uniformly bounded in row and column sums for a T sequence although $\mathbb{J}_T(\rho)$ is, it is desirable to work with $\mathbb{J}_T(\rho)$ in (5) instead of $H^{-1}_{T-1}(\rho)$ in (4) for analysis.

(5) $$\ln L_{w,nT}(\theta_1) = -\frac{n(T-1)}{2}\ln\left(2\pi\sigma_e^2\right) + (T-1)\left(\ln|S_{n1}(\lambda_1)| + \ln|S_{n2}(\lambda_2)| - \ln|B_{n2}(\delta_2)|\right) - \frac{n}{2}\ln|H_{T-1}(\rho)| - \frac{1}{2\sigma_e^2}\mathbf{V}'_{nT}(\theta_1)(\mathbb{J}_T(\rho)\otimes I_n)\mathbf{V}_{nT}(\theta_1),$$

with its score and information matrix in Appendix B.1. For the evaluation of this log-likelihood function, it involves inverses of $S_{n1}(\lambda_1)$, $S_{n2}(\lambda_2)$, and $B_{n2}(\delta_2)$ and their determinants, which would be the same as those for a conventional SAR model.[14]

Although the estimation of the equations in (3) is based on the difference of sample observations, i.e., $\mathbf{Y}^d_{n,T-1}$, one may interpret the estimation via the likelihood function in (5) as a "within" estimation based on the use of sample observations deviated from time averages, as in the traditional estimation of a fixed effects panel regression model in Maddala (1971). This is so as follows. Let $y = (y_1, \ldots, y_T)'$ be a vector of random variables of dimension T. Consider the difference transformation $L_{T-1,T}$ and the deviation from the time average transformation $J_T = I_T - l_T l'_T/T$. The $L_{T-1,T}y = (y_2 - y_1, \ldots, y_T - y_{T-1})'$ is a $(T-1)$ -dimensional vector and the $J_T y = (y_1 - \bar{y}_., \ldots, y_T - \bar{y}_.)$ is a T-dimensional vector, where random variables of $J_T y$ are linearly dependent. To construct the likelihood function of $J_T y$, it is known that one may drop any single observation of $J_T y$ even though there is serial correlation among y (e.g., Kiefer, 1980). Thus, the likelihood function of $J_T y$ can be constructed from $(y_2 - \bar{y}_., \ldots, y_T - \bar{y}_.)$. There are relations between the time-deviated observations and the observations in difference: We have the identities $y_t - \bar{y}_. = \sum_{l=0}^{T-(t+1)}(y_{t+l} - y_{t+1+l}) - \frac{1}{T}\sum_{s=1}^{T-1} s(y_s - y_{s+1})$ for $t = 2, \ldots, T$, and, conversely, $y_t - y_{t-1} = (y_t - \bar{y}_.) - (y_{t-1} - \bar{y}_.)$ for $t = 3, \ldots, T$, and $y_2 - y_1 = (y_2 - \bar{y}_.) + \sum_{t=2}^{T}(y_t - \bar{y}_.)$. These provide a one-to-one linear transformation of the random vector $(y_2 - y_1, \cdots, y_T - y_{T-1})$ onto $(y_2 - \bar{y}_., \ldots, y_T - \bar{y}_.)$. Thus, the density function of $(y_2 - y_1, \cdots, y_T - y_{T-1})$ is the same as that of $(y_2 - \bar{y}_., \ldots, y_T - \bar{y}_.)$. For the equations of the spatial panel model in (2), with the time difference, one has the transformed equation

$$\Delta Y_{nt} = \lambda_{10} W_{n1}\Delta Y_{nt} + \Delta X_{nt}\beta_0 + \Delta U_{nt},$$

for $t = 2, \ldots, T$ in (3), but, with time deviation,

(6) $$Y_{nt} - \bar{Y}_{n.} = \lambda_{10} W_{n1}(Y_{nt} - \bar{Y}_{n.}) + (X_{nt} - \bar{X}_{n.})\beta_0 + (U_{nt} - \bar{U}_{n.}),$$

for $t = 2, \ldots, T$.[15] As $(\Delta U_{n2}, \cdots, \Delta U_{nT})$ and $(U_{n2} - \bar{U}_{n.}, \ldots, U_{nT} - \bar{U}_{n.})$ have the same density function, the likelihood functions of these two transformed equation systems are the same. In the time deviation version, observations $Y_{nt} - \bar{Y}_{n.}$ and $X_{nt} - \bar{X}_{n.}$ for $t = 2, \ldots, T$ provide explicit information of variations within each spatial unit. In the time difference version, ΔY_{nt} and ΔX_{nt} for $t = 2, \ldots, T$, have implicitly provided such information. For the panel regression model, the least squares dummy variable (LSDV) estimator of the regression coefficients has been termed the within estimator. As in Maddala (1971), this follows from within group regression. For our case, the within group information is captured in the likelihood function in (5). The estimation method via such a likelihood function can also be interpreted as a "within" estimation, even though the estimator is not simply a least square estimator. According to Lancaster (2000), this likelihood function does not involve individual effects, and it is a partial likelihood function.

For our asymptotic analysis, we make the following assumptions.

[14] If W_{n1}, W_{n2}, and M_{n2} are diagonalizable, they can be evaluated with their eigenvalues and eigenvectors (see, e.g., Ord, 1975). For example, as $M_{n2} = q_n \Lambda_n q_n^{-1}$ where Λ_n is the diagonal eigenvalue matrix and q_n is the corresponding eigenvector matrix, $B_{n2}^{-1} = q_n(I_n + \delta_{20}\Lambda_n)^{-1}q_n^{-1}$.

[15] Note that in the transformed equations with either time difference or time deviation, linear dependence of the transformed random variables has been removed by ignoring the corresponding $t = 1$ observation for each spatial unit.

ASSUMPTION 1. *W_{nj} for $j=1,2,3$ and M_{nj} for $j=2,3$ are nonstochastic spatial weights matrices with zero diagonals.*

ASSUMPTION 2. *The disturbances $\{e_{nt,i}\}$, $i=1,2,\ldots,n$ and $t=2,3,\ldots,T$, are i.i.d. across i and t with zero mean, variance σ_{e0}^2,* and *$E|e_{nt,i}|^{4+\eta}<\infty$ for some $\eta>0$; also, they are independent with $V_{n1}\sim(0,(\sigma_{e0}^2/(1-\rho_0^2))I_n)$.*

ASSUMPTION 3. *$S_{nj}(\lambda_j)$ for $j=1,2,3$ and $B_{nj}(\delta_j)$ for $j=2,3$ are invertible for all $\lambda_j\in\Lambda_j$ and $\delta_j\in\Delta_j$, and $\rho\in\mathbb{P}$, where Λ_j and Δ_j are compact intervals and $\mathbb{P}$ is a compact subset in $(-1,1)$. Furthermore, λ_{j0}, δ_{j0}, and ρ_0 are, respectively, in the interiors of Λ_j, Δ_j, and $\mathbb{P}$.*

ASSUMPTION 4. *W_{nj} for $j=1,2,3$ and M_{nj} for $j=2,3$ are uniformly bounded in both row and column sums in absolute value (for short, UB).[16] Also, $S_{nj}^{-1}(\lambda_j)$ and $B_{nj}^{-1}(\delta_j)$ are UB, uniformly in $\lambda_j\in\Lambda_j$ and $\delta_j\in\Delta_j$.*

ASSUMPTION 5. *n is large, where T is finite.*

ASSUMPTION 6. *Elements of the $n\times k_x$ matrix of regressors X_{nt} are nonstochastic and bounded, uniformly in n and t. Also, under the asymptotic setting in Assumption 5, the limit of $\frac{1}{n(T-1)}\sum_{t=1}^{T}\mathbf{X}_{nT}'(\mathbb{J}_T\otimes S_{n2}'B_{n2}'^{-1}B_{n2}^{-1}S_{n2})\mathbf{X}_{nT}$ exists and is nonsingular.*

The zero diagonal assumption in Assumption 1 helps the interpretation of the spatial effect, as self-influence shall be excluded in practice. In many empirical applications, each of the rows of W_{nj} (and M_{nj}) sums to 1, which ensures that all the weights are between 0 and 1. In general, our estimation and analysis for the model do not require the feature of row-normalization on spatial weights matrices. We note that W_{nj}'s and M_{nj}'s may or may not be the same in practice. Assumption 2 specifies an i.i.d. assumption for $e_{nt,i}$. If there were unknown heteroskedasticity in e_{nt}, the maximum likelihood estimator (MLE) would not be consistent. Methods such as the generalized method of moments in Lin and Lee (2010) and that in Kelejian and Prucha (2010) would be designed for that situation. Invertibility of $S_{nj}(\lambda_j)$ and $B_{nj}(\delta_j)$ in Assumption 3 guarantees that we have a valid reduced form for the SAR representation. Also, compactness of parameter spaces is a convenient condition for theoretical analysis on nonlinear functions.[17] When W_{nj} (and M_{nj}) is row-normalized, a compact subset of $(-1,1)$ has often been taken as the parameter space for λ_j (and δ_j) in theory. Assumption 4 is originated by Kelejian and Prucha (1998, 2001) and also used in Lee (2004, 2007). That W_{nj}, M_{nj}, $S_{nj}^{-1}(\lambda_j)$, and $B_{nj}^{-1}(\delta_j)$ are UB is a condition to limit the spatial correlation to a manageable degree. Assumption 5 is assumed for the short-panel data case.[18] We note that the serial correlation coefficient ρ_0 can be consistently estimated under large n even when T is small, where the inference of ρ_0 is obtained mainly from the cross-sectional variation of the data.[19] The case with a finite n but a large T is of less interest as spatial models are mainly designed for the large n case. When exogenous variables X_{nt} are included in the model, it is convenient to assume that they are uniformly bounded

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

[17] Due to the nonlinearity of λ_j and δ_j in the reduced form of the model, compactness of Λ_j and Δ_j is needed. However, the compactness of β and σ_e^2 is not necessary because the β and σ_e^2 estimates given λ_j and δ_j are least squares-type estimates.

[18] In a previous version of this article, we also consider the large T case where additional time effects are introduced. We show that the within estimate is asymptotically as efficient as the random effects estimate as one would expect.

[19] This is similar to a short dynamic panel data model with a proper treatment of the initial observation for each unit. Hsiao et al. (2002) have studied MLE estimation of such a model.

as in Assumption 6. If X_{nt} is allowed to be stochastic and unbounded, appropriate moment conditions can be imposed instead.

We also make assumptions to establish the consistency and asymptotic distribution of the within estimate.

ASSUMPTION 7. *Either (a) the conditions (D.3) and (D.4) in Appendix D.1 hold or (b) the condition (D.5) holds if (a) fails.*

ASSUMPTION 8. *The limit of the information matrix (B.1) is nonsingular.*

Assumption 7 specifies identification conditions of the within equation. Part (a) of Assumption 7 represents the possible identification of λ_{10} and β_0 through the deterministic part of the reduced form equation in (3), and λ_{20}, δ_{20}, ρ_0, and σ_{e0}^2 from the SAR process of ΔU_{nt} in (3). Part (b) of Assumption 7 states the identification through the SAR process of the reduced form of disturbances of ΔY_{nt} in (3) in the event that the identification in (a) is not possible. In (2), the disturbance process of U_{nt} is a spatial autogressive and moving average process in a general form. As in the time-series literature, a special case with restrictions on coefficients of such a process would have the autoregressive and moving average operators cancelled out so that the coefficients are not identifiable. This would also be the case for the spatial process when $M_{n2} = W_{n2}$ and $\delta_{20} = -\lambda_{20}$. The conditions (D.4) and (D.5) would rule out such an underidentification case.[20] Assumption 8 is for the nonsingularity of the limit of the information matrix for the within equation.

PROPOSITION 1. *Under Assumptions*[21] *1–7(a), or Assumptions 1–6, 7(b), and 8, the within estimate $\hat{\theta}_{w1}$ of θ_{10} under the fixed effects specification from* (5) *is consistent and asymptotically normal:*

$$\sqrt{n}(\hat{\theta}_{w1} - \theta_{10}) \xrightarrow{d} N\left(0, \lim_{n\to\infty} \Sigma_{w,nT}^{-1}(\Sigma_{w,nT} + \Omega_{w,nT})\Sigma_{w,nT}^{-1}\right),$$

where $\Sigma_{w,nT}$ in (B.1) is the information matrix, and $\Omega_{w,nT}$ in (B.2) is related to the third and fourth moments of $e_{nt,i}$ ($\Omega_{w,nT} = 0$ *under normality*).

2.2. *Random Individual Effects.* For the random effects specification of μ_n, the SAR and SMA features in μ_n could be regarded as permanent (global and local) spillover effects as described in Baltagi et al. (2007a). Hence, the model (1) can be written in the vector form with nT observations as

$$\mathbf{Y}_{nT} = l_T \otimes z_n b_0 + \lambda_{10}(I_T \otimes W_{n1})\mathbf{Y}_{nT} + \mathbf{X}_{nT}\beta_0 + \boldsymbol{\xi}_{nT}, \tag{7}$$

where $\boldsymbol{\xi}_{nT} = l_T \otimes S_{n3}^{-1}B_{n3}\mathbf{c}_{n0} + (I_T \otimes S_{n2}^{-1}B_{n2})\mathbf{V}_{nT}$ is the overall disturbance. The variance matrix of $\boldsymbol{\xi}_{nT}$ is

$$\Omega_{nT} = \sigma_{c0}^2\left[l_T l_T' \otimes S_{n3}^{-1}B_{n3}B_{n3}'S_{n3}'^{-1}\right] + \sigma_{e0}^2\left[\Sigma_{T,\rho_0} \otimes S_{n2}^{-1}B_{n2}B_{n2}'S_{n2}'^{-1}\right], \tag{8}$$

[20] Whenever $\delta_{20} = -\lambda_{20}$ and $M_{n2} = W_{n2}$, we have $S_{n2}^{-1}(\lambda_2)B_{n2}(\delta_2) = I_n$ and $p_{n,T-1}(\lambda_{10}, \lambda_2, \delta_2, \rho_0) = \sigma_{e0}^2$ in (D.4) and (D.5). Thus, the equations in (D.4) and (D.5) would become zero.

[21] We do not need $j = 3$ for Assumptions 1, 3, and 4 under the fixed effects model.

because $E(\mathbf{V}_{nT}\mathbf{V}'_{nT}) = \sigma^2_{e0}\Sigma_{T,\rho_0} \otimes I_n$ with $\Sigma_{T,\rho_0} = \Sigma_T(\rho_0)$ in (A.1). Hence, the log quasi-likelihood for (7) is

$$\text{(9)} \quad \ln L_{r,nT}(\theta) = -\frac{nT}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega_{nT}(\theta)| + T\ln|S_{n1}(\lambda_1)| - \frac{1}{2}\boldsymbol{\xi}'_{nT}(\theta)\Omega^{-1}_{nT}(\theta)\boldsymbol{\xi}_{nT}(\theta),$$

where $\boldsymbol{\xi}_{nT}(\theta) = (I_T \otimes S_{n1}(\lambda_1))\mathbf{Y}_{nT} - \mathbf{X}_{nT}\beta - l_T \otimes z_n b$. The score and information matrix are in Appendix B.2.

To evaluate the log-likelihood function (9), we need to compute the determinant and inverse of the $nT \times nT$ variance matrix Ω_{nT}. By Lemma 2.2 in Magnus (1982), $|\Omega_{nT}| = |\Sigma_{T,\rho_0}|^n \cdot |\sigma^2_{e0}S^{-1}_{n2}B_{n2}B'_{n2}S'^{-1}_{n2}|^{T-1} \cdot |Z^{-1}_{n0}|$ and

$$\begin{aligned}\Omega^{-1}_{nT} &= \frac{1}{d^2(1-\rho_0)^2}\Sigma^{-1}_{T,\rho_0} l_T l'_T \Sigma^{-1}_{T,\rho_0} \otimes Z_{n0} \\ &\quad + \left(\Sigma^{-1}_{T,\rho_0} - \frac{1}{d^2(1-\rho_0)^2}\Sigma^{-1}_{T,\rho_0} l_T l'_T \Sigma^{-1}_{T,\rho_0}\right) \otimes \left(\frac{1}{\sigma^2_{e0}} S'_{n2}B'^{-1}_{n2}B^{-1}_{n2}S_{n2}\right),\end{aligned}$$

where $Z_{n0} = [d^2(1-\rho_0)^2\sigma^2_{c0}S^{-1}_{n3}B_{n3}B'_{n3}S'^{-1}_{n3} + \sigma^2_{e0}S^{-1}_{n2}B_{n2}B'_{n2}S'^{-1}_{n2}]^{-1}$ with $d^2 = l'^{\rho}_T l^{\rho}_T = \frac{1+\rho_0}{1-\rho_0} + (T-1)$ and $l^{\rho}_T = (\sqrt{\frac{1+\rho_0}{1-\rho_0}}, 1, \ldots, 1)'$. Thus, the computation of the determinant and inverse of Ω_{nT} involves only $n \times n$ matrices.

ASSUMPTION 9. *$\mathbf{c}_{n0} \sim (0, \sigma^2_{c0}I_n)$ and $\mathbf{e}_{nT}$ are i.i.d. and independent of $\mathbf{X}_{nT}$ and z_n, where $\mathbf{e}_{nT} = (\sqrt{1-\rho^2_0}V'_{n1}, e'_{n2}, \ldots, e'_{nT})'$. Also, $\mathbf{c}_{n0}$ is independent of $\mathbf{e}_{nT}$.*

ASSUMPTION 10. *Elements of the $n \times k_x$ matrix of regressors X_{nt} and the $n \times k_z$ matrix z_n are nonstochastic and bounded, uniformly in n and t. Also, under the asymptotic setting in Assumption 5, the limit of $\frac{1}{nT}\sum_{t=1}^T \mathbf{Z}'_{nT}\Omega^{-1}_{nT}\mathbf{Z}_{nT}$ exists and is nonsingular where $\mathbf{Z}_{nT} = [l_T \otimes z_n, \mathbf{X}_{nT}]$.*

ASSUMPTION 11. *Either (a) the conditions (D.7) and (D.8) in Appendix D.2 hold or (b) the condition (D.9) holds if (a) fails.*

ASSUMPTION 12. *The limit of the information matrix (B.3) is nonsingular.*

Part (a) of Assumption 11 represents the possible identification of λ_{10}, b_0, and β_0 through the deterministic part of the reduced form equation of (7) and the identification of the rest of parameters from overall disturbances in (7). Part (b) of Assumption 11 states the identification through the SAR process of reduced form disturbances of Y_{nt} in (7) in the event that the identification via (a) fails. Assumption 12 is a condition for the nonsingularity of the limiting information matrix for the random effects model.

PROPOSITION 2. *Under Assumptions 1–5, 9, 10, 11(a), or Assumptions 1–5, 9, 10, 11(b), and 12, the estimates $\hat{\theta}_r = (\hat{\theta}'_{r1}, \hat{\theta}'_{r2})'$ of θ_0 from (9) of the random effects model are consistent and asymptotically normal:*

$$\sqrt{n}(\hat{\theta}_r - \theta_0) \xrightarrow{d} N\left(0, \lim_{n\to\infty} \Sigma^{-1}_{r,nT}(\Sigma_{r,nT} + \Omega_{r,nT})\Sigma^{-1}_{r,nT}\right),$$

where $\Sigma_{r,nT}$ in (B.3) is the information matrix, and $\Omega_{r,nT}$ in (B.4) is related to the third and fourth moments of $e_{nt,i}$ and $c_{n0,i}$ ($\Omega_{r,nT} = 0$ under normality).

2.3. *The Between Equation.* The likelihood of the random effects model can be written as a product of two likelihoods—the within likelihood from (3) and the between likelihood as shown below. The within model has $n(T-1)$ sample observations as ΔY_{nt} for $t=2,\dots,T$. Consider the remaining $Y_{n1} = z_n b_0 + \lambda_{10} W_{n1} Y_{n1} + X_{n1}\beta_0 + \mu_n + U_{n1}$. The covariance of its disturbances with those in the within model is $\mathrm{Cov}(V_{n1}, \mathbf{e}'^{d}_{n,T-1}) = -\frac{\sigma^2_{e0}}{1+\rho_0}[e'_{T-1}\otimes I_n]$, where $e_{T-1}=(1,0,\dots,0)'$. As $e'_{T-1}H^{-1}_{T-1} = [1+(T-1)\frac{1-\rho_0}{1+\rho_0}]^{-1}(T-1, T-2,\dots,1)$, we have

$$
\begin{aligned}
V_{n1} &= -\frac{\sigma^2_{e0}}{1+\rho_0}\left(e'_{T-1}\left(\sigma^2_{e0}H_{T-1}\right)^{-1}\otimes I_n\right)\mathbf{e}^d_{n,T-1} + \tilde{V}_{n1} \\
&= -\frac{1}{T-(T-2)\rho_0}\left((T-1)\Delta V_{n2} + \sum_{t=3}^{T}(T+1-t)\Delta e_{nt}\right) + \tilde{V}_{n1},
\end{aligned}
$$

with $\tilde{V}_{n1}$ being the residual vector uncorrelated with ΔV_{n2} and Δe_{nt} for $t=3,\dots,T$, $E(\tilde{V}_{n1})=0$, and

$$
\begin{aligned}
\mathrm{Var}(\tilde{V}_{n1}) &= \frac{\sigma^2_{e0}}{1-\rho_0^2}I_n - \left(\frac{\sigma^2_{e0}}{1+\rho_0}\right)^2 (e'_{T-1}\otimes I_n)\left(\left(\sigma^2_{e0}H_{T-1}\right)^{-1}\otimes I_n\right)(e'_{T-1}\otimes I_n)' \\
&= \frac{\sigma^2_{e0}}{1-\rho_0^2}I_n - \frac{\sigma^2_{e0}}{(1+\rho_0)^2}\left(e'_{T-1}H^{-1}_{T-1}e_{T-1}\right)I_n = \sigma_1^2 I_n,
\end{aligned}
\tag{10}
$$

where $\sigma_1^2 = \frac{\sigma^2_{e0}}{(1-\rho_0)(T-(T-2)\rho_0)}$. Thus, the conditional likelihood of V_{n1}, conditional on $\mathbf{e}^d_{n,T-1}$ (under normality), can be constructed. By rearranging conditioned elements, we have

$$
S_{n1}\vec{Y}_{nT} = z_n b_0 + \vec{X}_{nT}\beta_0 + \mu_n + S^{-1}_{n2}B_{n2}\vec{V}_{nT}, \tag{11}
$$

where $\vec{V}_{nT} = \tilde{V}_{n1}$ can be simplified into $\vec{V}_{nT} = [T-(T-2)\rho_0]^{-1}[V_{n1} + (1-\rho_0)\sum_{t=2}^{T-1} V_{nt} + V_{nT}]$ and similarly for $\vec{Y}_{nT}$ and $\vec{X}_{nT}$. We may interpret (11) as a "between" equation, which captures the cross-sectional variation across spatial units, because outcomes for each unit have been properly aggregated over time. When $\rho_0=0$, $\vec{Y}_{nT}=\bar{Y}_{nT}$, $\vec{X}_{nT}=\bar{X}_{nT}$, and $\vec{V}_{nT}=\bar{V}_{nT}$ are time averages, and (11) becomes $S_{n1}\bar{Y}_{nT} = z_n b_0 + \bar{X}_{nT}\beta_0 + \mu_n + S^{-1}_{n2}B_{n2}\bar{V}_{nT}$ in the familiar form. Any spatial structure on μ_n is captured in this between equation. As the within equation does not involve μ_n, identification of the spatial structure of μ_n will solely depend on the between equation. This between equation highlights the main distinction of the random components model and the within equation.

The variance matrix of the overall disturbances in (11) is

$$
\Omega_{n1} = \mathrm{Var}\left(\mu_n + S^{-1}_{n2}B_{n2}\vec{V}_{nT}\,\middle|\,\mathbf{Y}^d_{n,T-1}\right) = \sigma^2_{c0}S^{-1}_{n3}B_{n3}B'_{n3}S'^{-1}_{n3} + \sigma_1^2 S^{-1}_{n2}B_{n2}B'_{n2}S'^{-1}_{n2},
$$

and the likelihood function is $L_{b,n}(\theta) = (2\pi)^{-n/2}|\Omega_{n1}(\theta)|^{-1/2}|S_{n1}(\lambda_1)|\exp(-\frac{1}{2}\boldsymbol{\xi}'_n(\theta)\Omega^{-1}_{n1}(\theta)\boldsymbol{\xi}_n(\theta))$, where $\boldsymbol{\xi}_n(\theta) = S_{n1}(\lambda_1)\vec{Y}_{nT}(\rho) - \vec{X}_{nT}(\rho)\beta - z_n b$. Hence, the log-likelihood of the between equation is

$$
\ln L_{b,n}(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega_{n1}(\theta)| + \ln|S_{n1}(\lambda_1)| - \frac{1}{2}\boldsymbol{\xi}'_n(\theta)\Omega^{-1}_{n1}(\theta)\boldsymbol{\xi}_n(\theta), \tag{12}
$$

with its score and information matrix in Appendix B.3. For parameters' identification, b_0, β_0, λ_{10}, and ρ_0 can be identified from the main equation in (11), but the variance matrix Ω_{n1} will be the sole source for the identification of λ_{20}, λ_{30}, δ_{20}, δ_{30}, σ^2_{e0}, and σ^2_{c0}.

