Heteroskedasticity Consistent Covariance Matrix Estimators for the GMME of Spatial Autoregressive Models*

Süleyman Taşpınar† Osman Doğan‡

June 11, 2016

Abstract

In the presence of heteroskedasticity, the conventional test statistics, based on the ordinary least square estimator, lead to incorrect inference results in the linear regression model. Given that heteroskedasticity is common in cross-sectional data, the test statistics based on various forms of heteroskedasticity consistent covariance matrices (HCCMs) have been developed in the literature. Heteroskedasticity is a more serious problem for spatial econometric models, generally causing inconsistent estimators. We investigate the finite sample properties of a heteroskedasticity robust generalized method of moments estimator for a spatial econometric model with an unknown form of hetereoskedasticity. We develop various HCCM-type corrections to improve the finite sample properties of the GMME and the conventional Wald test. Our Monte Carlo experiments indicate that the HCCM-type corrections produce more accurate inference results for the model parameters and the effects estimates.

JEL-Classification: C13, C21, C31.
Keywords: Spatial Autoregressive Models, SARAR, GMM, Heteroskedasticity, HCCME, Asymptotic Variance, Efficiency, Inference.

---

*We would like to thank the conference participants at 26^{th} $(\text{EC})^2$ Conference on Theory and Practice of Spatial Econometrics at Heriot-Watt University, Edinburgh, UK for helpful comments. 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.

†Economics Program, Queens College, The City University of New York, United States, email: STaspinar@qc.cuny.edu.

‡Project Department, Istanbul Ulasim A.S., Istanbul, Turkey, email: ODogan10@gmail.com.

1 Introduction

An unknown form of heteroskedasticity in the disturbance terms of a spatial autoregressive model can yield inconsistent extremum estimators. The robust generalized method of moments estimators (GMMEs) proposed by Kelejian and Prucha, (2010), Lin and Lee, (2010) and Debarsy et al., (2015) have the virtue of being consistent under both heteroskedasticity and homoskedasticity. Despite this desirable property, these estimators are inefficient as the best set of moment functions is generally not available when the model involves an unknown form of heteroskedasticity. Furthermore, there is not much known on inference based on these estimators in finite samples. An exception is Kelejian and Prucha, (2010) who provide results on the size properties of the standard Wald test based on their multi-step estimator. It remains open to investigate the properties of the robust estimation approach in terms of inference in finite samples. To this end, we consider an SARAR$(1, 1)$ model with an unknown form of heteroskedasticity in this study.

First, we revisit the estimation approach of Lin and Lee, (2010) for our SARAR$(1, 1)$ specification and investigate the form of the best set of moment functions following the idea in Lee, (2007). Our findings are in line with the findings of Debarsy et al., (2015). The best GMM estimator formulated from the best set of moment functions is not feasible as these moments involve an unknown covariance matrix that cannot be estimated consistently. More importantly, our main objective is to derive heteroskedasticity consistent covariance matrix (HCCM)-type corrections for the robust GMME. To this end, we suggest various HCCM estimators (HCCMEs) based on two quasi hat matrices and investigate their effects on the finite sample properties of the robust GMME as well as on the finite sample properties of the Wald test.

Originally suggested by Eicker, (1967) and White, (1980), HCCMEs are common tools to improve finite sample properties of the conventional tests of significance in linear regression models and generalized estimating equations (Bera et al., 2002; Cribari-Neto, 2004; Cribari-Neto et al., 2007; Kauermann and Carroll, 2001; Long and Ervin, 2000; MacKinnon and White, 1985). It has been well documented in the literature that the Wald test based on the original HCCME suggested in White, (1980) has serious size distortions. Therefore, various modifications to the original HCCME have been proposed over the years. MacKinnon and White, (1985) suggest alternative HCCMEs formulated from the leverage-adjusted residuals. Chesher and Jewitt, (1987), Chesher, (1989), Chesher and Austin, (1991) and Kauermann and Carroll, (2001) indicate that the standard Wald tests based on the HCCMEs suggested in MacKinnon and White, (1985) can still have poor finite sample properties when there are high leverage points in the design matrix. Cribari-Neto, (2004) and Cribari-Neto et al., (2007), therefore, propose modified HCCMEs to remove the effect of high leverage points. For a comprehensive review, see MacKinnon, (2013).

Lin and Chou, (2015) (LC hereafter) complement the literature by providing a methodology to formulate HCCMEs based on leverage-adjusted residuals within the GMM framework for non-linear regression models. Our contribution is extending LC's methodology to a spatial autoregressive model with an unknown form of heteroskedasticity to formulate various HCCMEs within the GMM framework. This extension is not straightforward mainly due to two complications arising from the spatial dependence in our model. First, our set of moments involve moment functions that are linear and quadratic in disturbance terms, whereas the set of moments in LC contains only linear moment functions. The presence of quadratic moment functions complicates the formulation of a hat matrix. Second, LC extend the idea of the leverage adjusted-residuals in MacKinnon and White, (1985) to a non-linear regression model. In essence, various HCCMs are based on a relationship derived at the observational level between the leverage-adjusted residuals and the individual variance under homoskedasticity assumption. In the presence of spatial dependence, such a relationship can not be established at the observational level. Instead, it has to be established at the sample level which complicates the derivation of a hat matrix.

In a simulation study, we investigate the finite sample properties of the GMME based on various finite sample correction methods formulated from two (quasi) hat matrices for a SARAR$(1,1)$ specification. These correction methods affect both the bias and the estimated standard errors of the GMME in finite samples. Our simulation results show that the bias properties of the GMME are similar across the correction methods. That is, the GMME formulated from each of the suggested correction method produce similar point estimates in finite samples. However, our results show that the estimated standard errors of the GMME are quite different across the correction methods. Especially, we show that the usual estimated standard errors (formulated from $SHC0$) differ from the empirical counterpart substantially, which in turn results in large size distortions for the standard Wald test. Our results indicate that the estimated standard error based on the correction methods are much closer to their empirical counterparts, and hence can lead to more accurate inference within the context of our spatial model.

This paper is organized in the following way. Section 2 presents the spatial autoregressive model, underlying assumptions and reviews the robust GMM estimation approach to lay out the details of the estimation approach for the SARAR$(1,1)$ specification. Section 3 deals with various methods of heteroskedasticity-consistent covariance matrix estimation in the GMM framework. Section 4 presents details of the derivation of the quasi-hat matrix. Section 5 lays out the details of the Monte Carlo design and presents the results. Section 6 closes with concluding remarks. Some of the technical derivations are relegated to an appendix.

2 SARAR(1,1) specification, assumptions and the robust GMME

Using the standard notation, the SARAR$(1,1)$ specification is given by

$$Y_n = \lambda_0 W_n Y_n + X_n\beta_0 + u_n, \quad u_n = \rho_0 M_n u_n + \varepsilon_n, \tag{2.1}$$

where $Y_n = (Y_{1n}, \ldots, Y_{nn})^{'}$ 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 matching parameter vector β_0. Furthermore, W_n and M_n are the $n \times n$ spatial weight matrices of known constants with zero diagonal elements, λ_0 and ρ_0 are the spatial autoregressive parameters, $u_n = (u_{1n}, \ldots, u_{nn})^{'}$ is the $n \times 1$ vector of regression disturbance terms and $\varepsilon_n = (\varepsilon_{1n}, \ldots, \varepsilon_{nn})^{'}$ is the $n \times 1$ vector of disturbances (or innovations). Let Θ be the parameter space of the model. In order to distinguish the true parameter vector from other possible values in Θ, we state the model with the true parameter vector $\theta_0 = (\rho_0, \lambda_0, \beta_0^{'})^{'}$. Furthermore, for notational simplicity, we let $S_n(\lambda) = (I_n - \lambda W_n)$, $R_n(\rho) = (I_n - \rho M_n)$, $G_n(\lambda) = W_n S_n^{-1}(\lambda)$, $H_n(\rho) = M_n R_n^{-1}(\rho)$, $\overline{G}_n(\rho, \lambda) = R_n(\rho) G_n(\lambda) R_n^{-1}(\rho)$ and $\overline{X}_n(\rho) = R_n(\rho) X_n$. Also, at (ρ_0, λ_0), we denote $S_n(\lambda_0) = S_n$, $R_n(\rho_0) = R_n$, $G_n(\lambda_0) = G_n$, $H_n(\rho_0) = H_n$, $\overline{G}_n(\rho_0, \lambda_0) = \overline{G}_n$ and $\overline{X}_n(\rho_0) = \overline{X}_n$.

We maintain Assumption 1 and 2 with respect to innovations and weight matrices.

Assumption 1. — The innovations ε_{in}s are distributed independently, and satisfy $\mathrm{E}(\varepsilon_{in}) = 0$, $\mathrm{E}\left(\varepsilon_{in}^2\right) = \sigma_{in}^2$, and $\mathrm{E}\,|\varepsilon_{in}|^{4+\eta} < \infty$ for some $\eta > 0$ 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}, R_n^{-1}, $S_n^{-1}(\lambda)$ and $R_n^{-1}(\rho)$ exist and are uniformly bounded in row and column sums in absolute value for all values of ρ and λ 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, denoted respectively by μ_3 and μ_4, of ε_{in} exist for all i and n. Assumption 2 also implies that the model in (2.1) represents an equilibrium relation for the dependent variable, that is, $Y_n = S_n^{-1}X_n\beta_0 + S_n^{-1}R_n^{-1}\varepsilon_n$.

For the model in (2.1), we consider a GMME based on a combination of linear and quadratic moment functions (Lee, 2007; Lin and Lee, 2010). The combined vector of moment functions is given by $g_n(\theta_0) = \left(\varepsilon_n^{'}P_{1n}\varepsilon_n, \ldots, \varepsilon_n^{'}P_{mn}\varepsilon_n, \varepsilon_n^{'}Q_n\right)^{'}$. Moment functions formulated with the $n \times n$ constant matrices P_{jn} for $j = 1, \ldots, m$ are called the quadratic moment functions. The remaining moment function $Q_n^{'}\varepsilon_n$ 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 diagonal elements. The quadratic moment functions formulated with matrices from $\mathcal{P}_n$ satisfy the orthogonality conditions when disturbance terms are independent.

In the following, Assumptions 3 and 4 states regularity conditions for moment matrices and regressors. Assumption 5 characterizes the parameter space.[1]

Assumption 3. — 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 4. — The regressors matrix X_n is an $n \times k$ matrix consisting of uniformly bounded constant elements. It has full column rank. Moreover, $\lim_{n\to\infty} \frac{1}{n}X_n^{'}X_n$ exists and is nonsingular.

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

The GMME is obtained by exploiting the sample moment counterparts of population moment conditions implied by the model specification. For our specification, the GMME is based on a set of quadratic and linear moment functions formulated from the orthogonality conditions implied by $R_nY_n = \lambda_0R_nW_nY_n + R_nX_n\beta_0 + \varepsilon_n = Z_n\delta_0 + \varepsilon_n$, where $Z_n = (R_nW_nY_n,\, R_nX_n)$ and $\delta_0 = \left(\lambda_0,\, \beta_0^{'}\right)^{'}$. The linear moment matrix Q_n is constructed from the expectation of $Z_n = \left(R_nW_nY_n,\, R_nX_n\right)$, and implies the population moment function of $Q_n^{'}\varepsilon_n$. The quadratic moment functions are formulated to exploit the information in the stochastic part of Z_n, which can be written as $R_nW_nY_n = R_nG_nX_n\beta_0 + R_nG_nR_n^{-1}\varepsilon_n$. The stochastic variables, denoted by $P_{jn}\varepsilon_n$ for $i = 1, \ldots, m$, are used to instrument the stochastic part $R_nG_nR_n^{-1}\varepsilon_n$ of $R_nW_nY_n$, which produce the quadratic moment functions $\varepsilon_n^{'}P_{jn}\varepsilon_n$. Hence, we have the following vector of moment functions $g_n(\theta_0) = \left(\varepsilon_n^{'}P_{1n}\varepsilon_n, \ldots, \varepsilon_n^{'}P_{mn}\varepsilon_n, \varepsilon_n^{'}Q_n\right)^{'}$ for the GMM estimation.

