Forthcoming in Regional Science and Urban Economics

Bias Correction and Refined Inferences 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.*

August 12, 2016

Abstract

This paper first presents simple methods for conducting up to third-order bias and variance corrections for the quasi maximum likelihood (QML) estimators of the spatial parameter(s) in the fixed effects spatial panel data (FE-SPD) models. Then, it shows how the bias and variance corrections lead to refined t-ratios for spatial effects and for covariate effects. The implementation of these corrections depends on the proposed bootstrap methods of which validity is established. Monte Carlo results reveal that (i) the QML estimators of the spatial parameters can be quite biased, (ii) a second-order bias correction effectively removes the bias, and (iii) the proposed t-ratios are much more reliable than the usual t-ratios.

Key Words: Bias correction, Variance correction, Refined t-ratios, Bootstrap, Wild bootstrap, Spatial panels, Fixed effects.

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

1 Introduction

Panel data models with spatial and social interactions have received a belated but recently increasing attention by econometricians, since Anselin (1988).[1] Spatial panel data (SPD) models are differentiated by whether they are static or dynamic and whether they contain random effects or fixed effects. The quasi maximum likelihood (QML) and the generalized method of moments (GMM) are the popular methods for estimation and inference of these models. See Lee and Yu (2010a, 2015) and Anselin et al. (2008) for general accounts on issues related to SPD model specifications, parameter estimation, etc.

It has been recognized through the studies of spatial regression models that QML estimators of the spatial parameter(s), although efficient, can be quite biased (Lee, 2004; Bao

*We thank the Editor Daniel McMillen, two referees, and the participants of the 26th (EC)^2 Conference–Theory and Practice of Spatial Econometrics for their helpful comments. Zhenlin Yang gratefully acknowledges the financial support from Singapore Management University under Grant C244/MSS14E002.

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

[1] See, among others, Baltagi et al. (2003, 2013), Kapoor et al. (2007), Yu et al. (2008, 2012), Yu and Lee (2010), Lee and Yu (2010a,b), Baltagi and Yang (2013a,b), and Su and Yang (2015).

and Ullah, 2007; Bao, 2013; Yang, 2015), and more so with a denser spatial weight matrix (Yang, 2015; Liu and Yang, 2015a). As a result the subsequent model inferences (based on t-ratios) can be seriously affected. Methods of bias-correcting the QML estimators of the spatial parameter(s) have been given for the spatial lag (SL) model (Bao and Ullah, 2007; Bao, 2013; Yang, 2015), the spatial error (SE) model (Liu and Yang, 2015a), and the spatial lag and error (SLE) model (Liu and Yang, 2015b). The improved t-ratios for the SL effect is given in Yang (2015), and improved t-ratios for the covariate effects are given in Liu and Yang (2015b) for the SL, SE and SLE models, respectively.

Evidently, the QML estimators of the SPD models are subjected to the same issues on the finite sample bias and finite sample performance of subsequent inferences, but these important issues have not been addressed.[2] Given the popularity of the SPD models among the applied researchers, it is highly desirable to have a set of simple and reliable methods for parameter estimation and model inference. In this paper, we focus on the SPD models with fixed effects to provide methods for bias and variance corrections (up to third-order) by extending the methods of Yang (2015),[3] and then to show how the bias and variance corrections lead to improved t-ratios for spatial and covariate effects. Lee and Yu (2010b) investigate the asymptotic properties for the QML estimation of this model based on *direct* and *transformation* approaches. The latter approach is more attractive as it provides consistent estimators for all the common parameters, which is crucial in the development of the methods for finite sample bias-corrections and refined inferences.[4]

We note that while the general stochastic expansions of Yang (2015) for nonlinear estimators are applicable to different models including the SPD models considered in this paper, the detailed developments of bias corrections, variance corrections and corrections on t-ratio vary from one model to another. Furthermore, the transformation approach induces errors that may no longer be independent and identically distributed (iid) even if the original errors are. Thus, the bootstrap method proposed by Yang (2015) under iid errors, may not be directly applicable. We demonstrate in this paper that when the original error distribution is not far from normality, the standard iid bootstrap method can still provide an excellent approximation, due to the fact that the transformed errors are

[2]The importance of bias correction for models with nonlinear parameters is seen from the large literature on the regular dynamic panels (see, e.g., Nickell (1981), Kiviet (1995), Hahn and Kuersteiner (2002), Hahn and Newey (2004), Bun and Carree (2005), Hahn and Moon (2006), and Arellano and Hahn (2005).

[3]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 for different random effects models considered in, e.g., Anselin (1988), Kapoor et al. (2007) and Baltagi et al. (2013). However, fixed effects model encounters incidental parameter problem (Neyman and Scott, 1948; Lancaster, 2000).