As Ω_{n1} is the sum of two variance matrices, i.e., those of the spatial individual effects component and the remaining disturbance, the identification of its parameters would crucially depend on distinctive structures of the two processes. For the error components in Kapoor et al. (2007) and Fingleton (2008), the two spatial processes are similar and have $S_{n3}^{-1}B_{n3} = S_{n2}^{-1}B_{n2}$. Under such a specification, Ω_{n1} will be reduced into a single piece, and it is apparent that only the sum, $\sigma_{c0}^2 + \sigma_1^2$, can be identified but not the separate parameters σ_{c0}^2 and σ_{e0}^2. However, the identification of σ_{c0}^2 in the random effects model is possible, as the identification of σ_{e0}^2 can be from the within equation. That is, σ_{e0}^2 can be identified from the within equation as well as the random effects model but may not be identified solely from the between equation. Such a kind of irregularity will have implications on the Hausman-type specification test (Section 3.1) for random versus fixed effects specifications. For the general model with two different spatial processes, one for the individual effects and the other for the remaining disturbance, due to the implied complicated variance in Ω_{n1}, those parameters might not be easily estimated from the between equation, which is essentially a cross-sectional spatial one.

The relative complexity under different specifications might be seen from matrix expansions of Ω_{n1} below with some heuristic arguments. We shall consider the specifications without the SMA in disturbances in order to understand the distinction, in particular the Anselin (1988) versus Kapoor et al. (2007) specifications via the between equation (in the following, we denote the two specifications as Anselin and KKP, respectively). Without SMA disturbances, $\Omega_{n1} = \sigma_{c0}^2 S_{n3}^{-1} S_{n3}'^{-1} + \sigma_1^2 S_{n2}^{-1} S_{n2}'^{-1}$.

(i) **General random specification with $W_{n3} = W_{n2} (= W_n)$:** From Ω_{n1}, all the four parameters λ_{20}, λ_{30}, σ_{c0}^2, and σ_1^2 can be identified as long as $\lambda_2 \neq \lambda_3$, but its nonlinearity might reveal a "weak" identification scenario. Assume that $|\lambda_j| < 1$ for $j = 2, 3$ and W_n is row-normalized. Thus, inverse of $S_{nj}(\lambda_j)$ can be expanded into $S_{nj}^{-1}(\lambda_j) = I_n + \lambda_j W_n + \lambda_j^2 W_n^2 + \lambda_j^3 W_n^3 + O(\lambda_j^4)$. It follows that

$$\begin{aligned}\Omega_{n1}(\theta) = {} & (\sigma_c^2 + \sigma_1^2) I_n + (\sigma_c^2 \lambda_3 + \sigma_1^2 \lambda_2)(W_n + W_n') + (\sigma_c^2 \lambda_3^2 + \sigma_1^2 \lambda_2^2)(W_n^2 + W_n'^2 + W_n W_n') \\ & + (\sigma_c^2 \lambda_3^3 + \sigma_1^2 \lambda_2^3)(W_n^3 + W_n'^3 + W_n W_n'^2 + W_n^2 W_n') + O((|\lambda_3| + |\lambda_2|)^4).\end{aligned}$$

Because we have four parameters in $(\lambda_2, \lambda_3, \sigma_c^2, \sigma_1^2)$, at least the coefficients of four leading terms are needed in order to identify them. However, if values λ_2 and λ_3 are small, high-order coefficients of those leading terms would also be small, and these parameters would be intuitively difficult to estimate. In addition, there is difficulty in the identification if λ_2 and λ_3 are close to each other because the above expansion can be rewritten as

$$\begin{aligned}\Omega_{n1}(\theta) = {} & \sigma_*^2 I_n + (\sigma_c^2(\lambda_3 - \lambda_2) + \sigma_*^2 \lambda_2)(W_n + W_n') + (\sigma_c^2(\lambda_3^2 - \lambda_2^2) + \sigma_*^2 \lambda_2^2)(W_n^2 + W_n'^2 + W_n W_n') \\ & + (\sigma_c^2(\lambda_3^3 - \lambda_2^3) + \sigma_*^2 \lambda_2^3)(W_n^3 + W_n'^3 + W_n W_n'^2 + W_n^2 W_n') + O((|\lambda_3| + |\lambda_2|)^4),\end{aligned}$$

where $\sigma_*^2 = \sigma_c^2 + \sigma_1^2$. When the difference of λ_2 and λ_3 is small, σ_c^2 would not be easily estimated, and so would be σ_1^2 in consequence. From these, the estimates of λ_3 and σ_c^2 could be possible in the random effects likelihood simply due to the fact that λ_2 and σ_1^2 can be consistently estimated from the within equation.

(ii) **General random specification with $W_{n3} \neq W_{n2}$:** For such a general case, we have

$$\begin{aligned}\Omega_{n1}(\theta) = {} & (\sigma_c^2 + \sigma_1^2) I_n + \sigma_c^2 \lambda_3 (W_{n3} + W_{n3}') + \sigma_1^2 \lambda_2 (W_{n2} + W_{n2}') + \sigma_c^2 \lambda_3^2 (W_{n3}^2 + W_{n3}'^2 + W_{n3} W_{n3}') \\ & + \sigma_1^2 \lambda_2^2 (W_{n2}^2 + W_{n2}'^2 + W_{n2} W_{n2}') + O((|\lambda_3| + |\lambda_2|)^3).\end{aligned}$$

Here, the four parameters can be identified from the expansion up to the second order. Thus, when W_{n2} and W_{n3} can be distinguished from each other, the estimation of the between equation is easier.

(iii) **Anselin's specification:** For this specification, $\Omega_{n1}(\theta) = \sigma_c^2 I_n + \sigma_1^2 S_{n2}^{-1}(\lambda_2) S_{n2}'^{-1}(\lambda_2)$ can be written as

$$\Omega_{n1}(\theta) = (\sigma_c^2 + \sigma_1^2) I_n + (\sigma_c^2\lambda_3 + \sigma_1^2\lambda_2)(W_n + W_n') + (\sigma_c^2\lambda_3^2 + \sigma_1^2\lambda_2^2)(W_n^2 + W_n'^2 + W_n W_n') + (\sigma_c^2\lambda_3^3 + \sigma_1^2\lambda_2^3)(W_n^3 + W_n'^3 + W_n W_n'^2 + W_n^2 W_n') + O((|\lambda_3| + |\lambda_2|)^4).$$

Hence, the expansion is needed for up to the second order only.

(iv) **KKP's specification:** Here, we have $\lambda_2 = \lambda_3$ and $\Omega_{n1}(\theta) = \sigma_*^2 I_n + \sigma_*^2 \lambda_2 (W_n + W_n') + O(\lambda_2^2)$. Hence, λ_2 and σ_*^2 can be identified from the expansion up to the first order (although σ_{c0}^2 and σ_1^2 cannot be separately identified). Therefore, the between equation of the KKP model would be easier to be estimated than Anselin's specification and generalized ones.

2.4. *Pooling of Estimates*. The likelihood function of the random effects model has the whole parameter vector $\theta_0 = (\theta_{10}', \theta_{20}')'$, but that of the fixed effects model has only the subvector θ_{10}. The excluded parameters in θ_{20} would appear in the likelihood function of the between equation. Without loss of generality, assume that θ_{10} can be identified from both the within and between equations.[22] In order to compare the efficiency of estimates of the two models, one simple approach is to use the concentrated likelihood function $L_{r,nT}^c(\theta_1)$ of the random effects model with θ_2 concentrated out and compare it with $L_{w,nT}(\theta_1)$ of the within equation. Similarly, one can have the concentrated likelihood $L_{b,n}^c(\theta_1)$ of the between equation. Those concentrated likelihood functions of the random effects model and the between equation have the same common θ_1 as that of the within equation. The random effects estimate of θ_{10} can be interpreted as an asymptotically weighted average of the within and between estimates, closely analogous to Maddala (1971) for the panel regression model as shown below.

For the case where T is finite, the within estimate $\hat{\theta}_{w1}$ would be $\sqrt{n}$-consistent and $\sqrt{n}(\hat{\theta}_{w1} - \theta_{10}) = (-\frac{1}{n}\frac{\partial^2 \ln L_{w,nT}(\theta_{10})}{\partial\theta_1\partial\theta_1'})^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_{w,nT}(\theta_{10})}{\partial\theta_1} + o_p(1)$, and the between estimate $\hat{\theta}_{b1}$ would have $\sqrt{n}(\hat{\theta}_{b1} - \theta_{10}) = (-\frac{1}{n}\frac{\partial^2 \ln L_{b,n}^c(\theta_{10})}{\partial\theta_1\partial\theta_1'})^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_{b,n}^c(\theta_{10})}{\partial\theta_1} + o_p(1)$. On the other hand, the ML estimate of the random components model has $\sqrt{n}(\hat{\theta}_{r1} - \theta_{10}) = (-\frac{1}{n}\frac{\partial^2 L_{r,nT}^c(\theta_{10})}{\partial\theta_1\partial\theta_1'})^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_{r,nT}^c(\theta_{10})}{\partial\theta_1} + o_p(1)$. Because the nT sample observations in the random components model consist of $n(T-1)$ sample observations in the within equation and n sample observations in the between equation, we have the likelihood decomposition

$$L_{r,nT}^c(\theta_1) = L_{w,nT}(\theta_1) L_{b,n}^c(\theta_1).$$

Thus,

$$\frac{1}{\sqrt{n}}\frac{\partial \ln L_{r,nT}^c(\theta_1)}{\partial\theta_1} = \frac{1}{\sqrt{n}}\frac{\partial \ln L_{w,nT}(\theta_1)}{\partial\theta_1} + \frac{1}{\sqrt{n}}\frac{\partial \ln L_{b,n}^c(\theta_1)}{\partial\theta_1}$$

and

$$\frac{1}{n}\frac{\partial^2 \ln L_{r,nT}^c(\theta_1)}{\partial\theta_1\partial\theta_1'} = \frac{1}{n}\frac{\partial^2 \ln L_{w,nT}(\theta_1)}{\partial\theta_1\partial\theta_1'} + \frac{1}{n}\frac{\partial^2 \ln L_{b,n}^c(\theta_1)}{\partial\theta_1\partial\theta_1'}.$$

Hence,

$$\sqrt{n}(\hat{\theta}_{r1} - \theta_{10}) = A_{nT,1}\sqrt{n}(\hat{\theta}_{w1} - \theta_{10}) + A_{nT,2}\sqrt{n}(\hat{\theta}_{b1} - \theta_{10}) + o_p(1), \quad (13)$$

[22] If some of them can only be identified and estimated in one equation but not the other, we may consider the subset of common parameters that can be identified in both equations. In that case, relevant concentrated within and between likelihood functions will be used instead.

where

$$A_{nT,1} = \left(\frac{1}{n}\frac{\partial^2 \ln L^c_{r,nT}(\theta_{10})}{\partial\theta_1\partial\theta_1'}\right)^{-1} \frac{1}{n}\frac{\partial^2 \ln L_{w,nT}(\theta_{10})}{\partial\theta_1\partial\theta_1'}$$

and

$$A_{nT,2} = \left(\frac{1}{n}\frac{\partial^2 \ln L^c_{r,nT}(\theta_{10})}{\partial\theta_1\partial\theta_1'}\right)^{-1} \frac{1}{n}\frac{\partial^2 \ln L^c_{b,n}(\theta_{10})}{\partial\theta_1\partial\theta_1'}$$

are weights because $A_{nT,1} + A_{nT,2} = I_{k_{\theta_1}}$ with k_{θ_1} being the dimension of θ_1. Hence, the random effects estimate of θ_{10} can be interpreted as pooling within and between estimates. We note that the weighting above is valid even though those likelihood functions are quasi ones. However, when the likelihoods are only quasi-likelihoods, $A_{nT,1}$ and $A_{nT,2}$ are not necessarily interpreted as ratios of precision matrices of the within and between estimates relative to that of the random effects estimate.

From the relation of second-order derivatives of a concentrated log-likelihood function with those of the original log-likelihood (see, e.g., Amemiya, 1985), an alternative expression is[23]

$$\sqrt{n}(\hat{\theta}_{r1} - \theta_{10}) = B_{nT,1}\sqrt{n}(\hat{\theta}_{w1} - \theta_{10}) + B_{nT,2}\sqrt{n}(\hat{\theta}_{b1} - \theta_{10}) + o_p(1), \tag{14}$$

where

$$B_{nT,1} = \left[\frac{\partial^2 \ln L_{r,nT}}{\partial\theta_1\partial\theta_1'} - \frac{\partial^2 \ln L_{r,nT}}{\partial\theta_1\partial\theta_2'}\left(\frac{\partial^2 \ln L_{r,nT}}{\partial\theta_2\partial\theta_2'}\right)^{-1}\frac{\partial^2 \ln L_{r,nT}}{\partial\theta_2\partial\theta_1'}\right]^{-1}\left[\frac{\partial^2 \ln L_{w,nT}}{\partial\theta_1\partial\theta_1'}\right]$$

and

$$B_{nT,2} = \left[\frac{\partial^2 \ln L_{r,nT}}{\partial\theta_1\partial\theta_1'} - \frac{\partial^2 \ln L_{r,nT}}{\partial\theta_1\partial\theta_2'}\left(\frac{\partial^2 \ln L_{r,nT}}{\partial\theta_2\partial\theta_2'}\right)^{-1}\frac{\partial^2 \ln L_{r,nT}}{\partial\theta_2\partial\theta_1'}\right]^{-1} \times \left[\frac{\partial^2 \ln L_{b,n}}{\partial\theta_1\partial\theta_1'} - \frac{\partial^2 \ln L_{b,n}}{\partial\theta_1\partial\theta_2'}\left(\frac{\partial^2 \ln L_{b,n}}{\partial\theta_2\partial\theta_2'}\right)^{-1}\frac{\partial^2 \ln L_{b,n}}{\partial\theta_2\partial\theta_1'}\right].$$

3. TESTING

3.1. *The Hausman Specification Test*.

3.1.1. *The Hausman test under normality*. The likelihood decomposition provides a useful device for a Hausman-type test of random effects specification against the fixed effects specification where the individual effects could be correlated with exogenous regressors. Under the null hypothesis that $\mathbf{c}_{n0}$ is independent of regressors, the MLE $\hat{\theta}_{r1}$ of the random effects model is consistent and asymptotically efficient (assuming the likelihood function is correctly specified, in this case, normal disturbances). However, under the alternative that $\mathbf{c}_{n0}$ is correlated with the regressors, $\hat{\theta}_{r1}$ is inconsistent, although the within estimator $\hat{\theta}_{w1}$ is consistent under both the null and alternative hypotheses. The Hausman-type statistic is

$$n(\hat{\theta}_{r1} - \hat{\theta}_{w1})'\hat{\Omega}_n^+(\hat{\theta}_{r1} - \hat{\theta}_{w1}),$$

[23] Again, we note that this is valid even for quasi-likelihoods.

where $\hat{\Omega}_n$ is a consistent estimate of the limiting variance matrix of $\sqrt{n}(\hat{\theta}_{r1} - \hat{\theta}_{w1})$ under the null hypothesis, and $\hat{\Omega}_n^+$ is its generalized inverse. This test statistic will be asymptotically χ^2 distributed, and its degrees of freedom is the rank of the limiting matrix of Ω_n (see, e.g., Ruud, 2000).[24] Here, the rank of Ω_n needs special attention.

Suppose that B^{-1} is the limiting variance matrix of $\sqrt{n}(\hat{\theta}_{w1} - \theta_{10})$. As $\hat{\theta}_{r1}$ is asymptotically efficient relative to $\hat{\theta}_{w1}$, the limiting variance matrix of $\sqrt{n}(\hat{\theta}_{r1} - \theta_{10})$ can be written as $(B + C)^{-1}$ for some nonnegative definite matrix C. If C happens to be positive definite, $B^{-1} - (B + C)^{-1} = B^{-1}(B^{-1} + C^{-1})^{-1}B^{-1}$ is positive definite. In this case, the degrees of freedom of the χ^2 test is k_{θ_1}. However, if C is only positive semidefinite but not positive definite, the degrees of freedom could be smaller. Suppose that C is a positive semidefinite matrix of dimension k_{θ_1} with rank m, where $0 < m < k_{\theta_1}$. Let $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_{k_{\theta_1}}\}$ be the diagonal matrix of eigenvalues of C in the metric of B. That is, $B = QQ'$ and $C = Q\Lambda Q'$ for a nonsingular matrix Q, where eigenvalues in Λ are nonnegative and the number of positive eigenvalues corresponds to the rank of C (see, e.g., Proposition 62 in Dhrymes, 1978).[25] Let $\Lambda = \begin{pmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{pmatrix}$, where Λ_1 consists of the positive eigenvalues of C with m entries. Then,

$$B^{-1} - (B + C)^{-1} = (QQ')^{-1} - (QQ' + Q\Lambda Q')^{-1} = Q'^{-1}\text{diag}\left\{\frac{\lambda_1}{1+\lambda_1}, \dots, \frac{\lambda_m}{1+\lambda_m}, 0, \dots, 0\right\} Q^{-1}$$

with rank m.

As

$$\begin{pmatrix} \dfrac{\partial^2 \ln L_{r,nT}}{\partial\theta_1 \partial\theta_1'} & \dfrac{\partial^2 \ln L_{r,nT}}{\partial\theta_1 \partial\theta_2'} \\ \dfrac{\partial^2 \ln L_{r,nT}}{\partial\theta_2 \partial\theta_1'} & \dfrac{\partial^2 \ln L_{r,nT}}{\partial\theta_2 \partial\theta_2'} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial^2 \ln L_{w,nT}}{\partial\theta_1 \partial\theta_1'} + \dfrac{\partial^2 \ln L_{b,n}}{\partial\theta_1 \partial\theta_1'} & \dfrac{\partial^2 \ln L_{b,n}}{\partial\theta_1 \partial\theta_2'} \\ \dfrac{\partial^2 \ln L_{b,n}}{\partial\theta_2 \partial\theta_1'} & \dfrac{\partial^2 \ln L_{b,n}}{\partial\theta_2 \partial\theta_2'} \end{pmatrix} \tag{15}$$

from the likelihood decomposition, the asymptotic variance of $\hat{\theta}_{r1}$ from the random effects likelihood under normality would be $\{E(-\frac{\partial^2 \ln L_{w,nT}}{\partial\theta_1\partial\theta_1'}) + E(-\frac{\partial^2 \ln L_{b,n}}{\partial\theta_1\partial\theta_1'}) - E(-\frac{\partial^2 \ln L_{b,n}}{\partial\theta_1\partial\theta_2'})$ $[E(-\frac{\partial^2 \ln L_{b,n}}{\partial\theta_2\partial\theta_2'})]^{-1}E(-\frac{\partial^2 \ln L_{b,n}}{\partial\theta_2\partial\theta_1'})\}^{-1}$ and that of $\hat{\theta}_{w1}$ is $[E(-\frac{1}{n}\frac{\partial^2 \ln L_{w,nT}}{\partial\theta_1\partial\theta_1'})]^{-1}$. The matrix C would be the limit of

$$E\left(-\frac{1}{n}\frac{\partial^2 \ln L_{b,n}}{\partial\theta_1\partial\theta_1'}\right) - E\left(\frac{1}{n}\frac{\partial^2 \ln L_{b,n}}{\partial\theta_1\partial\theta_2'}\right)\left[E\left(-\frac{1}{n}\frac{\partial^2 \ln L_{b,n}}{\partial\theta_2\partial\theta_2'}\right)\right]^{-1} E\left(\frac{1}{n}\frac{\partial^2 \ln L_{b,n}}{\partial\theta_2\partial\theta_1'}\right). \tag{16}$$

Thus, the Hausman test statistic can be computed as

$$n(\hat{\theta}_{r1} - \hat{\theta}_{w1})'Q\text{diag}\left\{\frac{1+\lambda_1}{\lambda_1}, \dots, \frac{1+\lambda_m}{\lambda_m}, 0, \dots, 0\right\} Q'(\hat{\theta}_{r1} - \hat{\theta}_{w1}), \tag{17}$$

where Q is the eigenvector of C in the metric of $B = \lim_{n\to\infty} E(-\frac{1}{n}\frac{\partial^2 \ln L_{w,nT}}{\partial\theta_1\partial\theta_1'})$. In general, if θ_{10} can be identified and estimated from the between equation, C would have full rank k_{θ_1}. However, the KKP random components specification needs special attention, as a component of θ_{10} cannot be identified from the between equation.

[24] Asymptotically, this test is also equivalent to test the difference of the within and between estimates, because (13) implies that $\sqrt{n}(\hat{\theta}_{r1} - \hat{\theta}_{w1}) = A_{nT,2}\sqrt{n}(\hat{\theta}_{b1} - \hat{\theta}_{w1}) + o_p(1)$.

[25] To obtain Q, we can follow the following four steps: (i) Let R be the Cholesky decomposition of B^{-1} such that $R'R = B^{-1}$; (ii) let $A = RCR'$; (iii) find eigenvector matrix P and eigenvalues matrix Λ of A such that $P'AP = \Lambda$; (iv) let $Q = R^{-1}P$. By doing so, we see that $QQ' = R^{-1}PP'(R')^{-1} = R^{-1}(R')^{-1} = B$ and $Q\Lambda Q' = R^{-1}PP'APP'(R')^{-1} = R^{-1}A(R')^{-1} = R^{-1}RCR'(R')^{-1} = C$.

3.1.2. *The KKP model specification under normality.* The likelihood function for the KKP type model has $\lambda_2 = \lambda_3$ and $\delta_2 = \delta_3$ with $W_{n2} = W_{n3}$ and $M_{n2} = M_{n3}$ imposed. It is convenient to consider the reparameterization $\sigma_*^2 = \sigma_c^2 + \sigma_1^2$. Thus, σ_*^2 is used instead of σ_c^2 in the likelihood of the between equation. For this model, $\theta_1 = (\beta', \lambda_1, \lambda_2, \delta_2, \rho, \sigma_e^2)'$ and $\theta_2 = (b', \sigma_*^2)'$. Note that the parameter σ_e^2 is in $L_{w,nT}$ but not in $L_{b,n}$; on the other hand, σ_*^2 is in $L_{b,n}$ but not in $L_{w,nT}$. Thus, by denoting $\theta_1^* = (\beta', \lambda_1, \lambda_2, \delta_2, \rho)'$, the B matrix is the limit of

$$E\left(-\frac{1}{n}\frac{\partial^2 \ln L_{w,nT}}{\partial\theta_1^*\partial\theta_1^{*\prime}}\right) - E\left(\frac{\partial^2 \ln L_{w,nT}}{\partial\theta_1^*\partial\sigma_e^2}\right)\left[E\left(-\frac{1}{n}\frac{\partial^2 \ln L_{w,nT}}{\partial^2\sigma_e^2}\right)\right]^{-1} E\left(\frac{\partial^2 \ln L_{w,nT}}{\partial\partial\theta_1^{*\prime}}\right)$$

and $C = \lim_{n\to\infty}(E(-\frac{1}{n}\frac{\partial^2 \ln L_{b,n}}{\partial\theta_1^*\partial\theta_1^{*\prime}}) - E(\frac{\partial^2 \ln L_{b,n}}{\partial\theta_1^*\partial\theta_2'})[E(-\frac{1}{n}\frac{\partial^2 \ln L_{b,n}}{\partial\theta_2\partial\theta_2'})]^{-1}E(\frac{\partial^2 \ln L_{b,n}}{\partial\theta_2\partial\theta_1^{*\prime}}))$. The rank of this C is $(k_{\theta_1} - 1)$, the number of parameters of θ_1^*. This is intuitively understandable because there are no two separate estimates for σ_e^2 from the within and between equations to be compared with.

3.1.3. *The Hausman test under nonnormality.* When the disturbances are not normally distributed, the random effects estimate is no longer efficient; hence, the variance of the difference of $\hat\theta_{r1}$ and $\hat\theta_{w1}$ will not necessarily equal the difference of the variances of $\hat\theta_{r1}$ and $\hat\theta_{w1}$. Therefore, we need to compute the variance of $\sqrt{n}(\hat\theta_{r1} - \hat\theta_{w1})$ explicitly and set up a corresponding Wald-type robust test (Arellano, 1993). By using the relation $\ln L_{r,nT} = \ln L_{w,nT} + \ln L_{b,n}$, as is derived in Appendix C, $\sqrt{n}(\hat\theta_{r1} - \hat\theta_{w1}) = \frac{1}{\sqrt{n}}A'_{nT}(\frac{\partial \ln L_{w,nT}}{\partial\theta_1'}, \frac{\partial \ln L_{b,n}}{\partial\theta'})' + o_P(1)$, where $A'_{nT} = [(\Sigma_{w,nT} + C)^{-1} - \Sigma_{w,nT}^{-1},\ J'\Sigma_{r,nT}^{-1}]$ and $J = (I_{k_{\theta_1}}, \mathbf{0}_{k_{\theta_1}\times k_{\theta_2}})'$. Hence, when the disturbances are not normal, the asymptotic variance of $\sqrt{n}(\hat\theta_{r1} - \hat\theta_{w1})$ would be Ω_{nT}, where $\Omega_{nT} = (B^{-1} - (B + C)^{-1}) + A'_{nT}[\begin{smallmatrix}\Omega_{w,nT} & \Omega_{wb,nT}\\ \Omega'_{wb,nT} & \Omega_{b,n}\end{smallmatrix}]A_{nT}$. Here, $\Omega_{b,n}$ in (B.6) is the part due to the third and fourth moments of disturbances for the variance matrix of the score of the between equation, and $\Omega_{wb,nT}$ in (B.6) is for the covariance matrix of the scores of the within and between equations. The Hausman test statistic can be calculated as $\mathcal{H} = n(\hat\theta_{r1} - \hat\theta_{w1})'\Omega_{nT}^{+}(\hat\theta_{r1} - \hat\theta_{w1})$.

