Accepted version for: Econometrics 3, 376-411, 2015.

Asymptotic Distribution and Finite-Sample Bias Correction of QML Estimators for Spatial Error Dependence Model*

Shew Fan Liu and Zhenlin Yang†

School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903.

emails: shewfan.liu.2011@smu.edu.sg; zlyang@smu.edu.sg

May 12, 2015

Abstract

In studying the asymptotic and finite-sample properties of quasi-maximum likelihood (QML) estimators for the spatial linear regression models, much attention has been paid to the spatial lag dependence (SLD) model; little has been given to its companion, the spatial error dependence (SED) model. In particular, the effect of spatial dependence on the convergence rate of the QML estimators has not been formally studied, and methods for correcting finite-sample bias of the QML estimators have not been given. This paper fills in these gaps. Of the two, bias correction is particularly important to the applications of this model as it leads potentially to much improved inferences for the regression coefficients. Contrary to the common perceptions, both the large and small sample behaviours of the QML estimators for the SED model can be different from those for the SLD model in terms of the rate of convergence and the magnitude of bias. Monte Carlo results show that the bias can be severe and the proposed bias correction procedure is very effective.

Key Words: Asymptotics; Bias Correction; Bootstrap; Concentrated estimating equation; Monte Carlo; Spatial layout; Stochastic expansion.

JEL Classification: C10, C15, C21

1. Introduction

With the fast globalisation of economic activities and the concept of 'neighbour' ceasing to be merely the person next door, economists and econometricians alike have recognised the importance of modelling the spatial interaction of economic variables. As in time series where the concern is to alleviate the estimation problems caused by the lag in time, the analogous case in cross sectional data gives rise to a lag in space.

The conventional way to incorporate spatial autocorrelation in a regression model is to add a spatial lag of the dependent variable or a spatial lag of the error variable into the model, giving rise to a regression model with spatial lag dependence (SLD), or a regression model with spatial error dependence (SED). See, among the others, Cliff and Ord (1972, 1973), Ord (1975), Burridge (1980), Cliff and Ord (1981), Anselin (1980, 1988), Anselin and Bera (1998), Anselin (2001). These two models have over the years become the building blocks for spatial econometric

*The authors wish to thank the participants of the Asian Meeting of the Econometric Society 2013 and the Singapore Economic Review Conference 2013, and three anonymous referees for their useful comments and suggestions. Zhenlin Yang gratefully acknowledges the research support from Singapore Management University.

†Corresponding author: 90 Stamford Road, Singapore 178903. Phone: +65-6828-0852; Fax: +65-6828-0833.

modelling, and many more general spatial econometric models have been developed based on them. See, e.g., Anselin (2003), Das et al. (2003), Kelejian and Prucha (1998), and Lee and Liu (2010) for more general spatial regression models; Pinkse (1998) and Fleming (2004) for spatial discrete choices models; and Lee and Yu (2010) for a survey on spatial panel data models.

Of the methods available for spatial model estimation, the maximum likelihood (ML) or quasi-ML (QML) method remains attractive due to its efficiency. As a result of the fast increase in computing power allowing for easier manipulation of large matrices, the initial reluctance for the use of QML estimation as opposed to other easily implementable estimation methods alleviated.[1] As such there had been a growing interest in developing the theoretical aspects behind QML estimation in recent times which mainly identifies two intriguing issues related the QML estimation of spatial models: asymptotic distribution and finite-sample bias of the ML or QML estimators (MLEs or QMLEs). Of the two models, the SLD model has been extensively studied in terms of the asymptotic distributions of the MLEs or QMLEs (Lee, 2004); finite-sample bias corrections on MLEs or QMLEs (Bao and Ullah, 2007; Bao, 2013; Yang, 2015). A particularly interesting phenomenon revealed by Lee (2004) for the SLD model is that the spatial dependence may slow down the rate of convergence of QMLEs of certain model parameters, including the spatial parameter. An equally interesting phenomenon revealed by subsequent studies is that spatial dependence may cause QMLEs to be biased, and more so with heavier spatial dependence (Baltagi and Yang, 2013a,b; Yang, 2015; Liu and Yang, 2015).

Surprisingly, these issues have not been addressed in terms of the SED model. In particular, the effect of the *degree of spatial dependence* on the convergence rate of the QMLEs has not been formally studied, and methods for correcting finite-sample bias of the QMLEs for the SED model have not been given.[2] Built upon the works of Lee (2004) and Yang (2015), this paper fills in these gaps. Of the two, bias correction is particularly important to the applications of this model as it leads potentially to much improved inferences for the regression coefficients. Contrary to the common perceptions, both large and small sample behaviours of the QML estimators for the SED model can be different from those for the SLD model in terms of the rate of convergence and the magnitude of bias. In summary, the QMLE of the spatial parameter for the SED model *always* has a convergence rate slower than $\sqrt{n}$ whenever the degree of spatial dependence grows with the increase in sample size n, whereas the QMLEs of regression coefficient and error variance always have $\sqrt{n}$-rate of convergence whether or not the degree of spatial dependence increases with n. In contrast, the QMLEs of all the parameters in the SLD model have $\sqrt{n}$-rate of convergence when the spatially generated regressor is not asymptotically multicollinear with the original regressors (Lee, 2004, Assumption 8), and a slower than $\sqrt{n}$-rate of convergence occurs in some parameters for non-regular cases where the

[1]Other estimation methods include GMM (Kelejian and Robinson, 1993; Kelejian and Prucha, 1999; Lee, 2001, 2007; Fingleton, 2008), 2SLS (Kelejian and Prucha, 1998; Lee, 2003), IV estimation (Kelejian and Prucha, 2004), and OLS estimation (Lee, 2002).

[2]Here the *degree of spatial dependence* refers to, e.g., the number of neighbors each spatial unit has, or the connectivity in general. Jin and Lee (2013) studied asymptotic properties of models with both SLD and SED for the purpose of constructing Cox-type tests, but did not study these issues. Further, it is important to know the differences between the SLD model and the SED model in terms of asymptotic and finite sample behaviours, as they may provide a valuable guidance in the specification choice. See also Martellosio (2010) for a related work.

spatially generated regressor is asymptotically multicollinear with the original regressors and the degree of spatial dependence grows with the increase of n. Monte Carlo results show that the proposed bias correction procedure works very well for the SED model without compromising on the efficiency of the original QMLEs.

This paper is organised as follows. Section 2 presents results for consistency and asymptotic normality of the QMLEs for the SED model. Section 3 presents methods for finite sample bias correction. Section 4 extends the study to an alternative SED model where the spatial autoregressive (SAR) error is replaced by a spatial moving average (SMA) error; an undesirable feature of this alternative model specification is revealed. Section 5 presents Monte Carlo results and Section 6 concludes the paper.

2. Asymptotic Properties of QMLEs for SED Model

In this section, we examine the asymptotic properties of the QMLEs of the linear regression model with spatial error dependence, giving particular attention to the effect of spatial dependence on the rate of convergence of the QMLEs. We show that the QMLEs of the regression coefficients and the error variance always have the conventional $\sqrt{n}$-rate of convergence, whereas, the QMLE of the spatial parameter has the conventional $\sqrt{n}$-rate of convergence if the degree of spatial dependence does not grow with the increase in sample size, otherwise it has a slower rate. With an adjustment on the normalisation factor for the score component of the spatial parameter, we establish the joint asymptotic normality for the QMLEs of the model parameters. All proofs are given in Appendix A.

2.1 The model and the QML estimation

Consider the following linear regression model with spatial error dependence (SED), where the SED is specified as a spatial autoregressive (SAR) process:

$$Y_n = X_n\beta + u_n, \tag{1}$$

$$u_n = \rho W_n u_n + \epsilon_n, \tag{2}$$

where Y_n is an $n \times 1$ vector of observations on the dependent variable corresponding to n spatial units, X_n is an $n \times k$ matrix containing the values of k exogenous regressors, W_n is an $n \times n$ spatial weights matrix that summarises the interactions among the spatial units, ϵ_n is an $n \times 1$ vector of independent and identically distributed (iid) disturbances with mean zero and variance σ^2, ρ is the *spatial parameter*, and β denotes the $k \times 1$ vector of regression coefficients.

Let $\theta = (\beta', \sigma^2, \rho)'$ be the vector of model parameters and θ_0 be its true value. Denote $A_n(\rho) = I_n - \rho W_n$ and $A_n = A_n(\rho_0)$ where I_n is an $n \times n$ identity matrix. If A_n^{-1} exists, then Model (1) can be written as,

$$Y_n = X_n\beta_0 + A_n^{-1}\epsilon_n, \tag{3}$$

leading to $\text{Var}(u_n) = \text{Var}(A_n^{-1}\epsilon_n) = \sigma_0^2(A_n' A_n)^{-1}$.

The linear regression with spatial lag dependence (SLD) model has the form: $Y_n = \rho_0 W_n Y_n + X_n\beta_0 + \epsilon$, which can be rewritten as $Y_n = X_n\beta_0 + \rho_0 G_n X_n \beta_0 + A_n^{-1}\epsilon_n$, where $G_n = W_n A_n^{-1}$. While in both SED and SLD models, the spatial effects generate a non-spherical structure in the disturbance term, the SLD model has an extra *spatially generated regressor*, $G_n X_n \beta_0$. This spatial regressor plays an important role in the identification and estimation of the spatial parameter in the SLD model in a maximum likelihood estimation framework (Lee, 2004).

The first comprehensive treatment of maximum likelihood estimation for the SLD and SED models was given by Ord (1975). More formal results can be found in Anselin (1980). In particular, Anselin (1980) pointed out that the MLE of the SED model can be carried out as an application of the general framework of Magnus (1978) for non-spherical errors. See Anselin (1988); and Anselin and Bera (1998) for a detailed survey on the SLD and SED models.

While the SLD and SED models have been so fundamental and pivotal to the development of the spatial econometric models and methods, an important issue, which is perhaps unique to spatial econometrics models, the effect of the degree of spatial dependence on the asymptotic properties of the QMLEs, in particular the rate of convergence, was not addressed until Lee (2004) who clearly identified the situations where the rate of convergence can be affected when the spatial dependence increase with the number of observations. However, this issue has not been addressed in the context of SED models. Furthermore, as it will be seen from the following sections, the degree of spatial dependence also has a profound impact on the finite-sample performance of the spatial parameter estimates.

The quasi Gaussian log-likelihood function for the SED model is given by,

$$\ell_n(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) + \log|A_n(\rho)| - \frac{1}{2\sigma^2}(Y_n - X_n\beta)' A_n'(\rho) A_n(\rho)(Y_n - X_n\beta). \tag{4}$$

Maximizing $\ell_n(\theta)$ gives the MLE, $\hat\theta_n$ of θ if the errors are indeed Gaussian, otherwise the QMLE. Given ρ, the log-likelihood function $\ell_n(\theta)$ is partially maximized at,

$$\hat\beta_n(\rho) = [X_n' A_n'(\rho) A_n(\rho) X_n]^{-1} X_n' A_n'(\rho) A_n(\rho) Y_n, \text{ and} \tag{5}$$

$$\hat\sigma_n^2(\rho) = \tfrac{1}{n} Y_n' A_n'(\rho) M_n(\rho) A_n(\rho) Y_n, \tag{6}$$

where, $M_n(\rho) = I_n - A_n(\rho) X_n [X_n' A_n'(\rho) A_n(\rho) X_n]^{-1} X_n' A_n'(\rho)$. The concentrated log-likelihood function for ρ upon substituting the constrained QMLEs $\hat\beta_n(\rho)$ and $\hat\sigma_n^2(\rho)$ into (4):

$$\ell_n^c(\rho) = -\frac{n}{2}[\log(2\pi) + 1] + \log|A_n(\rho)| - \frac{n}{2}\log(\hat\sigma_n^2(\rho)). \tag{7}$$

Maximising $\ell_n^c(\rho)$ gives the unconstrained QMLE $\hat\rho_n$ of ρ, which in turn gives the unconstrained QMLEs of β and σ^2 as, $\hat\beta_n = \hat\beta_n(\hat\rho_n)$ and $\hat\sigma_n^2 = \hat\sigma_n^2(\hat\rho_n)$.

2.2 Consistency and asymptotic normality

The asymptotic properties of the QMLEs of the SED model are built upon the following basic regularity conditions:

Assumption 1: *The true ρ_0 is in the interior of the compact parameter set $\mathcal{P}$.*

Assumption 2: *$\{\epsilon_{n,i}\}$ are iid with mean 0, variance σ^2, and $\mathrm{E}|\epsilon_{n,i}|^{4+\delta} < \infty, \forall \delta > 0$.*

Assumption 3: *X_n has full column rank k, its elements are uniformly bounded constants, and $\lim_{n\to\infty} \frac{1}{n} X_n' A_n'(\rho) A_n(\rho) X_n$ exists and is non-singular for any ρ in a neighbourhood of ρ_0.*

Assumption 4: *The elements $\{w_{ij}\}$ of W_n are at most of order h_n^{-1} uniformly for all i and j, where h_n can be bounded or divergent but subject to $\lim_{n\to\infty} \frac{h_n}{n} = 0$; W_n is uniformly bounded in both row and column sums and its diagonal elements are zero.*

Assumption 5: *A_n is non-singular and A_n^{-1} is uniformly bounded in both row and column sums. Further, $A_n^{-1}(\rho)$ is uniformly bounded in either row or column sums, uniformly in $\rho \in \mathcal{P}$.*

We allow for the possibility that the degree of spatial dependence, quantified by h_n, grows with the sample size n, and the possibility that the error distribution is misspecified, i.e., the true error distribution is not normal. These conditions are similar to those in Lee (2004) to ascertain the $\sqrt{n/h_n}$-consistency of the QMLEs of the SLD model. All conditions but that on h_n are very general regularity conditions considered widely in the literature. Assumption 1 states that the spatial parameter ρ can only take values in a compact space such that the Jacobian term of the likelihood function, $\log|A_n(\rho)|$, is well defined.[3] The full rank condition of Assumption 3 is needed to guarantee that the model does not suffer from multicollinearity. Assumption 4 is based on Lee (2004) where extensive discussions can be found. Assumption 5 allows us to write the model in the reduced form (3). Uniform boundedness conditions given in Assumptions 4 and 5 are needed to limit the spatial correlation to a manageable degree. Boundedness on the regressors is not restrictive when analysing cross-sectional units, and in case of with stochastic regressors it can be replaced by certain finite moment conditions.

Identification of the model parameters requires that the expected log-likelihood function, $\bar{\ell}_n(\theta) = \mathrm{E}[\ell_n(\theta)]$, has identifiably unique maximisers that converge to θ_0 as $n \to \infty$. (White, 1994, Theorem 3.4; Lee, 2004). The expected log-likelihood function is,

$$\bar{\ell}_n(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) + \log|A_n(\rho)| - \frac{1}{2\sigma^2}\mathrm{E}\left[(Y_n - X_n\beta)' A_n'(\rho) A_n(\rho)(Y_n - X_n\beta)\right], \quad (8)$$

which, for a given ρ, is partially maximised at,

$$\begin{aligned}
\beta_n(\rho) &= (X_n' A_n'(\rho) A_n(\rho) X_n)^{-1} X_n' A_n'(\rho) A_n(\rho) \mathrm{E}(Y_n) = \beta_0, \quad \text{and} &(9)\\
\sigma_n^2(\rho) &= \tfrac{1}{n}\mathrm{E}\{[Y_n - X_n\beta_n(\rho)]' A_n'(\rho) A_n(\rho)[Y_n - X_n\beta_n(\rho)]\}\\
&= \tfrac{1}{n}\mathrm{E}\{\mathrm{tr}[\epsilon_n\epsilon_n' A_n'^{-1} A_n'(\rho) A_n(\rho) A_n^{-1}]\}\\
&= \tfrac{1}{n}\sigma_0^2 \mathrm{tr}[A_n'^{-1} A_n'(\rho) A_n(\rho) A_n^{-1}]. &(10)
\end{aligned}$$

[3] For this it is necessary that $|I_n - \rho W_n| = \prod_{i=1}^n (1-\rho\lambda_i) > 0$, where $\{\lambda_i\}$ are the eigenvalues of W_n. If the eigenvalues of W_n are all real, the parameter space $\mathcal{P}$ can be a closed interval contained in $(\lambda_{\min}^{-1}, \lambda_{\max}^{-1})$, where $\lambda_{\min}$ and $\lambda_{\max}$ are, respectively, the minimum and maximum eigenvalues. If W_n is row-normalised, then $\lambda_{\max} = 1$ and $-1 \le \lambda_{\min} < 0$ and $\mathcal{P}$ can be a closed interval contained in $(\lambda_{\min}^{-1}, 1)$, where the lower bound can be below -1 (Anselin, 1988). In general, the eigenvalues of W_n may not be all real and in this case Kelejian and Prucha (2010) suggested the interval $(-\tau_n^{-1}, \tau_n^{-1})$, where, $\tau_n = \max_i|\lambda_i|$ is the spectral radius of the weights matrix, and LeSage and Pace (2009, p. 88-89) suggested interval $(\lambda_s^{-1}, 1)$ where λ_s is the most negative real eigenvalue of W_n as only the real eigenvalues can affect the singularity of $I_n - \lambda W_n$.

The resulting concentrated expected log-likelihood function, $\bar{\ell}_n^c(\rho)$, takes the form,

$$\bar{\ell}_n^c(\rho) = \max_{\beta,\sigma^2} \bar{\ell}_n(\theta) = -\frac{n}{2}(\log(2\pi)+1) + \log|A_n(\rho)| - \frac{n}{2}\log(\sigma_n^2(\rho)). \tag{11}$$

From Assumption 3, it is clear that β and σ^2 are identified once ρ is. The latter is guaranteed if $\bar{\ell}_n^c(\rho)$ has an identifiably unique maximiser in $\mathcal{P}$ which converges to ρ_0 as $n \to \infty$, or $\lim_{n\to\infty} \frac{h_n}{n}[\bar{\ell}_n^c(\rho) - \bar{\ell}_n^c(\rho_0)] < 0$, $\forall \rho \neq \rho_0$. The global identification condition for the SED model thus simplifies to a condition on ρ alone.

Assumption 6: $\lim_{n\to\infty} \frac{h_n}{n} \left[\log|\sigma_0^2 A_n^{-1} A_n'^{-1}| - \log|\sigma_n^2(\rho) A_n^{-1}(\rho) A_n'^{-1}(\rho)|\right] \neq 0, \forall \rho \neq \rho_0$.

This differentiates the SED model from the SLD in the asymptotic behaviours of the QMLEs. The spatially generated regressor $G_n X_n \beta_0$ of the SLD model $Y_n = X_n\beta_0 + \rho_0 G_n X_n \beta_0 + A_n^{-1}\epsilon_n$ can help identifying ρ if it is not asymptotically multicollinear with the original regressors, giving the conventional $\sqrt{n}$-rate of convergence of $\hat{\rho}_n$ irrespective of whether h_n is bounded or unbounded. When $G_n X_n \beta_0$ is asymptotically collinear with X_n, the convergence rate of $\hat{\rho}_n$ becomes $\sqrt{n/h_n}$. In contrast, $\hat{\rho}_n$ for the SED model always has a $\sqrt{n/h_n}$-rate of convergence. Note that the variance of Y_n of (1) is $\sigma_0^2 A_n^{-1} A_n'^{-1}$ and hence the global identification condition given above ensures the uniqueness of the variance matrix. With this global identification condition and the uniform convergence of $\frac{h_n}{n}[\ell_n^c(\rho) - \bar{\ell}_n^c(\rho)]$ to zero in $\mathcal{P}$ which is proved in the Appendix, the consistency of $\hat{\rho}_n$ follows.

