Bias Correction for Fixed Effects Spatial Panel Data Models

Zhenlin Yang[a,*], Jihai Yu[b], and Shew Fan Liu[a]

[a] *School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903.*

[b] *Guanghua School of Management, Peking University, Beijing 100871, China.*

March 8, 2015

Abstract

This paper examines the finite sample properties of the quasi maximum likelihood (QML) estimators of the fixed effects spatial panel data (FE-SPD) models of Lee and Yu (2010). Following the general bias correction methods recently developed by Yang (2015), we derive up to third-order bias corrections for the QML estimators of the FE-SPD model, and propose a simple bootstrap method for their practical implementation. Monte Carlo results reveal that the QML estimators of the spatial parameters can be quite biased and that a second-order bias correction effectively removes the bias. The validity of the bootstrap method is established. Variance corrections are also considered, which together with bias corrections lead to improved inferences.

Key Words: Bias correction, Variance correction, Bootstrap, Spatial panel, Individual fixed effects, Time fixed effects, Quasi maximum likelihood, Spatial lag, Spatial error, Spatial ARAR.

JEL Classification: C10, C13, C21, C23, C15

1 Introduction

Recently there has been growing interest in panel data models with spatial interactions.[1] For the random effects specification, Anselin (1988) provides a panel regression model with error components and spatial autoregressive (SAR) disturbances, and Kapoor et al. (2007) propose a different specification with error components and an SAR structure in the overall disturbance. Baltagi et al. (2013) suggest an extended model without restrictions on implied SAR structures in the error component and the remaining disturbance, which nests the Anselin (1988) and Kapoor et al. (2007) models. As an alternative to the random effects specification, Lee and Yu (2010) investigate the asymptotic properties for quasi-maximum likelihood (QML) estimation of spatial panel models under fixed effects specification. The fixed effects model has the advantage of robustness because fixed effects are allowed to depend on included regressors. It also provides a unified model framework, because different random effects models in Anselin (1988), Kapoor et al. (2007) and Baltagi et al. (2013) reduce to the same fixed effects model.

However, finite sample properties of the QML estimators of fixed effects spatial panel data models are to be unveiled. When considering the finite sample properties of the QML estimators (QMLEs) of spatial panel models, one key feature to be recognized is the fact that the spatial parameters enter the

*Corresponding Author: 90 Stamford Road, Singapore 178903. Phone: +65-6828-0852; Fax: +65-6828-0833. E-mail: zlyang@smu.edu.sg.

[1] See, e.g., Anselin (1988), Baltagi et al. (2003, 2013), Kapoor et al. (2007), Yu et al. (2008, 2012), Yu and Lee (2010), Lee and Yu (2010), Baltagi and Yang (2013a,b), and Su and Yang (2015) for some related works on spatial panel models.

log-likelihood function in a highly non-linear manner that makes it the main source of bias. Direct bias and variance corrections are difficult as closed form evaluations of the expected values related to the bias and variance terms are in general extremely complicated, even with simplifying assumptions such as normality of disturbances. In the current paper, we adopt the approach in Yang (2015) to propose a simple and effective method for correcting the bias and variance of the QMLEs for fixed effects spatial panel data models. Yang (2015) employs a concentrated log-likelihood function of the spatial parameters only, applies a stochastic expansion (Rilestone et al. 1996) to capture bias up to third-order, and uses a simple bootstrap technique to estimate the expected values of higher order quantities that are difficult to derive analytically. Prior to Yang (2015), there are other approaches in the literature that try to tackle the bias problem in spatial linear regression models including Bao and Ullah (2007) and Bao (2013), and also the bias problem in standard panel data models such as Nickell (1981), Kiviet (1995), Hahn and Kuersteiner (2002), Hahn and Newey (2004), Bun and Carree (2005), Hahn and Moon (2006), Arellano and Hahn (2005), to name a few. The advantage of the method of Yang (2015) over the existing methods comes by way of ease of implementation, effectiveness, and generality. It is able to capture biases up to third-order, but typically a second-order bias correction effectively removes the bias. It can be applied to a more complicated model such as the model considered in the current paper.

Compared to Yang (2015), we need to consider the incidental parameter problem caused by the individual and time effects of the spatial panel model (Neyman and Scott, 1948; Lancaster, 2000). Following the ideas of Neyman and Scott (1948), Lee and Yu (2010) observe that when conducting a direct estimation using the likelihood function where all the common parameters and the fixed effects are estimated together, the estimate of the variance parameter is inconsistent when T is finite while n is large. Further, the direct approach is shown to yield consistent estimates for the spatial parameters and the regression coefficients. With data transformations to eliminate the fixed effects as in Lee and Yu (2010), one can avoid the incidental parameter problem, and the data can be pooled after this data transformation so that the ratio of n and T does not affect the asymptotic properties of estimates. The QMLEs derived after the transformation are shown to be consistent, and, except for the variance estimate, are identical to those from the direct approach. In this paper, we follow the transformation approach of Lee and Yu (2010) to examine the finite sample properties of the parameter estimates. Monte Carlo results reveal that the QMLEs of the spatial parameters can be quite biased, in particular for the models with spatial error dependence, and that a second-order bias correction effectively removes the bias. The validity of the bootstrap method is established. Variance corrections are also considered, which together with bias corrections lead to improved inferences.

The rest of the paper is organized as follows. Section 2 introduces the spatial panel data model allowing both spatial lag and spatial error, and both time-specific effects and individual-specific effects, and its QML estimation based on the transformed likelihood function. Section 3 presents a third-order stochastic expansion for the QML estimators of the spatial parameters, a third-order expansion for the bias, and a third-order expansion for the MSE or variance of the QML estimators of the spatial parameters. Section 3 also addresses issues on the bias of QMLEs of other model parameters, and on the inferences following bias and variance corrections. Section 4 introduces the bootstrap method for estimating various quantities in the expansions, and presents theories for the validity of such a method. Section 5 presents Monte Carlo results, and Section 6 concludes the paper.

2 The Model and Its QML Estimation

For the spatial panel data (SPD) model with fixed effects (FE), we can investigate the case with both spatial lag and spatial error, where n is large and T could be finite or large. We include both individual effects and time effects to have a robust specification. The FE-SPD model under consideration is

$$Y_{nt} = \lambda_0 W_{1n} Y_{nt} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + U_{nt}, \quad U_{nt} = \rho_0 W_{2n} U_{nt} + V_{nt}, \quad t = 1, 2, \ldots, T, \tag{2.1}$$

where, for a given t, $Y_{nt} = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ is an $n \times 1$ vector of observations on the response variable, X_{nt} is an $n \times k$ matrix containing the values of k nonstochastic, individually and time varying regressors,

$V_{nt} = (v_{1t}, v_{2t}, \ldots, v_{nt})'$ is an $n \times 1$ vector of errors where $\{v_{it}\}$ are independent and identically distributed (iid) for all i and t with mean 0 and variance σ_0^2, $\mathbf{c}_{n0}$ is an $n \times 1$ vector of fixed individual effects, and α_{t0} is the fixed time effect with l_n being an $n \times 1$ vector of ones. W_{1n} and W_{2n} are given $n \times n$ spatial weights matrices where W_{1n} generates the 'direct' spatial effects among the spatial units in their response values Y_{nt}, and W_{2n} generates cross-sectional dependence among the disturbances U_{nt}. In practice, W_{1n} and W_{2n} may be the same.

In Lee and Yu (2010), QML estimation of (2.1) is considered by using either a direct approach or a transformation approach. The direct approach is to estimate the regression parameters jointly with the individual and time effects, which yields a bias of order $O(T^{-1})$ due to the estimation of individual effects and a bias of order $O(n^{-1})$ due to the estimation of time effects. The transformation approach eliminates the individual and time effects and then implements the estimation, which yields consistent estimates of the common parameters when either n or T is large. In the current paper, we will follow the transformation approach so that it is free from the incidental parameter problem.

To eliminate the individual effects, define $J_T = (I_T - \frac{1}{T} l_T l_T')$ and let $[F_{T,T-1}, \frac{1}{\sqrt{T}} l_T]$ be the orthonormal eigenvector matrix of J_T, where $F_{T,T-1}$ is the $T \times (T-1)$ submatrix corresponding to the eigenvalues of one, I_T is a $T \times T$ identity matrix and l_T is a $T \times 1$ vector of ones.[2] To eliminate the time effects, let J_n and $F_{n,n-1}$ be similarly defined, and let W_{1n} and W_{2n} be row normalized.[3] For any $n \times T$ matrix $[Z_{n1}, \cdots, Z_{nT}]$, define the $(n-1) \times (T-1)$ transformed matrix as

$$[Z_{n1}^*, \ldots, Z_{n,T-1}^*] = F_{n,n-1}'[Z_{n1}, \ldots, Z_{nT}]F_{T,T-1}. \tag{2.2}$$

This leads to, for $t = 1, \ldots, T-1$, Y_{nt}^*, U_{nt}^*, V_{nt}^*, and $X_{nt,j}^*$ for the jth regressor. As in Lee and Yu (2010), let $X_{nt}^* = [X_{nt,1}^*, X_{nt,2}^*, \ldots, X_{nt,k}^*]$, and $W_{hn}^* = F_{n,n-1}' W_{hn} F_{n,n-1}, h = 1, 2$. The transformed model we will work on thus takes the form:

$$Y_{nt}^* = \lambda_0 W_{1n}^* Y_{nt}^* + X_{nt}^* \beta_0 + U_{nt}^*, \quad U_{nt}^* = \rho_0 W_{2n}^* U_{nt}^* + V_{nt}^*, \quad t = 1, \ldots, T-1. \tag{2.3}$$

After the transformations, the effective sample size becomes $N = (n-1)(T-1)$. Stacking the vectors and matrices in (2.3), i.e., letting $\mathbf{Y}_N = (Y_{n1}^{*\prime}, \ldots, Y_{n,T-1}^{*\prime})'$, $\mathbf{U}_N = (U_{n1}^{*\prime}, \ldots, U_{n,T-1}^{*\prime})'$, $\mathbf{V}_N = (V_{n1}^{*\prime}, \ldots, V_{n,T-1}^{*\prime})'$, $\mathbf{X}_N = (X_{n1}^{*\prime}, \ldots, X_{n,T-1}^{*\prime})'$, and denoting $\mathbf{W}_{hN} = I_{T-1} \otimes W_{hn}^*, h = 1, 2$, we have the following compact expression for the transformed model:

$$\mathbf{Y}_N = \lambda_0 \mathbf{W}_{1N} \mathbf{Y}_N + \mathbf{X}_N \beta_0 + \mathbf{U}_N, \quad \mathbf{U}_N = \rho_0 \mathbf{W}_{2N} \mathbf{U}_N + \mathbf{V}_N, \tag{2.4}$$

which is in form identical to the spatial autoregressive model with autoregressive errors (SARAR), showing that the QML estimation of the two-way fixed effects panel SARAR model is similar to that of the linear SARAR model. The key difference is that the elements of $\mathbf{V}_N$ may not be iid though they are uncorrelated and homoskedastic as shown below. This may have a certain impact on the bootstrap method (see next section for details).

It is easy to show that the transformed errors $\{v_{it}^*\}$ are uncorrelated for all i and t by using the identity $(V_{n1}^{*\prime}, \ldots, V_{n,T-1}^{*\prime})' = (F_{T,T-1}' \otimes F_{n,n-1}')(V_{n1}', \ldots, V_{nT}')'$,

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

Hence, $\{v_{it}^*\}$ are iid $N(0, \sigma_0^2)$ if the original errors $\{v_{it}\}$ are iid $N(0, \sigma_0^2)$. It follows that the (quasi) Gaussian log likelihood function for (2.3) is, letting $\zeta = (\beta', \lambda, \rho)'$, and $\theta = (\beta', \sigma^2, \lambda, \rho)'$,

$$\ell_N(\theta) = -\frac{N}{2} \ln(2\pi\sigma^2) + \ln|\mathbf{A}_N(\lambda)| + \ln|\mathbf{B}_N(\rho)| - \frac{1}{2\sigma^2} \mathbf{V}_N'(\zeta) \mathbf{V}_N(\zeta), \tag{2.5}$$

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

[3] When W_{jn} are not row normalized, the linear SARAR presentation of (2.4) for the spatial panel model will no longer hold. In that case, a likelihood formulation would not be feasible, and alternative estimation methods, such as the generalized method of moment, would be possible. Such an estimation approach is beyond the scope of this paper.

where $\mathbf{A}_N(\lambda) = I_N - \lambda \mathbf{W}_{1N}$, $\mathbf{B}_N(\rho) = I_N - \rho \mathbf{W}_{2N}$, and $\mathbf{V}_N(\zeta) = \mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y}_N - \mathbf{X}_N\beta]$.

Now, letting $\mathbf{Y}_N(\lambda) = \mathbf{A}_N(\lambda)\mathbf{Y}_N$ and $\mathbf{X}_N(\rho) = \mathbf{B}_N(\rho)\mathbf{X}_N$, the constrained QMLEs of β and σ^2, given λ and ρ, can be expressed in the following simple form:

$$\tilde{\beta}_N(\lambda, \rho) = [\mathbf{X}'_N(\rho)\mathbf{X}_N(\rho)]^{-1}\mathbf{X}'_N(\rho)\mathbf{B}_N(\rho)\mathbf{Y}_N(\lambda), \tag{2.6}$$

$$\tilde{\sigma}^2_N(\lambda, \rho) = N^{-1}\mathbf{Y}'_N(\lambda)\mathbf{M}_N(\rho)\mathbf{Y}_N(\lambda), \tag{2.7}$$

where $\mathbf{M}_N(\rho) = \mathbf{B}'_N(\rho)\{I_N - \mathbf{X}_N(\rho)[\mathbf{X}'_N(\rho)\mathbf{X}_N(\rho)]^{-1}\mathbf{X}'_N(\rho)\}\mathbf{B}_N(\rho)$. Substituting $\tilde{\beta}_N(\lambda, \rho)$ and $\tilde{\sigma}^2_N(\lambda, \rho)$ back into (2.5) gives the concentrated log likelihood function of (λ, ρ):

$$\ell^c_N(\lambda, \rho) = -\frac{N}{2}(\ln(2\pi) + 1) + \ln|\mathbf{A}_N(\lambda)| + \ln|\mathbf{B}_N(\rho)| - \frac{N}{2}\ln \tilde{\sigma}^2_N(\lambda, \rho). \tag{2.8}$$

Maximizing $\ell^c_N(\lambda, \rho)$ in (2.8) gives the unconstrained QMLEs $\hat{\lambda}_N$ and $\hat{\rho}_N$ of λ and ρ, and substituting $(\hat{\lambda}_N, \hat{\rho}_N)$ back into (2.6) and (2.7) gives the unconstrained QMLEs of β and σ^2 as $\hat{\beta}_N \equiv \tilde{\beta}_N(\hat{\lambda}_N, \hat{\rho}_N)$ and $\hat{\sigma}^2_N \equiv \tilde{\sigma}^2_N(\hat{\lambda}_N, \hat{\rho}_N)$.[4] Write $\hat{\theta}_N = (\hat{\beta}'_N, \hat{\lambda}_N, \hat{\rho}_N, \hat{\sigma}^2_N)'$. Lee and Yu (2010) show that $\hat{\theta}_N$ is $\sqrt{N}$-consistent under some mild conditions. These conditions are stated in Appendix A to facilitate the subsequent developments for the higher-order results. The $\sqrt{N}$-consistency of $\hat{\beta}_N$ and $\hat{\sigma}^2_N$ immediately follows. It follows that the QML estimators of any of the submodels discussed below will be $\sqrt{N}$-consistent as well where N can be $(n-1)(T-1)$, $n(T-1)$, $(n-1)T$, or nT.

The **linear SARAR representation** (2.4) is seen to have greatly facilitated the QML estimation of the general FE-SPD model. It will also be helpful for the subsequent developments in bias and variance corrections. Based on this representation, the results developed for this general model can easily be reduced to suit simpler models. For example, setting ρ or λ to zero in (2.4) gives an FE-SPD model with only the spatial lag (SL) effect or an FE-SPD model with only the spatial error (SE) effect; dropping either α_{t0} or $\mathbf{c}_{n0}$ in (2.1) (or dropping either $F_{n,n-1}$ or $F_{T,T-1}$ in (2.2)) leads to a submodel with only the individual-specific effects or a submodel with only the time-specific effects; finally, dropping both $\mathbf{c}_{n0}$ and α_{t0} in (2.1) leads to a model that is essentially the linear SARAR model discussed in Yang (2015, Sec. 4) for bias corrections. On the other hand, the spatial panel model considered in this paper can also be extended to include more spatial lag terms in both the response and the disturbance, in particular the former. Software can be developed to facilitate the end users of the methodologies developed in this paper.

3 Third-Order Bias and MSE for FE-SPD Model

3.1 Third-order stochastic expansions for nonlinear estimators

In a recent paper, Yang (2015) presents a general method for up to third-order bias and variance corrections on a set of nonlinear estimators based on stochastic expansions and bootstrap. The stochastic expansions provide tractable approximations to the bias and variance of the nonlinear estimators and the bootstrap make these expansions practically implementable. The method is demonstrated, through a linear SAR model, to be very effective in correcting the bias and improving inferences. It was emphasized in Yang (2015) that in estimating a model with both linear and nonlinear parameters the main source of bias and the main difficulty in correcting the bias are associated with the estimation of the nonlinear parameters, and hence one should focus on the concentrated estimation equations. By doing so, the dimensionality of the problem can be greatly reduced, and more importantly the additional variations from the estimation of the linear and scale parameters are captured in correcting the nonlinear estimators,

[4] Numerical maximization of $\ell^c_N(\lambda, \rho)$ can be computationally demanding if N is large due to the need of repeated calculations of the two determinants. Following simplifications help alleviate such a burden: $|\mathbf{A}_N(\lambda)| = \left|I_{n-1} - \lambda W^*_{1n}\right|^{T-1} = \left(\frac{1}{1-\lambda}|I_n - \lambda W_{1n}|\right)^{T-1} = \left(\frac{1}{1-\lambda}\prod_{i=1}^n(1 - \lambda\omega_{1i})\right)^{T-1}$, where ω_{1i} are the eigenvalues of W_{1n}, the middle equation from Lee and Yu (2010), and the last equation is from Griffith (1988). Similarly the determinant of $|\mathbf{B}_N(\rho)|$ is calculated.

thus making the bias and variance corrections more effective. In the current paper, we follow the approach of Yang (2015) to tackle the bias problem in a more complicated model, the FE-SPD model.

Let δ be the vector of nonlinear parameters of a model, and $\hat{\delta}_N$ defined as

$$\hat{\delta}_N = \arg\{\tilde{\psi}_N(\delta) = 0\}, \tag{3.1}$$

be its $\sqrt{N}$-consistent estimator, with $\tilde{\psi}_N(\delta)$ being referred to as the concentrated estimating function (CEF) and $\tilde{\psi}_N(\delta) = 0$ the concentrated estimating equation (CEE). Let $H_{rN}(\delta) = \nabla^r \tilde{\psi}_N(\delta), r = 1, 2, 3,$ where the partial derivatives are carried out sequentially and elementwise, with respect to δ'. Let $\tilde{\psi}_N \equiv \tilde{\psi}_N(\delta_0)$, $H_{rN} \equiv H_{rN}(\delta_0)$ and $H^{\circ}_{rN} = H_{rN} - \text{E}(H_{rN}), r = 1, 2, 3$. Note that here and hereafter the expectation operator 'E' corresponds to the true model parameters θ_0. Define $\Omega_N = -[\text{E}(H_{1N})]^{-1}$. Yang (2015), extending Rilstone et al. (1996) and Bao and Ullah (2007), gives a set of sufficient conditions for a third-order stochastic expansion of $\hat{\delta}_N = \arg\{\tilde{\psi}_N(\delta) = 0\}$, based a general CEF $\tilde{\psi}_N(\delta)$, which are restated here to facilitate the development of higher-order results for the FE-SPD model:

Assumption G1. *$\hat{\delta}_N$ solves $\tilde{\psi}_N(\delta) = 0$ and $\hat{\delta}_N - \delta_0 = O_p(N^{-1/2})$.*

Assumption G2. *$\tilde{\psi}_N(\delta)$ is differentiable up to the rth order for δ in a neighborhood of δ_0, $\text{E}(H_{rN}) = O(1)$, and $H^{\circ}_{rN} = O_p(N^{-1/2})$, $r = 1, 2, 3$.*

Assumption G3. *$[\text{E}(H_{1N})]^{-1} = O(1)$, and $H^{-1}_{1N} = O_p(1)$.*

Assumption G4. *$\|H_{rN}(\delta) - H_{rN}(\delta_0)\| \leq \|\delta - \delta_0\| U_N$ for δ in a neighborhood of δ_0, $r = 1, 2, 3$, and $\text{E}\,|U_N| \leq c < \infty$ for some constant c.*

Under these conditions, a third-order stochastic expansion for $\hat{\delta}_N$ takes the following form:

$$\hat{\delta}_N - \delta_0 = a_{-1/2} + a_{-1} + a_{-3/2} + O_p(N^{-2}), \tag{3.2}$$

where $a_{-s/2}$ represents a term of order $O_p(N^{-s/2})$ for $s = 1, 2, 3$, having the expressions

$$\begin{aligned}
a_{-1/2} &= \Omega_N \tilde{\psi}_N, \\
a_{-1} &= \Omega_N H^{\circ}_{1N} a_{-1/2} + \tfrac{1}{2}\Omega_N \text{E}(H_{2N})(a_{-1/2} \otimes a_{-1/2}), \\
a_{-3/2} &= \Omega_N H^{\circ}_{1N} a_{-1} + \tfrac{1}{2}\Omega_N H^{\circ}_{2N}(a_{-1/2} \otimes a_{-1/2}) \\
&\quad + \tfrac{1}{2}\Omega_N \text{E}(H_{2N})(a_{-1/2} \otimes a_{-1} + a_{-1} \otimes a_{-1/2}) \\
&\quad + \tfrac{1}{6}\Omega_N \text{E}(H_{3N})(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}),
\end{aligned}$$

where $\otimes$ denotes the Kronecker product. In moving from the stochastic expansion given in (3.2) to third-order expansions for the bias, MSE and variance of $\hat{\delta}_N$, it is assumed that $\text{E}(\tilde{\psi}_N) = O(N^{-1})$ and that *a quantity bounded in probability has a finite expectation.* The latter is a simplifying assumption to ensure that the remainders are of the stated order. A third-order expansion for the bias of $\hat{\delta}_N$ is

$$\text{Bias}(\hat{\delta}_N) = b_{-1} + b_{-3/2} + O(N^{-2}), \tag{3.3}$$

where $b_{-1} = \text{E}(a_{-1/2}+a_{-1})$ and $b_{-3/2} = \text{E}(a_{-3/2})$, being respectively the second- and third-order biases. Similarly, a third-order expansion for the mean squared error (MSE) of $\hat{\delta}_N$ is

$$\text{MSE}(\hat{\delta}_N) = m_{-1} + m_{-3/2} + m_{-2} + O(N^{-5/2}), \tag{3.4}$$

where $m_{-1} = \text{E}(a_{-1/2}a'_{-1/2})$, $m_{-3/2} = \text{E}(a_{-1/2}a'_{-1} + a_{-1}a'_{-1/2})$ and $m_{-2} = \text{E}(a_{-1}a'_{-1} + a_{-1/2}a'_{-3/2} + a_{-3/2}a'_{-1/2})$, and the third-order expansion for the variance of $\hat{\delta}_N$ is

$$\text{Var}(\hat{\delta}_N) = v_{-1} + v_{-3/2} + v_{-2} + O(N^{-5/2}), \tag{3.5}$$

where $v_{-1} = \text{Var}(a_{-1/2})$, $v_{-3/2} = \text{Cov}(a_{-1/2}, a_{-1}) + \text{Cov}(a_{-1}, a_{-1/2})$, and $v_{-2} = \text{Cov}(a_{-1/2}, a_{-3/2}) + \text{Cov}(a_{-3/2}, a_{-1/2}) + \text{Var}(a_{-1} + a_{-3/2})$; or simply $v_{-1} = m_{-1}$, $v_{-3/2} = m_{-3/2}$, and $v_{-2} = m_{-2} - b^2_{-1}$.

