GMM inference in spatial autoregressive models
Süleyman Taşpinar, Osman Doğan, and Wim P. M. Vijverberg

ABSTRACT
In this study, we investigate the finite sample properties of the optimal generalized method of moments estimator (OGMME) for a spatial econometric model with a first-order spatial autoregressive process in the dependent variable and the disturbance term (for short SARAR(1, 1)). We show that the estimated asymptotic standard errors for spatial autoregressive parameters can be substantially smaller than their empirical counterparts. Hence, we extend the finite sample variance correction methodology of Windmeijer (2005) to the OGMME for the SARAR(1, 1) model. Results from simulation studies indicate that the correction method improves the variance estimates in small samples and leads to more accurate inference for the spatial autoregressive parameters. For the same model, we compare the finite sample properties of various test statistics for linear restrictions on autoregressive parameters. These tests include the standard asymptotic Wald test based on various GMMEs, a bootstrapped version of the Wald test, two versions of the C(α) test, the standard Lagrange multiplier (LM) test, the minimum chi-square test (MC), and two versions of the generalized method of moments (GMM) criterion test. Finally, we study the finite sample properties of effects estimators that show how changes in explanatory variables impact the dependent variable.

KEYWORDS
Asymptotic variance; efficiency; GMM; inference; optimal GMME; SARAR; spatial autoregressive models; variance correction; Windmeijer correction

JEL CLASSIFICATION
C13; C21; C31

1. Introduction
The finite sample properties of generalized method of moments estimators (GMMEs) have been extensively studied in the literature on nonspatial econometric models (in the 1996 special issue of Journal of Business & Economic Statistics). For example, Hansen et al. (1996) show that the estimated asymptotic variance of their continuous updating GMME for the asset pricing models are downward biased in finite samples. For linear and nonlinear models of covariance structures, Altonji and Segal (1996) and Clark (1996) show that the correlation between the optimal weight matrix and sampling errors introduces finite sample bias and that the amount of bias is relatively greater when data have long-tailed and skewed distributions. Furthermore, the optimal GMMEs (OGMMEs) of models of covariance structures have smaller asymptotic standard errors, making inference and specification testing problematic. Hall and Horowitz (1996); Brown and Newey (2002) confirm that the first-order asymptotic theory often provides poor approximations to the distributions of test statistics obtained from GMM estimators. Similarly, with regard to dynamic panel data models and panel count data models, asymptotic standard errors for the OGMME may also be severely downward biased (more than 20%) relative to the empirical standard deviations (Arellano and Bond, 1991; Blundell and Bond, 1998; Windmeijer, 2005; Bond and Windmeijer, 2005; Windmeijer, 2008).

GMM estimators for the spatial autoregressive models may suffer from similar deficiencies because they are essentially multistep estimators. For instance, Kelejian and Prucha (1998, 2010) propose a
multistep estimation method that involves a combination of IV and GMM estimation for the spatial model with a spatial autoregressive process in the dependent variable and the disturbance term, commonly referred to as SARAR(1, 1). In terms of estimation, the Kelejian and Prucha approach is straightforward as computations do not require special techniques even for extremely large samples. In principle, the GMM/IV estimator is inefficient relative to the maximum likelihood estimator (MLE) (Prucha, 2014), but in finite samples the difference can be negligible (Das et al., 2003).

To increase the efficiency of GMMEs, Lee (2007a,b), Lin and Lee (2010), Liu et al. (2010), and Lee and Liu (2010) suggest a set of linear and quadratic moment functions. The reduced forms of spatial models specify expressions for spatial lags of the dependent variable (i.e., endogenous variables) in terms of exogenous variables and the disturbance term. Then, linear moment functions may be constructed on the basis of the deterministic part of the spatial lag terms, and quadratic moment functions may be formulated from the stochastic part of the spatial lag variables. Both linear and quadratic quadratic moment functions are chosen in such a way that the resulting GMME is asymptotically equivalent to the MLE when disturbances are independent and identically distributed (i.i.d.) with a normal density. When disturbances are simply i.i.d., Liu et al. (2010) and Lee and Liu (2010) show that the generalized method of moments (GMM) estimator can be more efficient than the quasi MLE, respectively, for the case of SARAR(1, 1) and SARAR(p, q).

The GMM estimation approach suggested in Lee (2007a,b), Lin and Lee (2010), and Liu et al. (2010) requires initial consistent parameter estimates to obtain not only the optimal weight matrix but also the linear and quadratic moment functions. The asymptotic theory in Lee (2007a,b), Lin and Lee (2010), and Liu et al. (2010) assumes that these estimated matrices are infinitely precise and therefore does not account for the variation stemming from the initial estimator used to construct these matrices. The correction method suggested in Windmeijer (2005) specifically accounts for the presence of these initial consistent parameter estimates in the optimal weight matrix. Therefore, it can provide improvements for statistical inference. The aforementioned studies on the GMM estimation of spatial models indicate that the GMMEs have desirable finite sample properties in terms of bias and root mean square error. However, their performance in terms of estimated asymptotic standard errors and related test statistics for statistical inference has not been explored. Hence, further research on the performance of the estimated asymptotic standard errors of the GMME in spatial autoregressive models is warranted.

In this study, we evaluate the performance of the asymptotic approximations suggested for the GMM estimators and MLE for the SARAR(1, 1) model. We extend the finite sample correction method of Windmeijer (2005) to our model. The correction is based on a Taylor approximation of the first-order conditions of the GMM objective function to account for the additional variation in the optimal weight matrix. Therefore, the correction method accounts for the variation stemming from the initial GMM estimation to obtain an estimate of the optimal weight matrix. In a Monte Carlo study, we compare the estimated asymptotic standard errors and the corrected ones with the empirical standard deviations under various scenarios. Our simulation results indicate that the estimates of the asymptotic standard errors based on the OGMME are downward biased for the spatial autoregressive parameters. The amount of bias is generally larger than 5%. For the correction method, the amount of bias is less than 5% in most cases suggesting that the correction method can be useful for statistical inference.

For the SARAR(1, 1) specification, we consider various test statistics for testing linear restrictions on spatial autoregressive parameters based on the GMM and ML methods. These tests include standard asymptotic Wald tests based on the MLE and the GMME, a bootstrapped version of the Wald test, two versions of the \(C(\alpha)\) test, the standard LM test, the minimum chi-square test (MC), and two versions of the GMM criterion test. We compare the performance of these tests through a Monte Carlo simulation for the two null hypotheses that there is no spatial dependence \((i)\) in the dependent variable or \((ii)\) in the disturbance term. Our results show that the standard asymptotic Wald test based on the OGMME tends to over-reject, whereas the Wald test based on the corrected standard errors has proper size properties. The GMM criterion tests and the \(C(\alpha)\) test can be useful for testing the null hypothesis that there is no spatial dependence in the disturbance terms. Our results also show that the \(C(\alpha)\) test can be useful to detect the presence of spatial dependence in the dependent variable.
For the same specification, we study the finite sample properties of scalar measures suggested by LeSage and Pace (2009) for the partial derivatives that show how changes in explanatory variables impact the dependent variable. We consider (i) the simulation method of LeSage and Pace (2009) and (ii) the Delta method to study the dispersions of the direct, indirect and total effects. Our results indicate that both methods perform similarly and the dispersions calculated with the corrected standard errors for the effects estimators can be useful as they have relatively smaller bias.

This article is organized in the following way. Section 2 presents the spatial autoregressive model under consideration and discusses its assumptions. Section 3 reviews the GMM estimation approach and lays out the details of the finite sample correction method for the SARAR(1, 1) specification. Section 4 reviews scalar measures suggested for the interpretation of parameters, and shows how their dispersions can be calculated. Section 5 reviews the details of test statistics under consideration in the GMM framework. Section 6 lays out the details of the Monte Carlo design and presents the results. Section 7 closes with concluding remarks. Some of the technical derivations and Monte Carlo results are relegated to a web appendix.

2. The model specification and assumptions

The SARAR(1, 1) specification is given by

\[ Y_n = \lambda_0 W_n Y_n + X_n \beta_0 + u_n, \quad u_n = \rho_0 M_n u_n + \epsilon_n, \]

(2.1)

where \( Y_n \) is the \( n \times 1 \) vector of a dependent variable, \( X_n \) is the \( n \times k \) matrix of non-stochastic exogenous variables with a conformable parameter vector \( \beta_0 \), \( W_n \) and \( M_n \) are the \( n \times n \) spatial weight matrices of known constants with zero diagonal elements, and \( \epsilon_n \) is the \( n \times 1 \) vector of disturbances (or innovations).

\( W_n Y_n \) and \( M_n u_n \) are known as, respectively, the spatial lag of the dependent variable and the spatial lag of the disturbance term. \( \lambda_0 \) and \( \rho_0 \) are the spatial autoregressive parameters. Because spatial data is characterized with triangular arrays, the variables in (2.1) have the subscript \( n \) (Kelejian and Prucha, 2010). Let \( \Theta \) be the parameter space of the model. In order to distinguish the true parameter vector from other possible values in \( \Theta \), we state the model with the true parameter vector \( \theta_0 = (\rho_0, \lambda_0, \beta_0') \).

Furthermore, for notational simplicity we let \( S_n(\lambda) = (I_n - \lambda W_n) \), \( R_n(\rho) = (I_n - \rho M_n) \), \( G_n(\lambda) = W_n S_n^{-1}(\lambda) \), \( H_n(\rho) = M_n R_n^{-1}(\rho) \), and \( \overline{G}_n(\rho, \lambda) = R_n(\rho) G_n(\lambda) R_n^{-1}(\rho) \).

Also, at the true parameter values \( (\rho_0, \lambda_0) \), we denote \( S_n(\lambda_0) = S_n \), \( R_n(\rho_0) = R_n \), \( G_n(\lambda_0) = G_n \), \( H_n(\rho_0) = H_n \), and \( \overline{G}_n = R_n G_n R_n^{-1} \).

The consistency and asymptotic normality of the GMME for (2.1) are established under the following assumptions.\(^1\)

**Assumption 1.** The elements \( \epsilon_{in} \) of the disturbance term \( \epsilon_n \) are distributed independently and identically with mean zero and variance \( \sigma^2_\epsilon \), and \( E|\epsilon_{in}|^{4+\nu} < \infty \) for some \( \nu > 0 \) for all \( n \) and \( i \).

**Assumption 2.** The spatial weight matrices \( M_n \) and \( W_n \) are uniformly bounded in row and column sums in absolute value. Moreover, \( S_n^{-1}, S_n^{-1}(\lambda), R_n^{-1}, \) and \( R_n^{-1}(\rho) \) exist and are uniformly bounded in row and column sums in absolute value for all values of \( \rho \) and \( \lambda \) in a compact parameter space.

The regularity conditions in Assumptions 1 and 2 are motivated to restrict the spatial autocorrelation in the model at a tractable level (Kelejian and Prucha, 1998). By this assumption, the third and fourth moments of \( \epsilon_{in} \), denoted respectively by \( \mu_3 \) and \( \mu_4 \), exist for all \( i \) and \( n \). Assumption 2 also implies that the model in (2.1) represents an equilibrium relation for the dependent variable. By this assumption, the reduced form of the model can be written as \( Y_n = S_n^{-1} X_n \beta_0 + S_n^{-1} R_n^{-1} \epsilon_n \).\(^2\)

\(^1\)For interpretations and implications of these assumptions, see Lee (2007a) and Kelejian and Prucha (2010).

\(^2\)The uniform boundedness of \( S_n^{-1}(\lambda) \) and \( R_n^{-1}(\rho) \) is required for the ML estimator. In the case of GMM estimators, only the uniform boundedness of \( S_n^{-1} \) and \( R_n^{-1} \) is required (Liu et al., 2010).
a function of the sample size, such that the sequence of weight matrices \( \{W_n\} \) is uniformly bounded in both row and column sums and its elements \( w_{n,j} \) are \( \mathcal{O}\left(\frac{1}{\ln n}\right) \). The sequence \( \{h_n\} \) can be bounded or divergent with the property that \( \lim_{n \to 0} \frac{h_n}{n} = 0 \), which implies that \( h_n \) is allowed to diverge only at a rate slower than that of \( n \). For large group interactions for which \( \lim_{n \to \infty} \frac{h_n}{n} \neq 0 \), the estimators might not be consistent. In this study, we assume that \( h_n \) is bounded.