Theorem 1: *Under Assumptions 1-6, the QMLE $\hat{\rho}_n$ is a consistent estimator of ρ_0.*

Theorem 1 and Assumption 3 lead immediately to the consistency of $\hat{\beta}_n$ and $\hat{\sigma}_n^2$. However, Theorem 1 reveals nothing about the rate of convergence of $\hat{\rho}_n$, and hence the rates of convergence of $\hat{\beta}_n$ and $\hat{\sigma}_n^2$ remain unknown as well. To reveal the exact convergence rates, and at the same time to derive the asymptotic distributions of the QMLEs, consider the score function,

$$S_n(\theta) \equiv \frac{\partial \ell_n(\theta)}{\partial \theta} = \begin{cases} \frac{1}{\sigma^2} X_n' A_n'(\rho) A_n(\rho) u_n(\beta), \\ \frac{1}{2\sigma^4} u_n'(\beta) A_n'(\rho) A_n(\rho) u_n(\beta) - \frac{n}{2\sigma^2}, \\ \frac{1}{\sigma^2} u_n'(\beta) A_n'(\rho) W_n u_n(\beta) - \mathrm{tr}[G_n(\rho)], \end{cases} \tag{12}$$

where, $u_n(\beta) = Y_n - X_n\beta$ and $G_n(\rho) = W_n A_n^{-1}(\rho)$. It is known that for likelihood-based inferences, the normalized score $\frac{1}{\sqrt{n}} S_n(\theta_0)$ at the true parameter value would be asymptotically normal. Indeed, under Assumptions 1-5 one can easily show that this is true for β and σ^2 components of $\frac{1}{\sqrt{n}} S_n(\theta_0)$. However, the normalized score for ρ is $O_p(\frac{1}{\sqrt{h_n}})$, see Lemmas A.2 and A.3 in Appendix. This means that when h_n is divergent, the likelihood function with respect to ρ is too flat so that its normalized score converges to a degenerate distribution. As a result $\hat{\rho}_n$ converges to ρ_0 at a slower rate than the conventional $\sqrt{n}$-rate. A similar phenomenon is observed by Lee (2004) for the spatial parameter as well as the regression coefficients in the SLD model, in the 'non-regular cases' where the spatially generated regressor $G_n X_n \beta_0$, is asymptotically collinear with the regular regressors. This motivate us to consider the following modification.

To account for the effect of spatial dependence on the asymptotic behaviour of the QMLE $\hat{\rho}_n$ of the spatial parameter ρ, and to jointly study the asymptotic distribution of the QMLE $\hat{\theta}_n$ of the model parameter vector θ, we consider the following modified score vector:

$$S_n^*(\theta) = K_n S_n(\theta),$$

where, $K_n = \text{diag}(I_k, 1, \sqrt{h_n})$. Hence, $\frac{1}{\sqrt{n}} S_n^*(\theta)$ would have a proper asymptotic behaviour whether h_n is divergent or bounded. Under Assumptions 1-5, the central limit theorem (CLT) for linear-quadratic forms of Kelejian and Prucha (2001) can be applied to prove the result,

$$\frac{1}{\sqrt{n}} S_n^*(\theta_0) \xrightarrow{D} N(0, \Gamma^*),$$

where, $\Gamma^* = \lim_{n\to\infty} \frac{1}{n}\Gamma_n^*$, $\Gamma_n^* = \text{Var}[S_n^*(\theta_0)] = K_n \Gamma_n K_n'$, $\Gamma_n = \text{Var}[S_n(\theta_0)]$, and

$$\Gamma_n = \begin{pmatrix} \frac{1}{\sigma_0^2} X_n' A_n' A_n X_n & \frac{1}{2\sigma_0^3}\gamma X_n' A_n' \iota_n & \frac{1}{\sigma_0}\gamma X_n' A_n' g_n \\ \frac{1}{2\sigma_0^3}\gamma \iota_n' A_n X_n & \frac{n}{4\sigma_0^4}(\kappa+2) & \frac{1}{2\sigma_0^2}(\kappa+2)\text{tr}(G_n) \\ \frac{1}{\sigma_0}\gamma g_n' A_n X_n & \frac{1}{2\sigma_0^2}(\kappa+2)\text{tr}(G_n) & \kappa g_n' g_n + \text{tr}(G_n^s G_n) \end{pmatrix},$$

where, ι_n is an $n\times 1$ vector of ones, $\gamma = \sigma_0^{-3}\text{E}(\epsilon_{n,i}^3)$ is the measure of skewness, $\kappa = \sigma_0^{-4}\text{E}(\epsilon_{n,i}^4) - 3$ is the measure of excess kurtosis, $g_n = \text{diag}(G_n)$, $G_n = G_n(\rho_0)$, and $G_n^s = G_n + G_n'$.

It is easy to see that the information matrix $\Sigma_n = -\text{E}\left(\frac{\partial^2}{\partial\theta\partial\theta'}\ell_n(\theta_0)\right)$, takes the form:

$$\Sigma_n = \begin{pmatrix} \frac{1}{\sigma_0^2} X_n' A_n' A_n X_n & 0 & 0 \\ 0 & \frac{n}{2\sigma_0^4} & \frac{1}{\sigma_0^2}\text{tr}(G_n) \\ 0 & \frac{1}{\sigma_0^2}\text{tr}(G_n) & \text{tr}(G_n^s G_n) \end{pmatrix},$$

which leads to the modified version of the information matrix, $\Sigma_n^* = K_n \Sigma_n K_n'$. One can show that Γ^* exists and its diagonal elements are non-zero and $\Sigma^* = \lim_{n\to\infty} \frac{1}{n}\Sigma_n^*$ exists and is positive definite irrespective of whether h_n is bounded or unbounded. In contrast,

$$\lim_{n\to\infty} \frac{1}{n}\Gamma_n = \begin{pmatrix} \frac{1}{\sigma_0^2} V_1 & \frac{\gamma}{2\sigma_0^3} V_2 & 0 \\ \frac{\gamma}{2\sigma_0^3} V_2' & \frac{1}{4\sigma_0^4}(\kappa+2) & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ and } \lim_{n\to\infty} \frac{1}{n}\Sigma_n = \begin{pmatrix} \frac{1}{\sigma_0^2} V_1 & 0 & 0 \\ 0 & \frac{1}{2\sigma_0^4} & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

if h_n is unbounded, where, $V_1 = \lim_{n\to\infty} \frac{1}{n} X_n' A_n' A_n X_n$ and $V_2 = \lim_{n\to\infty} \frac{1}{n} X_n' A_n' \iota_n$. Hence, without the adjustment factor K_n, we cannot derive the asymptotic normality results due to the singularity of the matrices required to compute the asymptotic variance-covariance matrix.

To see that Σ^* is non-singular under a general h_n, consider the determinant of Σ_n^*: $|\Sigma_n^*| = \frac{1}{2\sigma_0^6}\frac{1}{n}|X_n' A_n' A_n X_n|\frac{h_n}{n}[\text{tr}(G_n^s G_n) - \frac{2}{n}\text{tr}^2(G_n)]$. If h_n is bounded then by Assumptions 3, 4 and 5, $|\Sigma_n^*| = O(1)$. Now suppose h_n is unbounded where $\lim_{n\to\infty} h_n = \infty$ such that $\frac{h_n}{n} \to 0$, then $g_{n,ii}$, $\frac{1}{n}\text{tr}(G_n' G_n)$, $\frac{1}{n}\text{tr}(G_n^2)$, and $\frac{1}{n}\text{tr}(G_n)$ are all $O(h_n^{-1})$ and hence by Assumption 3, $|\Sigma_n^*| = O(1)$.

We have the following theorem for asymptotic normality of QMLE $\hat{\theta}_n$ of θ_0.

Theorem 2: *Under Assumptions 1-6, we have,*

$$\sqrt{n}K_n^{-1}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0,\ \Sigma^{*-1}\Gamma^{*}\Sigma^{*-1}),$$

where, $\Gamma^* = \lim_{n\to\infty}\frac{1}{n}\Gamma_n^*$ *and* $\Sigma^* = \lim_{n\to\infty}\frac{1}{n}\Sigma_n^*$. *If errors* $\{\epsilon_{n,i}\}$ *are normally distributed, then* $\sqrt{n}K_n^{-1}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0,\ \Sigma^{*-1})$.

Remark 1: For practical applications of the above result, it is important to note that h_n, the quantity characterising the degree of spatial dependence and affecting the rate of convergence of the QMLEs, is not known in general. However, inference concerning the model parameters does not depend on it, because $\Sigma_n^{*-1}\Gamma_n^*\Sigma_n^{*-1} = (K_n\Sigma_nK_n)^{-1}(K_n\Gamma_nK_n)(K_n\Sigma_nK_n)^{-1} = K_n^{-1}\Sigma_n^{-1}\Gamma_n\Sigma_n^{-1}K_n^{-1}$. Hence, $\text{AVar}(\hat{\theta}_n - \theta_0) = n^{-1}\Sigma_n^{-1}\Gamma_n\Sigma_n^{-1}$.

For the purpose of statistical inference, it might be useful to have the marginal asymptotic distributions of the QMLEs, in particular, the marginal asymptotic distribution of $\hat{\rho}_n$.

Corollary 1: *Under the assumptions of Theorem 2, we have,*

$$\begin{aligned}
\sqrt{n}(\hat{\beta}_n - \beta_0) &\xrightarrow{D} N\big(0,\ \sigma_0^2V_1^{-1}\big),\\
\sqrt{n}(\hat{\sigma}_n^2 - \sigma_0^2) &\xrightarrow{D} N\big[0,\ 2\sigma_0^4T_1 + \kappa\sigma_0^4(T_1 - 2T_2^2T_3)\big],\\
\sqrt{\tfrac{n}{h_n}}(\hat{\rho}_n - \rho_0) &\xrightarrow{D} N\big(0,\ T_4 + \kappa T_5\big);
\end{aligned}$$

where, $T_1 = \lim_{n\to\infty}\frac{tr(G_n^sG_n)}{tr(C_n^sC_n)}$, $T_2 = \lim\limits_{n\to\infty}\frac{tr(G_n)}{tr(C_n^sC_n)}$, $T_3 = \lim\limits_{n\to\infty}\frac{1}{n}\left[tr(G_n^sG_n) - 2g_n'g_n\right]$, $T_4 = \lim\limits_{n\to\infty}\frac{n}{h_n}tr^{-1}(C_n^sC_n)$, $T_5 = \lim\limits_{n\to\infty}\frac{n}{h_n}\frac{g_n'g_n - n^{-1}tr^2(G_n)}{tr^2(C_n^sC_n)}$, $C_n = G_n - \frac{tr(G_n)}{n}I_n$ *and* $C_n^s = C_n' + C_n$.

Corollary 1 clearly reveals that only the QMLE of the spatial parameter has a slower rate of convergence of $\sqrt{n/h_n}$ when h_n is unbounded, which says that the effect of a growing spatial dependence is that the effective sample size for estimating ρ is reduced to n/h_n; $\hat{\beta}_n$ and $\hat{\sigma}_n^2$ have the traditional $\sqrt{n}$-convergence rate whether h_n is bounded or unbounded. Intuitively this is correct since unlike in the SLD model where there is a lagged dependent variable W_nY_n, in the SED model, the spatial structure affects only the errors and hypothetically if ρ is known, the model in (1) can be simplified to a linear regression model.

We note that due to the block-diagonal structure of Σ_n and the fact that the skewness measure γ appears only in the off-diagonal blocks of Γ_n, the marginal asymptotic distributions do not depend upon γ. For general asymptotic inferences, γ and κ can be consistently estimated by $\hat{\gamma}_n = \frac{1}{n\hat{\sigma}_n^3}\sum_{i=1}^n \hat{\epsilon}_{n,i}^3$ and $\hat{\kappa}_n = \frac{1}{n\hat{\sigma}_n^4}\sum_{i=1}^n \hat{\epsilon}_{n,i}^4 - 3$, respectively, where $\hat{\epsilon}_{n,i}$ are the QML residuals. Thus, the estimates of Σ_n and Γ_n are obtained by plugging in $\hat{\theta}_n$, $\hat{\gamma}_n$ and $\hat{\kappa}_n$ into Σ_n and Γ_n. These discussions show that the asymptotic inferences for the SED model based on QML estimation are extremely simple. However, an important question remains: how do they perform in finite samples? Take a simple, and a very important special case where the inference concerns the regression coefficients β. While the bias of $\hat{\rho}_n$ does not have much impact on the bias of $\hat{\beta}_n$ and $\hat{\sigma}_n^2$, it does translate into the bias of the variance estimator of $\hat{\beta}_n$ through the term $\hat{V}_n = \frac{1}{n}X_n'A_n'(\hat{\rho}_n)A_n(\hat{\rho}_n)X_n$ (see the end of Section 4). This shows the importance of bias correction for the SED model, or perhaps for the more general models with non-spherical errors.

3. Finite-Sample Bias Correction for the QML Estimators

With the formal asymptotic results given in the earlier section, we are ready to study the more important issue: the finite sample properties of the QMLEs of the SED model. The problem of estimation bias, arising from the estimation of non-linear parameters has been widely recognized by econometricians (see, among others, Kiviet, 1995; Hahn and Kuersteiner, 2002; Hahn and Newey, 2004; Bun and Carree, 2005). Spatial econometricians too have recognized this issue in estimating spatial econometric models and have successfully tackled this problem for the SLD model (Bao and Ullah, 2007; Bao, 2013; Yang, 2015). However, no work has been done for the SED model and other spatial models. In a spatial regression context, spatial parameter(s) enter the regression model in a highly non-linear manner and spatial dependence maybe quite strong. As a result, the bias problem in estimating spatial parameter(s) may be quite severe, and hence it is very important to perform bias corrections on spatial estimator(s). Among the various methods for bias corrections, the *stochastic expansion method* of Rilstone et al. (1996) has recently gained more attention. With the introduction of the bootstrap method by Yang (2015), its applicability has been greatly expanded (See Efron, 1979, for a general introduction to the bootstrap method).

In this section, we derive the second- and third-order biases of the QMLE of the spatial parameter in the SED model, based on the technique of stochastic expansion (Rilstone et al., 1996) and bootstrap (Yang, 2015). As in Yang (2015), the key quantities involved in the terms related to the bias of a non-linear estimator are the derivatives of the concentrated log-likelihood function and their expectations. While deriving the analytical solutions of the higher-order derivatives may only be a matter of tedious algebraic manipulations, evaluation of their expectations can be very difficult if not impossible. We follow the general method introduced in Yang (2015) and propose a bootstrap procedure for implementing these bias corrections for the SED model. The validity of this procedure when applied to the SED model is established. Monte Carlo results show an excellent performance of the proposed bias-correction procedure. We argue that once the spatial estimator is bias-corrected, the estimators of the other models parameters become nearly unbiased. All proofs are given in Appendix B.

3.1 The general method for bias correction

In studying the finite sample properties of a parameter estimator, say $\hat{\theta}_n$, defined as $\hat{\theta}_n = \arg\{\psi_n(\theta) = 0\}$ for an estimating function $\psi_n(\theta)$, based on a sample of size n, Rilstone et al. (1996) and Bao and Ullah (2007) developed a stochastic expansion from which a bias-correction on $\hat{\theta}_n$ can be made. The vector of parameters θ may contain a set of *linear and scale parameters*, say δ, and a *non-linear parameter*, say ρ, in the sense that given ρ, the constrained estimator $\hat{\delta}_n(\rho)$ of the vector δ possesses an explicit expression and the estimation of ρ has to be done through numerical optimization. In this case, Yang (2015) argued that it is more effective to work with the *concentrated estimating function* (CEF): $\tilde{\psi}_n(\rho) = \psi_n(\hat{\delta}_n(\rho), \rho)$, and to perform a stochastic expansion on this CEF and hence do the bias correction only on the non-linear estimator defined by,

$$\hat{\rho}_n = \arg\{\tilde{\psi}_n(\rho) = 0\}. \tag{13}$$

In doing so, a multi-dimensional problem is reduced to a single-dimensional problem, and the additional variability from the estimation of the 'nuisance' parameters δ is taken into account in bias-correcting the estimate of the non-linear parameter ρ.

Let $H_{rn}(\rho) = \frac{d^r}{d\rho^r}\tilde{\psi}_n(\rho), r = 1, 2, 3$. Under some general smoothness conditions on $\tilde{\psi}_n(\rho)$, Yang (2015) presented a third-order, CEF-based, stochastic expansion for $\hat{\rho}_n$ at the true parameter value ρ_0 as,

$$\hat{\rho}_n - \rho_0 = a_{-1/2} + a_{-1} + a_{-3/2} + O_p(n^{-2}), \tag{14}$$

where, $a_{-s/2}$ represents terms of order $O_p(n^{-s/2})$ for $s = 1, 2, 3$, and they are,

$$a_{-1/2} = \Omega_n\tilde{\psi}_n, \qquad a_{-1} = \Omega_n H_{1n}^{\circ} a_{-1/2} + \tfrac{1}{2}\Omega_n \mathrm{E}(H_{2n})(a_{-1/2}^2) \qquad \text{and}$$
$$a_{-3/2} = \Omega_n H_{1n}^{\circ} a_{-1} + \tfrac{1}{2}\Omega_n H_{2n}^{\circ}(a_{-1/2}^2) + \Omega_n \mathrm{E}(H_{2n})(a_{-1/2}a_{-1}) + \tfrac{1}{6}\Omega_n \mathrm{E}(H_{3n})(a_{-1/2}^3),$$

where, $\tilde{\psi}_n \equiv \tilde{\psi}_n(\rho_0)$, $H_{rn} \equiv H_{rn}(\rho_0), r = 1, 2, 3$, $H_{rn}^{\circ} = H_{rn} - \mathrm{E}(H_{rn})$ and $\Omega_n = -[\mathrm{E}(H_{1n})]^{-1}$.

The above stochastic expansion leads to a second-order bias, $\mathrm{E}(a_{-1/2} + a_{-1})$, and a third-order bias, $\mathrm{E}(a_{-1/2} + a_{-1} + a_{-3/2})$, which may be used for performing bias corrections on $\hat{\rho}_n$, provided that analytical expressions of the various expected quantities in the expansion can be derived so that they can be estimated through a plug-in method. Several applications of this plug-in method have appeared in the literature including Bao and Ullah (2007) for the pure spatial autoregressive process, and Bao (2013) for the SLD model. The plug-in method may run into difficulty when the analytical expectations are not available or are difficult/impossible to derive as in the SED model we consider. To overcome this obstacle, Yang (2015) proposed a simple and yet a very effective bootstrap method to estimate the relevant expected values.

3.2 Bias of the QMLE of the spatial parameter of the SED model

Recall the concentrated log-likelihood function, defined in (7). Define the concentrated score function or the CEF for ρ as, $\tilde{\psi}_n(\rho) = \frac{\partial}{\partial\rho}\frac{h_n}{n}\ell_n^c(\rho)$, then,

$$\tilde{\psi}_n(\rho) = -h_n T_{0n}(\rho) + h_n R_{1n}(\rho), \tag{15}$$

where, $T_{0n}(\rho) = \frac{1}{n}\mathrm{tr}(G_n(\rho))$ and

$$R_{1n}(\rho) = \frac{Y_n' A_n'(\rho) M_n(\rho) G_n(\rho) M_n(\rho) A_n(\rho) Y_n}{Y_n' A_n'(\rho) M_n(\rho) A_n(\rho) Y_n}, \tag{16}$$

leading to $\hat{\rho}_n = \arg\{\tilde{\psi}_n(\rho) = 0\}$. Let $H_{rn}(\rho) = \frac{d^r}{d\rho^r}\tilde{\psi}_n(\rho), r = 1, 2, 3$, then,

$$\begin{aligned}
h_n^{-1}H_{1n}(\rho) &= -T_{1n}(\rho) - R_{2n}(\rho) + 2R_{1n}^2(\rho), && (17)\\
h_n^{-1}H_{2n}(\rho) &= -2T_{2n}(\rho) - R_{3n}(\rho) - 6R_{1n}(\rho)R_{2n}(\rho) + 8R_{1n}^3(\rho), && (18)\\
h_n^{-1}H_{3n}(\rho) &= -6T_{3n}(\rho) - R_{4n}(\rho) - 8R_{1n}(\rho)R_{3n}(\rho) + 6R_{2n}^2(\rho) \\
&\quad -48R_{1n}^2(\rho)R_{2n}(\rho) + 48R_{1n}^4(\rho), && (19)
\end{aligned}$$

where, $T_{rn}(\rho) = \frac{1}{n}tr(G_n^{r+1}(\rho)), r = 1, 2, 3,$ and

$$R_{jn}(\rho) = \frac{Y_n' A_n'(\rho) M_n(\rho) D_{jn}(\rho) M_n(\rho) A_n(\rho) Y_n}{Y_n' A_n'(\rho) M_n(\rho) A_n(\rho) Y_n}, \ j = 2, 3, 4. \tag{20}$$

The full expressions for $D_{jn}(\rho), j = 2, 3, 4$ are given in Appendix B. Clearly, $D_{1n}(\rho) = G_n(\rho)$ in $R_{1n}(\rho)$.