[4]Lee and Yu (2010b) 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. With data transformations to eliminate the fixed effects, the incidental parameter problem is avoided, and the ratio of n and T does not affect the asymptotic properties of estimates as the data are pooled. The QMLEs so derived are shown to be consistent, and, except for the variance estimate, are identical to those from the direct approach.

homoskedastic and uncorrelated. When the original errors are extremely non-normal, we show that the wild bootstrap method can improve the approximation. 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. Furthermore, Monte Carlo results show that inferences for spatial and covariate effects based on the regular t-ratios can be misleading, but those based on the proposed t-ratios are very reliable. We emphasize that while corrections on the bias and variance of a point estimator are important, it is more important to correct the t-ratios so that practical applications of the models are more reliable. The methods presented in this paper show a plausible way to do so. They are simple and yet quite general as the spatial regression models are embedded as special cases.

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 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 methods for estimating various quantities in the expansions, and presents theories for the validity of these methods. Section 5 presents Monte Carlo results. Section 6 discusses and concludes the paper.

2 The Model and Its QML Estimation

For the spatial panel data (SPD) model with fixed effects (FE), we 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,

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

for $t = 1, 2, \ldots, T$, 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 (2010b), 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 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.[5] To eliminate the time effects, let J_n and $F_{n,n-1}$ be similarly defined, and W_{1n} and W_{2n} be row normalized.[6] For any $n \times T$ matrix $[Z_{n1}, \cdots, Z_{nT}]$, define the $(n-1) \times (T-1)$ transformed matrix as

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

This leads to, for $t = 1, \dots, T-1$, Y_{nt}^*, U_{nt}^*, V_{nt}^*, and $X_{nt,j}^*$ for the jth regressor. As in Lee and Yu (2010b), let $X_{nt}^* = [X_{nt,1}^*, X_{nt,2}^*, \dots, X_{nt,k}^*]$, 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:

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

After the transformations, the effective sample size becomes $N = (n-1)(T-1)$. Stacking the vectors and matrices, i.e., letting $\mathbf{Y}_N = (Y_{n1}^{*\prime}, \dots, Y_{n,T-1}^{*\prime})'$, $\mathbf{U}_N = (U_{n1}^{*\prime}, \dots, U_{n,T-1}^{*\prime})'$, $\mathbf{V}_N = (V_{n1}^{*\prime}, \dots, V_{n,T-1}^{*\prime})'$, $\mathbf{X}_N = (X_{n1}^{*\prime}, \dots, X_{n,T-1}^{*\prime})'$, 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 independent although they are uncorrelated and homoskedastic. 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 $(V_{n1}^{*\prime}, \dots, V_{n,T-1}^{*\prime})' = (F_{T,T-1}' \otimes F_{n,n-1}')(V_{n1}', \dots, V_{nT}')'$. Then,

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

[5] As discussed in Lee and Yu (2010b, Footnote 12), the first difference and Helmert transformation have often been used to eliminate the individual effects. A special selection of $F_{T,T-1}$ gives rise to the Helmert transformation where $\{V_{nt}\}$ are 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.

[6] When W_{jn} are not row normalized, the linear SARAR representation of (2.4) for the spatial panel model will no longer hold. In that case, a likelihood formulation would not be feasible.

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,

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

where $\zeta = (\beta', \lambda, \rho)'$, $\theta = (\beta', \sigma^2, \lambda, \rho)'$, $\mathbf{A}_N(\lambda) = I_N - \lambda\mathbf{W}_{1N}$, $\mathbf{B}_N(\rho) = I_N - \rho\mathbf{W}_{2N}$, 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:

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

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

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 $\tilde\sigma_N^2(\lambda,\rho)$ back into (2.5) gives the concentrated log likelihood function of (λ,ρ):

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

Maximizing $\ell_N^c(\lambda,\rho)$ 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 $\hat\sigma_N^2 \equiv \tilde\sigma_N^2(\hat\lambda_N, \hat\rho_N)$.[7] Write $\hat\theta_N = (\hat\beta_N', \hat\lambda_N, \hat\rho_N, \hat\sigma_N^2)'$. Lee and Yu (2010b) show that $\hat\theta_N$ is $\sqrt{N}$-consistent and asymptotically normal under some mild conditions. These conditions and the asymptotic variance of $\hat\theta_N$ are given in Appendix A to facilitate the subsequent developments for the higher-order results. It follows that the QML estimators for any of the submodels discussed below will be $\sqrt{N}$-consistent and asymptotically normal, where N can be $(n-1)(T-1)$, $n(T-1)$, $(n-1)T$, or nT.