Throughout this study, the vector of moment functions we consider for the GMM estimator takes the form of \( g(\theta_0) = (\varepsilon_n P_{1n} \varepsilon_n^\prime, \ldots, \varepsilon_n P_{mn} \varepsilon_n^\prime, \varepsilon_n^\prime Q_n \varepsilon_n)^\prime \). Moment functions involving the \( n \times n \) constant matrices \( P_{jn} \) for \( j = 1, \ldots, m \) are known as quadratic moment functions. The moment function \( Q_n \varepsilon_n^\prime \varepsilon_n \) is a linear moment function where \( Q_n \) is an \( n \times r \) instrument matrix with \( r \geq k + 1 \) and has full column rank. The matrices \( P_{jn} \) and \( Q_n \) are chosen in such way that orthogonality conditions of population moment functions are not violated. Let \( \mathcal{P}_n \) be the class of \( n \times n \) constant matrices with zero trace. The quadratic moment functions involving matrices from \( \mathcal{P}_n \) satisfy the orthogonality conditions when disturbance terms are i.i.d. Assumption 4 states regularity conditions for these matrices. Assumption 5 characterizes the parameter space.\(^3\)

**Assumption 3.** The regressors matrix \( X_n \) is an \( n \times k \) matrix consisting of uniformly bounded constant elements. It has full column rank. Moreover, \( \lim_{n \to \infty} \frac{1}{n} X_n^\prime X_n \) exists and is nonsingular.

**Assumption 4.** Elements of the IV matrix \( Q_n \) are uniformly bounded. Matrices \( P_{jn} \) for \( j = 1, \ldots, m \) are uniformly bounded in row and column sums in absolute value.

**Assumption 5.** The parameter space \( \Theta \) is a compact subset of \( \mathbb{R}^{k+2} \), and \( \theta_0 \in \text{Int}(\Theta) \).

### 3. Spatial GMM and finite sample variance correction method

#### 3.1. The moment conditions

Equation (2.1) implies \( Y_n = S_n^{-1}(X_n \beta_0 + R_n^{-1} \varepsilon_n) \). Thus, the endogenous variable \( W_n Y_n \) on the right-hand side of Eq. (2.1) may be written as \( W_n Y_n = W_n S_n^{-1}(X_n \beta_0 + R_n^{-1} \varepsilon_n) = G_n X_n \beta_0 + G_n R_n^{-1} \varepsilon_n \). Equation (2.1) may also be written as \( R_n Y_n = \lambda_0 R_n W_n Y_n + R_n X_n \beta_0 + \varepsilon_n \). This expression is the basis for the choice of instruments (Lee, 2007a; Liu et al., 2010; Lee and Liu, 2010; Lin and Lee, 2010; Kelejian and Prucha, 2010): the linear moment matrix \( Q_n \) is constructed from the expectation of \( Z_n = (R_n W_n Y_n, R_n X_n) \), which implies that the orthogonality condition for the population moment function holds: \( E(Q_n' \varepsilon_n) = Q_n' E(\varepsilon_n') = 0_{1 \times 1} \). Given consistent initial estimates of \( \lambda_0, \rho_0 \), and \( \beta_0 \), the IV matrix \( Q_n \) becomes available. In particular, Lee (2003) shows that the 2SLS estimator based on \( Q_n \) is the best estimator in the sense that its asymptotic covariance matrix is the smallest among the class of 2SLS estimators based on linear moment conditions.

Instruments that lead to quadratic moment functions follow from Assumption 1, in particular that \( \varepsilon_{jn} \) is i.i.d. with variance \( \sigma_j^2 \). For any \( P_{jn} \in \mathcal{P}_n \), \( P_{jn} \varepsilon_n^\prime \) is an instrument that satisfies the orthogonality condition for a population moment function: \( E(\varepsilon_n P_{jn} \varepsilon_n^\prime) = E(\varepsilon_n P_{jn}^\prime \varepsilon_n^\prime) = \text{tr}(P_{jn}^\prime E(\varepsilon_n \varepsilon_n^\prime)) = \sigma_j^2 \text{tr}(P_{jn}) = 0 \) for \( j = 1, \ldots, m \), where \( \text{tr}() \) returns the trace of an input matrix. Moreover, when \( P_{jn} \) is suitably chosen, we have \( E((P_{jn} \varepsilon_n)^\prime R_n W_n Y_n) = \sigma_j^2 \text{tr}(P_{jn}^\prime R_n G_n R_n^{-1}) \neq 0: P_{jn} \varepsilon_n \) also instruments \( R_n W_n Y_n \).

The consistency of the GMM estimator does not depend on a particular \( P_{jn} \), but the asymptotic variance-covariance matrix is a function of utilized \( P_{jn} \) matrices. Therefore, the asymptotic efficiency of the estimator is used as a criterion for the selection of \( P_{jn} \) matrices from \( \mathcal{P}_n \). Liu et al. (2010) suggest a set of quadratic and linear moment matrices such that the resulting GMME is theoretically more efficient than the quasi MLE. In this approach, the diagonal elements of the best quadratic and linear moment matrices are also used to formulate additional moment functions to increase efficiency.

\(^3\)See Kelejian and Prucha (2010) for the specification of the parameter space of autoregressive parameters.
To describe the best choice of matrices $P_{jn}$, we define the following notation. For the regressors matrix, let $\bar{X}_n = R_nX_n = (R_nI_n, R_nX_n^*)$, where $I_n$ denotes $n \times 1$ vector of ones and $X_n^*$ is the submatrix of $X_n$ with the intercept column deleted. The submatrix $R_nX_n^* = \bar{X}_n^*$ is assumed to be $n \times k^*$ matrix with column vectors denoted by $\bar{X}_n^*$ for $l = 1, \ldots, k^*$. Let $D(\cdot)$ be the operator which either creates a matrix from the diagonal elements of an input matrix or returns a diagonal matrix if the input is a vector. Furthermore, let $\text{vec}(\cdot)$ be the operator that creates a column vector from the elements of an input matrix, and let $\text{vec}_{\mathcal{D}}(\cdot)$ be the operator that creates a column vector from the diagonal elements of an input matrix. Finally, for any $n \times n$ matrix $A_n$, let $A_n^{(l)} = A_n - \frac{1}{n} \text{tr}(A_n)I_n$, and let $A_n^{(s)} = A_n + A_n'$. Define the following quadratic column matrices: $P_{1n}^* = \bar{G}_{(l)}^1, P_{2n}^* = D(\bar{G}_{(l)}^1), P_{3n}^* = D(\bar{G}_{n\bar{X}_n\theta_0})^1(t), \ P_{4n}^* = H_n^1, P_{5n}^* = D(H_n^1), \text{and } P_{l+5,n}^* = D(\bar{X}_n^*)^l(t) \text{ for } l = 1, \ldots, k^*$. Note that $P_{jn}^* \in \mathcal{P}_n$ for $j = 1, \ldots, k^* + 5$. The linear moment matrices are $Q_n^* = (Q_{1n}^*, Q_{2n}^*, Q_{3n}^*, Q_{4n}^*, Q_{5n}^*) = (\bar{X}_n^*, \bar{G}_{n\bar{X}_n\theta_0}, I_n, \text{vec}_D(\bar{G}_{(l)}^1(t)), \text{vec}_D(H_n^1(t)))$, which, one might note, expands on the matrix $Q_n$ given above. Then, as shown by Liu et al. (2010), the best set of moment functions is given by $g_n^*(\theta) = \left(\epsilon_n' \theta (P_{1n}^* P_{1n}^*)^{-1} \epsilon_n, \ldots, \epsilon_n' \theta (P_{k^*+5,n}^* P_{k^*+5,n}^*)^{-1} \epsilon_n, \epsilon_n' \theta Q_n^* \epsilon_n\right)'$, where $\epsilon_n(\theta) = R_n(\rho)S_n(\lambda)Y_n - R_n(\rho)X_n\beta$. This set of moment functions is best in the sense that any other moment function that can be added to this set does not increase the asymptotic efficiency of the GMME.

Let $\Omega_n^* = E \left[ g_n^*(\theta) g_n^*(\theta)' \right]$ and $\Phi_n^* = -E \left[ \frac{\partial g_n^*(\theta)}{\partial \theta} \right]$. It can be shown that:

$$\Omega_n^* = \begin{bmatrix} (\mu_4 - 3\mu_3^2) \theta_{nm}' \theta_{nm} & \mu_3 \theta_{nm} Q_n'^* \ 0_{r \times m} \end{bmatrix} + \sigma_0^4 \begin{bmatrix} \Xi_{nm} & 0_{m \times r} \\ 0_{r \times m} & \sigma_0^{-2} Q_n'^* Q_n^* \end{bmatrix},$$

where $m = k^* + 5, r = k^* + 4, \Xi_{nm} = \left[ \text{vec}(P_{1n}^{(s)}), \ldots, \text{vec}(P_{k^*+5,n}^{(s)}) \right]' \left[ \text{vec}(P_{1n}^*), \ldots, \text{vec}(P_{k^*+5,n}^*) \right]$, and $\theta_{nm} = \left[ \text{vec}_D(P_{1n}^*), \ldots, \text{vec}_D(P_{k^*+5,n}^*) \right]$. The negative expectation of the gradient matrix is given by:

$$\Phi_n^* = \begin{bmatrix} \sigma_0^2 \text{tr}(H_n' P_{1n}^{(s)}) & \sigma_0^2 \text{tr}(\bar{G}_n' P_{1n}^{(s)}) & 0_{1 \times k} \\ \vdots & \vdots & \vdots \\ \sigma_0^2 \text{tr}(H_n' P_{k^*+5,n}^{(s)}) & \sigma_0^2 \text{tr}(\bar{G}_n' P_{k^*+5,n}^{(s)}) & 0_{1 \times k} \\ 0_{r \times 1} & 0_{r \times 1} & Q_n'^* \bar{G}_n' \bar{X}_n' \beta_0 & Q_n'^* \bar{X}_n' \end{bmatrix}.$$

The results in (3.1) and (3.2) indicate that consistent estimates of $\Omega_n^*$ and $\Phi_n^*$ require consistent estimates of $\theta_0, \sigma_0^2, \mu_3$, and $\mu_4$. In practice, an initial estimator of $\theta_0$ can be used to recover consistent estimates of these parameters. Let $\hat{\theta}_1$ be an initial GMME (IGMME). Let $\Omega_n(\hat{\theta}_1)$ denote the estimate of $\Omega_n^*$ recovered from $\hat{\theta}_1$. Then, the OGMME is defined by:

$$\hat{\theta}_n^* = \arg\min_{\theta \in \Theta} g_n^*(\theta) \Omega_n^{-1}(\hat{\theta}_1) g_n^*(\theta).$$

Under Assumptions 1–5, Liu et al. (2010) show that the GMM estimator defined in (3.3) is consistent and asymptotically normally distributed, namely:

$$\sqrt{n}(\hat{\theta}_n^* - \theta_0) \xrightarrow{d} \mathcal{N}(0_{(k+2) \times 1}, \lim_{n \rightarrow \infty} \frac{1}{n} E[\Phi_n^* \Omega_n^{-1}(\hat{\theta}_n^*)^{-1} \Phi_n^*]).$$

1 In the case where $X_0$ has no column of ones, $R_0 X_0 = R_0 X_0^* = \bar{X}_n^*$ and $k^* = k$.
2 Lemma 1 in Web Appendix A can be used to derive $\Omega_n^*$ matrices in this section.
3 An initial estimator of $\theta_0$ can be the one suggested in Prucha (2014). Another candidate is the initial GMM estimator (IGMME) suggested at the end of this section. In addition, the moment matrices also are functions of unknown parameters. With initial estimates, these matrices also became available.
4 The identification of parameters in the GMM framework requires $\lim_{n \rightarrow \infty} \frac{1}{n} E(g_n'(\theta_0)) = 0$ (Newey and McFadden, 1994). Lee (2007a) investigates this condition and state the identification conditions for parameters. Here, we simply assume that the parameters are identified.
An estimate for the covariance of \( \sqrt{n}(\hat{\theta}_n^* - \theta_0) \) is needed for statistical inference. The result in (3.4) suggests that an estimate may be obtained from \( \left[ \frac{1}{n} \Phi_n^*(\hat{\theta}_n^*) \Omega_n^{-1}(\hat{\theta}_1) \Phi_n^*(\hat{\theta}_n^*) \right]^{-1} \), where \( \Phi_n^*(\hat{\theta}_0) \) is an estimate of \( \Phi_n^* \) recovered from \( \hat{\theta}_n^* \). Alternatively, one might form an estimate from \( \left[ \frac{1}{n} \Phi_n^*(\hat{\theta}_0) \Omega_n^{-1}(\hat{\theta}_n^*) \Phi_n^*(\hat{\theta}_n^*) \right]^{-1} \), where \( \Omega_n(\hat{\theta}_n^*) \) is an estimate of \( \Omega_n^* \) recovered from \( \hat{\theta}_n^* \). Although these two procedures are asymptotically equivalent, little seems to be known about the relative merits of these two procedures in small samples (Newey and McFadden, 1994).