The above expressions show that the key quantities in the third-order stochastic expansion for $\hat{\rho}_n$ (the QMLE of the spatial parameter in the SED model), are those ratios of quadratic forms $R_{jn}(\rho), j = 1\ldots, 4$. Note that, in what follows, a function of ρ evaluated at $\rho = \rho_0$ is denoted by dropping the function argument, e.g., $\tilde{\psi}_n = \tilde{\psi}_n(\rho_0), A_n = A_n(\rho_0), G_n = G_n(\rho_0)$, $R_{jn} = R_{jn}(\rho_0)$, $H_{rn} = H_{rn}(\rho_0)$, $T_{rn} = T_{rn}(\rho_0)$. Now, some case specific conditions on R_{jn} are needed to regulate the limiting behaviour of H_{rn} so that the required quantities have finite limits in expectation.

Assumption 7: $\mathrm{E}\big(\frac{h_n}{n}\epsilon_n' M_n G_n M_n \epsilon_n \big(\frac{1}{\bar{\sigma}_n^4} - \frac{1}{\sigma_0^4}\big)(\hat{\sigma}_n^2 - \sigma_0^2)\big) = O\big((\frac{h_n}{n})^{\frac{1}{2}}\big)$, *where,* $\bar{\sigma}_n^2$ *lies between* σ_0^2 *and* $\hat{\sigma}_n^2$.

Assumption 8:
(*i*) $h_n^s \mathrm{E}[(R_{1n} - \mathrm{E}R_{1n})^s] = O\big((\frac{h_n}{n})^{\frac{1}{2}}\big), s = 2, 3, 4;$
(*ii*) $h_n^s \mathrm{E}[(R_{2n} - \mathrm{E}R_{2n})^s] = O\big((\frac{h_n}{n})^{\frac{1}{2}}\big), s = 1, 2;$
(*iii*) $h_n \mathrm{E}(R_{rn} - \mathrm{E}R_{rn}) = O((\frac{h_n}{n})^{\frac{1}{2}}), r = 3, 4;$
(*iv*) $h_n^{s+1} \mathrm{E}[(R_{1n} - \mathrm{E}R_{1n})^s (R_{2n} - \mathrm{E}R_{2n})] = O((\frac{h_n}{n})^{\frac{1}{2}}), s = 1, 2,$ and
(*v*) $h_n^2 \mathrm{E}[(R_{1n} - \mathrm{E}R_{1n})(R_{3n} - \mathrm{E}R_{3n})\big] = O((\frac{h_n}{n})^{\frac{1}{2}}).$

The following Lemma shows the bounded behaviour of the expectations of the quantities in the stochastic expansion.

Lemma 1: *Under Assumptions 1-7, (i)* $h_n R_{in} = O_p(1)$, *(ii)* $\mathrm{E}(h_n R_{in}) = O(1)$, *and (iii)* $h_n R_{in} = \mathrm{E}(h_n R_{in}) + O_p((\frac{h_n}{n})^{\frac{1}{2}}), i = 1, \ldots, 4.$

Given Lemma 1 and the regularity conditions, we can prove the following propositions:

Proposition 1: *Suppose the SED model specified by* (1) *and* (2) *satisfies Assumptions 1-8. Then, the third-order stochastic expansion given in* (14) *holds for the QMLE* $\hat{\rho}_n$ *of the spatial parameter in the model with* n *replaced by* n/h_n *for the stochastic order:*

$$\hat{\rho}_n - \rho_0 = c_{1n}'\zeta_n + c_{2n}'\zeta_n + c_{3n}'\zeta_n + O_p((\tfrac{h_n}{n})^2), \tag{21}$$

where, $c_{sn}'\zeta_n$ *are of stochastic order* $O((\frac{h_n}{n})^{\frac{s}{2}}), s = 1, 2, 3,$ *with,*
$\zeta_n = \{\tilde{\psi}_n,\ H_{1n}\tilde{\psi}_n,\ \tilde{\psi}_n^2,\ H_{1n}^2\tilde{\psi}_n,\ H_{2n}\tilde{\psi}_n^2,\ H_{1n}\tilde{\psi}_n^2,\ \psi_n^3\}'$,
$c_{1n} = \{\Omega_n,\ 0_{6\times 1}'\}'$, $\quad \Omega_n = -\mathrm{E}(H_{1n})^{-1}$, $\quad c_{2n} = \{\Omega_n,\ \Omega_n^2,\ \frac{1}{2}\Omega_n^3 E(H_{2n}),\ 0_{4\times 1}'\}'$, *and*
$c_{3n} = \{\Omega_n,\ 2\Omega_n^2,\ \Omega_n^3\mathrm{E}(H_{2n}),\ \Omega_n^3,\ \frac{1}{2}\Omega_n^3,\ \frac{3}{2}\Omega_n^4\mathrm{E}(H_{2n}),\ \frac{1}{2}\Omega_n^5\mathrm{E}(H_{2n})^2 + \frac{1}{6}\Omega_n^4\mathrm{E}(H_{3n})\}'$.

Remark 2: Note that by letting $C_{2n} = c_{1n} + c_{2n}$ and $C_{3n} = c_{1n} + c_{2n} + c_{3n}$, the stochastic expansions can be further simplified to $c_{1n}'\zeta_n$ (asymptotic), $C_{2n}'\zeta_n$ (second-order), and $C_{3n}'\zeta_n$ (third order), which are particularly helpful in the bootstrap work introduced later.

Proposition 2: *Under Assumptions 1-8 and further assuming that a quantity bounded in probability has a finite expectation, a third-order expansion for the bias of $\hat{\rho}_n$ is:*

$$\text{Bias}(\hat{\rho}_n) = C_{2n}'\text{E}(\zeta_n) + c_{3n}'\text{E}(\zeta_n) + O((\tfrac{h_n}{n})^2), \tag{22}$$

and the 2nd and 3rd order bias corrected QMLEs are:

$$\hat{\rho}_n^{bc2} = \hat{\rho}_n - \widehat{C}_{2n}'\widehat{\text{E}}(\zeta_n) \quad \textit{and} \quad \hat{\rho}_n^{bc3} = \hat{\rho}_n - \widehat{C}_{3n}'\widehat{\text{E}}(\zeta_n), \tag{23}$$

where, a quantity with a $\widehat{}$ *is the corresponding estimate of that quantity.*

Practical implementation of the bias corrections given in (23) depends on the availability of the estimates $\widehat{\text{E}}(\zeta_n)$, and $\widehat{C}_{2n}$ or $\widehat{C}_{3n}$. Note that ζ_n is defined in terms of $\tilde{\psi}_n$ and H_{rn}, and C_{2n} and C_{3n} are defined in terms of $\text{E}(H_{rn}), r = 1, 2, 3$. Given the complicated expressions for $\tilde{\psi}_n$ and H_{rn} defined in (15)-(19), the conventional method of estimation by deriving the analytical expectations for $\text{E}(\zeta_n)$, and C_{2n} or C_{3n} would be extremely difficult if not impossible. The method of using the sample analogue would not work either due to the fact that $\tilde{\psi}(\hat{\rho}_n) = 0$. These iterate the point raised in Yang (2015), and hence, the bootstrap method given in same is adopted for the estimation of the quantities in question.

3.3 Bootstrap method for implementing the bias-correction

From (15), and (17)-(19), we see that $\tilde{\psi}_n$ and H_{rn} are functions of only $R_{jn}, j = 1, \ldots, 4$, i.e., we need to individually estimate the following terms:

$$\begin{gathered}\text{E}(R_{1n}^i), i = 1, \ldots, 5; \quad \text{E}(R_{2n}^j), j = 1, 2; \quad \text{E}(R_{3n}); \quad \text{E}(R_{4n});\\ \text{E}(R_{1n}^i R_{2n}), i = 1, 2, 3; \quad \text{E}(R_{1n}R_{2n}^2); \quad \text{E}(R_{1n}^i R_{3n}), i = 1, 2.\end{gathered}$$

It is easy to see that,

$$R_{jn} \equiv R_{jn}(e_n, \rho_0) = \frac{e_n'\Lambda_{jn}(\rho_0)e_n}{e_n'M_n(\rho_0)e_n}, \tag{24}$$

where $e_n = \sigma_0^{-1}\epsilon_n$, $\Lambda_{jn}(\rho_0) = M_n(\rho_0)D_{jn}M_n(\rho_0)$ with $D_{1n} = G_n$ and $D_{jn}, j = 2, 3$ being defined at the beginning of Appendix B. It follows that all the necessary quantities whose expectations are required can be expressed in terms of e_n and ρ_0. In particular, we can write,

$$H_{rn} \equiv H_{rn}(e_n, \rho_0), \quad \text{and} \quad \zeta_n \equiv \zeta_n(e_n, \rho_0).$$

Thus, H_{rn} and ζ_n, and their distributions are *invariant* of β_0 and σ_0^2. The bootstrap procedure for estimating the expectations of the above quantities can be described as follows:

(1) Compute the QMLEs $\hat{\theta}_n = (\hat{\beta}_n', \hat{\sigma}_n^2, \hat{\rho}_n)'$ based on the original data,

(2) Compute the standardized QML residuals, $\hat{e}_n = \hat{\sigma}_n^{-1}A_n(\hat{\rho}_n)(Y_n - X_n\hat{\beta}_n)$.[4] Denote the

[4]Whether to bootstrap the standardized QML residuals $\hat{e}_n$ or the original QML residuals $\hat{\epsilon}_n = \hat{\sigma}_n\hat{e}_n$ does not make a difference as R_{jn} are invariant of σ_0. However, use of $\hat{e}_n$ makes the theoretical discussion easier.

empirical distribution function (EDF) of the centred $\hat{e}_n$ by $\mathcal{F}_n$,

(3) Draw a random sample of size n from $\mathcal{F}_n$, and denote it by $e^*_{n,b}$,

(4) Compute $R_{in}(e^*_{n,b}, \hat{\rho}_n)$, $i = 1, \ldots, 4$, and hence $H_{in}(e^*_{n,b}, \hat{\rho}_n)$, $i = 1, 2, 3$ and $\zeta_n(e^*_{n,b}, \hat{\rho}_n)$,

(5) Repeat steps (3) and (4) B times, and the bootstrap estimates of $\mathrm{E}(H_{in}), i = 1, 2, 3$, and $\mathrm{E}(\zeta_n)$ are given by:

$$\hat{\mathrm{E}}(H_{in}) = \tfrac{1}{B}\textstyle\sum_{b=1}^{B} H_{in}(e^*_{n,b}, \hat{\rho}_n), \text{ and } \hat{\mathrm{E}}(\zeta_{in}) = \tfrac{1}{B}\textstyle\sum_{b=1}^{B} \zeta_{in}(e^*_{n,b}, \hat{\rho}_n). \quad (25)$$

The proposed bootstrap procedure overcomes the difficulty of analytically evaluating the expectations of very complicated quantities, and is very straightforward since in every bootstrap iteration, no re-estimation of the model parameters is required. The question that remains is its validity, particularly the validity of using $\widehat{C}_{2n}\widehat{\mathrm{E}}(\xi_n)$ in the third-order bias corrections $\widehat{C}_{3n}\widehat{\mathrm{E}}(\xi_n) = \widehat{C}_{2n}\widehat{\mathrm{E}}(\xi_n) + \widehat{c}_{3n}\widehat{\mathrm{E}}(\xi_n)$. We now elaborate using the quantities R_{jn}.

Let $\mathcal{F}_0$ be the CDF of $e_{n,i}$. The EDF $\mathcal{F}_n$ is thus an estimate of $\mathcal{F}_0$. If ρ_0 and $\mathcal{F}_0$ were known, then $\mathrm{E}[R_{jn}(e_n, \rho_0)] \doteq \frac{1}{M}\sum_{m=1}^{M} R_{jn}(e_{n,m}, \rho_0)$, where $e_{n,m}$ is a random sample of size n drawn from $\mathcal{F}_0$ and M is an arbitrarily large number. If ρ is unknown but $\mathcal{F}_0$ is known, $\mathrm{E}[R_{jn}(e_n, \rho_0)]$ can be estimated by $\frac{1}{M}\sum_{m=1}^{M} R_{jn}(e_{n,m}, \hat{\rho}_n)$, giving the so-called Monte Carlo (or parametric bootstrap) estimates of an expectation. In reality, however, both ρ_0 and $\mathcal{F}_0$ are unknown. Hence, this Monte Carlo method does not work. The bootstrap analogue of Model (3) takes the form,

$$Y^*_{n,b} = X_n\hat{\beta}_n + \hat{\sigma}_n A_n^{-1}(\hat{\rho}_n)e^*_{n,b},$$

where $(\hat{\beta}_n, \hat{\sigma}^2_n, \hat{\rho}_n)$ are now treated as bootstrap parameters. Based on the generated bootstrap data $(Y^*_{n,b}, W_n, X_n)$ and the bootstrap parameter $\hat{\rho}_n$, one computes R_{jn} defined by (16) and (20), to give bootstrap analogues of R_{jn}, which are $R_{jn}(e^*_n, \hat{\rho}_n), j = 1, \ldots, 4$. The bootstrap estimates of $\mathrm{E}[R_{jn}(e_n, \rho_0)]$ are thus,

$$\mathrm{E}^*[R_{jn}(e^*_n, \hat{\rho}_n)] \doteq \tfrac{1}{B}\textstyle\sum_{b=1}^{B} R_{jn}(e^*_{n,b}, \hat{\rho}_n), \text{ for a large } B,$$

which takes the same form as the Monte Carlo estimate with a known $\mathcal{F}_0$. This gives a heuristic justification on the validity of the bootstrap method.

Formally, denote the second- and third-order bias terms by $b_2(\rho_0, \gamma_0) = C'_{2n}\mathrm{E}(\zeta_n)$ and $b_3(\rho_0, \gamma_0) = c'_{3n}\mathrm{E}(\zeta_n)$, respectively, where $\gamma_0 = \gamma(\mathcal{F}_0)$ denotes the higher (than 2nd) order moments of $\mathcal{F}_0$ that b_2 and b_3 may depend upon. In our QML estimation framework, γ_0 is unknown as $\mathcal{F}_0$ is specified up to only the first two moments. Following the arguments above, the bootstrap estimates of b_2 and b_3 must take the form: $\hat{b}_2 = b_2(\hat{\rho}_n, \hat{\gamma}_n)$ and $\hat{b}_3 = b_3(\hat{\rho}_n, \hat{\gamma}_n)$ where $\hat{\gamma}_n = \gamma(\hat{\mathcal{F}}_n)$. The validity of the bootstrap estimates of bias corrections is thus established.

Proposition 3: *Under Assumptions of Proposition 2 and further, assuming a quantity bounded in probability has a finite expectation, then,*

$$\mathrm{E}[b_2(\hat{\rho}_n, \hat{\gamma}_n)] = b_2(\rho_0, \gamma_0) + O((\tfrac{h_n}{n})^2), \quad \textit{and} \quad \mathrm{E}[b_3(\hat{\rho}_n, \hat{\gamma}_n)] = b_3(\rho_0, \gamma_0) + o_p((\tfrac{h_n}{n})^2).$$

It follows that $\mathrm{E}(\hat{\rho}_n^{\mathrm{bc2}}) = \rho_0 + O((\frac{h_n}{n})^{\frac{3}{2}})$ *and* $\mathrm{E}(\hat{\rho}_n^{\mathrm{bc3}}) = \rho_0 + O((\frac{h_n}{n})^2)$.

4. An Alternative Model Specification

As mentioned in Section 2, an alternative to the SED model with an SAR error process is the SED model with a spatial moving average (SMA) error process,

$$Y_n = X_n\beta + u_n, \quad u_n = \epsilon_n - \rho W_n \epsilon_n, \tag{26}$$

where, all the quantities are defined in a similar manner as (1). The model at the true parameters can be written as $Y_n = X_n\beta_0 + A_n\epsilon_n$, giving, $\mathrm{Var}(u_n) = \sigma_0^2 A_n A_n'$, suggesting a similar non-spherical error structure. The quasi Gaussian log-likelihood function for this model is,

$$\ell_n(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) - \log|A_n(\rho)| - \frac{1}{2\sigma^2}(Y_n - X_n\beta)' A_n'^{-1}(\rho) A_n^{-1}(\rho)(Y_n - X_n\beta) \tag{27}$$

Given ρ, the constrained QMLEs are,

$$\begin{aligned} \hat{\beta}_n(\rho) &= (X_n' A_n'^{-1}(\rho) A_n^{-1}(\rho) X_n)^{-1} X_n' A_n'^{-1}(\rho) A_n^{-1}(\rho) Y_n, \text{ and} \\ \hat{\sigma}_n^2(\rho) &= \tfrac{1}{n} Y_n' A_n'^{-1}(\rho) M_n(\rho) A_n^{-1}(\rho) Y_n, \end{aligned}$$

where, $M_n(\rho) = I_n - A_n^{-1}(\rho)X_n[X_n' A_n'^{-1}(\rho) A_n^{-1}(\rho) X_n]^{-1} X_n' A_n'^{-1}(\rho)$. This results in the following concentrated log-likelihood function by substituting $\hat{\beta}_n(\rho)$ and $\hat{\sigma}_n^2(\rho)$ into (27),

$$\ell_n^c(\rho) = -\frac{n}{2}[\log(2\pi) + 1] - \log|A_n(\rho)| - \frac{n}{2}\log(\hat{\sigma}_n^2(\rho)). \tag{28}$$

The unconstrained QMLE $\hat{\rho}_n$ of ρ maximises $\ell_n^c(\rho)$, and the unconstrained QMLEs of β and σ^2 are given as $\hat{\beta}_n \equiv \hat{\beta}_n(\hat{\rho}_n)$ and $\hat{\sigma}_n^2 \equiv \hat{\sigma}_n^2(\hat{\rho}_n)$, respectively as in Section 2.

The QMLE $\hat{\rho}_n$ of the SMA error model is likely to perform poorer than that of the SAR error model, because the parameter space $\mathcal{P}$ for ρ stays the same, but $\hat{\rho}_n$ now becomes upward biased by comparing (28) with (7). Thus, when ρ is positive, 0.5 say, $\hat{\rho}_n$ may hit the upper bound of $\mathcal{P}$ when n is small, causing difficulty in estimating ρ.[5] Monte Carlo results given in Section 5 confirm this point. See also Martellosio (2010) for related discussions.

Asymptotic Distribution: Consistency and asymptotic normality of $\hat{\theta}_n$ can be proved in a similar manner as in the SED model with SAR errors, under a similar set of regularity conditions. In particular, the Assumption 3 has to be modified as: $\lim_{n\to\infty} \frac{1}{n} X_n' A'^{-1}(\rho) A^{-1}(\rho) X_n$ *exists and is non-singular uniformly in* ρ *in a neighbourhood of* ρ_0; and replace Assumption 6, the identification condition by: *For any* $\rho \neq \rho_0$, $\lim_{n\to\infty} \frac{h_n}{n}\left[\log|\sigma_0^2 A_n' A_n| - \log|\sigma_n^2(\rho) A_n'(\rho) A_n(\rho)|\right] \neq 0$, *where,* $\sigma_n^2(\rho) = \frac{\sigma_0^2}{n}\mathrm{tr}[A_n' A_n'^{-1}(\rho) A_n^{-1}(\rho) A_n]$.

[5] A more natural parameterization for the SMA error model may be $u_n = \epsilon_n + \rho W_n \epsilon_n$, under which $\mathcal{P}$ becomes a closed interval contained in $(-1, -\lambda_{\min}^{-1})$, but the QMLE $\hat{\rho}_n$ is now downward biased, and hence when ρ_0 is negative and n is small $\hat{\rho}_n$ may hit the lower bound of $\mathcal{P}$, causing the numerical instability of $(I_n + \hat{\rho}_n W_n)^{-1}$.