The **linear SARAR representation** (2.4) has greatly facilitated the QML estimation of the general FE-SPD model. It is also very for the subsequent developments in bias and variance corrections. Obviously, it contains the spatial regression models as special cases. 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 SL dependence or an FE-SPD model with only SE dependence; 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 individual-specific effects or a submodel with only time-specific effects; and finally, dropping both $\mathbf{c}_{n0}$ and α_{t0} in (2.1) leads to a panel data model with SLE dependence. 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

[7] Numerical maximization of $\ell_N^c(\lambda,\rho)$ can be computationally demanding if N is large due to the need for repeated calculation of the two determinants. Following simplifications help alleviate this: $|\mathbf{A}_N(\lambda)| = |I_{n-1} - \lambda W_{1n}^*|^{T-1} = \left(\frac{1}{1-\lambda}|I_n - \lambda W_{1n}|\right)^{T-1} = \left(\frac{1}{1-\lambda}\prod_{i=1}^n(1-\lambda\omega_{1i})\right)^{T-1}$, 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.

disturbance, in particular the former.[8] 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 makes these expansions practically implementable. The method is demonstrated, through a linear SL 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 equation. By doing so, the dimensionality of the problem can be reduced, and more importantly, additional variations from the estimation of linear and scale parameters are captured in the stochastic expansions, thus making the bias and variance corrections more effective. The method is summarized as follows.

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 $H^{\circ}_{rN} = H_{rN} - \mathrm{E}(H_{rN}), r = 1, 2, 3$. Note that here and hereafter the expectation operator 'E' corresponds to the true model parameters θ_0. Define $\Omega_N = -[\mathrm{E}(H_{1N})]^{-1}$. 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 on 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 a rth order with respect to δ in a neighborhood of δ_0, $\mathrm{E}(H_{rN}) = O(1)$, and $H^{\circ}_{rN} = O_p(N^{-1/2})$, $r = 1, 2, 3$.*

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

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

[8] See Lee and Yu (2015, 2016) for more discussions and for the related issue on parameter identification.

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

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

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 $\mathrm{E}(\tilde{\psi}_N) = O(N^{-1})$ 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

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

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

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

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

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

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

Therefore, we can improve estimation and 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 $O(N^{-3/2})$ respectively. With estimated b_{-1} and $b_{-3/2}$, feasible δ_N^{bc2} and δ_N^{bc3} are constructed.

Similar procedures can be applied to increase the precision of the variance estimate. By (3.5), if $\hat{b}_{-1} - b_{-1} = O_p(N^{-3/2})$ 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 implementation of bias and variance corrections.

3.2 Third-order bias and variance for spatial estimators

As pointed out in the introduction, the general expansions summarized in Section 3.1 are applicable to different models including the FE-SPD model we consider in this paper, but the detailed developments for the corrections on bias, variance, and t-ratio vary from one model to another. Furthermore, the transformation approach induces errors to be no longer iid, rendering the bootstrap method of Yang (2015) for estimating the correction terms not directly applicable. In this subsection, we first derive all the quantities required for the third-order expansion of the FE-SPD model, and then discuss conditions under which the results (3.2)-(3.6) hold for 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 $\tilde{\psi}_N(\delta) = \frac{1}{N}\frac{\partial}{\partial\delta}\ell^c_N(\delta)$, which has the form:

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

where $T_{0N}(\lambda) = \frac{1}{N}\text{tr}(\mathbf{W}_{1N}\mathbf{A}^{-1}_N(\lambda))$, $K_{0N}(\rho) = \frac{1}{N}\text{tr}(\mathbf{W}_{2N}\mathbf{B}^{-1}_N(\rho))$, and $\mathbf{M}^{(1)}_N(\rho) = \frac{d}{d\rho}\mathbf{M}_N(\rho)$.[9]

To derive the rth order derivative, $H_{rN}(\delta)$, of $\tilde{\psi}_N(\delta)$ w.r.t. δ', $r = 1, 2, 3$, given in (3.2), define $T_{rN}(\lambda) = \frac{1}{N}\text{tr}[(\mathbf{W}_{1N}\mathbf{A}^{-1}_N(\lambda))^{r+1}]$, and $K_{rN}(\rho) = \frac{1}{N}\text{tr}[(\mathbf{W}_{2N}\mathbf{B}^{-1}_N(\rho))^{r+1}]$, $r = 0, 1, 2, 3$. Let $\mathbf{M}^{(k)}_N(\rho)$ be the kth derivative of $\mathbf{M}_N(\rho)$ w.r.t. ρ, $k = 1, 2, 3, 4$. Define

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

$$Q^\dagger_{kN}(\delta) = \frac{\mathbf{Y}'_N(\lambda)\mathbf{M}^{(k)}_N(\rho)\mathbf{W}_{1N}\mathbf{Y}_N}{\mathbf{Y}'_N(\lambda)\mathbf{M}_N(\rho)\mathbf{Y}_N(\lambda)}, \quad Q^\ddagger_{kN}(\delta) = \frac{\mathbf{Y}'_N\mathbf{W}'_{1N}\mathbf{M}^{(k)}_N(\rho)\mathbf{W}_{1N}\mathbf{Y}_N}{\mathbf{Y}'_N(\lambda)\mathbf{M}_N(\rho)\mathbf{Y}_N(\lambda)};$$