Therefore, we can improve the statistical inference in finite samples by correcting the bias and standard deviation of estimates. From (3.3), we can use

$$\delta_N^{\mathrm{bc2}} = \hat{\delta}_N - b_{-1} \quad \text{or} \quad \delta_N^{\mathrm{bc3}} = \hat{\delta}_N - b_{-1} - b_{-3/2},$$

to yield an estimator unbiased up to order $O(N^{-1})$ or an estimator unbiased up to order $O(N^{-3/2})$. With estimated b_{-1} and $b_{-3/2}$, feasible δ_N^{bc2} and δ_N^{bc3} can be constructed.

Similar procedures can be applied to increase the precision of variance estimate from (3.5). Under the assumption $\hat{b}_{-1} - b_{-1} = O_p(N^{-3/2})$ and $\hat{b}_{-3/2} - b_{-3/2} = O_p(N^{-2})$, we have

$$\mathrm{Var}(\delta_N^{\mathrm{bc3}}) = v_{-1} + v_{-3/2} + v_{-2} - 2\mathrm{ACov}(\hat{\delta}_N, \hat{b}_{-1}) + O(N^{-5/2}), \tag{3.6}$$

and $\mathrm{Var}(\delta_N^{\mathrm{bc2}}) = \mathrm{Var}(\delta_N^{\mathrm{bc3}}) + O(N^{-5/2})$, where ACov denotes asymptotic covariance. See Section 4 for details on the practical implementations of bias and variance corrections.

3.2 Third-order bias and variance for spatial estimators

In this subsection, we first derive all the quantities required for the third-order expansions for the FE-SPD model, and then discuss conditions under which the results (3.2)-(3.6) hold under the FE-SPD model instead of going through the detailed proofs of them. As seen from Section 2, the set of nonlinear parameters in the FE-SPD model are $\delta = (\lambda, \rho)'$. The CEF leading to the QMLE $\hat{\delta}_N = (\hat{\lambda}_N, \hat{\rho}_N)$ is $\tilde{\psi}_N(\delta) = \frac{1}{N}\frac{\partial}{\partial\delta}\ell_N^c(\delta)$, which is shown to have the form:

$$\tilde{\psi}_N(\delta) = \begin{cases} -T_{0N}(\lambda) + \dfrac{\mathbf{Y}_N'(\lambda)\mathbf{M}_N(\rho)\mathbf{W}_{1N}\mathbf{Y}_N}{\mathbf{Y}_N'(\lambda)\mathbf{M}_N(\rho)\mathbf{Y}_N(\lambda)}, \\ -K_{0N}(\rho) - \dfrac{\mathbf{Y}_N'(\lambda)\mathbf{M}_N^{(1)}(\rho)\mathbf{Y}_N(\lambda)}{2\mathbf{Y}_N'(\lambda)\mathbf{M}_N(\rho)\mathbf{Y}_N(\lambda)}, \end{cases} \tag{3.7}$$

where $T_{0N}(\lambda) = \frac{1}{N}\mathrm{tr}(\mathbf{W}_{1N}\mathbf{A}_N^{-1}(\lambda))$, $K_{0N}(\rho) = \frac{1}{N}\mathrm{tr}(\mathbf{W}_{2N}\mathbf{B}_N^{-1}(\rho))$, and $\mathbf{M}_N^{(1)}(\rho) = \frac{d}{d\rho}\mathbf{M}_N(\rho)$.[5] The conditions for the $\sqrt{N}$-consistency of $\hat{\delta}_N$ are given in Lee and Yu (2010), and also in Appendix A.

To derive the rth order derivative, $H_{rN}(\delta)$, of $\tilde{\psi}_N(\delta)$ w.r.t. δ', $r = 1, 2, 3$, for up to third-order bias correction, define $T_{rN}(\lambda) = \frac{1}{N}\mathrm{tr}[(\mathbf{W}_{1N}\mathbf{A}_N^{-1}(\lambda))^{r+1}]$, and $K_{rN}(\rho) = \frac{1}{N}\mathrm{tr}[(\mathbf{W}_{2N}\mathbf{B}_N^{-1}(\rho))^{r+1}]$, $r = 0, 1, 2, 3$. Let $\mathbf{M}_N^{(k)}(\rho)$ be the kth derivative of $\mathbf{M}_N(\rho)$ w.r.t. ρ, $k = 1, 2, 3, 4$. Define

$$R_{1N}(\delta) = \frac{\mathbf{Y}_N'(\lambda)\mathbf{M}_N(\rho)\mathbf{W}_{1N}\mathbf{Y}_N}{\mathbf{Y}_N'(\lambda)\mathbf{M}_N(\rho)\mathbf{Y}_N(\lambda)}, \qquad R_{2N}(\delta) = \frac{\mathbf{Y}_N'\mathbf{W}_{1N}'\mathbf{M}_N(\rho)\mathbf{W}_{1N}\mathbf{Y}_N}{\mathbf{Y}_N'(\lambda)\mathbf{M}_N(\rho)\mathbf{Y}_N(\lambda)};$$

$$Q_{kN}^{\dagger}(\delta) = \frac{\mathbf{Y}_N'(\lambda)\mathbf{M}_N^{(k)}(\rho)\mathbf{W}_{1N}\mathbf{Y}_N}{\mathbf{Y}_N'(\lambda)\mathbf{M}_N(\rho)\mathbf{Y}_N(\lambda)}, k = 1, 2, 3, \quad Q_{kN}^{\ddagger}(\delta) = \frac{\mathbf{Y}_N'\mathbf{W}_{1N}'\mathbf{M}_N^{(k)}(\rho)\mathbf{W}_{1N}\mathbf{Y}_N}{\mathbf{Y}_N'(\lambda)\mathbf{M}_N(\rho)\mathbf{Y}_N(\lambda)}, k = 1, 2;$$

$$S_{kN}(\delta) = \frac{\mathbf{Y}_N'(\lambda)\mathbf{M}_N^{(k)}(\rho)\mathbf{Y}_N(\lambda)}{\mathbf{Y}_N'(\lambda)\mathbf{M}_N(\rho)\mathbf{Y}_N(\lambda)}, k = 1, 2, 3, 4,$$

which have the following properties

$$\frac{\partial R_{1N}(\delta)}{\partial\lambda} = 2R_{1N}^2(\delta) - R_{2N}(\delta), \qquad \frac{\partial R_{2N}(\delta)}{\partial\lambda} = 2R_{1N}(\delta)R_{2N}(\delta),$$
$$\frac{\partial Q_{kN}^{\dagger}(\delta)}{\partial\lambda} = 2R_{1N}(\delta)Q_{kN}^{\dagger}(\delta) - Q_{kN}^{\ddagger}(\delta), \qquad \frac{\partial Q_{kN}^{\ddagger}(\delta)}{\partial\lambda} = 2R_{1N}(\delta)Q_{kN}^{\ddagger}(\delta),$$
$$\frac{\partial S_{kN}(\delta)}{\partial\lambda} = 2R_{1N}(\delta)S_{kN}(\delta) - 2Q_{kN}^{\dagger}(\delta);$$

$$\frac{\partial R_{1N}(\delta)}{\partial\rho} = Q_{1N}^{\dagger}(\delta) - R_{1N}(\delta)S_{1N}(\delta), \qquad \frac{\partial R_{2N}(\delta)}{\partial\rho} = Q_{1N}^{\ddagger}(\delta) - R_{2N}(\delta)S_{1N}(\delta),$$
$$\frac{\partial Q_{kN}^{\dagger}(\delta)}{\partial\rho} = Q_{k+1,N}^{\dagger}(\delta) - Q_{kN}^{\dagger}(\delta)S_{1N}(\delta), \qquad \frac{\partial Q_{kN}^{\ddagger}(\delta)}{\partial\rho} = Q_{k+1,N}^{\ddagger}(\delta) - Q_{kN}^{\ddagger}(\delta)S_{1N}(\delta),$$
$$\frac{\partial S_{kN}(\delta)}{\partial\rho} = S_{k+1,N}(\delta) - S_{kN}(\delta)S_{1N}(\delta).$$

[5] Lee and Yu (2010) provide another useful identity for calculating the inverse: $(I_{n-1} - \lambda W_{hn}^*)^{-1} = F_{n,n-1}'(I_{n-1} - \lambda W_{hn})^{-1}F_{n,n-1}$. Based on this, the inverses of $\mathbf{A}_N(\lambda)$ and $\mathbf{B}_N(\lambda)$ can easily be calculated as they are block-diagonal.

Write $\tilde{\psi}_N(\delta) = (\tilde{\psi}_{1N}(\delta), \tilde{\psi}_{2N}(\delta))'$ with $\tilde{\psi}_{1N}(\delta) = -T_{0N}(\lambda)+R_{1N}(\delta)$ and $\tilde{\psi}_{2N}(\delta) = -K_{0N}(\rho)-S_{1N}(\delta)$. Denote the partial derivatives of $\tilde{\psi}_{jN}(\delta)$ by adding superscripts λ and/or ρ sequentially, e.g., $\tilde{\psi}_{1N}^{\lambda\lambda}(\delta) = \frac{\partial^2}{\partial\lambda^2}\tilde{\psi}_{1N}(\delta)$, and $\tilde{\psi}_{2N}^{\lambda\rho\lambda}(\delta) = \frac{\partial^3}{\partial\lambda\partial\rho\partial\lambda}\tilde{\psi}_{2N}(\delta)$. Thus, $H_{1N}(\delta)$ has 1st row $\{\tilde{\psi}_{1N}^{\lambda}(\delta), \tilde{\psi}_{1N}^{\rho}(\delta)\}$ and 2nd row $\{\tilde{\psi}_{2N}^{\lambda}(\delta), \tilde{\psi}_{2N}^{\rho}(\delta)\}$, which gives

$$H_{1N}(\delta) = \begin{pmatrix} -T_{1N}(\lambda) - R_{2N}(\delta) + 2R_{1N}^2(\delta), & Q_{1N}^{\dagger}(\delta) - R_{1N}(\delta)S_{1N}(\delta) \\ Q_{1N}^{\dagger}(\delta) - R_{1N}(\delta)S_{1N}(\delta), & -K_{1N}(\rho) - \frac{1}{2}S_{2N}(\delta) + \frac{1}{2}S_{1N}^2(\delta) \end{pmatrix}.$$

$H_{2N}(\delta)$ has rows $\{\tilde{\psi}_{1N}^{\lambda\lambda}(\delta), \tilde{\psi}_{1N}^{\lambda\rho}(\delta), \tilde{\psi}_{1N}^{\rho\lambda}(\delta), \tilde{\psi}_{1N}^{\rho\rho}(\delta)\}$ and $\{\tilde{\psi}_{2N}^{\lambda\lambda}(\delta), \tilde{\psi}_{2N}^{\lambda\rho}(\delta), \tilde{\psi}_{2N}^{\rho\lambda}(\delta), \tilde{\psi}_{2N}^{\rho\rho}(\delta)\}$, where

$$\begin{aligned}
\tilde{\psi}_{1N}^{\lambda\lambda}(\delta) &= -2T_{2N}(\lambda) - 6R_{1N}(\delta)R_{2N}(\delta) + 8R_{1N}^3(\delta),\\
\tilde{\psi}_{1N}^{\lambda\rho}(\delta) &= -Q_{1N}^{\ddagger}(\delta) + 4R_{1N}(\delta)Q_{1N}^{\dagger}(\delta) + R_{2N}(\delta)S_{1N}(\delta) - 4R_{1N}^2(\delta)S_{1N}(\delta),\\
\tilde{\psi}_{1N}^{\rho\rho}(\delta) &= Q_{2N}^{\dagger}(\delta) - 2Q_{1N}^{\dagger}(\delta)S_{1N}(\delta) + 2R_{1N}(\delta)S_{1N}^2(\delta) - R_{1N}(\delta)S_{2N}(\delta),\\
\tilde{\psi}_{2N}^{\rho\rho}(\delta) &= -2K_{2N}(\rho) - \tfrac{1}{2}S_{3N}(\delta) + \tfrac{3}{2}S_{1N}(\delta)S_{2N}(\delta) - S_{1N}^3(\delta),\\
\tilde{\psi}_{2N}^{\lambda\lambda}(\delta) &= \tilde{\psi}_{1N}^{\rho\lambda}(\delta) = \tilde{\psi}_{1N}^{\lambda\rho}(\delta), \text{ and } \tilde{\psi}_{2N}^{\lambda\rho}(\delta) = \tilde{\psi}_{2N}^{\rho\lambda}(\delta) = \tilde{\psi}_{1N}^{\rho\rho}(\delta).
\end{aligned}$$

$H_{3N}(\delta)$ is obtained by taking partial derivatives w.r.t. δ' of every element of $H_{2N}(\delta)$. It has elements:

$$\begin{aligned}
\tilde{\psi}_{1N}^{\lambda\lambda\lambda}(\delta) &= -6T_{3N}(\lambda) + 6R_{2N}^2(\delta) - 48R_{1N}^2(\delta)R_{2N}(\delta) + 48R_{1N}^4(\delta),\\
\tilde{\psi}_{1N}^{\lambda\lambda\rho}(\delta) &= -6Q_{1N}^{\dagger}(\delta)R_{2N}(\delta) + 12R_{1N}(\delta)R_{2N}(\delta)S_{1N}(\delta) - 6R_{1N}(\delta)Q_{1N}^{\ddagger}(\delta),\\
&\quad +24R_{1N}^2(\delta)[Q_{1N}^{\dagger}(\delta) - R_{1N}(\delta)S_{1N}(\delta)],\\
\tilde{\psi}_{1N}^{\lambda\rho\lambda}(\delta) &= 2Q_{1N}^{\ddagger}(\delta)R_{1N}(\delta) + 12R_{1N}(\delta)R_{2N}(\delta)S_{1N}(\delta) - 6R_{1N}(\delta)Q_{1N}^{\dagger}(\delta)\\
&\quad +8R_{1N}^2(\delta)Q_{1N}^{\dagger}(\delta) - 20R_{1N}^3(\delta)S_{1N}(\delta),\\
\tilde{\psi}_{1N}^{\lambda\rho\rho}(\delta) &= -Q_{2N}^{\ddagger}(\rho) + 2Q_{1N}^{\ddagger}(\rho)S_{1N}(\delta) - 2R_{2N}(\delta)S_{1N}^2(\delta) + R_{2N}(\delta)S_{2N}(\delta) + 4Q_{1N}^{\dagger\,2}(\delta)\\
&\quad -16R_{1N}(\delta)S_{1N}(\delta)Q_{1N}^{\dagger}(\delta) + 4R_{1N}(\delta)Q_{2N}^{\dagger}(\delta) + 12R_{1N}^2(\delta)S_{1N}^2(\delta) - 4R_{1N}^2(\delta)S_{2N}(\delta),\\
\tilde{\psi}_{1N}^{\rho\rho\lambda}(\delta) &= -Q_{2N}^{\ddagger}(\delta) + 4Q_{2N}^{\dagger}(\delta)R_{1N}(\delta) + 2Q_{1N}^{\ddagger}(\delta)S_{1N}(\delta) + 4Q_{1N}^{\dagger\,2}(\delta) - 16R_{1N}(\delta)Q_{1N}^{\dagger}(\delta)S_{1N}(\delta)\\
&\quad -R_{2N}(\delta)S_{2N}(\delta) + 12R_{1N}^2(\delta)S_{1N}^2(\delta) - 2R_{2N}(\delta)S_{1N}^2(\delta) - 4S_{1N}^2(\delta)S_{2N}(\delta),\\
\tilde{\psi}_{1N}^{\rho\rho\rho}(\delta) &= Q_{3N}^{\dagger}(\delta) - 3Q_{2N}^{\dagger}(\delta)S_{1N}(\delta) + 6Q_{1N}^{\dagger}(\delta)S_{1N}^2(\delta) - 3Q_{1N}^{\dagger}(\delta)S_{2N}(\delta) - 6R_{1N}(\delta)S_{1N}^3(\delta)\\
&\quad +6R_{1N}(\delta)S_{1N}(\delta)S_{2N}(\delta) - R_{1N}(\delta)S_{3N}(\delta),\\
\tilde{\psi}_{2N}^{\rho\rho\lambda}(\delta) &= Q_{3N}^{\dagger}(\delta) - R_{1N}(\delta)S_{3N}(\delta) - 3Q_{1N}^{\dagger}(\delta)S_{2N}(\delta) + 6R_{1N}(\delta)S_{1N}(\delta)S_{2N}(\delta)\\
&\quad -3S_{1N}(\delta)Q_{2N}^{\dagger}(\delta) + 6S_{1N}^2(\delta)Q_{1N}^{\dagger}(\delta) - 6R_{1N}(\delta)S_{1N}^3(\delta),\\
\tilde{\psi}_{2N}^{\rho\rho\rho}(\delta) &= -6K_{3N}(\rho) - \tfrac{1}{2}S_{4N}(\delta) + 2S_{1N}(\delta)S_{3N}(\delta) + \tfrac{3}{2}S_{2N}^2(\delta) - 6S_{2N}(\delta)S_{1N}^2(\delta) + 3S_{1N}^4(\delta).\\
\tilde{\psi}_{1N}^{\rho\lambda\lambda}(\delta) &= \tilde{\psi}_{1N}^{\lambda\rho\lambda}(\delta) = \tilde{\psi}_{2N}^{\lambda\lambda\lambda}(\delta),\ \tilde{\psi}_{1N}^{\rho\lambda\rho}(\delta) = \tilde{\psi}_{1N}^{\lambda\rho\rho}(\delta) = \tilde{\psi}_{2N}^{\lambda\lambda\rho}(\delta),\\
\tilde{\psi}_{1N}^{\rho\rho\lambda}(\delta) &= \tilde{\psi}_{2N}^{\lambda\rho\lambda}(\delta) = \tilde{\psi}_{2N}^{\rho\lambda\lambda}(\delta), \text{ and } \tilde{\psi}_{1N}^{\rho\rho\rho}(\delta) = \tilde{\psi}_{2N}^{\lambda\rho\rho}(\delta) = \tilde{\psi}_{2N}^{\rho\lambda\rho}(\delta).
\end{aligned}$$

The expressions of $\mathbf{M}_N^{(k)}(\rho)$, ρ, $k = 1, 2, 3, 4$, are lengthy, and hence are relegated to Appendix B.

For the general results (3.2)-(3.6) to be valid when the CEF $\tilde{\psi}_N(\delta)$ corresponds to the FE-SPD model, it is sufficient that this function satisfies Assumptions G1-G4 listed in Section 3.1. First the $\sqrt{N}$-consistency of $\hat{\delta}_N$ in Assumption G1 is given in Theorem A.1 in Appendix A. The differentiability of $\tilde{\psi}_N(\delta)$ in Assumption G2 is obvious. From Section 4.1 we see that the R-, S- and Q-quantities at the true parameter values are all ratios of quadratic forms in $\mathbf{V}_N$, having the same denominator $\mathbf{V}_N'\mathbf{M}_N^{\diamond}\mathbf{V}_N$ where $\mathbf{M}_N^{\diamond} = I_N - \mathbf{X}_N(\rho_0)[\mathbf{X}_N'(\rho_0)\mathbf{X}_N(\rho_0)]^{-1}\mathbf{X}_N'(\rho_0)$. It can be shown that $\frac{1}{N}\mathbf{V}_N'\mathbf{M}_N^{\diamond}\mathbf{V}_N$ converges to $\sigma_0^2(>0)$ with probability one. Hence, with Assumptions A1-A8 in Appendix A, for the H-quantities to have proper stochastic behavior, it would typically require the existence of the 6th moment of v_{it} for the second-order bias correction, and the existence of the 10th moment of v_{it} for the third-order bias correction. Variance corrections have stronger moment requirements. However, these moment requirements are no more than those under a joint estimating equation with analytical approach. The condition $\mathrm{E}(\tilde{\psi}_N) = O(N^{-1})$ is required so that b_{-1} is truly $O(N^{-1})$. This condition is not restrictive as the asymptotic normality of $\hat{\delta}_N$, i.e., as $N \to \infty$, $\sqrt{N}(\hat{\delta}_N - \delta_0)$ converges to a centered bivariate normal distribution, established by Lee and Yu (2010), implies that $\mathrm{E}(\tilde{\psi}_N) = o(N^{-1/2})$. The other conditions are likely to hold by the FE-SPD model. With these and Assumptions A1-A8 in Appendix A, the results (3.2)-(3.6) are likely to hold. For these reasons, we do not present detailed proofs of the results (3.2)-(3.6) for the FE-SPD model, but rather focus on the validity of the bootstrap methods for the practical implementation of these bias and variance corrections.

3.3 Reduced models

Letting either $\rho = 0$ or $\lambda = 0$ leads to two important submodels, the FE-SPD model with SL dependence only and the FE-SPD model with SE dependence only. Bias and variance corrections become much simpler in these cases, in particular the former.

FE-SPD model with SL dependence. The necessary terms for up to third-order bias and variances correction for the FE-SPD model with only SL dependence are:

$$\begin{aligned}
R_{1N}(\lambda) &= \tfrac{\mathbf{Y}'_N(\lambda)\mathbf{M}_N^0\mathbf{W}_{1N}\mathbf{Y}_N}{\mathbf{Y}'_N(\lambda)\mathbf{M}_N^0\mathbf{Y}_N(\lambda)}, \quad R_{2N}(\lambda) = \tfrac{\mathbf{Y}'_N\mathbf{W}'_{1N}\mathbf{M}_N^0\mathbf{W}_{1N}\mathbf{Y}_N}{\mathbf{Y}'_N(\lambda)\mathbf{M}_N^0\mathbf{Y}_N(\lambda)},\\
\tilde{\psi}_N(\lambda) &= -T_{0N}(\lambda) + R_{1N}(\lambda),\\
H_{1N}(\lambda) &= -T_{1N}(\lambda) - R_{2N}(\lambda) + 2R_{1N}^2(\lambda),\\
H_{2N}(\lambda) &= -2T_{2N}(\lambda) - 6R_{1N}(\lambda)R_{2N}(\lambda) + 8R_{1N}^3(\lambda),\\
H_{3N}(\lambda) &= -6T_{3N}(\lambda) + 6R_{2N}^2(\lambda) - 48R_{1N}^2(\lambda)R_{2N}(\lambda) + 48R_{1N}^4(\lambda),
\end{aligned}$$

where $\mathbf{M}_N^0 \equiv \mathbf{M}_N(0) = I_N - \mathbf{X}_N(\mathbf{X}'_N\mathbf{X}_N)^{-1}\mathbf{X}'_N$. These results contain, as a special case, the results for linear SAR model considered in detail in Yang (2015), showing the usefulness of the linear SARAR representation for the fixed effects spatial panel data model given in (2.4).

FE-SPD model with SE dependence. The necessary terms for up to third-order bias and variances correction for the FE-SPD model with only SE dependence are:

$$\begin{aligned}
S_{kN}(\rho) &= \tfrac{\mathbf{Y}'_N\mathbf{M}_N^{(k)}(\rho)\mathbf{Y}_N}{\mathbf{Y}'_N\mathbf{M}_N(\rho)\mathbf{Y}_N}, \ k = 1,2,3,4,\\
\tilde{\psi}_N(\rho) &= -K_{0N}(\rho) - \tfrac{1}{2}S_{1N}(\rho),\\
H_{1N}(\rho) &= -K_{1N}(\rho) - \tfrac{1}{2}S_{2N}(\rho) + \tfrac{1}{2}S_{1N}^2(\rho),\\
H_{2N}(\rho) &= -2K_{2N}(\rho) - \tfrac{1}{2}S_{3N}(\rho) + \tfrac{3}{2}S_{1N}(\rho)S_{2N}(\rho) - S_{1N}^3(\rho),\\
H_{3N}(\rho) &= -6K_{3N}(\rho) - \tfrac{1}{2}S_{4N}(\delta) + 2S_{1N}(\delta)S_{3N}(\delta) + \tfrac{3}{2}S_{2N}^2(\delta)\\
&\quad -6S_{2N}(\delta)S_{1N}^2(\delta) + 3S_{1N}^4(\delta).
\end{aligned}$$

These results contain, as a special case, the results for the linear SED model considered in Liu and Yang (2014). Again, these results show the usefulness of the linear SASAR representation for the fixed effects spatial panel data model given in (2.4).