Theorem 3: *Under the modified Assumptions 1-6, we have,*

$$\sqrt{n}K_n^{-1}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0,\ \Sigma^{*-1}\Gamma^*\Sigma^{*-1}),$$

where, $\Gamma^* = \lim_{n\to\infty}\frac{1}{n}\Gamma_n^*$, $\Sigma^* = \lim_{n\to\infty}\frac{1}{n}\Sigma_n^*$, $\Gamma_n^* = K_n\Gamma_n K_n'$, $\Sigma_n^* = K_n\Sigma_n K_n'$,

$$\Gamma_n = \begin{pmatrix} \frac{1}{\sigma_0^2}X_n'A_n^{-1\prime}A_n^{-1}X_n & \frac{1}{2\sigma_0^3}\gamma X_n'A_n^{-1\prime}\iota_n & \frac{1}{\sigma_0}\gamma X_n'A_n^{-1\prime}g_n \\ \frac{1}{2\sigma_0^3}\gamma\iota_n'A_n^{-1}X_n & \frac{n}{4\sigma_0^4}(\kappa+2) & \frac{1}{2\sigma_0^2}(\kappa+2)\mathrm{tr}(G_n) \\ \frac{1}{\sigma_0}\gamma g_n'A_n^{-1}X_n & \frac{1}{2\sigma_0^2}(\kappa+2)\mathrm{tr}(G_n) & \kappa g_n'g_n + \mathrm{tr}(G_n^sG_n) \end{pmatrix},$$

$$\Sigma_n = \begin{pmatrix} \frac{1}{\sigma_0^2}X_n'A_n^{-1\prime}A_n^{-1}X_n & 0 & 0 \\ 0 & \frac{n}{2\sigma_0^4} & \frac{1}{\sigma_0^2}\mathrm{tr}(G_n) \\ 0 & \frac{1}{\sigma_0^2}\mathrm{tr}(G_n) & \mathrm{tr}(G_n^sG_n) \end{pmatrix},\ \textit{and } G_n = A_n^{-1}W_n.$$

Note that if the errors $\{\epsilon_{n,i}\}$ are normally distributed, then $\sqrt{n}K_n^{-1}(\hat{\theta}_n-\theta_0) \xrightarrow{D} N(0,\ \Sigma^{*-1})$. A similar set of results as in Corollary 1 can be obtained as well. Since the arguments for the proof Theorem 3 is very similar to that of Theorem 2, the explicit proof is omitted.

Finite-Sample Bias Correction. To simplify the exposition, we only present the necessary expressions for a second-order bias correction. The third-order results are available from the authors upon request. The derivatives of the *averaged* concentrated log-likelihood function $\frac{h_n}{n}\ell_n^c(\rho)$, up to a third-order, are:

$$\begin{aligned} \tilde{\psi}_n(\rho) &= h_nT_{0n}(\rho) - h_nR_{1n}(\rho), \\ h_n^{-1}H_{1n}(\rho) &= T_{1n}(\rho) - R_{2n}(\rho) + 2R_{1n}^2(\rho), \\ h_n^{-1}H_{2n}(\rho) &= 2T_{2n}(\rho) - R_{3n}(\rho) + 6R_{1n}(\rho)R_{2n}(\rho) - 8R_{1n}^3(\rho), \end{aligned}$$

where, $T_{rn}(\rho) = \frac{1}{n}\mathrm{tr}(G_n^{r+1}(\rho)), r = 0,1,2,$

$$R_{1n}(\rho) = \frac{Y_n'A_n'^{-1}(\rho)M_n(\rho)G_n(\rho)M_n(\rho)A_n^{-1}(\rho)Y_n}{Y_n'A_n'^{-1}(\rho)M_n(\rho)A_n^{-1}(\rho)Y_n}, \text{ and} \tag{29}$$

$$R_{jn}(\rho) = \frac{Y_n'A_n'^{-1}(\rho)M_n(\rho)D_{jn}(\rho)M_n(\rho)A_n^{-1}(\rho)Y_n}{Y_n'A_n'^{-1}(\rho)M_n(\rho)A_n^{-1}(\rho)Y_n},\ j = 2,3, \tag{30}$$

where, $D_{2n}(\rho)$ and $D_{3n}(\rho)$ are given in Appendix B.

Finally, with the clear definitions of the quantities $\tilde{\psi}_n(\rho)$, $h_n^{-1}H_{1n}(\rho)$ and $h_n^{-1}H_{2n}(\rho)$, the second-order bias correction of the QMLE $\hat{\rho}_n$ can be carried out using an identical bootstrap procedure as described in Section 3. The validity of the bootstrap procedure applied to this model can be proved in a similar manner. While the third-order bias correction can be carried out in the same manner, we found from the Monte Carlo experiments that the second-order bias corrections are more than satisfactory in all the cases considered.

Impact of bias correction. In connection with the discussion at the end of Section 2, we now offer some details on the impact of bias-correcting $\hat{\rho}_n$ on the subsequent inference for β in the form of testing $H_0 : c_0'\beta = 0$. The test statistic based on Corollary 1 is $t_n = c_0'\hat{\beta}_n/\sqrt{\hat{\sigma}_n^2 c_0'\hat{V}_n^{-1}c_0/n}$, where $\hat{V}_n = \frac{1}{n}X_n'A_n'(\hat{\rho}_n)A_n(\hat{\rho}_n)X_n = V_n - (\hat{\rho}_n - \rho_0)X_n'(W_n'A_n + A_n'W_n)X_n/n + (\hat{\rho}_n - \rho_0)^2 X_n'W_n'W_nX_n/n$. As $\hat{\rho}_n$ is downward biased, $\hat{V}_n$ tends to over estimate V_n, and hence $\hat{V}_n^{-1}$ tends to under estimate V_n^{-1}, causing t_n to be more variable and hence size distortions (over rejections). Our Monte Carlo results (unreported for brevity) show that simply replacing $\hat{\rho}_n$ in t_n by $\hat{\rho}_n^{\rm bc2}$ defined in (23) significantly reduces the size distortion. This shows that bias correction has a great potential for improving inferences for the regression coefficients. A formal study on this is interesting, but beyond the scope of this paper.

5. Simulation

The objective of the Monte Carlo simulations is to investigate the finite sample behaviour of $\hat{\rho}_n$ and the bias-corrected $\hat{\rho}_n$, under various spatial layouts, error distributions and the model parameters. The simulations are carried out based on the following data generation processes (DGP):

$$Y_n = \iota_n\beta_0 + X_{1n}\beta_1 + X_{2n}\beta_2 + u_n, \quad u_n = \rho W_n u_n + \epsilon_n,$$

where, ι_n is an $n\times 1$ vector of ones for the intercept term and X_{1n} and X_{2n} are the $n \times 1$ vectors containing the values of two fixed regressors. The parameters of the simulation are initially set to be as: $\beta = (5, 1, 1)'$, $\sigma^2 = 1$, ρ takes values form $\{-0.5, -0.25, 0, 0.25, 0.5\}$ and n take values from $\{50, 100, 200, 500\}$. Each set of Monte Carlo results is based on $M = 10,000$ Monte Carlo samples, and $B = 999 + \lfloor n^{0.75}\rfloor$ bootstrap samples within each Monte Carlo sample.

Spatial Weights Matrix: We use three different methods for generating the spatial weights matrix W_n: (*i*) `Rook Contiguity`, (*ii*) `Queen Contiguity`, and (*iii*) `Group Interaction`. The degree of spatial dependence specified by layouts in (*i*) and (*ii*) are fixed while in (*iii*) it grows with the increase in sample size. Specifically in (*iii*), W_n is block-diagonal, with k blocks (groups) of sizes $n_1, \ldots, n_k$. The rth block is an $n_r \times n_r$ matrix with off-diagonal elements $\frac{1}{n_r-1}$ and diagonal elements zero. In our Monte Carlo experiments, $k = \texttt{round}(n^\delta)$ with $\delta = .5$ or $.65$, and $\{n_r, r = 1, \ldots, k\}$ are k random draws from a discrete uniform distribution from $0.5m$ to $1.5m$ with $m = \texttt{round}(n/k)$. Clearly in this case the degree of spatial dependence, indicated by the average group size m, increases with n, and it is stronger when $\delta = .5$ than when $\delta = .65$. See Yang (2015) for a detailed description.

Regressors: The fixed regressors are generated by `REG1`: $\{x_{1i}, x_{2i}\} \overset{iid}{\sim} N(0, 1)/\sqrt{2}$ when Rook or Queen contiguity is followed; and according to either `REG1` or `REG2`: $\{x_{1,ir}, x_{2,ir}\} \overset{iid}{\sim} (2z_r + z_{ir})/\sqrt{10}$, where $(z_r, z_{ir}) \overset{iid}{\sim} N(0, 1)$ for $i = 1, \ldots, n_r$ and $r = 1, \ldots, k$, when group interaction scheme is followed. The `REG2` scheme gives non-iid regressors where the group means of the regressors' values are different, see Lee (2004). Note that both schemes give a signal-to-noise ratio of 1 when $\beta_1 = \beta_2 = \sigma = 1$.

Error Distribution: To generate $\epsilon_n = \sigma e_n$, three DGPs are considered: `DGP1:` $\{e_{n,i}\}$ are iid standard normal, `DGP2:` $\{e_{n,i}\}$ are iid standardized normal mixture with 10% of the values from $N(0,4)$ and the remaining from $N(0,1)$, and `DGP3:` $\{e_{n,i}\}$ iid standardized log-normal with parameters 0 and 1. Thus, the error distribution from `DGP2` is leptokurtic, and that of `DGP3` is both skewed and leptokurtic.

Partial Monte Carlo results are summarised in Tables 1-4, where in each table, the Monte Carlo means, root mean square errors (rmse) and the standard errors (se) of $\hat{\rho}_n$ and $\hat{\rho}_n^{bc2}$ are reported. The results for $\hat{\rho}_n^{bc3}$ are omitted as $\hat{\rho}_n^{bc2}$ provides satisfactory bias corrections for all the cases and the additional gain of using $\hat{\rho}_n^{bc3}$, although apparent, is quite marginal. Further, the case of queen contiguity (Table 2) is replicated by changing the β value to $(0.5, 0.1, 0.1)'$ (Table 5), and by changing the σ value to 3 (Table 6). We also give some partial results (Tables 7 and 8) for the SMA error model under the same set of parameters values set out at beginning of this section. It is useful to the note the following general characteristics of the results:

(i) $\hat{\rho}_n$ suffers from severe downward bias for almost all of the ρ values considered. The severity of the bias varies according to variations in (a) the sample size, (b) the spatial layout, and (c) the distribution of the errors considered.

(ii) $\hat{\rho}_n^{bc2}$ is almost unbiased in all the cases, even at considerably small sample sizes, which ascertains the effectiveness of the proposed bias correction procedure. These corrections can be attained without compromising the efficiency of the original QMLEs.

(iii) The spatial layout has a considerable impact on the finite sample performance of $\hat{\rho}_n$ in terms of the bias, rmse and se. A relatively sparse W_n, as in contiguity schemes, results in lower bias, rmse and se while a relatively dense W_n, as in group interaction scheme, results in the opposite.

(iv) The bias of the original QMLE seems to worsen as the error distribution deviates from normality. In contrast, $\hat{\rho}_n^{bc2}$ attains a similar level of accuracy in all the cases.

(v) The performance of $\hat{\rho}_n$ is not so sensitive to changes in the values of σ and β in terms of bias and the bias correction works well regardless of the true values set for the parameters.

(vi) The impact of the degree of spatial dependence on the rate of convergence is clearly revealed when comparing the results in Table 3 with those in Table 4 under the group interaction scheme. When the degree of spatial dependence is stronger as in the case where $k = n^{0.5}$, the rate of convergence is slower than in the case where $k = n^{0.65}$.

As expected, the magnitude of the bias, rmse and se are larger for small sample sizes. When considering the efficiency variations in terms of standard errors it can be seen that the efficiency of the estimators are sensitive to the sample size and the spatial layout. However, the different error distributions does not seem to have a significant effect on standard errors, reiterating the applicability of the proposed bias correction method in terms of robustness.

When the errors follow the SMA process, $u_n = (I_n - W_n)\epsilon_n$, the Monte Carlo results given in Tables 7 and 8 show that (i) the bias becomes positive, (ii) the QMLE $\hat{\rho}_n$ again can be severely biased, and (*iii*) the bias corrected $\hat{\rho}_n$ is almost unbiased. As discussed in Section 4, the Monte Carlo results indeed show that when ρ is positive (e.g., 0.5) and n is small (e.g., 50), $\hat{\rho}_n$ can be close to or can hit its upper bound, say 0.9999, causing numerical instability in calculating $A_n^{-1}(\hat{\rho}_n) = (I_n - \hat{\rho}_n W_n)^{-1}$, thus resulting in a poor performance of $\hat{\rho}_n$ and causing difficulty in bootstrapping the bias. This stands in contrast to the SED model with SAR errors where $\hat{\rho}_n$ is downward biased. However, with a larger $n(\geq 100)$, this problem disappears as seen from the results in Tables 7 and 8. Nevertheless, this does signal to a possible poor performance of the QMLE for an SMA error model when the sample size is not so large and the true spatial parameter value is positive and big.

Finally, compared to the Monte Carlo results presented in Yang (2015) for the SLD model, we see that the bias of $\hat{\rho}_n$ is more severe for the SED model, but does not spill over to $\hat{\beta}_n$ and $\hat{\sigma}_n^2$ that much.

6. Conclusions

This paper fills in some gaps in the literature by providing formal results for the asymptotic distribution as well as finite sample bias correction of the QMLEs for the spatial error dependence model. The primary concentration in the paper is a SED model with autoregressive errors of order 1. Comparable results for moving average errors of order 1 has been illustrated as well.

Consistency and the asymptotic normality of the QMLEs has been addressed with a specific attention given to the effect of the degree of spatial dependence on the rate of convergence of the QMLEs of the model parameters. Specifically when the degree spatial dependence, h_n, grows with the sample size n, the QMLE of the spatial parameter will have a lower rate of convergence (of $\sqrt{n/h_n}$) while the other QMLEs will have a $\sqrt{n}$-rate of convergence irrespective of the behaviour of h_n. Of the finite sample properties of spatial models, a specific attention has been given to the finite sample bias of the QMLE of the spatial parameter as it enters the model in a highly nonlinear manner and thus the estimation of it constitutes the main source of bias. Simulation studies indicate a prominent single direction bias in the estimation of the spatial parameter which in turn affects the subsequent inferences for the other model parameters. The severity of the bias increases as the spatial weights matrix becomes less sparse.

The finite sample results of this paper demonstrate again that stochastic expansions (Rilstone et al., 1996) coupled with bootstrap (Yang, 2015) provide a general and effective method for finite sample bias corrections of a nonlinear estimator. The suggested theories and methodologies are likely to be appealing to both theorists as well as practitioners alike who are dealing with the SED model or any other regression model that considers a spatial dependence structure in the error process (like SARAR, panel SARAR, spatial dynamic panel data models, etc). It would be interesting, as a future research, to address similar issues for these more complicated models. A formal study of the impacts of bias-correcting spatial/nonlinear estimators on the subsequent inferences for the regression coefficients is also in the agenda of our future research, in relation to a broader model of non-spherical errors.

Appendix A: Proofs of Asymptotic Results in Section 2:

The following lemmas are extended versions of some lemmas from Lee (2004) and Kelejian and Prucha (2001), which are needed in the proofs of the main results.

Lemma A.1: *Suppose the matrix of independent variables X_n has uniformly bounded elements and that the matrix A_n is defined s.t. Assumptions 3 and 5 are satisfied, then the projection matrices $M_n(\rho) = I_n - A_n(\rho)X_n[X_n' A_n'(\rho)A_n(\rho)X_n]^{-1}X_n' A_n'(\rho)$ and $P_n(\rho) = I_n - M_n(\rho)$ are uniformly bounded in both row and column sums, uniformly in $\rho \in \mathcal{P}$.*

Lemma A.2: *Let A_n be an $n \times n$ matrix, uniformly bounded in both row and column sums. Then for $M_n = M_n(\rho_0)$ defined in Lemma A.1,*
(*i*) $\mathrm{tr}(A_n^m) = O(n)$ *for* $m \geq 1$,
(*ii*) $\mathrm{tr}(A_n' A_n) = O(n)$,
(*iii*) $\mathrm{tr}((M_n A_n)^m) = \mathrm{tr}(A_n^m) + O(1)$ *for* $m \geq 1$ *and*
(*iv*) $\mathrm{tr}((A_n' M_n A_n)^m) = \mathrm{tr}((A_n' A_n)^m) + O(1)$ *for* $m \geq 1$.
Suppose further that B_n is an $n \times n$ matrix, uniformly bounded in both row and column sums, and C_n is a matrix s.t. the elements are of order $O(h_n^{-1})$, then,
(*iv*) *$A_n B_n$ is uniformly bounded in both row and column sums,*
(*v*) $A_n C_n = C_n A_n = O(h_n^{-1})$ *uniformly and*
(*vi*) $\mathrm{tr}(A_n C_n) = \mathrm{tr}(C_n A_n) = O(\frac{n}{h_n})$ *uniformly.*

Lemma A.3 (Moments and Limiting Distribution of Quadratic Forms): *Suppose the innovations $\{\epsilon_{ni}\}$ satisfy Assumption 2 and let γ and κ be respectively the measures of skewness and excess kurtosis of ϵ_{ni}. Further, let A_n be an $n \times n$ matrix with elements denoted by $a_{n,ij}$. Let, $Q_n = \epsilon_n' A_n \epsilon_n$, then,*
(*i*) $\mathrm{E}(Q_n) = \sigma_0^2 \mathrm{tr}(A_n)$ *and*
(*ii*) $\mathrm{Var}(Q_n) = \sigma_0^4[\mathrm{tr}(A_n' A_n + A_n^2) + \kappa \sum_{i=1}^n a_{n,ii}^2]$.
Now, if A_n is uniformly bounded either in row or column sums with the elements being of uniform order $O(\frac{1}{h_n})$, then,
(*iii*) $\mathrm{E}(Q_n) = O(\frac{n}{h_n})$,
(*iv*) $\mathrm{Var}(Q_n) = O(\frac{n}{h_n})$,
(*v*) $Q_n = O_p(\frac{n}{h_n})$,
(*vi*) $\frac{h_n}{n} Q_n - \frac{h_n}{n}\mathrm{E}(Q_n) = O_p\big((\frac{h_n}{n})^{\frac{1}{2}}\big) = o_p(1)$ *and*
(*vii*) $\mathrm{Var}(\frac{h_n}{n} Q_n) = O(\frac{h_n}{n}) = o(1)$.
Further, if the elements of A_n are uniformly bounded in both row and column sums and Assumption 4 is satisfied, then,
(*viii*) $\frac{Q_n - \mathrm{E}(Q_n)}{\mathrm{Var}(Q_n)} \xrightarrow{D} N(0,1)$.

Proof of Theorem 1: Following Theorem 3.4 of White (1994), it is sufficient to show that (*i*) the identification uniqueness condition: $\limsup_{n\to\infty} \max_{\rho \in \mathcal{N}_\epsilon^c(\rho_0)} \frac{h_n}{n}[\bar{\ell}_n^c(\rho) - \bar{\ell}_n^c(\rho_0)] < 0$ for any $\epsilon > 0$, where $\mathcal{N}_\epsilon^c(\rho_0)$ is the compliment of an open neighborhood of ρ_0 on $\mathcal{P}$ of radius ϵ, and (*ii*) the uniform convergence in probability: $\frac{h_n}{n}[\ell_n^c(\rho) - \bar{\ell}_n^c(\rho)] \xrightarrow{p} 0$ uniformly in $\rho \in \mathcal{P}$.

To show (i), first observing from (10) that $\sigma_n^2(\rho_0) = \sigma_0^2$, we have,

$$
\begin{aligned}
& \lim_{n\to\infty} \tfrac{h_n}{n}\big[\bar{\ell}_n^c(\rho) - \bar{\ell}_n^c(\rho_0)\big] \\
= \;& \lim_{n\to\infty} \big[\tfrac{h_n}{n}(\log|A_n(\rho)| - \log|A_n|) - \tfrac{h_n}{2}(\log\sigma_n^2(\rho) - \log\sigma_0^2)\big] \\
= \;& \lim_{n\to\infty} \big[\tfrac{h_n}{2n}(\log|A_n'(\rho)A_n(\rho)| - \log|A_n'A_n|) + \tfrac{h_n}{2n}(\log|\sigma_n^{-2}(\rho)I_n| - \log|\sigma_0^{-2}I_n|)\big] \\
\neq \;& 0 \text{ for } \rho \neq \rho_0, \text{ by Assumption 6.}
\end{aligned}
$$

Next, let $p_n(\theta) = \exp[\ell_n(\theta)]$ be the *quasi* joint pdf of $u_n(= Y_n - X_n\beta_0)$, and $p_n^0(\theta)$ the *true* joint pdf of u_n. Let E^q denote the expectation with respect to p_n, to differentiate from the usual notation E that corresponds to p_n^0. By Jensen's inequality (see Rao, 1973, p. 58), we have,

$$
0 = \log \mathrm{E}^q\Big(\frac{p_n(\theta)}{p_n(\theta_0)}\Big) \geq \mathrm{E}^q\Big[\log\Big(\frac{p_n(\theta)}{p_n(\theta_0)}\Big)\Big] = \mathrm{E}\Big[\log\Big(\frac{p_n(\theta)}{p_n(\theta_0)}\Big)\Big],
$$

where, the last equation follows from the fact that $\log p_n(\theta_0)$ and $\log p_n(\theta)$ are either a quadratic form or a linear-quadratic form of u_n, and hence their expectations w.r.t $p_n(\theta_0)$ are the same as those w.r.t. $p_n^0(\theta_0)$. It follows that $\mathrm{E}[\log p_n(\theta)] \leq \mathrm{E}[\log p_n(\theta_0)]$, and that,

$$
\bar{\ell}_n(\rho) = \max_{\beta,\sigma^2} \mathrm{E}[\log p_n(\theta)] \leq \mathrm{E}[\log p_n(\theta_0)] = \bar{\ell}_n(\rho_0), \quad \text{for } \rho \neq \rho_0.
$$

The identification uniqueness condition thus follows.