It proves helpful to introduce the following notation. Let $A^{(s)} = A_n + A_n^{'}$ for any matrix A_n. We denote the (i,j)th element, the ith row and jth column of A_n, respectively, by $A_{ij,n}$, $A_{i\bullet,n}$ and $A_{\bullet j,n}$. Hence, $A_{ij,n}^{(s)} = \left(A_{ij,n} + A_{ji,n}\right)$, $A_{i\bullet,n}^{(s)} = \left(A_{i\bullet,n} + A_{\bullet i,n}^{'}\right)$ and $A_{\bullet j,n}^{(s)} = \left(A_{\bullet j,n} + A_{j\bullet,n}^{'}\right)$. Also note that $A_{i\bullet,n}^{(s)} = A_{\bullet i,n}^{(s)'}$. Let $\mathrm{D}(\cdot)$ be a matrix operator that creates a matrix from the diagonal elements of an input matrix, and $\mathrm{vec}_D(\cdot)$ be a vector operator that returns a vector from the diagonal elements of an input matrix. We will denote $\mathrm{D}(\sigma_{1n}^2, \ldots, \sigma_{nn}^2)$ by Σ_n, which is the covariance matrix of the disturbance terms. Furthermore, let $\Omega_n = \mathrm{E}\left[g_n(\theta_0)g_n^{'}(\theta_0)\right]$ and $\Phi_n = \mathrm{E}\left[\partial g_n(\theta_0)/\partial\theta^{'}\right]$, which are functions of Σ_n.[2] Under our assumptions, we have $\frac{1}{n}\Omega_n = O(1)$ and $\frac{1}{n}\Phi_n = O(1)$. Let $\widehat{\varepsilon}_{in}$ be the ith residual of the model based on a consistent initial estimator $\widehat{\theta}_{1n}$ of θ_0, and let $\widehat{\Sigma}_n$ denote $\mathrm{D}(\widehat{\varepsilon}_{in}^2, \ldots, \widehat{\varepsilon}_{nn}^2)$. When Σ_n in Ω_n and Φ_n is replaced by $\widehat{\Sigma}_n$, the resulting matrices are denoted by $\widehat{\Omega}_n$ and $\widehat{\Phi}_n$, respectively. It can be shown that $\frac{1}{n}\widehat{\Omega}_n = \frac{1}{n}\Omega_n + o_p(1)$ and $\frac{1}{n}\widehat{\Phi}_n = \frac{1}{n}\Phi_n + o_p(1)$. Let $\widehat{\theta}_{1n}$ be an initial robust GMME (IRGMME) and $\widehat{\Omega}_{1n}$ be the estimate of Ω_n recovered from $\widehat{\theta}_{1n}$. Then, the optimal robust GMME (ORGMME) is given by $\widehat{\theta}_{2n} = \mathrm{argmin}_{\theta\in\Theta}\, g_n^{'}(\theta)\,\widehat{\Omega}_{1n}^{-1}\,g_n(\theta)$ and

[1] See Kelejian and Prucha, (2010) for the specification of the parameter space of autoregressive parameters.

[2] See Appendix C for their explicit forms.

furthermore it can be shown that[3]

$$\sqrt{n}\big(\widehat{\theta}_{2n}-\theta_0\big)\xrightarrow{d} N\Big(0_{(k+2)\times 1},\ \big[\lim_{n\to\infty}\frac{1}{n}\Phi_n^{'}\Omega_n^{-1}\Phi_n\big]^{-1}\Big). \tag{2.2}$$

An estimate of the variance-covariance matrix of $\sqrt{n}\big(\widehat{\theta}_{2n}-\theta_0\big)$ can be formulated from $\big[\frac{1}{n}\,\widehat{\Phi}_{2n}^{'}\,\widehat{\Omega}_{1n}^{-1}\,\widehat{\Phi}_{2n}\big]^{-1}$ where $\widehat{\Phi}_{2n}$ is an estimate of Φ_n recovered from $\widehat{\theta}_{2n}$.

The result in (2.2) indicates that the asymptotic efficiency of the GMME should be considered for the selection of the moment functions. As stated, the linear IVs are based on the expectation of $Z_n=[R_nW_nY_n,\ R_nX_n]$. Hence, the best IV matrix is given by $Q_n=\mathrm{E}(Z_n)=[R_nG_nX_n\beta_0,\ R_nX_n]$ (Lee, 2003). Selection of P_{jn}s in $\mathcal{P}_n$ can be made by investigating an upper bound for $\big[\Phi_n^{'}\Omega_n^{-1}\Phi_n\big]$. To this end, we can write

$$\Phi_n^{'}\Omega_n^{-1}\Phi_n=\begin{bmatrix}\mathcal{B}_{1\times 1} & \mathcal{D}_{1\times 1} & 0_{1\times k}\\ \mathcal{D}_{1\times 1}^{'} & \mathcal{G}_{1\times 1} & 0_{1\times k}\\ 0_{k\times 1} & 0_{k\times 1} & 0_{k\times k}\end{bmatrix} \tag{2.3}$$
$$+\begin{bmatrix}0_{1\times 1} & 0_{1\times 1} & 0_{1\times k}\\ 0_{1\times 1}^{'} & \beta_0^{'}\overline{X}_n^{'}\overline{G}_n^{'}Q_n\big(Q_n^{'}\Sigma_nQ_n\big)^{-1}Q_n^{'}\overline{G}_n\overline{X}_n\beta_0 & \beta_0^{'}\overline{X}_n^{'}\overline{G}_n^{'}Q_n\big(Q_n^{'}\Sigma_nQ_n\big)^{-1}Q_n^{'}\overline{X}_n\\ 0_{k\times 1} & \overline{X}_n^{'}Q_n\big(Q_n^{'}\Sigma_nQ_n\big)^{-1}Q_n^{'}\overline{G}_n\overline{X}_n\beta_0 & \overline{X}_n^{'}Q_n\big(Q_n^{'}\Sigma_nQ_n\big)^{-1}Q_n^{'}\overline{X}_n\end{bmatrix}$$

where $\mathcal{B}=\big[\mathrm{tr}\big(\Sigma_nH_n^{'}P_{1n}^{(s)}\big),\ldots,\mathrm{tr}\big(\Sigma_nH_n^{'}P_{mn}^{(s)}\big)\big]\mathcal{A}_n^{-1}\big[\mathrm{tr}\big(\Sigma_nH_n^{'}P_{1n}^{(s)}\big),\ldots,\mathrm{tr}\big(\Sigma_nH_n^{'}P_{mn}^{(s)}\big)\big]^{'}$, $\mathcal{G}=\big[\mathrm{tr}\big(\Sigma_n\overline{G}_n^{'}P_{1n}^{(s)}\big),\ldots,\mathrm{tr}\big(\Sigma_n\overline{G}_n^{'}P_{mn}^{(s)}\big)\big]\mathcal{A}_n^{-1}\big[\mathrm{tr}\big(\Sigma_n\overline{G}_n^{'}P_{1n}^{(s)}\big),\ldots,\mathrm{tr}\big(\Sigma_n\overline{G}_n^{'}P_{mn}^{(s)}\big)\big]^{'}$, $\mathcal{D}=\big[\mathrm{tr}\big(\Sigma_nH_n^{'}P_{1n}^{(s)}\big),\ldots,\mathrm{tr}\big(\Sigma_nH_n^{'}P_{mn}^{(s)}\big)\big]\mathcal{A}_n^{-1}\big[\mathrm{tr}\big(\Sigma_n\overline{G}_n^{'}P_{1n}^{(s)}\big),\ldots,\mathrm{tr}\big(\Sigma_n\overline{G}_n^{'}P_{mn}^{(s)}\big)\big]^{'}$ and $\mathcal{A}_n=\frac{1}{2}\big[\,\mathrm{vec}\,\big((\Sigma_nP_{1n})^{(s)}\big),\ldots,\mathrm{vec}\,\big((\Sigma_nP_{mn})^{(s)}\big)\big]^{'}\big[\,\mathrm{vec}\,\big((\Sigma_nP_{1n})^{(s)}\big),\ldots,\mathrm{vec}\,\big((\Sigma_nP_{mn})^{(s)}\big)\big]$. Note that when $P_jn\in\mathcal{P}_n\ \forall j$, the covariance between a quadratic linear moment function and the linear moment function is zero. That is, $\mathrm{Cov}\big(\varepsilon_n^{'}P_{jn}\varepsilon_n,\,Q_n^{'}\varepsilon_n\big)=Q_n^{'}\sum_{k=1}^{n}\sum_{l=1}^{n}P_{kl,jn}\mathrm{E}\big(\varepsilon_n\varepsilon_{kn}\varepsilon_{ln}\big)=\mu_3Q_n^{'}\,\mathrm{vec}_D\,\big(P_{jn}\big)=0_{n\times 1}$, since $\mathrm{vec}_D\,\big(P_{jn}\big)=0_{n\times 1}$ for all j (See Lemma 1). This result shows that the best P_{jn}s can be determined from the first matrix on the right hand side of (2.3) using the Schwartz inequality to determine upper bounds for its elements.

Claim 1. — Under our stated assumptions, the best P_n matrices for the quadratic moment functions are $P_{1n}=\Sigma_n^{-1}\big(\overline{G}_n-\mathrm{D}(\overline{G}_n)\big)$ and $P_{2n}=\Sigma^{-1}\big(H_n-\mathrm{D}(H_n)\big)$.

Proof. See Appendix C. □

The best quadratic moment matrices involve the unknown covariance matrix Σ_n which has an unknown form. In the case where there is an assumed parametric specification for the variance terms, Σ_n can be consistently estimated and the best quadratic moments will be available. Hence, under heteroskedasticity of an unknown form, the GMME based on the best quadratic moment moment matrices is not feasible. One can consider the GMME based on the quadratic moment matrices when the disturbance terms are simply i.i.d. In that case, Claim 1 implies that the best quadratic moment matrices are $P_{1n}=\overline{G}_n-\mathrm{D}(\overline{G}_n)$ and $P_{2n}=H_n-\mathrm{D}(H_n)$.

The optimal robust GMME requires an initial consistent estimates of the parameters. Among others, an IRGMME based on the quadratic moment matrices $P_{1n}=W_n^{'}W_n-\mathrm{D}\left(W_n^{'}W_n\right)$, $P_{2n}=M_n^{'}M_n-\mathrm{D}\left(M_n^{'}M_n\right)$ and the linear moment matrix $Q_n=\big[W_nM_nX_n,W_nX_n,M_nX_n,X_n\big]$ can be employed, when the disturbance terms satisfy Assumption 1.

[3]The asymptotic results in this section are proved in Dogan and Taspinar, (2013) along the lines of Lin and Lee, (2010).

3 Heteroskedasticity-Consistent Covariance Matrix Estimators

In this section, we consider various refinement methods suggested in the literature, and extend these methods for our spatial autoregressive model. We provide a general argument by considering the general vector of population moment functions $g_n(\theta_0) = \left(\varepsilon_n^{'} P_{1n}\varepsilon_n, \ldots, \varepsilon_n^{'} P_{mn}\varepsilon_n, \varepsilon_n^{'} Q_n\right)^{'}$ where Q_n is an $n \times r$ matrix of linear instruments, and $P_{jn} \in \mathcal{P}_n$ for $j = 1, \ldots, m$.

Following the similar notation of MacKinnon and White, (1985), we denote $\left[\frac{1}{n}\,\widehat{\Phi}_{2n}^{'}\,\widehat{\Omega}_{1n}^{-1}\,\widehat{\Phi}_{2n}\right]^{-1}$ by $SHC0$ when $\widehat{\Sigma}_n = \mathrm{D}(\widehat{\varepsilon}_{1n}^2, \ldots, \widehat{\varepsilon}_{nn}^2)$. Hinkley, (1977) consider another version in which individual residuals are scaled according to the degrees of freedom in the residual vector. This version of the estimated covariance, denoted by $SHC1$, is based on $\widehat{\Sigma}_{1n} = (n/(n-k))\,\mathrm{D}(\widehat{\varepsilon}_{1n}^2, \ldots, \widehat{\varepsilon}_{nn}^2)$.[4] Following Horn et al., (1975), MacKinnon and White, (1985) suggest an alternative approach for a linear regression model when the disturbance terms of the model are homeskedastic. This approach produces an unbiased estimator and is based on the diagonal elements of a matrix, called the hat matrix. The literature has provided various modifications based on the diagonal elements of the hat matrix (Bera et al., 2002; Cribari-Neto, 2004; Cribari-Neto et al., 2007; Kauermann and Carroll, 2001; Lin and Chou, 2015; Long and Ervin, 2000; MacKinnon, 2013; MacKinnon and White, 1985). We will consider the counterparts of these modified versions for our spatial model as well.