3.1.4. *The KKP model specification under nonnormality.* For the KKP model, we compare estimates of θ_1^* under the within and random specifications. Let $J^* = (I_{k_{\theta_1^*}}, \mathbf{0}_{k_{\theta_1}^*\times(k_{\theta_2}+1)})'$. As is derived in Appendix C, $\sqrt{n}(\hat\theta_{r1}^* - \hat\theta_{w1}^*) = \frac{1}{\sqrt{n}}A^{*\prime}_{nT}(\frac{\partial \ln L_{w,nT}}{\partial\theta_1'}, \frac{\partial \ln L_{b,n}}{\partial\theta'})' + o_P(1)$, where $A^{*\prime}_{nT} = [(\Sigma_{w,nT} + C)^{-1} - \Sigma_{w,nT}^{-1},\ J^{*\prime}\Sigma_{r,nT}^{-1}]$. Thus, the asymptotic variance of $\sqrt{n}(\hat\theta_{r1}^* - \hat\theta_{w1}^*)$ is[26]

$$A^{*\prime}_{nT}\left(\begin{bmatrix}\Sigma_{w,nT}^{-1} & \mathbf{0}_{k_{\theta_1^*}\times(k_{\theta_1^*}+k_{\theta_2})}\\ \mathbf{0}_{(k_{\theta_1^*}+k_{\theta_2})\times k_{\theta_1^*}} & \Sigma_{b,n}\end{bmatrix} + \begin{bmatrix}\Omega_{w,nT} & \Omega_{wb,nT}\\ \Omega'_{wb,nT} & \Omega_{b,n}\end{bmatrix}\right)A^*_{nT}.$$

3.2. *LM Test.* Mutl and Pfaffermayr (2011) consider the Hausman test of fixed effects versus random effects in the KKP spatial panel with 2SLS estimates. In order to investigate the power of the test, they specify a model similar to Mundlak (1978), where an individual effect depends on the time average of regressors. For the general spatial panel model, we can have the similar specification

$$\mathbf{c}_{n0} = \bar X_{nT}\pi_0 + \zeta_n, \tag{18}$$

where ζ_n is independent of $\bar X_{nT}$ and is assumed to be i.i.d. $N(0, \sigma_\zeta^2 I_n)$.

[26] For the KKP model, the information matrix of the between equation would be simpler than the general model, as well as the covariance matrix of the between and within scores. Its detailed formulae can be induced from the general model.

With the specification in (18), an alternative test can be based on the LM approach. The between equation will become

$$S_{n1}\vec{Y}_{nT} = z_n b_0 + \vec{X}_{nT}\beta_0 + S_{n3}^{-1}B_{n3}\bar{X}_{nT}\pi_0 + S_{n3}^{-1}B_{n3}\zeta_n + S_{n2}^{-1}B_{n2}\vec{V}_{nT}. \tag{19}$$

If $\pi_0 = 0$, this between equation would be the one in (11). This suggests that we may consider $\pi_0 = 0$ as the null hypothesis, and (19) will be useful for the construction of the LM test from the random effects log-likelihood or, equivalently, the sum of the log-likelihood functions of the within and between equations. The log-likelihood function for the between equation is $\ln L_{b,n}(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Omega_{n1}(\theta)| + \ln|S_{n1}(\lambda_1)| - \frac{1}{2}\boldsymbol{\xi}_n'(\theta)\Omega_{n1}^{-1}(\theta)\boldsymbol{\xi}_n(\theta)$, where $\boldsymbol{\xi}_n(\theta) = S_{n1}(\lambda_1)\vec{Y}_{nT}(\theta) - \vec{X}_{nT}(\theta)\beta - z_n b - S_{n3}^{-1}B_{n3}\bar{X}_{nT}\pi$ with θ extended to include π. For the random effects model, its likelihood is similar to (9), where $\boldsymbol{\xi}_{nT}(\theta) = (I_T \otimes S_{n1}(\lambda_1))\mathbf{Y}_{nT} - \mathbf{X}_{nT}\beta - l_T \otimes (z_n b + S_{n3}^{-1}B_{n3}\bar{X}_{nT}\pi)$.

The first-order derivatives of the random effects log-likelihood evaluated at the restricted $\hat{\theta}_r$ are zero except that with respect to π. The LM test statistic would be

$$\frac{\partial \ln L_{r,nT}(\hat{\theta}_r)}{\partial \pi'}\left[\left(-E\frac{\partial^2 \ln L_{r,nT}(\hat{\theta}_r)}{\partial\theta\partial\theta'}\right)^{-1}\right]_{\pi,\pi}\frac{\partial \ln L_{r,nT}(\hat{\theta}_r)}{\partial \pi}, \tag{20}$$

where $\frac{\partial \ln L_{r,nT}(\hat{\theta}_r)}{\partial \pi} = \frac{\partial \ln L_{b,n}(\hat{\theta}_r)}{\partial \pi} = (S_{n3}^{-1}(\hat{\lambda}_{3,r})B_{n3}(\hat{\delta}_{3,r})\bar{X}_{nT})'\Omega_{n1}^{-1}(\hat{\theta}_r)\boldsymbol{\xi}_n(\hat{\theta}_r)$ and $[(-E\frac{\partial^2 \ln L_{r,nT}(\hat{\theta}_r)}{\partial\theta\partial\theta'})^{-1}]_{\pi,\pi}$ is the diagonal block in $(-E\frac{\partial^2 \ln L_{r,nT}(\hat{\theta}_r)}{\partial\theta\partial\theta'})^{-1}$ corresponding to the π entry. The $E\frac{\partial^2 \ln L_{r,nT}(\hat{\theta}_r)}{\partial\theta\partial\theta'}$ would be similar to (B.3) but with additional derivatives involving π in $L_{b,n}$.

Here, one would be interested in comparing the type I error and power of this LM test with those of the Hausman-type test. The comparison is reported in Section 4 (see Tables 7 and 8).

4. MONTE CARLO

We conduct small Monte Carlo experiments to evaluate the performance of estimates under different settings. We have different data generating processes (DGPs) depending on spatial structures as well as correlated individual effects or serial correlation; correspondingly, we have different likelihoods. For notational purpose, we use, e.g., θ_{B_B}, to denote estimates where the first subscript is for the DGP and the second subscript is for the estimation method. We first focus on models without SMA structures in disturbances. We have totally 12 DGPs: A, B, K, AF, BF, KF, AI, BI, KI, AFI, BFI, and KFI, where A denotes the (restricted $\lambda_{30} = 0$) model in Anselin (1988), B the generalized random components DGP in Baltagi et al. (2007a), K the (restricted $\lambda_{20} = \lambda_{30}$) model in Kapoor et al. (2007), F the above cases with fixed individual effects, and I the cases without serial correlation. We have totally eight estimation methods: A, B, K, W, AI, BI, KI, and WI, where A, B, K denote corresponding estimates of random components models, W the within estimate, and I corresponding estimates without serial correlation.

4.1. *Estimation*.

The following items summarize the purpose of the simulation on those methods and specifications:

- Comparing the within and random approach estimates (Table 1)
  - Robustness of within estimates against those of the random components models (θ_{AF_W}, θ_{BF_W}, and θ_{KF_W} against θ_{AF_A}, θ_{BF_B}, and θ_{KF_K})
  - Efficiency of the estimates of the random components models against the within estimates (θ_{A_A}, θ_{B_B}, and θ_{K_K} against θ_{A_W}, θ_{B_W}, and θ_{K_W})
- Misspecification of serial correlation (Tables 2 and 3)

TABLE 1
COMPARING WITHIN AND RANDOM ESTIMATES

			θ_1					θ_2		
			β	λ_1	λ_2	ρ	σ_e^2	λ_3	σ_c^2	b
Robustness of within estimates										
(1)	θ_{BF_B}	Mean	1.0082	0.2089	0.1894	0.1986	0.9919	0.3069	1.1105	0.9828
		E-SD	0.0319	0.0619	0.0761	0.0380	0.0481	0.1513	0.1854	0.1930
		RMSE	0.0329	0.0626	0.0768	0.0380	0.0487	0.1776	0.2158	0.1938
	θ_{KF_K}	Mean	1.0083	0.2007	0.1980	0.2009	0.9934		1.0872	0.9952
		E-SD	0.0319	0.0605	0.0717	0.0377	0.0482		0.1780	0.1574
		RMSE	0.0329	0.0605	0.0717	0.0378	0.0487		0.1982	0.1575
	θ_{AF_A}	Mean	1.0083	0.1985	0.2052	0.2062	0.9962		1.0687	0.9992
		E-SD	0.0319	0.0572	0.0705	0.0425	0.0492		0.1895	0.1353
		RMSE	0.0329	0.0572	0.0707	0.0429	0.0493		0.2016	0.1353
	θ_{BF_W}	Mean	0.9998	0.1994	0.2015	0.2010	0.9928			
		E-SD	0.0319	0.0619	0.0743	0.0376	0.0482			
		RMSE	0.0319	0.0619	0.0743	0.0377	0.0488			
Efficiency of random estimates										
(2)	θ_{B_B}	Mean	1.0002	0.2080	0.1898	0.1977	0.9913	0.3026	1.0160	0.9825
		E-SD	0.0318	0.0615	0.0771	0.0373	0.0483	0.1460	0.1756	0.1880
		RMSE	0.0318	0.0620	0.0778	0.0374	0.0491	0.1755	0.1764	0.1888
	θ_{K_K}	Mean	1.0000	0.1995	0.1992	0.2009	0.9932		0.9869	0.9955
		E-SD	0.0317	0.0612	0.0723	0.0377	0.0483		0.1632	0.1532
		RMSE	0.0317	0.0612	0.0723	0.0377	0.0487		0.1637	0.1533
	θ_{A_A}	Mean	0.9999	0.1962	0.2077	0.2065	0.9962		0.9681	1.0012
		E-SD	0.0318	0.0583	0.0711	0.0416	0.0490		0.1709	0.1326
		RMSE	0.0318	0.0584	0.0715	0.0421	0.0491		0.1739	0.1326
	θ_{B_W}	Mean	0.9998	0.1994	0.2015	0.2010	0.9928			
		E-SD	0.0319	0.0619	0.0743	0.0376	0.0482			
		RMSE	0.0319	0.0619	0.0743	0.0377	0.0488			

NOTE: (i) The DGPs for the three models considered (A, B, and K) differ only on the specification of individual effects, which are wiped out by the first difference. Therefore, the within estimates for all three models are the same.
(ii) For DGP of B or BF, $(\beta_0, \lambda_{10}, \lambda_{20}, \rho_0, \sigma_{e0}^2, \lambda_{30}, \sigma_{c0}^2, b_0) = (1, 0.2, 0.2, 0.2, 1, 0.4, 1, 1)$.
(iii) For DGP of K or KF, $(\beta_0, \lambda_{10}, \lambda_{20}, \rho_0, \sigma_{e0}^2, \lambda_{30}, \sigma_{c0}^2, b_0) = (1, 0.2, 0.2, 0.2, 1, 0.2, 1, 1)$.
(iv) For DGP of A or AF, $(\beta_0, \lambda_{10}, \lambda_{20}, \rho_0, \sigma_{e0}^2, \lambda_{30}, \sigma_{c0}^2, b_0) = (1, 0.2, 0.2, 0.2, 1, 0, 1, 1)$.

  - Efficiency of correct restrictions on serial correlation ($\theta_{AI\_AI}, \theta_{BI\_BI}, \theta_{KI\_KI}$, and θ_{BFI_WI} against $\theta_{AI\_A}, \theta_{BI\_B}, \theta_{KI\_K}$, and θ_{BFI_W})
  - Misspecification of ignoring serial correlation ($\theta_{A\_A}, \theta_{B\_B}, \theta_{K\_K}$, and θ_{BF_W} against θ_{A_AI}, $\theta_{B\_BI}, \theta_{K\_KI}$, and θ_{BF_WI})

- Comparing random components models (Tables 4 and 5)
  - Sensitivity of random components misspecifications (θ_{B_B} against θ_{B_K} and θ_{B_A}; θ_{K_K} against θ_{K_A}; θ_{A_A} against θ_{A_K})
  - Efficiency of correctly restricted random components specifications (θ_{K_K} against θ_{K_B}; θ_{A_A} against θ_{A_B})

We first generate samples from (1), where the exogenous variable is from a standard normal distribution, and we include an intercept term in the model. The disturbances e_{nt} are generated from independent standard normal distribution ($\sigma_{e0}^2 = 1$), and $W_{n1} = W_{n2} = W_{n3}$ are all rook matrices. For the DGPs of Baltagi et al. (2007a), we have $\lambda_{10} = \lambda_{20} = 0.2$ and $\lambda_{30} = 0.4$; for the DGPs of Kapoor et al. (2007), $\lambda_{10} = \lambda_{20} = \lambda_{30} = 0.2$; for the DGPs of Anselin (1988), $\lambda_{10} = \lambda_{20} = 0.2$ and $\lambda_{30} = 0$. Under all these cases, $b_0 = 1$, $\beta_0 = 1$, and $\rho_0 = 0.2$ when serial correlation is present. Under the random effects specification, individual effects are generated from a standard normal distribution ($\sigma_{c0}^2 = 1$) independent of $\mathbf{X}_{nT}$. Under the fixed effects model, individual effects are generated from a standard normal distribution plus the time average of

TABLE 2
SERIAL CORRELATION SPECIFICATIONS: ESTIMATES

			θ_1					θ_2		
			β	λ_1	λ_2	ρ	σ_e^2	λ_3	σ_c^2	b
Efficiency of correct specification										
(1)	θ_{BI_B}	Mean	0.9995	0.1988	0.2034	0.0034	0.9942	0.3658	0.9778	0.9934
		E-SD	0.0329	0.0649	0.0782	0.0381	0.0485	0.1470	0.1572	0.1891
		RMSE	0.0329	0.0649	0.0783	0.0383	0.0489	0.1509	0.1587	0.1893
	θ_{BI_BI}	Mean	0.9995	0.1951	0.2078		0.9945	0.3907	0.9677	0.9980
		E-SD	0.0329	0.0632	0.0753		0.0476	0.1417	0.1517	0.1879
		RMSE	0.0329	0.0634	0.0757		0.0479	0.1420	0.1551	0.1879
(2)	θ_{KI_K}	Mean	0.9999	0.1997	0.1987	0.0005	0.9929		0.9899	0.9951
		E-SD	0.0328	0.0633	0.0736	0.0373	0.0481		0.1572	0.1530
		RMSE	0.0328	0.0633	0.0736	0.0373	0.0487		0.1576	0.1530
	θ_{KI_KI}	Mean	0.9999	0.1999	0.1987		0.9939		0.9882	0.9947
		E-SD	0.0328	0.0631	0.0738		0.0478		0.1543	0.1523
		RMSE	0.0328	0.0631	0.0738		0.0481		0.1547	0.1523
(3)	θ_{AI_A}	Mean	0.9998	0.1962	0.2047	0.0011	0.9931		0.9880	1.0009
		E-SD	0.0328	0.0594	0.0728	0.0369	0.0482		0.1561	0.1309
		RMSE	0.0328	0.0595	0.0730	0.0369	0.0487		0.1566	0.1309
	θ_{AI_AI}	Mean	0.9998	0.1963	0.2044		0.9938		0.9891	1.0007
		E-SD	0.0328	0.0595	0.0730		0.0479		0.1545	0.1309
		RMSE	0.0328	0.0596	0.0732		0.0483		0.1549	0.1309
(4)	θ_{BFI_W}	Mean	0.9997	0.1998	0.2010	0.0009	0.9925			
		E-SD	0.0330	0.0641	0.0764	0.0365	0.0481			
		RMSE	0.0330	0.0641	0.0764	0.0366	0.0487			
	θ_{BFI_WI}	Mean	0.9997	0.1998	0.2010		0.9934			
		E-SD	0.0330	0.0641	0.0763		0.0478			
		RMSE	0.0330	0.0641	0.0763		0.0482			
Misspecification of serial correlation										
(5)	θ_{B_B}	Mean	1.0002	0.2080	0.1898	0.1977	0.9913	0.3026	1.0160	0.9825
		E-SD	0.0318	0.0615	0.0771	0.0373	0.0483	0.1460	0.1756	0.1880
		RMSE	0.0318	0.0620	0.0778	0.0374	0.0491	0.1755	0.1764	0.1888
	θ_{B_BI}	Mean	0.9996	0.1951	0.2080		0.9857	0.3820	1.0183	0.9988
		E-SD	0.0327	0.0640	0.0768		0.0486	0.1447	0.1612	0.1932
		RMSE	0.0327	0.0642	0.0772		0.0506	0.1458	0.1622	0.1932
(6)	θ_{K_K}	Mean	1.0000	0.1995	0.1992	0.2009	0.9932		0.9869	0.9955
		E-SD	0.0317	0.0612	0.0723	0.0377	0.0483		0.1632	0.1532
		RMSE	0.0317	0.0612	0.0723	0.0377	0.0487		0.1637	0.1533
	θ_{K_KI}	Mean	1.0001	0.1997	0.1992		0.9851		1.0381	0.9953
		E-SD	0.0326	0.0632	0.0743		0.0485		0.1624	0.1550
		RMSE	0.0326	0.0632	0.0743		0.0507		0.1668	0.1550
(7)	θ_{A_A}	Mean	0.9999	0.1962	0.2077	0.2065	0.9962		0.9681	1.0012
		E-SD	0.0318	0.0583	0.0711	0.0416	0.0490		0.1709	0.1326
		RMSE	0.0318	0.0584	0.0715	0.0421	0.0491		0.1739	0.1326
	θ_{A_AI}	Mean	1.0000	0.1976	0.2034		0.9851		1.0401	0.9993
		E-SD	0.0327	0.0593	0.0734		0.0486		0.1626	0.1334
		RMSE	0.0327	0.0593	0.0734		0.0508		0.1675	0.1334
(8)	θ_{BF_W}	Mean	0.9998	0.1994	0.2015	0.2010	0.9928			
		E-SD	0.0319	0.0619	0.0743	0.0376	0.0482			
		RMSE	0.0319	0.0619	0.0743	0.0377	0.0488			
	θ_{BF_WI}	Mean	0.9999	0.1996	0.2015		0.9846			
		E-SD	0.0329	0.0642	0.0767		0.0485			
		RMSE	0.0329	0.0642	0.0768		0.0508			

NOTE: (i) The DGPs for the three models considered (A, B, and K) differ only on the specification of individual effects, which are wiped out by the first difference. Therefore, the within estimates for all three models are the same.
(ii) The value of $(\beta_0, \lambda_{10}, \lambda_{20}, \rho_0, \sigma_{e0}^2, \lambda_{30}, \sigma_{c0}^2, b_0)$ are the same as Table 1.

TABLE 3
SERIAL CORRELATION SPECIFICATIONS: LR TESTS

(1)	The type I error under different significance levels			
	Significance level	0.01	0.05	0.1
	LR test with the B random effects specification	0.017	0.073	0.132
(2)	The type I error under different significance levels			
	Significance level	0.01	0.05	0.1
	LR test with the K random effects	0.006	0.045	0.087
(3)	The type I error under different significance levels			
	Significance level	0.01	0.05	0.1
	LR test with the A random effects	0.006	0.044	0.094
(4)	The type I error under different significance levels			
	Significance level	0.01	0.05	0.1
	LR test with the fixed effects model	0.005	0.046	0.092
(5)	The power under different significance levels			
	Significance level	0.01	0.05	0.1
	LR test with the B random effects specification with serial correlation	0.991	0.997	0.998
(6)	The power under different significance levels			
	Significance level	0.01	0.05	0.1
	LR test with the K random effects with serial correlation	0.996	1.00	1.00
(7)	The power under different significance levels			
	Significance level	0.01	0.05	0.1
	LR test with the A random effects with serial correlation	0.991	0.999	1.00
(8)	The power under different significance levels			
	Significance level	0.01	0.05	0.1
	LR test with the fixed effects model with serial correlation	0.996	0.999	1.00

$\mathbf{X}_{nT}$. We use $n = 100$, $T = 10$, and the repetition is 1,000. We report the mean (Mean), empirical standard deviation (E-SD), and corresponding root mean square error (RMSE).

From item (1) in Table 1, we see that the within estimate is robust to different DGPs; given the fixed effects DGPs, random effects estimates have a larger bias in β. From item (2) where the DGP is under the random effects specification, we see that random effects estimates are more efficient than the within estimates for β, λ_1, and λ_2 (but not always for ρ and σ_e^2).[27]

From items (1)–(4) in Table 2, correctly restricted estimates without serial correlation are more efficient on average than unrestricted ones, both for random effects and within estimates; however, the improvement in efficiency is not obvious. From items (5)–(8), the RMSE is higher for estimates under misspecifications, and biases of estimates of σ_e^2 are larger for all the cases. From Table 3 for the likelihood ratio (LR) test of serial correlation, we see that the type I error is close to the theoretical value and the power of the LR test is high.

From items (1)–(3) in Table 4, we see that MSEs are larger with the misspecification of random components; however, it is not apparent for item (3). From items (4) and (5), we see that correctly restricted random components models yield more efficient estimates; however, it is not necessarily for σ_e^2 and σ_c^2 in item (5). From Table 5 for the random specification test, we see that the type I error is close to the theoretical value. The power is moderate for testing the Anselin specification against the generalized Baltagi specification, but the power is low to detect the misspecification of the KKP specification against the generalized Baltagi specification unless the λ_{30} takes a large value.[28]

[27] From (13), the between estimate is important for the efficiency of random effects estimates versus the fixed effects estimates. The difficulty for the estimates of ρ_0 and σ_{0e}^2 might be due to their hidden roles in the between equation. Similar to the classical linear panel data model, the role of the information in the between equation would be more important when T is smaller, the variance of individual effects is smaller, and the cross-variation of $\vec{X}_{nT}$ is larger. In subsequent analysis, we consider an additional model design where it has a smaller σ_{0c}^2 and larger variation of $\vec{X}_{nT}$. The corresponding results are reported in Tables 9 and 10. Comparing the estimates from Tables 7 and 8 with those from Tables 9 and 10, the efficiency of the random effects estimates is more apparent.

[28] When we have a larger $\lambda_{30} = 0.7$, the power increases to 0.75.

TABLE 4
COMPARING RANDOM EFFECTS MODELS: ESTIMATES

			θ_1					θ_2		
			β	λ_1	λ_2	ρ	σ_e^2	λ_3	σ_c^2	b
Sensitivity of random misspecifications										
(1)	θ_{B_B}	Mean	1.0002	0.2080	0.1898	0.1977	0.9913	0.3026	1.0160	0.9825
		E-SD	0.0318	0.0615	0.0771	0.0373	0.0483	0.1460	0.1756	0.1880
		RMSE	0.0318	0.0620	0.0778	0.0374	0.0491	0.1755	0.1764	0.1888
	θ_{B_K}	Mean	1.0000	0.1994	0.2172	0.2010	0.9914		1.0345	0.9932
		E-SD	0.0318	0.0614	0.0721	0.0377	0.0482		0.1716	0.1894
		RMSE	0.0318	0.0614	0.0741	0.0377	0.0490		0.1750	0.1895
	θ_{B_A}	Mean	1.0015	0.2540	0.1556	0.2106	0.9978		1.0389	0.9253
		E-SD	0.0318	0.0598	0.0752	0.0442	0.0500		0.1932	0.1783
		RMSE	0.0318	0.0806	0.0873	0.0454	0.0500		0.1970	0.1933
(2)	θ_{K_K}	Mean	1.0000	0.1995	0.1992	0.2009	0.9932		0.9869	0.9955
		E-SD	0.0317	0.0612	0.0723	0.0377	0.0483		0.1632	0.1532
		RMSE	0.0317	0.0612	0.0723	0.0377	0.0487		0.1637	0.1533
	θ_{K_A}	Mean	1.0010	0.2229	0.1839	0.2084	0.9970		0.9740	0.9664
		E-SD	0.0317	0.0592	0.0729	0.0421	0.0492		0.1763	0.1497
		RMSE	0.0317	0.0635	0.0747	0.0430	0.0493		0.1782	0.1534
(3)	θ_{A_A}	Mean	0.9999	0.1962	0.2077	0.2065	0.9962		0.9681	1.0012
		E-SD	0.0318	0.0583	0.0711	0.0416	0.0490		0.1709	0.1326
		RMSE	0.0318	0.0584	0.0715	0.0421	0.0491		0.1739	0.1326
	θ_{A_K}	Mean	1.0000	0.1993	0.1835	0.2010	0.9951		0.9989	0.9973
		E-SD	0.0318	0.0616	0.0733	0.0377	0.0483		0.1663	0.1336
		RMSE	0.0318	0.0616	0.0751	0.0377	0.0485		0.1663	0.1337
Efficiency of correctly restricted random specification										
(4)	θ_{K_K}	Mean	1.0000	0.1995	0.1992	0.2009	0.9932		0.9869	0.9955
		E-SD	0.0317	0.0612	0.0723	0.0377	0.0483		0.1632	0.1532
		RMSE	0.0317	0.0612	0.0723	0.2044	0.0487		0.1637	0.1533
	θ_{K_B}	Mean	1.0001	0.2033	0.1972	0.2049	0.9950	0.1596	0.9676	0.9908
		E-SD	0.0318	0.0611	0.0751	0.0390	0.0483	0.1467	0.1655	0.1523
		RMSE	0.0318	0.0611	0.0751	0.0393	0.0486	0.2816	0.1686	0.1526
(5)	θ_{A_A}	Mean	0.9999	0.1962	0.2077	0.2065	0.9962		0.9681	1.0012
		E-SD	0.0318	0.0583	0.0711	0.0416	0.0490		0.1709	0.1326
		RMSE	0.0318	0.0584	0.0715	0.0421	0.0491		0.1739	0.1326
	θ_{A_B}	Mean	0.9998	0.1998	0.2018	0.2093	0.9972	0.0049	0.9460	0.9968
		E-SD	0.0318	0.0625	0.0753	0.0413	0.0489	0.1692	0.1627	0.1358
		RMSE	0.0318	0.0625	0.0753	0.0423	0.0490	0.4298	0.1714	0.1358

NOTE: The value of $(\beta_0, \lambda_{10}, \lambda_{20}, \rho_0, \sigma_{e0}^2, \lambda_{30}, \sigma_{c0}^2, b_0)$ are the same as Table 1.