To show (ii), note that $\frac{h_n}{n}[\ell_n^c(\rho) - \bar{\ell}_n^c(\rho)] = -\frac{h_n}{2}[\log(\hat{\sigma}_n^2(\rho)) - \log(\sigma_n^2(\rho))]$. By the mean value theorem, $h_n[\log(\hat{\sigma}_n^2(\rho)) - \log(\sigma_n^2(\rho))] = \frac{h_n}{\tilde{\sigma}_n^2(\rho)}[\hat{\sigma}_n^2(\rho) - \sigma_n^2(\rho)]$ where $\tilde{\sigma}_n^2(\rho)$ lies between $\hat{\sigma}_n^2(\rho)$ and $\sigma_n^2(\rho)$. Note that,

$$
\begin{aligned}
\hat{\sigma}_n^2(\rho) &= \tfrac{1}{n}Y_n'A_n'(\rho)M_n(\rho)A_n(\rho)Y_n = \tfrac{1}{n}\epsilon_n'A_n'^{-1}A_n'(\rho)M_n(\rho)A_n(\rho)A_n^{-1}\epsilon_n \\
&= \tfrac{1}{n}\epsilon_n'A_n'^{-1}A_n'(\rho)A_n(\rho)A_n^{-1}\epsilon_n - \varDelta_n(\rho)
\end{aligned}
$$

where, $\varDelta_n(\rho) \equiv \frac{1}{n}\epsilon_n'A_n'^{-1}A_n'(\rho)P_n(\rho)A_n(\rho)A_n^{-1}\epsilon_n$.

By Assumption 3, $V_{1n}(\rho) \equiv \frac{1}{n}X_n'A_n'(\rho)A_n(\rho)X_n = O(1)$. In addition from Lemma A.2, $\frac{1}{n}\mathrm{tr}(W_nA_n^{-1}) \equiv \frac{1}{n}\mathrm{tr}(G_n) = O(\frac{1}{h_n})$ and using $A_n(\rho) = A_n + (\rho_0 - \rho)W_n$, we have,

$$
\begin{aligned}
\varDelta_n^*(\rho) &= \tfrac{1}{\sqrt{n}}X_n'A_n'(\rho)A_n(\rho)A_n^{-1}\epsilon_n \\
&= \tfrac{1}{\sqrt{n}}\big[X_n'A_n'\epsilon_n + (\rho_0 - \rho)X_n'(W_n' + A_n'G_n)\epsilon_n + (\rho_0 - \rho)^2X_n'W_n'G_n\epsilon_n\big] = O_p(\tfrac{1}{h_n}).
\end{aligned}
$$

Hence, $\varDelta_n(\rho) = \frac{1}{n}\varDelta_n^{*\prime}(\rho)V_{1n}^{-1}(\rho)\varDelta_n^*(\rho) = o_p(1)$, uniformly in $\rho \in \mathcal{P}$. It follows by Lemma A.3(vi) that, $h_n[\hat{\sigma}_n^2(\rho) - \sigma_n^2(\rho)] = \frac{h_n}{n}[\epsilon_n'A_n'^{-1}A_n'(\rho)A_n(\rho)A_n^{-1}\epsilon_n - \sigma_0^2\mathrm{tr}[A_n'^{-1}A_n'(\rho)A_n(\rho)A_n^{-1}] + o_p(1) = o_p(1)$, uniformly in $\rho \in \mathcal{P}$.

It left to show that $\sigma_n^2(\rho)$ is uniformly bounded away from zero, which is done by a counter argument. Suppose $\sigma_n^2(\rho)$ is not uniformly bounded away from zero in $\mathcal{P}$. Then there exists a sequence $\rho_n \in \mathcal{P}$ s.t. $\sigma_n^2(\rho_n) \to 0$ as $n \to \infty$. Consider a simpler model by setting β in (1) to 0. The Gaussian log-likelihood is $\ell_{t,n}(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) + \log|A_n(\rho)| - \frac{1}{2\sigma^2}Y_n'A_n'(\rho)A_n(\rho)Y_n$. Then

$\bar{\ell}_{t,n}(\rho) = \max_{\sigma^2} \mathrm{E}[\ell_{t,n}(\theta)] = -\frac{n}{2}[\log(2\pi)+1] - \frac{n}{2}\log(\sigma_n^2(\rho)) + \log|A_n(\rho)|$. By Jensen's inequality, $\bar{\ell}_{t,n}(\theta) \leq \mathrm{E}[\ell_{t,n}(\theta_0)] = \bar{\ell}_{t,n}(\rho_0), \forall\rho$. This implies $\frac{1}{n}[\bar{\ell}_{t,n}(\theta) - \bar{\ell}_{t,n}(\theta_0)] \leq 0$ and $-\frac{1}{2}\log(\sigma_n^2(\rho)) \leq -\frac{1}{2}\log(\sigma_0^2) + \frac{1}{n}(\log|A_n(\rho_0)| - \log|A_n(\rho)|) = O(1)$ using the Lemma A.2, that is, $-\log(\sigma_n^2(\rho))$ is bounded from above which is a contradiction. Hence, $\sigma_n^2(\rho)$ is bounded away from zero uniformly in $\rho \in \mathcal{P}$, and $\log(\sigma_n^2(\rho))$ is well defined $\forall\rho \in \mathcal{P}$.

Since $\sigma_n^2(\rho)$ is bounded away from zero and $h_n[\hat{\sigma}_n^2(\rho) - \sigma_n^2(\rho)] = o_p(1)$, $\hat{\sigma}_n^2(\rho)$ is bounded away from zero uniformly in probability in $\mathcal{P}$ as well. Collecting all these results together along with the mean value theorem, we have $h_n|\log(\hat{\sigma}_n^2(\rho)) - \log(\sigma_n^2(\rho))| = o_p(1)$ uniformly in $\rho \in \mathcal{P}$. Hence $\sup_{\rho\in\mathcal{P}} \frac{h_n}{n}|[\ell_n^c(\rho) - \bar{\ell}_n^c(\rho)]| = o_p(1)$.

Proof of Theorem 2: By applying the mean value theorem on the modified first order condition, we have,

$$\begin{aligned} 0 = \tfrac{1}{\sqrt{n}} S_n^*(\hat{\theta}_n) &= \tfrac{1}{\sqrt{n}} S_n^*(\theta_0) + \tfrac{1}{\sqrt{n}} \tfrac{\partial}{\partial\theta'} S_n^*(\tilde{\theta}_n)(\hat{\theta}_n - \theta_0) \\ &= \tfrac{1}{\sqrt{n}} S_n^*(\theta_0) - \tfrac{1}{n} K_n H_n(\tilde{\theta}_n) K_n . \sqrt{n} K_n^{-1}(\hat{\theta}_n - \theta_0) \end{aligned} \tag{A-1}$$

where $\tilde{\theta}_n$ lies between the line segment joining θ_0 and $\hat{\theta}_n$, thus $\tilde{\theta} \xrightarrow{p} \theta_0$. Here $H_n(\theta)$ is the negative Hessian matrix and K_n is as defined in section 2.2.

Under Assumptions 1-5, the central limit theorem for linear-quadratic forms of Kelejian and Prucha (2001) is applicable, which gives $\frac{1}{\sqrt{n}} S_n^*(\theta_0) = \frac{K_n}{\sqrt{n}} \frac{\partial}{\partial\theta} \ell(\theta_0) \xrightarrow{D} N(0, \Gamma^*)$, where, $\Gamma^* = \lim_{n\to\infty} \frac{1}{n}\Gamma_n^*$ and $\Gamma_n^* = \mathrm{Var}[S_n^*(\theta_0)]$. The asymptotic normality of $\hat{\theta}_n$ thus follows from: (i) $\frac{1}{n} K_n H_n(\tilde{\theta}_n) K_n - \frac{1}{n} K_n H_n(\theta_0) K_n = o_p(1)$ and (ii) $\frac{1}{n} K_n H_n(\theta_0) K_n - \frac{1}{n} K_n \Sigma_n K_n = o_p(1)$, where, $\Sigma_n = \mathrm{E}[H_n(\theta_0)]$ is the information matrix given in section 2.2. To show (i), note that $H_n(\theta) =$

$$\begin{pmatrix} \frac{1}{\sigma^2} X_n' A_n'(\rho) A_n(\rho) X_n & \frac{1}{\sigma^4} X_n' A_n'(\rho) \epsilon_n(\delta) & \frac{2}{\sigma^2} X_n' A_n'(\rho) G_n'(\rho) \epsilon_n(\delta) \\ \frac{1}{\sigma^4} \epsilon_n'(\delta) A_n(\rho) X_n & \frac{1}{2\sigma^6}(2\epsilon_n'(\delta)\epsilon_n(\delta) - n\sigma^2) & \frac{1}{\sigma^4} \epsilon_n'(\delta) G_n'(\rho) \epsilon_n(\delta) \\ \frac{2}{\sigma^2} \epsilon_n'(\delta) G_n(\rho) A_n(\rho) X_n & \frac{1}{\sigma^4} \epsilon_n'(\delta) G_n(\rho) \epsilon_n(\delta) & \frac{1}{\sigma^2}[\epsilon_n'(\delta) G_n'(\rho) G_n(\rho) \epsilon_n(\delta) + \sigma^2 \mathrm{tr}(G_n^2(\rho))] \end{pmatrix}$$

where $\delta = (\beta', \rho)'$. Let $\tilde{A}_n = A_n(\tilde{\rho}_n)$. Under Assumption 3 and using $\tilde{\theta}_n \xrightarrow{p} \theta_0$, we have,

$$\begin{aligned} \tfrac{1}{n}\big(\tfrac{\partial^2}{\partial\beta\partial\beta'} \ell_n(\tilde{\theta}_n) - \tfrac{\partial^2}{\partial\beta\partial\beta'} \ell_n(\theta_0)\big) &= \tfrac{1}{n}\big(\tfrac{1}{\sigma_0^2} X_n' A_n' A_n X_n - \tfrac{1}{\tilde{\sigma}_n^2} X_n' \tilde{A}_n' \tilde{A}_n X_n\big) \\ &= \big(\tfrac{1}{\sigma_0^2} - \tfrac{1}{\tilde{\sigma}_n^2}\big) \tfrac{1}{n} X_n' A_n' A_n X_n + o_p(1) = o_p(1), \end{aligned}$$

noticing that $A_n' A_n - \tilde{A}_n' \tilde{A}_n = (\tilde{\rho}_n - \rho_0)(W_n + W_n') - (\tilde{\rho}_n^2 - \rho_0^2) W_n' W_n$.

Similarly, it can be shown that, letting $\tilde{\epsilon}_n = \epsilon_n(\tilde{\rho}_n)$,

$$\begin{aligned} \tfrac{1}{n}\big(\tfrac{\partial^2}{\partial(\sigma^2)^2} \ell_n(\tilde{\theta}_n) - \tfrac{\partial^2}{\partial(\sigma^2)^2} \ell_n(\theta_0)\big) &= \tfrac{1}{n\sigma_0^6} \epsilon_n' \epsilon_n - \tfrac{1}{n\tilde{\sigma}_n^6} \tilde{\epsilon}_n' \tilde{\epsilon}_n - \tfrac{1}{2}\big(\tfrac{1}{\sigma_0^4} - \tfrac{1}{\tilde{\sigma}_n^4}\big) \\ &= \tfrac{1}{n\sigma_0^6}(\epsilon_n' \epsilon_n - \tilde{\epsilon}_n' \tilde{\epsilon}_n) + o_p(1) = o_p(1), \end{aligned}$$

since $\tilde{\epsilon}_n' \tilde{\epsilon}_n - \epsilon_n' \epsilon_n = 2(\rho_0 - \tilde{\rho}_n) \epsilon_n' G_n \epsilon_n + 2\epsilon_n' A_n X_n(\beta_0 - \tilde{\beta}_0) + (\rho_0 - \tilde{\rho}_0)^2 \epsilon_n' G_n' G_n \epsilon_n + 2(\rho_0 - \tilde{\rho}_n) \epsilon_n' W_n X_n(\beta_0 - \tilde{\beta}_n) + 2(\rho_0 - \tilde{\rho}_n) \epsilon_n' G_n' A_n X_n(\beta_0 - \tilde{\beta}_n) + (\beta_0 - \tilde{\beta}_n)' X_n' A_n' A_n X_n(\beta_0 - \tilde{\beta}_n) +$

$2(\rho_0 - \tilde{\rho}_n)^2 \epsilon_n' G_n' W_n X_n(\beta_0 - \tilde{\beta}_n) + 2(\rho_0 - \tilde{\rho}_n)(\beta - \tilde{\beta}_n)' X_n' A_n' W_n X_n(\beta_0 - \tilde{\beta}_n) + (\rho_0 - \tilde{\rho}_n)^2(\beta_0 - \tilde{\beta}_n)' X_n' W_n' W_n X_n(\beta_0 - \tilde{\beta}_n) = o_p(1)$.

Now by the mean value theorem, $\mathrm{tr}(G_n^2(\tilde{\rho}_n)) = \mathrm{tr}(G_n^2) + 2\mathrm{tr}[G_n^3(\bar{\rho}_n)](\tilde{\rho}_n - \rho_0)$, where $\bar{\rho}_n$ lies between ρ_0 and $\tilde{\rho}_n$. By Lemma A.2, and Assumptions 4 and 5, $\mathrm{tr}[G_n^3(\bar{\rho}_n)] = O\big(\frac{n}{h_n}\big)$. Hence, $\frac{h_n}{n}[\mathrm{tr}(G_n^2(\tilde{\rho}_n)) - \mathrm{tr}(G_n^2)] = o_p(1)$ since $\tilde{\rho}_n \xrightarrow{p} \rho_0$.

Further, $\epsilon_n' G_n' G_n \epsilon_n = Y_n' W_n' W_n Y_n - 2Y_n' W_n' W_n X_n \beta_0 + \beta_0' X_n' W_n' W_n X_n \beta_0 = O_p\big(\frac{n}{h_n}\big)$ by Lemmas A.2(i) and A.3(v). Hence, $\frac{h_n}{n}[\tilde{\epsilon}_n' \tilde{G}_n' \tilde{G}_n \tilde{\epsilon}_n - \epsilon_n' G_n' G_n \epsilon_n] = \frac{h_n}{n}[(\beta_0 - \tilde{\beta}_n)' X_n' W_n' W_n X_n(\beta_0 - \tilde{\beta}_n) - 2\epsilon_n' G_n' W_n X_n(\beta_0 - \tilde{\beta}_n)] = o_p(1)$, hence,

$$
\begin{aligned}
\tfrac{h_n}{n}\big(\tfrac{\partial^2}{\partial\rho^2}\ell_n(\tilde{\theta}_n) - \tfrac{\partial^2}{\partial\rho^2}\ell_n(\theta_0)\big) &= \tfrac{h_n}{n}\big(\tfrac{1}{\sigma_0^2}\epsilon_n' G_n' G_n \epsilon_n - \tfrac{1}{\tilde{\sigma}_n^2}\tilde{\epsilon}_n' \tilde{G}_n' \tilde{G}_n \tilde{\epsilon}_n + \mathrm{tr}(G_n^2) - \mathrm{tr}(\tilde{G}_n^2)\big) \\
&= \tfrac{h_n}{n}\big(\tfrac{1}{\sigma_0^2} - \tfrac{1}{\tilde{\sigma}_n^2}\big)\epsilon_n' G_n' G_n \epsilon_n + o_p(1) = o_p(1).
\end{aligned}
$$

Using similar arguments, the convergence in probability to zero of the rest of the terms in the modified Hessian can be shown:

$$
\begin{aligned}
\tfrac{\sqrt{h_n}}{n}\big(\tfrac{\partial^2}{\partial\beta\partial\rho}\ell_n(\tilde{\theta}_n) - \tfrac{\partial^2}{\partial\beta\partial\rho}\ell_n(\theta_0)\big) &= \tfrac{2\sqrt{h_n}}{n\sigma_0^2}(X_n' W_n' \epsilon_n - X_n' W_n' \tilde{\epsilon}_n) + o_p(1) = o_p(1), \\
\tfrac{1}{n}\big(\tfrac{\partial^2}{\partial\beta\partial\sigma^2}\ell_n(\tilde{\theta}_n) - \tfrac{\partial^2}{\partial\beta\partial\sigma^2}\ell_n(\theta_0)\big) &= \tfrac{1}{n\sigma_0^4}[(X_n' A_n' \epsilon_n) - (X_n' \tilde{A}_n' \tilde{\epsilon}_n)] + o_p(1) = o_p(1), \quad \text{and} \\
\tfrac{\sqrt{h_n}}{n}\big(\tfrac{\partial^2}{\partial\sigma^2\partial\rho}\ell_n(\tilde{\theta}_n) - \tfrac{\partial^2}{\partial\sigma^2\partial\rho}\ell_n(\theta_0)\big) &= \tfrac{\sqrt{h_n}}{n\sigma^4}(\epsilon_n' G_n' \epsilon_n - \tilde{\epsilon}_n' \tilde{G}_n' \tilde{\epsilon}_n) + o_p(1) \\
&= \tfrac{\sqrt{h_n}}{n\sigma^4}[\epsilon_n' W_n(Y_n - X_n\beta_n) - \tilde{\epsilon}_n' W_n(Y_n - X_n\tilde{\beta}_n)] + o_p(1) \\
&= \tfrac{\sqrt{h_n}}{n\sigma^4}[(\epsilon_n' - \tilde{\epsilon}_n')W_n Y_n - \epsilon_n' W_n X_n \beta_n + \tilde{\epsilon}_n' W_n X_n \tilde{\beta}_n] + o_p(1) \\
&= o_p(1).
\end{aligned}
$$

Proof of (ii) is more straightforward, as the differences of the corresponding elements of $\frac{1}{n}K_n H_n(\theta_0) K_n$ and $\frac{1}{n}K_n \Sigma_n K_n$ are, respectively, 0, $\frac{1}{n\sigma^4}(X_n' A_n' \epsilon_n) = o_p(1)$, $\frac{1}{2n\sigma^6}(2\epsilon_n'\epsilon_n - n\sigma^2) - \frac{1}{2\sigma_0^4} = \frac{1}{n\sigma^6}\epsilon_n'\epsilon_n = o_p(1)$, $\frac{2\sqrt{h_n}}{n\sigma_0^2}X_n' A_n' G_n' = o_p(1)$, $\frac{\sqrt{h_n}}{n\sigma^4}\epsilon_n' G_n \epsilon_n - \frac{\sqrt{h_n}}{n\sigma_0^2}\mathrm{tr}(G_n) = o_p(1)$, and $\frac{h_n}{n\sigma_0^2}(\epsilon_n' G_n' G_n \epsilon_n + \sigma^2 \mathrm{tr}(G_n^2)) - \frac{h_n}{n}\mathrm{tr}(G_n^s G_n) = \frac{h_n}{n\sigma_0^2}\epsilon_n' G_n' G_n \epsilon_n = o_p(1)$.