Next, we derive alternative HCCMEs formulated from a hat matrix by extending the refinement methodology of Lin and Chou, (2015) for our spatial model. The extension is not trivial mainly due to complications arising from the spatial structure of our model. First, moment functions that are quadratic in the disturbance terms complicate a direct extension of Lin and Chou, (2015). Second, their methodology is an extension of the idea of the leverage adjusted-residuals in MacKinnon and White, (1985) to a non-linear regression model. In essence, various HCCMEs are based on the leverage-adjusted residuals relation, stated as $\mathrm{E}\left(\widehat{\varepsilon}_{in}^2\right) = \sigma_0^2\left(1 - \mathcal{H}_{ii,n}\right)$. Here, $\widehat{\varepsilon}_{in}^2$ is the ith residual based on a consistent estimator and $\mathcal{H}_{ii,n}$ is the (i,i)th element of a matrix $\mathcal{H}_n$. In the presence of spatial dependence, such a relationship between the residuals and the individual variance cannot be established at the observational level. Instead, such a relationship needs to be established at the sample level in the form of $\mathrm{E}\left(\widehat{\varepsilon}_n\widehat{\varepsilon}_n^{'}\right) = \sigma_0^2\left(I_n - \mathcal{H}_n\right)$. In the following, we present the details on how this relationship can be established for our spatial model.

By the mean value theorem, we can write $\varepsilon_n(\widehat{\theta}_n) = \varepsilon_n(\theta_0) + \frac{\partial \varepsilon_n(\bar{\theta}_n)}{\partial \theta'}\left(\widehat{\theta}_n - \theta_0\right)$ where $\bar{\theta}_n$ lies between $\widehat{\theta}_n$ and θ_0. Let $\epsilon_n \equiv \widehat{\varepsilon}_n(\widehat{\theta}_{1n})$, where $\widehat{\varepsilon}_n(\widehat{\theta}_{1n})$ is the residual vector recovered by using the initial estimator $\widehat{\theta}_{1n}$. Then, the outer product of ϵ_n is given by

$$\begin{aligned}\epsilon_n\epsilon_n^{'} = \varepsilon_n(\theta_0)\varepsilon_n^{'}(\theta_0) &+ \frac{\partial \varepsilon_n(\bar{\theta}_n)}{\partial \theta'}\left(\widehat{\theta}_{1n} - \theta_0\right)\left(\widehat{\theta}_{1n} - \theta_0\right)^{'}\frac{\partial \varepsilon_n^{'}(\bar{\theta}_n)}{\partial \theta} + \frac{\partial \varepsilon_n(\bar{\theta}_n)}{\partial \theta'}\left(\widehat{\theta}_{1n} - \theta_0\right)\varepsilon_n^{'}(\theta_0)\\ &+ \varepsilon_n(\theta_0)\left(\widehat{\theta}_{1n} - \theta_0\right)^{'}\frac{\partial \varepsilon_n^{'}(\bar{\theta}_n)}{\partial \theta}. \end{aligned} \tag{3.1}$$

Now, replacing $\bar{\theta}_n$ with θ_0 and taking the expectation of (3.1) under homoskedasticity assumption,

[4] In the context of non-spatial linear regression models, both $HC0$ and $HC1$ are consistent, but generally biased under both homoskedasticity and heteroskedasticity (Bera et al., 2002).

we obtain

$$\mathrm{E}\big(\epsilon_n\epsilon_n^{'}\big) \approx \sigma_0^2 I_n + \mathrm{E}\bigg(\frac{\partial\varepsilon_n(\theta_0)}{\partial\theta'}\big(\widehat{\theta}_{1n}-\theta_0\big)\big(\widehat{\theta}_{1n}-\theta_0\big)'\frac{\partial\varepsilon_n^{'}(\theta_0)}{\partial\theta}\bigg) \\ + \mathrm{E}\bigg(\frac{\partial\varepsilon_n(\theta_0)}{\partial\theta'}\big(\widehat{\theta}_{1n}-\theta_0\big)\varepsilon_n^{'}(\theta_0)\bigg) + \mathrm{E}\bigg(\varepsilon_n(\theta_0)\big(\widehat{\theta}_{1n}-\theta_0\big)'\frac{\partial\varepsilon_n^{'}(\theta_0)}{\partial\theta}\bigg). \tag{3.2}$$

The above representation, implicitly, suggests a quasi-hat matrix, which can be recovered from $\mathrm{E}\big(\epsilon_n\epsilon_n^{'}\big) \approx \sigma_0^2\big(I_n - \mathcal{H}_{1n}\big)$, where

$$\mathcal{H}_{1n} = -\bigg[\frac{1}{\sigma_0^2}\mathrm{E}\bigg(\frac{\partial\varepsilon_n(\theta_0)}{\partial\theta'}\big(\widehat{\theta}_{1n}-\theta_0\big)\big(\widehat{\theta}_{1n}-\theta_0\big)'\frac{\partial\varepsilon_n^{'}(\theta_0)}{\partial\theta}\bigg) + \frac{1}{\sigma_0^2}\mathrm{E}\bigg(\frac{\partial\varepsilon_n(\theta_0)}{\partial\theta'}\big(\widehat{\theta}_{1n}-\theta_0\big)\varepsilon_n^{'}(\theta_0)\bigg) \\ + \frac{1}{\sigma_0^2}\mathrm{E}\bigg(\varepsilon_n(\theta_0)\big(\widehat{\theta}_{1n}-\theta_0\big)'\frac{\partial\varepsilon_n^{'}(\theta_0)}{\partial\theta}\bigg)\bigg]. \tag{3.3}$$

First order asymptotic results for $\big(\widehat{\theta}_{1n}-\theta_0\big)$ can be used to determine the expectation of each term in (3.3). Let Ψ_n be an arbitrary non-stochastic weighting matrix for the GMM objective function. Then, an initial GMME is defined by $\widehat{\theta}_{1n} = \operatorname{argmin}_{\theta\in\Theta}\, g_n^{'}(\theta)\,\Psi_n^{-1}\,g_n(\theta)$. The first order condition of the objective function is $\frac{\partial g_n{}'(\widehat{\theta}_{1n})}{\partial\theta}\Psi_n^{-1}g_n(\widehat{\theta}_{1n}) = 0$. By the mean value theorem at $\bar{\theta}_n$, we have

$$\sqrt{n}\big(\widehat{\theta}_{1n}-\theta_0\big) = -\bigg(\frac{1}{n}\frac{\partial g_n^{'}(\widehat{\theta}_{1n})}{\partial\theta}\Psi_n^{-1}\frac{1}{n}\frac{\partial g_n(\bar{\theta}_{1n})}{\partial\theta'}\bigg)^{-1}\frac{1}{n}\frac{\partial g_n^{'}(\widehat{\theta}_{1n})}{\partial\theta}\Psi_n^{-1}\frac{1}{\sqrt{n}}g_n(\theta_0), \tag{3.4}$$

where $\frac{1}{n}\frac{\partial g_n(\theta)}{\partial\theta'} = \frac{1}{n}\big(P_{1n}^s\varepsilon_n(\theta), P_{2n}^s\varepsilon_n(\theta), \ldots, P_{mn}^s\varepsilon_n(\theta), Q_n\big)'\frac{\partial\varepsilon_n(\theta)}{\partial\theta'}$. Under our regularity conditions, we have $\frac{1}{n}\frac{\partial g_n(\widehat{\theta}_{1n})}{\partial\theta'} = \frac{1}{n}\mathrm{E}\big(\frac{\partial g_n(\theta_0)}{\partial\theta'}\big) + o_p(1) = \frac{1}{n}\Phi_n + o_p(1)$. Therefore, we have

$$\sqrt{n}\big(\widehat{\theta}_{1n}-\theta_0\big) = -\bigg(\frac{1}{n}\Phi_n^{'}\Psi_n^{-1}\frac{1}{n}\Phi_n\bigg)^{-1}\frac{1}{n}\Phi_n^{'}\Psi_n^{-1}\frac{1}{\sqrt{n}}g_n(\theta_0) + o_p(1) = \mathcal{Z}_n\frac{1}{\sqrt{n}}g_n(\theta_0) + o_p(1) \tag{3.5}$$

where $\mathcal{Z}_n = -\big(\frac{1}{n}\Phi_n^{'}\Psi_n^{-1}\frac{1}{n}\Phi_n\big)^{-1}\frac{1}{n}\Phi_n^{'}\Psi_n^{-1}$ is a $(k+2)\times(m+r)$ matrix. For $\frac{\partial\varepsilon(\theta_0)}{\partial\theta'}$ in (3.3), we have

$$\frac{\partial\varepsilon(\theta_0)}{\partial\theta'} = -\big[M_n\big(S_nY_n - X_n\beta_0\big),\, R_nW_nY_n,\, R_nX_n\big]. \tag{3.6}$$

Let $\mathcal{K}_n \equiv \big[M_n\big(S_nY_n - X_n\beta_0\big),\, R_nW_nY_n,\, R_nX_n\big]$ and let E_i, for $i = 1, 2$, denote a $(k+2)\times(k+2)$ square matrix with zero elements except the $(1, i)$th element, which equals 1. Also, let E_3 be a $(k+2)\times(k+2)$ square matrix with zero elements except the elements from the $(1, 3)$th element through $(1, k+2)$th element, which equal 1. It will be convenient to write (3.6) in the following way:

$$\frac{\partial\varepsilon(\theta_0)}{\partial\theta'} = -\big(\mathcal{K}_nE_1 + \mathcal{K}_nE_2 + \mathcal{K}_nE_3\big). \tag{3.7}$$

From (3.3), (3.4) and (3.7), it follows that

$$
\begin{aligned}
\mathcal{H}_{1n} = & -\frac{1}{n^2}\frac{1}{\sigma_0^2}\Big[\mathrm{E}\Big(\big(\mathcal{K}_nE_1+\mathcal{K}_nE_2+\mathcal{K}_nE_3\big)\mathcal{Z}_ng_n(\theta_0)g_n^{'}(\theta_0)\mathcal{Z}_n^{'}\big(\mathcal{K}_nE_1+\mathcal{K}_nE_2+\mathcal{K}_nE_3\big)^{'}\Big)\Big] \\
& +\frac{1}{n}\frac{1}{\sigma_0^2}\mathrm{E}\Big(\big(\mathcal{K}_nE_1+\mathcal{K}_nE_2+\mathcal{K}_nE_3\big)\mathcal{Z}_ng_n(\theta_0)\varepsilon_n^{'}(\theta_0)\Big) \\
& +\frac{1}{n}\frac{1}{\sigma_0^2}\mathrm{E}\Big(\varepsilon_n^{'}(\theta_0)g_n^{'}(\theta_0)\mathcal{Z}_n^{'}\big(\mathcal{K}_nE_1+\mathcal{K}_nE_2+\mathcal{K}_nE_3\big)^{'}\Big). \qquad (3.8)
\end{aligned}
$$

The result in (3.8) indicates that the quasi-hat matrix will be available when all the expectation terms are evaluated. We will elaborate on how to evaluate these expectation terms in Section 4. We will show that an estimate of $\mathcal{H}_{1n}$ can be recovered from the initial consistent estimates of θ_0, σ_0^2, $\mu_3=\mathrm{E}\big(\varepsilon_{in}^3\big)$ and $\mu_4=\mathrm{E}\big(\varepsilon_{in}^4\big)$. We will denote the resulting estimate of $\mathcal{H}_{1n}$ by $\mathcal{H}_{1n}(\widehat{\theta}_{1n})$, where $\widehat{\theta}_{1n}$ is an initial consistent estimator of θ_0.