TABLE 5
COMPARING RANDOM EFFECTS MODELS: LR TESTS

(a)	The power under different significance levels			
	Significance level	0.01	0.05	0.1
	LR test for model K against B	0.051	0.140	0.199
(b)	The power under different significance levels			
	Significance level	0.01	0.05	0.1
	LR test for model A against B	0.352	0.585	0.690
(c)	The type I error under different significance levels			
	Significance level	0.01	0.05	0.1
	LR test for model K against B	0.010	0.036	0.074
(d)	The type I error under different significance levels			
	Significance level	0.01	0.05	0.1
	LR test for model A against B	0.032	0.083	0.129

NOTE: (i) The powers in (a) and (b) are obtained from item (1) in Table 4.
(ii) The type I errors in (c) and (d) are obtained from items (4) and (5) in Table 4, respectively.

We then investigate models with SMA errors. We have three DGPs with $M_{n2} = M_{n3} = W_{n1}$: (i) the general random model with $\lambda_{10} = \lambda_{20} = 0.2$, $\lambda_{30} = 0.4$, and $\delta_{20} = \delta_{30} = 0.3$; (ii) the KKP random model with $\lambda_{10} = \lambda_{20} = \lambda_{30} = 0.2$ and $\delta_{20} = \delta_{30} = 0.3$; and (iii) the Anselin random model with $\lambda_{10} = \lambda_{20} = 0.2$, $\lambda_{30} = 0$, $\delta_{20} = 0.3$, and $\delta_{30} = 0$. We use $n = 100$ and $T = 10$ where the repetition is 1,000. Similar to previous simulations, we assign $b_0 = 1$ for the constant term in the DGP and $\beta_0 = 1$ for the X_{nt}. For each DGP, we report corresponding random estimates with both SAR and SMA disturbances, with SAR disturbances only, and those with relevant SMA disturbances only. Also, within estimates for the three DGPs are reported.

From Table 6, when the true DGP is a general random effects model with both SAR and SMA disturbances, the random effects estimate in item (1) has small bias for the common parameter θ_1, although it has some bias for λ_3 and δ_3. For the within estimate in item (4), we see that the bias is small. When we have the model misspecification where the relevant SMA disturbances are omitted, it results in an upward bias for the estimate of λ_2 for both random and within estimates; when we have the model misspecification where the relevant SAR disturbances are omitted, it results in an upward bias for the estimate of δ_2 for both random and within estimates. Similar results are found for both the KKP and Anselin models in items (2) and (3).[29]

4.2. *Testing*. We also conduct simulations to check the performance of the Hausman and LM tests for the KKP and Anselin DGPs. We use $n = 100$ and $T = 5$, where the number of repetitions is 1,000. We assign $b_0 = 1$ for the intercept term in the DGP and $\beta_0 = 1$ for X_{nt} as before. The spatial effects coefficients are $\lambda_{10} = \lambda_{20} = \lambda_{30} = 0.3$ for the KKP model, and $\lambda_{10} = \lambda_{20} = 0.3$ with $\lambda_{30} = 0$ for the Anselin model. For the KKP model in Table 7, we first use the random effects DGP and obtain the within, random, and between equations estimates. From the within and/or random effects estimates, we obtain the LM and Hausman tests and their type I errors. We then use fixed effects DGP where individual effects are generated by (18) with standard normally distributed ζ_n and $\pi_0 = \sqrt{5}$. From corresponding within and random estimates, we obtain the LM and Hausman tests and their power. Similarly, we obtain those estimates and test statistics for the Anselin model in Table 8. We also estimate the between equation in order to investigate the quality of the between estimate, which might provide us an intuitive view on the performance of these tests because the Hausman test, by comparing random and within estimates, is asymptotically equivalent to comparison of within and between estimates.

From Table 7 for the KKP model, we see that within estimates are unbiased under both DGPs of random and fixed effects. When the true DGP is a random effects model, the random effects estimate has a smaller bias, a smaller (empirical) standard deviation, and thus a smaller RMSE; also, biases of between estimates are not large, except for that of ρ_0. The type I errors for the LM and Hausman tests are close to the theoretical values at the conventional 1%, 5%, and 10% levels of significance. When the true DGP is a fixed effects model, the random effects estimate has a larger bias and a larger RMSE, and the estimate of σ_{c0}^2 is not reliable; also, the between estimate has large bias, large standard deviation, and large RMSE. The powers of the LM and Hausman tests at the three conventional levels of significance are 1.

From Table 8 for the Anselin model, we have the same within estimate as the KKP model because underlying processes are different only in the specification of individual effects (which is eliminated in the within equation). When the true DGP is a random effects model, the random effects estimate has a smaller (empirical) standard deviation and a smaller RMSE, but, the between estimate has a large bias, a large standard deviation, and a large RMSE, especially for variance parameters σ_c^2 and σ_e^2 (for comparison, although in the KKP model in Table 7,

[29] We see that estimates of λ_2 and δ_2 have large standard deviations for the general model under both random and fixed effects DGPs. This implies difficulty in the numerical search for $(\lambda_{20}, \delta_{20})$ as $B_{n2}^{-1}(\delta_2)S_{n2}(\lambda_2)$ might not be easily separated especially when $M_{n2} = W_{n2}$. When the sample size increases along with a larger λ_{20} and δ_{20}, the empirical standard deviations of estimates for λ_{20} and δ_{20} would be smaller from our Monte Carlo results (which are not reported in the table to save space).

TABLE 6
SPATIAL PANELS WITH SMA ERRORS

		θ_1						θ_2			
		β	λ_1	λ_2	δ_2	ρ	σ_e^2	λ_3	δ_3	σ_c^2	b
(1) DGP is a general random model with SMA errors											
θ_{B_B}	Mean	1.0057	0.2358	0.1962	0.2476	0.1912	0.9800	0.1398	0.3322	1.0980	0.9462
	E-SD	0.0326	0.0717	0.2208	0.2150	0.0418	0.0683	0.2652	0.2891	0.2213	0.2360
	RMSE	0.0331	0.0801	0.2209	0.2213	0.0427	0.0712	0.3716	0.2909	0.2421	0.2420
$\theta_{B\_B\_SAR}$	Mean	1.0064	0.2407	0.4237		0.1968	0.9281	0.4960		0.9570	0.9397
	E-SD	0.0312	0.0598	0.0694		0.0386	0.0470	0.1720		0.1859	0.2262
	RMSE	0.0319	0.0723	0.2342		0.0387	0.0859	0.1970		0.1908	0.2341
$\theta_{B\_B\_SMA}$	Mean	1.0086	0.2475		0.4285	0.1964	1.0286		0.5176	1.1278	0.9316
	E-SD	0.0312	0.0601		0.0659	0.0400	0.0552		0.1920	0.1975	0.2270
	RMSE	0.0323	0.0767		0.1444	0.0401	0.0621		0.2902	0.2353	0.2371
(2) DGP is a KKP random model with SMA errors											
θ_{K_K}	Mean	1.0011	0.2043	0.1596	0.3260	0.1997	1.0482			1.0548	0.9882
	E-SD	0.0328	0.0621	0.1684	0.1440	0.0386	0.0700			0.1880	0.1916
	RMSE	0.0329	0.0622	0.1732	0.1463	0.0386	0.0850			0.1958	0.1920
$\theta_{K\_K\_SAR}$	Mean	1.0047	0.2245	0.4403		0.2010	0.9287			0.9245	0.9632
	E-SD	0.0311	0.0501	0.0526		0.0378	0.0462			0.1547	0.1802
	RMSE	0.0314	0.0557	0.2460		0.0378	0.0850			0.1722	0.1839
$\theta_{K\_K\_SMA}$	Mean	1.0069	0.2329		0.4433	0.2009	1.0859			1.0757	0.9526
	E-SD	0.0321	0.0547		0.0547	0.0377	0.0621			0.1785	0.1828
	RMSE	0.0328	0.0639		0.1533	0.0377	0.1060			0.1939	0.1888
(3) DGP is an Anselin random model with SMA errors											
θ_{A_A}	Mean	1.0002	0.2015	0.2104	0.2773	0.1971	0.9871			1.0043	0.9955
	E-SD	0.0319	0.0596	0.1756	0.1567	0.0381	0.0625			0.1662	0.1405
	RMSE	0.0319	0.0597	0.1759	0.1584	0.0382	0.0638			0.1663	0.1405
$\theta_{A\_A\_SAR}$	Mean	1.0033	0.2169	0.4475		0.2014	0.9278			0.9888	0.9761
	E-SD	0.0311	0.0498	0.0538		0.0386	0.0465			0.1660	0.1314
	RMSE	0.0313	0.0526	0.2533		0.0386	0.0858			0.1664	0.1336
$\theta_{A\_A\_SMA}$	Mean	1.0055	0.2260		0.4503	0.2017	1.0395			0.9863	0.9649
	E-SD	0.0311	0.0525		0.0548	0.0390	0.0541			0.1635	0.1323
	RMSE	0.0316	0.0586		0.1600	0.0390	0.0670			0.1641	0.1369
(4) Within estimates for above models											
θ_{B_W}	Mean	0.9999	0.2001	0.1821	0.3118	0.2008	0.9969				
	E-SD	0.0323	0.0619	0.1762	0.1534	0.0375	0.0642				
	RMSE	0.0323	0.0619	0.1771	0.1539	0.0375	0.0643				
$\theta_{B\_W\_SAR}$	Mean	1.0041	0.2223	0.4440		0.2010	0.9278				
	E-SD	0.0314	0.0519	0.0547		0.0376	0.0463				
	RMSE	0.0317	0.0565	0.2501		0.0377	0.0857				
$\theta_{B\_W\_SMA}$	Mean	1.0065	0.2320		0.4455	0.2008	1.0368				
	E-SD	0.0314	0.0552		0.0562	0.0377	0.0541				
	RMSE	0.0321	0.0638		0.1560	0.0377	0.0654				

NOTE: (i) $\theta_{B\_B\_SAR}$ denotes estimates for models with only SAR disturbances.
(ii) $\theta_{B\_B\_SMA}$ denotes estimates for models with only SMA disturbances.
(iii) For all the DGPs, $(\beta_0, \lambda_{10}, \lambda_{20}, \delta_{20}, \rho_0, \sigma_{e0}^2, \sigma_{c0}^2, b_0) = (1, 0.2, 0.2, 0.3, 0.2, 1, 1, 1)$. For the DGP of B, $(\lambda_{30}, \delta_{30}) = (0.4, 0.3)$; for DGP of K, $(\lambda_{30}, \delta_{30}) = (0.2, 0.3)$; for DGP of A, $(\lambda_{30}, \delta_{30}) = (0, 0)$.

estimates of σ_c^2 and σ_e^2 under random effects DGP perform better).[30] The type I error for the LM test is slightly higher than the theoretical value, although that for the Hausman test is much larger than the theoretical value. When the true DGP is a fixed effects model, the random effects estimate has a larger bias, but not necessarily larger standard deviation or RMSE, although the between estimate has a large bias, a large standard deviation, and a large RMSE, and they are

[30] This can be seen from the matrix expansion of $\Omega_{n1}(\theta)$ in Section 2.3, where the estimation for the KKP model is easier than that for the Anselin model.

TABLE 7
KKP MODEL SPECIFICATION: TESTS FOR RANDOM VERSUS FIXED EFFECTS

		θ_1					θ_2		
		β	λ_1	λ_2	ρ	σ_e^2	σ_*^2	σ_c^2	b
(1) DGP with random effects									
θ_w	Mean	0.9974	0.2984	0.2936	0.3032	0.9902			
	E-SD	0.0473	0.0898	0.1109	0.0794	0.0802			
	RMSE	0.0474	0.0898	0.1111	0.0795	0.0808			
θ_r	Mean	0.9976	0.2978	0.2914	0.3022	0.9914		0.9591	1.0035
	E-SD	0.0468	0.0871	0.1039	0.0780	0.0797		0.2071	0.2021
	RMSE	0.0468	0.0871	0.1043	0.0781	0.0801		0.2111	0.2022
θ_b	Mean	0.9955	0.2835	0.2630	0.2102		1.2697		1.0265
	E-SD	0.2659	0.2448	0.2548	0.4390		0.1949		0.3952
	RMSE	0.2659	0.2587	0.2625	0.4391		0.2102		0.3961
The type I error under different significance levels									
Significance level				0.01	0.05	0.10			
LM test				0.010	0.048	0.100			
Hausman test				0.024	0.071	0.118			
(2) DGP with fixed effects									
θ_w	Mean	0.9974	0.2984	0.2936	0.3032	0.9902			
	E-SD	0.0473	0.0898	0.1109	0.0794	0.0802			
	RMSE	0.0474	0.0898	0.1111	0.0795	0.0808			
θ_r	Mean	1.0382	0.3064	0.2817	0.2994	0.9918		1.9197	0.9988
	E-SD	0.0481	0.0868	0.1062	0.0810	0.0808		0.3445	0.2476
	RMSE	0.0614	0.0871	0.1077	0.0811	0.0812		0.9821	0.2476
θ_b	Mean	3.1956	0.4605	0.0955	0.0312		1.2539		0.7719
	E-SD	0.2719	0.1236	0.1875	0.2256		0.1904		0.2266
	RMSE	2.2124	0.2883	0.2147	0.2818		0.2126		0.3216
The power under different significance levels									
Significance level				0.01	0.05	0.10			
LM test				1.00	1.00	1.00			
Hausman test				1.00	1.00	1.00			

NOTE: (i) For the KKP DGP, $(\beta_0, \lambda_{10}, \lambda_{20}, \rho_0, \sigma_{e0}^2, \sigma_{c0}^2, b_0) = (1, 0.3, 0.3, 0.3, 1, 1, 1)$.
(ii) For the between estimates θ_b, the implied parameter $\sigma_*^2 = \sigma_c^2 + \sigma_1^2$ has true value 1.3484. Also, the initial search value for θ_1 is from the within estimates.

not reliable. We see that the power of the LM and Hausman tests of the three conventional significance levels are high.

For Tables 7 and 8, we see that the LM and Hausman tests for the KKP and Anselin models are different, and the performance of the tests are better in the KKP model setting. This might be due to the weak identification of π_0 in the (extended) between equation (19) for the Anselin model. In the Anselin model, $\lambda_{30} = \delta_{30} = 0$ so that $S_{n3}^{-1} B_{n3} \bar{X}_{nT}$ is reduced to $\bar{X}_{nT}$, which is highly multicollinear with $\vec{X}_{nT}$, especially T is not small. However, in the KKP model, $\lambda_{30} = \delta_{30} = 0.3 \neq 0$ so that $\bar{X}_{nT}$ and $S_{n3}^{-1} B_{n3} \bar{X}_{nT}$ are not multicollinear. This might explain the better performance of LM tests for the KKP model than that for the Anselin model. For the Hausman test, as explained in footnote 26, the random effects estimates would be more efficient with a smaller σ_{c0}^2 and a larger cross-sectional variation of $\vec{X}_{nT}$, which might also help its performance. To confirm those possibilities, we use a different DGP, where $\sigma_{c0}^2 = 0.5$ and the exogenous variables are generated from $x_{it} = \xi_i + z_{it}$, where ξ_i and z_{it} are independent standard normal random variables. The estimates and tests based on this model design are reported in Tables 9 and 10. With a smaller σ_{c0}^2 and a larger variation of $\vec{X}_{nT}$, we see that the efficiency of the random estimates are more apparent for both KKP and Anselin models, and the performance of the Hausman test for the Anselin model is improved. However, due to the weak identification of π_0 under the Anselin model, the LM test for the KKP model still outperforms that in the Anselin model even though the Hausman's test is now good for both the KKP and Anselin models.

TABLE 8
ANSELIN MODEL SPECIFICATION: TESTS FOR RANDOM VERSUS FIXED EFFECTS

		θ_1					θ_2	
		β	λ_1	λ_2	ρ	σ_e^2	σ_c^2	b
(1) DGP with random effects								
θ_w	Mean	0.9974	0.2984	0.2936	0.3032	0.9902		
	E-SD	0.0473	0.0898	0.1109	0.0794	0.0802		
	RMSE	0.0474	0.0898	0.1111	0.0795	0.0808		
θ_r	Mean	0.9977	0.2945	0.2947	0.2986	0.9897	0.9684	1.0080
	E-SD	0.0466	0.0796	0.1053	0.0773	0.0796	0.2130	0.1692
	RMSE	0.0467	0.0798	0.1054	0.0773	0.0802	0.2154	0.1694
θ_b	Mean	0.9990	0.2124	0.1650	0.2081	3.8826	0.3116	1.1270
	E-SD	0.2712	0.2393	0.3426	0.4423	4.4705	0.3625	0.3751
	RMSE	0.2712	0.2548	0.3682	0.4517	5.3193	0.7780	0.3960
The type I error under different significance levels								
Significance level				0.01	0.05	0.10		
LM test				0.028	0.096	0.161		
Hausman test				0.064	0.114	0.160		
(2) DGP with fixed effects								
θ_w	Mean	0.9974	0.2984	0.2936	0.3032	0.9902		
	E-SD	0.0473	0.0898	0.1109	0.0794	0.0802		
	RMSE	0.0474	0.0898	0.1111	0.0795	0.0808		
θ_r	Mean	1.0388	0.3074	0.2831	0.3006	0.9938	1.9140	0.9950
	E-SD	0.0472	0.0697	0.0973	0.0752	0.0784	0.3260	0.1896
	RMSE	0.0611	0.0701	0.0987	0.0752	0.0787	0.9704	0.1896
θ_b	Mean	3.2215	0.2841	0.0930	0.0324	4.0085	0.4445	1.0238
	E-SD	0.2759	0.1432	0.3466	0.2270	2.8255	0.3734	0.2530
	RMSE	2.2386	0.1440	0.4037	0.3509	4.1273	0.6693	0.2541
The power under different significance levels								
Significance level				0.01	0.05	0.10		
LM test				1.00	1.00	1.00		
Hausman test				1.00	1.00	1.00		

NOTE: (i) For the Anselin DGP, $(\beta_0, \lambda_{10}, \lambda_{20}, \rho_0, \sigma_{e0}^2, \sigma_{c0}^2, b_0) = (1, 0.3, 0.3, 0.3, 1, 1, 1)$.
(ii) For the between estimates θ_b, the initial search value for θ_1 is from the within estimates.

4.3. *Consequences of Omitting Various Spatial and/or Serial Correlations*. Finally, we investigate more consequences of misspecifications with reference to an empirical study of a spatial panel model. In empirical applications of spatial panel models, it seems typical for investigators to impose only some limited spatial structures, e.g., Moscone et al. (2007), who study spatial dependence of heath expenditure in British counties. Moscone et al. (2007) estimate a random effects spatial panel model without serial correlation, where it includes either a spatial lag or a spatial error, but not both. It is of interest to see with a general model, where spatial lag, spatial error, and also serial correlation are present, whether a certain omission of spatial or serial correlation would have significant misspecification effects on estimates of included variables. In generating data for such a general model, we use $n = 148$ and $T = 6$ following the panel dimensions in Moscone et al. (2007). The spatial weights matrix is their contiguity matrix after row-normalization, and relevant variables are generated from standard normal distributions. The repetition is 1,000. Table 11 presents the results where the DGP is a random effects Anselin model whereas Table 12 has the results under a fixed effects DGP, where the fixed individual effects are generated by (18) with standard normally distributed ζ_n and $\pi_0 = \sqrt{6}$. Instead of only estimating a random effects model as in Moscone et al. (2007), we will investigate how the following various misspecifications would cause estimation and testing problems: misspecifications of omitting spatial lag or spatial error, misspecifications of omitting serial correlation, and misspecifications of omitting both. Under correct specification, we see that both random effects

TABLE 9
KKP MODEL SPECIFICATION: TESTS FOR RANDOM VERSUS FIXED EFFECTS

		θ_1					θ_2		
		β	λ_1	λ_2	ρ	σ_e^2	σ_*^2	σ_c^2	b
(1) DGP with random effects									
θ_w	Mean	0.9988	0.3035	0.2913	0.2982	0.9865			
	E-SD	0.0463	0.0907	0.1133	0.0792	0.0756			
	RMSE	0.0463	0.0907	0.1136	0.0792	0.0768			
θ_r	Mean	0.9990	0.3006	0.2918	0.2963	0.9893		0.4803	0.9985
	E-SD	0.0405	0.0795	0.0998	0.0780	0.0755		0.1383	0.1702
	RMSE	0.0405	0.0795	0.1001	0.0781	0.0762		0.1397	0.1702
θ_b	Mean	0.9935	0.2909	0.2658	0.2054		0.8142		1.0126
E-SD		0.0852	0.1599	0.2023	0.3768		0.1155		0.2664
RMSE		0.0854	0.1840	0.2128	0.3769		0.1205		0.2667
The type I error under different significance levels									
Significance level				0.01	0.05	0.10			
LM test				0.007	0.047	0.102			
Hausman test				0.030	0.072	0.123			
(2) DGP with fixed effects									
θ_w	Mean	0.9988	0.3035	0.2913	0.2982	0.9865			
	E-SD	0.0463	0.0907	0.1133	0.0792	0.0756			
	RMSE	0.0463	0.0907	0.1136	0.0792	0.0768			
θ_r	Mean	1.0705	0.5407	0.4348	0.4812	0.9886		6.1297	0.9066
	E-SD	0.0444	0.0819	0.1145	0.0917	0.0794		1.1094	0.4974
	RMSE	0.0833	0.2543	0.1768	0.2031	0.0802		5.7380	0.5061
θ_b	Mean	3.1528	0.7541	0.1496	0.0800		1.0141		0.4907
	E-SD	0.0978	0.0356	0.1382	0.1885		0.1430		0.1380
	RMSE	2.1550	0.5552	0.1471	0.2234		0.2189		0.5277
The power under different significance levels									
Significance level				0.01	0.05	0.10			
LM test				1.00	1.00	1.00			
Hausman test				1.00	1.00	1.00			

NOTE: (i) For the KKP DGP, $(\beta_0, \lambda_{10}, \lambda_{20}, \rho_0, \sigma_{e0}^2, \sigma_{c0}^2, b_0) = (1, 0.3, 0.3, 0.3, 1, 0.5, 1)$.
(ii) For the between estimates θ_b, the implied parameter $\sigma_*^2 = \sigma_{c0}^2 + \sigma_1^2$ has true value 0.8484. Also, the initial search value for θ_1 is from the within estimates.
(iii) The exogenous variables are generated from $x_{it} = \xi_i + z_{it}$, where ξ_i and z_{it} are independent standard normal random variables.

and fixed effects estimates perform well with small biases and small standard deviations; also, the Hausman test has an accurate size.

When we omit the spatial error while keeping the spatial lag and serial correlation in the model, biases for the estimates of β increase slightly, which results in a higher RMSE. The estimates of λ_1 would be biased upward by about 50% for both random and fixed effects estimates. Also, the type I error of the Hausman test remains accurate. When we omit the spatial lag and still include the spatial error and serial correlation in the regression, the estimates of β are biased downward by 10% and the estimates of λ_1 are biased upward by about 30%. For the Hausman test, there is an overrejection because the type I error is very high. For the above two misspecifications, the estimates of b are not reliable for either of them.

If we include both spatial lag and spatial error in the model but omit the serial correlation, the consequences for the estimates are not substantial, as there are only slight increases in the RMSEs, but, the Hausman test has an overrejection of random effects DGP. However, if we omit either spatial error or spatial lag, while at the same time omitting the serial correlation, we find that the increases in the biases of β and λ_j for $j = 1, 2$ are larger than those of omitting spatial error only or spatial lag only.