Simplifications to a one-way fixed effects model are easily done by dropping either $F_{n,n-1}$ or $F_{T,T-1}$ in defining the transformed variables Y_{nt}^*, U_{nt}^*, and V_{nt}^*, and the transformed matrices X_{nt}^* and W_{hn}^*, $h = 1, 2$. Obviously, when the model contains only individual-specific effects, $t = 1, \ldots, T-1$ and $N = n(T-1)$, and when model contains only the time-specific effects, $t = 1, \ldots, T$ and $N = (n-1)T$.

3.4 Bias correction for non-spatial estimators

Note that $\hat{\beta}_N = \tilde{\beta}_N(\hat{\delta}_N)$ and $\hat{\sigma}^2_N = \tilde{\sigma}^2_N(\hat{\delta}_N)$, where $\tilde{\beta}_N(\delta)$ and $\tilde{\sigma}^2_N(\delta)$ are the constrained QMLEs of β and σ^2 defined in (2.6) and (2.7). As $\tilde{\beta}_N(\delta_0)$ is an unbiased estimator of β, and $\frac{N}{N-k}\tilde{\sigma}^2_N(\delta_0)$ is an unbiased estimator of σ^2, it is natural to expect that, with a bias-corrected QMLE $\hat{\delta}^{\rm bc}_N$ of δ, $\hat{\beta}^{\rm bc}_N = \tilde{\beta}_N(\hat{\delta}^{\rm bc}_N)$ and $\hat{\sigma}^{2,{\rm bc}}_N = \frac{N}{N-k}\tilde{\sigma}^2_N(\hat{\delta}^{\rm bc}_N)$ would be much less biased than the original QMLEs. Thus, with a bias-corrected nonlinear estimator, the QMLEs of the linear and scale parameters may be automatically bias-corrected, making the overall bias correction much easier. This is another point stressed by Yang (2015) in supporting the arguments that one should use CEE to perform bias correction on nonlinear parameters. We now present some results to support this point.

First, from (2.6), $\hat{\beta}_N \equiv \tilde{\beta}_N(\hat{\delta}_N) = \mathbf{F}_N(\hat{\rho}_N)\mathbf{Y}_N(\hat{\lambda}_N)$, where $\mathbf{F}_N(\rho) = [\mathbf{X}'_N(\rho)\mathbf{X}_N(\rho)]^{-1}\mathbf{X}'_N(\rho)\mathbf{B}_N(\rho)$. Let $\tilde{\beta}^{(k)}_N(\delta)$ be the kth derivative of $\tilde{\beta}_N(\delta)$ w.r.t. δ', and $\mathbf{F}^{(k)}_N(\rho)$ the kth derivative of $\mathbf{F}_N(\rho)$ w.r.t. ρ. A notational convention is followed: $\tilde{\beta}_N \equiv \tilde{\beta}_N(\delta_0)$, $\tilde{\beta}^{(k)}_N \equiv \tilde{\beta}^{(k)}_N(\delta_0)$, $\mathbf{F}_N \equiv \mathbf{F}_N(\rho_0)$, $\mathbf{A}_N = \mathbf{A}_N(\lambda_0)$, $\mathbf{B}_N = \mathbf{B}_N(\rho_0)$, etc. Assume $\mathrm{E}(\tilde{\beta}^{(k)}_N)$ exists and $\tilde{\beta}^{(k)}_N - \mathrm{E}(\tilde{\beta}^{(k)}_N) = O_p(N^{-1/2}), k = 1, 2$. By a Taylor series expansion, we obtain,

$$\begin{aligned}\tilde{\beta}_N(\hat{\delta}_N) &= \tilde{\beta}_N + \tilde{\beta}^{(1)}_N(\hat{\delta}_N - \delta_0) + \tfrac{1}{2}\tilde{\beta}^{(2)}_N[(\hat{\delta}_N - \delta_0)\otimes(\hat{\delta}_N - \delta_0)] + O_p(N^{-3/2}), \qquad (3.8)\\ &= \tilde{\beta}_N + \mathrm{E}(\tilde{\beta}^{(1)}_N)(\hat{\delta}_N - \delta_0) + b_N a_{-1/2} + \tfrac{1}{2}\mathrm{E}(\tilde{\beta}^{(2)}_N)(a_{-1/2}\otimes a_{-1/2}) + O_p(N^{-3/2}),\end{aligned}$$

where $\mathrm{E}(\tilde{\beta}^{(1)}_N) = [-\mathbf{F}_N\mathbf{G}_N\mathbf{X}_N\beta_0,\ \mathbf{F}^{(1)}_N\mathbf{X}_N\beta_0]$, $\mathbf{G}_N = \mathbf{W}_{1N}\mathbf{A}^{-1}_N$, $b_N = [-\mathbf{F}_N\mathbf{G}_N\mathbf{B}^{-1}_N\mathbf{V}_N,\ \mathbf{F}^{(1)}_N\mathbf{B}^{-1}_N\mathbf{V}_N]$, and $\mathrm{E}(\tilde{\beta}^{(2)}_N) = [0_{k\times 1},\ -\mathbf{F}^{(1)}_N\mathbf{G}_N\mathbf{X}_N\beta_0,\ -\mathbf{F}^{(1)}_N\mathbf{G}_N\mathbf{X}_N\beta_0,\ \mathbf{F}^{(2)}_N\mathbf{X}_N\beta_0]$. Recall $a_{-1/2} = \Omega_N\tilde{\psi}_N$. It is easy to see that the expansion (3.8) holds when $\hat{\delta}_N$ is replaced by $\hat{\delta}^{\rm bc2}_N$. It follows that

$$\begin{aligned}\mathrm{Bias}(\hat{\beta}_N) &= \mathrm{E}(\tilde{\beta}^{(1)}_N)\mathrm{Bias}(\hat{\delta}_N) + \mathrm{E}(b_N a_{-1/2}) + \tfrac{1}{2}\mathrm{E}(\tilde{\beta}^{(2)}_N)\mathrm{E}(a_{-1/2}\otimes a_{-1/2}) + O(N^{-3/2}),\\ \mathrm{Bias}(\hat{\beta}^{\rm bc2}_N) &= \mathrm{E}(b_N a_{-1/2}) + \tfrac{1}{2}\mathrm{E}(\tilde{\beta}^{(2)}_N)\mathrm{E}(a_{-1/2}\otimes a_{-1/2}) + O(N^{-3/2}). \qquad (3.9)\end{aligned}$$

The key term $\mathrm{E}(\tilde{\beta}^{(1)}_N)\mathrm{Bias}(\hat{\delta}_N)$ of order $O(N^{-1})$ in the bias of $\tilde{\beta}_N(\hat{\delta}_N)$ is absorbed into the error term when $\hat{\delta}_N$ is replaced by $\hat{\delta}^{\rm bc2}_N$ in defining the estimator for β_0. Thus, it can be expected that the resulting bias reduction can be big, and the estimator $\hat{\beta}^{\rm bc2}_N = \tilde{\beta}_N(\hat{\delta}^{\rm bc2}_N)$ is essentially second-order bias-corrected, if $\mathrm{E}(b_N a_{-1/2}) + \frac{1}{2}\mathrm{E}(\tilde{\beta}^{(2)}_N)\mathrm{E}(a_{-1/2}\otimes a_{-1/2})$ is 'small'. In general, using (3.9), $\hat{\beta}^{\rm bc2}_N$ can easily be further bias-corrected to be 'truly' second-order unbiased. However, our Monte Carlo results given in Section 5 suggest that this may not be necessary. Finally, $\mathbf{F}^{(k)}_N(\rho), k = 1, 2$, can be easily derived.

Now, from (2.7), $\hat{\sigma}^2_N = \tilde{\sigma}^2_N(\hat{\delta}_N) = \frac{1}{N}\mathbf{Y}'_N(\hat{\lambda}_N)\mathbf{M}_N(\hat{\rho}_N)\mathbf{Y}_N(\hat{\lambda}_N) \equiv \frac{1}{N}Q_N(\hat{\delta}_N)$. Let $Q^{(k)}_N(\delta)$ be the kth partial derivative of $Q_N(\delta)$ w.r.t. δ', and similarly $Q^{(k)}_N \equiv Q^{(k)}_N(\delta_0)$. Assume $\frac{1}{N}\mathrm{E}(Q^{(k)}_N) = O(1)$ and $\frac{1}{N}[Q^{(k)}_N - \mathrm{E}(Q^{(k)}_N)] = O_p(N^{-1/2})$ for $k = 1, 2$. A Taylor series expansion gives,

$$\begin{aligned}\tilde{\sigma}^2_N(\hat{\delta}_N) &= \tilde{\sigma}^2_N + \tfrac{1}{N}Q^{(1)}_N(\hat{\delta}_N - \delta_0) + \tfrac{1}{2N}Q^{(2)}_N[(\hat{\delta}_N - \delta_0)\otimes(\hat{\delta}_N - \delta_0)] + O_p(N^{-3/2}),\\ &= \tilde{\sigma}^2_N + \tfrac{1}{N}\mathrm{E}(Q^{(1)}_N)(\hat{\delta}_N - \delta_0) + q_N a_{-1/2} + \tfrac{1}{2N}\mathrm{E}(Q^{(2)}_N)(a_{-1/2}\otimes a_{-1/2}) + O_p(N^{-3/2}) \qquad (3.10)\end{aligned}$$

where the exact expressions for q_N and $\mathrm{E}(Q^{(k)}_N), k = 1, 2$, are given in Appendix B. It is easy to see that the expansion (3.10) holds when $\hat{\delta}_N$ is replaced by $\hat{\delta}^{\rm bc2}_N$. It follows that

$$\begin{aligned}\mathrm{Bias}[\tfrac{N}{N-k}\tilde{\sigma}^2_N(\hat{\delta}_N)] &= \tfrac{1}{N-k}\mathrm{E}(Q^{(1)}_N)\mathrm{Bias}(\hat{\delta}_N) + \tfrac{N}{N-k}\mathrm{E}(q_N a_{-1/2})\\ &\quad + \tfrac{1}{2(N-k)}\mathrm{E}(Q^{(2)}_N)\mathrm{E}(a_{-1/2}\otimes a_{-1/2}) + O(N^{-3/2}),\\ \mathrm{Bias}[\tfrac{N}{N-k}\tilde{\sigma}^2_N(\hat{\delta}^{\rm bc2}_N)] &= \tfrac{N}{N-k}\mathrm{E}(q_N a_{-1/2}) + \tfrac{1}{2(N-k)}\mathrm{E}(Q^{(2)}_N)\mathrm{E}(a_{-1/2}\otimes a_{-1/2}) + O(N^{-3/2}). \qquad (3.11)\end{aligned}$$

Again, the key bias term $\frac{1}{N-k}\mathrm{E}(Q^{(1)}_N)\mathrm{Bias}(\hat{\delta}_N)$ is removed when $\hat{\delta}_N$ is replaced by $\hat{\delta}^{\rm bc2}_N$ in defining the estimator for σ^2_0, and our Monte Carlo results in Section 5 show that $\frac{N}{N-k}\tilde{\sigma}^2_N(\hat{\delta}^{\rm bc2}_N)$ is nearly unbiased for σ^2_0. In any case, one can always use (3.11) to carry out further bias correction on $\frac{N}{N-k}\tilde{\sigma}^2_N(\hat{\delta}^{\rm bc2}_N)$.

3.5 Inferences following bias and variance corrections

The impacts of bias correction for spatial estimators on the estimation of the regression coefficients and error standard deviation were investigated in the earlier subsection. It would be interesting to further investigate the impacts of bias and variance corrections for spatial estimators on the statistical inferences concerning the spatial parameters or the regression coefficients. One of the most interesting type of inferences for a spatial model would be the testing for the existence of spatial effects. With the availability of QMLEs $\hat{\delta}_N$ and its asymptotic variance $\Omega_N \mathrm{E}(\tilde{\psi}_N \tilde{\psi}_N')\Omega_N$, one can easily carry out a Wald test. However, given the fact that $\hat{\delta}_N$ can be quite biased, it is questionable that this asymptotic test would be reliable when N is not large. With the bias and variance correction results presented in Section 3, one can easily construct various 'bias-corrected' Wald tests. For testing $H_0 : \lambda = \rho = 0$, i.e., the joint non-existence of both types of spatial effects, we have,

$$\mathcal{W}_{N,jk}^{\mathrm{SARAR}} = (\hat{\delta}_N^{\mathrm{bc}j})' \mathrm{Var}_k^{-1}(\hat{\delta}_N^{\mathrm{bc}j})\hat{\delta}_N^{\mathrm{bc}j}, \tag{3.12}$$

where $\hat{\delta}_N^{\mathrm{bc}j}$ is the jth-order bias-corrected $\hat{\delta}_N$ and $\mathrm{Var}_k(\hat{\delta}_N^{\mathrm{bc}j})$ is the kth-order corrected variance of $\hat{\delta}_N^{\mathrm{bc}j}$. When $j = k = 1$, $\hat{\delta}_N^{\mathrm{bc}1} = \hat{\delta}_N$, $\mathrm{Var}_1^{-1}(\hat{\delta}_N^{\mathrm{bc}1}) = \Omega_N \mathrm{E}(\tilde{\psi}_N \tilde{\psi}_N')\Omega_N$, and the test is an asymptotic Wald test. The details on estimating $\mathrm{Var}_k(\hat{\delta}_N^{\mathrm{bc}j})$, in particular, $\mathrm{Var}_3(\hat{\delta}_N^{\mathrm{bc}3})$, are given at the end of Section 4.

Similarly, for testing the non-existence of one type of spatial effects, allowing the existence of the other type of spatial effects, i.e., $H_0 : \lambda = 0$, allowing ρ, or $H_0 : \rho = 0$ allowing λ, we have, respectively,

$$\mathcal{W}_{N,jk}^{\mathrm{SAR}} = \hat{\lambda}_N^{\mathrm{bc}j} / \sqrt{\mathrm{Var}_{11,k}(\hat{\delta}_N^{\mathrm{bc}j})} \quad \text{or} \quad \mathcal{W}_{N,jk}^{\mathrm{SED}} = \hat{\rho}_N^{\mathrm{bc}j} / \sqrt{\mathrm{Var}_{22,k}(\hat{\delta}_N^{\mathrm{bc}j})}, \tag{3.13}$$

where $\mathrm{Var}_{ii,k}(\hat{\delta}_N^{\mathrm{bc}j})$ denotes the i-th diagonal element of $\mathrm{Var}_k(\hat{\delta}_N^{\mathrm{bc}j})$. Furthermore, we can easily construct improved tests for testing the non-existence of spatial effect in the two reduced models, i.e., testing $H_0 : \lambda = 0$, given $\rho = 0$, or $H_0 : \rho = 0$, given $\lambda = 0$:

$$\mathcal{T}_{N,jk}^{\mathrm{SAR}} = \hat{\lambda}_N^{\mathrm{bc}j} / \sqrt{\mathrm{Var}_k(\hat{\lambda}_N^{\mathrm{bc}j})} \quad \text{or} \quad \mathcal{T}_{N,jk}^{\mathrm{SED}} = \hat{\rho}_N^{\mathrm{bc}j} / \sqrt{\mathrm{Var}_k(\hat{\rho}_N^{\mathrm{bc}j})}, \tag{3.14}$$

where $\mathrm{Var}_k(\hat{\lambda}_N^{\mathrm{bc}j})$ and $\mathrm{Var}_k(\hat{\rho}_N^{\mathrm{bc}j})$ are the k-order corrected variances of the jth-order bias-corrected estimators based on the corresponding reduced models described in Section 3.3.

Another important type of inference is the testing or confidence interval construction for $c'\beta_0$, a linear combination of the regression parameters. For an improved inference, we need the bias-corrected variance estimator for $\hat{\beta}_N^{\mathrm{bc}2}$. By (3.8) with $\hat{\delta}_N$ replaced by $\hat{\delta}_N^{\mathrm{bc}2}$, we have,

$$\mathrm{Var}(\hat{\beta}_N^{\mathrm{bc}2}) = \mathrm{Var}\big[\tilde{\beta}_N + \mathrm{E}(\tilde{\beta}_N^{(1)})(a_{-1/2} + a_{-1}) + b_N a_{-1/2} + \tfrac{1}{2}\mathrm{E}(\tilde{\beta}_N^{(2)})(a_{-1/2} \otimes a_{-1/2})\big] + O_p(N^{-2}).$$

This variance can be easily estimated based on the bootstrap method described at the end of Section 4. For testing $H_0 : c'\beta_0 = 0$, the following two statistics may be used:

$$\mathcal{T}_{N,11} = c'\hat{\beta}_N / \sqrt{c'\widehat{\mathrm{AVar}}(\hat{\beta}_N)c}, \quad \text{and} \quad \mathcal{T}_{N,22} = c'\hat{\beta}_N^{\mathrm{bc}2} / \sqrt{c'\widehat{\mathrm{Var}}(\hat{\beta}_N^{\mathrm{bc}2})c}, \tag{3.15}$$

where $\widehat{\mathrm{AVar}}(\hat{\beta}_N)$ is the estimate of the asymptotic variance of $\hat{\beta}_N$ and $\widehat{\mathrm{Var}}(\hat{\beta}_N^{\mathrm{bc}2})$ is the bootstrap estimate of $\mathrm{Var}(\hat{\beta}_N^{\mathrm{bc}2})$ (see the end of Section 4). These results can easily be simplified for the two simpler models.

4 Bootstrap for Feasible Bias and Variance Corrections

For practical purpose, we need to evaluate the expectations of $a_{-s/2}$ for $s = 1, 2, 3$, and the expectations of their cross products. Thus, we need to compute expectations of all the R-, S-, and Q-ratios of quadratic forms defined below (3.7), expectations of their powers, and expectations of cross products of powers, which seem impossible analytically. The use of a joint estimating equation (JEE) as in Bao and

Ullah (2007) and Bao (2013) may offer a possibility. However, even for a second-order bias correction of a simple SAR model (Bao, 2013), the formulae are seen to be very complicated already. Furthermore, the analytical approach runs into another problem with variance corrections and higher-order bias corrections – it may involve higher than fourth moments of the errors of which estimation may not be stable numerically. In the current paper, we follow Yang (2015) to use the CEE, $\tilde{\psi}_N(\delta) = 0$, which not only reduces the dimensionality but also captures additional bias and variability from the estimation of linear and scale parameters, making the bias correction more effective. We then use bootstrap to estimate these expectations involved in the bias and variance corrections, which overcomes the difficulty in analytically evaluating the expectations of ratios of quadratic forms and avoids the direct estimation of higher-order moments of the errors.

4.1 The bootstrap method

We follow Yang (2015) and propose a bootstrap procedure for the FE-SPD model with SARAR effects. Note $\mathbf{Y}_N(\lambda_0) = \mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}(\rho_0)\mathbf{V}_N$, $\mathbf{W}_{1N}\mathbf{Y}_N = \mathbf{G}_N[\mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}(\rho_0)\mathbf{V}_N]$, where $\mathbf{G}_N \equiv \mathbf{G}_N(\lambda_0) = \mathbf{W}_{1N}\mathbf{A}^{-1}(\lambda_0)$, and $\mathbf{M}_N(\rho)\mathbf{X}_N = 0$. The R-ratios, S-ratios and Q-ratios at $\delta = \delta_0$ defined below (3.7) can all be written as functions of $\zeta_0 = (\beta_0', \delta_0')'$ and $\mathbf{V}_N$, given $\mathbf{X}_N$ and $\mathbf{W}_{jN}, j = 1, 2$:

$$R_{1N}(\zeta_0, \mathbf{V}_N) = \frac{\mathbf{V}_N'\mathbf{B}_N'^{-1}\mathbf{M}_N\mathbf{G}_N(\mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}\mathbf{V}_N)}{\mathbf{V}_N'\mathbf{M}_N^{\diamond}\mathbf{V}_N}, \tag{4.1}$$

$$R_{2N}(\zeta_0, \mathbf{V}_N) = \frac{(\mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}\mathbf{V}_N)'\mathbf{G}_N'\mathbf{M}_N\mathbf{G}_N(\mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}\mathbf{V}_N)}{\mathbf{V}_N'\mathbf{M}_N^{\diamond}\mathbf{V}_N}, \tag{4.2}$$

$$Q_{kN}^{\dagger}(\zeta_0, \mathbf{V}_N) = \frac{(\mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}\mathbf{V}_N)'\mathbf{M}_N^{(k)}\mathbf{G}_N(\mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}\mathbf{V}_N)}{\mathbf{V}_N'\mathbf{M}_N^{\diamond}\mathbf{V}_N}, \tag{4.3}$$

$$Q_{kN}^{\ddagger}(\zeta_0, \mathbf{V}_N) = \frac{(\mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}\mathbf{V}_N)'\mathbf{G}_N'\mathbf{M}_N^{(k)}\mathbf{G}_N(\mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}\mathbf{V}_N)}{\mathbf{V}_N'\mathbf{M}_N^{\diamond}\mathbf{V}_N}, \tag{4.4}$$

$$S_{kN}(\zeta_0, \mathbf{V}_N) = \frac{(\mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}\mathbf{V}_N)'\mathbf{M}_N^{(k)}(\mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}\mathbf{V}_N)}{\mathbf{V}_N'\mathbf{M}_N^{\diamond}\mathbf{V}_N}, \tag{4.5}$$

where $\mathbf{M}_N^{\diamond} = I_N - \mathbf{X}_N(\rho_0)[\mathbf{X}_N'(\rho_0)\mathbf{X}_N(\rho_0)]^{-1}\mathbf{X}_N'(\rho_0)$, and $\mathbf{M}_N^{(k)} \equiv \mathbf{M}_N^{(k)}(\rho_0)$. It follows that $\tilde{\psi}_N = \tilde{\psi}_N(\zeta_0, \mathbf{V}_N)$ and $H_{rN} = H_{rN}(\zeta_0, \mathbf{V}_N), r = 1, 2, 3$.

Now, define the QML estimate of the error vector $\mathbf{V}_N$ in the FE-SPD model (2.4):

$$\hat{\mathbf{V}}_N = \mathbf{B}_N(\hat{\rho}_N)[\mathbf{A}(\hat{\lambda}_N)\mathbf{Y}_N - \mathbf{X}_N\hat{\beta}_N]. \tag{4.6}$$

Let $\hat{\mathbf{V}}_N^{\star}$ be a bootstrap sample based on $\hat{\mathbf{V}}_N$. The bootstrap analogs of various quantities are simply

$$\tilde{\psi}_N^{\star} \equiv \tilde{\psi}_N(\hat{\zeta}_N, \mathbf{V}_N^{\star}) \ \text{ and } \ H_{rN}^{\star} \equiv H_{rN}(\hat{\zeta}_N, \mathbf{V}_N^{\star}),\ r = 1, 2, 3.$$

Thus, the bootstrap estimates of the quantities in bias and variance corrections are,[6] for example,

$$\widehat{\mathrm{E}}(\tilde{\psi}_N \otimes H_{rN}) = \mathrm{E}^{\star}\big[\tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}_N^{\star}) \otimes H_{rN}(\hat{\zeta}_N, \hat{\mathbf{V}}_N^{\star})\big], \text{ and}$$
$$\widehat{\mathrm{E}}(\tilde{\psi}_N \otimes \tilde{\psi}_N \otimes \tilde{\psi}_N) = \mathrm{E}^{\star}\big[\tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}_N^{\star}) \otimes \tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}_N^{\star}) \otimes \tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}_N^{\star})\big],$$

where $\mathrm{E}^{\star}$ denotes the expectation with respect to the bootstrap distribution. The bootstrap estimates of other quantities are defined in the same manner. To make these bootstrap expectations practically feasible, we first follow Yang (2015) and propose the following *iid bootstrap* procedure:

[6] To facilitate the bootstrapping, the $a_{-s/2}$ in (3.2) can be re-expressed so that the random quantities are put together, using the well-known properties of Kronecker product: $(A \otimes B)(C \otimes D) = AC \otimes BD$ and $\mathrm{vec}(ACB) = (B' \otimes A)\mathrm{vec}(C)$, where 'vec' vectorizes a matrix by stacking its columns. For example, $H_{1N}\Omega_N\tilde{\psi}_N = (\psi_N' \otimes H_{1N})\mathrm{vec}(\Omega_N)$, and $a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2} = (\Omega_N \otimes \Omega_N \otimes \Omega_N)(\tilde{\psi}_N \otimes \tilde{\psi}_N \otimes \tilde{\psi}_N)$. Alternatively, one can follow the 'two-step' procedure given in Yang (2015, Sec. 4).