Let $\widehat{\mathcal{H}}_{ii,1n}$ be the ith diagonal element of $\mathcal{H}_{1n}(\widehat{\theta}_{1n})$ for $i=1,\ldots,n$. In analogous to the non-spatial literature, we use the diagonal elements of this hat matrix to define some other HCCME versions. Corresponding to $HC2$ and $HC3$ of MacKinnon and White, (1985), we formulate $SHC2^{\star}$ and $SHC3^{\star}$ based on the following matrices:

$$
\widehat{\Sigma}_{2n}^{\star}=\mathrm{D}\left(\frac{\widehat{\varepsilon}_{1n}^2(\widehat{\theta}_{2n})}{1-\widehat{\mathcal{H}}_{11,1n}},\ldots,\frac{\widehat{\varepsilon}_{nn}^2(\widehat{\theta}_{2n})}{1-\widehat{\mathcal{H}}_{nn,1n}}\right), \qquad (3.9)
$$

$$
\widehat{\Sigma}_{3n}^{\star}=\mathrm{D}\left(\frac{\widehat{\varepsilon}_{1n}^2(\widehat{\theta}_{2n})}{\big(1-\widehat{\mathcal{H}}_{11,1n}\big)^2},\ldots,\frac{\widehat{\varepsilon}_{nn}^2(\widehat{\theta}_{2n})}{\big(1-\widehat{\mathcal{H}}_{nn,1n}\big)^2}\right). \qquad (3.10)
$$

Corresponding to $HC4$ of Cribari-Neto, (2004), we formulate another covariance estimate denoted by $SHC4^{\star}$, with the following matrix:

$$
\widehat{\Sigma}_{4n}^{\star}=\mathrm{D}\left(\frac{\widehat{\varepsilon}_{1n}^2(\widehat{\theta}_{2n})}{\big(1-\widehat{\mathcal{H}}_{11,2n}\big)^{\nu_1}},\ldots,\frac{\widehat{\varepsilon}_{nn}^2(\widehat{\theta}_{2n})}{\big(1-\widehat{\mathcal{H}}_{nn,2n}\big)^{\nu_n}}\right), \qquad (3.11)
$$

where $\nu_i=\min\big\{\frac{n\widehat{\mathcal{H}}_{ii,1n}}{\sum_{i=1}^n\widehat{\mathcal{H}}_{ii,1n}},\,4\big\}$ for $i=1,\ldots,n$. Using the fact that $\sum_{i=1}^n\widehat{\mathcal{H}}_{ii,1n}=\mathrm{tr}\big(\widehat{\mathcal{H}}_{1n}\big)=k$, we can simply define $\nu_i=\min\big\{\frac{n\widehat{\mathcal{H}}_{ii,1n}}{k},\,4\big\}$. In (3.11), observations that have high leverage are more inflated by the corresponding discount factors. The truncation at 4 for the discount factors is twice what is used in the definition of $SHC3$. When $\widehat{\mathcal{H}}_{ii,1n}>4k/n$, $\nu_i=4$. Cribari-Neto et al., (2007) also suggest a modified version of $HC4$ which we will denote with $HC5$. Our analogous version $SHC5^{\star}$ is formulated with

$$
\widehat{\Sigma}_{5n}^{\star}=\mathrm{D}\left(\frac{\widehat{\varepsilon}_{1n}^2(\widehat{\theta}_{2n})}{\big(1-\widehat{\mathcal{H}}_{11,1n}\big)^{\alpha_1}},\ldots,\frac{\widehat{\varepsilon}_{nn}^2(\widehat{\theta}_{2n})}{\big(1-\widehat{\mathcal{H}}_{nn,1n}\big)^{\alpha_n}}\right), \qquad (3.12)
$$

where $\alpha_i=\min\big\{\frac{n\widehat{\mathcal{H}}_{ii,1n}}{\sum_{i=1}^n\widehat{\mathcal{H}}_{ii,1n}},\,\max\big\{\frac{n\kappa\widehat{\mathcal{H}}_{\max}}{\sum_{i=1}^n\widehat{\mathcal{H}}_{ii,1n}},\,4\big\}\big\}$. Here, $\kappa\in(0,\,1)$ is a predefined constant, and $\mathcal{H}_{\max}=\max\big\{\widehat{\mathcal{H}}_{11,1n},\ldots,\widehat{\mathcal{H}}_{nn,1n}\big\}$. The literature on linear regression models shows that $HC0$ can be substantially downward biased in finite sample, especially when there are are high leverage points in the design matrix (Chesher, 1989; Chesher and Jewitt, 1987)[5]. Both ν_i and α_i determine

[5]For a non-spatial linear regression model, the hat matrix is given by $H=X(X^{'}X)^{-1}X^{'}$. A value of H_{ii} greater

how much the ith residual should be inflated to adjust the ith observation leverage. For non-spatial linear regression models, Cribari-Neto, (2004) and Cribari-Neto et al., (2007) show that $HC4$ and $HC5$ can yield reliable inference results, even under extremely leveraged data. If $\kappa = 0$, $\widehat{\Sigma}_{5n}$ reduces to $\widehat{\Sigma}_{4n}$. The simulation results in Cribari-Neto et al., (2007) indicate that the setting of $\kappa = 0.7$ provides reliable inference in finite samples.

We will close this section by considering a naive approach which will yield another hat matrix. For a given value of $\delta = (\rho, \lambda)^{'}$, the model in (2.1) can be written as $R_n(\rho)S_n(\lambda)Y_n = R_n(\rho)X_n\beta + \varepsilon$. The OLS estimator from this equation is given by $\widehat{\beta}_n = \left(X_n^{'}R_n^{'}(\rho)R_n(\rho)X_n\right)^{-1}X_n^{'}R_n^{'}(\rho)R_n(\rho)S_n(\lambda)Y_n$. For a given value of δ, we have $\widehat{\varepsilon}_n(\delta) = R_n(\rho)S_n(\lambda)Y_n - R_n(\rho)X_n\widehat{\beta}_n = \mathcal{M}_n(\rho)R_n(\rho)S_n(\lambda)Y_n$, where $\mathcal{M}_n(\rho) = \big[I_n - R_n(\rho)X_n\big(X_n^{'}R_n^{'}(\rho)R_n(\rho)X_n\big)^{-1}X_n^{'}R_n^{'}(\rho)\big]$ is an idempotent residual maker type matrix. Under the assumption of homoskedasticity, we have

$$\mathrm{E}\big(\widehat{\varepsilon}_n(\delta)\widehat{\varepsilon}_n^{'}(\delta)\big) = \mathcal{M}_n(\rho)\mathrm{E}\big(\varepsilon_n\varepsilon_n^{'}\big)\mathcal{M}_n(\rho) = \sigma_0^2\mathcal{M}_n(\rho) = \sigma_0^2\big(I_n - \mathcal{H}_{2n}(\rho)\big), \tag{3.13}$$

where $\mathcal{H}_{2n}(\rho) = R_n(\rho)X_n\big(X_n^{'}R_n^{'}(\rho)R_n(\rho)X_n\big)^{-1}X_n^{'}R_n^{'}(\rho)$ can be considered as a quasi hat matrix. We can use (3.13) to replace $\widehat{\varepsilon}_{in}^2$ in $\widehat{\Sigma}_n$. Analogous to (3.9), an estimate of Σ_n, denoted by $\widehat{\Sigma}_{2n}$, can be formulated using $\widehat{\varepsilon}_{1n}^2(\widehat{\delta}_n)$ and the diagonal elements of $\widehat{\mathcal{H}}_{2n}$. Here, $\widehat{\delta}_n$ is a consistent estimator of δ_0. We will refer to the covariance estimate formulated with $\widehat{\Sigma}_{2n}$ by $SHC2$. Note also that we can determine the bias $\mathrm{E}\left(\widehat{\varepsilon}_{in}^2(\delta)\right) - \sigma_{in}^2$ when $\mathrm{E}\left(\varepsilon_n\varepsilon_n^{'}\right) = \Sigma_n$ for a given δ (Bera et al., 2002; Chesher and Jewitt, 1987). We have

$$\begin{aligned}\mathrm{E}\big(\widehat{\varepsilon}_{in}^2(\delta)\big) &= \mathcal{M}_{\bullet i,n}^{'}(\rho)\mathrm{E}\big(\varepsilon_n\varepsilon_n^{'}\big)\mathcal{M}_{\bullet i,n}(\rho) = \mathcal{M}_{\bullet i,n}^{'}(\rho)\Sigma_n\mathcal{M}_{\bullet i,n}(\rho)\\ &= \sigma_{in}^2 - 2\mathcal{H}_{\bullet i,2n}^{'}(\rho)\mathcal{H}_{\bullet i,2n}(\rho)\sigma_{in}^2 + \mathcal{H}_{\bullet i,2n}^{'}(\rho)\Sigma_n\mathcal{H}_{\bullet i,2n}(\rho)\end{aligned} \tag{3.14}$$

where the last equality follows from the fact that $\mathcal{H}_{2n}(\rho)$ is symmetric and idempotent. The result in(3.14) implies the bias of $\mathrm{E}\left(\widehat{\varepsilon}_{in}^2(\delta)\right) - \sigma_{in}^2 = \mathcal{H}_{\bullet i,2n}^{'}(\rho)\left(\Sigma_n - 2I_n\sigma_{in}^2\right)\mathcal{H}_{\bullet i,2n}(\rho)$ for a given δ. Note that when $\mathrm{E}\big(\varepsilon_n\varepsilon_n^{'}\big) = \sigma_0^2 I_n$, we have $\mathrm{E}\left(\widehat{\varepsilon}_{in}^2(\delta)\right) - \sigma_0^2 = -\sigma_0^2\mathcal{H}_{ii,2n}(\rho)$ for a given δ. Hence, $\mathrm{E}\big(\widehat{\varepsilon}_{in}^2(\delta)/\big[1-\mathcal{H}_{ii,2n}(\rho)\big]\big) = \sigma_0^2$ for a given δ. Similarly, we can define counterparts of (3.10) through (3.12) using $\widehat{\varepsilon}_n^2(\widehat{\delta}_n)$ and $\widehat{\mathcal{H}}_{2n}$. We will denote the respective covariance estimates with $SHC3$, $SHC4$ and $SHC5$.

4 The Quasi-Hat Matrix

In this section, we lay out the details on how to evaluate each expression stated in (3.8). The latter two terms in (3.8) are relatively easier to deal with and we will start with these terms. First, we consider (i) $\mathrm{E}\big(\mathcal{K}_nE_1\mathcal{Z}_ng_n(\theta_0)\varepsilon_n^{'}(\theta_0)\big) = H_n\mathrm{E}\big(\varepsilon_n\mathcal{Z}_{1\bullet,n}g_n(\theta_0)\varepsilon_n^{'}\big) = H_n\mathrm{E}\big(\mathcal{D}_{1n}\big)$ where $\mathcal{Z}_{1\bullet,n}$ is the first row of $\mathcal{Z}_n$ and $\mathcal{D}_{1n} = \varepsilon_n\mathcal{Z}_{1\bullet,n}g_n(\theta_0)\varepsilon_n^{'}$. Let e_i be the ith elementary vector in $\mathbb{R}^n$. Then, the expectation of the (s,s)th element of $\mathcal{D}_{1n}$ is given by $\mathrm{E}\big(e_s^{'}\mathcal{D}_{1n}e_s\big) = \mathcal{Z}_{1\bullet,n}\mathrm{E}\left(g_n(\theta_0)\varepsilon_{sn}^2\right)$, where $\mathrm{E}\big(g_n(\theta_0)\varepsilon_{sn}^2\big) = \big[0_{1\times m},\, \mu_3Q_{s\bullet,n}\big]^{'}$ by Lemma 2. Similarly, by using elementary vectors, the expectation of the (s,t)th element in $\mathcal{D}_{1n}$ is given by $\mathrm{E}\big(e_s^{'}\mathcal{D}_{1n}e_t\big) = \mathcal{Z}_{1\bullet,n}\mathrm{E}\big(g_n(\theta_0)\varepsilon_{sn}\varepsilon_{tn}\big)$, where by Lemma 2 we have $\mathrm{E}\big(g_n(\theta_0)\varepsilon_{sn}\varepsilon_{tn}\big) = \big[\sigma_0^4\mathcal{V}_{st},\, 0_{1\times r}\big]^{'}$ and $\mathcal{V}_{st} = \big[P_{st,1n}^{(s)}, \ldots, P_{st,mn}^{(s)}\big]$.