When the DGP is a fixed effects model, as we see from Table 12, there are similar misspecification consequences of estimates and testings due to omitting spatial structures or serial

TABLE 10
ANSELIN MODEL SPECIFICATION: TESTS FOR RANDOM VERSUS FIXED EFFECTS

		θ_1					θ_2	
		β	λ_1	λ_2	ρ	σ_e^2	σ_c^2	b
(1) DGP with random effects								
θ_w	Mean	0.9988	0.3035	0.2913	0.2982	0.9865		
	E-SD	0.0463	0.0907	0.1133	0.0792	0.0756		
	RMSE	0.0463	0.0907	0.1136	0.0792	0.0768		
θ_r	Mean	0.9991	0.2976	0.2957	0.2950	0.9884	0.4840	1.0025
	E-SD	0.0403	0.0737	0.0999	0.0767	0.0751	0.1379	0.1507
	RMSE	0.0403	0.0737	0.1000	0.0769	0.0760	0.1388	0.1507
θ_b	Mean	0.9942	0.2791	0.1412	0.2044	2.1963	0.2476	1.0292
	E-SD	0.0859	0.1573	0.3611	0.3769	2.4083	0.2431	0.2540
	RMSE	0.0861	0.1587	0.3945	0.3889	2.6891	0.3504	0.2557
The type I error under different significance levels								
Significance level				0.01	0.05	0.10		
LM test				0.020	0.072	0.125		
Hausman test				0.030	0.063	0.102		
(2) DGP with fixed effects								
θ_w	Mean	0.9988	0.3035	0.2913	0.2982	0.9865		
	E-SD	0.0463	0.0907	0.1133	0.0792	0.0756		
	RMSE	0.0463	0.0907	0.1136	0.0792	0.0768		
θ_r	Mean	1.0703	0.5281	0.4565	0.4750	0.9883	6.0779	0.9390
	E-SD	0.0442	0.0595	0.0905	0.0934	0.0795	1.0724	0.2929
	RMSE	0.0831	0.2358	0.1808	0.1983	0.0803	5.680	0.2992
θ_b	Mean	3.2301	0.4960	0.3971	0.0845	3.1404	0.3491	1.0062
	E-SD	0.0987	0.0605	0.2558	0.1903	2.0477	0.3080	0.2076
	RMSE	2.2323	0.2051	0.2736	0.2876	2.9622	0.3430	0.2077
The power under different significance levels								
Significance level				0.01	0.05	0.10		
LM test				1.00	1.00	1.00		
Hausman test				1.00	1.00	1.00		

NOTE: (i) For the Anselin DGP, $(\beta_0, \lambda_{10}, \lambda_{20}, \rho_0, \sigma_{e0}^2, \sigma_{c0}^2, b_0) = (1, 0.3, 0.3, 0.3, 1, 0.5, 1)$.
(ii) For the between estimates θ_b, the initial search value for θ_1 is from the within estimates.
(iii) The exogenous variables are generated from $x_{it} = \xi_i + z_{it}$, where ξ_i and z_{it} are independent standard normal random variables.

correlations. However, we see that the powers of the Hausman test are high even under various misspecifications.

Thus, we recommend running various specifications on given data including the general model with spatial lag, spatial error, and serial correlation. When different specifications yield different empirical results so that robustness is a priority over efficiency, the result from a general specification should be preferable.

5. CONCLUSION

This article investigates spatial panel data models with a space–time filter in disturbances. We estimate the model by both fixed effects and random effects specifications. With a between equation properly defined, the random effects model can be decomposed into a within equation and a between equation. The within equation corresponds to the fixed effects model with individual effects eliminated for estimation, the between equation illustrates differences of various random components specifications, and the random effects estimate is the pooling of the within and between estimates. A Hausman-type specification test and an LM test are proposed. Monte Carlo experiments are conducted to investigate the performance of the ML estimation and tests of various model specifications. Consequences of misspecifications of omitting spatial and/or serial correlations are also investigated via Monte Carlo experiments.

TABLE 11

CONSEQUENCES OF VARIOUS MODEL MISSPECIFICATIONS: DGP IS A RANDOM EFFECTS MODEL WITH SPATIAL LAG, SPATIAL ERROR, AND SERIAL CORRELATION

		θ_1					θ_2	
		β	λ_1	λ_2	ρ	σ_e^2	σ_c^2	b
(1) Correct specification								
θ_r	Mean	0.9993	0.4952	0.4984	0.4937	0.9904	0.9999	1.0142
	E-SD	0.0320	0.0588	0.0714	0.0503	0.0580	0.1970	0.1972
	RMSE	0.0320	0.0590	0.0715	0.0507	0.0588	0.1970	0.1977
θ_w	Mean	0.9991	0.4956	0.4993	0.4953	0.9904		
	E-SD	0.0324	0.0636	0.0741	0.0535	0.0584		
	RMSE	0.0324	0.0638	0.0741	0.0537	0.0592		
Type I error of Hausman test (theoretical 0.01, 0.05, 0.10)						0.020	0.051	0.084
(2) Misspecification omitting spatial error								
θ_r	Mean	0.9717	0.7221		0.4832	1.0235	0.9706	0.5566
	E-SD	0.0329	0.0241		0.0545	0.0592	0.1955	0.0980
	RMSE	0.0434	0.2234		0.0570	0.0636	0.1977	0.4541
θ_w	Mean	0.9703	0.7313		0.4816	1.0177		
	E-SD	0.0332	0.0256		0.0558	0.0589		
	RMSE	0.0446	0.2327		0.0588	0.0615		
Type I error of Hausman test						0.012	0.047	0.091
(3) Misspecification omitting spatial lag								
θ_r	Mean	0.8913		0.8213	0.5174	0.9907	0.7969	2.0087
	E-SD	0.0292		0.0200	0.0551	0.0591	0.2016	0.3038
	RMSE	0.1126		0.3219	0.0578	0.0598	0.2861	1.0534
θ_w	Mean	0.8904		0.8205	0.4757	0.9648		
	E-SD	0.0295		0.0211	0.0540	0.0574		
	RMSE	0.1135		0.3212	0.0592	0.0674		
Type I error of Hausman test						0.270	0.437	0.558
(4) Misspecification omitting serial correlation								
θ_r	Mean	1.0026	0.5297	0.4674		0.9664	1.3851	0.9434
	E-SD	0.0356	0.0577	0.0785		0.0612	0.1860	0.1882
	RMSE	0.0357	0.0649	0.0850		0.0698	0.4276	0.1965
θ_w	Mean	0.9980	0.4931	0.4997		0.9648		
	E-SD	0.0366	0.0749	0.0861		0.0611		
	RMSE	0.0366	0.0753	0.0861		0.0705		
Type I error of Hausman test						0.182	0.233	0.263
(5) Misspecification omitting serial correlation and spatial error								
θ_r	Mean	0.9697	0.7341			0.9885	1.3204	0.5320
	E-SD	0.0365	0.0256			0.0612	0.1765	0.0969
	RMSE	0.0474	0.2355			0.0623	0.3659	0.4779
θ_w	Mean	0.9680	0.7447			0.9826		
	E-SD	0.0369	0.0270			0.0607		
	RMSE	0.0488	0.2462			0.0632		
Type I error of Hausman test						0.002	0.024	0.057
(6) Misspecification omitting serial correlation and spatial lag								
θ_r	Mean	0.8904		0.8345		0.9479	1.4339	2.0054
	E-SD	0.0326		0.0227		0.0608	0.2420	0.3149
	RMSE	0.1143		0.3353		0.0801	0.4969	1.0536
θ_w	Mean	0.8902		0.8176		0.9391		
	E-SD	0.0331		0.0231		0.0601		
	RMSE	0.1146		0.3185		0.0856		
Type I error of Hausman test						0.745	0.837	0.864

NOTE: For the DGP, $(\beta_0, \lambda_{10}, \lambda_{20}, \rho_0, \sigma_{e0}^2, \sigma_{c0}^2, b_0) = (1, 0.5, 0.5, 0.5, 1, 1, 1)$.

TABLE 12
CONSEQUENCES OF VARIOUS MODEL MISSPECIFICATIONS: DGP IS A FIXED EFFECTS MODEL WITH SPATIAL LAG, SPATIAL ERROR, AND SERIAL CORRELATION

		θ_1					θ_2	
		β	λ_1	λ_2	ρ	σ_e^2	σ_c^2	b
(1) Misspecification as a random effects model								
θ_r	Mean	1.0231	0.5065	0.4875	0.4925	0.9908	1.9885	0.9926
	E-SD	0.0321	0.0542	0.0701	0.0536	0.0585	0.3339	0.2053
	RMSE	0.0396	0.0546	0.0712	0.0542	0.0592	1.0433	0.2055
Power of Hausman test						1	1	1
(2) Additional misspecification omitting spatial error								
θ_r	Mean	0.9976	0.7164		0.4838	1.0277	1.9436	0.5689
	E-SD	0.0330	0.0242		0.0553	0.0595	0.3175	0.1078
	RMSE	0.0331	0.2178		0.0577	0.0656	0.9956	0.4444
Power of Hausman test						1	1	1
(3) Additional misspecification omitting spatial lag								
θ_r	Mean	0.9102		0.8250	0.5360	1.0001	1.7460	2.0124
	E-SD	0.0296		0.0199	0.0638	0.0623	0.3945	0.3634
	RMSE	0.0946		0.3256	0.0732	0.0623	0.8439	1.0756
Power of Hausman test						1	1	1
(4) Additional misspecification omitting serial correlation								
θ_r	Mean	1.0315	0.5308	0.4652		0.9671	2.3671	0.9418
	E-SD	0.0358	0.0528	0.0746		0.0613	0.3105	0.1951
	RMSE	0.0477	0.0612	0.0824		0.0696	1.4019	0.2036
Power of Hausman test						1	1	1
(5) Additional misspecification omitting serial correlation and spatial error								
θ_r	Mean	1.0020	0.7281			0.9929	2.2948	0.5446
	E-SD	0.0366	0.0257			0.0615	0.3054	0.1049
	RMSE	0.0367	0.2296			0.0619	1.3304	0.4673
Power of Hausman test						1	1	1
(6) Additional misspecification omitting serial correlation and spatial lag								
θ_r	Mean	0.9116		0.8327		0.9460	2.7434	2.0092
	E-SD	0.0330		0.0237		0.0609	0.4570	0.3730
	RMSE	0.0943		0.3335		0.0814	1.8023	1.0759
Power of Hausman test						1	1	1

NOTE: (i) The significance levels for Hausman test are 0.01, 0.05, 0.10, respectively.
(ii) For the DGP, $(\beta_0, \lambda_{10}, \lambda_{20}, \rho_0, \sigma_{e0}^2, \sigma_{c0}^2, b_0) = (1, 0.5, 0.5, 0.5, 1, 1, 1)$.
(iii) The fixed individual effects are generated by $\mathbf{c}_{n0} = \sqrt{6}\bar{X}_{nT} + \zeta_n$ from (18).
(iv) For the fixed effects estimates, they are the same as those in Table 11 and hence not reported here.

APPENDIX

A. *Notations, Some Important Matrices, and Algebra.*

A.1. *Notations*. The following list summarizes some frequently used notations in the article:
For $j = 1, 2, 3$, $S_{nj}(\lambda_j) = I_n - \lambda_j W_{nj}$ for any possible λ_j and $G_{nj}(\lambda_j) = W_{nj} S_{nj}^{-1}(\lambda_j)$.
For $j = 2, 3$, $B_{nj}(\delta_j) = (I_n + \delta_j M_{nj})$ for any possible δ_j and $K_{nj}(\delta_j) = B_{nj}^{-1}(\delta_j) M_{nj}$.
$S_{nj} = I_n - \lambda_{j0} W_{nj}$ and $G_{nj} = W_{nj} S_{nj}^{-1}$; $B_{nj} = (I_n + \delta_{j0} M_{nj})$ and $K_{nj} = B_{nj}^{-1} M_{nj}$.
$\mathcal{A}^s = \mathcal{A}' + \mathcal{A}$ for any square matrix $\mathcal{A}$.
$vec_D(\mathcal{A})$ is the column vector formed by diagonal elements of any square matrix $\mathcal{A}$.
$\mathbf{W}_{nT,j} = I_T \otimes W_{nj}$ and $\mathbf{G}_{nT,j} = I_T \otimes G_{nj}$.
$e_{n1} \equiv \sqrt{1 - \rho_0^2} V_{n1}$ and $\mathbf{e}_{nT} = (e'_{n1}, e'_{n2}, \ldots, e'_{nT})'$.
$\mathbf{V}_{nT} = (V'_{n1}, \ldots, V'_{nT})'$ and $\mathbf{e}^d_{n,T-1} = (\Delta V'_{n2}, \Delta e'_{n3}, \ldots, \Delta e'_{nT})'$.

A.2. *A list of some matrices*. For an AR(1) process $v_t = \rho v_{t-1} + e_t$, where e_t's are i.i.d. with zero mean and a unit variance, the following matrices in (A.1)–(A.3) are important.

The $T \times T$ variance matrix of $(v_1, \dots, v_T)$ is

$$\Sigma_T(\rho) = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{pmatrix}. \tag{A.1}$$

The inverse matrix of $\Sigma_T(\rho)$ is

$$\Sigma_T^{-1}(\rho) = \begin{pmatrix} 1 & -\rho & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1+\rho^2 & -\rho \\ 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}. \tag{A.2}$$

A decomposition: $\Sigma_T^{-1}(\rho) = P'_T(\rho)P_T(\rho)$, where

$$P_T(\rho) = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & \cdots & 0 & 0 \\ -\rho & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}. \tag{A.3}$$

The $(T-1) \times T$ difference operator:

$$L_{T-1,T} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}. \tag{A.4}$$

The $(T-1) \times (T-1)$ quasi-difference operator:

$$Q_{T-1}(\rho) = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & \cdots & 0 & 0 \\ 0 & -\rho & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ddots & -\rho & 1 \end{pmatrix}. \tag{A.5}$$

The variance matrix of $\mathbf{e}^d_{n,T-1}$ is $\sigma^2_{e0} H_{T-1} \otimes I_n$, where $H_{T-1} = H_{T-1}(\rho_0)$ with

$$H_{T-1}(\rho) = \begin{pmatrix} \frac{2}{1+\rho} & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix}. \tag{A.6}$$

The inverse matrix of (A.6) is

(A.7)

$$H^{-1}_{T-1}(\rho) = \frac{1-\rho}{T-(T-2)\rho}$$
$$\times \begin{pmatrix} T-1 & (T-2) & (T-3) & \cdots & 2 & 1 \\ (T-2) & (T-2)\omega & (T-3)\omega & \cdots & 2\omega & \omega \\ (T-3) & (T-3)\omega & (T-3)(2\omega-1) & \cdots & 2(2\omega-1) & (2\omega-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 2 & 2\omega & 2(2\omega-1) & \cdots & 2[(T-3)\omega-(T-4)] & (T-3)\omega-(T-4) \\ 1 & \omega & (2\omega-1) & \cdots & [(T-3)\omega-(T-4)] & (T-2)\omega-(T-3) \end{pmatrix},$$

where $\omega = \frac{2}{1+\rho}$.

An important weighting matrix for the within estimation:

$$\mathbb{J}_T(\rho) = \Sigma_T^{-1}(\rho) - \frac{1-\rho}{T-(T-2)\rho} \begin{pmatrix} 1 \\ 1-\rho \\ \vdots \\ 1-\rho \\ 1 \end{pmatrix} (1, 1-\rho, \ldots, 1-\rho, 1). \tag{A.8}$$

A.3. *Algebra about* $H_{T-1}(\rho)$ *and* $\mathbb{J}_T(\rho)$. From Hsiao et al. (2002), there exist $A_{T-1}(\rho)$ and $D_{T-1}(\rho)$ such that $A_{T-1}(\rho)H_{T-1}(\rho)A'_{T-1}(\rho) = D_{T-1}(\rho)$, where $D_{T-1}(\rho) = \operatorname{diag}\{a_0a_1, a_1a_2, \ldots, a_{T-2}a_{T-1}\}$,

$$A_{T-1}(\rho)=\begin{pmatrix} a_0 & 0 & 0 & \cdots & 0 \\ a_0 & a_1 & 0 & \ddots & \vdots \\ a_0 & a_1 & a_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ a_0 & a_1 & a_2 & \cdots & a_{T-2} \end{pmatrix} \quad \text{with } A_{T-1}^{-1}(\rho)=\begin{pmatrix} \frac{1}{a_0} & 0 & 0 & \cdots & 0 \\ -\frac{1}{a_1} & \frac{1}{a_1} & 0 & \ddots & \vdots \\ 0 & -\frac{1}{a_2} & \frac{1}{a_2} & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & -\frac{1}{a_{T-2}} & \frac{1}{a_{T-2}} \end{pmatrix},$$

$a_s = 1 + s(\omega - 1)$ for $s = 0, \ldots, T-1$ and $\omega(\rho) = \frac{2}{1+\rho}$ (for notational simplicity, we simplify $a_s(\rho)$ as a_s and $\omega(\rho)$ as ω). Thus, $H_{T-1}^{-1}(\rho) = A'_{T-1}(\rho)D_{T-1}^{-1}(\rho)A_{T-1}(\rho)$. With $|H_{T-1}(\rho)| = 1 + (T-1)(\omega-1)$, we have the explicit expression for $H_{T-1}^{-1}(\rho)$ in (A.7).

To derive $\mathbb{J}_T(\rho) = L'_{T-1,T}Q'_{T-1}(\rho)H_{T-1}^{-1}(\rho)Q_{T-1}(\rho)L_{T-1,T}$, we have

$$A_{T-1}(\rho)Q_{T-1}(\rho)L_{T-1,T}$$
$$=\begin{pmatrix} -1 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ -\frac{1-\rho}{1+\rho} & -1 & a_1 & 0 & 0 & \cdots & 0 & 0 \\ -\frac{1-\rho}{1+\rho} & -\frac{(1-\rho)^2}{1+\rho} & a_1-(1+\rho)a_2 & a_2 & 0 & \cdots & 0 & 0 \\ -\frac{1-\rho}{1+\rho} & -\frac{(1-\rho)^2}{1+\rho} & -\frac{(1-\rho)^2}{1+\rho} & a_2-(1+\rho)a_3 & a_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{1-\rho}{1+\rho} & -\frac{(1-\rho)^2}{1+\rho} & -\frac{(1-\rho)^2}{1+\rho} & -\frac{(1-\rho)^2}{1+\rho} & -\frac{(1-\rho)^2}{1+\rho} & \cdots & a_{T-3} & 0 \\ -\frac{1-\rho}{1+\rho} & -\frac{(1-\rho)^2}{1+\rho} & -\frac{(1-\rho)^2}{1+\rho} & -\frac{(1-\rho)^2}{1+\rho} & -\frac{(1-\rho)^2}{1+\rho} & \cdots & a_{T-3}-(1+\rho)a_{T-2} & a_{T-2} \end{pmatrix}.$$

Thus, with $D_{T-1}(\rho) = \text{diag}\{a_0a_1, a_1a_2, \ldots, a_{T-2}a_{T-1}\}$, elements of the first row of $\mathbb{J}_T(\rho)$ are

$$\mathbb{J}_T(1,1) = \frac{1}{a_0a_1} + \left(\frac{1-\rho}{1+\rho}\right)^2\left(\frac{1}{a_1a_2} + \cdots \frac{1}{a_{T-2}a_{T-1}}\right)$$
$$\mathbb{J}_T(1,2) = -\frac{1}{a_0a_1} + \frac{1-\rho}{1+\rho}\frac{1}{a_1a_2} + \frac{(1-\rho)^3}{(1+\rho)^2}\left(\frac{1}{a_2a_3} + \cdots \frac{1}{a_{T-2}a_{T-1}}\right)$$
$$\mathbb{J}_T(1,3) = -a_1\frac{1-\rho}{1+\rho}\frac{1}{a_1a_2} - \frac{1-\rho}{1+\rho}(a_1-(1+\rho)a_2)\frac{1}{a_1a_2} + \frac{(1-\rho)^3}{(1+\rho)^2}\left(\frac{1}{a_3a_4} + \cdots \frac{1}{a_{T-2}a_{T-1}}\right)$$
$$\mathbb{J}_T(1,4) = -a_2\frac{1-\rho}{1+\rho}\frac{1}{a_2a_3} - \frac{1-\rho}{1+\rho}(a_2-(1+\rho)a_3)\frac{1}{a_3a_4} + \frac{(1-\rho)^3}{(1+\rho)^2}\left(\frac{1}{a_4a_5} + \cdots \frac{1}{a_{T-2}a_{T-1}}\right)$$
$$\vdots$$
$$\mathbb{J}_T(1,T) = -\frac{1-\rho}{1+\rho}a_{T-2}\left(\frac{1}{a_{T-2}a_{T-1}}\right).$$

For the second row, we have

$$\mathbb{J}_T(2,1) = \mathbb{J}_{T-1}(1,2)$$

$$\mathbb{J}_T(2,2) = \frac{1}{a_0a_1} + \frac{1}{a_1a_2} + \frac{(1-\rho)^4}{(1+\rho)^2}\left(\frac{1}{a_2a_3} + \cdots \frac{1}{a_{T-2}a_{T-1}}\right)$$

$$\mathbb{J}_T(2,3) = -a_1\frac{1}{a_1a_2} - \frac{(1-\rho)^2}{1+\rho}(a_1-(1+\rho)a_2)\frac{1}{a_2a_3} + \frac{(1-\rho)^4}{(1+\rho)^2}\left(\frac{1}{a_3a_4} + \cdots \frac{1}{a_{T-2}a_{T-1}}\right)$$

$$\mathbb{J}_T(2,4) = -a_2\frac{(1-\rho)^2}{1+\rho}\frac{1}{a_2a_3} - \frac{(1-\rho)^2}{1+\rho}(a_2-(1+\rho)a_3)\frac{1}{a_3a_4} + \frac{(1-\rho)^4}{(1+\rho)^2}\left(\frac{1}{a_4a_5} + \cdots \frac{1}{a_{T-2}a_{T-1}}\right)$$

$$\vdots$$

$$\mathbb{J}_T(2,T) = -\frac{(1-\rho)^2}{1+\rho}a_{T-2}\left(\frac{1}{a_{T-2}a_{T-1}}\right).$$

The rest of the rows can be derived similarly. By using $\frac{1}{a_ta_{t+1}} + \cdots \frac{1}{a_{T-2}a_{T-1}} = \frac{1+\rho}{1-\rho}(\frac{1}{a_t} - \frac{1}{a_{T-1}})$ for $t < T-1$ and the values of a_s for $s = 0, \ldots, T-1$, these above items can be simplified. Thus, we have the explicit expression of $\mathbb{J}_T(\rho)$ in (A.8). It is apparent that $\mathbb{J}_T(\rho)$ is UB as $|\rho| < 1$.

B. *Appendix for section 2*.

B.1. *Algebra for the within equation (5)*.