The estimator defined in (3.3) is the best GMM estimator in the class of OGMME formulated from the set of linear and quadratic moment functions. When disturbances are i.i.d. normal, the best set of moment functions is given by \( P_{1n} = \overline{G}_n(t) \), \( P_{2n} = H_n(t) \), and \( Q_n = (\overline{G}_n, \overline{X}_n \beta_0, \overline{X}_n) \). Liu et al. (2010) show that any other moment function that can be added to this set is redundant. Furthermore, the MLE is also characterized by this set of moment functions, which implies that the GMME based on this set is asymptotically equivalent to the MLE. When disturbances are simply i.i.d., the additional moment functions formulated from \( D(\overline{G}_n(t)), D(\overline{G}_n, \overline{X}_n \beta_0(t)) \), \( D(H_n(t)), D(\overline{G}_n(t)) \), \( D(H_n(t)) \), and \( D(\overline{X}_n(t)) \) for \( l = 1, \ldots, k^* \) increase asymptotic efficiency. The resulting GMME based on the bigger set \( g_n^*(\theta) \) is asymptotically more efficient than the quasi MLE (Liu et al., 2010). Note that, the efficiency gain from these additional moment functions depends on the specification of weight matrices and regressors.

For the case of SARAR(1, 0), the best set of moment functions follows from the more general case above. Let \( \delta_0 = (\lambda_0, \beta_0')' \) be the true parameter vector. Then, the best set of moments \( g_n^*(\delta) \) uses \( P_{1n} = \overline{G}_n(t) \), \( P_{2n} = D(\overline{G}_n(t)) \), \( P_{3n} = D(G_n \overline{X}_n \beta_0(t)) \), \( P_{4n} = D(X_n(t)) \) for \( l = 1, \ldots, k^* \), and \( Q_n = (Q_{1n}, Q_{2n}, Q_{3n}, Q_{4n}) = (\overline{X}_n, G_n \overline{X}_n \beta_0, l_n, vec(p(G_n(t)))) \), where \( \epsilon_n(\delta) = S_n(\lambda) Y_n - X_n \beta \). Similarly, the best set of the moment functions for the case of SARAR(0, 1) may be obtained. Let \( \Psi_0 = (\rho_0, \beta_0')' \) be the true parameter vector. Then, the best set of moments \( g_n^*(\eta) \) uses \( P_{1n} = H_n(t) \), \( P_{2n} = D(H_n(t)) \), \( P_{3n} = D(\overline{X}_n(t)) \) for \( l = 1, \ldots, k^* \), and \( Q_n = (Q_{1n}, Q_{2n}, Q_{3n}, Q_{4n}) = (\overline{X}_n, l_n, vec(H_n(t))) \) where \( \epsilon_n(\eta) = R_n(\rho) Y_n - R_n(\rho) X_n \beta \).

Note that \( P_{j\star}^n \) and \( Q_{j\star}^n \) also involve unknown parameters. In practice, an initial consistent estimator is used to construct these matrices. For example, an IGMME of \( \theta_0 \) based on quadratic moment matrices \( P_{1n} = W_n' W_n - D(W_n' W_n) \), \( P_{2n} = M_n' M_n - D(M_n' M_n) \), \( P_{3n} = (W_n' W_n(t)) \), and \( P_{4n} = (M_n' M_n(t)) \), and the linear moment matrix \( Q_n = (W_n M_n X_n, W_n X_n, M_n X_n, X_n) \) may be used to construct these matrices when disturbance terms satisfy Assumption 1. Hence, the optimal GMM estimator defined in (3.4) becomes feasible when there is an initial consistent estimator of \( \theta_0 \). The optimal weight, the quadratic, and linear matrices are constructed from the initial estimator. The replacement of \( P_{j\star}^n \), \( Q_{j\star}^n \), and \( \Omega_{j\star}^n \) with their consistent estimates does not affect the asymptotic properties of the optimal GMM estimators.\(^8\) However, it may affect the small-sample behavior of the estimators.

### 3.2. Finite sample variance correction method

In this section, the finite sample correction methodology of Windmeijer (2005) is extended for our spatial autoregressive model. We provide a general argument by considering the general set of population moment functions \( g_n(\theta_0) = (\epsilon_n P_{1n} \epsilon_n, \ldots, \epsilon_n P_{mn} \epsilon_n, \epsilon_n Q_n)' \), where \( Q_n \) is an \( n \times r \) matrix of linear instruments, and \( P_{mn} \in \mathcal{P}_n \) for \( j = 1, \ldots, m \).

The extension of the finite sample variance correction method to our case is not straightforward for mainly two reasons. First, our generic set of moment functions cannot be written at the observational level in such a way that the resulting moment functions are independent across observations. This arises simply because of the spatial dependence. For the spatial framework considered in Windmeijer (2005), the set of moment functions is defined at the observational level such that \( g_{ln}(\theta) \) is the sample moment function for the \( l \)th observation. Within this framework, a good candidate for an estimate of \( \Omega_{ln} \) is

\(^8\)Lee (2007a, Proposition 2) and Liu et al. (2010, Proposition 2, p. 306) give the asymptotic arguments.
Then, ignoring the remainder terms, a first-order Taylor series approximation of (3.6) around the definition of the GMM objective function, we have

\[ \text{\hat{\Psi}}_1. \]

The IGMM estimator is obtained by minimizing the objective function with respect to estimator is defined as \( \hat{\theta} \). Once we have a consistent estimator, we can construct an estimate of \( \theta \) as possible to zero. Furthermore, assume that \( \Psi_n \) converges to a constant positive definite matrix \( \Psi \). The IGMME estimator is obtained by minimizing the objective function with respect to \( \theta \). Let \( Q_n(\theta) = g_n(\theta)^{-1}g_n(\theta) \) be the objective function when the weight matrix is \( \Psi_n \). Then, the initial estimator is defined as \( \hat{\theta}_1 = \text{argmin}_\theta Q_n(\theta). \)

Under our stated assumptions, we have \( \sqrt{n}(\hat{\theta}_1 - \theta_0) \xrightarrow{d} N(0_{(k+2)\times 1}, \gamma) \), where

\[ \gamma = \lim_{n \to \infty} \left( \frac{1}{n} \Phi_n' \Psi_n^{-1} \frac{1}{n} \Phi_n \right)^{-1} \left( \frac{1}{n} \Phi_n' \right) \Psi_n^{-1} \left( \frac{1}{n} \Omega_n \right) \Psi_n^{-1} \left( \frac{1}{n} \Phi_n \right) \left( \frac{1}{n} \Phi_n' \Psi_n^{-1} \frac{1}{n} \Phi_n \right)^{-1}. \] (3.5)

Once we have a consistent estimator, we can construct an estimate of \( \Omega_n \) for the optimal GMM estimation. Let \( \Omega_n(\hat{\theta}_1) \) be the estimate of \( \Omega_n \) constructed from \( \hat{\theta}_1 \). Similarly, we will denote the unknown \( \Omega_n \) with \( \Omega_n(\theta_0) \) in the rest of this section. Then the OGMME is defined as \( \hat{\theta}_2 = \text{argmin}_\theta Q_{\Omega_n(\hat{\theta}_1)} \).

Let \( C_n(\theta) = \frac{\partial g_n(\theta)}{\partial \theta} \) be the matrix of first derivatives and \( F_n(\theta) = \frac{\partial^2 g_n(\theta)}{\partial \theta \partial \theta'} \) be the matrix of second derivatives. The first and second derivatives of the GMM objective function are given by

\[ B_{\Psi_n}(\theta_0) = \frac{1}{2} \left. \frac{\partial Q_n}{\partial \theta} \right|_{\theta_0} = C_n'(\theta_0) \Psi_n^{-1} g_n(\theta_0), \] (3.6)

\[ A_{\Psi_n}(\theta_0) = \frac{1}{2} \left. \frac{\partial^2 Q_n}{\partial \theta \partial \theta'} \right|_{\theta_0} = C_n'(\theta_0) \Psi_n^{-1} C_n(\theta_0) + F_n'(\theta_0) \left[ I_{k+2} \otimes \Psi_n^{-1} g_n(\theta_0) \right], \] (3.7)

where \( B_{\Psi_n}(\theta_0) \) is a \((k+2) \times 1\) vector and \( A_{\Psi_n}(\theta_0) \) is a \((k+2) \times (k+2)\) matrix. When the weight matrix is \( \Omega_n \), the objective function, these matrices are denoted by \( B_{\Omega_n}(\theta_0) \) and \( A_{\Omega_n}(\theta_0) \), respectively. From the definition of the GMM objective function, we have \( B_{\Omega_n(\hat{\theta}_1)}(\hat{\theta}_2) = 0_{(k+2)\times 1} \) and \( B_{\Psi_n}(\hat{\theta}_1) = 0_{(k+2)\times 1}. \) Then, ignoring the remainder terms, a first-order Taylor series approximation of (3.6) around \( \theta_0 \) for the OGMME and IGMME yields, respectively,

\[ \hat{\theta}_2 - \theta_0 = -A_{\Omega_n(\hat{\theta}_1)}^{-1}(\theta_0) B_{\Omega_n(\hat{\theta}_1)}(\theta_0), \] (3.8)

\[ \hat{\theta}_1 - \theta_0 = -A_{\Psi_n}^{-1}(\theta_0) B_{\Psi_n}(\theta_0). \] (3.9)

Note that the expansion in (3.8) is also a function of \( \hat{\theta}_1 \). Therefore, a further expansion of \( \hat{\theta}_1 \) around \( \theta_0 \) yields

\[ \hat{\theta}_2 - \theta_0 = -A_{\Omega_n(\theta_0)}^{-1}(\theta_0) B_{\Omega_n(\theta_0)}(\theta_0) + T_{\Omega_n(\theta_0)}(\theta_0) (\hat{\theta}_1 - \theta_0), \] (3.10)

---

9For notional simplicity, we suppress the subscript \( n \) for the estimators. In order to make a difference between the IGMME and the OGMME, we use \( \hat{\theta}_1 \) for the IGMME and \( \hat{\theta}_2 \) for the OGMME in this section.