The next term that we consider is (ii) $\mathrm{E}\big(\mathcal{K}_nE_2\mathcal{Z}_ng_n(\theta_0)\varepsilon_n^{'}\big) = \overline{G}_n\overline{X}_n\beta_0\mathcal{Z}_{2\bullet,n}\mathrm{E}\big(\mathcal{D}_{2n}\big) + \overline{G}_n\mathrm{E}\big(\mathcal{D}_{3n}\big)$ where $\mathcal{D}_{2n} = g_n(\theta_0)\varepsilon_n^{'}$, $\mathcal{D}_{3n} = \varepsilon_n\mathcal{Z}_{2\bullet,n}g_n(\theta_0)\varepsilon_n^{'}$ and $\mathcal{Z}_{2\bullet,n} = \big(\mathcal{Z}_{21,n}, \ldots, \mathcal{Z}_{2(m+r),n}\big)$ is the second

than $\frac{2}{n}\mathrm{tr}(H) = \frac{2k}{n}$ or $\frac{3}{n}\mathrm{tr}(H) = \frac{3k}{n}$ is considered as a high leverage point (Judge et al., 1988).

row of $\mathcal{Z}_n$. First, we shall evaluate the expectation of $\mathcal{D}_{2n}$. The independence of ε_{in}s implies that $\mathrm{E}\big(\mathcal{D}_{2n}\big) = \big[0_{n\times m},\, \sigma_0^2 Q_n\big]^{'}$. Coming to the expectation of $\mathcal{D}_{3n}$, the (s,s)th and (s,t)th elements of $\mathrm{E}\big(\mathcal{D}_{3n}\big)$ are respectively given by $\mathrm{E}\big(e_s^{'}\mathcal{D}_{3n}e_s\big) = \mathcal{Z}_{2\bullet,n}\big[0_{1\times m},\, \mu_3 Q_{s\bullet,n}\big]^{'}$ and $\mathrm{E}\big(e_s^{'}\mathcal{D}_{3n}e_t\big) = \mathcal{Z}_{2\bullet,n}\big[\sigma_0^4\mathcal{V}_{st},\, 0_{1\times r}\big]^{'}$, where we use Lemma 2. Let $\mathbb{Z}_{3n} = \big(\mathcal{Z}_{3\bullet,n}^{'}, \ldots, \mathcal{Z}_{(k+2)\bullet,n}^{'}\big)^{'}$ be the $k\times(m+r)$ matrix. The last term we need to evaluate in the latter two terms in (3.8) is $\mathrm{E}\big(\mathcal{K}_n E_3 \mathcal{Z}_n g_n(\theta_0)\varepsilon_n^{'}\big) = R_n X_n \mathbb{Z}_{3n}\mathrm{E}\big(\mathcal{D}_{2n}\big)$. Then, we obtain $\mathrm{E}\big(\mathcal{K}_n E_3 \mathcal{Z}_n g_n(\theta_0)\varepsilon_n^{'}\big) = R_n X_n \mathbb{Z}_{3n}\big[0_{n\times m}\,, \sigma_0^2 Q_n\big]^{'}$ by the independence of ε_{in}s.

Next, we shall return to the first term on the right hand side in (3.8) which involves expectation expressions for six unique terms. We start with (iv) $\mathrm{E}\big(\mathcal{K}_n E_1 \mathcal{Z}_n g_n(\theta_0) g_n(\theta_0)^{'}\mathcal{Z}_n^{'}E_1^{'}\mathcal{K}_n^{'}\big)$. The integrand of this term is given by $H_n\varepsilon_n\mathcal{Z}_{1\bullet,n}g_n(\theta_0)g_n(\theta_0)^{'}\mathcal{Z}_{1\bullet,n}^{'}\varepsilon_n^{'}H_n^{'}$. For notational conevenience, let $\mathcal{F}_n$ denote $g_n(\theta_0)g_n^{'}(\theta_0)$ and let $\mathcal{U}_{1n}$ denote $\varepsilon_n\mathcal{Z}_{1\bullet,n}\mathcal{F}_n\mathcal{Z}_{1\bullet,n}^{'}\varepsilon_n^{'}$. Then,

$$\mathrm{E}\big(\mathcal{K}_n E_1 \mathcal{Z}_n g_n(\theta_0) g_n(\theta_0)^{'}\mathcal{Z}_n^{'}E_1^{'}\mathcal{K}_n^{'}\big) = H_n\mathrm{E}\left(\mathcal{U}_{1n}\right)H_n^{'} \tag{4.1}$$

Then, the (s,s)th element of $\mathrm{E}\left(\mathcal{U}_{1n}\right)$ is $\mathcal{Z}_{1\bullet,n}\mathrm{E}\big(\mathcal{F}_n\varepsilon_{sn}^2\big)\mathcal{Z}_{1\bullet,n}^{'}$. Using Lemma 2, we can show that

$$\mathrm{E}\big(\mathcal{F}_n\varepsilon_{sn}^2\big) = \sigma_0^6\begin{bmatrix} \Xi_{nm} & 0_{m\times r} \\ 0_{r\times m} & 0_{r\times r}\end{bmatrix} + \begin{bmatrix} \left(\sigma_0^2\mu_4 - \sigma_0^6\right)\mathcal{V}_s^{'}\mathcal{V}_s & \mu_3\sigma_0^2\mathcal{V}_s^{'}Q_n \\ \mu_3\sigma_0^2 Q_n^{'}\mathcal{V}_s & \sigma_0^4 Q_n^{'}Q_n + \left(\mu_4 - \sigma_0^4\right)Q_{s\bullet,n}^{'}Q_{s\bullet,n}\end{bmatrix} \tag{4.2}$$

where $\Xi_{nm} = \big[\mathrm{vec}(P_{1n}^{(s)}), \ldots, \mathrm{vec}(P_{mn}^{(s)})\big]^{'}\big[\mathrm{vec}(P_{1n}), \ldots, \mathrm{vec}(P_{mn})\big]$, $\mathcal{V}_s = \big[P_{\bullet s,1n}^{(s)}, \ldots, P_{\bullet s,mn}^{(s)}\big]$ and $P_{\bullet s,jn}^{(s)} = P_{s\bullet,jn}^{'} + P_{\bullet s,jn}$. Similarly, the expectation of the (s,t)th element of $\mathcal{U}_{1n}$ is $\mathcal{Z}_{1\bullet,n}\mathrm{E}\left(\mathcal{F}_n\varepsilon_{sn}\varepsilon_{tn}\right)\mathcal{Z}_{1\bullet,n}^{'}$. Then, using Lemma 2 again, we obtain

$$\mathrm{E}\big(\mathcal{F}_n\varepsilon_{sn}\varepsilon_{tn}\big) = \begin{bmatrix} \mu_3^2\mathcal{V}_{st}^{'}\mathcal{V}_{st} & \mu_3\sigma_0^2\mathcal{V}_{st}^{'}\mathcal{Q}_{st} \\ \mu_3\sigma_0^2\mathcal{Q}_{st}^{'}\mathcal{V}_{st} & \sigma_0^4\left(Q_{s\bullet,n}^{'}Q_{t\bullet,n} + Q_{t\bullet,n}^{'}Q_{s\bullet,n}\right)\end{bmatrix} \tag{4.3}$$

where $\mathcal{V}_{st} = \big[P_{st,1n}^{(s)}, \ldots, P_{st,mn}^{(s)}\big]$, $P_{st,jn}^{(s)} = P_{st,jn} + P_{ts,jn}$ and $\mathcal{Q}_{st} = Q_{s\bullet,n} + Q_{t\bullet,n}$.

Another term in (3.8) is (vii) $\mathrm{E}\left(\mathcal{K}_n E_2 \mathcal{Z}_n g_n(\theta_0) g_n(\theta_0)^{'}\mathcal{Z}_n^{'}E_2^{'}\mathcal{K}_n^{'}\right)$, which can be written as

$$\begin{aligned}
&\mathrm{E}\left(\mathcal{K}_n E_2 \mathcal{Z}_n g_n(\theta_0) g_n(\theta_0)^{'}\mathcal{Z}_n^{'}E_2^{'}\mathcal{K}_n^{'}\right) = \big(\overline{G}_n\overline{X}_n\beta_0\big)\mathcal{Z}_{2\bullet,n}\mathrm{E}\big(\mathcal{F}_n\big)\mathcal{Z}_{2\bullet,n}^{'}\big(\overline{G}_n\overline{X}_n\beta_0\big)^{'} \\
&\quad + \overline{G}_n\mathrm{E}\big(\varepsilon_n\mathcal{Z}_{2\bullet,n}\mathcal{F}_n\mathcal{Z}_{2\bullet,n}^{'}\varepsilon_n^{'}\big)\overline{G}_n^{'} + \big(\overline{G}_n\overline{X}_n\beta_0\big)\mathcal{Z}_{2\bullet,n}\mathrm{E}\big(\mathcal{F}_n\mathcal{Z}_{2\bullet,n}^{'}\varepsilon_n^{'}\big)\overline{G}_n^{'} \\
&\quad + \overline{G}_n\mathrm{E}\big(\varepsilon_n\mathcal{Z}_{2\bullet,n}\mathcal{F}_n\big)\mathcal{Z}_{2\bullet,n}^{'}\big(\overline{G}_n\overline{X}_n\beta_0\big)^{'}.
\end{aligned} \tag{4.4}$$

We will evaluate each term in (4.4) separately. Let $\mathrm{Diag}\left(\cdot\right)$ be a generalized block diagonal matrix operator that forms a block diagonal matrix from the list of input matrices. Then, it follows from Lemma 1 that

$$\begin{aligned}
&\big(\overline{G}_n\overline{X}_n\beta_0\big)\mathcal{Z}_{2\bullet,n}\mathrm{E}\big(\mathcal{F}_n\big)\mathcal{Z}_{2\bullet,n}^{'}\big(\overline{G}_n\overline{X}_n\beta_0\big)^{'} \\
&\quad = \big(\overline{G}_n\overline{X}_n\beta_0\big)\mathcal{Z}_{2\bullet,n}\,\mathrm{Diag}\left(\sigma_0^4\Xi_{nm}, \sigma_0^2 Q_n^{'}Q_n\right)\mathcal{Z}_{2\bullet,n}^{'}\big(\overline{G}_n\overline{X}_n\beta_0\big)^{'}
\end{aligned} \tag{4.5}$$

where $\Xi_{nm} = \big[\mathrm{vec}(P_{1n}^{(s)}), \ldots, \mathrm{vec}(P_{mn}^{(s)})\big]^{'}\big[\mathrm{vec}(P_{1n}), \ldots, \mathrm{vec}(P_{mn})\big]$. The next term we shall consider is $\overline{G}_n\mathrm{E}\big(\varepsilon_n\mathcal{Z}_{2\bullet,n}\mathcal{F}_n\mathcal{Z}_{2\bullet,n}^{'}\varepsilon_n^{'}\big)\overline{G}_n^{'} = \overline{G}_n\mathrm{E}\big(\mathcal{T}_{1n}\big)\overline{G}_n^{'}$, where $\mathcal{T}_{1n} = \varepsilon_n\mathcal{Z}_{2\bullet,n}\mathcal{F}_n\mathcal{Z}_{2\bullet,n}^{'}\varepsilon_n^{'}$. Then, the (s,s)th element of $\mathrm{E}\left(\mathcal{T}_{1n}\right)$ is $\mathrm{E}\left(e_s^{'}\mathcal{T}_{1n}e_s\right) = \mathcal{Z}_{2\bullet,n}\mathrm{E}\left(\mathcal{F}_n\varepsilon_{sn}^2\right)\mathcal{Z}_{2\bullet,n}^{'}$ where $\mathrm{E}\left(\mathcal{F}_n\varepsilon_{sn}^2\right)$ is given in (4.2). Similarly, the (s,t)th element of $\mathrm{E}\left(\mathcal{T}_{1n}\right)$ is $\mathrm{E}\left(e_s^{'}\mathcal{T}_{1n}e_t\right) = \mathcal{Z}_{2\bullet,n}\mathrm{E}\left(\mathcal{F}_n\varepsilon_{sn}\varepsilon_{tn}\right)\mathcal{Z}_{2\bullet,n}^{'}$ where $\mathrm{E}\left(\mathcal{F}_n\varepsilon_{sn}\varepsilon_{tn}\right)$ is given in (4.3).