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

which have the following properties,

$$\begin{aligned} &\tfrac{\partial R_{1N}(\delta)}{\partial\lambda} = 2R^2_{1N}(\delta) - R_{2N}(\delta), && \tfrac{\partial R_{2N}(\delta)}{\partial\lambda} = 2R_{1N}(\delta)R_{2N}(\delta), \\ &\tfrac{\partial Q^\dagger_{kN}(\delta)}{\partial\lambda} = 2R_{1N}(\delta)Q^\dagger_{kN}(\delta) - Q^\ddagger_{kN}(\delta), && \tfrac{\partial Q^\ddagger_{kN}(\delta)}{\partial\lambda} = 2R_{1N}(\delta)Q^\ddagger_{kN}(\delta), \\ &\tfrac{\partial S_{kN}(\delta)}{\partial\lambda} = 2R_{1N}(\delta)S_{kN}(\delta) - 2Q^\dagger_{kN}(\delta); \end{aligned}$$

$$\begin{aligned} &\tfrac{\partial R_{1N}(\delta)}{\partial\rho} = Q^\dagger_{1N}(\delta) - R_{1N}(\delta)S_{1N}(\delta), && \tfrac{\partial R_{2N}(\delta)}{\partial\rho} = Q^\ddagger_{1N}(\delta) - R_{2N}(\delta)S_{1N}(\delta), \\ &\tfrac{\partial Q^\dagger_{kN}(\delta)}{\partial\rho} = Q^\dagger_{k+1,N}(\delta) - Q^\dagger_{kN}(\delta)S_{1N}(\delta), && \tfrac{\partial Q^\ddagger_{kN}(\delta)}{\partial\rho} = Q^\ddagger_{k+1,N}(\delta) - Q^\ddagger_{kN}(\delta)S_{1N}(\delta), \\ &\tfrac{\partial S_{kN}(\delta)}{\partial\rho} = S_{k+1,N}(\delta) - S_{kN}(\delta)S_{1N}(\delta). \end{aligned}$$

[9]Lee and Yu (2010b) provide a useful identity: $(I_{n-1} - \lambda W^*_{hn})^{-1} = F'_{n,n-1}(I_{n-1} - \lambda W_{hn})^{-1}F_{n,n-1}$. Based on this, the inverses of $\mathbf{A}_N(\lambda)$ and $\mathbf{B}_N(\lambda)$ can easily be calculated as they are block-diagonal. The conditions for the $\sqrt{N}$-consistency of $\hat{\delta}_N$ are given in Lee and Yu (2010b), and also in Appendix A.

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., $\tilde{\psi}^{\lambda\lambda}_{1N}(\delta) = \frac{\partial^2}{\partial\lambda^2}\tilde{\psi}_{1N}(\delta)$, and $\tilde{\psi}^{\lambda\rho\lambda}_{2N}(\delta) = \frac{\partial^3}{\partial\lambda\partial\rho\partial\lambda}\tilde{\psi}_{2N}(\delta)$. Thus, $H_{1N}(\delta)$ has 1st row $\{\tilde{\psi}^{\lambda}_{1N}(\delta), \tilde{\psi}^{\rho}_{1N}(\delta)\}$ and 2nd row $\{\tilde{\psi}^{\lambda}_{2N}(\delta), \tilde{\psi}^{\rho}_{2N}(\delta)\}$, which gives

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

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

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

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

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

The expressions of $\mathbf{M}^{(k)}_N(\rho)$, 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 given 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 (2010b), implies that $\mathrm{E}(\tilde{\psi}_N) = o(N^{-1/2})$. The other conditions are likely to hold for 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.[10] 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:

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

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 SL dependence model considered in detail in Yang (2015), showing the usefulness of the linear SARAR representation (2.4) for the FE-SPD model.

[10] Both can be extended by adding a spatial Durbin term $\mathbf{W}_N\mathbf{X}_N$ to give the so-called spatial Durbin model and the spatial Durbin error model; see, e.g., Elhorst (2014). See also Sec. 6 for more discussions.

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:

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

These results contain, as a special case, the results for the linear SED model considered in Liu and Yang (2015a). Again, these results show the usefulness of the linear SASAR representation for the FE-SPD 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, $t = 1,\dots,T-1$ and $N = n(T-1)$, and when model contains only time-specific effects, $t = 1,\dots,T$ and $N = (n-1)T$.