10In Web Appendix B, we provide the explicit expressions for \( C_n(\theta) \) and \( F_n(\theta) \) for our generic set of moment functions.
where $\mathcal{T}_{\Omega_n}(\theta_0)(\theta_0) = \frac{\partial \left( A_{\Omega_n(\theta)}(\theta_0) \Omega_{\theta}(\theta_0) \right)}{\partial \theta^T} \bigg|_{\theta_0}$ is a $(k + 2) \times (k + 2)$ matrix. The $j$th column of $\mathcal{T}_{\Omega_n}(\theta_0)(\theta_0)$ is given by (see Windmeijer, 2005, p. 28)

$$
\mathcal{T}_{\Omega_n}(\theta_0)(\theta_0)[\bullet, j] = -A_{\Omega_n(\theta)}^{-1}(\theta_0) \left( C_n(\theta_0) \Omega_n^{-1}(\theta_0) \frac{\partial \Omega_n(\theta)}{\partial \theta_j} \bigg|_{\theta_0} \Omega_n^{-1}(\theta_0) C_n(\theta_0) \right)
$$

$$
+ F_n(\theta_0) \left( I_{k+2} \otimes \Omega_n^{-1}(\theta_0) \frac{\partial \Omega_n(\theta)}{\partial \theta_j} \bigg|_{\theta_0} \Omega_n^{-1}(\theta_0) g_n(\theta) \right) \right) \right)
$$

$$
+ A_{\Omega_n(\theta)}^{-1}(\theta_0) C_n'(\theta_0) \Omega_n^{-1}(\theta_0) \frac{\partial \Omega_n(\theta)}{\partial \theta_j} \bigg|_{\theta_0} \Omega_n^{-1}(\theta_0) g_n(\theta).
$$

(3.11)

Notice that the first term in $\mathcal{T}_{\Omega_n}(\theta_0)(\theta_0)[\bullet, j]$ is a function of $A_{\Omega_n(\theta)}^{-1}(\theta_0) B_{\Omega_n(\theta)}(\theta_0)$. This result indicates that the term $A_{\Omega_n(\theta)}^{-1}(\theta_0) B_{\Omega_n(\theta)}(\theta_0)$ is the bias of an infeasible GMM estimator based on the unknown optimal weight matrix $\Omega_n(\theta_0) = E\left[ g_n(\theta_0) g_n'(\theta_0) \right]$. Hence, the first term is a function of this bias, which tends to be small and generally remains constant as the number of moment functions increases. The remaining term in $\mathcal{T}_{\Omega_n}(\theta_0)(\theta_0)[\bullet, j]$ is the dominant term, and it becomes larger as the number of moment functions increases (Windmeijer, 2005, p. 28).

The result in (3.10) may be used to formulate a variance formula for $\hat{\theta}_2$. The term $\mathcal{T}_{\Omega_n}(\theta_0)(\theta_0)(\hat{\theta}_1 - \theta_0)$ in (3.10) has an order of $O_p\left( \frac{1}{n} \right)$ and vanishes as $n \to \infty$, but it does not equal zero in finite samples. Therefore, accounting for this term in a variance formula for $\hat{\theta}_2$ may improve inference in finite samples. Equations (3.9) and (3.10) result in the following finite sample variance $\hat{\theta}_2$:

$$
\text{Var}_c(\hat{\theta}_2) = A_{\Omega_n(\theta)}^{-1}(\theta_0) \left( C_n(\theta_0) \Omega_n^{-1}(\theta_0) C_n(\theta_0) A_{\Omega_n(\theta)}^{-1}(\theta_0) \right)
$$

$$
+ \mathcal{T}_{\Omega_n}(\theta_0)(\theta_0) A_{\Omega_n(\theta)}^{-1}(\theta_0) C_n(\theta_0) \Omega_n^{-1}(\theta_0) \frac{\partial \Omega_n(\theta)}{\partial \theta_j} \bigg|_{\theta_0} \Omega_n^{-1}(\theta_0) C_n(\theta_0) A_{\Omega_n(\theta)}^{-1}(\theta_0) \right)
$$

$$
+ A_{\Omega_n(\theta)}^{-1}(\theta_0) C_n'(\theta_0) \Omega_n^{-1}(\theta_0) C_n(\theta_0) A_{\Omega_n(\theta)}^{-1}(\theta_0) \right)
$$

$$
+ \mathcal{T}_{\Omega_n}(\theta_0)(\theta_0) \text{Var}(\hat{\theta}_1) T' \mathcal{T}_{\Omega_n}(\theta_0)(\theta_0).
$$

(3.12)

The variance formula in (3.12) is infeasible as it involves the unknown $\theta_0$. Windmeijer (2005) circumvents this problem by simply replacing the unknown parameters with their consistent estimates. The same approach cannot be directly adopted for our case as the functional form for the estimate of $\Omega_n$ is now different. An estimate for the unknown term $\mathcal{T}_{\Omega_n}(\theta_0)(\theta_0)[\bullet, j]$ is recovered from the estimates of $\hat{\theta}_1$ and $\hat{\theta}_2$. The first term in $\mathcal{T}_{\Omega_n}(\theta_0)(\theta_0)[\bullet, j]$ vanishes as $B_{\Omega_n(\theta)}(\theta_0) = 0_{(k+2) \times 1}$ for any consistent estimator $\hat{\theta}_n$. Hence, the first term of (3.11) does not play any role in the calculation of the variance of $\hat{\theta}_2$. Then, a consistent estimate of the $j$th column of $\mathcal{T}_{\Omega_n(\theta)}(\theta_0)$ for $j = 1, \ldots, k + 2$ is given by

$$
\mathcal{T}_{\Omega_n}(\theta_n)(\theta_0)[\bullet, j] = A_{\Omega_n(\theta)}^{-1}(\theta_0) C_n'(\theta_2) \Omega_n^{-1}(\theta_2) \frac{\partial \Omega_n(\theta)}{\partial \theta_j} \bigg|_{\theta_0} \Omega_n^{-1}(\theta_0) g_n(\theta_2).
$$

(3.13)

More explicitly, the terms in (3.13) are given by

$$
\frac{1}{n} \frac{\partial \Omega_n(\theta)}{\partial \theta_j} \bigg|_{\theta_0} = E\left[ \frac{\partial g_n(\theta)}{\partial \theta_j} + g_n(\theta) \frac{\partial g_n'(\theta)}{\partial \theta_j} \bigg|_{\theta_0} \right] + O_p(1).
$$

(3.14)

$$
A_{\Omega_n(\theta)}(\theta_2) = C_n'(\theta_2) \Omega_n^{-1}(\theta_2) g_n(\theta_2),
$$

(3.15)

where $C_n(\theta_2) = \frac{\partial g_n(\theta)}{\partial \theta_j} \bigg|_{\theta_0}$ and $g_n(\theta_2)$ are simply obtained by evaluating $C_n(\theta), F_n(\theta)$, and $g_n(\theta)$ at the consistent estimate $\hat{\theta}_2$. On the other
hand, the consistent estimates $\Omega_n(\hat{\theta}_1)$ and $\Omega_{G_n}(\theta_0)\big|_{\theta_2}$ are recovered from the corresponding results derived under the expectation operator.

Next, we provide explicit expressions for $E\left[ \frac{\partial g_n(\theta)}{\partial \lambda} g'_n(\theta) \big|_{\theta_0} + g_n(\theta) \frac{\partial g'_n(\theta)}{\partial \lambda} \big|_{\theta_0} \right]$ for $j = 1, \ldots, k + 2$. Let $bk(\cdot)$ be the matrix operator that creates a generalized block diagonal matrix from a given list of matrices. For example, if $c = [A_{r \times k}, B_{r \times k}]$, then $bk(c) = \left( A_{r \times k} \ 0_{r \times i} \ B_{r \times k} \right)$. With this new notation, we have the following result with respect to $\lambda$:

$$
E \left[ \frac{\partial g_n(\theta)}{\partial \lambda} g'_n(\theta) \big|_{\theta_0} + g_n(\theta) \frac{\partial g'_n(\theta)}{\partial \lambda} \big|_{\theta_0} \right]
= - \left[ \begin{array}{c}
(\mu_4 - 3\sigma_0^4) \left[ c_{1nm}', c_{2nm} \right]^{(s)} \\
\mu_3 \left[ Q_n' \overline{G}_n c_{2nm} + (I_m' \otimes Q_n) bk(c_{1nm}) \right] \end{array} \right] \begin{array}{c}
\sigma_0^{-2} \left[ I_m \otimes (\overline{G}_n X_n \beta_0)' \right] \begin{array}{c}
\sigma_0^{-2} Q_n' \overline{G}_n Q_n \end{array} \end{array} \left[ \begin{array}{c}
\sigma_0^{-2} I_m \otimes (\overline{G}_n X_n \beta_0)' \end{array} \right] \begin{array}{c}
\sigma_0^{-2} Q_n' \overline{G}_n Q_n \end{array} \right],
$$

(3.16)

where $c_{1nm} = [\text{vec}_D(P_{1n}^{(s)} \overline{G}_n), \ldots, \text{vec}_D(P_{mn}^{(s)} \overline{G}_n)]$, $c_{2nm} = [\text{vec}_D(P_{1n}), \ldots, \text{vec}_D(P_{mn})]$, $c_{3nm} = [\text{vec}(P_{1n}^{(s)} \overline{G}_n), \ldots, \text{vec}(P_{mn}^{(s)} \overline{G}_n)]$, $d_{1nm} = [P_{1n}^{(s)}, \ldots, P_{mn}^{(s)}]$, $I_m$ is the $m \times m$ identity matrix, $l_m$ is the $m \times 1$ vector of ones, and $\otimes$ denotes the kronecker product. In a similar fashion, with respect to $\rho$ we obtain

$$
E \left[ \frac{\partial g_n(\theta)}{\partial \rho} g'_n(\theta) \big|_{\theta_0} + g_n(\theta) \frac{\partial g'_n(\theta)}{\partial \rho} \big|_{\theta_0} \right]
= - \left[ \begin{array}{c}
(\mu_4 - 3\sigma_0^4) \left[ d_{2nm}', c_{2nm} \right]^{(s)} \\
\mu_3 Q_n' \left[ H_n c_{2nm} + d_{2nm} \right] \end{array} \right] \begin{array}{c}
\sigma_0^{-4} \left[ d_{3nm} \right]^{(s)} \end{array} \left[ \begin{array}{c}
\sigma_0^{-4} Q_n' H_n^{(s)} Q_n \end{array} \right],
$$

(3.17)

where $d_{2nm} = [\text{vec}_D(P_{1n}^{(s)} H_n), \ldots, \text{vec}_D(P_{mn}^{(s)} H_n)]$, and $d_{3nm} = [\text{vec}(P_{1n}^{(s)} H_n), \ldots, \text{vec}(P_{mn}^{(s)} H_n)]$. Finally, with respect to $\beta_j$ for $j = 1, \ldots, k$, we obtain

$$
E \left[ \frac{\partial g_n(\theta)}{\partial \beta_j} g'_n(\theta) \big|_{\theta_0} + g_n(\theta) \frac{\partial g'_n(\theta)}{\partial \beta_j} \big|_{\theta_0} \right]
= - \left[ \begin{array}{c}
\mu_3 \left[ \left( I_m \otimes (R_n X_n) \right)' \right] \left[ d_{1nm}' c_{2nm} \right]^{(s)} \\
\sigma_0^2 \left[ (l_m' \otimes Q_n)' \right] \left[ I_m \otimes (R_n X_n) \right] \end{array} \right] \begin{array}{c}
\sigma_0^2 Q_n' H_n^{(s)} Q_n \end{array} \left[ \begin{array}{c}
\sigma_0^{-2} Q_n' H_n^{(s)} Q_n \end{array} \right],
$$

(3.18)

where $X_n$ is the $j$th column of $X_n$. Note that under our stated assumptions, estimates of (3.16), (3.17), and (3.18) are recovered by replacing $(\theta_0', \sigma_0^2, \mu_3, \mu_4)'$ with their consistent estimates $(\hat{\theta}_n', \hat{\sigma}_n^2, \hat{\mu}_3, \hat{\mu}_4)'$.