The last term we shall evaluate in (4.4) is $\big(\overline{G}_n\overline{X}_n\beta_0\big)\mathcal{Z}_{2\bullet,n}\mathrm{E}\big(\mathcal{F}_n\mathcal{Z}_{2\bullet,n}^{'}\varepsilon_n^{'}\big)\overline{G}_n^{'} = \big(\overline{G}_n\overline{X}_n\beta_0\big)\mathcal{Z}_{2\bullet,n}\mathrm{E}\big(\mathcal{T}_{2n}\big)\overline{G}_n^{'}$ where $\mathcal{T}_{2n} = \mathcal{F}_n\mathcal{Z}_{2\bullet,n}^{'}\varepsilon_n^{'}$. Let $\mathbf{e}_s$ be the sth elementary vector in $\mathbb{R}^{m+r}$ (and e_t is the tth elementary vector in $\mathbb{R}^n$). Then, the (s,t)th element of $\mathrm{E}\big(\mathcal{T}_{2n}\big)$ is given by $\mathrm{E}\big(\mathbf{e}_s^{'}\mathcal{T}_{2n}e_t\big) = \mathbf{e}_s^{'}\mathrm{E}\big(\mathcal{F}_n\varepsilon_{tn}\big)\mathcal{Z}_{2\bullet,n}^{'}$. By Lemma 2, we have

$$\mathrm{E}\big(\mathcal{F}_n\varepsilon_{tn}\big) = \begin{bmatrix} \sigma_0^2\mu_3\mathcal{O}_t^{'}\mathcal{O}_t & \sigma_0^4\mathcal{O}_t^{'}Q_n \\ \sigma_0^4Q_n^{'}\mathcal{O}_t & \mu_3Q_{t\bullet,n}^{'}Q_{t\bullet,n} \end{bmatrix} \tag{4.6}$$

where $\mathcal{O}_t = \big[\mathcal{O}_{t1},\mathcal{O}_{t2},\ldots,\mathcal{O}_{tm}\big]$ with $\mathcal{O}_{tj} = P_{\bullet t,jn}^{(s)} = \big[P_{1t,jn}^{(s)},P_{2t,jn}^{(s)},\ldots,P_{nt,jn}^{(s)}\big]^{'}$ for $j = 1,\ldots,m$.

Next, we shall work on (viii) $\mathrm{E}\big(\mathcal{K}_nE_3\mathcal{Z}_ng_n(\theta_0)g_n(\theta_0)^{'}\mathcal{Z}_n^{'}E_3^{'}\mathcal{K}_n^{'}\big) = R_nX_n\mathbb{Z}_{3n}\mathrm{E}\big(\mathcal{F}_n\big)\mathbb{Z}_{3n}^{'}X_n^{'}R_n^{'}$, where $\mathbb{Z}_{3n} = \big(\mathcal{Z}_{3\bullet,n}^{'},\ldots,\mathcal{Z}_{(k+2)\bullet,n}^{'}\big)^{'}$. By Lemma 1, we have

$$\mathrm{E}\big(\mathcal{F}_n\big) = \mathrm{Diag}\left(\sigma_0^4\Xi_{nm},\sigma_0^2Q_n^{'}Q_n\right).$$

Another term in (3.8) that we need to consider is (ix) $\mathrm{E}\big(\mathcal{K}_nE_2\mathcal{Z}_ng_n(\theta_0)g_n(\theta_0)^{'}\mathcal{Z}_n^{'}E_1^{'}\mathcal{K}_n^{'}\big)$, which can be written as

$$\begin{aligned}&\mathrm{E}\big(\mathcal{K}_nE_2\mathcal{Z}_ng_n(\theta_0)g_n(\theta_0)^{'}\mathcal{Z}_n^{'}E_1^{'}\mathcal{K}_n^{'}\big) = \big(\overline{G}_n\overline{X}_n\beta_0\big)\mathcal{Z}_{2\bullet,n}\mathrm{E}\big(\mathcal{F}_n\mathcal{Z}_{1\bullet,n}^{'}\varepsilon_n^{'}\big)H_n^{'}\\ &\quad+\overline{G}_n\mathrm{E}\big(\varepsilon_n\mathcal{Z}_{2\bullet,n}\mathcal{F}_n\mathcal{Z}_{1\bullet,n}^{'}\varepsilon_n^{'}\big)H_n^{'} = \big(\overline{G}_n\overline{X}_n\beta_0\big)\mathcal{Z}_{2\bullet,n}\mathrm{E}\left(\mathcal{T}_{3n}\right)H_n^{'} + \overline{G}_n\mathrm{E}\left(\mathcal{T}_{4n}\right)H_n^{'},\end{aligned}$$

where $\mathcal{T}_{3n} = \mathcal{F}_n\mathcal{Z}_{1\bullet,n}^{'}\varepsilon_n^{'}$ and $\mathcal{T}_{4n} = \varepsilon_n\mathcal{Z}_{2\bullet,n}\mathcal{F}_n\mathcal{Z}_{1\bullet,n}^{'}\varepsilon_n^{'}$. We start with $\mathrm{E}\left(\mathcal{T}_{3n}\right)$. The expectation of the (s,t)th element of $\mathcal{T}_{3n}$ for $s = 1,\ldots,m+r$ and $t = 1,\ldots,n$ is given by $\mathrm{E}\big(\mathbf{e}_s^{'}\mathcal{T}_{3n}e_t\big) = \mathbf{e}_s^{'}\mathrm{E}\left(\mathcal{F}_n\varepsilon_{tn}\right)\mathcal{Z}_{1\bullet,n}^{'}$, where $\mathrm{E}\left(\mathcal{F}_n\varepsilon_{tn}\right)$ is given in (4.6). Next, we shall evaluate the term involving $\mathcal{T}_{4n}$. Then, the (s,s)th element of $\mathrm{E}\left(\mathcal{T}_{4n}\right)$ is $\mathrm{E}\big(e_s^{'}\mathcal{T}_{4n}e_s\big) = \mathcal{Z}_{2\bullet,n}\mathrm{E}\left(\mathcal{F}_n\varepsilon_{sn}^2\right)\mathcal{Z}_{1\bullet,n}^{'}$, where $\mathrm{E}\left(\mathcal{F}_n\varepsilon_{sn}^2\right)$ is given in (4.2). Similarly, the (s,t)th element of $\mathrm{E}\left(\mathcal{T}_{4n}\right)$ is $\mathrm{E}\big(e_s^{'}\mathcal{T}_{4n}e_t\big) = \mathcal{Z}_{2\bullet,n}\mathrm{E}\left(\mathcal{F}_n\varepsilon_{sn}\varepsilon_{tn}\right)\mathcal{Z}_{1\bullet,n}^{'}$, where $\mathrm{E}\left(\mathcal{F}_n\varepsilon_{sn}\varepsilon_{tn}\right)$ is given in (4.3).

Another term in (3.8) that we need to consider is (x) $\mathrm{E}\left(\mathcal{K}_nE_3\mathcal{Z}_ng_n(\theta_0)g_n(\theta_0)^{'}\mathcal{Z}_n^{'}E_1^{'}\mathcal{K}_n^{'}\right)$. The expectation of this term is $R_nX_n\mathbb{Z}_{3n}\mathrm{E}\big(\mathcal{T}_{3n}\big)H_n^{'}$ where $\mathcal{T}_{3n} = \mathcal{F}_n\mathcal{Z}_{1\bullet,n}^{'}\varepsilon_n^{'}$ and $\mathbb{Z}_{3n} = \big(\mathcal{Z}_{3\bullet,n}^{'},\ldots,\mathcal{Z}_{(k+2)\bullet,n}^{'}\big)^{'}$. The calculation of the (s,t)th element of $\mathrm{E}\big(\mathcal{T}_{3n}\big)$ for $s = 1,\ldots,m+r$ and $t = 1,\ldots,n$ is illustrated in the preceding paragraph.

The last term we shall evaluate in (3.8) is (xi) $\mathrm{E}\left(\mathcal{K}_nE_2\mathcal{Z}_ng_n(\theta_0)g_n(\theta_0)^{'}\mathcal{Z}_n^{'}E_3^{'}\mathcal{K}_n^{'}\right)$. The expectation of this term is

$$\begin{aligned}\mathrm{E}\big(\mathcal{K}_nE_2\mathcal{Z}_ng_n(\theta_0)g_n(\theta_0)^{'}\mathcal{Z}_n^{'}E_3^{'}\mathcal{K}_n^{'}\big) &= \big(\overline{G}_n\overline{X}_n\beta_0\big)\mathcal{Z}_{2\bullet,n}\mathrm{E}\big(\mathcal{F}_n\big)\mathbb{Z}_{3n}^{'}X_n^{'}R_n^{'}\\ &\quad+\overline{G}_n\mathrm{E}\big(\varepsilon_n\mathcal{Z}_{2\bullet,n}\mathcal{F}_n\big)\mathbb{Z}_{3n}^{'}X_n^{'}R_n^{'}.\end{aligned} \tag{4.7}$$

In the first term on the right hand side of (4.7), we have $\mathrm{E}\left(\mathcal{F}_n\right) = \mathrm{Diag}\left(\sigma_0^4\Xi_{nm},\sigma_0^2Q_n^{'}Q_n\right)$. For the second term, let $\mathcal{T}_{5n} = \varepsilon_n\mathcal{Z}_{2\bullet,n}\mathcal{F}_n$. Furthermore, let $\mathbf{e}_t$ be tth elementary vector in $\mathbb{R}^{m+r}$ (and e_s is the sth elementary vector in $\mathbb{R}^n$). Then, the (s,t)th element of $\mathrm{E}\left(\mathcal{T}_{5n}\right)$ for $s = 1,\ldots,n$ and

$t = 1, \ldots, m+r$ is given by $\mathrm{E}\big(e_s^{'} \mathcal{T}_{5n} \mathbf{e}_t\big) = \mathcal{Z}_{2\bullet,n} \mathrm{E}\left(\mathcal{F}_n \varepsilon_{sn}\right) \mathbf{e}_t$. By Lemma 2, we obtain

$$\mathrm{E}\left(\mathcal{F}_n \varepsilon_{sn}\right) = \begin{bmatrix} \sigma_0^2 \mu_3 \mathcal{O}_s^{'} \mathcal{O}_s & \sigma_0^4 \mathcal{O}_s^{'} Q_n \\ \sigma_0^4 Q_n^{'} \mathcal{O}_s & \mu_3 Q_{s\bullet,n}^{'} Q_{s\bullet,n} \end{bmatrix} \tag{4.8}$$

where $\mathcal{O}_s = \left[\mathcal{O}_{s1}, \mathcal{O}_{s2}, \ldots, \mathcal{O}_{sm}\right]$ and $\mathcal{O}_{sj} = P_{\bullet s,jn}^{(s)} = \left[P_{1s,jn}^{(s)}, P_{2s,jn}^{(s)}, \ldots, P_{ns,jn}^{(s)}\right]^{'}$ for $j = 1, \ldots, m$.

The evaluations provided in the preceding paragraphs indicate that a consistent estimate of $\mathcal{H}_{1n}$ can be obtained once we have consistent estimates of θ_0, σ_0^2, $\mu_3 = \mathrm{E}\big(\varepsilon_{in}^3\big)$ and $\mu_4 = \mathrm{E}\big(\varepsilon_{in}^4\big)$. Hence, $\mathcal{H}_{1n}$ will be available once we have an initial robust GMME.

5 A Monte Carlo Study

5.1 Design

In order to study the finite sample properties of the suggested refinement methods, we design an extensive Monte Carlo study. For the model given in (2.1), we consider three regressors $X_n = (X_{1n}, X_{2n}, X_{3n})$ that are mutually independent vectors of independent standard normal random variables. We set $(\beta_{01}, \beta_{02}, \beta_{03})^{'} = (1\,, -1.2\,, -0.2)^{'}$ for all experiments. For the spatial autoregressive parameters, we employ combinations of $\{0.2,\, 0.6\}$ to allow for weak and strong spatial interactions. The weights matrix W_n and M_n are block diagonal matrices where each block is the row normalized contiguity matrix W_o from Anselin (1988)'s study of crimes across 49 districts of Columbus, Ohio. We consider 3 cases: (i) $W_n = M_n = W_o$, (ii) $W_n = M_n = I_2 \otimes W_o$, and (iii) $W_n = M_n = I_5 \otimes W_o$. These three cases yield, respectively, sample sizes of 49, 98 and 245.