3.4 Bias correction for non-spatial estimators

As $\tilde{\beta}_N(\delta_0)$ in (2.6) is an unbiased estimator of β, and $\frac{N}{N-k}\tilde{\sigma}_N^2(\delta_0)$ in (2.7) is an unbiased estimator of σ^2, it is natural to expect that, with a bias-corrected QMLE $\hat{\delta}_N^{\text{bc}}$ of δ, $\hat{\beta}_N^{\text{bc}} = \tilde{\beta}_N(\hat{\delta}_N^{\text{bc}})$ and $\hat{\sigma}_N^{2,\text{bc}} = \frac{N}{N-k}\tilde{\sigma}_N^2(\hat{\delta}_N^{\text{bc}})$ 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, $\hat{\beta}_N \equiv \tilde{\beta}_N(\hat{\delta}_N) = \mathbf{F}_N(\hat{\rho}_N)\mathbf{Y}_N(\hat{\lambda}_N)$, where $\mathbf{F}_N(\rho) = [\mathbf{X}_N'(\rho)\mathbf{X}_N(\rho)]^{-1}\mathbf{X}_N'(\rho)\mathbf{B}_N(\rho)$, by (2.6). Let $\tilde{\beta}_N^{(k)}(\delta)$ be the kth derivative of $\tilde{\beta}_N(\delta)$ w.r.t. δ', and $\mathbf{F}_N^{(k)}(\rho)$ the kth derivative of $\mathbf{F}_N(\rho)$ w.r.t. ρ. A notational convention is followed: $\tilde{\beta}_N \equiv \tilde{\beta}_N(\delta_0)$, $\tilde{\beta}_N^{(k)} \equiv \tilde{\beta}_N^{(k)}(\delta_0)$, $\mathbf{F}_N \equiv \mathbf{F}_N(\rho_0)$, $\mathbf{A}_N = \mathbf{A}_N(\lambda_0)$, $\mathbf{B}_N = \mathbf{B}_N(\rho_0)$, etc. Assume $\mathrm{E}(\tilde{\beta}_N^{(k)})$ exists and $\tilde{\beta}_N^{(k)} - \mathrm{E}(\tilde{\beta}_N^{(k)}) = O_p(N^{-1/2}), k = 1,2$. By a Taylor series expansion, we obtain,

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

where $\mathrm{E}(\tilde{\beta}_N^{(1)}) = [-\mathbf{F}_N\mathbf{G}_N\mathbf{X}_N\beta_0,\ \mathbf{F}_N^{(1)}\mathbf{X}_N\beta_0]$, $\mathbf{G}_N = \mathbf{W}_{1N}\mathbf{A}_N^{-1}$, $b_N = [-\mathbf{F}_N\mathbf{G}_N\mathbf{B}_N^{-1}\mathbf{V}_N,\ \mathbf{F}_N^{(1)}\mathbf{B}_N^{-1}\mathbf{V}_N]$, and $\mathrm{E}(\tilde{\beta}_N^{(2)}) = [0_{k\times 1},\ -\mathbf{F}_N^{(1)}\mathbf{G}_N\mathbf{X}_N\beta_0,\ -\mathbf{F}_N^{(1)}\mathbf{G}_N\mathbf{X}_N\beta_0,\ \mathbf{F}_N^{(2)}\mathbf{X}_N\beta_0]$. 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 $\hat{\delta}_N^{\rm bc2}$. Thus,

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

The key term $\text{E}(\tilde{\beta}_N^{(1)})\text{Bias}(\hat{\delta}_N)$ 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 $\hat{\delta}_N^{\rm bc2}$ in defining the estimator for β_0. Thus, it can be expected that the resulting bias reduction can be big, and the estimator $\hat{\beta}_N^{\rm bc2} = \tilde{\beta}_N(\hat{\delta}_N^{\rm bc2})$ is essentially second-order bias-corrected, if $\text{E}(b_N a_{-1/2}) + \frac{1}{2}\text{E}(\tilde{\beta}_N^{(2)})\text{E}(a_{-1/2} \otimes a_{-1/2})$ is 'small'. In general, using (3.9), $\hat{\beta}_N^{\rm bc2}$ 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, $\mathbf{F}_N^{(k)}(\rho), k = 1, 2$, can be easily derived.

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

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

where the exact expressions for q_N and $\text{E}(Q_N^{(k)}), k = 1, 2$, are given in Appendix B.

It is easy to see that (3.10) holds when $\hat{\delta}_N$ is replaced by $\hat{\delta}_N^{\rm bc2}$. It follows that

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

Again, the key bias term $\frac{1}{N-k}\text{E}(Q_N^{(1)})\text{Bias}(\hat{\delta}_N)$ is removed when $\hat{\delta}_N$ is replaced by $\hat{\delta}_N^{\rm bc2}$ in defining the estimator for σ_0^2, and our Monte Carlo results in Section 5 show that $\frac{N}{N-k}\tilde{\sigma}_N^2(\hat{\delta}_N^{\rm bc2})$ is nearly unbiased for σ_0^2. In any case, one can always use (3.11) to carry out further bias correction on $\frac{N}{N-k}\tilde{\sigma}_N^2(\hat{\delta}_N^{\rm bc2})$.

3.5 Inferences following bias and variance corrections

The impact 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. The latter issue is of a great practical relevance, as being able to assess the covariate effects in a reliable manner may be the most desirable feature in any econometric modelling activity. Unfortunately, this issue has not been addressed for spatial panel data regression models.

One of the most interesting type of inferences for a spatial model would be 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 $j = k = 1$, $\hat{\delta}_N^{\mathrm{bc1}} = \hat{\delta}_N$, $\mathrm{Var}_1(\hat{\delta}_N^{\mathrm{bc1}}) = \Omega_N \mathrm{E}(\tilde{\psi}_N \tilde{\psi}_N')\Omega_N$, and the test is an asymptotic Wald test. The details on estimating $\mathrm{Var}_k(\hat{\delta}_N^{\mathrm{bc}j})$, in particular, $\mathrm{Var}_3(\hat{\delta}_N^{\mathrm{bc3}})$, are given at the end of Section 4.