The results (i) and (ii) give $0 = \frac{1}{\sqrt{n}}S_n^* - \frac{1}{n}\Sigma_n^* \cdot \sqrt{n}K_n^{-1}(\hat{\theta}_n - \theta_0) + o_p(1)$, and it follows that,

$$
\sqrt{n}K_n^{-1}(\hat{\theta}_n - \theta_0) = \Sigma_n^{*-1} S_n^* \xrightarrow{D} N(0,\ \Sigma^{*-1}\Gamma^*\Sigma^{*-1}).
$$

Proof of Corollary 1: By using the block diagonal nature of Σ_n,

$$
\Sigma_n^{-1} = \begin{pmatrix} \sigma_0^2(X_n' A_n' A_n X_n)^{-1} & 0 & 0 \\ 0 & \frac{2\sigma_0^4}{n}T_{1n} & -\frac{2\sigma_0^2}{n}T_{2n} \\ 0 & -\frac{2\sigma_0^2}{n}T_{2n} & \frac{h_n}{n}T_{4n} \end{pmatrix}
$$

where, $T_{1n} = \frac{\mathrm{tr}(G_n^s G_n)}{\mathrm{tr}(C_n^s C_n)}$, $T_{2n} = \frac{\mathrm{tr}(G_n)}{\mathrm{tr}(C_n^s C_n)}$, $T_{4n} = \frac{n}{h_n}\mathrm{tr}^{-1}(C_n^s C_n)$. Then deriving $\Sigma_n^{*-1}\Gamma_n^*\Sigma_n^{*-1} = K_n^{-1}\Sigma_n^{-1}\Gamma_n\Sigma_n^{-1}K_n^{-1}$ is just a matter of matrix multiplication.

Appendix B: Proofs of Higher-Order Results in Section 3

We prove the higher-order results given in Section 3. First, we present the full expressions for $D_{jn}(\rho), j=2,3,4$, which are required in the expressions for $R_{jn}(\rho)$ given in (20):

$$\begin{aligned}
D_{2n}(\rho) &= G'_n(\rho)M_n(\rho)G_n(\rho) - 2G_n(\rho)P_n(\rho)G_n(\rho) - G_n(\rho)P_n(\rho)G'_n(\rho),\\
D_{3n}(\rho) &= \dot{D}_{2n}(\rho) + G_n(\rho)P_n(\rho)D_{2n}(\rho) + D_{2n}(\rho)P_n(\rho)G'_n(\rho)\\
&\quad -G'_n(\rho)M_n(\rho)D_{2n}(\rho) - D_{2n}(\rho)M_n(\rho)G_n(\rho),\\
D_{4n}(\rho) &= \dot{D}_{3n}(\rho) + G_n(\rho)P_n(\rho)D_{3n}(\rho) + D_{3n}(\rho)P_n(\rho)G'_n(\rho)\\
&\quad -G'_n(\rho)M_n(\rho)D_{3n}(\rho) - D_{3n}(\rho)M_n(\rho)G_n(\rho),
\end{aligned}$$

where $P_n(\rho) = I_n - M_n(\rho)$ and $\dot{D}_{jn}(\rho) = \frac{d}{d\rho}D_{jn}(\rho), j=2,3$. Note that a predictable pattern emerges from $D_{3n}(\rho)$ onwards. Using the fact that $\frac{d}{d\rho}G_n^i = G_n^{i+1}$ for $i=1,2,\ldots$, we have,

$$\begin{aligned}
\dot{D}_{2n}(\rho) &= G_n'^2(\rho)M_n(\rho)G_n(\rho) + G'_n(\rho)\dot{M}_n(\rho)G_n(\rho) + G'_n(\rho)M_n(\rho)G_n^2(\rho)\\
&\quad -2G_n^2(\rho)P_n(\rho)G_n(\rho) + 2G_n(\rho)\dot{M}_n(\rho)G_n(\rho) - 2G_n(\rho)P_n(\rho)G_n^2(\rho)\\
&\quad -G_n^2(\rho)P_n(\rho)G'_n(\rho) + G_n(\rho)\dot{M}_n(\rho)G'_n(\rho) - G_n(\rho)P_n(\rho)G_n'^2(\rho),\\
\dot{M}_n(\rho) &= P_n(\rho)G'_n(\rho)M_n(\rho) + M_n(\rho)G_n(\rho)P_n(\rho),\\
\dot{D}_{3n}(\rho) &= G_n'^3(\rho)M_n(\rho)G_n(\rho) + 2G_n'^2(\rho)\dot{M}_n(\rho)G_n(\rho) + 2G_n'^2(\rho)M_n(\rho)G_n^2(\rho)\\
&\quad +G'_n(\rho)\ddot{M}_n(\rho)G_n(\rho) + 2G'_n(\rho)\dot{M}_n(\rho)G_n^2(\rho) + G'_n(\rho)M_n(\rho)G_n^3(\rho)\\
&\quad -2G_n^3(\rho)P_n(\rho)G_n(\rho) + 4G_n^2(\rho)\dot{M}_n(\rho)G_n(\rho) - 4G_n^2(\rho)P_n(\rho)G_n^2(\rho)\\
&\quad +2G_n(\rho)\ddot{M}_n(\rho)G_n(\rho) + 4G_n(\rho)\dot{M}_n(\rho)G_n^2(\rho) - 2G_n(\rho)P_n(\rho)G_n^3(\rho)\\
&\quad -G_n^3(\rho)P_n(\rho)G'_n(\rho) + 2G^2(\rho)\dot{M}_n(\rho)G'_n(\rho) - 2G_n^2(\rho)P_n(\rho)G_n'^2(\rho)\\
&\quad +G_n(\rho)\ddot{M}_n(\rho)G'_n(\rho) + 2G_n(\rho)\dot{M}_n(\rho)G_n'^2(\rho) - G_n(\rho)P_n(\rho)G_n'^3(\rho),
\end{aligned}$$

$$\begin{aligned}
\ddot{M}_n(\rho) &= 2P_n(\rho)G'_n(\rho)P_n(\rho)G'_n(\rho)M_n(\rho) + 2P_n(\rho)G'_n(\rho)M_n(\rho)G_n(\rho)P_n(\rho)\\
&\quad +2M_n(\rho)G_n(\rho)P_n(\rho)G_n(\rho)P_n(\rho) - 2M_n(\rho)G_n(\rho)P_n(\rho)G'_n(\rho)M_n(\rho).
\end{aligned}$$

For the SED model with SMA errors, the additional quantities required by (30) are,

$$\begin{aligned}
D_{2n}(\rho) &= G'_n(\rho)M_n(\rho)G_n(\rho) + 2G_n(\rho)M_n(\rho)G_n(\rho) - G_n(\rho)P_n(\rho)G'_n(\rho),\\
D_{3n}(\rho) &= \dot{D}_{2n}(\rho) - G_n(\rho)P_n(\rho)D_{2n}(\rho) - D_{2n}(\rho)P_n(\rho)G'_n(\rho)\\
&\quad +G'_n(\rho)M_n(\rho)D_{2n}(\rho) + D_{2n}(\rho)M_n(\rho)G_n(\rho),\\
\dot{D}_{2n}(\rho) &= G_n'^2(\rho)M_n(\rho)G_n(\rho) + G'_n(\rho)\dot{M}_n(\rho)G_n(\rho) + G'_n(\rho)M_n(\rho)G_n^2(\rho)\\
&\quad +2G_n^2(\rho)M_n(\rho)G_n(\rho) + 2G_n(\rho)\dot{M}_n(\rho)G_n(\rho) + 2G_n(\rho)M_n(\rho)G_n^2(\rho)\\
&\quad -G_n^2(\rho)P_n(\rho)G'_n(\rho) + G_n(\rho)\dot{M}_n(\rho)G'_n(\rho) - G_n(\rho)P_n(\rho)G_n'^2(\rho),\\
\dot{M}_n(\rho) &= -P_n(\rho)G'_n(\rho)M_n(\rho) - M_n(\rho)G_n(\rho)P_n(\rho), \text{ and } P_n = I_n - M_n.
\end{aligned}$$

Proof of Lemma 1: Note, $\hat{\sigma}_n^2(\rho_0) \equiv \hat{\sigma}_{n0}^2 = \frac{1}{n}Y_n'A_n'M_nA_nY_n = \frac{1}{n}\epsilon_n'M_n\epsilon_n$. By the moments for quadratic forms, we have, $\mathrm{Var}(\hat{\sigma}_{n0}^2) = \frac{1}{n^2}O(n) = O(\frac{1}{n})$. Now by the generalised Chebyshev's inequality, $\mathrm{P}(\sqrt{n}|\hat{\sigma}_{n0}^2 - \sigma_0^2| \geq \delta) \leq \frac{1}{\delta^2}n\mathrm{Var}(\hat{\sigma}_{n0}^2) = O(1)$. Hence, by the definition of order of magnitudes[6] for stochastic components we have, $\hat{\sigma}_{n0}^2 = \sigma_0^2 + O_p(\frac{1}{\sqrt{n}})$.

In order to prove that $\hat{\sigma}_{n0}^{-2}$ is $\sqrt{n}$-consistent, by the Mean Value Theorem, we have, $\frac{1}{\hat{\sigma}_{n0}^2} - \frac{1}{\sigma_0^2} = -\frac{1}{\bar{\sigma}_{n0}^4}(\hat{\sigma}_{n0}^2 - \sigma_0^2)$, which can be written as, $\frac{1}{\hat{\sigma}_{n0}^2} = \frac{1}{\sigma_0^2} - \frac{1}{\sigma_0^4}(\hat{\sigma}_{n0}^2 - \sigma_0^2) - \big(\frac{1}{\bar{\sigma}_{n0}^4} - \frac{1}{\sigma_0^4}\big)(\hat{\sigma}_{n0}^2 - \sigma_0^2)$, where $\bar{\sigma}^2$ lies between $\hat{\sigma}_{n0}^2$ and σ_0^2. Hence, $\bar{\sigma}_{n0}^2 = \sigma_0^2 + O_p(\frac{1}{\sqrt{n}})$, $\bar{\sigma}_{n0}^4 = \big(\sigma_0^2 + O_p(\frac{1}{\sqrt{n}})\big)^2 = \sigma_0^4 + O_p(\frac{1}{\sqrt{n}})$, and $\bar{\sigma}_{n0}^{-4} = \big(\sigma_0^4 + O_p(\frac{1}{\sqrt{n}})\big)^{-1} = \sigma_0^{-4} + O_p(\frac{1}{\sqrt{n}})$. Therefore, we conclude that $\hat{\sigma}_{n0}^{-2} = \sigma_0^{-2} + O_p(\frac{1}{\sqrt{n}})$.

Now consider, $h_nR_{1n} = \frac{h_n}{n\hat{\sigma}_{n0}^2}\epsilon_n'M_nG_nM_n\epsilon_n$. By Lemma A.3(v), $\frac{h_n}{n}\epsilon_n'M_nG_nM_n\epsilon_n = O_p(1)$. Hence,

$$h_nR_{1n} = \frac{1}{\sigma_0^2}\frac{h_n}{n}\epsilon_n'M_nG_nM_n\epsilon_n + O_p(\tfrac{1}{\sqrt{n}}) = O_p(1). \tag{B-1}$$

Using the expression for $\hat{\sigma}_{n0}^{-2}$, $\mathrm{E}(h_nR_{1n}) = \frac{1}{\sigma_0^2}\mathrm{E}\big(\frac{h_n}{n}\epsilon_n'M_nG_nM_n\epsilon_n\big) - \frac{1}{\sigma_0^4}\mathrm{E}\big(\frac{h_n}{n}\epsilon_n'M_nG_nM_n\epsilon_n(\hat{\sigma}_{n0}^2 - \sigma_0^2)\big) - \mathrm{E}\big(\frac{h_n}{n}\epsilon_n'M_nG_nM_n\epsilon_n\big(\frac{1}{\bar{\sigma}_{n0}^4} - \frac{1}{\sigma_0^4}\big)(\hat{\sigma}_{n0}^2 - \sigma_0^2)\big)$. The first term is, $\frac{h_n}{\sigma_0^2 n}\mathrm{E}(\epsilon_n'\epsilon_n)\mathrm{tr}(M_nG_nM_n) = O(1)$. The third term is, $O((\frac{h_n}{n})^{\frac{1}{2}})$ by Assumption 7. For the second term note that, $\mathrm{E}(\hat{\sigma}_{n0}^2) = \sigma_0^2 + O(\frac{1}{n})$ and $\mathrm{E}(\epsilon_n'M_nG_nM_n\epsilon_n) = \sigma_0^2\mathrm{tr}(M_nG_nM_n) = O(\frac{n}{h_n})$. Then by Cauchy-Schwartz inequality,

$$\begin{aligned}
& |\mathrm{E}\big(\epsilon_n'M_nG_nM_n\epsilon_n(\hat{\sigma}_{n0}^2 - \sigma_0^2)\big)| \\
= \; & |\mathrm{E}\big([\epsilon_n'M_nG_nM_n\epsilon_n - \mathrm{E}(\epsilon_n'M_nG_nM_n\epsilon_n) + \mathrm{E}(\epsilon_n'M_nG_nM_n\epsilon_n)](\hat{\sigma}_{n0}^2 - \sigma_0^2)\big)| \\
\leq \; & |\mathrm{E}\big([\epsilon_n'M_nG_nM_n\epsilon_n - \sigma_0^2\mathrm{tr}(M_nG_nM_n)](\hat{\sigma}_{n0}^2 - \sigma_0^2)\big)| + \sigma_0^2|\mathrm{tr}(M_nG_nM_n)\mathrm{E}\big(\hat{\sigma}_{n0}^2 - \sigma_0^2\big)| \\
= \; & |\mathrm{Cov}\big([\epsilon_n'M_nG_nM_n\epsilon_n - \sigma_0^2\mathrm{tr}(M_nG_nM_n)], (\hat{\sigma}_{n0}^2 - \sigma_0^2)\big)| + O(\tfrac{1}{h_n}) \\
\leq \; & \tfrac{1}{n}\big(\mathrm{Var}(\epsilon_n'M_nG_nM_n\epsilon_n)\mathrm{Var}(\epsilon_n'M_n\epsilon)\big)^{\frac{1}{2}} + O(\tfrac{1}{h_n}) = \tfrac{1}{n}\big(O(\tfrac{n}{h_n})O(n)\big)^{\frac{1}{2}} + O(\tfrac{1}{h_n}) = O(\tfrac{1}{\sqrt{h_n}}),
\end{aligned}$$

where we have used the results for quadratic forms. Then, $\frac{1}{\sigma_0^4}\mathrm{E}\big[\frac{h_n}{n}\epsilon_n'M_nG_nM_n\epsilon_n(\hat{\sigma}_{n0}^2 - \sigma_0^2)\big] = O(\frac{\sqrt{h_n}}{n})$, which implies,

$$\mathrm{E}(h_nR_{1n}) = \mathrm{Max}\big\{O(1), O\big(\tfrac{\sqrt{h_n}}{n}\big), O\big((\tfrac{h_n}{n})^{\frac{1}{2}}\big)\big\} = O(1). \tag{B-2}$$

By (B-1) and (B-2), $h_nR_{1n} - \mathrm{E}(h_nR_{1n}) = \frac{h_n}{\sigma_0^2 n}\epsilon_n'M_nG_nM_n\epsilon_n - \frac{h_n}{\sigma_0^2 n}\mathrm{E}(\epsilon_n'\epsilon_n)\mathrm{tr}(M_nG_nM_n) + O_p(\frac{1}{\sqrt{n}}) - O(\frac{\sqrt{h_n}}{n}) - O((\frac{h_n}{n})^{\frac{1}{2}}) = O((\frac{h_n}{n})^{\frac{1}{2}})$.

By Lemma A.2 the remaining parts can be proved in a similar fashion noting that, $D_{jn} = O(\frac{n}{h_n})$, of the sandwich forms of R_{jn} for $j = 2, 3, 4$, of the higher order derivatives of the concentrated estimating equation.

Proof of Proposition 1: We go on to prove the proposition using Lemma 1. To that effect

[6] If $\forall\epsilon > 0, \exists c \geq 0, n_0 > 0$ s.t. $\mathrm{P}(|x_n| > cf_n) < \epsilon, \forall n \geq n_0$ then $x_n = O_p(f_n)$

consider the Taylor series expansion of $\tilde{\psi}_n(\rho)$ around ρ_0,

$$\begin{aligned} 0 &= \tilde{\psi}_n(\hat{\rho}_n) \\ &= \tilde{\psi}_n + H_{1n}(\hat{\rho}_n-\rho_0) + \tfrac{1}{2}H_{2n}(\hat{\rho}_n-\rho_0)^2 + \tfrac{1}{6}H_{3n}(\hat{\rho}_n-\rho_0)^3 + \tfrac{1}{6}[H_{3n}(\bar{\rho})-H_{3n}](\hat{\rho}_n-\rho_0)^3, \end{aligned}$$

where the last two terms sums up the mean value form of the remainder term with $\bar{\rho}$ lying between ρ_0 and $\hat{\rho}_n$. We have already shown that $\hat{\rho}_n - \rho_0 \to_p \big(\frac{h_n}{n}\big)^{\frac{1}{2}}$. Next, note that $h_n T_{rn} = O(1)$ for $r = 0,1,2,3$ by Assumptions 4 and 5. Now, in order to prove the result of the proposition, we need to establish the following conditions:

(*i*) $\tilde{\psi}_n = O_p\big((\frac{h_n}{n})^{\frac{1}{2}}\big)$ and $\mathrm{E}(\tilde{\psi}_n) = O(\frac{h_n}{n})$,
(*ii*) $\mathrm{E}(H_{rn}) = O(1)$ and $H_{rn} - \mathrm{E}(H_{rn}) = O_p\big((\frac{h_n}{n})^{\frac{1}{2}}\big)$ for $r = 1,2,3$,
(*iii*) $H_{1n}^{-1} = O_{pu}(1)$ and $\mathrm{E}(H_{1n})^{-1} = O(1)$ and
(*iv*) $H_{3n}(\bar{\rho}) - H_{3n} = O_p\big((\frac{h_n}{n})^{\frac{1}{2}}\big)$.

For (*i*), by Lemma A.2, $\epsilon_n' M_n G_n M_n \epsilon_n - \sigma_0^2 \mathrm{tr}(M_n G_n M_n) = O_p\big((\frac{n}{h_n})^{\frac{1}{2}}\big)$ and

$$\mathrm{tr}(M_n G_n M_n) = \mathrm{tr}(G_n) + O(1) = nT_{0n} + O(1). \tag{B-3}$$

Therefore, $\tilde{\psi}_n = -h_n T_{0n} + h_n R_{1n} = -h_n T_{0n} + \frac{h_n}{\sigma_0^2 n}\epsilon_n' M_n G_n M_n \epsilon_n + O_p(\frac{1}{\sqrt{n}}) = -h_n T_{0n} + \frac{h_n}{\sigma_0^2 n}\big[\sigma_0^2 \mathrm{tr}(G_n) + O_p((\frac{n}{h_n})^{\frac{1}{2}})\big] + O_p(\frac{1}{\sqrt{n}}) = O_p\big((\frac{h_n}{n})^{\frac{1}{2}}\big)$ and $\mathrm{E}(\tilde{\psi}_n) = -h_n T_{0n} + \frac{h_n}{n}\mathrm{tr}(M_n G_n M_n) + O\big((\frac{h_n}{n})^{\frac{1}{2}}\big) = -h_n T_{0n} + \frac{h_n}{n}(\mathrm{tr}(G_n) + O(1)) + O\big((\frac{h_n}{n})^{\frac{1}{2}}\big) = O(\frac{h_n}{n})$.

For (*ii*), Lemma 1 implies, $(h_n R_{1n})^s = \mathrm{E}(h_n R_{1n})^s + O_p((\frac{h_n}{n})^{\frac{1}{2}})$ for $s = 2,3,4$, $(h_n R_{2n})^2 = \mathrm{E}(h_n R_{2n})^2 + O_p((\frac{h_n}{n})^{\frac{1}{2}})$, $(h_n R_{1n})^s h_n R_{2n} = \mathrm{E}(h_n R_{1n})^s \mathrm{E}(h_n R_{2n}) + O_p((\frac{h_n}{n})^{\frac{1}{2}})$ for $s = 1,2$, and $h_n R_{1n} h_n R_{3n} = \mathrm{E}(h_n R_{1n})\mathrm{E}(h_n R_{3n}) + O_p((\frac{h_n}{n})^{\frac{1}{2}})$.