---

The asymptotic argument is similar to the one given in the proof of Proposition 2 of Liu et al. (2010). Hence, it is omitted.
When $\hat{\theta}_2$ is used to estimate (3.16), (3.17), and (3.18), the resulting estimate of $E\left[\frac{\partial g_n(\theta)}{\partial \beta_j}g_n(\theta)\right]_{\theta_0} + g_n(\theta)\frac{\partial g_n(\theta)}{\partial \beta_j}\bigg|_{\theta_0}$ is denoted by $\frac{\partial \Omega_n(\theta)}{\partial \beta_j}$. Then, inserting these results into (3.12) yields an estimate of the variance of $\hat{\theta}_2$ that incorporates the variation in the IGMME $\hat{\theta}_1$:

$$
\hat{\text{Var}}(\hat{\theta}_2) = A^{-1}_{\Omega_n(\hat{\theta}_1)}(\hat{\theta}_2)C_n(\hat{\theta}_2)\Omega_n^{-1}(\hat{\theta}_1)C_n(\hat{\theta}_2)A^{-1}_{\Omega_n(\hat{\theta}_1)}(\hat{\theta}_2) + T_{\Omega_n(\hat{\theta}_1)}(\hat{\theta}_2)A_{\Psi_n}(\hat{\theta}_1)C_n(\hat{\theta}_2)\Omega_n^{-1}(\hat{\theta}_1)A^{-1}_{\Omega_n(\hat{\theta}_1)}(\hat{\theta}_2)
$$

$$
+ A^{-1}_{\Omega_n(\hat{\theta}_1)}(\hat{\theta}_2)C_n(\hat{\theta}_2)\Psi_n^{-1}(\hat{\theta}_1)A_{\Psi_n}(\hat{\theta}_1)T_{\Omega_n(\hat{\theta}_1)}(\hat{\theta}_2)
$$

$$
+ T_{\Omega_n(\hat{\theta}_1)}(\hat{\theta}_2)\hat{\text{Var}}(\hat{\theta}_1)T_{\Omega_n(\hat{\theta}_1)}(\hat{\theta}_2).
$$

(3.19)

Using (3.9), the estimate of variance of $\hat{\theta}_1$ is obtained as

$$
\hat{\text{Var}}(\hat{\theta}_1) = A_{\Psi_n}(\hat{\theta}_1)C_n(\hat{\theta}_1)\Psi_n^{-1}(\hat{\theta}_1)C_n(\hat{\theta}_1)A_{\Psi_n}(\hat{\theta}_1).
$$

(3.20)

The estimate of the variance of $\hat{\theta}_2$ in (3.19) accounts for the variation that stems from the initial estimator $\hat{\theta}_1$. Windmeijer (2005, 2008) shows that this finite sample correction provides a better approximation to the empirical standard deviations of the estimator when the estimator itself is not substantially biased. The Monte Carlo studies in Lee (2007a), Lin and Lee (2010), and Liu et al. (2010) indicate that the bias of the best GMME of the spatial autoregressive model is negligible. Hence, the correction formula provided in (3.19) can be useful for statistical inference.

4. Direct, indirect and total effects

In this section, we describe the marginal effects (impact measures) accounting within the context of our spatial model. The partial derivative of the dependent variable with respect to $k$th explanatory variable yields the following matrix of marginal effects:

$$
\frac{\partial Y_n}{\partial X_{nk}} = S_n^{-1}\beta_{k0},
$$

(4.1)

where $\beta_{k0}$ is the $k$th component of $\beta_0$. The result in (4.1) indicates that the marginal effects of a change in $X_{nk}$ is represented by an $n \times n$ matrix. The diagonal elements of this matrix ($\partial Y_{ni}/\partial X_{nk}$) contain the own-partial derivatives, while the off-diagonal elements represent the cross-partial derivatives ($\partial Y_{nj}/\partial X_{nk}$). LeSage and Pace (2009) define the average of the main diagonal elements of this matrix as a scalar summary measure of direct effects, and the average of off-diagonal elements as a scalar summary measure of indirect effects. The sum of direct and indirect effects is labeled as the total effects.

We consider two methods for the calculation of dispersions of these impact measures: (i) a simulation approach suggested by LeSage and Pace (2009), and (ii) the Delta method (Debarsy et al., 2015). The simulation approach utilizes the parameter estimates and the estimated covariance matrix of a consistent estimator. Let $L_n$ be a lower-triangular matrix recovered from the Cholesky decomposition of $\text{Var}(\hat{\theta}_n)$, and $\theta_n$ be a random vector that has a multivariate standard normal distribution. Then, random draws of the parameter vector are generated according to

$$
\theta^r = L_n\theta^r_n + \hat{\theta}_n, \quad \text{for } r = 1, \ldots, R.
$$

(4.2)

For each $\theta^r$, the direct, indirect, and total effects can be calculated using (4.1). The mean and the standard deviation calculated from each sequence of impact measures can be used as the point estimate and the standard error of the corresponding impact measure.
The second approach is the conventional method based on the Delta method. The result in (4.1) indicates that the estimate of direct effect is \( \frac{1}{n} \text{tr}(S_n^{-1}(\hat{\lambda}_n) \hat{\beta}_{nk}) \). By the mean value theorem,

\[
\frac{1}{\sqrt{n}} \left[ \text{tr} \left( S_n^{-1}(\hat{\lambda}_n) \hat{\beta}_{nk} \right) - \text{tr} \left( S_n^{-1} \beta_{k0} \right) \right] = A_{1n} \times \sqrt{n} \left( \hat{\lambda}_n - \lambda_0, \hat{\beta}_{nk} - \beta_{k0} \right) + o_p(1)
\]

\[
d \to N \left( 0, \lim_{n \to \infty} A_{1n} B_n A_{1n}' \right),
\]

(4.3)

where \( A_{1n} = \left[ \frac{1}{n} \text{tr} \left( S_n^{-1} G_n \beta_{k0} \right), \frac{1}{n} \text{tr} \left( S_n^{-1} \right) \right] \), and \( B_n \) is the asymptotic covariance of \( \sqrt{n} (\hat{\lambda}_n - \lambda_0, \hat{\beta}_{nk} - \beta_{k0})' \). The result in (4.3) indicates that the asymptotic variance of direct effects can be estimated by \( \frac{1}{n} \hat{A}_{1n} \hat{B}_n \hat{A}_{1n}' \), where \( \hat{A}_{1n} = \left[ \frac{1}{n} \text{tr} \left( S_n^{-1}(\hat{\lambda}_n) G_n(\hat{\lambda}_n) \hat{\beta}_{nk} \right), \frac{1}{n} \text{tr}(S_n^{-1}(\hat{\lambda}_n)) \right] \), and \( \hat{B}_n \) is the estimated asymptotic covariance of \( \sqrt{n}(\hat{\lambda}_n - \lambda_0, \hat{\beta}_{nk} - \beta_{k0})' \).

Applying the mean value theorem to the estimate of total effects \( \frac{1}{n} \hat{\beta}_{nk} l_n' n^{-1}(\hat{\lambda}_n) l_n \) yields

\[
\frac{1}{\sqrt{n}} \left[ \hat{\beta}_{nk} l_n' S_n^{-1}(\hat{\lambda}_n) l_n - \beta_{k0} l_n' S_n^{-1} l_n \right] = A_{2n} \times \sqrt{n} \left( \hat{\lambda}_n - \lambda_0, \hat{\beta}_{nk} - \beta_{k0} \right) + o_p(1)
\]

\[
d \to N \left( 0, \lim_{n \to \infty} A_{2n} B_n A_{2n}' \right),
\]

(4.4)

where \( A_{2n} = \left[ \frac{1}{n} \beta_{k0} l_n' S_n^{-1} G_n l_n, \frac{1}{n} l_n' S_n^{-1} l_n \right] \). Hence, \( \text{Var}(\frac{1}{n} \hat{\beta}_{nk} l_n' n^{-1}(\hat{\lambda}_n) l_n) \) can be estimated by

\[
\frac{1}{n} \hat{A}_{2n} \hat{B}_n \hat{A}_{2n}' \text{, where } A_{2n} = \left[ \frac{1}{n} \hat{\beta}_{nk} l_n' S_n^{-1}(\hat{\lambda}_n) G_n(\hat{\lambda}_n) l_n, \frac{1}{n} l_n' S_n^{-1}(\hat{\lambda}_n) l_n \right].
\]

The estimate of indirect effects is given by \( \frac{1}{n} \left[ \hat{\beta}_{nk} l_n' S_n^{-1}(\hat{\lambda}_n) l_n - \text{tr}(S_n^{-1}(\hat{\lambda}_n) \hat{\beta}_{nk}) \right] \). The results in (4.3) and (4.4) imply that

\[
\frac{1}{\sqrt{n}} \left[ \left( \hat{\beta}_{nk} l_n' S_n^{-1}(\hat{\lambda}_n) l_n - \text{tr}(S_n^{-1}(\hat{\lambda}_n) \hat{\beta}_{nk}) \right) - (\beta_{k0} l_n' S_n^{-1} l_n - \text{tr}(S_n^{-1}) \beta_{k0}) \right]
\]

\[
= (A_{2n} - A_{1n}) \times \sqrt{n} \left( \hat{\lambda}_n - \lambda_0, \hat{\beta}_{nk} - \beta_{k0} \right) + o_p(1) \text{ } d \to N \left( 0, \lim_{n \to \infty} (A_{2n} - A_{1n}) B_n (A_{2n} - A_{1n})' \right).
\]

(4.5)

Hence, an estimate of \( \text{Var}(\frac{1}{n} \left[ \hat{\beta}_{nk} l_n' S_n^{-1}(\hat{\lambda}_n) l_n - \text{tr}(S_n^{-1}(\hat{\lambda}_n) \hat{\beta}_{nk}) \right]) \) is given by \( \frac{1}{n} \left( \hat{A}_{2n} - \hat{A}_{1n} \right) \hat{B}_n \left( \hat{A}_{2n} - \hat{A}_{1n} \right)' \).

### 5. Hypothesis testing with the GMME

In this section, we consider a Wald test procedure that is familiar within the general GMM framework for testing linear and nonlinear restrictions on the parameter vector \( \theta_0 \). We provide details about how to adjust the formulation of this test statistic for the spatial SARAR(1, 1) model.

Let \( r : \mathbb{R}^p \to \mathbb{R}^q \) be a twice continuously differentiable function, and assume that its derivative matrix \( R = \partial r(\theta)/\partial \theta' \) has rank \( q \).\(^{12}\) Let \( \hat{\theta}_n \) be an unrestricted optimal GMME. Then, the Wald test statistic for \( q \) linear restrictions of the form \( r(\theta_0) = 0 \) has a limiting chi-square distribution with \( q \) degrees of freedom, namely

\[
\text{Wald} = \hat{r}_n' \left( \hat{R}_n \text{Var}(\hat{\theta}_n) \hat{R}_n \right)^{-1} \hat{r}_n \text{ } d \to \chi^2_q,
\]

(5.1)

where \( \hat{r}_n = r(\hat{\theta}_n) \) and \( \hat{R}_n = \left. \frac{\partial r(\theta)}{\partial \theta'} \right|_{\theta = \hat{\theta}_n} \). The Wald test statistic may be calculated with the variance formulas given (3.4), (3.5), and (3.19).