Heteroskedasticity is incorporated using a skedastic function that maps household income values taken from the same Anselin, (1988) study onto $(0, \infty)$. More explicitly, let Income_{in} denote household income value (measured in thousand dollars) for the ith observation. Then, the disturbance terms are generated as $\varepsilon_{in} = \sigma_{in}\xi_{in}$ where $\xi_{in} \sim \text{i.i.d}\, N(0,1)$ and $\sigma_{in}^2 = \exp\left(0.1 + 0.05 \cdot \text{Income}_{in}\right)$. For the sample sizes 98 and 245, household income values are sampled randomly with replacement. Following Chesher and Jewitt, (1987), we measure the degree of heteroskedasticity as the ratio $\zeta = \max_i \left(\sigma_{in}^2\right) / \min_i \left(\sigma_{in}^2\right)$. Our data generating process yields a ζ value around 3.77.[6]

We use the following expression to measure the level of signal-to-noise in this set up (Pace et al., 2012):

$$R^2 = 1 - \frac{\mathrm{tr}\big(R_n^{-1'} S_n^{-1'} S_n^{-1} R_n^{-1} \Sigma_n\big)}{\beta_0^{'} X_n^{'} S_n^{-1'} S_n^{-1} X_n \beta_0 + \mathrm{tr}\big(R_n^{-1'} S_n^{-1'} S_n^{-1} R_n^{-1} \Sigma_n\big)}. \tag{5.1}$$

Our setup yields an R^2 value about 0.5, which is a reasonable level of goodness-of-fit. Resampling is carried out for 2000 times.

5.2 Simulation Results on Model Parameters

Our suggested SHC-corrections affect the point estimates of GMME through the weight matrix used in the GMM objective function. Therefore, we first evaluate the finite sample bias properties of the GMME based on various SHCs. The simulation results for the bias properties are presented

[6] MacKinnon, (2013) generates individual variances by $\sigma_i = z(\gamma)\left(\beta_1 + \sum_{k=2}^{5} \beta_k X_{ik}\right)^{\gamma}$, where $0 \leq \gamma \leq 2$ is a parameter used to determine the degree of heteroskedasticity. MacKinnon, (2013) states that $\gamma = 0$ implies homoskedasticity and $\gamma \geq 1$ implies extreme heteroskedasticity. Thus, a moderate degree of heteroskedasticity can be obtained by setting $\gamma = 0.5$, which generates a value of ζ around 4.

in Tables 1–2. The absolute average biases across different corrections methods are generally similar and small for all values of (λ_0, ρ_0). In all cases, $\widehat{\beta}_3$ reports relatively smaller bias. The results for the autoregressive parameters in Table 2 show that the estimators of these parameters report very low and similar biases across all methods and cases.

Next, we provide simulation results for the estimated asymptotic standard errors and the empirical standard deviations for each method. These results are provided in Tables 3–4. The results are easily interpretable if we highlight the difference between the estimated standard errors and the corresponding empirical deviations. To this end, we compute the percentage deviation of the mean absolute deviations of the estimated asymptotic standard errors from the corresponding empirical standard deviations.[7] In the following, we will refer to these measures simply as the percentage deviations. A small percentage deviation for an estimator suggests that its assumed distribution approximates the true finite sample distribution well enough.

The percentage deviations reported in Tables 3–4 are generally larger in the case of $SHC0$. In particular, the GMME of λ_0 and ρ_0 based on $SHC0$ reports relatively larger percentage deviations in all cases. The percentage deviations get smaller as the sample size gets larger in all cases. To give an overall picture, we can calculate the average percentage deviations across all λ_0 and ρ_0 values from the results presented in Tables 3–4 for each method. For example, for the GMME of β_1, the average percentage deviations are 8.3% for $SHC0$, 6.8% for $SHC1$, 6.1% for $SHC2$, 4.2% for $SHC3$, 4.6% for $SHC4$, 4.6% for $SHC5$, 9.1% for $SHC2^\star$, 2.3% for $SHC3^\star$, 2.8% for $SHC4^\star$ and 2.9% for $SHC5^\star$. For the GMME of λ_0, these averages are 17.9% for $SHC0$, 16.8% for $SHC1$, 15.7% for $SHC2$, 16.1% for $SHC3$, 16.1% for $SHC4$, 16.1% for $SHC5$, 16% for $SHC2^\star$, 12.3% for $SHC3^\star$, 11.9% for $SHC4^\star$ and 12% for $SHC5^\star$. Finally, for the GMME of ρ, these averages are 11.5% for $SHC0$, 11.7% for $SHC1$, 11.2% for $SHC2$, 10.7% for $SHC3$, 10.5% for $SHC4$, 10.5% for $SHC5$, 11.3% for $SHC2^\star$, 10.3% for $SHC3^\star$, 10.5% for $SHC4^\star$ and 10.6% for $SHC5^\star$. These results indicate that the small-sample corrections $SHC3^\star$, $SHC4^\star$ and $SHC5^\star$ perform relatively better than the other methods.

We use the P value discrepancy plots to illustrate the size properties of standard Wald test formulated from the corrections methods. Figures 1 through 5 display the discrepancy between the actual size of the Wald test and its nominal size. In these figures, the nominal size values, depicted on the x-axis, span from 1% to 10%, and the discrepancies are reported for our three sample size next to each other in the same plot. For the null hypotheses $H_0 : \beta_1 = 1$, $H_0 : \beta_2 = -1.2$ and $H_0 : \beta_3 = -0.2$, there are large size distortions for the Wald tests based on $SHC0$ when $n = 49$ and $n = 98$. Figures 1 through 3 indicate that the Wald tests for the coefficients of the exogenous variables, generally, over reject under all methods and in all cases. However, the rejection rates based on the finite-sample corrections $SHC2^\star - SHC5^\star$ are much closer to the nominal sizes than the other methods in all cases. This conclusion is consistent with the results presented in Tables 3 through 4, where the percentage deviations reported are relatively smaller in the case of $SHC2^\star - SHC5^\star$. Finally, the performance of $SHC1 - SHC5$ is, generally, better than $SHC0$, but worse than $SHC2^\star - SHC5^\star$.

The P value discrepancy plots for the Wald tests of autoregressive parameters are given in Figures 4 and 5. The rejection rates reported in these figures are larger than the corresponding nominal sizes, especially when $n = 49$ and $n = 98$. In Figure 4, the correction methods $SHC3^\star - SHC5^\star$ outperform the other methods in all cases. Hence, these methods can be useful for testing λ_0. The P value discrepancy plots for the null hypotheses involving ρ_0 are given in Figure 5. When $n = 49$ and $n = 98$, the correction methods $SHC3^\star - SHC5^\star$ outperform the other methods in

[7] In our Monte Carlo set up, let y_i be the estimated standard errors for an estimator in the ith repetition and y be the calculated empirical standard deviation of the same estimator across all resamples. Then, we compute this scalar measure by $100 \times |\text{Median}\,(y_i) - y|\,/y$.

Table 1: Bias Properties of $\widehat{\beta}_1$, $\widehat{\beta}_2$ and $\widehat{\beta}_3$

					Bias of $\widehat{\beta}_1$						
n	$\rho_0 - \lambda_0$	$SHC0$	$SHC1$	$SHC2$	$SHC3$	$SHC4$	$SHC5$	$SHC2^{\star}$	$SHC3^{\star}$	$SHC4^{\star}$	$SHC5^{\star}$
49	0.2–0.2	-0.0176	-0.0176	-0.0179	-0.0179	-0.0173	-0.0173	-0.0183	-0.0181	-0.0200	-0.0200
	0.2–0.6	-0.0094	-0.0081	-0.0089	-0.0087	-0.0076	-0.0076	-0.0078	-0.0080	-0.0077	-0.0084
	0.6–0.2	-0.0251	-0.0250	-0.0220	-0.0227	-0.0211	-0.0211	-0.0231	-0.0259	-0.0326	-0.0322
	0.6–0.6	-0.0195	-0.0193	-0.0185	-0.0184	-0.0177	-0.0177	-0.0230	-0.0230	-0.0233	-0.0205
98	0.2–0.2	-0.0202	-0.0201	-0.0202	-0.0202	-0.0206	-0.0206	-0.0199	-0.0198	-0.0204	-0.0204
	0.2–0.6	-0.0034	-0.0034	-0.0036	-0.0037	-0.0037	-0.0037	-0.0036	-0.0037	-0.0035	-0.0035
	0.6–0.2	-0.0226	-0.0220	-0.0209	-0.0211	-0.0211	-0.0211	-0.0214	-0.0210	-0.0207	-0.0207
	0.6–0.6	-0.0158	-0.0160	-0.0160	-0.0152	-0.0182	-0.0182	-0.0160	-0.0155	-0.0173	-0.0167
245	0.2–0.2	-0.0065	-0.0065	-0.0065	-0.0064	-0.0065	-0.0065	-0.0065	-0.0065	-0.0064	-0.0064
	0.2–0.6	-0.0027	-0.0027	-0.0026	-0.0027	-0.0027	-0.0027	-0.0026	-0.0027	-0.0027	-0.0027
	0.6–0.2	-0.0031	-0.0030	-0.0030	-0.0033	-0.0031	-0.0031	-0.0031	-0.0033	-0.0034	-0.0031
	0.6–0.6	-0.0045	-0.0045	-0.0046	-0.0046	-0.0049	-0.0049	-0.0045	-0.0046	-0.0045	-0.0044
					Bias of $\widehat{\beta}_2$						
49	0.2–0.2	0.0237	0.0243	0.0241	0.0233	0.0236	0.0236	0.0244	0.0237	0.0238	0.0238
	0.2–0.6	0.0252	0.0251	0.0249	0.0258	0.0248	0.0248	0.0245	0.0241	0.0250	0.0244
	0.6–0.2	0.0272	0.0265	0.0262	0.0279	0.0273	0.0273	0.0279	0.0303	0.0380	0.0382
	0.6–0.6	0.0391	0.0365	0.0381	0.0358	0.0369	0.0369	0.0391	0.0427	0.0404	0.0378
98	0.2–0.2	0.0122	0.0119	0.0119	0.0117	0.0117	0.0117	0.0117	0.0117	0.0116	0.0116
	0.2–0.6	0.0125	0.0129	0.0126	0.0125	0.0125	0.0125	0.0123	0.0127	0.0125	0.0125
	0.6–0.2	0.0075	0.0067	0.0058	0.0064	0.0055	0.0055	0.0072	0.0073	0.0061	0.0061
	0.6–0.6	0.0196	0.0219	0.0200	0.0200	0.0222	0.0222	0.0220	0.0209	0.0200	0.0198
245	0.2–0.2	0.0056	0.0056	0.0056	0.0056	0.0056	0.0056	0.0056	0.0056	0.0056	0.0056
	0.2–0.6	0.0030	0.0030	0.0030	0.0030	0.0031	0.0031	0.0030	0.0030	0.0030	0.0030
	0.6–0.2	-0.0000	-0.0000	-0.0000	0.0000	0.0001	0.0001	0.0001	0.0001	-0.0000	-0.0001
	0.6–0.6	0.0026	0.0026	0.0026	0.0025	0.0026	0.0026	0.0026	0.0025	0.0027	0.0026
					Bias of $\widehat{\beta}_3$						
49	0.2–0.2	0.0087	0.0090	0.0087	0.0088	0.0088	0.0088	0.0089	0.0088	0.0088	0.0088
	0.2–0.6	-0.0009	-0.0011	-0.0010	-0.0008	-0.0015	-0.0015	-0.0005	-0.0012	-0.0011	-0.0011
	0.6–0.2	0.0068	0.0065	0.0059	0.0058	0.0056	0.0056	0.0059	0.0061	0.0090	0.0095
	0.6–0.6	0.0042	0.0034	0.0039	0.0034	0.0032	0.0032	0.0044	0.0064	0.0101	0.0101
98	0.2–0.2	0.0034	0.0033	0.0034	0.0034	0.0035	0.0035	0.0033	0.0032	0.0033	0.0033
	0.2–0.6	-0.0007	-0.0006	-0.0006	-0.0005	-0.0005	-0.0005	-0.0007	-0.0006	-0.0004	-0.0004
	0.6–0.2	0.0018	0.0018	0.0022	0.0022	0.0023	0.0023	0.0023	0.0027	0.0023	0.0023
	0.6–0.6	0.0035	0.0044	0.0046	0.0044	0.0045	0.0045	0.0049	0.0043	0.0052	0.0051
245	0.2–0.2	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015	0.0015
	0.2–0.6	0.0027	0.0028	0.0027	0.0027	0.0027	0.0027	0.0027	0.0028	0.0028	0.0028
	0.6–0.2	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004
	0.6–0.6	0.0050	0.0050	0.0050	0.0050	0.0050	0.0050	0.0049	0.0049	0.0049	0.0050