Algorithm 4.1 *(iid Bootstrap)*

1. *Compute* $\hat{\zeta}_N$ *and* $\hat{\mathbf{V}}_N$, *and center* $\hat{\mathbf{V}}_N$,
2. *Draw a bootstrap sample* $\hat{\mathbf{V}}^{\star}_{N,b}$, *i.e., make* N *random draws from the elements of centered* $\hat{\mathbf{V}}_N$,
3. *Compute* $\tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b})$ *and* $H_{rN}(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b}), r = 1, 2, 3$,
4. *Repeat steps 2-3 for* B *times to give approximate bootstrap estimates as*

$$\begin{aligned}\widehat{\mathrm{E}}(\tilde{\psi}_N \otimes H_{rN}) &\doteq \tfrac{1}{B}\textstyle\sum_{b=1}^{B}\big[\tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b}) \otimes H_{rN}(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b})\big], \text{ and} \\ \widehat{\mathrm{E}}(\tilde{\psi}_N \otimes \tilde{\psi}_N \otimes \tilde{\psi}_N) &\doteq \tfrac{1}{B}\textstyle\sum_{b=1}^{B}\big[\tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b}) \otimes \tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b}) \otimes \tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b})\big].\end{aligned}$$

Note that the approximation in the last step of Algorithm (4.1) can be made arbitrarily accurate by choosing an arbitrarily large B, and that the scale parameter σ_0^2 and its QMLE $\hat{\sigma}_N^2$ do not play a role in the bootstrap process as they are hidden in either $\mathbf{V}_N$ or $\hat{\mathbf{V}}_N$.

The iid bootstrap procedure requires that the underlining error vector $\mathbf{V}_N$ contains iid elements, which apparently may not be true in general if the original errors are not normal. However, the fact that the elements of $\mathbf{V}_N$ are uncorrelated and homoskedastic suggests that applying the iid bootstrap may give a very good approximation although it may not be strictly valid. Nevertheless, when the original errors are nonnormal, the following *wild bootstrap* or *perturbation* procedure can be used.

Algorithm 4.2 *(Wild Bootstrap)*

1. *Compute* $\hat{\zeta}_N$ *and* $\hat{\mathbf{V}}_N$, *and center* $\hat{\mathbf{V}}_N$,
2. *Compute* $\hat{\mathbf{V}}^{\star}_{N,b} = \hat{\mathbf{V}}_N \odot \boldsymbol{\varepsilon}_b$, *where* $\odot$ *denotes the Hadamard product, and* $\boldsymbol{\varepsilon}_b$ *is an* N*-vector of iid draws from a distribution of mean zero and all higher moments 1, and is independent of* $\hat{\mathbf{V}}_N$.[7]
3. *Compute* $\tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b})$ *and* $H_{rN}(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b}), r = 1, 2, 3$,
4. *Repeat steps 2-3 for* B *times to give approximate bootstrap estimates as*

$$\begin{aligned}\widehat{\mathrm{E}}(\tilde{\psi}_N \otimes H_{rN}) &\doteq \tfrac{1}{B}\textstyle\sum_{b=1}^{B}\big[\tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b}) \otimes H_{rN}(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b})\big], \text{ and} \\ \widehat{\mathrm{E}}(\tilde{\psi}_N \otimes \tilde{\psi}_N \otimes \tilde{\psi}_N) &\doteq \tfrac{1}{B}\textstyle\sum_{b=1}^{B}\big[\tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b}) \otimes \tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b}) \otimes \tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b})\big].\end{aligned}$$

Note that the common applications of the wild bootstrap method are to handle the problem of unknown heteroskedasticity, which clearly is not the main purpose of this paper. In our model, the (transformed) errors are homoskedastic in the usual sense, i.e., variances are constant. Also, the errors are uncorrelated. However, the transformed errors are, strictly speaking, heteroskedastic in the sense that their third and higher order moments may not be constant. The wild bootstrap here aims to capture these non-constant higher-order moments. Also, there may be higher-order dependence, which the wild bootstrap is not able to capture. We see in the next section that this can be ignored.

4.2 Validity of the bootstrap method

In discussing the validity of the bootstrap method, we concentrate on the bias corrections. The fact that the elements of the transformed errors $\mathbf{V}_N = \{v_{it}^*\}$ are uncorrelated and homoskedastic (up to second moment) across i and t, and its observed counterpart $\hat{\mathbf{V}}_N$ is consistent provide the theoretical base for the proposed iid bootstrap method. However, these may not be sufficient in general for the classical iid bootstrap method to be strictly valid, as our estimation requires matching of the higher-order bootstrap moments with those of v_{it}^*. There are important special cases under which the classical iid bootstrap method is strictly valid.

[7] We are unaware of the existence of such a distribution. However, the two-point distribution suggested by Mammen (1993): $\varepsilon_{b,i} = -(\sqrt{5}-1)/2$ or $(\sqrt{5}+1)/2$ with probability $(\sqrt{5}+1)/(2\sqrt{5})$ or $(\sqrt{5}-1)/(2\sqrt{5})$, has mean zero, and second and third moments 1. Another two-point distribution: $\varepsilon_{b,i} = -1$ or 1 with equal probability, has all the odd moments zero and even moments 1. See Liu (1988) and Davidson and Flachaire (2008) for more details on wild bootstrap.

First, we note that the original errors $\{v_{it}\}$ are iid normal, the transformed errors $\{v^*_{it}\}$ are again iid normal. Further, Lemma 4.1 shows that if the original errors $\{v_{it}\}$ are iid with mean zero, variance σ_0^2, and cumulants $k_r = 0, r = 3, 4, \ldots$, then the transformed errors $\{v^*_{it}\}$ will also have mean zero, variance σ_0^2, and rth cumulant being zero for $r = 3, 4, \ldots$. Furthermore, the rth order joint cumulants of the transformed errors are also zero. The iid bootstrap procedure essentially falls into the general framework of Yang (2015) and hence its validity is fully established. We have the following proposition.

Proposition 4.1 *Suppose the conditions leading to the third-order bias expansion* (3.3) *are satisfied by the FE-SPD model. Assume further that the* r*th cumulant* k_r *of* $\{v_{it}\}$ *is 0,* $r = 3, \ldots, 10$*. Then the iid bootstrap method stated in Algorithm 4.1 is valid, i.e., Bias*$(\hat{\delta}_N^{bc2}) = O(N^{-3/2})$ *and Bias*$(\hat{\delta}_N^{bc3}) = O(N^{-2})$*.*

Second, for the important submodel with individual effects only and small T, the transformed errors, $[V^*_{n1}, \ldots, V^*_{n,T-1}] = [V_{n1}, \ldots, V_{n,T}]F_{T,T-1}$ are iid across i, i.e., the rows of the matrix $[V^*_{n1}, \ldots, V^*_{n,T-1}]$ are iid whether the original errors are normal or nonnormal, where $N = n(T-1)$. As T is small and fixed, the asymptotics depend only on n. The bootstrap thus proceeds by randomly drawing the rows of the QML estimate of $[V^*_{n1}, \ldots, V^*_{n,T-1}]$. We have the following proposition.

Proposition 4.2 *Suppose the conditions leading to the third-order bias expansion* (3.3) *are satisfied by the FE-SPD model with only individual effects. Assume further that the* r*th cumulant* k_r *of* $\{v_{it}\}$ *exists,* $r = 3, \ldots, 10$*, and* T *is fixed. Then the bootstrap method making iid draws from the rows of the QML estimates of* $[V^*_{n1}, \ldots, V^*_{n,T-1}]$ *is valid, i.e., Bias*$(\hat{\delta}_N^{bc2}) = O(N^{-3/2})$ *and Bias*$(\hat{\delta}_N^{bc3}) = O(N^{-2})$*.*

For the general FE-SPD model with two-way fixed effects, T being small or large, and the original errors being iid but not necessarily normal, the classical iid bootstrap may not be strictly valid, because the transformed errors (on which the iid bootstrap depend) are not guaranteed to be iid, although they are uncorrelated with mean zero and constant variance σ_0^2. In particular, the transformed errors may not be independent, and their higher-order moments (3rd-order and higher) may not be constant. On the other hand, making random draws from the empirical distribution function (EDF) of the centered $\hat{\mathbf{V}}_N$ gives bootstrap samples that are of iid elements. Thus, the classical iid bootstrap does not fully *mimic* or *recreate* the random structure of $\mathbf{V}_N$, rendering its validity questionable. The following proposition says that the wild bootstrap described in Algorithm 4.2 is valid.

Proposition 4.3 *Suppose the conditions leading to the third-order bias expansion* (3.3) *are satisfied by the FE-SPD model. Assume further that the* r*th cumulant* k_r *of* $\{v_{it}\}$ *exists for* $r = 3, \ldots, 10$*. Then the wild bootstrap method stated in Algorithm 4.2 is valid for the general FE-SPD model, provided that the joint cumulants of the transformed errors* $\{v^*_{it}\}$ *up to* r*th order,* $r = 3, \ldots, 10$*, are negligible.*

Proof: We now present a collective discussion/proof of the Propositions 4.1-4.3. Very importantly, we want to 'show' that the classical iid bootstrap method can give a very good approximation in cases it is not strictly valid, i.e., the 'missing parts' can be ignored numerically.

Let $\mathbb{V}_{nT} = (V'_{n1}, \ldots, V'_{nT})'$ be the vector of original errors in Model (2.1), which contains iid elements of mean zero, variance σ_0^2, cumulative distribution function (CDF) $\mathcal{F}$, and cumulants $k_r, r = 3, 4, \ldots, 10$. Let $\mathbb{F}_{nT,N} = F_{T,T-1} \otimes F_{n,n-1}$ be the $nT \times N$ transformation matrix. We have

$$\mathbf{V}_N = \mathbb{F}'_{nT,N}\mathbb{V}_{nT}. \tag{4.7}$$

For convenience, denote the elements of $\mathbf{V}_N$ by $\mathbf{v}_i$, and the ith column of $\mathbb{F}_{nT,N}$ by $\mathbf{f}_i, i = 1, \ldots, N$. Let $\kappa_r(\cdot)$ denote the rth cumulant of a random variable, and $\kappa(\cdot, \ldots, \cdot)$ the joint cumulants of random variables. Let $\odot$ denote the Hadamard product. A vector raised to rth power is operated elementwise.

From the definition of the bias terms $b_{-s/2}, s = 2, 3$, we see that $b_{-s/2} \equiv b_{-s/2}(\zeta_0, \boldsymbol{\kappa}_N)$ where $\boldsymbol{\kappa}_N$ contains the cumulants or joint cumulants of $\{\mathbf{v}_i\}$. From (4.1)-(4.6), it is clear that the bootstrap estimates of $b_{-s/2}$ are such that $\hat{b}_{-s/2} \equiv b_{-s/2}(\hat{\zeta}_N, \hat{\boldsymbol{\kappa}}^\star_N)$ where $\hat{\boldsymbol{\kappa}}^\star_N$ contains the cumulants of $\{\mathbf{v}^\star_i\}$ w.r.t. the bootstrap distribution. With the $\sqrt{N}$-consistency of $\hat{\theta}_N$, how the set $\hat{\boldsymbol{\kappa}}^\star_N$ match the set $\boldsymbol{\kappa}_N$, becomes central to the validity of the bootstrap method. Following lemmas reveal their relationship.

Lemma 4.1 *If the elements of $\mathbb{V}_{nT}$ are iid with mean zero, variance σ_0^2, CDF $\mathcal{F}$, and higher-order cumulants $k_r, r = 3, 4, \ldots$, then,*

(*a*) $\kappa_1(\mathbf{v}_i) = 0,\ \kappa_2(\mathbf{v}_i) = \sigma_0^2,$ *and* $\kappa_r(\mathbf{v}_i) = k_r\ a_{r,i},\ r \geq 3,\ i = 1, \ldots, N,$

(*b*) $\kappa(\mathbf{v}_i, \mathbf{v}_j) = 0$ *for* $i \neq j$, *and* $\kappa(\mathbf{v}_{i_1}, \ldots, \mathbf{v}_{i_r}) = k_r\ a_{i_1,\ldots,i_r},\ r \geq 3,$

where $a_{r,i} = l'_{nT}\mathbf{f}_i^r,\ a_{i_1,\ldots,i_r} = l'_{nT}(\mathbf{f}_{i_1} \odot \cdots \odot \mathbf{f}_{i_r})$, *and* $\{i_1, \ldots, i_r\}$ *are not all the same.*

Lemma 4.1 shows clearly how the cumulants behave when $\mathbf{v}_i's$ deviate from being iid. First, $a_{r,i}$ are constant across i only when $r = 1$ and 2, i.e., $a_{1,i} = 0$ and $a_{2,i} = 1$. Thus, $\kappa_r(\mathbf{v}_i), r \geq 3$, are not constant across i unless $k_r = 0$. Second, $\mathbf{v}_i's$ are not independent as $a_{i_1,\ldots,i_r} \neq 0$ for $r \geq 3$. However, simple calculation shows that $a_{i_1,\ldots,i_r}$ is very small for any $r \geq 3$ and any choice of $\{i_1, \ldots, i_r\}$ with at least two different elements. The larger the r, the smaller is $a_{i_1,\ldots,i_r}$. These suggest that the higher-order dependence among $\{\mathbf{v}_i\}$ can be ignored. The question left is how well the two sets of cumulants match.

Lemma 4.2 *Let $\mathbf{v}^\star$ be a random draw from $\{\mathbf{v}_i, i = 1, \ldots, N\}$. Then, under the conditions of Lemma 4.1, we have*

$$\kappa_1^\star(\mathbf{v}^\star) = 0,\ \kappa_2^\star(\mathbf{v}^\star) = \sigma_0^2 + O_p(N^{-1/2}),\ \text{and}\ \kappa_r^\star(\mathbf{v}^\star) = k_r\bar{a}_r + O_p(N^{-1/2}),\ r \geq 3,$$

where $\bar{a}_r = \frac{1}{N}\sum_{i=1}^N a_{r,i}$, *and* $\kappa_r^\star(\cdot)$ *denotes rth cumulant w.r.t. the EDF* $\mathcal{G}_N$ *of* $\{\mathbf{v}_i, i = 1, \ldots, N\}$.

Lemma 4.2 shows that the *iid bootstrap* is able to capture, to a certain degree, the higher-order moments of $\mathbf{v}_i$ (a_r versus $a_{r,i}$), but is unable to capture the higher-order dependence. However, as shown by Lemma 4.1, the latter does not have a significant effect as such dependence is weak and negligible. As both $\{a_{r,i}\}$ and their variability are not big and get smaller as r increases, the results of Lemmas 4.1-4.3 strongly suggest that the simple iid bootstrap method may be able to give a good approximation in the situations where the original errors are not far from normal.

Lemma 4.3 *Suppose Assumptions A1-A8 and the conditions of Lemma 4.1 hold. Let $\hat{\mathbf{v}}^\star$ be a random drawn from the EDF $\hat{\mathcal{G}}_N$ of $\{\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_N\}$, and $\mathbf{v}^\star$ a random draw from the EDF $\mathcal{G}_N$ of $\{\mathbf{v}_1 \ldots, \mathbf{v}_N\}$. Then,*

$$\kappa_r^\star(\hat{\mathbf{v}}^\star) = \kappa_r^\star(\mathbf{v}^\star) + O_p(N^{-1/2}),\ \text{or}\ \kappa_r(\hat{\mathcal{G}}_N) = \kappa_r(\mathcal{G}_N) + O_p(N^{-1/2}),\ r \geq 3,$$

where $\kappa_r^\star(\hat{\mathbf{v}}^\star)$ *is the rth cumulant of* $\hat{\mathbf{v}}^\star$ *w.r.t.* $\hat{\mathcal{G}}_N$, *and* $\kappa_r^\star(\mathbf{v}^\star)$ *is the rth cumulant of* $\mathbf{v}^\star$ *w.r.t.* $\mathcal{G}_N$.

In case of severe nonnormality, it may be more important to be able to match the even moments, in particular the kurtosis, than the odd moments as a_r is typically very small when r is odd. This point is also reflected by the fact that the variance of the joint score function (given in Theorem A.1) is free from the third cumulant of the original error. In this spirit, the simple two-point distribution with equal probability described in Footnote 7 may provide satisfactory results.

Lemma 4.4 *Suppose Assumptions A1-A8 and the conditions of Lemma 4.1 hold. Let $\hat{\mathbf{v}}_i^\star = \hat{\mathbf{v}}_i\varepsilon^\star$, where $\varepsilon^\star$ is independent of $\hat{\mathbf{v}}_i$, having a distribution with mean 0 and rth moment 1, $r \geq 2$. Then,*

$$\mathrm{E}^\star(\hat{\mathbf{v}}_i^\star) = 0,\ \text{and}\ \mathrm{E}^\star[(\hat{\mathbf{v}}_i^\star)^r] = \hat{\mathbf{v}}_i^r,\ r \geq 2,$$

where $\mathrm{E}^\star$ *corresponds to the distribution of* $\varepsilon^\star$.

Lemma 4.3 shows that moving from the model errors to their observed counterparts introduces errors of smaller order and hence can be ignored asymptotically. With the results of Lemma 4.4, the validity of the wild bootstrap follows. The proofs of Lemmas 4.1-4.4 are given in Appendix C.

Variance corrections. A final note is given to the variance correction before ending this section. Note that the bootstrap estimate of a bias term or a variance term typically has a bias of order $O(N^{-1})$ multiplied by the order of that term, i.e., $\text{Bias}(\hat{b}_{-1}) = O(N^{-2})$, $\text{Bias}(\hat{v}_{-1}) = O(N^{-2})$,

$\text{Bias}(\hat{v}_{-3/2}) = O(N^{-5/2})$, etc. This is sufficient for achieving a third-order bias correction, but not for a third-order variance correction. Thus, to achieve a third-order variance correction (up to $O(N^{-2})$), a further correction on the bootstrap estimate $\hat{v}_{-1}$ of v_{-1} is desirable. Yang (2015) proposed a method based on the first-order variance term obtained from the joint estimating function. To avoid algebraic complications, in the current paper, we adopt a simple approximation method: replacing $\hat{v}_{-1}$ evaluated at the original QMLE $\hat{\theta}_N$, by $\hat{v}_{-1}^{\text{bc}}$ evaluated at the second-order bias-corrected QMLE $\hat{\theta}_N^{\text{bc2}}$. Monte Carlo results given in the next section show that this approximation works well.

To have a third-order variance correction for $\hat{\delta}_N^{\text{bc3}}$, we also need to estimate $\text{ACov}(\hat{\delta}_N, \hat{b}_{-1})$ in (3.6). Following Yang (2015), we write $\text{ACov}(\hat{\delta}_N, \hat{b}_{-1}) = \text{ACov}(\hat{\delta}_N, \hat{\zeta}_N)\text{E}(b'_{-1,\zeta_0})$, where b_{-1,ζ_0} is the partial derivative of b_{-1} with respect to ζ_0', and $\text{ACov}(\hat{\delta}_N, \hat{\zeta}_N)$ is the submatrix of

$$\text{E}\big(\tfrac{\partial}{\partial\theta_0'}\psi_N(\theta_0)\big)^{-1}\text{Var}\big(\psi_N(\theta_0)\big)\text{E}\big(\tfrac{\partial}{\partial\theta_0'}\psi_N(\theta_0)\big)^{-1},$$

where $\psi_N(\theta) = \frac{\partial}{\partial\theta'}\ell_N(\theta)$. The detailed expressions of $\psi_N(\theta) = \frac{\partial}{\partial\theta'}\ell_N(\theta)$, $\text{Var}\big(\psi_N(\theta_0)\big)$, and $\text{E}\big(\frac{\partial}{\partial\theta_0'}\psi_N(\theta_0)\big)$ are given in Theorem A.1 in Appendix A. We estimate $\text{E}(b_{-1,\zeta_0})$ by $\hat{b}_{-1,\hat{\zeta}_N}$, the numerical derivatives. $\text{E}(\frac{\partial}{\partial\theta_0'}\psi_N(\theta_0))$ can simply be estimated by the plug-in method as it involves only the parameter-vector θ_0. $\text{Var}(\frac{\partial}{\partial\theta_0}\ell_N(\theta_0))$ involves k_4, the fourth cumulant of the original errors, besides the parameter-vector θ_0. The results of Lemmas 4.1-4.3 suggest that k_4 can be consistently estimated by

$$\hat{k}_4 = \bar{a}_4^{-1}\kappa_4(\hat{\mathbf{V}}_N),$$

where $\kappa_4(\hat{\mathbf{V}}_N)$ is the fourth sample cumulant of the QML residuals $\hat{\mathbf{V}}_N$, and $\bar{a}_4$ is give in Lemma 4.2.

Finally, to estimate $\widehat{\text{Var}}(\hat{\beta}_N^{\text{bc2}})$ in (3.15): (i) calculate the estimates of all the non-stochastic quantities with analytical expressions by plugging in $\hat{\delta}_N^{\text{bc2}}$ and $\hat{\beta}_N^{\text{bc2}}$ for δ_0 and β_0, (ii) calculate the new QML residuals based on $\hat{\delta}_N^{\text{bc2}}$ and $\hat{\beta}_N^{\text{bc2}}$, and (iii) bootstrap the new residuals to give bootstrap estimates of the other quantities in $\text{Var}(\hat{\beta}_N^{\text{bc2}})$, including Ω_N and $\text{E}(H_{2N})$, and hence the final estimate $\widehat{\text{Var}}(\hat{\beta}_N^{\text{bc2}})$ of $\text{Var}(\hat{\beta}_N^{\text{bc2}})$. For simplicity, the estimates of Ω_N and $\text{E}(H_{2N})$ from the early stage bootstrap based on the original QMLEs $\hat{\delta}_N$ and $\hat{\beta}_N$ can be directly used.

5 Monte Carlo Study

We present Monte Carlo results to show (i) the finite sample performance of the QMLE $\hat{\delta}_N$ and the bias-corrected QMLEs $\hat{\delta}_N^{\text{bc2}}$ and $\hat{\delta}_N^{\text{bc3}}$, (ii) the impact of bias corrections for $\hat{\delta}_N$ on the estimations for β and σ^2, and (iii) the impact of bias and variance correction on the inferences for spatial or regression coefficients. The simulations are carried out based on the following data generation process (DGP):

$$Y_{nt} = \lambda_0 W_{1n}Y_{nt} + X_{1nt}\beta_{10} + X_{2nt}\beta_{20} + \mathbf{c}_{n0} + \alpha_{t0}l_n + U_{nt}, \quad U_{nt} = \rho_0 W_{2n}U_{nt} + V_{nt}, \quad t = 1, \ldots, T.$$

For all the Monte Carlo experiments, $\beta_0 = (\beta_{10}, \beta_{20})'$ is set to $(1, 1)'$, $\sigma_0^2 = 1$, λ_0 and ρ_0 take values form $\{-0.5, -0.25, 0, 0.25, 0.5\}$, n takes values from $\{25, 50, 100, 200, 500\}$, and $T = \{3, 10\}$. Each set of Monte Carlo results is based on $M = 5000$ Monte Carlo samples, and $B = 999$ bootstrap samples within each Monte Carlo sample. The weight matrices, the regressors, and the idiosyncratic errors are generated as follows.

Weights Matrices. We use four different methods for generating the spatial weights matrices W_{1n} and W_{2n}: (i) `Rook contiguity`, (ii) `Queen contiguity`, (iii) `Circular neighbors`, and (iv) `Group Interaction`. The degree of spatial dependence specified by layouts $(i) - (iii)$ are all fixed while in (iv) it may grow with the increase in sample size. This is attained by relating the number of groups, k, to the sample size n, e.g., $k = n^{0.5}$. In this case, the degree of spatial dependence is reflected by the average group size n/k. For more details on generating spatial weights matrices see Yang (2015).

Regressors. The exogenous regressors are generated according to `REG1`: $\{X_{knt}\} \overset{iid}{\sim} N(0,1)/\sqrt{2}$, and are independent across $k = 1, 2$, and $t = 1, \ldots T$. In case when the spatial dependence is in the form of `group interaction`, the regressors can also be generated according to `REG2`: the ith value of the kth regressor in the gth group is such that $X_{kt,ig} \overset{iid}{\sim} (2z_g + z_{ig})/\sqrt{10}$, where $(z_g, z_{ig}) \overset{iid}{\sim} N(0,1)$ when group interaction scheme is followed; $\{X_{kt,ig}\}$ are independent across k and t, $\{z_g\}$ iid, and $\{z_{ig}\}$ iid.