Therefore, Assumption 8 implies, $\mathrm{E}[(h_n R_{1n})^s] = \mathrm{E}(h_n R_{1n})^s + O((\frac{h_n}{n})^{\frac{1}{2}})$ for $s = 2,3,4$, $\mathrm{E}[(h_n R_{2n})^2] = \mathrm{E}(h_n R_{2n})^2 + O((\frac{h_n}{n})^{\frac{1}{2}})$, $\mathrm{E}[(h_n R_{1n})^s h_n R_{2n}] = \mathrm{E}(h_n R_{1n})^s \mathrm{E}(h_n R_{2n}) + O((\frac{h_n}{n})^{\frac{1}{2}})$ for $s = 1,2$, and $\mathrm{E}[h_n R_{1n} h_n R_{3n}] = \mathrm{E}(h_n R_{1n})\mathrm{E}(h_n R_{3n}) + O((\frac{h_n}{n})^{\frac{1}{2}})$. Combining these results with (B-3) and Lemma 1, we reach to the conclusion that, $H_{rn} - \mathrm{E}(H_{rn}) = O_p((\frac{h_n}{n})^{\frac{1}{2}})$ and $\mathrm{E}(H_{rn}) = O(1)$ for $r = 1,2,3$.

For (*iii*), by Lemma 1 and $\mathrm{E}[(h_n R_{1n})^2] = \mathrm{E}(h_n R_{1n})^2 + O\big((\frac{h_n}{n})^{\frac{1}{2}}\big)$,

$$\begin{aligned} \mathrm{E}(H_{1n}) &= \tfrac{2}{h_n}\mathrm{E}[(h_n R_{1n})^2] - h_n T_{1n} - \mathrm{E}(h_n R_{2n}) \\ &= \tfrac{2}{h_n}\big(\tfrac{h_n}{n}\mathrm{tr}(M_n G_n M_n) + O\big((\tfrac{h_n}{n})^{\frac{1}{2}}\big)\big)^2 - h_n T_{1n} - \big(\tfrac{h_n}{n}\mathrm{tr}(M_n D_{2n} M_n) + O\big((\tfrac{h_n}{n})^{\frac{1}{2}}\big)\big) \\ &= \tfrac{2}{h_n}\big(\big(\tfrac{h_n}{n}\mathrm{tr}(M_n G_n M_n)\big)^2\big) - h_n T_{1n} - \tfrac{h_n}{n}\mathrm{tr}(M_n D_{2n} M_n) + O\big((\tfrac{h_n}{n})^{\frac{1}{2}}\big) \\ &= \tfrac{2}{h_n}\big(\tfrac{h_n}{n}\mathrm{tr}(G_n)\big)^2 - \tfrac{h_n}{n}\mathrm{tr}(G_n^2) - \tfrac{h_n}{n}\mathrm{tr}(G_n' G_n) + O\big((\tfrac{h_n}{n})^{\frac{1}{2}}\big) \\ &= -\tfrac{h_n}{n}\big(\mathrm{tr}(G_n^2) + \mathrm{tr}(G_n' G_n) - 2T_{0n}^2 \mathrm{tr}(I_n)\big) + O\big((\tfrac{h_n}{n})^{\frac{1}{2}}\big) \\ &= -\tfrac{h_n}{n}\big(\mathrm{tr}(G_n - T_{0n} I_n)^2 + \mathrm{tr}(G_n - T_{0n} I_n)'(G_n - T_{0n} I_n)\big) + O\big((\tfrac{h_n}{n})^{\frac{1}{2}}\big). \end{aligned}$$

That is, $\mathrm{E}(H_{1n})$ is negative for sufficiently large n and it is finite. Therefore, $\mathrm{E}(H_{1n})^{-1} = O(1)$. Also by, $H_{1n} = \mathrm{E}(H_{1n}) + O_p((\frac{h_n}{n})^{\frac{1}{2}})$, we have, $H_{1n}^{-1} = O_p(1)$.

Finally for (iv), consider equation (19) evaluated at $\bar{\rho}_n$. By the mean value theorem, $h_n T_{3n}(\bar{\rho}) = \frac{h_n}{n}\text{tr}(G_n^4(\bar{\rho})) = \frac{h_n}{n}\text{tr}(G_n^4) + 4\frac{h_n}{n}\text{tr}(G_n^5(\tilde{\rho}))(\bar{\rho}-\rho_0)$, where, $\tilde{\rho}$ lies between $\bar{\rho}$ and ρ_0. By repeatedly applying the mean value theorem we can find a $\tilde{\rho}$ which is much closer to the true value ρ_0. For such $\tilde{\rho}$, $\frac{h_n}{n}\text{tr}(G_n^5(\tilde{\rho})) = O(1)$ by Assumptions 4 and 5. Combining with the $(\frac{n}{h_n})^{1/2}$-convergence of $\bar{\rho}$ to the true value we have, $h_n T_{3n}(\bar{\rho}) = O(1)$.

Now consider $\hat{\sigma}_n^2(\bar{\rho}) = \frac{1}{n}Y_n' A_n'(\bar{\rho})M_n(\bar{\rho})A_n(\bar{\rho})Y_n$ and $\hat{\sigma}_{n0}^2 = \frac{1}{n}Y_n' A_n' M_n A_n Y_n$. Similarly, by the mean value theorem we have, $\hat{\sigma}_n^2(\bar{\rho}) = \hat{\sigma}_{n0}^2 - \frac{2}{n}(\bar{\rho}-\rho_0)Y_n' A_n'(\tilde{\rho})M_n(\tilde{\rho})G_n(\tilde{\rho})M_n(\tilde{\rho})A_n(\tilde{\rho})Y_n = \hat{\sigma}_{n0}^2 - 2(\bar{\rho}-\rho_0)O_p(h_n^{-1}) = \hat{\sigma}_{n0}^2 + O_p((nh_n)^{-1/2})$. By continuity of $\hat{\sigma}_{n0}^{-2}$, it can be deduced that, $\hat{\sigma}_n^{-2}(\bar{\rho}) = \big(\hat{\sigma}_{n0}^2 + O_p((nh_n)^{-1/2})\big)^{-1} = \hat{\sigma}_{n0}^{-2} + O_p((nh_n)^{-1/2})$. Now,

$$
\begin{aligned}
h_n R_{1n}(\bar{\rho}) &= \hat{\sigma}_n^{-2}(\bar{\rho})\tfrac{h_n}{n}Y_n' A_n'(\bar{\rho})M_n(\bar{\rho})G_n(\bar{\rho})M_n(\bar{\rho})A_n(\bar{\rho})Y_n \\
&= \hat{\sigma}_n^{-2}(\bar{\rho})\tfrac{h_n}{n}\big[Y_n' A_n' M_n G_n M_n A_n Y_n - (\bar{\rho}-\rho_0)Y_n' A_n'(\tilde{\rho})M_n(\tilde{\rho})D_{2n}(\tilde{\rho})M_n(\tilde{\rho})A_n(\tilde{\rho})Y_n\big] \\
&= \big(h_n R_{1n} + O_p\big((\tfrac{1}{nh_n})^{\frac{1}{2}}\big)\big) - O_p((\tfrac{h_n}{n})^{\frac{1}{2}}) = h_n R_{1n} + O_p(\tfrac{h_n}{n})^{\frac{1}{2}} \qquad \text{(B-4)}
\end{aligned}
$$

Using a similar set of arguments it can be shown that, $h_n R_{kn}(\bar{\rho}) = h_n R_{kn} + O_p((\frac{h_n}{n})^{\frac{1}{2}})$ for $k = 2, 3, 4$. Then it follows that, $H_{3n}(\bar{\rho}) - H_{3n} = O_p((\frac{h_n}{n})^{\frac{1}{2}})$.

Proof of Proposition 2: Arguments are similar to that of Proposition 1.

Proof of Proposition 3: Note that $b_2(\rho_0,\gamma_0) = O((\frac{n}{h_n})^{-1})$ and that it is differentiable. It follows that $\frac{\partial}{\partial(\rho_0,\gamma_0)}b_2(\rho_0,\gamma_0) = O((\frac{n}{h_n})^{-1})$. As $\hat{\rho}_n$, the QMLE of ρ defined at the beginning of Section 2, is $\sqrt{n/h_n}$-consistent, it can be shown that $\hat{\gamma}_n = \gamma(\hat{\mathcal{F}}_n)$ is also $\sqrt{n/h_n}$-consistent. We have, under the additional assumptions in Proposition 3,

$$
b_2(\hat{\rho}_n,\hat{\gamma}_n) = b_2(\rho_0,\gamma_0) + \tfrac{\partial}{\partial\rho_0}b_2(\rho_0,\gamma_0)(\hat{\rho}_n-\rho_0) + \tfrac{\partial}{\partial\gamma_0}b_2(\rho_0,\gamma_0)(\hat{\gamma}_n-\gamma_0) + O_p((\tfrac{n}{h_n})^{-2}).
$$

Thus, $\text{E}[b_2(\hat{\rho}_n,\hat{\gamma}_n)] = b_2(\rho_0,\gamma_0) + \frac{\partial}{\partial\rho_0}b_2(\rho_0,\gamma_0)\text{E}(\hat{\rho}_n-\rho_0) + \frac{\partial}{\partial\gamma_0}b_2(\rho_0,\gamma_0)\text{E}(\hat{\gamma}_n-\gamma_0) + O((\frac{n}{h_n})^{-2})]$. As $\text{E}(\hat{\rho}_n-\rho_0) = O(\frac{h_n}{n})$, it can be shown that $\text{E}(\hat{\gamma}_n-\gamma_0) = O(\frac{h_n}{n})$. These lead to $\text{E}[b_2(\hat{\rho}_n,\hat{\gamma}_n)] = b_2(\rho_0,\gamma_0) + O((\frac{n}{h_n})^{-2})$. Similarly, we show that $\text{E}[b_3(\hat{\rho}_n,\hat{\gamma}_n)] = b_3(\rho_0,\gamma_0) + o((\frac{n}{h_n})^{-2})$, noting that $b_3(\rho_0,\gamma_0) = O((\frac{n}{h_n})^{-3/2})$.

Clearly, our bootstrap estimate has two step approximations, one is that described above, and the other is the bootstrap approximations to the various expectations in (25) given $\hat{\rho}_n$, e.g.,

$$
\hat{\text{E}}(H_{1n}\tilde{\psi}_n) = \tfrac{1}{B}\textstyle\sum_{b=1}^{B} H_{1n}(e_{n,b}^*,\hat{\rho}_n)\tilde{\psi}_n(e_{n,b}^*,\hat{\rho}_n).
$$

However, these approximations can be made arbitrarily accurate, for a given $\hat{\rho}_n$ and $\mathcal{F}_n$, by choosing an arbitrarily large B. The result of Proposition 3 thus follows.

References

Anselin, L., 1980. *Estimation methods for spatial autoregressive structures: A study in spatial econometrics.* Regional Science Dissertation and Monograph Series 8.

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

Anselin, L., 2001. *Spatial Econometrics. A Companion to Theoretical Econometrics, edited by Badi H. Baltagi.* Blackwell Publishing.

Anselin, L., 2003. Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review* 26, 153-166.

Anselin, L., Bera, A. K., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: *Handbook of Applied Economic Statistics, edited by Aman Ullah and David E. A. Giles.* New York: Marcel Dekker.

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

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

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

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

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

Burridge, P. 1980. On the Cliff-Ord test for spatial autocorrelation. *Journal of the Royal Statistical Society* B 42, 107-8.

Cliff, A., Ord, J. K., 1972. Testing for spatial autocorrelation among regression residuals. *Geographical Analysis* 4, 267-284.

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

Cliff, A. D., Ord, J. K., 1981. *Spatial Process, Models and Applications.* London: Pion.

Das, D., Kelejian, H. H., Prucha, I. R., 2003. Finite sample properties of spatial autoregressive models with autoregressive disturbances. *Papers in Regional Science* 82, 1-26.

Efron, B., 1979. Bootstrap method: another look at the jackknife. *The Annals of Statistics* 7, 1-26.

Fingleton, B., 2008. A generalized method of moments estimator for a spatial model with moving average errors, with application to real estate prices. *Empirical Economics* 33, 35-57.

Fleming, M. M., 2004. Techniques for Estimating Spatially Dependent Discrete Choice Models *Advances in Spatial Econometrics, edited by Luc Anselin, Raymond J. G. M. Florax and Sergio J. Rey.* Springer Berlin Heidelberg.

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

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

Jin, F., Lee, L. F., 2013. Cox-type tests for competing spatial autoregressive models with spatial autoregressive disturbances. *Regional Science and Urban Economics* 43, 590-616.

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

Kelejian H. H., Prucha, I. R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40, 509-533.

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

Kelejian H. H., Prucha, I. R., 2004. Instrumental variable estimation of a spatial autoregressive models with autoregressive disturbances: Large and small sample results. In *Spatial and Spatiotemporal Econometrics, Advances in Econometrics edited by LeSage, J., and Pace, R. K.* New York: Elsevier 18, 163-198.

Kelejian H. H., Prucha, I. R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157 53-67.

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

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

Lee, L. F., 2001. Generalised method of moments estimation of spatial autoregressive processes - Unpublished manuscript.

Lee, L. F., 2002. Consistency and efficiency of least squares estimation for mixed regressive spatial autoregressive models. *Econometric Theory* 18, 252-277.

Lee, L. F., 2003. Best spatial two-stage least squares estimators for a spatial autoregressive models with autoregressive disturbances. *Econometric Reviews* 22, 307-335.

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

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

Lee, L. F., Liu, X., 2010. Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. *Econometric Theory* 26 187-230.

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

LeSage, J., Pace, R.K., 2009. *Introduction to Spatial Econometrics.* CRC Press, Taylor & Francis Group, London.

Liu, S. F., Yang, Z. L. 2015. Modified QML estimation of spatial autoregressive models with unknown heteroskedasticity and nonnormality. *Regional Science and Urban Economics* 52, 50-70.

Magnus, J. R., 1978. Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix. *Journal of Econometrics*, 7, 281-312.

Martellosio, F., 2010. Power properties of invariant tests for spatial autocorrelation in linear regression. *Econometric Theory* 26, 152-186.

Ord, J., 1975. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70, 120-126.

Pinkse, J. 1998. Asymptotic properties of the Moran and related tests and a test for spatial correlation in probit models. Department of Economics, University of British Columbia.

Rao, C. R., 1973. Linear Statistical Inference and Its Applications. New York: John Wiley and Sons.

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

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

Yang, Z. L., 2015. A general method for third order bias and variance corrections on a nonlinear estimator. *Journal of Econometrics* 186, 178-200.

Accepted version for: Econometrics 3, 376-411, 2015.

Table 1

Empirical Mean[rmse](sd) of Estimators of ρ for SED Model with SAR Errors - Rook Contiguity, REG-1

		Normal Errors		Mixed Normal Errors		Log-Normal Errors	
ρ	n	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$
.50	50	.440[.175](.164)	.495[.169](.169)	.445[.166](.157)	.499[.161](.161)	.452[.152](.144)	.503[.147](.147)
	100	.472[.116](.112)	.501[.114](.114)	.471[.112](.108)	.499[.110](.110)	.473[.104](.101)	.500[.102](.102)
	200	.487[.079](.077)	.501[.078](.078)	.486[.077](.075)	.500[.076](.076)	.487[.072](.071)	.500[.071](.071)
	500	.495[.049](.049)	.501[.049](.049)	.495[.049](.048)	.500[.049](.049)	.495[.046](.046)	.500[.046](.046)
.25	50	.202[.192](.186)	.248[.195](.195)	.203[.182](.176)	.248[.184](.184)	.207[.169](.163)	.250[.170](.170)
	100	.228[.130](.128)	.252[.131](.131)	.225[.127](.124)	.248[.127](.127)	.228[.119](.117)	.251[.120](.120)
	200	.239[.091](.090)	.251[.091](.091)	.239[.090](.090)	.250[.090](.090)	.240[.085](.084)	.251[.085](.085)
	500	.246[.057](.057)	.250[.057](.057)	.246[.057](.057)	.251[.058](.058)	.246[.055](.055)	.251[.055](.055)
.00	50	-.032[.192](.189)	.002[.201](.201)	-.035[.184](.181)	-.002[.191](.191)	-.033[.178](.175)	-.002[.184](.184)
	100	-.021[.135](.133)	-.004[.137](.137)	-.018[.131](.130)	.000[.133](.133)	-.019[.124](.123)	-.003[.126](.126)
	200	-.010[.097](.096)	-.001[.098](.098)	-.008[.093](.093)	.001[.094](.094)	-.010[.089](.088)	-.002[.089](.089)
	500	-.005[.060](.060)	-.001[.060](.060)	-.005[.059](.059)	-.001[.059](.059)	-.004[.058](.058)	.001[.058](.058)
-.25	50	-.270[.180](.179)	-.252[.191](.191)	-.273[.171](.170)	-.255[.181](.181)	-.274[.169](.168)	-.257[.178](.178)
	100	-.262[.127](.126)	-.252[.130](.130)	-.261[.124](.123)	-.251[.127](.127)	-.262[.120](.119)	-.252[.123](.123)
	200	-.255[.090](.090)	-.250[.091](.091)	-.255[.088](.088)	-.250[.089](.089)	-.255[.087](.087)	-.250[.088](.088)
	500	-.253[.057](.057)	-.250[.058](.058)	-.252[.057](.057)	-.250[.058](.058)	-.253[.056](.056)	-.250[.057](.057)
-.50	50	-.503[.152](.152)	-.502[.163](.163)	-.503[.144](.144)	-.500[.153](.153)	-.509[.144](.143)	-.507[.153](.153)
	100	-.504[.107](.107)	-.502[.111](.111)	-.503[.104](.104)	-.501[.108](.108)	-.504[.103](.103)	-.502[.106](.106)
	200	-.502[.076](.076)	-.501[.077](.077)	-.502[.074](.074)	-.501[.076](.076)	-.503[.074](.074)	-.502[.075](.075)
	500	-.501[.048](.048)	-.500[.049](.049)	-.501[.047](.047)	-.500[.048](.048)	-.501[.046](.046)	-.501[.047](.047)

Table 2

Empirical Mean[rmse](sd) of Estimators of ρ for SED Model with SAR Errors - Queen Contiguity, REG-1

		Normal Errors		Mixed Normal Errors		Log-Normal Errors	
ρ	n	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$
.50	50	.390[.244](.218)	.492[.215](.215)	.395[.232](.206)	.493[.204](.204)	.406[.207](.184)	.501[.181](.181)
	100	.445[.153](.143)	.499[.140](.140)	.449[.145](.135)	.501[.133](.133)	.451[.133](.124)	.501[.122](.122)
	200	.474[.099](.095)	.500[.095](.095)	.474[.098](.095)	.500[.094](.094)	.476[.091](.087)	.500[.087](.087)
	500	.491[.059](.058)	.501[.058](.058)	.490[.059](.058)	.500[.058](.058)	.490[.056](.055)	.500[.055](.055)
.25	50	.144[.270](.248)	.248[.250](.250)	.153[.255](.236)	.254[.238](.238)	.153[.239](.218)	.250[.219](.219)
	100	.196[.179](.171)	.253[.169](.169)	.194[.177](.168)	.249[.166](.166)	.197[.165](.156)	.250[.154](.154)
	200	.221[.121](.117)	.248[.117](.117)	.222[.118](.115)	.249[.114](.114)	.225[.110](.107)	.250[.107](.107)
	500	.240[.073](.073)	.250[.073](.073)	.240[.075](.074)	.250[.074](.074)	.241[.069](.068)	.251[.068](.068)
.00	50	-.101[.294](.276)	-.002[.285](.285)	-.095[.277](.260)	.003[.268](.268)	-.095[.259](.241)	-.001[.247](.247)
	100	-.059[.200](.192)	-.002[.192](.192)	-.059[.197](.188)	-.002[.189](.189)	-.055[.181](.172)	.001[.172](.172)
	200	-.027[.135](.132)	.001[.133](.133)	-.026[.132](.130)	.002[.130](.130)	-.027[.124](.121)	-.002[.121](.121)
	500	-.011[.083](.082)	-.001[.082](.082)	-.011[.082](.081)	.000[.081](.081)	-.010[.079](.079)	.001[.079](.079)
-.25	50	-.339[.299](.285)	-.248[.300](.300)	-.338[.284](.270)	-.249[.283](.283)	-.337[.265](.250)	-.251[.261](.261)
	100	-.308[.211](.203)	-.252[.206](.206)	-.303[.202](.195)	-.248[.198](.198)	-.307[.194](.185)	-.254[.188](.188)
	200	-.277[.142](.140)	-.251[.141](.141)	-.274[.140](.138)	-.249[.139](.139)	-.275[.132](.129)	-.250[.130](.130)
	500	-.262[.089](.089)	-.252[.089](.089)	-.260[.088](.088)	-.250[.088](.088)	-.261[.084](.083)	-.251[.084](.084)
-.50	50	-.576[.291](.281)	-.499[.301](.301)	-.577[.283](.272)	-.502[.290](.290)	-.584[.268](.255)	-.511[.271](.270)
	100	-.548[.208](.203)	-.498[.209](.209)	-.550[.201](.195)	-.501[.201](.201)	-.547[.193](.188)	-.499[.193](.193)
	200	-.524[.144](.142)	-.501[.144](.144)	-.524[.141](.139)	-.501[.141](.141)	-.521[.136](.134)	-.498[.136](.136)
	500	-.511[.090](.089)	-.502[.090](.089)	-.510[.089](.089)	-.501[.089](.089)	-.509[.086](.086)	-.500[.086](.086)

Accepted version for: Econometrics 3, 376-411, 2015.