Similarly, for testing the non-existence of one type of spatial effect, allowing the existence of the other type of spatial effect, 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 concerns the covariate effects, i.e., 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 $\hat{\beta}_N^{\mathrm{bc2}}$. By (3.8) with $\hat{\delta}_N$ being replaced by $\hat{\delta}_N^{\mathrm{bc2}}$, we have,

$$\mathrm{Var}(\hat{\beta}_N^{\mathrm{bc2}}) = \mathrm{Var}\big[\tilde{\beta}_N + \mathrm{E}(\tilde{\beta}_N^{(1)})(a_{-1/2} + a_{-1}) + b_N a_{-1/2} + \tfrac{1}{2}\mathrm{E}(\tilde{\beta}_N^{(2)})(a_{-1/2} \otimes a_{-1/2})\big] + O_p(N^{-2}).$$

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:

$$\mathcal{T}_{N,11} = c'\hat{\beta}_N / \sqrt{c'\widehat{\mathrm{AVar}}(\hat{\beta}_N)c}, \quad \text{and} \quad \mathcal{T}_{N,22} = c'\hat{\beta}_N^{\mathrm{bc2}} / \sqrt{c'\widehat{\mathrm{Var}}(\hat{\beta}_N^{\mathrm{bc2}})c}, \tag{3.15}$$

where $\widehat{\mathrm{AVar}}(\hat{\beta}_N)$ is the estimate of the asymptotic variance of $\hat{\beta}_N$ and $\widehat{\mathrm{Var}}(\hat{\beta}_N^{\mathrm{bc2}})$ is the bootstrap estimate of $\mathrm{Var}(\hat{\beta}_N^{\mathrm{bc2}})$ (see the end of Section 4). These results can easily be simplified to suit the simpler models. Monte Carlo results presented in Section 5 show that inferences based on $\mathcal{T}_{N,22}$ are much more reliable than inferences based on $\mathcal{T}_{N,11}$.

4 Bootstrap for Feasible Bias and Variance Corrections

For practical implementation of the methods given in the previous section, 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 the 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 the pure SL dependent model without regressors (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 of 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 SLE dependence given in (2.4). 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 $\mathbf{M}_N^{\diamond} = I_N - \mathbf{X}_N(\rho_0)[\mathbf{X}_N'(\rho_0)\mathbf{X}_N(\rho_0)]^{-1}\mathbf{X}_N'(\rho_0)$ given at the end of Section 3.2, 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, for example,

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

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.[11] To make these bootstrap expectations practically feasible, we first follow Yang (2015) and propose the following *iid bootstrap* procedure:

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

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

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.

[11] 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, $H_{1N}\Omega_N\tilde{\psi}_N = (\tilde{\psi}_N' \otimes H_{1N})\mathrm{vec}(\Omega_N)$, 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 4.2 *(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* $\hat{\mathbf{V}}_N$*.*[12]
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*

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

Note that the common application of the wild bootstrap method is 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 unable 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. 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}. However, there are important special cases under which the classical iid bootstrap method is strictly valid.

First, we note that when 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 give the following proposition.

[12] 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.

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 rth 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 only individual effects 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 irrespective of 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 rth 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 depends) 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 strict 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 rth 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 rth 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 that the higher-order cumulants or joint cumulants of $\{\mathbf{v}_i\}$ are proportional to the higher-order cumulants k_r of the original errors $\{v_{it}\}$. This suggests that when $k_r = 0, r = 3, \ldots, 10$, $\{\mathbf{v}_i\}$ are essentially iid and hence the conclusion of Proposition 4.1 holds in light of the results of Yang (2015) for the iid bootstrap. Similarly, the conclusion of Proposition 4.2 also holds.

When $k_r \neq 0$ for some or all $r = 3, \ldots, 10$, $\{\mathbf{v}'_i\}$ are no longer 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 $r \geq 3$. The latter may cause more problem as it is known that the iid bootstrap is unable to capture dependence. However, noting that the proportionality constants $a_{i_1,\ldots,i_r}$ are all pure numbers, being the sum of elementwise products of the orthonormal vectors $\{\mathbf{f}_i\}$, intuitively they should be small, and the larger the r, the smaller the $a_{i_1,\ldots,i_r}$.[13] These suggest that the higher-order dependence among $\{\mathbf{v}_i\}$ can largely be ignored. The question left is how well the two sets of cumulants match.