Error distributions. $v_{it} = \sigma_0 e_{it}$ are generated according to `DGP1`: $\{e_{it}\}$ are iid standard normal; `DGP2`: $\{e_{n,i}\}$ are iid normal mixture with 10% of values from $N(0,4)$ and the remaining from $N(0,1)$, standardized to have mean 0 and variance 1; and `DGP3`: $\{e_{n,i}\}$ iid log-normal (i.e., $\log e_{it} \overset{iid}{\sim} N(0,1)$), standardized to have mean 0 and variance 1.

The estimators of spatial parameters. The finite sample performance of the QMLEs and bias-corrected QMLEs of the spatial parameters is investigated. Monte Carlo results are summarized in Tables 1a, 1b, 2, 3a and 3b, where Tables 1a-1b correspond to the model with $\rho = 0$, i.e., the spatial lag dependence model; Table 2 the model with $\lambda = 0$, i.e., the spatial error dependence model; and Tables 3a-3b the general model. All the reported results correspond to the iid bootstrap method given in Algorithm 4.1. The results (unreported for brevity) using the wild bootstrap method described in Algorithm 4.2 show that the wild bootstrap gives almost identical results as the iid bootstrap, consistent with remarks below Lemma 4.2.

From Tables 1a and 1b, we see that regular QMLEs of the spatial parameters can be very biased, depending on the spatial layouts, the true values of the parameters, and the way that the regressors are generated. First, when the number of cross sectional units increases from 50 to 500, the magnitude of the bias becomes small. The bias is apparent for $n = 50$ and negligible for $n = 500$, which implies that bias correction is especially needed for the data with a small sample size. Also, when the spatial weights matrix becomes denser (from the queen matrix to the group interaction matrix), the bias of regular QMLEs becomes larger. When the true value of spatial effect parameter becomes larger in absolute value, the bias becomes larger. Either reducing the magnitude of the regression parameters β or increasing the value of the error standard deviation increases the bias of the QMLE of the spatial parameter. The magnitude of the bias is also influenced by the way that the regressors are generated. The DGPs with normal errors and lognormal errors give a smaller bias than the DGP with normal mixture errors. For the bias correction, we see that our bias correction procedure works very well, independent of the spatial layouts, model parameters, and the way the regressors being generated. We see that even for the small sample case of $n = 50$, the bias correction procedure produces nearly unbiased estimates. By comparing $\hat{\lambda}_n^{\text{bc2}}$ and $\hat{\lambda}_n^{\text{bc3}}$, we see that in most of the situations considered, a second-order bias correction has essentially removed the bias of the QMLEs and the third-order bias correction might not be needed.

The results in Table 2 show that the patterns observed from the spatial lag model for the regular QMLEs and bias corrections generally hold for the spatial error model. A noticeable difference is that the regular QMLE of the spatial error parameter can be much more biased and the bias can be much more persistent than the QMLE of the spatial lag parameter in the spatial lag model. Therefore, the bias correction procedures developed in the current paper works even more effectively for the spatial error model. Furthermore, unlike the case of spatial lag model, the magnitude of β and σ does not affect the performance of $\hat{\rho}_N$ much.

From Tables 3a and 3b where the third-order bias correction results are omitted for brevity, we see that the general patterns we observed for the two special models hold for the general model as well. However, we observe that the QMLE of the spatial error parameter can be much more biased than the QMLE of the spatial lag parameter, in particular when the regressors are generated in a non-iid manner. The bias of the QMLE of the spatial error parameter can be very persistent and even when $n = 500$, there can still exist very noticeable bias.

The results show that in general the QMLEs of the spatial panel data models need to be bias-corrected even when sample size is not small, and that the proposed bias correction method is very effective in removing the bias. As far as the bias correction is concerned, a simple iid bootstrap may well serve the purpose. The method can easily be applied and thus is recommended to the practitioners.

The estimators of non-spatial parameters. The finite sample properties of $\hat{\beta}_N$ and $\hat{\sigma}^2_N$, and their bias-corrected versions $\hat{\beta}^{\text{bc}}_N$ and $\hat{\sigma}^{2,\text{bc}}_N$ defined in Section 3.4 are investigated. Monte Carlo results reveal some interesting phenomena. The biases of the non-spatial estimators $\hat{\beta}_N$ and $\hat{\sigma}^2_N$ depend very much on whether $\hat{\lambda}_N$ is biased, not much on whether $\hat{\rho}_N$ is biased. In general the biases of $\hat{\beta}_N$ and $\hat{\sigma}^2_N$ are not problems of serious concern (at most 6-7% for the experiments considered). Consistent with the discussions in Section 3.4, $\hat{\beta}^{\text{bc}}_N$ is nearly unbiased in general. When the error distribution is skewed, $\hat{\sigma}^{2,\text{bc}}_N$ may still encounter a bias of less than 5% when $n = 50$ and $T = 3$, and in this case the method given in Section 3.4 can be applied for further bias correction. Partial results are summarized in Table 4.

Inferences following bias and variance corrections. To demonstrate the potential gains from bias and variance corrections, we present Monte Carlo results concerning the finite sample performance of various tests for spatial effects, and the tests concerning the regression coefficients, presented in Section 3.5. Partial results are summarized in Tables 5a-5c, and 6. More comprehensive results are available from the authors.

Table 5a presents the empirical sizes of, respectively, the joint tests for the lack of both SLD and SED effects given in (3.12), and the one-directional tests for the lack of SLD effect allowing the presence of SED effect or the lack of SED effect allowing the presence of SLD effect, given in (3.13). The results show that the third-order bias and variance corrections on the spatial estimators lead to tests that can have a much better finite sample performance over the tests based on the original estimates and asymptotic variances. The tests based on second-order corrections offer improvements over the asymptotic ones but may not be satisfactory. All the reported results are based on the wild bootstrap with the perturbation distribution being the simple two-point (1 and -1) distribution with equal probability. Consistent with the results of Section 4.2, in case of severe nonnormality such as the lognormal errors, the wild bootstrap perform better than the iid bootstrap; in case of normal errors, the iid bootstrap performs slightly better than the wild bootstrap and both show excellent performance of the third-order corrected Wald tests. Due to its robustness, the wild bootstrap may be a better choice in the case of testing for spatial effects. Tables 5b and 5c present the empirical sizes of the tests given in (3.14) for the two simpler models, from which the same conclusions are drawn.

Table 6 presents partial results for the empirical sizes of the tests for the equality of the two regression slopes given in (3.15), based on iid bootstrap. The results show that the tests with merely second-order bias and variance corrections significantly outperforms the standard tests with the original estimate and asymptotic variance. With smaller values of the slope parameters, the size distortion for the standard tests becomes more persistent. The results (unreported for brevity) shows that when the spatial dependence becomes milder the performance of the asymptotic test improves, but is still outperformed by the proposed bias-corrected test.

6 Conclusion and Discussion

We have introduced a general method for correcting the finite sample bias of QMLEs of the two-way fixed effects spatial panel data models where the spatial interactions can be in the form of either spatial lag or spatial error, or both, and the panels can be either short or long. The proposed method follows that of Yang (2015), and is seen to be very easy to implement, and very effective. If only bias-correction is of concern, a second-order correction using iid bootstrap suffices. For improved inferences for the spatial parameters, a third-order variance correction seems necessary and a wild bootstrap method seems to perform better. However, for improved inferences concerning the regression coefficients, the second-order bias and variance corrections seem sufficient, and the resulting inferences can be much more reliable than those based on the standard asymptotic methods. All the methods proposed in the current paper can easily be built into the standard statistical software to facilitate the practical applications.

Appendix A: Some First-Order Results

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

- $\delta = (\lambda, \rho)'$, and δ_0 is its true value.
- For an integer m, $J_m = I_m - \frac{1}{m} l_m l_m'$ where l_m is an $m \times 1$ vector of ones. $[F_{m,m-1}, \frac{1}{\sqrt{m}} l_m]$ is the eigenvector matrix of J_m, where $F_{m,m-1}$ corresponds to eigenvalue of ones.
- $W_{hn}^* = F_{n,n-1}' W_{hn} F_{n,n-1}$, $h = 1, 2$.
- $A_n(\lambda) = I_n - \lambda W_{1n}$ and $B_n(\rho) = I_n - \rho W_{2n}$.
- $[Z_{n1}^*, \ldots, Z_{n,T-1}^*] = F_{n,n-1}' [Z_{n1}, \ldots, Z_{nT}] F_{T,T-1}$ for any $n \times T$ matrix $[Z_{n1}, \cdots, Z_{nT}]$.
- $\mathbf{Y}_N = (Y_{n1}^{*\prime}, \ldots, Y_{n,T-1}^{*\prime})'$, $\mathbf{X}_N = (X_{n1}^{*\prime}, \ldots, X_{n,T-1}^{*\prime})'$, and $\mathbf{W}_{hN} = I_{T-1} \otimes W_{hn}^*$, $h = 1, 2$.
- $\mathbf{A}_N(\lambda) = I_N - \lambda \mathbf{W}_{1N}$, and $\mathbf{B}_N(\rho) = I_N - \rho \mathbf{W}_{2N}$.
- $\mathbf{M}_N(\rho) = \mathbf{B}_N'(\rho)\{I_N - \mathbf{X}_N(\rho)[\mathbf{X}_N'(\rho)\mathbf{X}_N(\rho)]^{-1}\mathbf{X}_N'(\rho)\}\mathbf{B}_N(\rho)$.

The following set of regularity conditions from Lee and Yu (2010) are sufficient for the $\sqrt{N}$-consistency of the QMLE $\hat{\delta}_{nT}$ defined by maximizing (2.8), and hence the $\sqrt{N}$-consistency of the QMLEs $\hat{\beta}_N$ and $\hat{\sigma}_N^2$ of β and σ^2, which are clearly essential for the development of the higher-order results for the QMLEs.

Assumption A1. *W_{1n} and W_{2n} are row-normalized nonstochastic spatial weights matrices with zero diagonals.*

Assumption A2. *The disturbances $\{v_{it}\}$, $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$, are iid across i and t with zero mean, variance σ_0^2 and $\mathrm{E}|v_{it}|^{4+\eta} < \infty$ for some $\eta > 0$.*

Assumption A3. *$A_n(\lambda)$ and $B_n(\rho)$ are invertible for all $\lambda \in \Lambda$ and $\rho \in \mathbb{P}$, where Λ and $\mathbb{P}$ are compact intervals. Furthermore, λ_0 is in the interior of Λ, and ρ_0 is in the interior of $\mathbb{P}$.*[8]

Assumption A4. *The elements of X_{nt} are nonstochastic, and are bounded uniformly in n and t. Under the setting in Assumption A6, the limit of $\frac{1}{N}\mathbf{X}_N'\mathbf{X}_N$ exists and is nonsingular.*

Assumption A5. *W_{1n} and W_{2n} are uniformly bounded in both row and column sums in absolute value (for short, UB).*[9] *Also $A_n^{-1}(\lambda)$ and $B_n^{-1}(\rho)$ are UB, uniformly in $\lambda \in \Lambda$ and $\rho \in \mathbb{P}$.*

Assumption A6. *n is large, where T can be finite or large.*[10]

Assumption A7. *Either (a): $\lim_{n\to\infty} \mathcal{H}_N(\rho)$ is nonsingular $\forall \rho \in \mathbb{P}$ and $\lim_{n\to\infty} \mathcal{Q}_{1n}(\rho) \neq 0$ for $\rho \neq \rho_0$; or (b): $\lim_{n\to\infty} \mathcal{Q}_{2n}(\delta) \neq 0$ for $\delta \neq \delta_0$, where*

$$\begin{aligned}
\mathcal{H}_N(\rho) &= \tfrac{1}{N}(\mathbf{X}_N, \mathbf{W}_{1N}\mathbf{A}_N^{-1}\mathbf{X}_N\beta_0)'\mathbf{B}_N'(\rho)\mathbf{B}_N(\rho)(\mathbf{X}_N, \mathbf{W}_{1N}\mathbf{A}_N^{-1}\mathbf{X}_N\beta_0), \\
\mathcal{Q}_{1n}(\rho) &= \tfrac{1}{n-1}\left(\ln\left|\sigma_0^2 B_n^{-1\prime} J_n B_n^{-1}\right| - \ln\left|\sigma_n^2(\rho) B_n^{-1}(\rho)' J_n B_n^{-1}(\rho)\right|\right), \\
\mathcal{Q}_{2n}(\delta) &= \tfrac{1}{n-1}\left(\ln\left|\sigma_0^2 B_n^{-1\prime} A_n^{-1\prime} J_n A_n^{-1} B_n^{-1}\right| - \ln\left|\sigma_n^2(\delta) B_n^{-1}(\rho)' A_n^{-1}(\lambda)' J_n A_n^{-1}(\lambda) B_n^{-1}(\rho)\right|\right), \\
\sigma_n^2(\delta) &= \tfrac{\sigma_0^2}{n-1}\mathrm{tr}[(B_n(\rho)A_n(\lambda)A_n^{-1}B_n^{-1})' J_n (B_n(\rho)A_n(\lambda)A_n^{-1}B_n^{-1})], \text{ and } \sigma_n^2(\rho) = \sigma_n^2(\delta)|_{\lambda=\lambda_0}.
\end{aligned}$$

Assumption A8. *The limit of $\frac{1}{(n-1)^2}\left[\mathrm{tr}(C_n^s C_n^s)\mathrm{tr}(D_n^s D_n^s) - \mathrm{tr}^2(C_n^s D_n^s)\right]$ is strictly positive, where $C_n = J_n \ddot{G}_n - \frac{\mathrm{tr} J_n \ddot{G}_n}{n-1} J_n$ and $D_n = J_n H_n - \frac{\mathrm{tr} J_n H_n}{n-1} J_n$, with $H_n = W_{2n} B_n^{-1}$ and $\ddot{G}_n = B_n(W_{1n} A_n^{-1}) B_n^{-1}$.*

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

[9] A (sequence of $n \times n$) matrix P_n is said to be uniformly bounded in row and column sums in absolute value if $\sup_{n\geq 1}\|P_n\|_\infty < \infty$ and $\sup_{n\geq 1}\|P_n\|_1 < \infty$, where $\|P_n\|_\infty = \sup_{1\leq i\leq n}\sum_{j=1}^n |p_{ij,n}|$ and $\|P_n\|_1 = \sup_{1\leq j\leq n}\sum_{i=1}^n |p_{ij,n}|$ are, respectively, the row sum and column sum norms.

[10] The case with a finite n and large T is of less interest as the incidental parameter problem does not occur in this model. The consistency and asymptotic normality of QML estimate still hold under a finite n and a large T.

Theorem A.1 *(Lee and Yu, 2010) Under Assumptions A1-A8, we have* $\hat{\theta}_N \xrightarrow{p} \theta_0$*, and*

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{D} N\big[0,\ \lim_{N\to\infty} \Sigma_N^{-1}(\theta_0)\Gamma_N(\theta_0)\Sigma_N^{-1}(\theta_0)\big], \tag{A.1}$$

where $\Sigma_N(\theta_0) = \frac{1}{N}\mathrm{E}[\frac{\partial^2}{\partial\theta_0\partial\theta_0'}\ell_N(\theta_0)]$ *assumed to be positive definite for large enough* N*, and* $\Gamma_N(\theta_0) = \frac{1}{N}\mathrm{E}[(\frac{\partial}{\partial\theta_0}\ell_N(\theta_0))(\frac{\partial}{\partial\theta_0}\ell_N(\theta_0))']$ *assumed to exist.*

The results of Theorem A.1 serve two purposes: one is the $\sqrt{N}$-consistency of $\hat{\theta}_N$, which is crucial for the higher-order results developed in this paper, and the other is the asymptotic VC matrix of $\hat{\theta}_N$, which is needed in the third-order variance correction. With the set of compact notations introduced in Section 2, the component $\Sigma_N(\theta_0)$ of the VC matrix takes the following form:

$$\Sigma_N(\theta_0) = \begin{pmatrix} \frac{1}{N\sigma_0^2}\mathbf{X}_N'\mathbf{B}_N'\mathbf{B}_N\mathbf{X}_N, & 0, & \frac{1}{N\sigma_0^2}\mathbf{X}_N'\mathbf{B}_N'\boldsymbol{\eta}_N, & 0 \\ \sim, & \frac{1}{2\sigma_0^4}, & \frac{1}{N\sigma_0^2}\mathrm{tr}(\mathbf{B}_N'^{-1}\mathbf{G}_N\mathbf{B}_N), & \frac{1}{N\sigma_0^2}\mathrm{tr}(\mathbf{W}_{2N}\mathbf{B}_N^{-1}) \\ \sim, & \sim, & T_{1N} + T_{1N}^* + \frac{1}{N\sigma_0^2}\boldsymbol{\eta}_N'\boldsymbol{\eta}_N, & T_{2N}^* \\ \sim, & \sim, & \sim, & K_{1N} + K_{1N}^* \end{pmatrix},$$

where $\boldsymbol{\eta}_N = \mathbf{G}_N\mathbf{X}_N\beta_0$, $T_{1N}^* = \frac{1}{N}\mathrm{tr}(\mathbf{B}_N'^{-1}\mathbf{G}_N'\mathbf{B}_N'\mathbf{B}_N\mathbf{G}_N\mathbf{B}_N^{-1})$, $K_{1N}^* = \frac{1}{N}\mathrm{tr}(\mathbf{B}_N'^{-1}\mathbf{W}_{2N}'^{-1}\mathbf{W}_{2N}^{-1}\mathbf{B}_N^{-1})$, and $T_{2N}^* = \frac{1}{N}\mathrm{tr}(\mathbf{B}_N'^{-1}\mathbf{G}_N'\mathbf{W}_{2N} + \mathbf{B}_N'^{-1}\mathbf{G}_N'\mathbf{B}_N'\mathbf{W}_{2N}\mathbf{B}_N^{-1})$.

To obtain the other component $\Gamma_N(\theta_0)$ of the VC matrix, it is helpful to express the score vector in terms of the original errors using (4.7):

$$\frac{1}{N}\frac{\partial \ell_N(\theta_0)}{\partial\theta_0} = \begin{cases} \frac{1}{N\sigma_0^2}\mathbb{A}_{1nT}'\mathbb{V}_{nT} \\ -\frac{1}{2\sigma_0^2} + \frac{1}{2N\sigma_0^4}\mathbb{V}_{nT}'\mathbb{A}_{2nT}'\mathbb{V}_{nT} \\ -T_{0N} + \frac{1}{N\sigma_0^2}\mathbb{V}_{nT}'\mathbb{A}_{3nT}'\mathbb{V}_{nT} + \frac{1}{N\sigma_0^2}\mathbf{b}_{nT}'\mathbb{V}_{nT} \\ -K_{0N} + \frac{1}{N\sigma_0^2}\mathbb{V}_{nT}'\mathbb{A}_{4nT}'\mathbb{V}_{nT} \end{cases}$$

where $\mathbf{b}_{nT} = \mathbb{F}_{nT,N}\mathbf{B}_N\boldsymbol{\eta}_N$, $\mathbb{A}_{1nT} = \mathbb{F}_{nT,N}\mathbf{B}_N\mathbf{X}_N$, $\mathbb{A}_{2nT} = \mathbb{F}_{nT,N}\mathbb{F}_{nT,N}'$, $\mathbb{A}_{3nT} = \mathbb{F}_{nT,N}\mathbf{B}_N\mathbf{G}_N\mathbf{B}_N^{-1}\mathbb{F}_{nT,N}'$, and $\mathbb{A}_{4nT} = \mathbb{F}_{nT,N}\mathbf{W}_{2N}\mathbf{B}_N^{-1}\mathbb{F}_{nT,N}'$. Letting $\mathbf{a}_{inT}$ be the diagonal vector of $\mathbb{A}_{inT}$, and denoting

$$\Pi_{ij} = \tfrac{1}{N}\mathrm{tr}[\mathbb{A}_{inT}(\mathbb{A}_{jnT} + \mathbb{A}_{jnT}')] + \tfrac{1}{N}k_4\mathbf{a}_{inT}'\mathbf{a}_{jnT},$$

we obtain, referring to Lemma A.4 of Lee and Yu (2010) and its proof,

$$\Gamma_N(\theta_0) = \begin{pmatrix} \frac{1}{N\sigma_0^2}\mathbf{X}_N'\mathbf{B}_N'\mathbf{B}_N\mathbf{X}_N, & 0, & \frac{1}{N\sigma_0^2}\mathbf{A}_{1nT}'\mathbf{b}_{nT}, & 0 \\ \sim, & \frac{1}{4\sigma_0^4}\Pi_{22}, & \frac{1}{2\sigma_0^2}\Pi_{23}, & \frac{1}{2\sigma_0^2}\Pi_{24} \\ \sim, & \sim, & \Pi_{33} + \frac{1}{N\sigma_0^2}\mathbf{b}_{nT}'\mathbf{b}_{nT}, & \Pi_{34} \\ \sim, & \sim, & \sim, & \Pi_{44} \end{pmatrix}.$$

Appendix B: Some Higher-Order Results

Derivatives of $\mathbf{M}_N(\rho)$ defined below (2.7)**.**

We have $\mathbf{M}_N(\rho) = C_N(\rho) - C_N(\rho)\mathbf{X}_N D_N(\rho)\mathbf{X}_N' C_N(\rho)$ where $C_N(\rho) = \mathbf{B}_N'(\rho)\mathbf{B}_N(\rho)$ and $D_N(\rho) = [\mathbf{X}_N' C_N(\rho)\mathbf{X}_N]^{-1}$. Let $C_N^{(k)}(\rho)$ and $D_N^{(k)}(\rho)$ be, respectively, the kth order partial derivatives of $C_N(\rho)$ and $D_N(\rho)$ w.r.t. ρ. The derivatives of $\mathbf{M}_N(\rho)$ are given as follows,

$$\begin{aligned}\mathbf{M}_N^{(1)}(\rho) = {} & C_N^{(1)}(\rho) - C_N^{(1)}(\rho)\mathbf{X}_N D_N(\rho)\mathbf{X}_N' C_N(\rho) - C_N(\rho)\mathbf{X}_N D_N^{(1)}(\rho)\mathbf{X}_N' C_N(\rho) \\ & - C_N(\rho)\mathbf{X}_N D_N(\rho)\mathbf{X}_N' C_N^{(1)}(\rho),\end{aligned}$$

$$\begin{aligned}\mathbf{M}_N^{(2)}(\rho) = & C_N^{(2)}(\rho) - C_N^{(2)}(\rho)\mathbf{X}_N D_N(\rho)\mathbf{X}_N' C_N(\rho) - 2C_N^{(1)}(\rho)\mathbf{X}_N D_N^{(1)}(\rho)\mathbf{X}_N' C_N(\rho) \\ & -2C_N^{(1)}(\rho)\mathbf{X}_N D_N(\rho)\mathbf{X}_N' C_N^{(1)}(\rho) - 2C_N(\rho)\mathbf{X}_N D_N^{(1)}(\rho)\mathbf{X}_N' C_N^{(1)}(\rho) \\ & -C_N(\rho)\mathbf{X}_N D_N^{(2)}(\rho)\mathbf{X}_N' C_N(\rho) - C_N(\rho)\mathbf{X}_N D_N(\rho)\mathbf{X}_N' C_N^{(2)}(\rho)\end{aligned}$$

$$\begin{aligned}\mathbf{M}_N^{(3)}(\rho) = & -3C_N^{(2)}(\rho)\mathbf{X}_N D_N^{(1)}\mathbf{X}_N' C_N(\rho) - 3C_N^{(2)}(\rho)\mathbf{X}_N D_N(\rho)\mathbf{X}_N' C_N^{(1)}(\rho) \\ & -3C_N^{(1)}(\rho)\mathbf{X}_N D_N^{(2)}(\rho)\mathbf{X}_N' C_N(\rho) - 6C_N^{(1)}(\rho)\mathbf{X}_N D_N^{(1)}(\rho)\mathbf{X}_N' C_N^{(1)}(\rho) \\ & -3C_N^{(1)}(\rho)\mathbf{X}_N D_N(\rho)\mathbf{X}_N' C_N^{(2)}(\rho) - 3C_N(\rho)\mathbf{X}_N D_N^{(2)}(\rho)\mathbf{X}_N' C_N^{(1)}(\rho) \\ & -3C_N(\rho)\mathbf{X}_N D_N^{(1)}(\rho)\mathbf{X}_N' C_N^{(2)}(\rho) - C_N(\rho)\mathbf{X}_N D_N^{(3)}(\rho)\mathbf{X}_N' C_N(\rho)\end{aligned}$$

$$\begin{aligned}\mathbf{M}_N^{(4)}(\rho) = & -6C_N^{(2)}(\rho)\mathbf{X}_N D_N^{(2)}\mathbf{X}_N' C_N(\rho) - 12C_N^{(2)}(\rho)\mathbf{X}_N D_N^{(1)}(\rho)\mathbf{X}_N' C_N^{(1)}(\rho) \\ & -6C_N^{(2)}(\rho)\mathbf{X}_N D_N(\rho)\mathbf{X}_N' C_N^{(2)}(\rho) - 4C_N^{(1)}(\rho)\mathbf{X}_N D_N^{(3)}(\rho)\mathbf{X}_N' C_N(\rho) \\ & -4C_N(\rho)\mathbf{X}_N D_N^{(3)}(\rho)\mathbf{X}_N' C_N^{(1)}(\rho) - 12C_N^{(1)}(\rho)\mathbf{X}_N D_N^{(2)}(\rho)\mathbf{X}_N' C_N^{(1)}(\rho) \\ & -12C_N^{(1)}(\rho)\mathbf{X}_N D_N^{(1)}(\rho)\mathbf{X}_N' C_N^{(2)}(\rho) - 6C_N(\rho)\mathbf{X}_N D_N^{(2)}(\rho)\mathbf{X}_N' C_N^{(2)}(\rho) \\ & -C_N(\rho)\mathbf{X}_N D_N^{(4)}(\rho)\mathbf{X}_N' C_N(\rho).\end{aligned}$$