We also consider a bootstrapping method for the distribution of the Wald test statistic in our GMM framework for the null hypotheses of \( H_0 : \lambda_0 = 0 \) and \( H_0 : \rho_0 = 0 \). Hall and Horowitz (1996), Brown and Newey (2002), and Bond and Windmeijer (2005) outline bootstrap methods to determine finite sampling distribution of Wald statistics. In the spatial econometrics literature, examples of these methods may be

\(^{12}\)Note that for our spatial model, \( p = k + 2 \).
found in Anselin (1988), Pinkse and Slade (1998), Fingleton (2009), Burridge and Fingleton (2010), and Jin and fei Lee (2013, 2015). According to the bootstrap approach, the empirical distribution of a test statistic is constructed by resampling data. For our model, sampling from $Y_n$, $W_n Y_n$, and $X_n$ cannot be done without preserving the inherent spatial dependence structure of the model. Instead, Anselin (1988) considers bootstrapping of residuals when disturbances of the model are i.i.d. Given a consistent estimator $\hat{\theta}_n$, the vector of residuals is $\hat{e}_n = R_n(\hat{\theta}_n)\left[S_n(\hat{\theta}_n) Y_n - X_n \hat{\beta}_n\right]$. Under Assumption 1, we demean residuals to obtain $\bar{e}_n = J_n \hat{e}_n$, where $J_n = I_n - \frac{1}{n} I_n' I_n$. The bootstrap sample of residuals, denoted by $\hat{e}_n^*$, is generated with random draws from the empirical distribution of the demeaned residuals. Then, the vector of random draws $\hat{e}_n^*$ is used to generate $Y_n^* = S_n^{-1}(\hat{\lambda}_n) [X_n \hat{\beta}_n + R^{-1}(\hat{\rho}_n)\hat{e}_n^*]$. As a result, a bootstrap sample of $\{(Y_n^*, X_n, i^*): i = 1, \ldots, n\}$, where $X_n, i^*$ is the $i$th row of $X_n$, is obtained. For our GMM set up, the bootstrapped Wald test of $H_0: \lambda_0 = 0$ is implemented with the following procedure:

(i) Compute the best GMME $\hat{\theta}_n$ stated in (3.3), and the resulting $\bar{e}_n = J_n R_n(\hat{\theta}_n)\left[S_n(\hat{\lambda}_n) Y_n - X_n \hat{\beta}_n\right]$. Calculate $t_{\lambda} = \hat{\lambda}_n / Se(\hat{\lambda}_n)$, where $Se(\hat{\lambda}_n)$ is the estimated asymptotic standard errors of $\hat{\lambda}_n$.

(ii) Draw $\hat{e}_n^*$ from $\bar{e}_n$ through sampling with replacement. Generate $Y_n^* = S_n^{-1}(\hat{\lambda}_n) [X_n \hat{\beta}_n + R^{-1}(\hat{\rho}_n)\hat{e}_n^*]$.

(iii) Formulate an initial GMME with the bootstrap version of moment functions: $g_n^*(\theta) = (\hat{e}_n^*(\theta) P_1 \hat{e}_n^*(\theta), \ldots, \hat{e}_n^*(\theta) P_n \hat{e}_n^*(\theta), \hat{e}_n^*(\theta) Q_n)'$, where $P_1 = W_n' W_n - D(W_n' W_n)$, $P_2 = M_n' M_n - D(M_n' M_n)$, $P_3 = (W_n' W_n)^{(t)}$, $P_4 = (M_n' M_n)^{(t)}$, $Q_n = (W_n M_n X_n, W_n X_n, M_n X_n, X_n)$, and $e_n(\theta) = R_n(\rho) S_n(\lambda) Y_n - R_n(\rho) X_n \hat{\beta}$. Define $\hat{\phi}_1 = \text{argmin}_g \phi_n^*(\theta) g_n^*(\theta)$.

(iv) Using $\hat{\phi}_1$ and corresponding residuals, construct an estimate of $\Omega_n^*$ given in (3.1) and denote it with $\bar{\Omega}_n^*$. Using $\hat{\phi}_1$, generate the best set of moment functions $g_n^*(\theta)$. Define the bootstrap optimal GMME by $\hat{\theta}_{b2} = \text{argmin}_g g_n^*(\theta) \bar{\Omega}_n^* g_n^*(\theta)$.

(v) Using $\hat{\phi}_{b2}$ and corresponding residuals, find an estimate of $\Phi_n^*$ and denote it with $\hat{\Phi}_n^*$. Obtain an estimate of the variance of $\hat{\phi}_{b2}$ from $(\hat{\Phi}_n^* \bar{\Omega}_n^* \hat{\Phi}_n^*)^{-1}$. Obtain the bootstrap standard error of $\hat{\phi}_{b2}$ and denote it with $Se(\hat{\phi}_{b2})$. Then, calculate $t_{\phi} = (\hat{\phi}_{b2} - \hat{\phi}_2) / Se(\hat{\phi}_{b2})$.

(vi) Repeat steps (ii)–(v) $B$ times and calculate the bootstrap $p$-value of $p_{\phi} = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}(|t_{\phi}^b| > |t_{\phi}|)$.

As for a test of $H_0: \rho_0 = 0$, simply compute $t_{\phi}$ in step (i) and $t_{\phi}$ in step (vi), and evaluate these values similar to step (vi).

Given the result in (3.4), $t_\lambda$ has an asymptotic standard normal distribution, and hence the inference is based on this asymptotic distribution. On the other hand, the distribution function of $t_\phi$ obtained from the bootstrap method is an approximation to the exact finite sample distribution function of $t_\phi$ and can differ from the asymptotic distribution.\footnote{The Wald test is not the only test statistic that could be used for hypothesis testing. We refer the reader to Web Appendix C for the formulation of five other test statistics (the LM test, the $C(\alpha)$ test, the minimum chi-square test (MC), and two criterion-based tests ($D_{NW}$ and $D_{RU}$)), and we report on their performance in Web Appendix E.}

### 6. A Monte Carlo study

#### 6.1. Design

For the model specified in (2.1), we assume that there are are two regressors and no intercept term such that $X_n = [X_{1,n}, X_{2,n}]$ and $\beta_0 = (\beta_{10}, \beta_{20})' = (1, 1)'$, where $X_{1,n}$ and $X_{2,n}$ are $n \times 1$ vectors of variables. The $i$-th observation of disturbance term $e_{in}$ is i.i.d. Normal$(0, 1)$. The spatial autoregressive parameters $\lambda_0$ and $\rho_0$ take values from the set $\mathcal{X} = \{-0.6, -0.3, 0, 0.3, 0.6\}$ in order to allow for both weak and strong spatial interactions.

As for $X_{1,n}$ and $X_{2,n}$, we use the county-level data set of Pace and Barry (1997) on the 1980 presidential election. We set $X_{1,n}$ equal to the standardized value of income per-capita and $X_{2,n}$ equal to the
standardized value of the homeownership rate. The data set describes 3,107 U.S. counties, of which we use the first \(n\) observations in the Monte Carlo study.

The weight matrix \(W_n\) is based on the interaction scenario described in Arraiz et al. (2010). The observations are distributed across four quadrants of a space in such a way that the number of observations in each quadrant can be arranged to allow for sparse or dense quadrants. The location of each unit across the space is determined by the \(xy\)-coordinates over a square grid. Let \(m\) and \(\bar{m}\) be two integers. Then, the units in the northeast quadrant of the space have discrete coordinates satisfying \(m + 1 \leq x \leq \bar{m}\) and \(m + 1 \leq y \leq \bar{m}\), with an increment value of 0.5. For the other quadrants, the location coordinates are integers satisfying \(1 \leq x \leq m\), \(1 \leq y \leq m\), and \(1 \leq x \leq \bar{m}\), \(1 \leq y \leq m\). The distance \(d_{ij}\) between any two units \(i\) and \(j\), located respectively at \((x_1, y_1)\) and \((x_2, y_2)\), is measured by the Euclidean distance given by \(d_{ij} = [(x_1 - x_2)^2 + (y_1 - y_2)^2]^{1/2}\). Then, the \((i,j)\)th element of the weight matrix \(W_n\), \(w_{n,ij}\), equals 1 if \(0 \leq d_{ij} \leq 1\) and zero otherwise.

Varying the values for \(m\) and \(\bar{m}\) leads to a different sample size and a different share of units in the dense northeast quadrant. For our small samples, we consider combinations \((m, \bar{m}) = (5, 15)\) and \((m, \bar{m}) = (14, 20)\). The first combination produces a sample size of 486 and locates 75\% of units in the northeast quadrant, whereas the second combination generates a sample size of 485 and locates 25\% of the observation in the northeast quadrant. To generate large sample sizes, we specify \((m, \bar{m}) = (7, 21)\) and \((m, \bar{m}) = (20, 28)\). Here, the first combination generates a sample size of 974 with 75\% of observation located in the northeast quadrant, and the second one produces a sample size of 945 with 25\% units in the northeast quadrant. Let the percentage of units located in the northeast quadrant be denoted by \(d\). Then our Monte Carlo experiments are based on the combinations \((m, \bar{m}, n, d) = \{(5, 15, 486, 75\%), (14, 20, 485, 25\%), (7, 21, 974, 75\%), (20, 28, 945, 25\%)\}\). and we refer to these combinations by their \((n, d)\) values.

For the spatial weight matrix \(M_n\), we consider a distance-based binary matrix. Let \(J_i\) be the set of the nearest 10 observations to the \(i\)th observation based on \(d_{ij}\). Then, the \((i,j)\)th element of \(M_n\), \(m_{n,ij}\), equals 1 if \(j \in J_i\) and zero otherwise.

Both \(W_n\) and \(M_n\) are row normalized. For each specification, resampling is repeated 1,000 times. We examine the following estimators: (i) the IGMME, (ii) the OGMME, and (iii) the MLE. For each estimator, we provide bias, estimated asymptotic standard errors (OGMME-Asy.), and empirical standard errors (Emp.). Besides these results, we also provide the corrected standard errors (OGMME-Crc.) in the case of OGMME.

### 6.2. Estimation results

The results are easily interpretable when presented in graphical form, and in doing so, we highlight the standard error estimators. To this end, we compute the percentage deviation of the estimated asymptotic and corrected standard errors from the corresponding empirical standard deviations; i.e., we compute \(100 \times (\text{Asy.} - \text{Emp.})/\text{Emp.}\) and \(100 \times (\text{Crc.} - \text{Emp.})/\text{Emp.}\). We then plot these percentage deviations against 25 combinations of \((\lambda_0, \rho_0)\) separately for each \((n, d)\) value. Following Bond and Windmeijer (2005, p. 18), we use the term “bias” for these deviations. If the asymptotic distribution of an estimator performs well enough, the bias should be near zero. Conversely, a large bias suggests that, for the given estimator, the asymptotic distribution approximates the finite sample distribution poorly.

Figure 1 presents the results for \(\hat{\rho}_n\). There is a downward bias in the case of OGMME-Asy., OGMME-Crc., and MLE. The direction of the bias for the IGMME changes from positive to negative as the northeast quadrant density changes from low to high. The estimated asymptotic standard errors of IGMME can exceed the corresponding empirical standard errors by more than 10\%. In the case of the MLE, spikes for \(\rho_0 = -0.6\) result in a downward bias of up to 50\%. Except for these spikes, the estimated asymptotic standard errors of the MLE are relatively closer than those of OGMME to the corresponding

---

14Asy. and Crc. are the median values of their respective distributions. For the MLE, we use the analytical result for the covariance matrix to calculate standard errors.
Figure 1. Percentage deviations of the estimated standard errors from the empirical standard deviations: $\hat{\rho}_n$.

empirical standard deviations. In the case of OGMME, the correction method provides standard errors that are closer to the empirical standard deviations. In Fig. 1(b), where the density of the northeast quadrant is high, all estimators have a downward bias. The size of the bias is the smallest in the case of the correction method for all combinations of $(\lambda_0, \rho_0)$. In Figs. 1(c) and 1(d), where the results for the large samples are presented, the same pattern is also observed. That is, the standard errors based on the correction method are relatively closer to the corresponding empirical standard deviations.

The results for $\hat{\lambda}_n$ are presented in Fig. 2. In Figs. 2(a) and 2(b), the magnitude of the percentage deviations are similar for the OGMME-Asy. and the OGMME-Crc. In both figures, the bias is generally less than 5%, and it is relatively larger in Fig. 2(a). As can be seen from these figures, the correction method generally provides an improvement for the statistical inference as they are closer to the empirical standard deviations. Again for the combinations of autoregressive parameters where $\rho_0 = -0.6$, the MLE results display spikes, but this time with smaller amplitudes than for $\hat{\rho}_n$. The asymptotic method based on the IGMME provides standard error estimates that are biased by 25% or more. As the northeast quadrant density increases, biases become smaller. Biases are generally less than 5% in the case of OGMME-Asy., OGMME-Crc., and MLE for all $(\lambda_0, \rho_0)$ combinations.
We summarize the main findings of our simulation experiment in the following list with a focus on results reported for the OGMME-Crc.\textsuperscript{15}

(i) Across the various \((\lambda_0, \rho_0)\) combinations, the three estimators of \(\theta_0\) have a negligible bias.