[13]We are unable to further characterize these quantities. However, as they are pure numbers depending on n and T through $F_{T,T-1}$ and $F_{n,n-1}$, it should be indicative to present some of their values. With the eigenvector-based transformations defined above (2.2) and calculated using Matlab `eig` function, we have, for $n = 100$ and $T = 3$, $a_{1,2,3} = -5.6e^{-5}$, $a_{1,2,3,4} = 3.4e^{-5}$, and $a_{1,2,3,4,5} = -3.7e^{-7}$; and for $n = 200$, the same set of numbers become $2.3e^{-5}$, $-3.8e^{-6}$ and $1.3e^{-8}$. With Helmert transformations (see Footnote 5), these numbers become much smaller ($< 1.0e^{-19}$).

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 $\mathbf{v}_i$ ($\bar{a}_r$ versus $a_{r,i}$), but is unable to capture the higher-order dependence. However, as argued below 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,[14] 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 draw 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 of the original errors so that the transformed errors are far from being iid, it may be more important to be able to match the even moments, in particular the kurtosis, than the odd moments, as $a_{r,i}$ is typically small on average with moderate variability when r is odd, see Footnote 13. 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 errors. In this spirit, the simple two-point distribution with equal probability described in Footnote 12 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* $\varepsilon^\star$.

[14] Again, we are unable to further characterize these pure constants. To have some concrete idea, we have calculated the mean and standard deviation of $\{a_{r,i}\}$ for $n = 100, T = 3$ and $r = 3,4,5,6$: $(-.0020, .0827)$, $(.1245, .0679)$, $(-.0010, .0425)$, $(.0308, .0299)$. When $n = 500$, the same set of values becomes: $(.0008, .0751)$, $(.1141, .0714)$, $(.0010, .0360)$, $(.0263, .0281)$. With Helmert transformations, these numbers become slightly bigger.

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 on 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} w.r.t. 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 given in Lemma 4.2.

Finally, to estimate $\widehat{\text{Var}}(\hat{\beta}_N^{\text{bc2}})$ in (3.15): we need to (i) calculate the estimates of all the non-stochastic quantities using 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

Extensive Monte Carlo experiments are conducted to investigate (i) the finite sample performance of the QMLE $\hat{\delta}_N$ and the bias-corrected QMLEs $\hat{\delta}_N^{\rm bc2}$ and $\hat{\delta}_N^{\rm bc3}$, (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):

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

For all the Monte Carlo experiments, $\beta_0 = (\beta_{10}, \beta_{20})'$ is set to $(1,1)'$ or $(.5,.5)'$, $\sigma_0^2 = 1$, λ_0 and ρ_0 take values from $\{-0.5, -0.25, 0, 0.25, 0.5\}$, $n = \{25, 50, 100, 200, 500\}$, 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 $F_{T,T-1}$ and $F_{n,n-1}$ defined above (2.2) are used and calculated using Matlab `eig` function. The weights 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 interactions (number of neighbors each unit has) specified by layouts (i)-(iii) are all fixed while in (iv) it may grow with the sample size. This is attained by relating the number of groups k to the sample size n, e.g., $k = n^{0.5}$ (Lee, 2004). In this case, the degree of spatial interactions is reflected by the group sizes generated from $U(0.5m, 1.5m)$, where $m = n/k$ is roughly the average group size. 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. Partial Monte Carlo results are reported in Tables 1a, 1b, 2, 3a and 3b, where Tables 1a-1b correspond to the model with $\rho = 0$, i.e., the spatial lag (SL) dependence model; Table 2 the model with $\lambda = 0$, i.e., the spatial error (SE) dependence model; and Tables 3a-3b the general model with both SL and SE (SLE) dependence. 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 very similar results as the iid bootstrap, consistent with remarks below Lemma 4.2.[15]

In general, The Monte Carlo results reveal that the regular QMLEs of the spatial parameters can be very biased, depending on the sample size, the degree of spatial interactions, the parameter values, and the way the regressors are generated, etc. A larger sample gives a smaller bias; a denser spatial weight matrix results in a larger bias; a smaller $|\beta|$ or a larger σ may lead to a larger bias; and non-iid regressors lead to a larger bias compared with the case of iid regressors. In stark contrast, the bias-corrected estimators are nearly unbiased in all situations, and 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. Finally, the error distribution does not much affect the finite sample performance of the estimators. Some details are as follows.

First, for each model, comparing the results for the QMLEs (reported and unreported) under `Queen Contiguity` and `Group Interaction`, we see that a denser spatial weight matrix leads to significantly larger biases, and non-iid regressors (`REG2`) lead to noticeably larger biases compared with the iid regressors (`REG1`). With the two combined, the biases of the QMLEs can be noticeable even when $n = 500$ (see Table 1b, case (b); Table 2: case (b); Table 3b). Second, comparing the results in Table 1a, where $\beta = (1\ 1)'$, with the corresponding results in Table 1b, where $\beta = (.5\ .5)'$ and thus the *signal-to-noise ratio* (SNR), defined as $|\mathrm{E}(Y|X)|/\mathrm{sd}(Y|X)$, is halved, we see that reducing the SNR significantly increases the bias of the QMLE $\hat{\lambda}_N$ of the spatial lag parameter λ in the SL model, in particular in the case of group interaction spatial layout. Monte Carlo results (unreported for brevity) reveal a similar phenomenon for $\hat{\lambda}_N$ in the SLE model. However, changing SNR value by changing β and σ does affect the QMLE $\hat{\rho}_N$ of the spatial error parameter ρ in the SE model, and has little effects on $\hat{\rho}_N$ in the SLE model. This is because ρ is a pure 'noise parameter' and estimation of it is invariant of β and σ, at least in the case of SE model (see Liu and Yang, 2015a). In contrast, λ affects both the signal and the noise parts of the model, and hence $\hat{\lambda}_N$ is sensitive to changes in either β or σ (see Yang, 2015).