For the derivatives of $C_N(\rho)$, we have $C_N^{(1)}(\rho) = -\mathbf{W}_{2N}'\mathbf{B}_N(\rho) - \mathbf{B}_N'(\rho)\mathbf{W}_{2N}$, $C_N^{(2)}(\rho) = 2\mathbf{W}_{2N}'\mathbf{W}_{2N}$, and $C_N^{(k)}(\rho) = 0, k \geq 3$. For the derivatives of $D_N(\rho)$, denoting $P_N(\rho) = \mathbf{X}_N' C_N(\rho)\mathbf{X}_N$ and its kth derivative $P_N^{(k)}(\rho)$, we have,

$$D_N^{(1)}(\rho) = -D_N(\rho)P_N^{(1)}(\rho)D_N(\rho),$$

$$D_N^{(2)}(\rho) = -D_N^{(1)}(\rho)P_N^{(1)}(\rho)D_N(\rho) - D_N(\rho)P_N^{(2)}(\rho)D_N(\rho) - D_N(\rho)P_N^{(1)}(\rho)D_N^{(1)}(\rho),$$

$$\begin{aligned}D_N^{(3)}(\rho) = & -D_N^{(2)}(\rho)P_N^{(1)}(\rho)D_N(\rho) - D_N(\rho)P_N^{(1)}(\rho)D_N^{(2)}(\rho) - 2D_N^{(1)}(\rho)P_N^{(2)}(\rho)D_N(\rho) \\ & -2D_N^{(1)}(\rho)P_N^{(1)}(\rho)D_N^{(1)}(\rho) - 2D_N(\rho)P_N^{(2)}(\rho)D_N^{(1)}(\rho),\end{aligned}$$

$$\begin{aligned}D_N^{(4)}(\rho) = & -D_N^{(3)}(\rho)P_N^{(1)}(\rho)D_N(\rho) - D_N(\rho)P_N^{(1)}(\rho)D_N^{(3)}(\rho) - 3D_N^{(2)}(\rho)P_N^{(2)}(\rho)D_N(\rho) \\ & -3D_N^{(2)}(\rho)P_N^{(1)}(\rho)D_N^{(1)}(\rho) - 3D_N^{(1)}(\rho)P_N^{(1)}(\rho)D_N^{(2)}(\rho) - 3D_N(\rho)P_N^{(2)}(\rho)D_N^{(2)}(\rho) \\ & -6D_N^{(1)}(\rho)P_N^{(2)}(\rho)D_N^{(1)}(\rho).\end{aligned}$$

Clearly, $P_N^{(k)}(\rho)$ can be obtained from $C_N^{(k)}(\rho)$, and both are zero when $k \geq 3$.

Additional quantities required in (3.10)**.**

Letting $\mathrm{E}(Q_N^{(1)}) = (s_1, s_2)$, $q_N = (s_3, s_4)$ and $\mathrm{E}[Q_N^{(2)}(\delta_0) = (s_5, s_6, s_7, s_8)$, we have

$$\begin{aligned}
s_1 &= -2\beta_0'\mathbf{X}_N'\mathbf{G}_{1N}'\mathbf{M}_N\mathbf{X}_N\beta_0 - 2\sigma_0^2\mathrm{tr}[\mathbf{G}_N\mathbf{M}_N(\mathbf{B}_N'\mathbf{B}_N)^{-1}],\\
s_2 &= 2\beta_0'\mathbf{X}_N'\mathbf{M}_N^{(1)}\mathbf{X}_N\beta_0 + \sigma_0^2\mathrm{tr}[\mathbf{M}_N^{(1)}(\mathbf{B}_N'\mathbf{B}_N)^{-1}],\\
s_3 &= -4\beta_0'\mathbf{X}_N'\mathbf{G}_{1N}'\mathbf{M}_N\mathbf{B}_N^{-1}\mathbf{V}_N - 2\mathbf{V}_N'\mathbf{B}_N'\mathbf{G}_N\mathbf{M}_N\mathbf{B}_N^{-1}\mathbf{V}_N + 2\sigma_0^2\mathrm{tr}[\mathbf{G}_N\mathbf{M}_N(\mathbf{B}_N'\mathbf{B}_N)^{-1}],\\
s_4 &= 2\beta_0'\mathbf{X}_N'\mathbf{M}_N^{(1)}\mathbf{B}_N^{-1}\mathbf{V}_N + \mathbf{V}_N'\mathbf{B}_N'^{-1}\mathbf{M}_N^{(1)}\mathbf{B}_N^{-1}\mathbf{V}_N - \sigma_0^2\mathrm{tr}[\mathbf{M}_N^{(1)}(\mathbf{B}_N'\mathbf{B}_N)^{-1}],\\
s_5 &= 2\beta_0'\mathbf{X}_N'\mathbf{G}_{1N}'\mathbf{M}_N\mathbf{G}_N\mathbf{X}_N\beta_0 + 2\sigma_0^2\mathrm{tr}[\mathbf{G}_{1N}'\mathbf{M}_N\mathbf{G}_N(\mathbf{B}_N'\mathbf{B}_N)^{-1}],\\
s_6 &= q_7 = -2\beta_0'\mathbf{X}_N'\mathbf{G}_{1N}'\mathbf{M}_N^{(1)}\mathbf{X}_N\beta_0 - 2\sigma_0^2\mathrm{tr}[\mathbf{G}_N\mathbf{M}_N^{(1)}(\mathbf{B}_N'\mathbf{B}_N)^{-1}],\\
s_8 &= \beta_0'\mathbf{X}_N'\mathbf{M}_N^{(2)}\mathbf{X}_N\beta_0 + \sigma_0^2\mathrm{tr}[\mathbf{M}_N^{(2)}(\mathbf{B}_N'\mathbf{B}_N)^{-1}],
\end{aligned}$$

where $\mathbf{M}_N \equiv \mathbf{M}_N(\rho_0)$ and $\mathbf{M}_N^{(k)} \equiv \mathbf{M}_N^{(k)}(\rho_0)$.

Appendix C: Proofs for Section 4

Proof of Lemma 4.1: The results of (a) follows from the following properties of cumulants: for two independent random variables X and Y and a constant c, (i) $\kappa_1(X+c) = \kappa_1(X) + c$, (ii) $\kappa_r(X+c) = \kappa_r(X), r \geq 2$, (iii) $\kappa_r(cX) = c^r\kappa_r(X)$, and (iv) $\kappa_r(X+Y) = \kappa_r(X) + \kappa_r(Y)$. See, e.g., Kendall and

Stuart (1969, Sec. 3.12). The results of (b) follows from the definition of the joint cumulants, and some straightforward but tedious derivations.

Proof of Lemma 4.2: Note that the rth cumulant w.r.t. the EDF $\mathcal{G}_N$ of $\{\mathbf{v}_i, i = 1, \ldots, N\}$ is just the rth sample cumulant of $\{\mathbf{v}_i, i = 1, \ldots, N\}$. This immediately gives $\kappa_1^\star(\mathbf{v}^\star) = \frac{1}{N}\sum_{i=1}^N \mathbf{v}_i = 0$.

To show $\kappa_2^\star(\mathbf{v}^\star) = \sigma_0^2 + O_p(N^{-1/2})$, note that $\mathrm{E}(\kappa_2^\star(\mathbf{v}^\star)) = \frac{1}{N}\mathrm{E}(\mathbf{V}_N'\mathbf{V}_N) = \sigma_0^2$. From Lemma 4.1, we have $\mathrm{Var}(\mathbf{v}_i^2) = k_4 a_{4,i} + 2\sigma_0^4$, $\mathrm{Cov}(\mathbf{v}_i^2, \mathbf{v}_j^2) = k_4 a_{i,i,j,j} = k_4 \sum_{m=1}^N f_{mi}^2 f_{mj}^2$, and thus

$$\begin{aligned}
\mathrm{Var}(\tfrac{1}{N}\mathbf{V}_N'\mathbf{V}_N) &= \tfrac{1}{N^2}\textstyle\sum_{i=1}^N \mathrm{Var}(\mathbf{v}_i^2) + \tfrac{2}{N^2}\sum_{i=1}^N\sum_{j\neq i}^N \mathrm{Cov}(\mathbf{v}_\mathrm{i}^2, \mathbf{v}_\mathrm{j}^2) \\
&= \tfrac{1}{N}(k_4\bar{a}_4 + 2\sigma_0^4) + \tfrac{2}{N^2}k_4 \textstyle\sum_{i=1}^N\sum_{j\neq i}^N\sum_{m=1}^N f_{mi}^2 f_{mj}^2 \\
&= \tfrac{1}{N}(k_4\bar{a}_4 + 2\sigma_0^4) + \tfrac{2}{N^2}k_4 \textstyle\sum_{i=1}^N\sum_{j=1}^N\sum_{m=1}^N f_{mi}^2 f_{mj}^2 - \tfrac{2}{N}k_4\bar{a}_4 \\
&= \tfrac{1}{N}(k_4\bar{a}_4 + 2\sigma_0^4) + \tfrac{2}{N^2}k_4 \textstyle\sum_{m=1}^N(\sum_{i=1}^N f_{mi}^2)(\sum_{j=1}^N f_{mj}^2) - \tfrac{2}{N}k_4\bar{a}_4 \\
&= O(N^{-1}),
\end{aligned}$$

due to the fact that $\sum_{i=1}^N f_{mi}^2$ is bounded, uniformly in $m = 1, 2, \ldots, nT$. It follows by the generalized Chebyshev's inequality that $\kappa_2^\star(\mathbf{v}^\star) = \sigma_0^2 + O_p(N^{-1/2})$.

For the general results with $r \geq 3$, it is easy to verify that $\mathrm{E}(\kappa_r^\star(\mathbf{v}^\star)) = k_r\bar{a}_r + O(N^{-1/2})$. By the results of Lemma 4.1 and the fact that $\sum_{i=1}^N |f_{mi}|^r$ is bounded, uniformly in $m = 1, 2, \ldots, nT$, it is straightforward, though tedious, to show that $\mathrm{Var}(\kappa_r^\star(\mathbf{v}^\star)) = O(N^{-1})$. The result thus follows.

Proof of Lemma 4.3: As $\hat{\mathbf{V}}_N$ is defined by replacing θ_0 in $\mathbf{V}_N$ by $\hat{\theta}_N$, the result follows directly from the $\sqrt{N}$-consistency of $\hat{\theta}_N$.

Proof of Lemma 4.4: The roof is trivial.

References

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

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

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

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

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

Baltagi, B., Egger, P., Pfaffermayr, 2013. A generalised spatial panel data model with random effects. *Econometric Reviews* 32, 650-685.

Baltagi, B. H., Yang, Z. L., 2013a. Standardized LM tests for spatial error dependence in linear or panel regressions. *The Econometrics Journal* 16, 103-134.

Baltagi, B. H., Yang, Z. L., 2013b. Heteroskedasticity and non-normality robust LM tests for spatial dependence. *Regional Science and Urban Economics* 43, 725-739.

Bao, Y., 2013. Finite sample bias of the QMLE in spatial autoregressive models. *Econometric Theory* 29, 68-88.

Bao, Y., Ullah, A., 2007. Finite sample properties of maximum likelihood estimator in spatial models. *Journal of Econometrics* 137, 396-413.

Bun, M. J. G., Carree, M. A., 2005. Bias-corrected estimation in dynamic panel data models. *Journal of Business and Economic Statistics* 23, 200-210.

Davidson, R., Flachaire, E., 2008. The wild bootstrap, tamed at last. *Journal of Econometrics* 146, 162

Debarsy, N., C. Ertur, 2010. Testing for spatial autocorrelation in a fixed effects panel data model. *Regional Science and Urban Economics* 40, 453-470.

Griffith, D. A. 1988. *Advanced Spatial Statistics*. Kluwer, Dordrecht, the Netherlands.

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

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

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

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

Kendall, M. G., Stuart, A., 1969. *Advanced Theory of Statistics*, Vol. 1, 3rd Ed. Griffin, London.

Kiviet, J. F., 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* 68, 53-78.

Lancaster, T., 2000. Incidental parameter problem since 1948. *Journal of Econometrics* 95, 391-413.

Lee, L. F., Yu, J., 2010. Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics* 154, 165-185.

Liu, R.Y., 1988. Bootstrap procedures under some non-I.I.D. models. *Annals of Statistics* 16, 1696

Liu, S. F., Yang, Z. L., 2014. Asymptotic distribution and finite-sample bias correction of QML estimators for spatial error dependence model. *Working Paper, Singapore Management University.*

Mammen, E., 1993. Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics* 21, 255

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

Nickell, S., 1981. Biases in dynamic models with fixed effects. *Econometrica* 39, 383-396.

Rilstone, P., Srivastava, V. K., Ullah, A., 1996. The second-order bias and mean squared error of non-linear estimators. *Journal of Econometrics* 75, 369-395.

Su, L. J., Yang Z. L., 2015. QML estimation of dynamic panel data models with spatial errors. *Journal of Econometrics* 185, 230-258.

Yang, Z. L., 2015. A general method for third-order bias and variance corrections on a nonlinear estimator. *Journal of Econometrics.* http://dx.doi.org/10.1016/j.jeconom.2014.07.003.

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

Yu, J., de Jong, R., Lee, L. F., 2012. Estimation of spatial dynamic panel data with fixed effects: The case of spatial cointegration. *Journal of Econometrics* 167, 16-37.

Yu, J., Lee, L. F., 2010. Estimation of unit root spatial dynamic panel data models. *Econometric Theory* 26, 1332-1362.

Table 1a. Empirical Mean[rmse](sd) of Estimators of λ, 2FE-SPD Model with SLD, $T = 3, \beta = (\mathbf{1}, \mathbf{1})', \sigma = 1$

λ	$\hat{\lambda}_N$	$\hat{\lambda}_N^{\text{bc2}}$	$\hat{\lambda}_N^{\text{bc3}}$	$\hat{\lambda}_N$	$\hat{\lambda}_N^{\text{bc2}}$	$\hat{\lambda}_N^{\text{bc3}}$
	(a) Queen Contiguity, REG1			(b) Group Interaction, REG2		
	Normal Error, n=50					
.50	.484[.120](.119)	.502[.120](.120)	.502[.120](.120)	.469[.095](.089)	.497[.088](.088)	.499[.088](.088)
.25	.234[.142](.141)	.248[.143](.143)	.250[.143](.143)	.210[.130](.124)	.250[.123](.123)	.251[.123](.123)
.00	-.010[.158](.158)	.001[.161](.161)	.002[.161](.161)	-.049[.167](.159)	-.001[.160](.160)	.001[.160](.160)
-.25	-.258[.161](.161)	-.251[.164](.164)	-.250[.165](.165)	-.303[.189](.182)	-.250[.184](.184)	-.248[.184](.184)
-.50	-.504[.163](.163)	-.503[.166](.166)	-.502[.167](.167)	-.565[.214](.204)	-.509[.208](.208)	-.507[.208](.208)
	Normal Mixture, n=50					
.50	.483[.119](.117)	.500[.118](.118)	.501[.118](.118)	.470[.091](.086)	.498[.084](.084)	.499[.084](.084)
.25	.238[.139](.139)	.253[.141](.141)	.254[.141](.141)	.209[.128](.121)	.248[.120](.120)	.249[.120](.120)
.00	-.013[.155](.154)	-.002[.157](.157)	-.001[.157](.157)	-.048[.160](.152)	-.001[.153](.153)	.001[.153](.153)
-.25	-.257[.158](.158)	-.251[.161](.161)	-.250[.162](.162)	-.301[.188](.181)	-.248[.182](.182)	-.247[.183](.183)
-.50	-.504[.163](.163)	-.503[.166](.166)	-.503[.167](.167)	-.556[.206](.199)	-.500[.203](.203)	-.498[.203](.203)
	Lognormal Error, n=50					
.50	.485[.111](.110)	.501[.111](.111)	.502[.111](.111)	.470[.090](.085)	.497[.083](.083)	.498[.083](.083)
.25	.239[.133](.133)	.253[.134](.134)	.254[.134](.134)	.212[.122](.116)	.249[.115](.115)	.251[.115](.115)
.00	-.010[.146](.146)	.001[.149](.149)	.002[.149](.149)	-.045[.154](.147)	.000[.147](.147)	.002[.147](.147)
-.25	-.255[.151](.151)	-.249[.154](.154)	-.248[.154](.154)	-.302[.178](.171)	-.251[.173](.173)	-.250[.173](.173)
-.50	-.498[.152](.152)	-.499[.155](.155)	-.499[.156](.156)	-.556[.204](.196)	-.503[.200](.200)	-.501[.200](.200)
	Normal Error, n=100					
.50	.493[.079](.078)	.502[.078](.078)	.502[.078](.078)	.482[.067](.065)	.500[.064](.064)	.501[.064](.064)
.25	.243[.095](.095)	.251[.095](.095)	.252[.095](.095)	.222[.096](.092)	.248[.092](.092)	.248[.092](.092)
.00	-.007[.110](.109)	.000[.110](.110)	.000[.110](.110)	-.031[.123](.119)	.000[.120](.120)	.001[.120](.120)
-.25	-.255[.114](.114)	-.250[.115](.115)	-.250[.115](.115)	-.289[.146](.141)	-.254[.143](.143)	-.253[.143](.143)
-.50	-.503[.117](.117)	-.501[.118](.118)	-.501[.118](.118)	-.538[.162](.158)	-.503[.162](.162)	-.503[.162](.162)
	Normal Mixture, n=100					
.50	.490[.078](.078)	.499[.078](.078)	.500[.078](.078)	.482[.067](.065)	.500[.065](.065)	.500[.065](.065)
.25	.241[.095](.095)	.249[.095](.095)	.250[.095](.095)	.224[.095](.091)	.250[.091](.091)	.250[.091](.091)
.00	-.006[.106](.106)	.001[.107](.107)	.002[.107](.107)	-.034[.122](.117)	-.002[.118](.118)	-.002[.118](.118)
-.25	-.255[.112](.112)	-.250[.113](.113)	-.250[.113](.113)	-.286[.144](.140)	-.251[.142](.142)	-.250[.142](.142)
-.50	-.502[.117](.117)	-.499[.119](.119)	-.499[.119](.119)	-.535[.160](.156)	-.500[.159](.159)	-.500[.159](.159)
	Lognormal Error, n=100					
.50	.492[.075](.075)	.501[.075](.075)	.501[.075](.075)	.482[.065](.062)	.500[.062](.062)	.500[.062](.062)
.25	.242[.091](.091)	.250[.091](.091)	.250[.091](.091)	.225[.093](.090)	.250[.090](.090)	.250[.090](.090)
.00	-.006[.102](.102)	.001[.103](.103)	.001[.103](.103)	-.029[.116](.113)	.001[.113](.113)	.002[.113](.113)
-.25	-.255[.110](.110)	-.250[.111](.111)	-.250[.111](.111)	-.283[.138](.134)	-.249[.136](.136)	-.248[.136](.136)
-.50	-.503[.112](.112)	-.500[.113](.113)	-.500[.113](.113)	-.526[.157](.154)	-.492[.159](.159)	-.495[.159](.159)
	Normal Error, n=500					
.50	.498[.033](.033)	.500[.033](.033)	.500[.033](.033)	.495[.034](.033)	.500[.033](.033)	.500[.033](.033)
.25	.249[.040](.040)	.251[.041](.041)	.251[.041](.041)	.242[.050](.049)	.249[.049](.049)	.249[.049](.049)
.00	-.001[.047](.047)	.000[.047](.047)	.000[.047](.047)	-.009[.065](.064)	.000[.065](.065)	.000[.065](.065)
-.25	-.252[.050](.050)	-.251[.050](.050)	-.251[.050](.050)	-.260[.080](.079)	-.249[.079](.079)	-.249[.079](.079)
-.50	-.501[.050](.050)	-.501[.050](.050)	-.501[.050](.050)	-.514[.096](.095)	-.501[.095](.095)	-.501[.095](.095)
	Normal Mixture, n=500					
.50	.498[.033](.033)	.500[.033](.033)	.500[.033](.033)	.495[.034](.033)	.500[.033](.033)	.500[.033](.033)
.25	.249[.040](.040)	.250[.040](.040)	.250[.040](.040)	.242[.050](.049)	.249[.049](.049)	.249[.049](.049)
.00	-.002[.045](.045)	-.001[.045](.045)	-.001[.045](.045)	-.007[.066](.066)	.002[.066](.066)	.002[.066](.066)
-.25	-.251[.048](.048)	-.250[.048](.048)	-.250[.048](.048)	-.261[.081](.081)	-.250[.081](.081)	-.250[.081](.081)
-.50	-.501[.050](.050)	-.500[.050](.050)	-.500[.050](.050)	-.514[.095](.094)	-.501[.094](.094)	-.501[.094](.094)
	Lognormal Error, n=500					
.50	.498[.032](.032)	.500[.032](.032)	.500[.032](.032)	.496[.034](.034)	.501[.034](.034)	.501[034](.034)
.25	.248[.040](.040)	.250[.040](.040)	.250[.040](.040)	.243[.050](.049)	.250[.049](.049)	.250[.049](.049)
.00	-.003[.046](.046)	-.001[.046](.046)	-.001[.046](.046)	-.009[.065](.064)	.000[.064](.064)	.000[.064](.064)
-.25	-.250[.048](.048)	-.249[.048](.048)	-.249[.048](.048)	-.259[.080](.080)	-.248[.080](.080)	-.248[.080](.080)
-.50	-.501[.049](.049)	-.501[.049](.049)	-.501[.049](.049)	-.514[.095](.094)	-.501[.095](.095)	-.501[.095](.095)

Table 1b. Empirical Mean[rmse](sd) of Estimators of λ, 2FE-SPD Model with SLD, $T = 3, \beta = (.\mathbf{5}, .\mathbf{5})', \sigma = 1$