(ii) The size of the northeast quadrant in our experiment affects the precision of estimators. When \(d\) is larger, the empirical standard deviations of the IGMME for \(\hat{\rho}_n\) and \(\hat{\lambda}_n\) are relatively larger. The MLE for \(\hat{\lambda}_n\), \(\hat{\beta}_1n\), and \(\hat{\beta}_2n\) has relatively smaller empirical standard deviations, and the standard errors of the OGMME-Crc. for \(\hat{\rho}_n\) is relatively smaller.

(iii) For \(\hat{\rho}_n\), when \(n = 485\) or \(n = 486\), our results indicate that the downward bias in the corrected standard errors is less than 5%, while the downward bias in the estimated asymptotic standard errors is generally larger than 5%. When \(n = 945\) or \(n = 974\), the corrected standard errors still perform better than the estimated asymptotic standard errors and again they are less than 5% in almost all cases. In order to give an overall picture, we compute for each estimator the proportions of percentage deviations that fall within a \((-5\%, 5\%)\) interval across all combinations of \((\lambda_0, \rho_0)\)

\textsuperscript{15}The results for all cases are provided in tables presented in Web Appendix D.
and \((n, d)\) values. These proportions are (i) 0.33 for the IGMME, (ii) 0.40 for the OGMME-Asy., (iii) 0.89 for the OGMME-Crc., and (iv) 0.67 for the MLE.

(iv) The corrected standard errors for \(\hat{\lambda}_n\) has relatively less downward bias than the estimated asymptotic standard errors. If we consider the proportions of percentage deviations that fall within a \((-5\%, 5\%)\) interval for each estimator, we have (i) 0.27 for the IGMME, (ii) 0.80 for the OGMME-Asy., (iii) 0.88 for the OGMME-Crc., and (iv) 0.84 for the MLE.

(v) For \(\hat{\beta}_{1n}\), both the OGMME-Asy. and the OGMME-Crc. have a downward bias of more than 5% in some cases. The proportion of percentage deviations that fall within a \((-5\%, 5\%)\) interval equals (i) 0.57 for the IGMME, (ii) 0.36 for the OGMME-Asy., (iii) 0.24 for the OGMME-Crc., and (iv) 0.93 for the MLE. It is obvious that the MLE is performing best in this case.

(vi) On \(\hat{\beta}_{2n}\), both the OGMME-Asy. and OGMME-Crc. provide similar results with a bias amount less than 5% in many cases. The proportion of percentage deviations that fall in a \((-5\%, 5\%)\) interval is (i) 0.48 for the IGMME, (ii) 0.83 for the OGMME-Asy., (iii) 0.95 for the OGMME-Crc., and (iv) 0.92 for the MLE.

### 6.3. Inference results

For each test, we consider the empirical rejection frequencies under the null hypothesis of the form \(H_0: \lambda_0 = 0\) or \(H_0: \rho_0 = 0\). We use \(p\)-value plots suggested by Davidson and MacKinnon (1998) to illustrate the actual size of these tests under the null hypothesis for various nominal size values. Figures 3 and 4 are the \(p\)-value plots for the null hypothesis of no spatial dependence, respectively, in the disturbance term and in the dependent variable. We present the results for the Wald tests and its bootstrap version for two representative cases and leave the results for other cases and other test statistics to Web Appendix E. In the figures, \(W_1, W_2, W_{2c}, W_m,\) and \(W_{2b}\) denote respectively the Wald tests based on the IGMME, OGMME-Asy, OGMME-Crc, MLE, and the bootstrap version.

Conventional significance levels in hypothesis tests are often chosen to be 10% or smaller. Therefore, we use nominal size values that are less than or equal to 12% in these figures. If the asymptotic approximation used for the distributions of these test statistics perform well, we would expect that the line graphs in these figures to follow closely the 45\(^\circ\) line. Hence, a discrepancy from the 45\(^\circ\) line indicates that the sampling distribution of the test statistic differs from the asymptotic distribution that is used to establish the critical values.

First, we evaluate properties of tests under the null of \(H_0: \rho_0 = 0\). For the null hypothesis under consideration, we only present the \(p\)-value plots for the case of \(\lambda_0 = 0.3\) in Fig. 3. The \(p\)-value plots for all other cases are provided in the web appendix. The salient properties of these tests are summarized as follows:

(i) The Wald test based on the IGMME (\(W_1\)) in Fig. 3 is not properly sized. It is undersized when the northeast density is low and is over-sized when the northeast density is high. The increase in the sample size does not improve the size property of \(W_1\) but rather makes it worse for the low density case.

(ii) The standard Wald test based on the OGMME-Asy (\(W_2\)) is over-sized in all cases which confirms previous findings in the literature on non-spatial econometric models. Therefore, inference based on this test could be very misleading. For example, at the nominal size of 6%, this test falsely reject the null in around 10% of the replications. This test performs worse than the \(W_m\) and \(W_{2c}\) tests in all cases. This result is principally due to the downward bias in the estimated asymptotic standard errors of the OGMME reported in Section 6.2.

(iii) For large sample sizes of \(n = 945\) and \(n = 974\), \(W_2\) seems to improve but is over-sized relative to \(W_m\) and \(W_{2c}\) tests. The size of the northeast quadrant appears to affect the performance of this test in small sample sizes.

(iv) The corrected version based on the OGMME-Crc (\(W_{2c}\)) outperforms \(W_2\) in all cases, which confirms our results reported in Section 6.2. Moreover, it performs as well as the MLE-based
Figure 3. Wald tests: $H_0: \rho_0 = 0$ (for $\lambda_0 = 0.3$).

Thus, the correction method improves the size properties of the Wald test, which is consistent with the simulation results presented in Section 6.2.

(v) The MLE based Wald test ($W_m$) has a proper size in all cases: the asymptotic distribution derived under the null hypothesis provides a good approximation to the finite sample distribution.

(vi) As the bootstrapped version of the Wald test ($W_{2b}$) is computationally intensive, we have computed results only for the OGMME and only for the cases where $n = 485$ and $n = 486$. The $W_{2b}$ test performs better than the $W_2$ test and is slightly over-sized for larger nominal size values. In comparison with the $W_m$ and $W_{2c}$ tests, the size is slightly worse.

(vii) The size properties of remaining tests, $LM$, $MC$, $D_{NW}$, $D_{RU}$, and $C(\alpha)$, are evaluated in detail in Web Appendix. Overall, the $D_{NW}$ and $D_{RU}$ tests perform better than the other tests. The $C(\alpha)$ test is properly sized in many cases and performs better than $LM$ and $MC$.

Next, we evaluate the size properties of these tests under the null of $H_0 : \lambda_0 = 0$. For the null hypothesis under consideration, we only present the $p$-value plots for the case of $\rho_0 = 0.3$ in Fig. 4.

---

16 For each resample, the number of bootstrap samples is limited to 99. This choice for the number of bootstrap samples satisfies the rule of thumb provided by Davidson and Mackinnon (1998): $(99 + 1) \times 0.01$ is a positive integer. A larger number of bootstrap samples is computationally challenging for the spatial model under consideration.

17 Here, $MC$ denotes the minimum chi-square test, and $D_{NW}$ and $D_{RU}$ denote criterion based tests. For details, see Web Appendix C.
The $p$-value plots for all other cases are provided in the Web Appendix. The salient properties of these tests are as follows:

(i) The $W_1$ test is oversized in all cases regardless of the density of the northeast quadrant and the sample size. For example, in Fig. 4(d), when the nominal size is 6% this test rejects the true null more than 10% of the replications. As shown in Figs. 4(a) to 4(d), the rejection frequencies do not converge to the true ones as the sample size increases.

(ii) The standard Wald test $W_2$ is slightly oversized at low nominal sizes when $d = 0.25$ and more generally so when $d = 0.75$.

(iii) The corrected version $W_{2c}$ achieves the correct sizes implied by the asymptotic theory in small samples. In large samples, it is slightly undersized when $d = 0.25$ and slightly oversized when $d = 0.75$.

(iv) The results for the bootstrapped version are again provided only for the Wald test with the OGMME when $n = 485$ and $n = 486$. As seen, this test is over-sized and its deviation from the $45^\circ$ line becomes larger for the large nominal size values. For example, for the nominal sizes of 6% and 10% the test statistic produces actual sizes of 9% and 11%, respectively in Fig. 4(a). Note, though, that the number of bootstrap samples is limited to only 99: this evidence may yet be subject to random fluctuations in the resampling process.
The MLE based Wald test is undersized when \((n, d) = (485, 25\%)\) and is slightly over-sized when \((n, d) = (945, 25\%)\) for larger nominal sizes. As the northeast quadrant density increases, the actual sizes converge to the nominal sizes.

The size properties of remaining tests are examined in the Web Appendix. Both LM and MC tests are over-sized in all cases and are not recommended for applied research to test this null hypothesis. The two criterion based tests, \(D_{NW}\) and \(D_{RU}\), are properly sized only when there exist negative spatial dependence in the disturbance term. The \(p\)-value plots for the \(C(\alpha)\) test are close to the \(45^\circ\) line in many cases, indicating that this test is properly sized.

### 6.4. Results on the effects estimators

We use the same Monte Carlo setup described in Section 6.1 to study the finite sample properties of the estimators of the impact measures defined in Section 4. For the sake of brevity, we consider only \((n, d) = \{(486, 75\%), (485, 46\%), (485, 25\%), (974, 75\%), (945, 46\%), (945, 25\%)\}). Besides the IGGME, OGMME, and MLE, we also consider the performance of the multi-step estimator (GS2SLSE) of Drukker et al. (2013) in this analysis.\(^{18}\) For each impact measure, we provide the bias, the estimated

---

\(^{18}\)Drukker et al. (2013) extend the GMM/IV estimation method of Kelejian and Prucha (1998) to a SARAR(1,1) specification that includes endogenous regressors. Although they study the finite sample properties of estimators and the Wald tests, the performance of impact measures is not explored. For our discussion of their estimators, see Web Appendix F.
asymptotic standard errors, the empirical standard deviations, and the rejection rates of Wald test for a nominal size value of 0.05.

The simulation results are given in 36 tables presented in Web Appendix G. Here, we summarize these results by providing scalar measures. We start by comparing the bias properties of effect estimates across estimators. To evaluate the performance of estimators in terms of bias, we calculate the average absolute bias for each impact measure over 25 combinations of \((\lambda_0, \rho_0)\). These results are given in Table 1. In general, the average absolute bias reported by each estimator is less than 0.05 for each impact measure. The IGMME and GS2SLE impose slightly larger bias on impact measures. Table 1 also indicates that the Delta method performs slightly better than the simulation method. The magnitude of bias is slightly larger when the density of the northeast quadrant is high.