Table 3

Empirical Mean[rmse](sd) of Estimators of ρ for SED Model with SAR Errors - Group Interaction, $k = n^{0.5}$, REG-2

		Normal Errors		Mixed Normal Errors		Log-Normal Errors	
ρ	n	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$
.50	50	.277[.403](.335)	.523[.223](.222)	.287[.395](.332)	.524[.223](.222)	.303[.354](.294)	.532[.194](.192)
	100	.375[.233](.197)	.512[.148](.148)	.377[.233](.198)	.511[.149](.149)	.384[.214](.180)	.515[.136](.136)
	200	.424[.160](.141)	.502[.116](.116)	.430[.152](.134)	.506[.111](.111)	.432[.143](.126)	.507[.104](.104)
	500	.454[.106](.096)	.502[.085](.085)	.455[.105](.095)	.502[.085](.085)	.456[.100](.090)	.502[.080](.080)
.25	50	-.082[.548](.437)	.291[.325](.322)	-.078[.541](.431)	.288[.318](.315)	-.061[.507](.401)	.296[.296](.293)
	100	.051[.345](.281)	.268[.220](.219)	.052[.342](.278)	.265[.218](.218)	.068[.309](.249)	.275[.196](.194)
	200	.129[.239](.206)	.259[.171](.171)	.127[.236](.201)	.256[.168](.168)	.131[.220](.184)	.257[.154](.153)
	500	.176[.160](.141)	.254[.126](.126)	.175[.161](.142)	.253[.127](.127)	.179[.153](.135)	.255[.120](.120)
.00	50	-.433[.679](.523)	.040[.419](.417)	-.432[.672](.514)	.034[.412](.411)	-.400[.620](.474)	.055[.378](.375)
	100	-.270[.448](.357)	.018[.288](.288)	-.260[.435](.347)	.020[.280](.280)	-.251[.409](.324)	.025[.263](.261)
	200	-.172[.315](.264)	.009[.223](.223)	-.171[.312](.261)	.008[.221](.221)	-.162[.295](.246)	.012[.209](.209)
	500	-.107[.215](.186)	.002[.167](.167)	-.106[.213](.185)	.002[.166](.166)	-.100[.199](.173)	.006[.156](.155)
-.25	50	-.758[.767](.575)	-.210[.487](.485)	-.746[.753](.567)	-.209[.483](.481)	-.723[.708](.527)	-.195[.448](.445)
	100	-.573[.534](.425)	-.227[.354](.353)	-.574[.530](.420)	-.233[.350](.350)	-.563[.490](.377)	-.228[.314](.313)
	200	-.467[.394](.329)	-.242[.282](.282)	-.466[.382](.315)	-.242[.271](.271)	-.455[.356](.291)	-.236[.250](.250)
	500	-.383[.263](.227)	-.240[.205](.204)	-.381[.263](.228)	-.246[.206](.206)	-.379[.250](.215)	-.245[.194](.194)
-.50	50	-1.057[.828](.614)	-.456[.553](.551)	-1.059[.828](.611)	-.467[.550](.549)	-1.040[.782](.566)	-.454[.505](.503)
	100	-.880[.612](.480)	-.481[.409](.409)	-.875[.598](.465)	-.482[.397](.396)	-.857[.562](.434)	-.472[.369](.368)
	200	-.753[.451](.374)	-.487[.325](.325)	-.751[.445](.369)	-.487[.320](.320)	-.746[.422](.344)	-.487[.299](.299)
	500	-.655[.308](.267)	-.493[.242](.242)	-.659[.311](.267)	-.497[.243](.243)	-.652[.294](.251)	-.492[.228](.228)

Accepted version for: Econometrics 3, 376-411, 2015.

Table 4

Empirical Mean[rmse](sd) of Estimators of ρ for SED Model with SAR Errors - Group Interaction, $k = n^{0.65}$, REG-2

		Normal Errors		Mixed Normal Errors		Log-Normal Errors	
ρ	n	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$
.50	50	.435[.155](.140)	.504[.119](.119)	.440[.147](.134)	.507[.114](.114)	.441[.133](.119)	.506[.101](.101)
	100	.458[.110](.101)	.502[.091](.091)	.460[.105](.097)	.502[.087](.087)	.462[.094](.086)	.503[.077](.077)
	200	.477[.077](.073)	.503[.069](.068)	.475[.077](.073)	.501[.068](.068)	.478[.069](.065)	.503[.061](.061)
	500	.486[.053](.051)	.501[.050](.050)	.485[.053](.051)	.500[.049](.049)	.487[.050](.048)	.502[.046](.046)
.25	50	.148[.213](.186)	.257[.166](.166)	.151[.205](.179)	.257[.160](.160)	.154[.189](.162)	.257[.144](.144)
	100	.182[.156](.140)	.252[.129](.129)	.183[.151](.135)	.252[.124](.124)	.185[.139](.123)	.252[.112](.112)
	200	.209[.113](.105)	.252[.099](.099)	.211[.109](.102)	.253[.096](.096)	.209[.104](.095)	.250[.090](.090)
	500	.228[.076](.073)	.252[.070](.070)	.227[.077](.073)	.251[.070](.070)	.227[.072](.068)	.251[.066](.066)
.00	50	-.129[.253](.218)	.006[.205](.205)	-.127[.244](.208)	.006[.195](.195)	-.119[.222](.187)	.011[.175](.174)
	100	-.087[.191](.170)	.005[.159](.159)	-.088[.187](.165)	.003[.155](.154)	-.081[.169](.148)	.007[.138](.138)
	200	-.056[.144](.133)	.003[.126](.126)	-.056[.140](.128)	.002[.122](.122)	-.052[.131](.120)	.005[.114](.114)
	500	-.033[.101](.096)	-.001[.093](.093)	-.034[.100](.094)	-.001[.091](.091)	-.030[.093](.088)	.002[.086](.086)
-.25	50	-.395[.273](.231)	-.248[.227](.227)	-.389[.260](.220)	-.244[.216](.216)	-.384[.241](.201)	-.242[.196](.196)
	100	-.351[.218](.193)	-.244[.184](.184)	-.353[.215](.189)	-.247[.180](.180)	-.349[.197](.170)	-.246[.162](.162)
	200	-.319[.170](.156)	-.248[.149](.149)	-.321[.169](.154)	-.251[.147](.147)	-.317[.155](.140)	-.249[.134](.134)
	500	-.290[.122](.115)	-.249[.112](.112)	-.291[.122](.115)	-.251[.112](.112)	-.289[.114](.107)	-.250[.104](.104)
-.50	50	-.647[.276](.234)	-.499[.241](.241)	-.644[.269](.228)	-.499[.236](.236)	-.639[.252](.210)	-.497[.215](.215)
	100	-.616[.241](.212)	-.497[.205](.205)	-.609[.234](.207)	-.492[.200](.200)	-.610[.219](.189)	-.495[.183](.183)
	200	-.580[.193](.176)	-.499[.170](.170)	-.579[.191](.174)	-.499[.168](.168)	-.579[.179](.161)	-.500[.156](.156)
	500	-.547[.141](.133)	-.500[.129](.129)	-.545[.139](.131)	-.498[.128](.128)	-.544[.131](.124)	-.497[.121](.121)

Accepted version for: Econometrics 3, 376-411, 2015.

Table 5

Replication of Table 2 for $\beta = (.5, .1, .1)$

		Normal Errors		Mixed Normal Errors		Log-Normal Errors	
ρ	n	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$
.50	50	.395[.242](.218)	.499[.213](.213)	.396[.230](.205)	.497[.200](.200)	.404[.210](.187)	.501[.182](.182)
	100	.446[.150](.140)	.500[.138](.138)	.447[.149](.139)	.499[.137](.137)	.451[.135](.125)	.501[.123](.123)
	200	.474[.100](.096)	.500[.096](.096)	.475[.096](.093)	.500[.092](.092)	.476[.091](.087)	.500[.087](.087)
	500	.490[.059](.058)	.500[.058](.058)	.490[.059](.058)	.500[.058](.058)	.491[.056](.055)	.501[.055](.055)
.25	50	.137[.282](.258)	.246[.258](.258)	.145[.263](.241)	.251[.240](.240)	.152[.246](.225)	.253[.224](.224)
	100	.195[.182](.173)	.252[.172](.172)	.196[.173](.165)	.252[.163](.163)	.195[.162](.152)	.249[.151](.151)
	200	.224[.121](.118)	.250[.118](.118)	.224[.118](.115)	.251[.115](.115)	.226[.111](.108)	.251[.108](.108)
	500	.241[.072](.071)	.251[.071](.071)	.240[.072](.071)	.251[.071](.071)	.241[.070](.070)	.251[.070](.070)
.00	50	-.104[.297](.279)	.004[.286](.286)	-.106[.285](.264)	-.002[.270](.270)	-.098[.269](.250)	.004[.255](.255)
	100	-.059[.201](.192)	-.002[.193](.193)	-.058[.196](.187)	-.001[.188](.188)	-.054[.181](.173)	.002[.173](.173)
	200	-.027[.134](.131)	.001[.132](.132)	-.028[.133](.131)	-.002[.131](.131)	-.027[.124](.121)	-.001[.121](.121)
	500	-.010[.082](.081)	.002[.082](.082)	-.012[.083](.082)	-.001[.082](.082)	-.011[.079](.078)	-.001[.078](.078)
-.25	50	-.352[.305](.288)	-.253[.302](.302)	-.351[.294](.276)	-.254[.289](.289)	-.346[.279](.262)	-.252[.273](.273)
	100	-.302[.208](.202)	-.247[.205](.205)	-.304[.203](.196)	-.249[.199](.199)	-.304[.192](.185)	-.251[.187](.187)
	200	-.275[.142](.140)	-.250[.141](.141)	-.280[.139](.136)	-.255[.137](.137)	-.277[.134](.131)	-.252[.132](.132)
	500	-.261[.090](.089)	-.251[.089](.089)	-.261[.088](.087)	-.251[.088](.088)	-.259[.085](.085)	-.249[.085](.085)
-.50	50	-.591[.300](.286)	-.506[.307](.307)	-.592[.290](.276)	-.508[.294](.294)	-.588[.280](.265)	-.506[.282](.282)
	100	-.549[.207](.201)	-.500[.208](.208)	-.554[.203](.195)	-.506[.201](.201)	-.548[.193](.187)	-.500[.192](.192)
	200	-.524[.144](.142)	-.501[.144](.144)	-.522[.141](.140)	-.499[.142](.142)	-.523[.136](.134)	-.501[.136](.136)
	500	-.509[.091](.090)	-.500[.091](.091)	-.508[.090](.089)	-.499[.090](.090)	-.510[.087](.086)	-.500[.087](.087)

Accepted version for: Econometrics 3, 376-411, 2015.

Table 6

Replication of Table 2 for $\sigma = 3$

		Normal Errors		Mixed Normal Errors		Log-Normal Errors	
ρ	n	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$
.50	50	.392[.243](.217)	.499[.210](.210)	.396[.234](.209)	.499[.202](.202)	.404[.212](.189)	.505[.182](.182)
	100	.449[.150](.141)	.501[.139](.139)	.449[.147](.137)	.499[.135](.135)	.452[.134](.125)	.501[.123](.123)
	200	.474[.098](.095)	.500[.094](.094)	.475[.097](.094)	.500[.093](.093)	.474[.091](.087)	.499[.087](.087)
	500	.489[.060](.059)	.499[.059](.059)	.490[.060](.059)	.500[.058](.058)	.490[.056](.055)	.500[.055](.055)
.25	50	.139[.282](.259)	.253[.257](.257)	.136[.271](.246)	.247[.243](.243)	.147[.249](.227)	.255[.224](.223)
	100	.196[.180](.172)	.250[.171](.171)	.195[.174](.165)	.249[.165](.165)	.202[.159](.152)	.253[.151](.151)
	200	.220[.120](.116)	.247[.116](.116)	.225[.119](.116)	.251[.116](.116)	.226[.110](.107)	.251[.107](.107)
	500	.240[.074](.073)	.250[.073](.073)	.240[.072](.071)	.251[.071](.071)	.240[.070](.070)	.250[.070](.070)
.00	50	-.114[.307](.285)	.001[.291](.291)	-.111[.297](.275)	.001[.280](.280)	-.109[.279](.256)	-.001[.259](.259)
	100	-.053[.195](.188)	.003[.189](.189)	-.053[.192](.184)	.001[.185](.185)	-.051[.177](.170)	.002[.171](.171)
	200	-.027[.134](.131)	-.001[.132](.132)	-.028[.132](.129)	-.002[.129](.129)	-.027[.123](.120)	-.002[.121](.121)
	500	-.010[.083](.083)	.001[.083](.083)	-.011[.082](.082)	-.001[.082](.082)	-.011[.079](.078)	-.001[.078](.078)
-.25	50	-.364[.312](.291)	-.258[.306](.305)	-.356[.298](.278)	-.250[.291](.291)	-.355[.286](.266)	-.252[.276](.276)
	100	-.300[.209](.203)	-.248[.207](.207)	-.302[.202](.195)	-.252[.199](.199)	-.297[.187](.181)	-.248[.183](.183)
	200	-.277[.143](.141)	-.252[.142](.142)	-.275[.139](.137)	-.249[.138](.138)	-.274[.134](.132)	-.249[.132](.132)
	500	-.259[.088](.087)	-.249[.087](.087)	-.262[.088](.087)	-.252[.087](.087)	-.260[.085](.085)	-.250[.085](.085)
-.50	50	-.593[.305](.290)	-.501[.312](.312)	-.596[.292](.276)	-.504[.296](.296)	-.599[.281](.263)	-.509[.280](.280)
	100	-.548[.207](.201)	-.503[.208](.208)	-.547[.198](.193)	-.502[.199](.199)	-.543[.192](.187)	-.499[.192](.192)
	200	-.522[.145](.143)	-.499[.145](.145)	-.525[.142](.140)	-.503[.142](.142)	-.522[.136](.134)	-.500[.136](.136)
	500	-.509[.091](.091)	-.500[.091](.091)	-.511[.089](.088)	-.502[.089](.089)	-.510[.086](.086)	-.501[.086](.086)

Table 7

Empirical Mean[rmse](sd) of Estimators of ρ for SED Model with SMA Errors - Queen Contiguity, REG-1

		Normal Errors		Mixed Normal Errors		Log-Normal Errors	
ρ	n	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$
.50	100	.554[.154](.145)	.509[.418](.418)	.552[.151](.142)	.509[.318](.318)	.553[.149](.139)	.506[.140](.140)
	200	.527[.101](.097)	.501[.096](.096)	.528[.099](.095)	.502[.095](.095)	.527[.096](.093)	.501[.092](.092)
	500	.510[.059](.058)	.500[.058](.058)	.510[.059](.058)	.500[.058](.058)	.510[.059](.058)	.500[.058](.058)
.25	100	.302[.184](.176)	.256[.178](.178)	.301[.180](.173)	.255[.171](.171)	.292[.171](.166)	.247[.163](.163)
	200	.275[.121](.119)	.251[.117](.117)	.273[.120](.118)	.250[.116](.116)	.274[.115](.112)	.251[.111](.111)
	500	.259[.074](.073)	.250[.073](.073)	.261[.073](.072)	.252[.072](.072)	.260[.071](.071)	.251[.070](.070)
.00	100	.041[.204](.200)	-.001[.196](.196)	.040[.197](.193)	-.002[.188](.188)	.039[.187](.183)	-.001[.179](.179)
	200	.019[.136](.134)	-.002[.132](.132)	.022[.133](.131)	.002[.129](.129)	.021[.129](.127)	.001[.125](.125)
	500	.009[.083](.083)	.001[.083](.083)	.009[.082](.082)	.001[.081](.081)	.008[.081](.080)	.000[.080](.080)
-.25	100	-.214[.217](.214)	-.249[.208](.208)	-.217[.210](.208)	-.251[.202](.202)	-.222[.197](.195)	-.254[.189](.189)
	200	-.234[.145](.144)	-.250[.142](.142)	-.233[.143](.142)	-.249[.140](.140)	-.235[.138](.137)	-.251[.134](.134)
	500	-.245[.089](.089)	-.251[.089](.089)	-.245[.089](.089)	-.251[.089](.089)	-.245[.086](.086)	-.251[.086](.086)
-.50	100	-.472[.218](.216)	-.498[.209](.209)	-.475[.214](.212)	-.500[.205](.205)	-.479[.201](.200)	-.502[.193](.193)
	200	-.489[.149](.149)	-.501[.146](.146)	-.492[.146](.146)	-.503[.143](.143)	-.490[.139](.138)	-.500[.136](.136)
	500	-.495[.092](.092)	-.500[.091](.091)	-.495[.089](.089)	-.500[.089](.089)	-.496[.087](.087)	-.500[.086](.086)

Accepted version for: Econometrics 3, 376-411, 2015.

Table 8

Empirical Mean[rmse](sd) of Estimators of ρ for SED Model with SMA Errors - Group Interaction, $k = n^{0.5}$, REG-1

		Normal Errors		Mixed Normal Errors		Log-Normal Errors	
ρ	n	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$
.50	100	.549[.129](.120)	.508[.128](.127)	.548[.126](.117)	.507[.124](.124)	.548[.121](.111)	.507[.118](.118)
	200	.534[.106](.100)	.503[.104](.104)	.534[.104](.098)	.502[.102](.102)	.533[.099](.094)	.502[.097](.097)
	500	.519[.078](.076)	.501[.078](.078)	.520[.079](.077)	.502[.079](.079)	.519[.077](.074)	.502[.076](.076)
.25	100	.309[.184](.174)	.254[.183](.183)	.310[.179](.169)	.256[.177](.177)	.306[.167](.158)	.253[.165](.165)
	200	.292[.148](.142)	.252[.147](.147)	.292[.147](.141)	.252[.146](.146)	.294[.140](.133)	.254[.138](.138)
	500	.277[.116](.113)	.252[.116](.116)	.276[.116](.113)	.252[.116](.116)	.275[.111](.108)	.251[.111](.111)
.00	100	.071[.234](.223)	.005[.234](.234)	.069[.228](.217)	.004[.227](.227)	.065[.211](.200)	.002[.209](.209)
	200	.051[.197](.190)	.001[.198](.198)	.053[.192](.185)	.004[.192](.192)	.052[.180](.172)	.004[.178](.178)
	500	.032[.152](.149)	-.001[.154](.154)	.032[.150](.146)	.001[.150](.150)	.034[.145](.141)	.003[.145](.145)
-.25	100	-.168[.281](.269)	-.246[.282](.282)	-.174[.269](.258)	-.251[.270](.270)	-.172[.254](.242)	-.246[.253](.253)
	200	-.194[.234](.227)	-.253[.236](.236)	-.187[.233](.225)	-.245[.233](.233)	-.192[.221](.214)	-.249[.222](.222)
	500	-.210[.188](.184)	-.248[.189](.189)	-.211[.188](.184)	-.249[.189](.189)	-.213[.178](.174)	-.251[.179](.179)
-.50	100	-.411[.321](.308)	-.500[.324](.324)	-.408[.315](.302)	-.495[.316](.316)	-.417[.294](.282)	-.503[.296](.296)
	200	-.427[.276](.266)	-.496[.276](.276)	-.427[.272](.262)	-.495[.273](.273)	-.436[.256](.247)	-.502[.257](.257)
	500	-.456[.219](.215)	-.501[.221](.221)	-.453[.223](.218)	-.498[.224](.224)	-.456[.213](.208)	-.501[.214](.214)