λ	$\hat{\lambda}_N$	$\hat{\lambda}_N^{\text{bc2}}$	$\hat{\lambda}_N^{\text{bc3}}$	$\hat{\lambda}_N$	$\hat{\lambda}_N^{\text{bc2}}$	$\hat{\lambda}_N^{\text{bc3}}$
	(a) Queen Contiguity, REG1			(b) Group Interaction, REG2		
	Normal Error, n=50					
.50	.477[.133](.132)	.500[.133](.133)	.500[.132](.132)	.449[.122](.111)	.498[.105](.105)	.500[.105](.105)
.25	.231[.157](.156)	.251[.159](.159)	.252[.158](.158)	.179[.171](.156)	.248[.150](.150)	.250[.150](.150)
.00	-.015[.176](.175)	.000[.180](.180)	.002[.180](.180)	-.086[.214](.196)	-.002[.191](.191)	.001[.191](.191)
-.25	-.261[.180](.180)	-.252[.185](.185)	-.251[.185](.185)	-.348[.247](.227)	-.252[.224](.224)	-.249[.224](.224)
-.50	-.505[.185](.184)	-.502[.190](.190)	-.501[.190](.190)	-.609[.283](.262)	-.504[.261](.261)	-.502[.262](.262)
	Normal Mixture, n=50					
.50	.478[.133](.132)	.501[.133](.133)	.500[.132](.132)	.449[.120](.109)	.498[.103](.103)	.500[.103](.103)
.25	.229[.158](.157)	.248[.159](.159)	.249[.159](.159)	.180[.168](.153)	.248[.147](.147)	.250[.147](.147)
.00	-.017[.174](.173)	-.002[.177](.177)	.000[.177](.177)	-.088[.212](.193)	-.003[.188](.188)	.000[.188](.188)
-.25	-.260[.176](.176)	-.251[.181](.181)	-.250[.181](.181)	-.346[.247](.227)	-.250[.224](.224)	-.247[.225](.225)
-.50	-.502[.181](.181)	-.499[.186](.186)	-.499[.186](.186)	-.608[.281](.260)	-.503[.260](.260)	-.500[.260](.260)
	Lognormal Error, n=50					
.50	.480[.123](.122)	.502[.123](.123)	.502[.122](.122)	.454[.112](.102)	.502[.097](.097)	.504[.097](.097)
.25	.229[.148](.147)	.249[.150](.149)	.250[.149](.149)	.184[.157](.143)	.251[.138](.138)	.254[.138](.138)
.00	-.013[.162](.161)	.002[.165](.165)	.003[.165](.165)	-.079[.193](.176)	.003[.172](.172)	.006[.172](.172)
-.25	-.258[.168](.167)	-.248[.172](.172)	-.247[.172](.172)	-.341[.225](.206)	-.247[.203](.203)	-.244[.203](.203)
-.50	-.504[.173](.172)	-.501[.177](.177)	-.501[.178](.178)	-.598[.258](.239)	-.495[.239](.239)	-.493[.240](.240)
	Normal Error, n=100					
.50	.490[.090](.090)	.502[.090](.090)	.502[.089](.089)	.469[.087](.081)	.499[.079](.079)	.500[.079](.079)
.25	.242[.108](.108)	.253[.109](.109)	.253[.109](.109)	.205[.127](.119)	.248[.117](.117)	.248[.117](.117)
.00	-.003[.122](.122)	.006[.123](.123)	.006[.123](.123)	-.058[.166](.155)	-.004[.153](.153)	-.003[.153](.153)
-.25	-.256[.130](.129)	-.250[.131](.131)	-.249[.131](.131)	-.313[.192](.181)	-.249[.179](.179)	-.249[.179](.179)
-.50	-.505[.131](.131)	-.503[.133](.133)	-.503[.133](.133)	-.578[.223](.209)	-.506[.209](.208)	-.506[.209](.209)
	Normal Mixture, n=100					
.50	.491[.088](.088)	.502[.088](.088)	.502[.088](.088)	.470[.087](.082)	.500[.080](.080)	.500[.079](.079)
.25	.241[.105](.105)	.252[.106](.106)	.252[.106](.106)	.207[.124](.116)	.249[.113](.113)	.250[.113](.113)
.00	-.010[.120](.120)	-.002[.121](.121)	-.001[.121](.121)	-.056[.160](.150)	-.001[.148](.148)	-.001[.148](.148)
-.25	-.254[.129](.129)	-.248[.131](.131)	-.247[.131](.131)	-.314[.195](.184)	-.251[.182](.182)	-.250[.182](.182)
-.50	-.503[.130](.130)	-.500[.131](.131)	-.500[.132](.132)	-.567[.217](.207)	-.496[.206](.206)	-.495[.206](.206)
	Lognormal Error, n=100					
.50	.490[.084](.084)	.502[.084](.084)	.502[.084](.084)	.470[.084](.079)	.500[.077](.077)	.500[.077](.077)
.25	.235[.102](.101)	.246[.102](.102)	.246[.102](.102)	.208[.120](.113)	.250[.110](.110)	.251[.110](.110)
.00	-.005[.116](.116)	.004[.117](.117)	.004[.117](.117)	-.050[.151](.143)	.003[.141](.141)	.004[.141](.141)
-.25	-.258[.121](.121)	-.252[.123](.123)	-.252[.123](.123)	-.316[.185](.172)	-.253[.171](.171)	-.253[.171](.171)
-.50	-.502[.125](.125)	-.499[.126](.126)	-.499[.126](.126)	-.565[.208](.197)	-.495[.197](.197)	-.495[.197](.197)
	Normal Error, n=500					
.50	.498[.039](.039)	.500[.039](.039)	.500[.039](.039)	.490[.050](.049)	.501[.048](.048)	.501[.048](.048)
.25	.247[.048](.048)	.250[.048](.048)	.250[.048](.048)	.234[.073](.071)	.250[.071](.071)	.250[.071](.071)
.00	-.001[.055](.055)	.001[.055](.055)	.001[.055](.055)	-.021[.097](.094)	.000[.094](.094)	.000[.094](.094)
-.25	-.251[.058](.058)	-.250[.058](.058)	-.250[.058](.058)	-.275[.117](.114)	-.249[.113](.113)	-.249[.113](.113)
-.50	-.500[.060](.060)	-.499[.061](.061)	-.499[.061](.061)	-.530[.139](.136)	-.500[.135](.135)	-.500[.135](.135)
	Normal Mixture, n=500					
.50	.499[.039](.039)	.501[.039](.039)	.501[.039](.039)	.490[.048](.047)	.501[.047](.047)	.501[.047](.047)
.25	.247[.048](.048)	.249[.048](.048)	.249[.048](.048)	.233[.074](.072)	.249[.071](.071)	.249[.071](.071)
.00	.000[.054](.054)	.002[.055](.055)	.002[.055](.055)	-.020[.095](.093)	.002[.092](.092)	.002[.092](.092)
-.25	-.250[.059](.059)	-.249[.059](.059)	-.249[.059](.059)	-.279[.119](.116)	-.253[.115](.115)	-.253[.115](.115)
-.50	-.501[.059](.059)	-.500[.060](.060)	-.500[.060](.060)	-.529[.137](.134)	-.499[.133](.133)	-.499[.133](.133)
	Lognormal Error, n=500					
.50	.497[.037](.037)	.500[.037](.037)	.500[.037](.037)	.491[.047](.046)	.502[.046](.046)	.502[.046](.046)
.25	.248[.048](.048)	.250[.048](.048)	.250[.048](.048)	.234[.072](.070)	.251[.069](.069)	.251[.069](.069)
.00	-.002[.053](.053)	.000[.053](.053)	.000[.053](.053)	-.020[.094](.092)	.001[.091](.091)	.001[.091](.091)
-.25	-.252[.057](.057)	-.251[.058](.058)	-.251[.058](.058)	-.277[.116](.112)	-.250[.112](.112)	-.251[.112](.112)
-.50	-.499[.059](.059)	-.499[.059](.059)	-.499[.059](.059)	-.530[.139](.136)	-.498[.135](.135)	-.499[.135](.135)

Table 2. Empirical Mean[rmse](sd) of Estimators of λ - 2FE-SPD Model with SED, $T = 3, \beta = (1, 1)', \sigma = 1$

λ	$\hat{\lambda}_N$	$\hat{\lambda}_N^{\text{bc2}}$	$\hat{\lambda}_N^{\text{bc3}}$	$\hat{\lambda}_N$	$\hat{\lambda}_N^{\text{bc2}}$	$\hat{\lambda}_N^{\text{bc3}}$
	(a) Queen Contiguity, REG1			(b) Group Interaction, REG2		
	Normal Error, n=50					
.50	.481[.144](.142)	.500[.143](.143)	.500[.142](.142)	.457[.139](.132)	.503[.116](.116)	.503[.115](.115)
.25	.233[.171](.170)	.252[.171](.171)	.254[.171](.171)	.177[.202](.188)	.258[.167](.167)	.260[.167](.166)
.00	-.018[.190](.189)	-.001[.190](.190)	.001[.191](.190)	-.115[.266](.240)	-.004[.221](.221)	-.001[.220](.220)
-.25	-.271[.202](.201)	-.255[.203](.203)	-.254[.204](.204)	-.382[.299](.268)	-.250[.256](.256)	-.249[.256](.256)
-.50	-.516[.203](.202)	-.503[.205](.205)	-.502[.206](.206)	-.637[.321](.290)	-.496[.287](.287)	-.497[.288](.288)
	Normal Mixture, n=50					
.50	.480[.139](.138)	.500[.138](.138)	.500[.137](.137)	.458[.137](.130)	.504[.114](.114)	.504[.113](.113)
.25	.233[.166](.165)	.252[.166](.166)	.251[.166](.166)	.168[.210](.194)	.251[.172](.172)	.250[.171](.171)
.00	-.016[.186](.185)	.002[.186](.186)	.003[.186](.186)	-.108[.258](.234)	.004[.214](.214)	.003[.214](.214)
-.25	-.267[.195](.194)	-.252[.196](.196)	-.250[.197](.197)	-.381[.293](.262)	-.248[.251](.251)	-.249[.251](.251)
-.50	-.511[.198](.197)	-.498[.200](.200)	-.498[.201](.201)	-.636[.313](.282)	-.493[.280](.280)	-.495[.281](.281)
	Lognormal Error, n=50					
.50	.483[.135](.133)	.504[.134](.134)	.503[.133](.133)	.454[.136](.128)	.502[.112](.112)	.502[.111](.111)
.25	.237[.160](.159)	.256[.161](.160)	.255[.160](.160)	.174[.196](.181)	.257[.160](.160)	.256[.160](.160)
.00	-.012[.179](.179)	.006[.180](.180)	.005[.180](.180)	-.105[.242](.218)	.009[.199](.199)	.002[.199](.199)
-.25	-.264[.186](.186)	-.248[.188](.188)	-.249[.188](.188)	-.368[.273](.247)	-.233[.235](.235)	-.239[.236](.235)
-.50	-.512[.191](.191)	-.499[.194](.194)	-.499[.194](.194)	-.632[.305](.275)	-.489[.272](.272)	-.489[.274](.273)
	Normal Error, n=100					
.50	.490[.096](.095)	.500[.095](.095)	.500[.095](.095)	.467[.107](.102)	.501[.093](.093)	.501[.093](.093)
.25	.241[.119](.119)	.251[.119](.119)	.251[.118](.118)	.196[.152](.142)	.252[.132](.132)	.251[.132](.132)
.00	-.011[.132](.132)	-.001[.132](.132)	.000[.132](.132)	-.074[.192](.177)	-.002[.171](.171)	-.002[.171](.171)
-.25	-.259[.141](.140)	-.249[.141](.141)	-.249[.141](.141)	-.333[.215](.199)	-.255[.199](.199)	-.255[.199](.199)
-.50	-.510[.142](.142)	-.501[.143](.143)	-.501[.143](.143)	-.574[.220](.207)	-.500[.215](.215)	-.500[.215](.215)
	Normal Mixture, n=100					
.50	.489[.095](.094)	.500[.094](.094)	.500[.094](.094)	.465[.104](.098)	.500[.090](.090)	.500[.090](.090)
.25	.240[.118](.117)	.250[.117](.117)	.250[.117](.117)	.196[.149](.139)	.253[.130](.130)	.253[.130](.130)
.00	-.010[.130](.130)	.001[.130](.130)	.001[.130](.130)	-.073[.189](.174)	.000[.168](.168)	.000[.168](.168)
-.25	-.260[.138](.138)	-.250[.138](.138)	-.249[.138](.138)	-.327[.211](.196)	-.249[.197](.197)	-.249[.197](.197)
-.50	-.510[.138](.138)	-.501[.139](.139)	-.501[.139](.139)	-.569[.220](.209)	-.495[.219](.219)	-.495[.219](.219)
	Lognormal Error, n=100					
.50	.494[.088](.088)	.505[.088](.088)	.505[.088](.088)	.465[.107](.101)	.501[.092](.092)	.500[.092](.092)
.25	.240[.110](.110)	.251[.110](.110)	.251[.110](.110)	.198[.145](.135)	.256[.126](.126)	.256[.126](.125)
.00	-.006[.126](.126)	.004[.127](.126)	.003[.127](.126)	-.064[.174](.162)	.010[.156](.156)	.010[.156](.156)
-.25	-.259[.136](.136)	-.250[.136](.136)	-.249[.136](.136)	-.320[.200](.188)	-.239[.189](.188)	-.239[.189](.189)
-.50	-.508[.135](.135)	-.500[.136](.136)	-.500[.136](.136)	-.561[.214](.205)	-.485[.215](.215)	-.486[.215](.215)
	Normal Error, n=500					
.50	.497[.041](.041)	.499[.041](.041)	.499[.041](.041)	.487[.060](.059)	.500[.057](.057)	.500[.057](.057)
.25	.249[.051](.051)	.251[.051](.051)	.251[.051](.051)	.226[.090](.087)	.249[.083](.083)	.249[.083](.083)
.00	-.003[.058](.058)	-.001[.058](.058)	-.001[.058](.058)	-.033[.121](.116)	.000[.112](.112)	.000[.112](.112)
-.25	-.252[.062](.061)	-.250[.062](.062)	-.250[.062](.062)	-.292[.148](.142)	-.249[.137](.137)	-.249[.137](.137)
-.50	-.500[.063](.063)	-.499[.063](.063)	-.499[.063](.063)	-.549[.170](.162)	-.499[.158](.158)	-.499[.158](.158)
	Normal Mixture, n=500					
.50	.498[.040](.040)	.500[.040](.040)	.500[.040](.040)	.485[.060](.058)	.499[.056](.056)	.499[.056](.056)
.25	.247[.051](.051)	.250[.051](.051)	.250[.051](.051)	.226[.091](.088)	.250[.084](.084)	.249[.084](.084)
.00	-.001[.058](.058)	.001[.058](.058)	.001[.058](.058)	-.035[.120](.114)	-.001[.110](.110)	-.002[.110](.110)
-.25	-.252[.062](.062)	-.250[.062](.062)	-.250[.062](.062)	-.291[.146](.140)	-.249[.136](.136)	-.249[.136](.136)
-.50	-.504[.063](.063)	-.502[.063](.063)	-.502[.063](.063)	-.551[.173](.165)	-.500[.161](.161)	-.500[.161](.161)
	Lognormal Error, n=500					
.50	.498[.040](.040)	.500[.040](.040)	.500[.040](.040)	.485[.062](.060)	.500[.058](.058)	.499[.058](.058)
.25	.249[.050](.050)	.251[.050](.050)	.251[.050](.050)	.227[.088](.085)	.252[.081](.081)	.252[.081](.081)
.00	-.003[.057](.056)	-.001[.056](.056)	-.001[.056](.056)	-.030[.112](.108)	.006[.104](.104)	.005[.104](.104)
-.25	-.251[.060](.060)	-.249[.060](.060)	-.249[.060](.060)	-.290[.141](.135)	-.245[.131](.130)	-.246[.130](.130)
-.50	-.503[.062](.062)	-.501[.062](.062)	-.501[.062](.062)	-.545[.168](.162)	-.492[.158](.157)	-.493[.158](.157)

Table 3a. Empirical Mean[rmse](sd) of Estimators of λ and ρ, 2FE-SPD Model with SARAR, $T = 3, \beta = (1, 1)', \sigma = 1$, `Queen Contiguity, REG-1`

λ	ρ	$\hat{\lambda}_N$	$\hat{\lambda}_N^{\mathrm{bc2}}$	$\hat{\rho}_N$	$\hat{\rho}_N^{\mathrm{bc2}}$	$\hat{\lambda}_N$	$\hat{\lambda}_N^{\mathrm{bc2}}$	$\hat{\rho}_N$	$\hat{\rho}_N^{\mathrm{bc2}}$
		(a) Normal Error, $n = 50$				(b) Lognormal Error, $n = 50$			
.50	.50	.484[.116](.115)	.500[.116](.116)	.483[.143](.142)	.500[.143](.143)	.486[.105](.104)	.502[.105](.105)	.484[.131](.130)	.502[.131](.131)
	.25	.484[.119](.117)	.501[.118](.118)	.226[.176](.174)	.242[.175](.175)	.485[.114](.113)	.501[.113](.113)	.233[.162](.161)	.250[.161](.161)
	.00	.483[.118](.116)	.500[.117](.117)	-.019[.192](.191)	-.002[.192](.192)	.486[.110](.109)	.503[.110](.110)	-.015[.177](.176)	.002[.177](.177)
	-.25	.482[.124](.122)	.500[.123](.123)	-.267[.202](.202)	-.251[.203](.203)	.487[.112](.111)	.503[.112](.112)	-.265[.193](.193)	-.249[.193](.193)
	-.50	.484[.125](.123)	.500[.124](.124)	-.513[.208](.208)	-.498[.209](.209)	.489[.111](.110)	.505[.111](.111)	-.514[.195](.194)	-.499[.196](.196)
-.50	.50	-.502[.158](.158)	-.500[.161](.161)	.486[.144](.143)	.504[.144](.144)	-.502[.145](.145)	-.500[.148](.148)	.486[.132](.131)	.504[.132](.131)
	.25	-.506[.165](.165)	-.504[.168](.168)	.232[.174](.173)	.249[.174](.174)	-.505[.152](.151)	-.503[.155](.154)	.233[.161](.160)	.250[.160](.160)
	.00	-.501[.163](.163)	-.499[.167](.167)	-.006[.187](.187)	.010[.187](.187)	-.499[.159](.159)	-.497[.162](.162)	-.018[.180](.179)	-.001[.180](.180)
	-.25	-.500[.164](.164)	-.498[.168](.168)	-.262[.209](.209)	-.246[.210](.210)	-.501[.152](.152)	-.499[.155](.155)	-.263[.197](.197)	-.246[.197](.197)
	-.50	-.506[.169](.169)	-.505[.172](.172)	-.518[.207](.206)	-.503[.208](.208)	-.498[.157](.157)	-.497[.160](.160)	-.513[.194](.194)	-.498[.195](.195)
		(c) Normal Error, $n = 100$				(d) Lognormal Error, $n = 100$			
.50	.50	.494[.078](.077)	.502[.078](.078)	.490[.096](.096)	.499[.096](.096)	.490[.078](.078)	.499[.078](.078)	.493[.090](.090)	.502[.090](.090)
	.25	.490[.080](.080)	.499[.080](.080)	.244[.117](.116)	.253[.117](.117)	.491[.081](.080)	.500[.080](.080)	.243[.111](.111)	.252[.111](.111)
	.00	.493[.083](.083)	.502[.083](.083)	-.011[.132](.131)	-.002[.131](.131)	.494[.079](.079)	.503[.079](.079)	-.009[.126](.126)	.001[.126](.126)
	-.25	.491[.084](.083)	.500[.083](.083)	-.258[.142](.142)	-.249[.142](.142)	.490[.077](.077)	.499[.077](.077)	-.264[.138](.137)	-.254[.138](.137)
	-.50	.490[.079](.078)	.499[.078](.078)	-.509[.142](.141)	-.499[.142](.142)	.493[.077](.077)	.501[.077](.077)	-.509[.137](.137)	-.499[.137](.137)
-.50	.50	-.494[.118](.118)	-.493[.119](.119)	.492[.094](.094)	.501[.094](.094)	-.503[.106](.106)	-.503[.107](.107)	.491[.089](.088)	.500[.088](.088)
	.25	-.501[.119](.119)	-.500[.121](.121)	.242[.117](.117)	.251[.117](.117)	-.502[.112](.112)	-.501[.113](.113)	.240[.111](.111)	.249[.111](.111)
	.00	-.496[.115](.115)	-.495[.117](.117)	-.008[.133](.133)	.001[.133](.133)	-.498[.114](.114)	-.498[.115](.115)	-.007[.129](.129)	.003[.128](.128)
	-.25	-.505[.118](.118)	-.504[.120](.120)	-.258[.143](.143)	-.248[.143](.143)	-.497[.112](.112)	-.496[.113](.113)	-.257[.136](.136)	-.248[.136](.136)
	-.50	-.501[.118](.118)	-.500[.120](.120)	-.504[.148](.148)	-.495[.149](.149)	-.505[.109](.109)	-.504[.110](.110)	-.507[.137](.137)	-.498[.138](.137)
		(e) Normal Error, $n = 500$				(f) Lognormal Error, $n = 500$			
.50	.50	.497[.033](.033)	.499[.033](.033)	.499[.041](.041)	.501[.041](.041)	.499[.030](.030)	.501[.030](.030)	.497[.040](.040)	.499[.040](.040)
	.25	.497[.033](.033)	.499[.033](.033)	.247[.052](.052)	.249[.052](.052)	.499[.032](.032)	.501[.032](.032)	.249[.050](.050)	.250[.050](.050)
	.00	.499[.033](.033)	.501[.033](.033)	.001[.057](.057)	.003[.058](.057)	.498[.033](.033)	.500[.033](.033)	-.001[.057](.057)	.001[.057](.057)
	-.25	.498[.033](.032)	.499[.033](.033)	-.254[.062](.062)	-.252[.062](.062)	.498[.033](.033)	.500[.033](.033)	-.250[.061](.061)	-.248[.061](.061)
	-.50	.498[.032](.032)	.500[.032](.032)	-.503[.062](.062)	-.501[.062](.062)	.497[.032](.032)	.499[.032](.032)	-.501[.062](.062)	-.499[.062](.062)
-.50	.50	-.502[.049](.049)	-.501[.049](.049)	.498[.041](.041)	.500[.041](.041)	-.499[.049](.049)	-.499[.049](.049)	.498[.040](.040)	.500[.040](.040)
	.25	-.503[.051](.051)	-.502[.051](.051)	.249[.051](.051)	.250[.051](.051)	-.500[.051](.051)	-.499[.051](.051)	.248[.050](.050)	.250[.050](.050)
	.00	-.501[.050](.050)	-.501[.050](.050)	-.001[.060](.060)	.001[.060](.060)	-.501[.051](.051)	-.500[.052](.052)	-.002[.058](.058)	.000[.058](.058)
	-.25	-.502[.051](.050)	-.502[.051](.051)	-.253[.061](.061)	-.251[.061](.061)	-.499[.051](.051)	-.498[.051](.051)	-.252[.062](.062)	-.250[.062](.062)
	-.50	-.500[.049](.049)	-.499[.049](.049)	-.501[.063](.063)	-.499[.064](.064)	-.500[.048](.048)	-.500[.049](.049)	-.503[.062](.062)	-.502[.062](.062)

Table 3b. Empirical Mean[rmse](sd) of Estimators of λ and ρ, 2FE-SPD Model with SARAR, $T = 3, \beta = (1,1)', \sigma = 1$, `Group Interaction, REG-2`