B.1.1. *The score and information matrix*. From the DGP, $\mathbf{W}_{nT,1}\mathbf{Y}_{nT} = (I_T \otimes G_{n1}(\lambda_1))\mathbf{X}_{nT}\beta + (I_T \otimes G_{n1}(\lambda_1)S_{n2}^{-1}(\lambda_2)B_{n2}(\delta_2))\mathbf{V}_{nT}(\theta_1)$. Denoting $\ddot{\mathbf{X}}_{nT} = (I_T \otimes B_{n2}^{-1}S_{n2})\mathbf{X}_{nT}$, $\ddot{G}_{n1} = B_{n2}^{-1}S_{n2}G_{n1}S_{n2}^{-1}B_{n2}$ and $\ddot{G}_{n2} = B_{n2}^{-1}G_{n2}B_{n2}$, the score of (5) is

$$\frac{\partial \ln L_{w,nT}(\theta_{10})}{\partial \beta} = \frac{1}{\sigma_{e0}^2}\ddot{\mathbf{X}}'_{nT}(\mathbb{J}_T \otimes I_n)\mathbf{V}_{nT},$$

$$\frac{\partial \ln L_{w,nT}(\theta_{10})}{\partial \lambda_1} = \frac{1}{\sigma_{e0}^2}(\ddot{\mathbf{X}}_{nT}\beta_0)'(\mathbb{J}_T \otimes \ddot{G}'_{n1})\mathbf{V}_{nT} + \frac{1}{\sigma_{e0}^2}\mathbf{V}'_{nT}(\mathbb{J}_T \otimes \ddot{G}_{n1})\mathbf{V}_{nT} - (T-1)tr(G_{n1}),$$

$$\frac{\partial \ln L_{w,nT}(\theta_{10})}{\partial \lambda_2} = \frac{1}{\sigma_{e0}^2}\mathbf{V}'_{nT}(\mathbb{J}_T \otimes \ddot{G}_{n2})\mathbf{V}_{nT} - (T-1)tr(G_{n2}),$$

$$\frac{\partial \ln L_{w,nT}(\theta_{10})}{\partial \delta_2} = \frac{1}{\sigma_{e0}^2}\mathbf{V}'_{nT}(\mathbb{J}_T \otimes K_{n2})\mathbf{V}_{nT} - (T-1)tr(K_{n2}),$$

$$\frac{\partial \ln L_{w,nT}(\theta_{10})}{\partial \rho} = -\frac{1}{2\sigma_{e0}^2}\mathbf{V}'_{nT}(\frac{\partial \mathbb{J}_T}{\partial \rho} \otimes I_n)\mathbf{V}_{nT} + \frac{n(T-1)}{1+\rho_0}\frac{1}{T-(T-2)\rho_0},$$

$$\frac{\partial \ln L_{w,nT}(\theta_{10})}{\partial \sigma_e^2} = \frac{1}{2\sigma_{e0}^4}\mathbf{V}'_{nT}(\mathbb{J}_T \otimes I_n)\mathbf{V}_{nT} - \frac{n(T-1)}{2}\frac{1}{\sigma_{e0}^2}.$$

Also, we have the information matrix

(B.1)
$$\Sigma_{w,nT} = \begin{bmatrix} \mathbf{H}_{w,nT} & * \\ \mathbf{0}_{4\times(k_x+1)} & \mathbf{0}_{4\times 4}\end{bmatrix} + \frac{1}{n}\times \begin{bmatrix} \mathbf{0}_{k_x\times k_x} & * & * & * & * & * \\ \mathbf{0}_{1\times k_x} & tr(\ddot{G}_{n1}'\ddot{G}_{n1}^s) & * & * & * & * \\ \mathbf{0}_{1\times k_x} & tr(\ddot{G}_{n1}'\ddot{G}_{n2}^s) & 0 & * & * & * \\ \mathbf{0}_{1\times k_x} & tr(\ddot{G}_{n1}'K_{n2}^s) & 0 & 0 & * & * \\ \mathbf{0}_{1\times k_x} & -\frac{2}{1+\rho_0}\frac{1}{T-(T-2)\rho_0}tr\ddot{G}_{n1} & 0 & 0 & 0 & * \\ \mathbf{0}_{1\times k_x} & \frac{1}{\sigma_{e0}^2}tr(\ddot{G}_{n1}) & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$+\frac{1}{n}\times \begin{bmatrix} \mathbf{0}_{k_x\times k_x} & * & * & * & * & * \\ \mathbf{0}_{1\times k_x} & 0 & * & * & * & * \\ \mathbf{0}_{1\times k_x} & 0 & tr(\ddot{G}_{n2}'\ddot{G}_{n2}^s) & * & * & * \\ \mathbf{0}_{1\times k_x} & 0 & tr(\ddot{G}_{n2}'K_{n2}^s) & tr(K_{n2}'K_{n2}^s) & * & * \\ \mathbf{0}_{1\times k_x} & 0 & -\frac{2}{1+\rho_0}\frac{1}{T-(T-2)\rho_0}tr\ddot{G}_{n2} & -\frac{2}{1+\rho_0}\frac{1}{T-(T-2)\rho_0}tr(K_{n2}) & -\frac{1}{(T-1)}E\frac{\partial^2 \ln L_{w,nT}(\theta_{10})}{\partial\rho^2} & * \\ \mathbf{0}_{1\times k_x} & 0 & \frac{1}{\sigma_{e0}^2}tr(\ddot{G}_{n2}) & \frac{1}{\sigma_{e0}^2}tr(K_{n2}) & -\frac{n}{2\sigma_{e0}^2}\frac{2}{1+\rho_0}\frac{1}{T-(T-2)\rho_0} & \frac{n}{2}\frac{1}{\sigma_{e0}^4} \end{bmatrix},$$

where $\mathbf{H}_{w,nT} = \frac{1}{\sigma_{e0}^2 n(T-1)}[\ddot{\mathbf{X}}_{nT}, \ddot{\mathbf{G}}_{nT,1}\ddot{\mathbf{X}}_{nT}\beta_0]'(\mathbb{J}_T\otimes I_n)[\ddot{\mathbf{X}}_{nT}, \ddot{\mathbf{G}}_{nT,1}\ddot{\mathbf{X}}_{nT}\beta_0]$ is of dimension $(k_x+1)\times(k_x+1)$ and

$$-E\frac{\partial^2 \ln L_{w,nT}(\theta_{10})}{\partial\rho^2} = \frac{2n}{(1+\rho_0)}\frac{T-1}{(T-(T-2)\rho_0)} \times\left[\frac{(T-1)}{(1+\rho_0)(T-(T-2)\rho_0)} + \frac{T-2}{2} - \frac{1}{(1-\rho_0)} + \frac{1}{T-1}\frac{1-\rho_0^{T-1}}{(1-\rho_0)^2}\right].$$

B.1.2. *Variance matrix of the score.* From the score $\frac{\partial \ln L_{w,nT}(\theta_{10})}{\partial\theta_1}$, we have $E\frac{1}{n(T-1)}\times \frac{\partial \ln L_{w,nT}(\theta_{10})}{\partial\theta_1}\frac{\partial \ln L_{w,nT}(\theta_{10})}{\partial\theta_1'} = \Sigma_{w,nT}+\Omega_{w,nT}$, where $\Sigma_{w,nT}$ is the information matrix, $\Omega_{w,nT}$ is related to the third and fourth moments of $e_{nt,i}$ ($\mu_{3,e}$ and $\mu_{4,e}$) and is equal to zero when $e_{nt,i}$'s are normal. Let $vec_D(\mathcal{A})$ be a column vector formed by diagonal elements of a square matrix $\mathcal{A}$. Denote $\mathcal{P} = [\mathcal{P}_1, \mathcal{P}_2]$, where $\mathcal{P}_1 = vec_D(P_T'^{-1}\mathbb{J}_T P_T^{-1}\otimes \ddot{G}_{n1})$ and

$$\mathcal{P}_2 = \Big[vec_D\big(P_T'^{-1}\mathbb{J}_T P_T^{-1}\otimes\ddot{G}_{n2}\big), vec_D\big(P_T'^{-1}\mathbb{J}_T P_T^{-1}\otimes K_{n2}\big), -\frac{1}{2}vec_D\left(P_T'^{-1}\frac{\partial\mathbb{J}_T}{\partial\rho}P_T^{-1}\otimes I_n\right), \frac{1}{2\sigma_{e0}^2}vec_D\big(P_T'^{-1}\mathbb{J}_T P_T^{-1}\otimes I_n\big)\Big]$$

with $P_T = P_T(\rho_0)$ in (A.3). By using $\mathbf{e}_{nT} = (P_T\otimes I_n)\mathbf{V}_{nT}$ and Lemma 3 in Yu et al. (2008), we have

(B.2)
$$\Omega_{w,nT} = \frac{\mu_{4,e}-3\sigma_{e0}^4}{\sigma_{e0}^4}\begin{bmatrix}\mathbf{0}_{k_x\times k_x} & * \\ \mathbf{0}_{5\times k_x} & \mathcal{P}'\mathcal{P}\end{bmatrix} + \frac{\mu_{3,e}}{\sigma_{e0}^4}\begin{bmatrix}\mathbf{0}_{k_x\times k_x} & * & * \\ \mathcal{P}_1'(\mathbb{J}_T\otimes I_n)\ddot{\mathbf{X}}_{nT} & 2\mathcal{P}_1'(\mathbb{J}_T\otimes\ddot{G}_{n1})\ddot{\mathbf{X}}_{nT}\beta_0 & * \\ \mathcal{P}_2'(\mathbb{J}_T\otimes I_n)\ddot{\mathbf{X}}_{nT} & \mathcal{P}_2'(\mathbb{J}_T\otimes\ddot{G}_{n1})\ddot{\mathbf{X}}_{nT}\beta_0 & \mathbf{0}_{4\times 4}\end{bmatrix}.$$

B.2. *Algebra for the Random Effects equation (9).*

B.2.1. *The score and information matrix.* The variance matrix Ω_{nT} in (8) is a function of $\phi_0 = (\lambda_{20}, \lambda_{30}, \delta_{20}, \delta_{30}, \rho_0, \sigma_{c0}^2, \sigma_{e0}^2)'$. For components in Ω_{nT}, we have $\frac{\partial \Sigma_{T,\rho_0}}{\partial \rho} = \frac{1}{1-\rho_0^2}(2\rho_0 \Sigma_{T,\rho_0} + F_{\rho_0})$, where F_{ρ_0} is a symmetric matrix as

$$F_{\rho_0} = \begin{pmatrix} 0 & * & * & \cdots & * \\ 1 & 0 & * & \cdots & * \\ 2\rho_0 & 1 & 0 & & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (T-1)\rho_0^{T-2} & (T-2)\rho_0^{T-3} & (T-3)\rho_0^{T-4} & \cdots & 0 \end{pmatrix};$$

also, $\frac{\partial S_{n2}^{-1}(\lambda_2) B_{n2}(\delta_2) B'_{n2}(\delta_2) S'^{-1}_{n2}(\lambda_2)}{\partial \lambda_2} = S_{n2}^{-1}(\lambda_2)(G_{n2}(\lambda_2) B_{n2}(\delta_2) B'_{n2}(\delta_2))^s S'^{-1}_{n2}(\lambda_2)$, which is denoted as $\mathcal{R}_1$ at true values. Similarly, at true parameter values, we define $\mathcal{C}_1 \equiv \frac{\partial S_{n3}^{-1} B_{n3} B'_{n3} S'^{-1}_{n3}}{\partial \lambda_3}$, $\mathcal{R}_3 \equiv \frac{\partial S_{n2}^{-1} B_{n2} B'_{n2} S'^{-1}_{n2}}{\partial \delta_2}$, and $\mathcal{C}_3 \equiv \frac{\partial S_{n3}^{-1} B_{n3} B'_{n3} S'^{-1}_{n3}}{\partial \delta_3}$. Hence,

$$\frac{\partial \Omega_{nT}}{\partial \lambda_2} = \sigma_{e0}^2 \Sigma_{T,\rho_0} \otimes \mathcal{R}_1, \quad \frac{\partial \Omega_{nT}}{\partial \lambda_3} = \sigma_{c0}^2 l_T l'_T \otimes \mathcal{C}_1, \quad \frac{\partial \Omega_{nT}}{\partial \delta_2} = \sigma_{e0}^2 \Sigma_{T,\rho_0} \otimes \mathcal{R}_3,$$
$$\frac{\partial \Omega_{nT}}{\partial \delta_3} = \sigma_{c0}^2 l_T l'_T \otimes \mathcal{C}_3, \quad \frac{\partial \Omega_{nT}}{\partial \rho} = \sigma_{e0}^2 \left[\frac{\partial \Sigma_{T,\rho_0}}{\partial \rho} \otimes \left(S'_{n2} B'^{-1}_{n2} B^{-1}_{n2} S_{n2}\right)^{-1} \right],$$
$$\frac{\partial \Omega_{nT}}{\partial \sigma_c^2} = l_T l'_T \otimes \left(S'_{n3} B'^{-1}_{n3} B^{-1}_{n3} S_{n3}\right)^{-1}, \quad \text{and} \quad \frac{\partial \Omega_{nT}}{\partial \sigma_e^2} = \Sigma_{T,\rho_0} \otimes \left(S'_{n2} B'^{-1}_{n2} B^{-1}_{n2} S_{n2}\right)^{-1}.$$

By denoting $\mathbf{Z}_{nT} = [l_T \otimes z_n, \mathbf{X}_{nT}]$ and $\gamma_0 = (b'_0, \beta'_0)$, the score vector has

$$\frac{\partial \ln L_{r,nT}(\theta_0)}{\partial \gamma} = \mathbf{Z}'_{nT} \Omega_{nT}^{-1} \boldsymbol{\xi}_{nT}$$
$$\frac{\partial \ln L_{r,nT}(\theta_0)}{\partial \lambda_1} = (\mathbf{G}_{nT,1} \mathbf{Z}_{nT} \gamma_0)' \Omega_{nT}^{-1} \boldsymbol{\xi}_{nT} + \boldsymbol{\xi}_{nT} \mathbf{G}'_{nT,1} \Omega_{nT}^{-1} \xi_{nT} - tr \mathbf{G}_{nT,1}$$
$$\frac{\partial \ln L_{r,nT}(\theta_0)}{\partial \phi_i} = \frac{1}{2} \left(\boldsymbol{\xi}'_{nT} \frac{\partial \Omega_{nT}^{-1}}{\partial \phi_i} \Omega_{nT}^{-1} \boldsymbol{\xi}_{nT} - tr \left(\frac{\partial \Omega_{nT}^{-1}}{\partial \phi_i} \Omega_{nT}^{-1} \right) \right),$$

where ϕ_i is an element in $\phi = (\lambda_2, \lambda_3, \delta_2, \delta_3, \rho, \sigma_c^2, \sigma_e^2)'$. By denoting $k = k_z + k_x$, the information matrix is

(B.3)

$$\Sigma_{r,nT} = \begin{bmatrix} \mathbf{H}_{r,nT} & * \\ \mathbf{0}_{7\times(k+1)} & \mathbf{0}_{7\times 7} \end{bmatrix} + \frac{1}{nT} \begin{bmatrix} \mathbf{0}_{k\times k} & * & * \\ \mathbf{0}_{1\times k} & T tr\left(G_{n1}^2\right) + tr\left(\mathbf{G}'_{nT,1} \Omega_{nT}^{-1} \mathbf{G}_{nT,1} \Omega_{nT}\right) & * \\ \mathbf{0}_{7\times k} & \Phi_{nT} & \Psi_{nT} \end{bmatrix},$$

where

$$\mathbf{H}_{r,nT} = \frac{1}{nT} \begin{bmatrix} \mathbf{Z}'_{nT} \Omega_{nT}^{-1} \mathbf{Z}_{nT} & * \\ (\mathbf{G}_{nT,1} \mathbf{Z}_{nT} \gamma_0)' \Omega_{nT}^{-1} \mathbf{Z}_{nT} & (\mathbf{G}_{nT,1} \mathbf{Z}_{nT} \gamma_0)' \Omega_{nT}^{-1} \mathbf{G}_{nT,1} \mathbf{Z}_{nT} \gamma_0 \end{bmatrix},$$

with $\gamma = (b', \beta')$, $\boldsymbol{\Phi}_{nT,i} = tr(\mathbf{G}'_{nT,1} \Omega_{nT}^{-1} \frac{\partial \Omega_{nT}}{\partial \phi_i})$, and $\Psi_{nT,ij} = -\frac{1}{2} tr(\frac{\partial \Omega_{nT}^{-1}}{\partial \phi_i} \frac{\partial \Omega_{nT}}{\partial \phi_j})$ for $i, j = 1, \ldots, 7$.

B.2.2. *Variance matrix of the score*. From the score $\frac{\partial \ln L_{r,nT}(\theta_0)}{\partial \theta}$, we have $E\frac{1}{nT} \times \frac{\partial \ln L_{r,nT}(\theta_0)}{\partial \theta}\frac{\partial \ln L_{r,nT}(\theta_0)}{\partial \theta'} = \Sigma_{r,nT} + \Omega_{r,nT}$, where $\Sigma_{r,nT}$ is the information matrix, and $\Omega_{r,nT}$ is related to the third and fourth moments of $e_{nt,i}$ and $c_{n0,i}$. Denote

$$
\begin{aligned}
\mathbb{A}_{c,\lambda_1} &= \big(l_T \otimes S_{n3}^{-1}B_{n3}\big)'\mathbf{G}'_{nT,1}\Omega_{nT}^{-1}\big(l_T \otimes S_{n3}^{-1}B_{n3}\big),\\
\mathbb{A}_{c,\phi_i} &= -\frac{1}{2}\big(l_T \otimes S_{n3}^{-1}B_{n3}\big)'\frac{\partial \Omega_{nT}^{-1}}{\partial \phi_i}\Omega_{nT}^{-1}\big(l_T \otimes S_{n3}^{-1}B_{n3}\big),\\
\mathbb{A}_{e,\lambda_1} &= \big(P_T^{-1} \otimes S_{n2}^{-1}B_{n2}\big)'\mathbf{G}'_{nT,1}\Omega_{nT}^{-1}\big(P_T^{-1} \otimes S_{n2}^{-1}B_{n2}\big),\\
\mathbb{A}_{e,\phi_i} &= -\frac{1}{2}\big(P_T^{-1} \otimes S_{n2}^{-1}B_{n2}\big)'\frac{\partial \Omega_{nT}^{-1}}{\partial \phi_i}\Omega_{nT}^{-1}\big(P_T^{-1} \otimes S_{n2}^{-1}B_{n2}\big),\\
\mathbb{B}_{c,\gamma} &= \big(l_T \otimes S_{n3}^{-1}B_{n3}\big)\Omega_{nT}^{-1}\mathbf{Z}_{nT}, \quad \mathbb{B}_{c,\lambda_1} = \big(l_T \otimes S_{n3}^{-1}B_{n3}\big)\Omega_{nT}^{-1}(\mathbf{G}_{nT,1}\mathbf{Z}_{nT}\gamma_0),\\
\mathbb{B}_{e,\gamma} &= \big(P_T^{-1} \otimes S_{n2}^{-1}B_{n2}\big)\Omega_{nT}^{-1}\mathbf{Z}_{nT}, \quad \mathbb{B}_{e,\lambda_1} = \big(P_T^{-1} \otimes S_{n2}^{-1}B_{n2}\big)\Omega_{nT}^{-1}(\mathbf{G}_{nT,1}\mathbf{Z}_{nT}\gamma_0),
\end{aligned}
$$

$\mathbb{P}_c = [\mathbb{P}_{c,\lambda_1}, \mathbb{P}_{c,\phi}]$ and $\mathbb{P}_e = [\mathbb{P}_{e,\lambda_1}, \mathbb{P}_{e,\phi}]$, where $\mathbb{P}_{c,\lambda_1} = vec_D(\mathbb{A}_{c,\lambda_1})$, $\mathbb{P}_{c,\phi} = [vec_D(\mathbb{A}_{c,\phi_1}), \dots, vec_D(\mathbb{A}_{c,\phi_7})]$, $\mathbb{P}_{e,\lambda_1} = vec_D(\mathbb{A}_{e,\lambda_1})$, and $\mathbb{P}_{e,\phi} = [vec_D(\mathbb{A}_{e,\phi_1}), \dots, vec_D(\mathbb{A}_{e,\phi_7})]$. Let $\mu_{3,c}$ and $\mu_{4,c}$ be the third and fourth moments of $c_{n0,i}$. By using Lemma 3 in Yu et al. (2008), we have

$$
\text{(B.4)}\quad \Omega_{r,nT} = \begin{bmatrix} \mathbf{0}_{k\times k} & * \\ \mathbf{0}_{8\times k} & \mu_{4,c}\mathbb{P}'_c\mathbb{P}_c + \mu_{4,e}\mathbb{P}'_e\mathbb{P}_e \end{bmatrix}
+ \begin{bmatrix} \mathbf{0}_{k\times k} & * & * \\ \mu_{3,c}\mathbb{P}'_{c,\lambda_1}\mathbb{B}_{c,\gamma} + \mu_{3,e}\mathbb{P}'_{e,\lambda_1}\mathbb{B}_{e,\gamma} & 2\mu_{3,c}\mathbb{P}'_{c,\lambda_1}\mathbb{B}_{c,\lambda_1} + 2\mu_{3,e}\mathbb{P}_{e,\lambda_1}\mathbb{B}_{e,\lambda_1} & * \\ \mu_{3,c}\mathbb{P}'_{c,\phi}\mathbb{B}_{c,\gamma} + \mu_{3,e}\mathbb{P}'_{e,\phi}\mathbb{B}_{e,\gamma} & \mu_{3,c}\mathbb{P}'_{c,\phi}\mathbb{B}_{c,\lambda_1} + \mu_{3,e}\mathbb{P}'_{e,\phi}\mathbb{B}_{e,\lambda_1} & \mathbf{0}_{7\times 7} \end{bmatrix}.
$$

B.2.3. *CLT for the random effects equation*. From the first-order condition of the random equation, the CLT involves the linear and quadratic forms of $\boldsymbol{\xi}_{nT} = l_T \otimes S_{n3}^{-1}B_{n3}\mathbf{c}_{n0} + (I_T \otimes S_{n2}^{-1}B_{n2})\mathbf{V}_{nT}$. The $\mathbf{V}_{nT}$ can be transformed into $\mathbf{e}_{nT}$ via $\mathbf{e}_{nT} = (P_T \otimes I_n)\mathbf{V}_{nT}$. Denote $\mathbf{U}_{n,T+1} = (\mathbf{c}'_{n0}, \mathbf{e}'_{nT})'$. By using $(l_T \otimes \mathbf{c}_{n0}) = (l_T \otimes I_n)\mathbf{c}_{n0}$, the object of the CLT has the form

$$
\begin{aligned}
&(l'_T \otimes \mathbf{c}'_{n0})\mathbf{A}_{nT}(l_T \otimes \mathbf{c}_{n0}) + \mathbf{e}'_{nT}\mathbf{B}_{nT}\mathbf{e}_{nT} + (l'_T \otimes \mathbf{c}'_{n0})\mathbf{C}_{nT}\mathbf{e}_{nT} + \mathbf{b}'_{n,T+1}\mathbf{U}_{n,T+1}\\
&\quad = \mathbf{U}'_{n,T+1}\begin{pmatrix} (l'_T \otimes I_n)\mathbf{A}_{nT}(l_T \otimes I_n) & \frac{1}{2}(l'_T \otimes I_n)\mathbf{C}_{nT} \\ \frac{1}{2}\mathbf{C}'_{nT}(l_T \otimes I_n) & \mathbf{B}_{nT} \end{pmatrix}\mathbf{U}_{n,T+1} + \mathbf{b}'_{n,T+1}\mathbf{U}_{n,T+1},
\end{aligned}
$$

where $\mathbf{A}_{nT}$, $\mathbf{B}_{nT}$, and $\mathbf{C}_{nT}$ are UB for n and T, and elements of $\mathbf{b}_{n,T+1}$ are uniformly bounded. Thus, as elements of $(\mathbf{c}'_{n0}, e'_{n1}, e'_{n2}, \dots, e'_{nT})$ are independent, the CLT in Kelejian and Prucha (2001) can be applied. Note that the CLT there requires that all elements in $\mathbf{c}_{n0}$, e_{nt} for all n and t are mutually independent but not necessarily identically distributed.

B.3. *Algebra for the between equation (12)*.

B.3.1. *The score and information matrix*. For $\Omega_{n1}(\phi)$, which is a function of $\phi = (\lambda_2, \lambda_3, \delta_2, \delta_3, \rho, \sigma_c^2, \sigma_e^2)'$, we have

$$\frac{\partial \Omega_{n1}}{\partial \lambda_2} = \sigma_1^2 \mathcal{R}_1, \quad \frac{\partial \Omega_{n1}}{\partial \lambda_3} = \sigma_{c0}^2 \mathcal{C}_1, \quad \frac{\partial \Omega_{n1}}{\partial \delta_2} = \sigma_1^2 \mathcal{R}_3, \quad \frac{\partial \Omega_{n1}}{\partial \delta_3} = \sigma_{c0}^2 \mathcal{C}_3,$$

$$\frac{\partial \Omega_{n1}}{\partial \rho} = \frac{2\sigma_{e0}^2 (T-1-(T-2)\rho_0)}{[(1-\rho_0)(T-(T-2)\rho_0)]^2} S_{n2}^{-1} B_{n2} B_{n2}' S_{n2}'^{-1},$$

$$\frac{\partial \Omega_{n1}}{\partial \sigma_c^2} = S_{n3}^{-1} B_{n3} B_{n3}' S_{n3}'^{-1}, \quad \text{and} \quad \frac{\partial \Omega_{n1}}{\partial \sigma_e^2} = \frac{1}{(1-\rho_0)(T-(T-2)\rho_0)} S_{n2}^{-1} B_{n2} B_{n2}' S_{n2}'^{-1}.$$

The score is

$$\frac{\partial \ln L_{b,n}(\theta_0)}{\partial b} = z_n' \Omega_{n1}^{-1} \boldsymbol{\xi}_n,$$

$$\frac{\partial \ln L_{b,n}(\theta_0)}{\partial \beta} = \vec{X}_{nT}' \Omega_{n1}^{-1} \boldsymbol{\xi}_n,$$

$$\frac{\partial \ln L_{b,n}(\theta_0)}{\partial \lambda_1} = -trG_{n1} + (W_{n1}\vec{Y}_{nT})' \Omega_{n1}^{-1} \boldsymbol{\xi}_n,$$

$$\frac{\partial \ln L_{b,n}(\theta_0)}{\partial \phi_i} = -\frac{1}{2} tr\left(\Omega_{n1}^{-1} \frac{\partial \Omega_{n1}}{\partial \phi_i}\right) + \frac{1}{2} \boldsymbol{\xi}_n' \Omega_{n1}^{-1} \frac{\partial \Omega_{n1}}{\partial \phi_i} \Omega_{n1}^{-1} \boldsymbol{\xi}_n \quad \text{for} \quad \phi_i \neq \rho,$$

$$\frac{\partial \ln L_{b,n}(\theta_0)}{\partial \rho} = -\frac{1}{2} tr\left(\Omega_{n1}^{-1} \frac{\partial \Omega_{n1}}{\partial \rho}\right) + \frac{1}{2} \boldsymbol{\xi}_n' \Omega_{n1}^{-1} \frac{\partial \Omega_{n1}}{\partial \rho} \Omega_{n1}^{-1} \boldsymbol{\xi}_n - \frac{\partial \boldsymbol{\xi}_n'}{\partial \rho} \Omega_{n1}^{-1} \boldsymbol{\xi}_n.$$