The above discussions about the biases of the QMLEs of the spatial parameters may also extend to the variabilities of the QMLEs: (*i*) a denser spatial weight matrix leads to higher variabilities of the QMLEs $\hat{\lambda}_N$ and $\hat{\rho}_N$ in general, (*ii*) reducing SNR, by reducing $|\beta|$ or increasing σ, increases the variability of $\hat{\lambda}_N$ but not that of $\hat{\rho}_N$, (*iii*) the variability of $\hat{\rho}_N$ increases as ρ decreases in general; the variability of $\hat{\lambda}_N$ increases as λ decreases for the SL model in general, but not for the SLE model.

It is interesting to note that in the general SLE model, the finite sample performance of $\hat{\rho}_N$ can be much worse than that of $\hat{\lambda}_N$ in terms of both bias and variability. Furthermore, under the same model setting, thus the same SNR, the $\hat{\rho}_N$ of $\rho(=0)$ for the SE model is more biased and more variable than $\hat{\lambda}_N$ of $\lambda(=0)$ for the SL model. It seems that the

[15] All the results, unreported but necessary for supporting the subsequent discussions are collected in a *Supplementary Material* available from `http://www.mysmu.edu/faculty/zlyang/`.

values of λ and ρ affect more on the variabilities of $\hat{\lambda}_N$ and $\hat{\rho}_N$ but less on their biases. The exact causes of these behaviors are unclear and deserves some future research.

In summary, the results show that in general the QMLEs of the spatial panel data models may need to be bias-corrected even when the sample size is not small, and that the proposed bias correction method is very effective in removing the bias. As far as bias correction is concerned, a simple iid bootstrap may well serve the purpose. The method can easily be applied and thus is recommended to 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 $\hat{\beta}^{\rm bc}_N$ and $\hat{\sigma}^{2,\rm bc}_N$ 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, $\hat{\beta}^{\rm bc}_N$ is nearly unbiased in general. When the error distribution is skewed, $\hat{\sigma}^{2,\rm bc}_N$ 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. More can be found in the *Supplementary Material* (see Footnote 15).

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 can be found in the *Supplementary Material* (see Footnote 15).

Table 5a presents the empirical sizes of, respectively, the joint tests for the lack of both SL and SE effects given in (3.12), and the one-directional tests for the lack of SL effect allowing the presence of SE effect or the lack of SE effect allowing the presence of SL 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. The results based on iid bootstrap are collected in the *Supplementary Material* (see Footnote 15). Consistent with the results of Section 4.2, in case of severe nonnormality such as the lognormal errors ($n = 500$), the wild bootstrap performs better than the iid bootstrap; in case of normal errors, the iid bootstrap performs 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 the 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. With a weaker spatial dependence (results unreported), 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 finite sample bias and variance corrections of the 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. We have demonstrated that bias and variance corrections lead to refined inferences for the spatial effects as well as covariate effects. The proposed methods are 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 covariate effects), 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. The latter observation is perhaps the most important one in this study, as being able to assess the covariate effects in a reliable manner may be the most desirable feature of econometric modelling activities. All the methods proposed in the current paper can easily be built into the standard statistical software to facilitate practical applications (an empirical illustration is given in the *Supplementary Material* and the corresponding matlab codes are available therein). The results presented in this paper reinforce that the general methodology of bias and variance corrections of Yang (2015), based on stochastic expansion and bootstrap, is indeed a promising approach in handling the bias issues, and in providing refined inference methods.

Further extensions of the proposed methods are desirable, and are possible at least in several directions: (*i*) adding spatial Durbin effects, (*ii*) studying the direct and indirect effects, and (*iii*) adding higher-order spatial effects. As indicated in Footnote 10, the spatial Durbin model and spatial Durbin error model are simple extensions of the SL and SE models by replacing $\mathbf{X}_N$ by $[\mathbf{X}_N \;\; \mathbf{W}_N\mathbf{X}_N]$, and hence can be studied using the same techniques developed for the SL and SE models. What is interesting here may be that the tests given in (3.15) can be adapted to give improved tests for the existence of spatial Durbin effects, and that the ideas behind the developments of the improved inferences for the regression coefficients, i.e., (3.8), (3.9), and (3.15), may be extended to give improved inferences for the direct and indirect effects, as the measures of direct and indirect effects