λ	ρ	$\hat{\lambda}_N$	$\hat{\lambda}_N^{\mathrm{bc2}}$	$\hat{\rho}_N$	$\hat{\rho}_N^{\mathrm{bc2}}$	$\hat{\lambda}_N$	$\hat{\lambda}_N^{\mathrm{bc2}}$	$\hat{\rho}_N$	$\hat{\rho}_N^{\mathrm{bc2}}$
		(a) Normal Error, $n = 50$				(b) Lognormal Error, $n = 50$			
.50	.50	.484[.095](.094)	.499[.092](.092)	.453[.156](.149)	.500[.129](.129)	.484[.089](.088)	.500[.087](.087)	.456[.146](.140)	.505[.121](.121)
	.25	.480[.103](.101)	.497[.099](.099)	.162[.238](.221)	.248[.194](.194)	.484[.096](.095)	.501[.093](.093)	.161[.237](.220)	.251[.193](.193)
	.00	.481[.104](.102)	.498[.100](.100)	-.120[.298](.272)	.001[.243](.243)	.486[.097](.096)	.501[.093](.093)	-.120[.301](.276)	.005[.247](.247)
	-.25	.481[.104](.102)	.496[.100](.100)	-.408[.362](.326)	-.257[.299](.299)	.488[.097](.096)	.502[.094](.094)	-.407[.365](.330)	-.252[.306](.306)
	-.50	.484[.099](.098)	.498[.096](.096)	-.685[.400](.354)	-.512[.335](.334)	.491[.095](.095)	.504[.093](.093)	-.682[.413](.370)	-.506[.354](.354)
-.50	.50	-.527[.218](.216)	-.499[.218](.218)	.453[.158](.150)	.501[.130](.130)	-.522[.214](.213)	-.494[.215](.215)	.458[.147](.141)	.507[.123](.122)
	.25	-.534[.237](.235)	-.505[.237](.236)	.164[.235](.219)	.251[.191](.191)	-.524[.226](.225)	-.495[.227](.227)	.171[.220](.205)	.259[.179](.179)
	.00	-.532[.239](.237)	-.504[.239](.239)	-.117[.301](.277)	.004[.249](.249)	-.528[.239](.237)	-.501[.239](.239)	-.114[.293](.270)	.010[.242](.242)
	-.25	-.530[.237](.235)	-.504[.237](.237)	-.407[.357](.320)	-.257[.295](.295)	-.519[.240](.240)	-.494[.241](.241)	-.396[.349](.317)	-.243[.293](.293)
	-.50	-.524[.233](.232)	-.500[.233](.233)	-.689[.403](.355)	-.518[.337](.336)	-.528[.251](.250)	-.505[.252](.252)	-.661[.399](.364)	-.489[.345](.345)
		(c) Normal Error, $n = 250$				(d) Lognormal Error, $n = 250$			
.50	.50	.497[.044](.044)	.501[.044](.044)	.477[.082](.079)	.500[.074](.074)	.497[.043](.043)	.500[.042](.042)	.477[.081](.078)	.500[.073](.073)
	.25	.497[.043](.043)	.500[.043](.043)	.209[.124](.117)	.250[.110](.110)	.497[.042](.042)	.500[.042](.042)	.209[.119](.112)	.250[.105](.105)
	.00	.497[.041](.040)	.499[.040](.040)	-.056[.161](.151)	.001[.142](.142)	.498[.040](.040)	.500[.039](.039)	-.056[.158](.148)	.002[.138](.138)
	-.25	.498[.038](.038)	.500[.038](.038)	-.327[.204](.189)	-.253[.178](.178)	.498[.038](.038)	.500[.038](.038)	-.322[.194](.180)	-.247[.169](.169)
	-.50	.499[.035](.035)	.500[.035](.035)	-.590[.232](.214)	-.501[.203](.203)	.500[.035](.035)	.501[.035](.035)	-.588[.229](.211)	-.497[.200](.200)
-.50	.50	-.508[.123](.122)	-.498[.122](.122)	.476[.082](.078)	.499[.073](.073)	-.509[.122](.121)	-.498[.121](.121)	.476[.081](.078)	.500[.073](.073)
	.25	-.510[.118](.118)	-.502[.118](.118)	.213[.121](.115)	.253[.108](.108)	-.504[.118](.118)	-.496[.118](.118)	.210[.120](.113)	.251[.106](.106)
	.00	-.507[.116](.116)	-.500[.116](.116)	-.063[.167](.155)	-.005[.146](.146)	-.509[.113](.113)	-.502[.113](.113)	-.058[.161](.150)	.000[.140](.140)
	-.25	-.502[.105](.105)	-.497[.105](.105)	-.326[.201](.186)	-.252[.175](.175)	-.507[.105](.105)	-.502[.105](.105)	-.320[.192](.179)	-.245[.169](.169)
	-.50	-.506[.099](.099)	-.502[.099](.099)	-.592[.235](.216)	-.503[.204](.204)	-.503[.100](.100)	-.499[.100](.100)	-.589[.234](.217)	-.498[.205](.205)
		(e) Normal Error, $n = 500$				(f) Lognormal Error, $n = 500$			
.50	.50	.498[.030](.030)	.500[.030](.030)	.484[.065](.063)	.500[.060](.060)	.498[.030](.030)	.500[.030](.030)	.484[.065](.063)	.501[.060](.060)
	.25	.499[.029](.029)	.500[.029](.029)	.220[.098](.093)	.248[.089](.089)	.498[.029](.029)	.500[.029](.029)	.223[.096](.092)	.252[.087](.087)
	.00	.500[.027](.027)	.501[.027](.027)	-.040[.128](.122)	.001[.116](.116)	.500[.027](.027)	.501[.027](.027)	-.044[.128](.120)	-.001[.114](.114)
	-.25	.500[.025](.025)	.501[.025](.025)	-.303[.160](.151)	-.249[.144](.144)	.500[.025](.025)	.501[.025](.025)	-.305[.158](.148)	-.249[.141](.141)
	-.50	.499[.023](.023)	.500[.023](.023)	-.562[.187](.176)	-.496[.168](.168)	.499[.022](.022)	.500[.022](.022)	-.565[.192](.180)	-.497[.172](.172)
-.50	.50	-.505[.087](.087)	-.500[.087](.087)	.485[.065](.063)	.500[.060](.060)	-.505[.085](.085)	-.499[.085](.085)	.484[.064](.062)	.501[.059](.059)
	.25	-.507[.082](.082)	-.503[.082](.082)	.220[.098](.094)	.248[.089](.089)	-.504[.081](.081)	-.500[.081](.081)	.223[.096](.092)	.252[.088](.088)
	.00	-.503[.075](.075)	-.500[.075](.075)	-.041[.131](.124)	.000[.118](.118)	-.502[.075](.075)	-.499[.075](.075)	-.044[.127](.119)	-.001[.113](.113)
	-.25	-.504[.070](.070)	-.502[.070](.070)	-.303[.161](.152)	-.249[.145](.145)	-.501[.071](.071)	-.499[.071](.071)	-.303[.159](.150)	-.248[.143](.143)
	-.50	-.501[.065](.065)	-.499[.065](.065)	-.569[.192](.179)	-.503[.171](.171)	-.502[.065](.065)	-.500[.065](.065)	-.562[.187](.176)	-.494[.168](.168)

Table 4. Empirical Means of the Non-Spatial Estimators, 2FE-SPD Model with SLD

Group Interaction, REG2, $T = 3$

λ	$\hat{\beta}_{1N}$	$\hat{\beta}_{2N}$	$\hat{\sigma}^2_N$	$\hat{\beta}^{\rm bc}_{1N}$	$\hat{\beta}^{\rm bc}_{2N}$	$\hat{\sigma}^{2,\rm bc}_N$	$\hat{\beta}_{1N}$	$\hat{\beta}_{2N}$	$\hat{\sigma}^2_N$	$\hat{\beta}^{\rm bc}_{1N}$	$\hat{\beta}^{\rm bc}_{2N}$	$\hat{\sigma}^{2,\rm bc}_N$
	(a) $\beta = (1,1)', \sigma = 1$						(b) $\beta = (.5,.5)', \sigma = 1$					
	Normal Error, n=50											
.50	1.041	1.035	0.984	0.996	0.998	0.992	0.533	0.530	0.985	0.496	0.499	0.991
.25	1.039	1.030	0.982	0.997	0.995	0.992	0.532	0.524	0.981	0.498	0.496	0.991
.00	1.035	1.023	0.980	0.997	0.992	0.992	0.529	0.519	0.978	0.498	0.494	0.991
-.25	1.032	1.023	0.978	0.997	0.995	0.992	0.524	0.519	0.975	0.496	0.496	0.992
-.50	1.030	1.019	0.974	0.999	0.994	0.989	0.527	0.514	0.970	0.501	0.494	0.990
	Normal Mixture, n=50											
.50	1.040	1.031	0.975	0.996	0.994	0.982	0.532	0.520	0.981	0.495	0.490	0.988
.25	1.041	1.030	0.973	1.000	0.996	0.982	0.531	0.523	0.973	0.497	0.495	0.983
.00	1.038	1.030	0.973	1.001	0.998	0.984	0.526	0.518	0.973	0.495	0.493	0.986
-.25	1.035	1.025	0.966	1.001	0.997	0.980	0.524	0.515	0.963	0.496	0.492	0.979
-.50	1.028	1.023	0.969	0.997	0.997	0.985	0.521	0.520	0.962	0.496	0.500	0.981
	Lognormal Error, n=50											
.50	1.036	1.031	0.944	0.994	0.995	0.951	0.529	0.523	0.946	0.493	0.493	0.952
.25	1.036	1.032	0.947	0.996	0.999	0.957	0.529	0.521	0.946	0.496	0.494	0.956
.00	1.028	1.020	0.936	0.992	0.990	0.947	0.525	0.519	0.944	0.495	0.494	0.957
-.25	1.029	1.019	0.942	0.996	0.992	0.955	0.522	0.517	0.943	0.494	0.494	0.959
-.50	1.026	1.017	0.940	0.996	0.993	0.956	0.518	0.514	0.926	0.494	0.494	0.945
	Normal Error, n=100											
.50	1.028	1.023	0.993	1.000	0.999	0.996	0.526	0.521	0.993	0.501	0.499	0.996
.25	1.027	1.019	0.991	1.000	0.996	0.995	0.524	0.517	0.990	0.500	0.496	0.995
.00	1.023	1.020	0.990	0.998	0.999	0.996	0.524	0.516	0.991	0.501	0.496	0.997
-.25	1.020	1.020	0.989	0.996	1.000	0.995	0.521	0.514	0.988	0.499	0.496	0.995
-.50	1.024	1.018	0.988	1.002	0.999	0.995	0.520	0.514	0.986	0.500	0.497	0.994
	Normal Mixture, n=100											
.50	1.026	1.022	0.990	0.998	0.998	0.993	0.523	0.518	0.988	0.497	0.497	0.991
.25	1.024	1.019	0.987	0.998	0.996	0.992	0.525	0.519	0.986	0.501	0.498	0.990
.00	1.022	1.018	0.985	0.997	0.996	0.990	0.522	0.515	0.985	0.499	0.496	0.991
-.25	1.023	1.018	0.987	1.000	0.998	0.994	0.523	0.517	0.983	0.501	0.499	0.991
-.50	1.022	1.019	0.982	1.000	1.001	0.989	0.518	0.515	0.983	0.498	0.498	0.992
	Lognormal Error, n=100											
.50	1.024	1.021	0.973	0.997	0.998	0.977	0.524	0.518	0.969	0.499	0.497	0.972
.25	1.025	1.023	0.964	1.000	1.002	0.968	0.522	0.516	0.966	0.498	0.496	0.971
.00	1.023	1.015	0.963	0.999	0.995	0.969	0.520	0.514	0.962	0.497	0.495	0.968
-.25	1.022	1.016	0.970	0.999	0.997	0.977	0.520	0.516	0.964	0.499	0.498	0.972
-.50	1.021	1.012	0.960	1.000	0.995	0.966	0.516	0.514	0.958	0.497	0.498	0.967
	Normal Error, n=250											
.50	1.011	1.010	0.997	0.999	0.998	0.999	0.512	0.512	0.997	0.499	0.499	0.998
.25	1.010	1.009	0.996	0.998	0.997	0.998	0.512	0.512	0.996	0.500	0.500	0.998
.00	1.009	1.009	0.996	0.998	0.997	0.998	0.509	0.509	0.996	0.497	0.497	0.998
-.25	1.009	1.010	0.996	0.997	0.998	0.999	0.508	0.511	0.995	0.497	0.500	0.998
-.50	1.009	1.010	0.995	0.998	0.999	0.998	0.511	0.510	0.994	0.500	0.499	0.997
	Normal Mixture, n=250											
.50	1.014	1.013	0.997	1.002	1.000	0.998	0.513	0.509	0.996	0.500	0.497	0.997
.25	1.012	1.010	0.993	1.000	0.998	0.995	0.512	0.511	0.995	0.500	0.498	0.996
.00	1.010	1.011	0.995	0.998	0.999	0.997	0.510	0.512	0.993	0.498	0.500	0.996
-.25	1.012	1.011	0.996	1.001	1.000	0.998	0.510	0.510	0.997	0.498	0.498	1.000
-.50	1.009	1.008	0.994	0.998	0.997	0.996	0.510	0.509	0.993	0.499	0.498	0.996
	Lognormal Error, n=250											
.50	1.011	1.010	0.986	0.999	0.998	0.987	0.511	0.511	0.982	0.498	0.498	0.983
.25	1.012	1.013	0.985	1.000	1.001	0.987	0.513	0.513	0.986	0.501	0.501	0.988
.00	1.010	1.009	0.983	0.998	0.998	0.985	0.511	0.511	0.984	0.499	0.499	0.987
-.25	1.010	1.009	0.982	0.999	0.997	0.985	0.512	0.510	0.984	0.500	0.498	0.987
-.50	1.007	1.007	0.985	0.996	0.997	0.987	0.509	0.508	0.983	0.498	0.497	0.986

Table 5a. Empirical Sizes: Two-Sided Tests of Spatial Dependence in SARAR Model

`Group Interaction, REG2,` $T=3, \beta=(1,1)', \sigma=1$

n	Test	10%	5%	1%	10%	5%	1%	10%	5%	1%
		Normal Errors			Normal Mixture			Lognormal Errors		
					$H_0: \lambda=\rho=0$					
50	$\mathcal{W}_{11}$	.1974	.1288	.0546	.1918	.1232	.0450	.1616	.1062	.0456
	$\mathcal{W}_{22}$	.1896	.1196	.0516	.1846	.1222	.0470	.1584	.1008	.0408
	$\mathcal{W}_{33}$	.1520	.0906	.0388	.1428	.0874	.0302	.1318	.0778	.0300
100	$\mathcal{W}_{11}$	.1732	.1048	.0348	.1652	.0964	.0384	.1416	.0860	.0286
	$\mathcal{W}_{22}$	.1754	.1116	.0366	.1684	.1070	.0388	.1416	.0858	.0284
	$\mathcal{W}_{33}$	.1290	.0764	.0224	.1228	.0734	.0266	.1192	.0676	.0208
250	$\mathcal{W}_{11}$	.1406	.0808	.0208	.1364	.0736	.0198	.1104	.0620	.0162
	$\mathcal{W}_{22}$	.1390	.0788	.0234	.1350	.0758	.0206	.1170	.0712	.0196
	$\mathcal{W}_{33}$	.1148	.0618	.0174	.1102	.0576	.0154	.1026	.0564	.0170
500	$\mathcal{W}_{11}$	.1334	.0740	.0176	.1168	.0682	.0142	.1128	.0630	.0136
	$\mathcal{W}_{22}$	.1358	.0752	.0178	.1270	.0674	.0176	.1338	.0730	.0196
	$\mathcal{W}_{33}$	.1088	.0548	.0128	.1000	.0528	.0118	.1096	.0552	.0118
					$H_0: \lambda=0$, (true $\rho=0$)					
50	$\mathcal{W}_{11}$	.1660	.1024	.0392	.1436	.0920	.0320	.1450	.0920	.0360
	$\mathcal{W}_{22}$	.1622	.1044	.0382	.1578	.0968	.0378	.1590	.0970	.0410
	$\mathcal{W}_{33}$	.1354	.0842	.0294	.1260	.0758	.0246	.1284	.0798	.0286
100	$\mathcal{W}_{11}$	.1362	.0798	.0256	.1352	.0812	.0268	.1302	.0734	.0230
	$\mathcal{W}_{22}$	.1532	.0908	.0282	.1494	.0906	.0294	.1332	.0758	.0230
	$\mathcal{W}_{33}$	.1174	.0668	.0212	.1162	.0686	.0202	.1186	.0670	.0178
250	$\mathcal{W}_{11}$	.1232	.0732	.0174	.1228	.0690	.0158	.1134	.0576	.0154
	$\mathcal{W}_{22}$	.1266	.0726	.0170	.1238	.0682	.0160	.1174	.0616	.0154
	$\mathcal{W}_{33}$	.1126	.0630	.0132	.1100	.0594	.0118	.1052	.0542	.0126
500	$\mathcal{W}_{11}$	.1108	.0578	.0142	.1094	.0556	.0116	.1116	.0616	.0138
	$\mathcal{W}_{22}$	.1198	.0588	.0148	.1120	.0576	.0128	.1198	.0662	.0160
	$\mathcal{W}_{33}$	.1050	.0530	.0122	.1030	.0524	.0098	.1070	.0572	.0130
					$H_0: \rho=0$ (true $\lambda=0$)					
50	$\mathcal{W}_{11}$	.1730	.1054	.0392	.1714	.1070	.0382	.1498	.0902	.0328
	$\mathcal{W}_{22}$	.1366	.0850	.0326	.1418	.0822	.0312	.1202	.0692	.0192
	$\mathcal{W}_{33}$	.1268	.0794	.0280	.1214	.0710	.0262	.1056	.0598	.0170
100	$\mathcal{W}_{11}$	.1604	.0980	.0268	.1478	.0856	.0250	.1292	.0710	.0198
	$\mathcal{W}_{22}$	.1302	.0758	.0252	.1274	.0732	.0260	.1142	.0672	.0220
	$\mathcal{W}_{33}$	.1124	.0630	.0198	.1056	.0612	.0196	.0952	.0568	.0164
250	$\mathcal{W}_{11}$	.1358	.0742	.0192	.1304	.0724	.0192	.1030	.0506	.0122
	$\mathcal{W}_{22}$	.1216	.0694	.0166	.1226	.0670	.0176	.1036	.0552	.0168
	$\mathcal{W}_{33}$	.1074	.0570	.0132	.1054	.0556	.0126	.0880	.0456	.0132
500	$\mathcal{W}_{11}$	.1306	.0704	.0158	.1126	.0600	.0140	.0976	.0514	.0124
	$\mathcal{W}_{22}$	.1208	.0682	.0170	.1110	.0590	.0150	.1154	.0616	.0146
	$\mathcal{W}_{33}$	.1030	.0528	.0114	.0928	.0466	.0106	.0966	.0478	.0116

Note: $\mathcal{W}_{jj}$ are defined in (3.12) for joint tests and (3.13) for one-directional tests.

Table 5b. Empirical Sizes: Two-Sided Tests of $H_0 : \lambda = 0$ in SLD Model
`Group Interaction, REG2,` $T = 3, \beta = (1,1)', \sigma = 1$. $\mathcal{T}_{jj}$ `are defined in` (3.14)

n	Test	10%	5%	1%	10%	5%	1%	10%	5%	1%
		Normal Errors			Normal Mixture			Lognormal Errors		
50	$\mathcal{T}_{11}$	.1422	.0850	.0232	.1254	.0676	.0190	.1068	.0552	.0140
	$\mathcal{T}_{22}$	.1348	.0808	.0212	.1154	.0586	.0162	.1042	.0586	.0134
	$\mathcal{T}_{33}$	.1120	.0616	.0146	.0992	.0472	.0126	.0918	.0484	.0102
100	$\mathcal{T}_{11}$	.1224	.0622	.0174	.1186	.0660	.0136	.1070	.0590	.0116
	$\mathcal{T}_{22}$	.1142	.0604	.0128	.1214	.0654	.0158	.1108	.0600	.0130
	$\mathcal{T}_{33}$	.1004	.0478	.0102	.1046	.0518	.0118	.0958	.0502	.0084
250	$\mathcal{T}_{11}$	.1148	.0584	.0176	.1042	.0540	.0112	.1006	.0512	.0142
	$\mathcal{T}_{22}$	.1130	.0622	.0172	.1128	.0604	.0128	.1140	.0572	.0150
	$\mathcal{T}_{33}$	.1006	.0526	.0130	.0946	.0506	.0086	.0996	.0466	.0124
500	$\mathcal{T}_{11}$	.1126	.0560	.0106	.1082	.0528	.0122	.0970	.0472	.0082
	$\mathcal{T}_{22}$	.1154	.0646	.0140	.1066	.0564	.0118	.1064	.0554	.0106
	$\mathcal{T}_{33}$	.1010	.0554	.0110	.0972	.0484	.0104	.0960	.0474	.0080

Table 5c. Empirical Sizes: Two-Sided Tests of $H_0 : \rho = 0$ in SED Model
`Group Interaction, REG2,` $T = 3, \beta = (1,1)', \sigma = 1$. $\mathcal{T}_{jj}$ `are defined in` (3.14)

n	Test	10%	5%	1%	10%	5%	1%	10%	5%	1%
		Normal Errors			Normal Mixture			Lognormal Errors		
50	$\mathcal{T}_{11}$	.1572	.0920	.0282	.1492	.0846	.0236	.1282	.0666	.0164
	$\mathcal{T}_{22}$	.1386	.0758	.0234	.1242	.0734	.0220	.1030	.0572	.0152
	$\mathcal{T}_{33}$	.1146	.0620	.0172	.1152	.0640	.0176	.0928	.0518	.0142
100	$\mathcal{T}_{11}$	.1420	.0798	.0224	.1324	.0738	.0142	.1170	.0598	.0126
	$\mathcal{T}_{22}$	.1274	.0736	.0202	.1248	.0700	.0160	.1010	.0550	.0140
	$\mathcal{T}_{33}$	.1116	.0594	.0154	.1054	.0540	.0112	.0840	.0444	.0116
250	$\mathcal{T}_{11}$	.1224	.0630	.0140	.1128	.0568	.0114	.1028	.0544	.0124
	$\mathcal{T}_{22}$	.1190	.0656	.0172	.1096	.0560	.0142	.1056	.0566	.0166
	$\mathcal{T}_{33}$	.1006	.0518	.0124	.0882	.0450	.0114	.0880	.0466	.0114
500	$\mathcal{T}_{11}$	.1124	.0578	.0120	.1126	.0526	.0098	.1004	.0518	.0116
	$\mathcal{T}_{22}$	.1136	.0624	.0142	.1202	.0604	.0148	.1164	.0610	.0178
	$\mathcal{T}_{33}$	.0952	.0492	.0098	.1004	.0482	.0108	.0982	.0476	.0126

Table 6. Empirical Sizes: Two-Sided Tests of $H_0 : \beta_1 = \beta_2$ in SARAR Model
`Group Interaction, REG2,` $T = 3, \sigma = 1, \lambda = \rho = 0$

n	Test	10%	5%	1%	10%	5%	1%	10%	5%	1%
		Normal Errors			Normal Mixture			Lognormal Errors		
50	$\mathcal{T}_{11}$	.1608	.1020	.0386	.1630	.1046	.0386	.1604	.0978	.0344
	$\mathcal{T}_{22}$	.1154	.0650	.0214	.1190	.0678	.0206	.1138	.0614	.0204
100	$\mathcal{T}_{11}$	.1334	.0744	.0228	.1344	.0794	.0218	.1334	.0782	.0218
	$\mathcal{T}_{22}$	.1012	.0546	.0138	.1042	.0536	.0126	.1032	.0534	.0120
250	$\mathcal{T}_{11}$	.1240	.0642	.0166	.1210	.0680	.0204	.1196	.0670	.0184
	$\mathcal{T}_{22}$	.1066	.0524	.0120	.1060	.0564	.0152	.1018	.0580	.0114
500	$\mathcal{T}_{11}$	.1092	.0548	.0116	.1100	.0564	.0140	.1154	.0616	.0200
	$\mathcal{T}_{22}$	.0958	.0472	.0092	.0978	.0472	.0100	.1022	.0536	.0146
50	$\mathcal{T}_{11}$	.1624	.1004	.0376	.1624	.1024	.0390	.1610	.0992	.0376
	$\mathcal{T}_{22}$	.1136	.0654	.0196	.1204	.0666	.0208	.1136	.0640	.0216
100	$\mathcal{T}_{11}$	.1282	.0742	.0196	.1394	.0810	.0208	.1420	.0808	.0250
	$\mathcal{T}_{22}$	.0968	.0496	.0114	.1068	.0540	.0090	.1060	.0564	.0118
250	$\mathcal{T}_{11}$	.1254	.0688	.0190	.1224	.0642	.0140	.1146	.0622	.0180
	$\mathcal{T}_{22}$	.1050	.0568	.0142	.1024	.0480	.0094	.0990	.0526	.0132
500	$\mathcal{T}_{11}$	.1240	.0626	.0152	.1130	.0594	.0130	.1220	.0650	.0160
	$\mathcal{T}_{22}$	.1102	.0502	.0124	.0978	.0482	.0096	.1084	.0552	.0122

Note: $\beta = (1,1)'$ for upper panel, and $(.5,.5)'$ for lower panel. $\mathcal{T}_{jj}$ are defined in (3.15).