Next, we compare the estimated standard errors with the empirical standard deviations for each estimator. For the total effects, we provide the percentage deviations of the estimated standard errors from the empirical standard deviations in Fig. 5 for the case of \((n, d) = (486, 75\%)\). Generally, the percentage bias falls within the \((-5\%, 5\%)\) interval for each estimator, except in the case of IGMME. To give an overall picture, we calculate the proportions of percentage deviations that fall within \((-5\%, 5\%)\)

![Graphs showing percentage deviations](image)

**Figure 5.** Total effects: Percentage deviations of the estimated standard errors from the empirical standard deviations for \(n = 486, d = 75\%).
across all combinations of \((\lambda_0, \rho_0)\) and \((n, d)\). For the simulation method, these proportions for the total effects of \(X_{1,n}\) are 0.48 for the IGMME, 0.746 for the OGMME-Asym., 0.68 for the OGMME-Crc., 0.94 for the MLE, and 0.933 for the GS2SLS, and for \(X_{2,n}\) these rates are 0.526 for the IGMME, 0.80 for the OGMME-Asym., 0.88 for the OGMME-Crc., 0.953 for the MLE, and 0.933 for the GS2SLS. For the Delta method, these proportions for the total effects of \(X_{1,n}\) are 0.453 for the IGMME, 0.70 for the OGMME-Asym., 0.74 for the OGMME-Crc., 0.913 for the MLE, and 0.926 for the GS2SLS, and for \(X_{2,n}\) these rates are 0.506 for the IGMME, 0.873 for the OGMME-Asym., 0.93 for the OGMME-Crc., 0.92 for the MLE, and 0.94 for the GS2SLS. While the IGMME performs relatively poorly, the MLE and the GS2SLS perform relatively better. In general, the OGMME-Crc performs better than OGMME-Asy, which is consistent with our results in Sections 6.2 and 6.3. The simulation method seems to be performing slightly better than the Delta method in many cases. The standard errors of impact estimators seem to be affected by the density of northeast quadrant, and the proportions of percentage deviations that falls in the \((-5\%, 5\%)\) interval are relatively higher for the case of \((n, d) = (485, 46%)\).

**Table 2.** Average size distortions of the Wald tests for impact measures.

|               | Total                | Direct                | Indirect               |
|---------------|----------------------|-----------------------|------------------------|
|               | Simulation Delta     | Simulation Delta      | Simulation Delta       |
| \(X_{1,n}\)   |                      | \(X_{2,n}\)           | \(X_{1,n}\)            |
| IGMME         | 0.0187               | 0.0216                | 0.0266                 |
| OGMME_Asy.    | 0.0132               | 0.0092                | 0.0165                 |
| OGMME_Crc.    | 0.0065               | 0.0061                | 0.0068                 |
| MLE           | 0.0082               | 0.0085                | 0.0100                 |
| GS2SLS        | 0.0054               | 0.0059                | 0.0064                 |
| \(X_{2,n}\)   |                      |                       |                        |
| IGMME         | 0.0218               | 0.0239                | 0.0219                 |
| OGMME_Asy.    | 0.0159               | 0.0436                | 0.0067                 |
| OGMME_Crc.    | 0.0121               | 0.0051                | 0.0063                 |
| MLE           | 0.0109               | 0.0122                | 0.0114                 |
| GS2SLS        | 0.0060               | 0.0063                | 0.0076                 |
| \(n = 486, d = 75\%) |                |                       |                        |
| IGMME         | 0.0222               | 0.0159                | 0.0192                 |
| OGMME_Asy.    | 0.0118               | 0.0078                | 0.0143                 |
| OGMME_Crc.    | 0.0079               | 0.0051                | 0.0076                 |
| MLE           | 0.0109               | 0.0092                | 0.0122                 |
| GS2SLS        | 0.0060               | 0.0063                | 0.0076                 |
| \(n = 485, d = 46\%) |                |                       |                        |
| IGMME         | 0.0212               | 0.0176                | 0.0218                 |
| OGMME_Asy.    | 0.0103               | 0.0069                | 0.0108                 |
| OGMME_Crc.    | 0.0056               | 0.0062                | 0.0058                 |
| MLE           | 0.0065               | 0.0052                | 0.0069                 |
| GS2SLS        | 0.0045               | 0.0051                | 0.0048                 |
| \(n = 485, d = 25\%) |                |                       |                        |
| IGMME         | 0.0265               | 0.0160                | 0.0268                 |
| OGMME_Asy.    | 0.0137               | 0.0091                | 0.0137                 |
| OGMME_Crc.    | 0.0074               | 0.0059                | 0.0075                 |
| MLE           | 0.0068               | 0.0074                | 0.0083                 |
| GS2SLS        | 0.0064               | 0.0069                | 0.0067                 |
| \(n = 974, d = 75\%) |                |                       |                        |
| IGMME         | 0.0265               | 0.0160                | 0.0268                 |
| OGMME_Asy.    | 0.0137               | 0.0091                | 0.0137                 |
| OGMME_Crc.    | 0.0074               | 0.0059                | 0.0075                 |
| MLE           | 0.0068               | 0.0074                | 0.0083                 |
| GS2SLS        | 0.0064               | 0.0069                | 0.0067                 |
| \(n = 945, d = 47\%) |                |                       |                        |
| IGMME         | 0.0194               | 0.0156                | 0.0189                 |
| OGMME_Asy.    | 0.0095               | 0.0082                | 0.0100                 |
| OGMME_Crc.    | 0.0050               | 0.0070                | 0.0049                 |
| MLE           | 0.0054               | 0.0060                | 0.0055                 |
| GS2SLS        | 0.0042               | 0.0045                | 0.0044                 |
In Table 2, we report the average size distortions of the Wald tests over 25 combinations of \((\lambda_0, \rho_0)\) for each impact measure for a nominal size of 0.05.19

The average size distortions are smaller than 0.03 in all cases, and the Wald tests based on IGMME have relatively larger average size distortions. Overall, the average size distortions of the Wald tests based on OGMME-Crc are smaller than those based on OGMME-Asy, which suggests that the correction method can lead to more accurate inference.

7. Conclusion

In this study, we review the GMM estimation approach for spatial autoregressive models and evaluate the finite sample properties of the GMM estimators through simulation studies. To the best of our knowledge, this study is the first extensive study to evaluate various GMM inference techniques for spatial autoregressive models. We extend the finite sample correction method of Windmeijer (2005) to the SARAR(1,1) model and formulate a variance formula to improve inference properties of the GMM estimator for detecting the presence of spatial dependence. We show that the corrected standard errors can be a good alternative to their estimated asymptotic counterparts.

Simulation studies confirm that the estimated asymptotic standard errors based on the OGMME can be substantially downward biased and can make inference problematic for the spatial autoregressive parameters. This result is consistent with the simulation results reported in the econometrics literature on the nonspatial models. More specifically, we evaluate the percentage deviations of the corrected standard errors from the empirical standard deviations under various scenarios. The results show that the bias in the corrected standard errors of OGMME and the effects estimators (or the impacts estimators) is mostly less than the bias in the estimated asymptotic standard errors of OGMME and the effects estimators. Hence, the correction method can provide more accurate inference in finite samples.

For the SARAR(1,1) model, we also compare the size properties of various test statistics through simulations. These tests included standard asymptotic Wald tests based on MLE and GMMEs, a bootstrapped version of Wald test, two versions of \(C(\alpha)\) test, the standard LM test, the minimum chi-square test, and two versions of GMM criterion test. Our results show that the standard asymptotic Wald test based on the asymptotic standard errors of the OGMME is generally oversized, whereas the Wald test based on the corrected standard errors of the OGMME is properly sized. Our findings also confirm that the GMM criterion tests and the \(C(\alpha)\) test can be useful for detecting the presence spatial dependence. All of these findings offer guidance to applied researchers who estimate and test spatial models with the GMM estimators.

Acknowledgments

We thank the editor, an associate editor, two anonymous referees and the conference participants at SEA2015 (Miami) and Camp Econometrics X (New York).

Funding

This research was supported, in part, by a grant of computer time from the City University of New York High Performance Computing Center under NSF Grants CNS-0855217 and CNS-0958379.

---

19We calculate the average size distortions over 25 combinations of \((\lambda_0, \rho_0)\) from 36 tables given in Web Appendix G by \(\frac{1}{25} \sum_{i=1}^{25} \left| \text{actual size}_i - 0.05 \right|\), where \(i\) corresponds to the \(i\)th combinations of \((\lambda_0, \rho_0)\).
References

Altonji, J. G., Segal, L. M. (1996). Small-sample bias in gmm estimation of covariance structures. *Journal of Business & Economic Statistics* 14(3):353–366.

Anselin, L. (1988). *Spatial econometrics: Methods and Models*. New York: Springer.

Arellano, M., Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies* 58(2):277–297.

Arroyo, I. et al. (2010). A spatial Cliff-Ord-type model with heteroskedastic innovations: Small and large sample results. *Journal of Regional Science* 50(2):592–614.

Blundell, R., Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87(1):115–143.

Bond, S., Windmeijer, F. (2005). Reliable inference for gmm estimators? finite sample properties of alternative test procedures in linear panel data models. *Econometric Reviews* 24(1):1–37.

Brown, B. W., Newey, W. K. (2002). Generalized method of moments, efficient bootstrapping, and improved inference. *Journal of Business & Economic Statistics* 20(4):507–517.

Burridge, P., Fingleton, B. (2010). Bootstrap inference in spatial econometrics: The J-test. *Spatial Economic Analysis* 5(1):93–119.

Clark, T. E. (1996). Small-sample properties of estimators of nonlinear models of covariance structure. *Journal of Business & Economic Statistics* 14(3):367–373.

Das, D., Kelejian, H. H., Prucha, I. R. (2003). Small sample properties of estimators of spatial autoregressive models with autoregressive disturbances. *Papers in Regional Science* 82:1–26.

Davidson, R., MacKinnon, J. G. (1998). Graphical methods for investigating the size and power of hypothesis tests. *The Manchester School* 66(1):1–26.

Debarsy, N., Jin, F., Lee, L. f. (2015). Large sample properties of the matrix exponential spatial specification with an application to FDI. *Journal of Econometrics* 188(1):1–21.

Drukker, D. M., Egger, P., Prucha, I. R. (2013). On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors. *Econometric Reviews* 32(5–6):686–733.

Fingleton, B. (2009). A generalized method of moments estimator for a spatial model with moving average errors, with application to real estate prices. In: Arbia, G., Baltagi, B. H., eds. *Spatial Econometrics*. Studies in Empirical Economics, pp. 35–57.

Hall, P., Horowitz, J. L. (1996). Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica* 64(4):891–916.

Hansen, L. P., Heaton, J., Yaron, A. (1996). Finite-sample properties of some alternative gmm estimators. *Journal of Business & Economic Statistics* 14(3):262–280.

Jin, F., Lee, L. f. (2013). Cox-type tests for competing spatial autoregressive models with spatial autoregressive disturbances. *Regional Science and Urban Economics* 43(4):590–616.

Jin, F., Lee, L. f. (2015). On the bootstrap for moran's test for spatial dependence. *Journal of Econometrics* 184(2):295–314.

Kelejian, H. H., Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics* 17(1):1899–1926.

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.

Lee, L.-f. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews* 22(4):307–335.

Lee, L.-f. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72(6):1899–1925.

Lee, L.-f. (2007a). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* 137(2):489–514.

Lee, L.-f. (2007b). The method of elimination and substitution in the GMM estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* 140:155–189.

Lee, L.-f., Liu, X. (2010). Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. *Econometric Theory*, 26(1):187–230.

LeSage, J., Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Statistics: A Series of Textbooks and Monographs. London: Chapman and Hall/CRC.

Lin, X., Lee, L.-f. (2010). GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics* 157(1):34–52.

Liu, X., Lee, L.-f., Bollinger, C. R. (2010). An efficient GMM estimator of spatial autoregressive models. *Journal of Econometrics* 159(2):303–319.

Newey, W. K., McFadden, D. (1994). *Large Sample Estimation and Hypothesis Testing*. In: Engle, R. F., McFadden, D. L., eds. Handbook of Econometrics, Vol. 4. Elsevier, pp. 2111–2245.

Pace, R. K., Barry, R. (1997). Quick computation of spatial autoregressive estimators. *Geographical Analysis* 29(3):232–247.
Pinkse, J., Slade, M. E. (1998). Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics* 85(1):125–154.

Prucha, I. R. (2014). Instrumental variables/method of moments estimation. In: Fischer, M. M., Nijkamp, P., eds. *Handbook of Regional Science*. Berlin, Heidelberg: Springer, pp. 1597–1617.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics* 126(1):25–51.

Windmeijer, F. (2008). GMM for panel data count models. In: Mátyás, L., Sevestre, P., eds. *The Econometrics of Panel Data*. Advanced Studies in Theoretical and Applied Econometrics, Vol. 46. Berlin, Heidelberg: Springer, pp. 603–624.