From these, we see that $\frac{\partial \ln L_{b,n}(\theta_0)}{\partial \rho}$ is different from $\frac{\partial \ln L_{b,n}(\theta_0)}{\partial \phi_i}$ for $\phi_i \neq \rho$, because ρ also appears in the regression equation. Thus, we have

$$\frac{\partial^2 \ln L_{b,n}(\theta_0)}{\partial^2 \phi_i} = -\frac{1}{2} tr\left(\frac{\partial \Omega_{n1}^{-1}}{\partial \phi_i} \frac{\partial \Omega_{n1}}{\partial \phi_i} + \Omega_{n1}^{-1} \frac{\partial^2 \Omega_{n1}}{\partial^2 \phi_i}\right) - \frac{1}{2} \boldsymbol{\xi}_n' \frac{\partial^2 \Omega_{n1}^{-1}}{\partial^2 \phi_i} \boldsymbol{\xi}_n \quad \text{for} \ \ \phi_i \neq \rho,$$

and

$$\frac{\partial^2 \ln L_{b,n}(\theta_0)}{\partial^2 \rho} = -\frac{1}{2} tr\left(\frac{\partial \Omega_{n1}^{-1}}{\partial \rho} \frac{\partial \Omega_{n1}}{\partial \rho} + \Omega_{n1}^{-1} \frac{\partial^2 \Omega_{n1}}{\partial^2 \rho}\right) - \frac{1}{2} \boldsymbol{\xi}_n' \frac{\partial^2 \Omega_{n1}^{-1}}{\partial^2 \rho} \boldsymbol{\xi}_n - 2 \frac{\partial \boldsymbol{\xi}_n'}{\partial \rho} \frac{\partial \Omega_{n1}^{-1}}{\partial \rho} \boldsymbol{\xi}_n - \frac{\partial^2 \boldsymbol{\xi}_n'}{\partial \rho^2} \Omega_{n1}^{-1} \boldsymbol{\xi}_n - \frac{\partial \boldsymbol{\xi}_n'}{\partial \rho} \Omega_{n1}^{-1} \frac{\partial \boldsymbol{\xi}_n}{\partial \rho}.$$

By denoting $\vec{Z}_{nT} = [z_n, \vec{X}_{nT}]$, the information matrix is

(B.5)
$$\Sigma_{b,n} = \begin{bmatrix} \mathbf{H}_{b,n} & * \\ \mathbf{0}_{7\times(k+1)} & \mathbf{0}_{7\times 7} \end{bmatrix} + \frac{1}{n} \begin{bmatrix} \mathbf{0}_{(k+5)\times(k+5)} & * & * \\ \mathbf{0}_{1\times(k+5)} & E\frac{\partial \boldsymbol{\xi}_n'}{\partial \rho} \Omega_{n1}^{-1} \frac{\partial \boldsymbol{\xi}_n}{\partial \rho} & * \\ \mathbf{0}_{2\times(k+5)} & \mathbf{0}_{2\times 1} & \mathbf{0}_{2\times 2} \end{bmatrix}$$
$$+ \frac{1}{n} \begin{bmatrix} \mathbf{0}_{k\times k} & * & * \\ \mathbf{0}_{1\times k} & tr(G_{n1}^2) + tr(G_{n1}'\Omega_{n1}^{-1} G_{n1} \Omega_{n1}) & * \\ \mathbf{0}_{7\times k} & \Phi_n & \Psi_n \end{bmatrix},$$

where

$$\mathbf{H}_{b,n} = \frac{1}{n}\begin{bmatrix} \vec{Z}'_{nT}\Omega_{n1}^{-1}\vec{Z}_{nT} & * \\ (G_{n1}\vec{Z}_{nT}\gamma_0)'\Omega_{n1}^{-1}\vec{Z}_{nT} & (G_{n1}\vec{Z}_{nT}\gamma_0)'\Omega_{n1}^{-1}G_{n1}\vec{Z}_{nT}\gamma_0 \end{bmatrix}$$

with $\Phi_{n,i} = tr(G'_{n1}\Omega_{n1}^{-1}\frac{\partial\Omega_{n1}}{\partial\phi_i})$, $\Psi_{n,ij} = -\frac{1}{2}tr(\frac{\partial\Omega_{n1}^{-1}}{\partial\phi_i}\frac{\partial\Omega_{n1}}{\partial\phi_j})$ for $i, j = 1,\ldots,7$, and $E\frac{\partial\xi'_n}{\partial\rho}\Omega_{n1}^{-1}\frac{\partial\xi_n}{\partial\rho} = \frac{tr(B'_{n2}S'^{-1}_{n2}\Omega_{n1}^{-1}S_{n2}^{-1}B_{n2})}{[T-(T-2)\rho_0]^4}\frac{2\sigma_{e0}^2}{(1-\rho_0^2)}[T^2 + T^2\rho_0^{T-1} - 2T + \frac{4\rho_0}{(1-\rho_0)}(T\rho_0^{T-1} - \frac{1-\rho_0^T}{1-\rho_0})]$.

B.3.2. *Steps to obtain* $E\frac{\partial^2\ln L_{b,n}(\theta_0)}{\partial^2\rho}$ *in the information matrix.* For $-E\frac{\partial^2\ln L_{b,n}(\theta_0)}{\partial^2\rho}$, (i) by using $\frac{\partial\Omega_{n1}^{-1}}{\partial\rho} = -\Omega_{n1}^{-1}\frac{\partial\Omega_{n1}}{\partial\rho}\Omega_{n1}^{-1}$ and $E\boldsymbol{\xi}_n\boldsymbol{\xi}'_n = \Omega_{n1}$, we have $E(-\frac{1}{2}tr(\frac{\partial\Omega_{n1}^{-1}}{\partial\rho}\frac{\partial\Omega_{n1}}{\partial\rho} + \Omega_{n1}^{-1}\frac{\partial^2\Omega_{n1}}{\partial^2\rho}) - \frac{1}{2}\boldsymbol{\xi}'_n\frac{\partial^2\Omega_{n1}^{-1}}{\partial^2\rho}\boldsymbol{\xi}_n) = -\frac{1}{2}tr\frac{\partial\Omega_{n1}^{-1}}{\partial\rho}\frac{\partial\Omega_{n1}}{\partial\rho}$. (ii) From $E\frac{\partial\ln L_{b,n}}{\partial\rho} = 0$, we have $E\frac{\partial\boldsymbol{\xi}'_n}{\partial\rho}\Omega_{n1}^{-1}\boldsymbol{\xi}_n = 0$. As $\frac{\partial\boldsymbol{\xi}_n}{\partial\rho} = \frac{S_{n2}^{-1}B_{n2}}{[T-(T-2)\rho_0]^2}[T(V_{n1} + V_{nT}) - 2\sum_{t=1}^T V_{nt}]$ and $\frac{\partial^2\boldsymbol{\xi}_n}{\partial\rho^2} = \frac{2(T-2)}{[T-(T-2)\rho_0]}\frac{\partial\boldsymbol{\xi}_n}{\partial\rho}$, $\frac{\partial^2\boldsymbol{\xi}_n}{\partial\rho^2}$ is proportional to $\frac{\partial\boldsymbol{\xi}_n}{\partial\rho}$, and hence $E\frac{\partial^2\boldsymbol{\xi}'_n}{\partial\rho^2}\Omega_{n1}^{-1}\boldsymbol{\xi}_n = 0$. (iii) By using $\frac{\partial\boldsymbol{\xi}_n}{\partial\rho}$ above, as $E\frac{\partial\boldsymbol{\xi}'_n}{\partial\rho}\Omega_{n1}^{-1}\frac{\partial\boldsymbol{\xi}_n}{\partial\rho} = tr[\Omega_{n1}^{-1}E(\frac{\partial\boldsymbol{\xi}_n}{\partial\rho}\frac{\partial\boldsymbol{\xi}'_n}{\partial\rho})]$, we have the results.

B.3.3. *Variance matrix of the score.* To compute the variance of the score for the between equation, $\boldsymbol{\xi}_n = \mu_n + S_{n2}^{-1}B_{n2}\vec{V}_{nT}$ can be written as $\boldsymbol{\xi}_n = S_{n3}^{-1}B_{n3}\mathbf{c}_{n0} + (\mathbf{p}'_T P_T^{-1} \otimes S_{n2}^{-1}B_{n2})\mathbf{e}_{nT}$, where $\mathbf{p}_T = (1, 1-\rho_0, \ldots, 1-\rho_0, 1)'$ and $\mathbf{p}'_T P_T^{-1} = l_T'^{\rho}$. Also, we have $\frac{\partial\boldsymbol{\xi}_n}{\partial\rho} = \frac{1}{(T-(T-2)\rho_0)^2}(\mathbf{q}'_T P_T^{-1} \otimes S_{n2}^{-1}B_{n2})\mathbf{e}_{nT}$, where $\mathbf{q}_T = (T-2, -2, \ldots, -2, T-2)'$. Thus, define

$$\begin{aligned}
\mathbb{C}_{c,\lambda_1} &= (S_{n3}^{-1}B_{n3})'G'_{n1}\Omega_{n1}^{-1}(S_{n3}^{-1}B_{n3}), \\
\mathbb{C}_{c,\phi_i} &= -\frac{1}{2}(S_{n3}^{-1}B_{n3})'\frac{\partial\Omega_{n1}^{-1}}{\partial\phi_i}(S_{n3}^{-1}B_{n3}) \quad \text{for} \quad i = 1,\ldots,7, \\
\mathbb{C}_{e,\lambda_1} &= (l_T'^{\rho} \otimes S_{n2}^{-1}B_{n2})'G_{n1}\Omega_{n1}^{-1}(l_T'^{\rho} \otimes S_{n2}^{-1}B_{n2}), \\
\mathbb{C}_{e,\phi_i} &= -\frac{1}{2}(l_T'^{\rho} \otimes S_{n2}^{-1}B_{n2})'\frac{\partial\Omega_{n1}^{-1}}{\partial\phi_i}(l_T'^{\rho} \otimes S_{n2}^{-1}B_{n2}), \quad \text{for} \quad \phi_i \neq \rho, \\
\mathbb{C}_{e,\phi_5} &= -\frac{1}{2}(l_T'^{\rho} \otimes S_{n2}^{-1}B_{n2})'\frac{\partial\Omega_{n1}^{-1}}{\partial\rho}(l_T'^{\rho} \otimes S_{n2}^{-1}B_{n2}) \\
&\quad - \frac{1}{[T-(T-2)\rho_0]^2}(\mathbf{q}'_T P_T^{-1} \otimes S_{n2}^{-1}B_{n2})'\Omega_{n1}^{-1}(l_T'^{\rho} \otimes S_{n2}^{-1}B_{n2}), \\
\mathbb{D}_{c,\gamma} &= (S_{n3}^{-1}B_{n3})'\Omega_{n1}^{-1}\vec{Z}_{nT}, \quad \mathbb{D}_{c,\lambda_1} = (S_{n3}^{-1}B_{n3})'\Omega_{n1}^{-1}G_{n1}\vec{Z}_{nT}\gamma_0, \\
\mathbb{D}_{e,\gamma} &= l_T^{\rho} \otimes (S_{n2}^{-1}B_{n2})'\Omega_{n1}^{-1}\vec{Z}_{nT}, \quad \mathbb{D}_{e,\lambda_1} = l_T^{\rho} \otimes (S_{n2}^{-1}B_{n2})'\Omega_{n1}^{-1}G_{n1}\vec{Z}_{nT}\gamma_0,
\end{aligned}$$

$\mathbb{Q}_c = [Q_{c,\lambda_1}, \mathbb{Q}_{c,\phi}]$, and $\mathbb{Q}_e = [Q_{e,\lambda_1}, \mathbb{Q}_{e,\phi}]$, where $\mathbb{Q}_{c,\lambda_1} = vec_D(\mathbb{C}_{c,\lambda_1})$, $\mathbb{Q}_{c,\phi} = [vec_D(\mathbb{C}_{c,\phi_1}), \ldots, vec_D(\mathbb{C}_{c,\phi_7})]$, $\mathbb{Q}_{e,\lambda_1} = vec_D(\mathbb{C}_{e,\lambda_1})$, and $\mathbb{Q}_{e,\phi} = [vec_D(\mathbb{C}_{e,\phi_1}), \ldots, vec_D(\mathbb{C}_{e,\phi_7})]$. We have $E\frac{1}{n}\frac{\partial\ln L_{b,n}(\theta_0)}{\partial\theta}\frac{\partial\ln L_{b,n}(\theta_0)}{\partial\theta'} = \Sigma_{b,n} + \Omega_{b,n}$, where $\Sigma_{b,n}$ is the information matrix and

$$\text{(B.6)} \quad \Omega_{b,n} = \begin{bmatrix} \mathbf{0}_{k\times k} & * \\ \mathbf{0}_{8\times k} & \mu_{4,c}\mathbb{Q}'_c\mathbb{Q}_c + \mu_{4,e}\mathbb{Q}'_e\mathbb{Q}_e \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{k\times k} & * & * \\ \mu_{3,c}\mathbb{Q}'_{c,\lambda_1}\mathbb{D}_{c,\gamma} + \mu_{3,e}\mathbb{Q}'_{e,\lambda_1}\mathbb{D}_{e,\gamma} & 2\mu_{3,c}\mathbb{Q}'_{c,\lambda_1}\mathbb{D}_{c,\lambda_1} + 2\mu_{3,e}\mathbb{Q}_{e,\lambda_1}\mathbb{D}_{e,\lambda_1} & * \\ \mu_{3,c}\mathbb{Q}'_{c,\phi}\mathbb{D}_{c,\gamma} + \mu_{3,e}\mathbb{Q}'_{e,\phi}\mathbb{D}_{e,\gamma} & \mu_{3,c}\mathbb{Q}'_{c,\phi}\mathbb{D}_{c,\lambda_1} + \mu_{3,e}\mathbb{Q}'_{e,\phi}\mathbb{D}_{e,\lambda_1} & \mathbf{0}_{7\times 7} \end{bmatrix}.$$

B.3.4. *Covariance matrix of the scores for the between and within equations*. The covariance matrix of the scores for the between and within equations is

$$\text{(B.7)} \qquad \Omega_{wb,nT} = \frac{\mu_{3,e}}{\sigma_0^2}\left[\begin{pmatrix} \mathbf{0}_{k_x\times(k+1)} & \mathbf{0}_{k_x\times 7} \\ \mathcal{P}'[\mathbb{D}_{e,\gamma}, \mathbb{D}_{e,\lambda_1}] & \mathbf{0}_{5\times 7}\end{pmatrix} + \begin{pmatrix} \mathbf{0}_{(k_x+1)\times k} & [\mathbb{D}_{e,\gamma}, \mathbb{D}_{e,\lambda_1}]'\mathbb{Q}_e \\ \mathbf{0}_{4\times k} & \mathbf{0}_{4\times 8}\end{pmatrix}\right] + \frac{\mu_{4,e} - 3\sigma_{e0}^4}{\sigma_{e0}^2}\begin{bmatrix} \mathbf{0}_{k_x\times k} & \mathbf{0}_{k_x\times 8} \\ \mathbf{0}_{5\times k} & \mathcal{P}'\mathbb{Q}_e\end{bmatrix}.$$

C. *Appendix for section 3.*

In this appendix, we derive the variance matrix of $\sqrt{n}(\hat\theta_{r1} - \hat\theta_{w1})$ when the disturbances are not normal.

C.1. *The general case*. From the Taylor expansion of the random effects estimate, we have $\sqrt{n}(\hat\theta_{r1} - \theta_{10}) = J'\Sigma_{r,nT}^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_{r,nT}(\theta_0)}{\partial\theta} + o_p(1)$. By (15) and

$$\text{(C.1)} \qquad \begin{pmatrix} \dfrac{\partial \ln L_{r,nT}}{\partial\theta_1} \\ \dfrac{\partial \ln L_{r,nT}}{\partial\theta_2}\end{pmatrix} = \begin{pmatrix} \dfrac{\partial \ln L_{w,nT}}{\partial\theta_1} + \dfrac{\partial \ln L_{b,n}}{\partial\theta_1} \\ \dfrac{\partial \ln L_{b,n}}{\partial\theta_2}\end{pmatrix},$$

we have $\sqrt{n}(\hat\theta_{r1} - \theta_{10}) = (\Sigma_{w,nT} + C)^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_{w,nT}(\theta_0)}{\partial\theta_1} + J'\Sigma_{r,nT}^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_{b,n}(\theta_0)}{\partial\theta} + o_p(1)$. Also, $\sqrt{n}(\hat\theta_{w1} - \theta_{10}) = \Sigma_{w,nT}^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_{w,nT}(\theta_{10})}{\partial\theta_1} + o_p(1)$. Hence, $\sqrt{n}(\hat\theta_{r1} - \hat\theta_{w1}) = \frac{1}{\sqrt{n}}A_{nT}'(\frac{\partial \ln L_{w,nT}}{\partial\theta_1'}, \frac{\partial \ln L_{b,n}}{\partial\theta'})' + o_p(1)$. Thus,

$$\begin{aligned}
\operatorname{Var}[\sqrt{n}(\hat\theta_{r1} - \hat\theta_{w1})] &= A_{nT}'\begin{bmatrix} \Sigma_{w,nT} + \Omega_{w,nT} & \Omega_{wb,nT} \\ \Omega_{wb,nT}' & \Sigma_{b,n} + \Omega_{b,n}\end{bmatrix} A_{nT} \\
&= A_{nT}'\left(\begin{bmatrix} \Sigma_{w,nT} & \mathbf{0}_{k_{\theta_1}\times k} \\ \mathbf{0}_{k\times k_{\theta_1}} & \Sigma_{b,n}\end{bmatrix} + \begin{bmatrix} \Omega_{w,nT} & \Omega_{wb,nT} \\ \Omega_{wb,nT}' & \Omega_{b,n}\end{bmatrix}\right) A_{nT} \\
&= B^{-1} - (B + C)^{-1} + A_{nT}'\begin{bmatrix} \Omega_{w,nT} & \Omega_{wb,nT} \\ \Omega_{wb,nT}' & \Omega_{b,n}\end{bmatrix} A_{nT}.
\end{aligned}$$

C.2. *The KKP model case*. For the KKP case, we compare estimates of θ_1^* under the within and random specifications. For the random effects model, by reparameterization, the parameters to be estimated are θ_{10}^*, σ_{e0}^2, and θ_{20}, where $\theta_1^* = (\beta', \lambda_1, \lambda_2, \delta_2, \rho)'$ and $\theta_2 = (b', \sigma_*^2)'$. Similar to the argument above, we have $\sqrt{n}(\hat\theta_{r1}^* - \hat\theta_{w1}^*) = \frac{1}{\sqrt{n}}A_{nT}^{*\prime}(\frac{\partial \ln L_{w,nT}}{\partial\theta_1'}, \frac{\partial \ln L_{b,n}}{\partial\theta'})' + o_p(1)$. Therefore, the asymptotic variance of $\sqrt{n}(\hat\theta_{r1}^* - \hat\theta_{w1}^*)$ is

$$A_{nT}^{*\prime}\begin{bmatrix} E\frac{1}{n}\frac{\partial \ln L_{w,nT}(\theta_0)}{\partial\theta_1}\frac{\partial \ln L_{w,nT}(\theta_0)}{\partial\theta_1'} & E\frac{1}{n}\frac{\partial \ln L_{w,nT}(\theta_0)}{\partial\theta_1}\frac{\partial \ln L_{b,n}(\theta_0)}{\partial\theta'} \\ E\frac{1}{n}\frac{\partial \ln L_{b,n}(\theta_0)}{\partial\theta_1'}\frac{\partial \ln L_{w,nT}(\theta_0)}{\partial\theta} & E\frac{1}{n}\frac{\partial \ln L_{b,n}(\theta_0)}{\partial\theta}\frac{\partial \ln L_{b,n}(\theta_0)}{\partial\theta'} \end{bmatrix} A_{nT}^{*}.$$

D. *Proofs for Propositions.*

D.1. *Proof for Proposition 1.*

D.1.1. *Uniform convergence.* From the DGP, denoting $D_n(\theta_1) = B_{n2}^{-1}(\delta_2) S_{n2}(\lambda_2)S_{n1}(\lambda_1)S_{n1}^{-1}S_{n2}^{-1}B_{n2}$, we have

$$\begin{aligned}\mathbf{V}_{nT}(\theta_1) &= \big(I_T\otimes B_{n2}^{-1}(\delta_2)S_{n2}(\lambda_2)\big)\big[\big(I_T\otimes S_{n1}(\lambda_1)S_{n1}^{-1}\big)\mathbf{X}_{nT}\beta_0-\mathbf{X}_{nT}\beta\big]\\ &\quad + l_T\otimes B_{n2}^{-1}(\delta_2)S_{n2}(\lambda_2)\big(S_{n1}(\lambda_1)S_{n1}^{-1}\mathbf{c}_{n0}-\mathbf{c}_n\big)+(I_T\otimes D_n(\theta_1))\mathbf{V}_{nT}.\end{aligned}$$

Because $l_T\otimes B_{n2}^{-1}(\delta_2)S_{n2}(\lambda_2)(S_{n1}(\lambda_1)S_{n1}^{-1}\mathbf{c}_{n0}-\mathbf{c}_n)$ is eliminated by $\mathbb{J}_T(\rho)\otimes I_n$, by denoting $\mathbf{b}_{nT}(\theta_1) = (I_T\otimes B_{n2}^{-1}(\delta_2)S_{n2}(\lambda_2))[(I_T\otimes S_{n1}(\lambda_1)S_{n1}^{-1})\mathbf{X}_{nT}\beta_0-\mathbf{X}_{nT}\beta]$,

$$\begin{aligned}&\mathbf{V}_{nT}'(\theta_1)(\mathbb{J}_T(\rho)\otimes I_n)\mathbf{V}_{nT}(\theta_1)-E\mathbf{V}_{nT}'(\theta_1)(\mathbb{J}_T(\rho)\otimes I_n)\mathbf{V}_{nT}(\theta_1)\\ &\quad=\mathbf{V}_{nT}'(\mathbb{J}_T(\rho)\otimes D_n'(\theta_1)D_n(\theta_1))\mathbf{V}_{nT}'-E\mathbf{V}_{nT}'(\mathbb{J}_T(\rho)\otimes D_n'(\theta_1)D_n(\theta_1))\mathbf{V}_{nT}'\\ &\qquad+2\mathbf{b}_{nT}'(\mathbb{J}_T(\rho)\otimes D_n(\theta_1))\mathbf{V}_{nT}.\end{aligned}$$

As $\mathbf{V}_{nT} = (P_T^{-1}\otimes I_n)\mathbf{e}_{nT}$, where elements of $\mathbf{e}_{nT}$ are independent $(0,\sigma_{e0}^2)$, $D_n(\theta_1)$ is UB in n, and $P_T'^{-1}\mathbb{J}_T(\rho)P_T^{-1}$ is UB in T, as both $\mathbb{J}_T(\rho)$ and P_T^{-1} are UB in T, we have $\frac{1}{n(T-1)}\ln L_{w,nT}(\theta_1) - E\frac{1}{n(T-1)}\ln L_{w,nT}(\theta_1)\xrightarrow{p} 0$ uniformly in $\theta_1\in\Theta_1$ (see Lee, 2004).

D.1.2. *Uniform equicontinuity.* As $S_{n1}(\lambda_1)S_{n1}^{-1} = I_n-(\lambda_1-\lambda_{10})G_{n1}$, by denoting $\mathcal{H}_{nT}(\lambda_2,\delta_2,\rho) = (\mathbf{X}_{nT},\mathbf{G}_{nT,1}\mathbf{X}_{nT}\beta_0)'(\mathbb{J}_T(\rho)\otimes S_{n2}'(\lambda_2)B_{n2}'^{-1}(\delta_2)B_{n2}^{-1}(\delta_2)S_{n2}(\lambda_2))(\mathbf{X}_{nT},\mathbf{G}_{nT,1}\mathbf{X}_{nT}\beta_0)$ and

$$p_{n,T-1}(\lambda_1,\lambda_2,\delta_2,\rho) = \frac{1}{n(T-1)}\sigma_{e0}^2 tr\big(P_T'^{-1}\mathbb{J}_T(\rho)P_T^{-1}\big)\cdot tr(D_n'(\theta_1)D_n(\theta_1)), \tag{D.1}$$

we have

(D.2)

$$\begin{aligned}&E\frac{1}{2\sigma_e^2}\frac{1}{n(T-1)}\mathbf{V}_{nT}'(\theta_1)(\mathbb{J}_T(\rho)\otimes I_n)\mathbf{V}_{nT}(\theta_1)\\ &\quad=\frac{1}{2\sigma_e^2}\frac{1}{n(T-1)}(\beta'-\beta_0',\lambda_1-\lambda_{10})\mathcal{H}_{nT}(\lambda_2,\delta_2,\rho)(\beta'-\beta_0',\lambda_1-\lambda_{10})'+\frac{1}{2\sigma_e^2}p_{n,T-1}(\lambda_1,\lambda_2,\delta_2,\rho).\end{aligned}$$

These terms are all polynomial functions of θ_1. Therefore, $E\frac{1}{n(T-1)}\ln L_{w,nT}(\theta_1)$ is uniformly equicontinuous on any bounded set of θ_1.

D.1.3. *Identification uniqueness.* By (D.2), $E\frac{1}{n(T-1)}\ln L_{w,nT}(\theta_1) - E\frac{1}{n(T-1)}\ln L_{w,nT}(\theta_{10}) = T_{1,n,T-1}(\lambda_1,\lambda_2,\delta_2,\rho,\sigma_e^2) - \frac{1}{2\sigma_e^2}T_{2,n,T-1}(\beta,\lambda_1,\lambda_2,\delta_2,\rho)$, where $T_{2,n,T-1}(\beta,\lambda_1,\lambda_2,\delta_2,\rho) = \frac{1}{n(T-1)}(\beta'-\beta_0',\lambda_1-\lambda_{10})\mathcal{H}_{nT}(\lambda_2,\delta_2,\rho)(\beta'-\beta_0',\lambda_1-\lambda_{10})'$ and

$$
\begin{aligned}
T_{1,n,T-1}\big(\lambda_1,\lambda_2,\delta_2,\rho,\sigma_e^2\big) = & -\frac{1}{2}\big(\ln\sigma_e^2-\ln\sigma_{e0}^2\big)\\
& +\frac{1}{n}[\ln|S_{n1}(\lambda_1)|+\ln|S_{n2}(\lambda_2)|-\ln|S_{n1}(\lambda_{10})|-\ln|S_{n2}(\lambda_{20})|]\\
& -\frac{1}{n}[\ln|B_{n2}(\delta_2)|-\ln|B_{n2}(\delta_{20})|]-\frac{1}{2(T-1)}(\ln|H_{T-1}(\rho)|-\ln|H_{T-1}|)\\
& -\frac{1}{2\sigma_e^2}\big(p_{n,T-1}(\lambda_1,\lambda_2,\delta_2,\rho)-\sigma_e^2\big).
\end{aligned}
$$

Consider a model without exogenous variables; after the first difference and quasi-difference, it is

$$
\begin{aligned}
&\Delta Y_{nt,\rho_0} = \lambda_1 W_{n1}\Delta Y_{nt,\rho_0}+\Delta U_{nt,\rho_0} \quad \text{with} \quad \Delta U_{nt,\rho_0}=\lambda_2 W_{n2}\Delta U_{nt,\rho_0}+B_{n2}\Delta e_{nt}, \quad t=3,\dots,T,\\
&\Delta Y_{n2} = \lambda_1 W_{n1}\Delta Y_{n2}+\Delta U_{n2} \quad \text{with} \quad \Delta U_{n2}=\lambda_2 W_{n2}\Delta U_{n2}+B_{n2}\Delta V_{n2} \quad \text{and}\\
&\Delta V_{n2} = e_{n2}-(1-\rho_0)V_{n1},
\end{aligned}
$$

where $Z_{nt,\rho_0}\equiv Z_{nt}-\rho_0 Z_{n,t-1}$ for any $n\times 1$ vector Z_{nt}. Its log-likelihood function is

$$
\begin{aligned}
\ln L_{w,nT}^{p}(\theta_1) = & -\frac{n(T-1)}{2}\ln 2\pi\\
& -\frac{n(T-1)}{2}\ln\sigma_e^2+(T-1)[\ln|S_{n1}(\lambda_1)|+\ln|S_{n2}(\lambda_2)|-\ln|B_{n2}(\delta_2)|]\\
& -\frac{n}{2}\ln|H_{T-1}(\rho)|-\frac{1}{2\sigma_e^2}\mathbf{V}_{nT}'(\theta_1)(\mathbb{J}_T(\rho)\otimes I_n)\mathbf{V}_{nT}(\theta_1).
\end{aligned}
$$

Using the information inequality from this model without exogenous variables, $T_{1,n,T-1}(\lambda_1,\lambda_2,\delta_2,\rho,\sigma_e^2)\le 0$ for any $(\lambda_1,\lambda_2,\delta_2,\rho,\sigma_e^2)$. Also, $T_{2,n,T-1}(\beta,\lambda_1,\lambda_2,\delta_2,\rho)$ is a quadratic function of β and λ_1 with a positive semidefinite matrix given λ_2, δ_2, and ρ. When

(D.3) $$\lim_{n\to\infty}\frac{1}{n(T-1)}\mathcal{H}_{nT}(\lambda_2,\delta_2,\rho) \text{ is nonsingular given any value of } \lambda_2,\delta_2, \text{ and } \rho,$$

then $T_{2,n,T-1}(\beta,\lambda_1,\lambda_2,\delta_2,\rho)>0$ given any λ_2, δ_2, and ρ whenever $(\beta,\lambda_1)\neq(\beta_0,\lambda_{10})$. Hence, (β_0,λ_{10}) is identified. Given λ_{10}, then λ_{20}, δ_{20}, ρ_0, and σ_{e0}^2 give the unique maximizer of $\lim_{n\to\infty}T_{1,n,T-1}(\lambda_{10},\lambda_2,\delta_2,\rho,\sigma_e^2)$ if

(D.4) $$
\begin{aligned}
&\lim_{n\to\infty}\bigg(\frac{1}{T-1}\ln|H_{T-1}|-\frac{1}{T-1}\ln\big|H_{T-1}(\rho)\big|+\frac{1}{n}\ln\Big|\sigma_{e0}^2B_{n2}'S_{n2}'^{-1}S_{n2}^{-1}B_{n2}\Big|\\
&\quad -\frac{1}{n}\ln\big|p_{n,T-1}(\lambda_2,\delta_2,\rho)B_{n2}'(\delta_2)S_{n2}^{-1}(\lambda_2)'S_{n2}^{-1}(\lambda_2)B_{n2}(\delta_2)\big|\bigg)\\
&\neq 0 \quad \text{for} \quad (\lambda_2,\delta_2,\rho)\neq(\lambda_{20},\delta_{20},\rho_0),
\end{aligned}
$$

where $p_{n,T-1}(\lambda_2,\delta_2,\rho)=p_{n,T-1}(\lambda_{10},\lambda_2,\delta_2,\rho)$. When $\lim_{n\to\infty}\frac{1}{n(T-1)}\mathcal{H}_{nT}(\lambda_2,\delta_2,\rho)$ is singular, β_0 and λ_{10} cannot be identified from $T_{2,n,T-1}(\beta,\lambda_1,\lambda_2,\delta_2,\rho)$. Identification requires that $\lim_{n\to\infty}T_{1,n,T-1}(\lambda_1,\lambda_2,\delta_2,\rho,\sigma_e^2)$ is strictly less than zero, which is equivalent to

$$\text{(D.5)}\quad \lim_{n\to\infty}\left(\frac{1}{T-1}\ln|H_{T-1}|-\frac{1}{T-1}\ln|H_{T-1}(\rho)|+\frac{1}{n}\ln\left|\sigma_{e0}^2B_{n2}'S_{n2}'^{-1}S_{n1}'^{-1}S_{n1}^{-1}S_{n2}^{-1}B_{n2}\right|\right)$$
$$-\lim_{n\to\infty}\left(\frac{1}{n}\ln\left|p_{n,T-1}(\lambda_1,\lambda_2,\delta_2,\rho)B_{n2}'(\delta_2)S_{n2}^{-1}(\lambda_2)'S_{n1}'^{-1}(\lambda_1)'S_{n1}^{-1}(\lambda_1)S_{n2}^{-1}(\lambda_2)B_{n2}(\delta_2)\right|\right)$$
$$\neq 0 \quad \text{for} \quad (\lambda_1,\lambda_2,\delta_2,\rho)\neq(\lambda_{10},\lambda_{20},\delta_{20},\rho_0).$$

When λ_{10} is identified, β_0 can be identified from $T_{2,n,T-1}(\beta,\lambda_1,\lambda_2,\delta_2,\rho)$.

Combined with uniform convergence and equicontinuity, the consistency follows.

D.1.4. *Asymptotic normality.* The score $\frac{\partial \ln L_{w,nT}(\theta_{10})}{\partial\theta_1}$ in Appendix B.1 has linear and quadratic forms of independent disturbances. As T is finite, the CLT for the linear-quadratic form in Kelejian and Prucha (2001) is applicable so that $\frac{1}{\sqrt{n(T-1)}}\frac{\partial \ln L_{w,nT}(\theta_{10})}{\partial\theta_1}\xrightarrow{d}N(0,\Sigma_{w,nT}+\Omega_{w,nT})$. By $-\frac{1}{n(T-1)}\frac{\partial^2\ln L_{w,nT}(\bar\theta_{w1})}{\partial\theta_1\partial\theta_1'}-\Sigma_{w,nT}\xrightarrow{p}0$ from (38)–(41) in Yu et al. (2008), where $\bar\theta_{w1}$ lies between $\hat\theta_{w1}$ and θ_{10}, the asymptotic distribution of the estimator follows from the Taylor expansion, $\sqrt{n(T-1)}(\hat\theta_{w1}-\theta_{10})=(-\frac{1}{n(T-1)}\frac{\partial^2\ln L_{w,nT}(\bar\theta_{w1})}{\partial\theta_1\partial\theta_1'})^{-1}\frac{1}{\sqrt{n(T-1)}}\frac{\partial \ln L_{w,nT}(\theta_{10})}{\partial\theta_1}$.

D.2. *Proof for Proposition 2.*

D.2.1. *Uniform convergence.* Regarding the likelihood in (9), we have $\boldsymbol{\xi}_{nT}(\theta)=\mathbf{S}_{nT,1}(\lambda_1)\mathbf{S}_{nT,1}^{-1}\mathbf{Z}_{nT}\gamma_0-\mathbf{Z}_{nT}\gamma+\mathbf{S}_{nT,1}(\lambda_1)\mathbf{S}_{nT,1}^{-1}\boldsymbol{\xi}_{nT}$, where $\mathbf{S}_{nT,1}(\lambda_1)=I_T\otimes S_{n1}(\lambda_1)$ and $\boldsymbol{\xi}_{nT}=l_T\otimes S_{n3}^{-1}B_{n3}\mathbf{c}_{n0}+(P_T^{-1}\otimes S_{n2}^{-1}B_{n2})\mathbf{e}_{nT}$. Hence, we have

$$\begin{aligned}&\frac{1}{nT}\boldsymbol{\xi}_{nT}'(\theta)\Omega_{nT}^{-1}(\phi)\boldsymbol{\xi}_{nT}(\theta)-E\frac{1}{nT}\boldsymbol{\xi}_{nT}'(\theta)\Omega_{nT}^{-1}(\phi)\boldsymbol{\xi}_{nT}(\theta)\\&\quad=\frac{1}{nT}\big(\mathbf{S}_{nT,1}(\lambda_1)\mathbf{S}_{nT,1}^{-1}\xi_{nT}'\big)\Omega_{nT}^{-1}(\phi)\big(\mathbf{S}_{nT,1}(\lambda_1)\mathbf{S}_{nT,1}^{-1}\boldsymbol{\xi}_{nT}\big)-p_{nT}(\lambda_1,\phi)\\&\qquad+\frac{2}{nT}\big(\mathbf{S}_{nT,1}(\lambda_1)\mathbf{S}_{nT,1}^{-1}\mathbf{Z}_{nT}\gamma_0-\mathbf{Z}_{nT}\gamma\big)'\Omega_{nT}^{-1}(\phi)\big(\mathbf{S}_{nT,1}(\lambda_1)\mathbf{S}_{nT,1}^{-1}\xi_{nT}\big),\end{aligned}$$

where $p_{nT}(\lambda_1,\phi)=\frac{1}{nT}tr[(\mathbf{S}_{nT,1}(\lambda_1)\mathbf{S}_{nT,1}^{-1})'\Omega_{nT}^{-1}(\phi)\mathbf{S}_{nT,1}(\lambda_1)\mathbf{S}_{nT,1}^{-1}\Omega_{nT}]$. As $\Omega_{nT}^{-1}(\phi)$, P_T^{-1}, and relevant spatial matrices are UB in T and n, $\frac{1}{nT}(\mathbf{S}_{nT,1}(\lambda_1)\mathbf{S}_{nT,1}^{-1}\boldsymbol{\xi}_{nT})'\Omega_{nT}^{-1}(\phi)\mathbf{S}_{nT,1}(\lambda_1)\mathbf{S}_{nT,1}^{-1}\boldsymbol{\xi}_{nT}-p_{nT}(\lambda_1,\phi)\xrightarrow{p}0$ uniformly in $\theta\in\Theta$ when T is either finite or large. As $\frac{1}{nT}(\mathbf{S}_{nT,1}(\lambda_1)\mathbf{S}_{nT,1}^{-1}\mathbf{Z}_{nT}\gamma_0-\mathbf{Z}_{nT}\gamma)'\Omega_{nT}^{-1}(\phi)(\mathbf{S}_{nT,1}(\lambda_1)\mathbf{S}_{nT,1}^{-1}\boldsymbol{\xi}_{nT})\xrightarrow{p}0$ for either finite or large T, $\frac{1}{nT}\ln L_{r,nT}(\theta)-E\frac{1}{nT}\ln L_{r,nT}(\theta)\xrightarrow{p}0$ uniformly in $\theta\in\Theta$.

D.2.2. *Uniform equicontinuity.* We have $E\frac{1}{nT}\ln L_{r,nT}(\theta)=-\frac{1}{2}\ln 2\pi-\frac{1}{2nT}\ln|\Omega_{nT}(\phi)|+\frac{1}{n}\ln|S_{n1}(\lambda_1)|-\frac{1}{2nT}E\boldsymbol{\xi}_{nT}'(\theta)\Omega_{nT}^{-1}(\phi)\xi_{nT}(\theta)$. Denoting $\mathcal{H}_{nT}(\phi)=(\mathbf{Z}_{nT},\mathbf{G}_{nT,1}\mathbf{Z}_{nT}\gamma_0)'\Omega_{nT}^{-1}(\phi)\times(\mathbf{Z}_{nT},\mathbf{G}_{nT,1}\mathbf{Z}_{nT}\gamma_0)$, it follows that

$$\text{(D.6)}\quad E\frac{1}{nT}\boldsymbol{\xi}_{nT}'(\theta)\Omega_{nT}^{-1}(\phi)\boldsymbol{\xi}_{nT}(\theta)=\frac{1}{nT}(\gamma'-\gamma_0',\lambda_1-\lambda_{10})\mathcal{H}_{nT}(\phi)(\gamma'-\gamma_0',\lambda_1-\lambda_{10})'+p_{nT}(\lambda_1,\phi),$$

which is a polynomial function of θ. Thus, $E\frac{1}{nT}\ln L_{r,nT}(\theta)$ is uniformly equicontinuous in $\theta\in\Theta$.

D.2.3. *Identification uniqueness.* By (D.6), $E\frac{1}{nT}\ln L_{r,nT}(\theta) - E\frac{1}{nT}\ln L_{r,nT}(\theta_0) = T_{1,nT}(\lambda_1, \phi) - \frac{1}{2}T_{2,nT}(\gamma,\lambda_1,\phi)$, where

$$T_{1,nT}(\lambda_1,\phi) = -\frac{1}{2nT}\ln|\Omega_{nT}(\phi)| + \frac{1}{n}\ln|S_{n1}(\lambda_1)| - \frac{1}{2}p_{nT}(\lambda_1,\phi) - \left(-\frac{1}{2nT}\ln|\Omega_{nT}| + \frac{1}{n}\ln|S_{n1}| - \frac{1}{2}\right)$$

with $T_{2,nT}(\gamma,\lambda_1,\phi) = \frac{1}{nT}(\gamma'-\gamma_0', \lambda_1-\lambda_{10})\mathcal{H}_{nT}(\phi)(\gamma'-\gamma_0', \lambda_1-\lambda_{10})'$. Consider a random effects model without exogenous variables $Y_{nt} = \lambda_{10}W_{n1}Y_{nt} + \mu_n + U_{nt}$, $t = 1,\ldots,T$, where the log-likelihood function is

$$\ln L^p_{r,nT}(\theta) = -\frac{nT}{2}\ln 2\pi - \frac{1}{2}\ln|\Omega_{nT}(\phi)| + T\ln|S_{n1}(\lambda_1)| - \frac{1}{2nT}\boldsymbol{\xi}'_{nT}(\theta)\Omega^{-1}_{nT}(\phi)\boldsymbol{\xi}_{nT}(\theta),$$

with $\boldsymbol{\xi}_{nT}(\theta) = \mathbf{S}_{nT,1}(\lambda_1)\mathbf{Y}_{nT}$. Using the information inequality for this model, $T_{1,nT}(\lambda_1,\phi)\le 0$ for any (λ_1,ϕ). Also, $T_{2,nT}(\gamma,\lambda_1,\phi)$ is a quadratic function of (γ,λ_1) with a positive semidefinite matrix given ϕ. When

$$\lim_{n,T\to\infty}\frac{1}{nT}\mathcal{H}_{nT}(\phi) \text{ is nonsingular given any value of } \phi, \tag{D.7}$$

then $T_{2,nT}(\gamma,\lambda_1,\phi) > 0$ given any ϕ whenever $(\gamma,\lambda_1)\neq(\gamma_0,\lambda_{10})$. Hence, (γ_0,λ_{10}) is identified. Given λ_{10}, ϕ_0 is the unique maximizer of $\lim_{n,T\to\infty}T_{1,nT}(\lambda_1,\phi)$ if

$$\lim_{n,T\to\infty}\left(\frac{1}{nT}\ln|\Omega_{nT}| + 1 - \left(\frac{1}{nT}\ln|\Omega_{nT}(\phi)| + p_{nT}(\phi)\right)\right) \neq 0 \quad \text{for} \quad \phi\neq\phi_0, \tag{D.8}$$

where $p_{nT}(\phi) = \frac{1}{nT}tr[\Omega^{-1}_{nT}(\phi)\Omega_{nT}]$. When $\lim_{n,T\to\infty}\frac{1}{nT}\mathcal{H}_{nT}(\phi)$ is singular, γ_0 and λ_{10} cannot be identified from $T_{2,nT}(\gamma,\lambda_1,\phi)$. Identification requires that $\lim_{n,T\to\infty}T_{1,nT}(\lambda_1,\phi)$ be strictly less than zero, which is equivalent to

$$\lim_{n,T\to\infty}\left(\frac{1}{nT}\ln\left|\mathbf{S}'^{-1}_{nT,1}\Omega_{nT}\mathbf{S}^{-1}_{nT,1}\right| + 1 - \left(\frac{1}{nT}\ln\left|\mathbf{S}^{-1}_{nT,1}(\lambda_1)\Omega_{nT}(\phi)\mathbf{S}'^{-1}_{nT,1}(\lambda_1)\right| + p_{nT}(\lambda_1,\phi)\right)\right) \neq 0 \quad \text{for} \quad (\lambda_1,\phi)\neq(\lambda_{10},\phi_0). \tag{D.9}$$

When λ_{10} is identified, γ_0 can be identified from $T_{2,nT}(\gamma,\lambda_1,\phi)$.

Combined with uniform convergence and equicontinuity, the consistency follows.

D.2.4. *Distribution.* The score is a linear and quadratic form of disturbances in $\boldsymbol{\xi}_{nT}$. Compared to the within equation case, we have two components $\mathbf{c}_{n0}$ and e_{nt}'s in $\boldsymbol{\xi}_{nT}$, where $\mathbf{c}_{n0}$ is time invariant and e_{nt} is time variant. By using the CLT for $\mathbf{c}_{n0}$ and e_{nt}'s (see Appendix B.2), where $\mathbf{c}_{n0}$ could be considered as e_{n0}, the score will asymptotically normal such that $\frac{1}{\sqrt{nT}}\frac{\partial \ln L_{r,nT}(\theta_0)}{\partial\theta} \xrightarrow{d} N(0, \Sigma_{r,nT} + \Omega_{r,nT})$, where $\Omega_{r,nT} = 0$ under normality. Combined with the consistency of $\hat{\theta}_r$ so that $-\frac{1}{nT}\frac{\partial^2 \ln L_{r,nT}(\bar{\theta})}{\partial\theta\partial\theta'} \xrightarrow{p} \Sigma_{r,nT}$, the distribution of $\hat{\theta}_r$ can be obtained from the Taylor expansion.

REFERENCES

AMEMIYA T., *Advanced Econometrics* (Cambridge, MA: Harvard University Press, 1985).

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

———, J. LE GALLO, AND H. JAYET, "Spatial Panel Econometrics," in L. Mátyás and P. Sevestre, eds., *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice* (Berlin and Heidelberg: Springer, 2008), 625–60.

ARELLANO, M., "On the Testing of Correlated Effects with Panel Data," *Journal of Econometrics* 59 (1993), 87–97.

BALTAGI, B., *Econometric Analysis of Panel Data* (Chichester, UK: John Wiley & Sons, Ltd., 2008).

——, AND Q. LI, "A Transformation that Will Circumvent the Problem of Autocorrelation in an Error-Component Model," *Journal of Econometrics* 48 (1991), 385–93.

———, G. BRESSON, AND A. PIROTTE, "Forecasting with Spatial Panel Data," *Computational Statistics and Data Analysis* 56 (2012), 3381–97.

———, P. EGGER, AND M. PFAFFERMAYR, "A Generalized Spatial Panel Data Model with Random Effects," Working Paper, Syracuse University, 2007a.

——, S. H. SONG, AND W. KOH, "Testing Panel Data Regression Models with Spatial Error Correlation," *Journal of Econometrics* 117 (2003), 123–50.

——, ———, B. C. JUNG, AND W. KOH, "Testing for Serial Correlation, Spatial Autocorrelation and Random Effects Using Panel Data," *Journal of Econometrics* 140 (2007b), 5–51.

BHARGAVA, A., L. FRANZINI, AND W. NARENDRANATHAN, "Serial Correlation and Fixed Effects Model," *Review of Economic Studies* 49 (1982), 533–49.

DHRYMES, P. J., *Introductory Econometrics* (New York: Springer Verlag, 1978).

ELHORST, J. P., "Unconditional Maximum Likelihood Estimation of Linear and Log-Linear Dynamic Models for Spatial Panels," *Geographical Analysis* 37 (2005), 85–106.

FINGLETON, B.,"A Generalized Method of Moments Estimators for a Spatial Panel Model with an Endogenous Spatial Lag and Spatial Moving Average Errors," *Spatial Economic Analysis* 3 (2008), 27–44.

HAUSMAN, J. A., "Specification Tests in Econometrics," *Econometrica* 46 (1978), 1251–71.

HSIAO, C., *Analysis of Panel Data* (New York: Cambridge University Press, 2003).

——, M. H. PESARAN, AND A. K. TAHMISCIOGLU, "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods," *Journal of Econometrics* 109 (2002), 107–50.

KAPOOR, M., H. H. KELEJIAN, AND I. R. PRUCHA, "Panel Data Models with Spatially Correlated Error Components," *Journal of Econometrics* 140 (2007), 97–130.

KELEJIAN, H. H., AND I. R. PRUCHA, "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) (1998), 99–121.

——, AND ——, "On the Asymptotic Distribution of the Moran I Test Statistic with Applications," *Journal of Econometrics* 104 (2001), 219–57.

——, AND ——, "Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances," *Journal of Econometrics* 157 (2010), 53–67.

KIEFER, N. M., "Estimation of Fixed Effects Models for Time Series of Cross Sections with Arbitrary Intertemporal Covariance," *Journal of Econometrics* 14 (1980), 195–202.

KORNIOTIS, G. M., "Estimating Panel Models with Internal and External Habit Formation," *Journal of Business and Economic Statistics* 28 (2010), 145–58.

LANCASTER, T., "The Incidental Parameter Problem since 1948," *Journal of Econometrics* 95 (2000), 391–413.

LEE, L. F., "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Econometric Models," *Econometrica* 72 (2004), 1899–925.

——, "GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models," *Journal of Econometrics* 137 (2007), 489–514.

——, AND J. YU, "Estimation of Spatial Autoregressive Panel Data Models with Fixed Effects," *Journal of Econometrics* 154 (2010), 165–85.

LESAGE, J., AND K. PACE, *Introduction to Spatial Econometrics* (Boca Raton, FL: CRC Press/Taylor & Francis Group, 2009).

LIN, X., AND L. F. LEE, "GMM Estimation of Spatial Autoregressive Models with Unknown Heteroskedasticity," *Journal of Econometrics* 157 (2010), 34–52.

MADDALA, G. S., "The Use of Variance Components Models in Pooling Cross Section and Time Series Data," *Econometrica* 39 (1971), 341–58.

MAGNUS, J. R., "Multivariate Error Components Analysis of Linear and Nonlinear Regression Models by Maximum Likelihood," *Journal of Econometrics* 19 (1982), 239–85.

MOSCONE, F., M. KNAPP, AND E. TOSETTI, “Mental Health Expenditure in England: A Spatial Panel Approach,” *Journal of Health Economics* 26 (2007), 842–64.

MUNDLAK, Y., “On Pooling Time Series and Cross Section Data,” *Econometrica* 46 (1978), 69–85.

MUTL, J., AND M. PFAFFERMAYR, “The Hausman Test in a Cliff and Ord Panel Model,” *Econometrics Journal* 14 (2011), 48–76.

NICKELL, S., “Biases in Dynamic Models with Fixed Effects,” *Econometrica* 49 (1981), 1417–26.

PARENT, O., AND J. P. LESAGE, “A Space-Time Filter for Panel Data Models Containing Random Effects,” Manuscript, University of Cincinnati, 2008.

RUUD, P. A., *An Introduction to Econometric Theory* (New York: Oxford University Press, 2000).

SU, L., AND Z. YANG, “QML Estimation of Dynamic Panel Data Models with Spatial Errors,” Mimeo, Singapore Management University, 2007.

YU, J., AND L. F. LEE, “Estimation of Unit Root Spatial Dynamic Panel Data Models,” *Econometric Theory* 26 (2010), 1332–62.

———, R. DE JONG, AND L. F. LEE, “Quasi-Maximum Likelihood Estimators for Spatial Dynamic Panel Data with Fixed Effects When Both n and T Are Large,” *Journal of Econometrics* 146 (2008), 118–34.

——, ——, AND ——, “Estimation for Spatial Dynamic Panel Data with Fixed Effects: The Case of Spatial Cointegration,” *Journal of Econometrics* 167 (2012), 16–37.