Figures 5(a) and 5(b), however there is no discernible differences across methods in Figures 5(c) and 5(d). This result indicates that the degree of spatial dependence in the disturbance term can

Table 2: Bias Properties of $\widehat{\lambda}$ and $\widehat{\rho}$

n	$\rho_0 - \lambda_0$	$SHC0$	$SHC1$	$SHC2$	$SHC3$	$SHC4$	$SHC5$	$SHC2^\star$	$SHC3^\star$	$SHC4^\star$	$SHC5^\star$
					Bias of $\widehat{\lambda}$						
49	0.2–0.2	0.0471	0.0454	0.0468	0.0469	0.0461	0.0461	0.0466	0.0461	0.0442	0.0446
	0.2–0.6	-0.0003	-0.0001	-0.0006	-0.0007	-0.0000	-0.0000	-0.0000	0.0003	0.0012	0.0027
	0.6–0.2	0.0054	0.0001	0.0074	0.0041	0.0093	0.0093	0.0056	-0.0029	-0.0205	-0.0213
	0.6–0.6	0.0385	0.0399	0.0384	0.0414	0.0400	0.0400	0.0352	0.0344	0.0315	0.0311
98	0.2–0.2	0.0101	0.0093	0.0092	0.0093	0.0095	0.0095	0.0093	0.0085	0.0075	0.0075
	0.2–0.6	-0.0092	-0.0094	-0.0088	-0.0098	-0.0090	-0.0090	-0.0083	-0.0091	-0.0093	-0.0089
	0.6–0.2	-0.0143	-0.0143	-0.0140	-0.0134	-0.0149	-0.0149	-0.0152	-0.0164	-0.0128	-0.0128
	0.6–0.6	0.0030	0.0026	0.0068	0.0057	0.0022	0.0022	0.0016	0.0028	0.0073	0.0089
245	0.2–0.2	0.0044	0.0044	0.0044	0.0043	0.0043	0.0043	0.0044	0.0044	0.0044	0.0044
	0.2–0.6	-0.0029	-0.0028	-0.0028	-0.0027	-0.0028	-0.0028	-0.0027	-0.0027	-0.0029	-0.0029
	0.6–0.2	0.0106	0.0110	0.0108	0.0101	0.0102	0.0102	0.0109	0.0109	0.0103	0.0106
	0.6–0.6	0.0050	0.0049	0.0050	0.0049	0.0048	0.0048	0.0050	0.0049	0.0049	0.0050
					Bias of $\widehat{\rho}$						
49	0.2–0.2	-0.0301	-0.0283	-0.0291	-0.0296	-0.0287	-0.0287	-0.0267	-0.0253	-0.0203	-0.0214
	0.2–0.6	0.0198	0.0201	0.0195	0.0208	0.0177	0.0177	0.0190	0.0216	0.0205	0.0148
	0.6–0.2	0.0181	0.0230	0.0173	0.0203	0.0171	0.0171	0.0196	0.0218	0.0405	0.0431
	0.6–0.6	-0.0098	-0.0114	-0.0089	-0.0123	-0.0111	-0.0111	-0.0030	0.0090	0.0138	0.0110
98	0.2–0.2	-0.0014	-0.0015	-0.0015	-0.0015	-0.0014	-0.0014	-0.0015	-0.0000	0.0002	0.0002
	0.2–0.6	0.0209	0.0213	0.0208	0.0214	0.0196	0.0196	0.0196	0.0201	0.0214	0.0214
	0.6–0.2	0.0149	0.0158	0.0157	0.0162	0.0163	0.0163	0.0156	0.0169	0.0145	0.0145
	0.6–0.6	0.0049	0.0061	0.0026	0.0032	0.0077	0.0077	0.0088	0.0077	-0.0008	-0.0012
245	0.2–0.2	-0.0046	-0.0046	-0.0047	-0.0047	-0.0047	-0.0047	-0.0047	-0.0048	-0.0048	-0.0048
	0.2–0.6	0.0089	0.0088	0.0088	0.0088	0.0088	0.0088	0.0089	0.0093	0.0093	0.0094
	0.6–0.2	-0.0022	-0.0023	-0.0023	-0.0021	-0.0020	-0.0020	-0.0023	-0.0022	-0.0021	-0.0023
	0.6–0.6	0.0037	0.0038	0.0038	0.0038	0.0039	0.0039	0.0038	0.0039	0.0041	0.0039

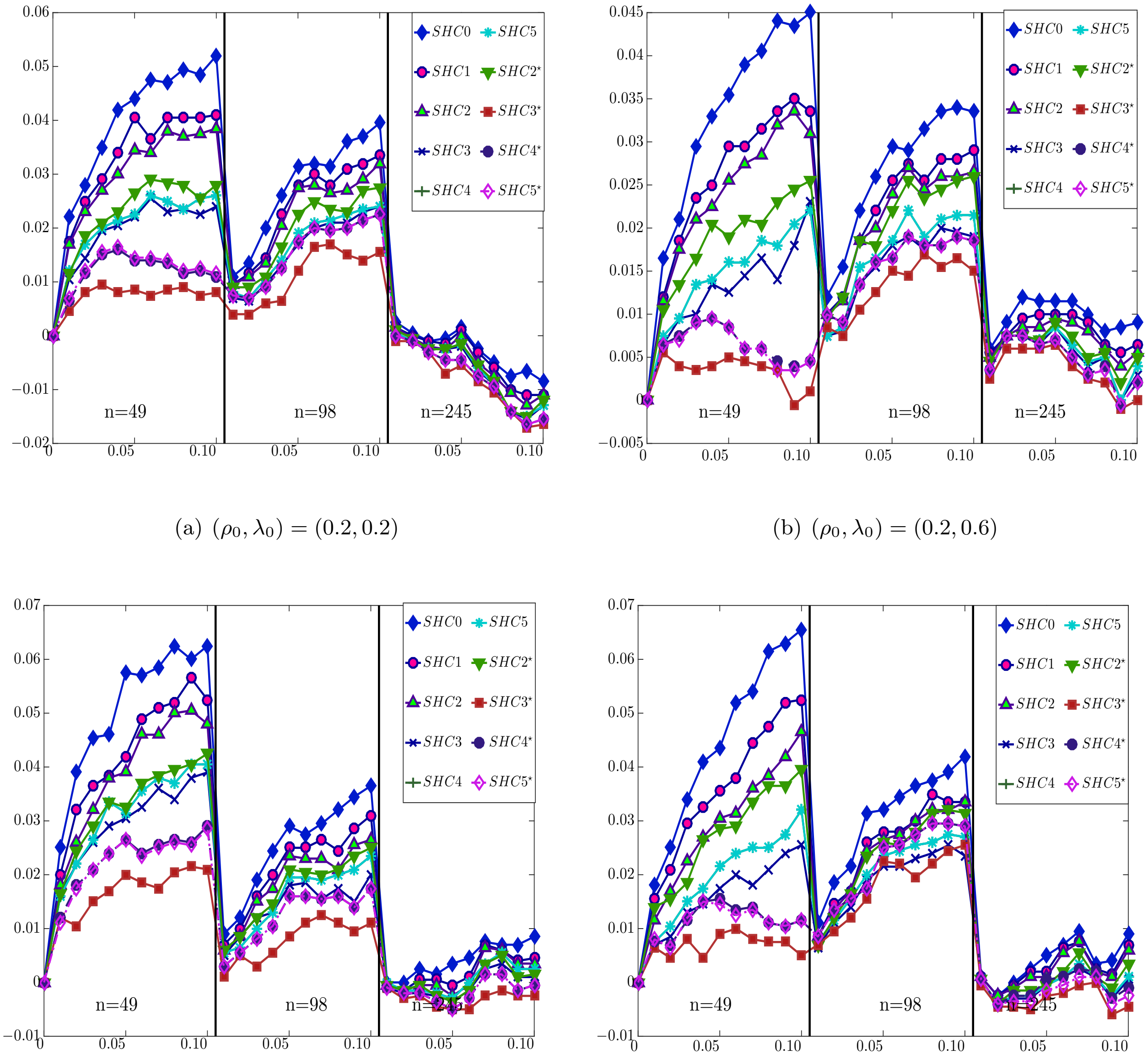
affect the size distortions across the correction methods.

Table 3: Percentage Deviations for $\widehat{\beta}_1$, $\widehat{\beta}_2$ and $\widehat{\beta}_3$

n	$\rho_0-\lambda_0$	$SHC0$	$SHC1$	$SHC2$	$SHC3$	$SHC4$	$SHC5$	$SHC2^\star$	$SHC3^\star$	$SHC4^\star$	$SHC5^\star$
Percentage of Mean Absolute Deviation of Estimated Standard Errors from Empirical Std: $\widehat{\beta}_1$											
49	0.2–0.2	13.2667	10.6371	9.2996	5.6339	6.4820	6.4820	6.4346	0.6529	0.0704	0.4051
	0.2–0.6	12.0230	9.1701	8.3114	4.4950	5.1852	5.1852	30.7138	1.4188	0.2796	0.2160
	0.6–0.2	17.0060	14.6927	12.3445	9.4550	10.4996	10.4996	10.3652	3.6481	7.1785	7.1735
	0.6–0.6	13.8752	11.0963	9.9654	5.9833	6.6256	6.6256	32.2774	0.7663	1.9994	3.1355
98	0.2–0.2	8.4393	7.0206	6.5489	4.3236	4.5755	4.5755	5.4098	2.4676	4.0313	3.9785
	0.2–0.6	8.7754	7.2755	6.7879	4.8435	5.2618	5.2618	6.1673	3.3320	4.3587	4.3684
	0.6–0.2	8.6986	7.2209	6.5813	5.1273	5.2276	5.2276	5.5981	2.0853	4.0842	4.0842
	0.6–0.6	9.7171	8.1310	7.7069	5.3481	6.0051	6.0051	7.1967	4.8111	5.8004	5.8433
245	0.2–0.2	2.1263	2.7605	3.0668	3.9815	3.8211	3.8211	3.6357	5.1416	4.4677	4.4679
	0.2–0.6	2.2666	1.6626	1.3679	0.4846	0.6286	0.6286	0.8679	0.5794	0.1492	0.1244
	0.6–0.2	1.4264	0.5209	0.2772	0.5725	0.1114	0.1114	0.2465	1.5986	0.9869	1.0486
	0.6–0.6	2.2041	1.6015	1.2886	0.3579	0.5033	0.5033	0.7795	0.6648	0.3325	0.0465
Percentage of Mean Absolute Deviation of Estimated Standard Errors from Empirical Std: $\widehat{\beta}_2$											
49	0.2–0.2	13.8142	11.3223	9.7158	4.9737	3.2435	3.2435	7.0803	0.4837	1.6411	1.6597
	0.2–0.6	15.7168	13.3847	12.1344	7.0007	5.9131	5.9131	34.7495	2.6743	3.6777	3.6465
	0.6–0.2	15.1191	12.8104	10.8487	7.2079	4.5892	4.5892	8.8242	2.2385	5.5089	5.4121
	0.6–0.6	13.9850	11.5546	9.8517	5.1401	3.4030	3.4030	32.0458	0.5686	0.1174	0.6514
98	0.2–0.2	9.5390	8.2922	7.7338	5.3648	5.4997	5.4997	6.3064	3.3708	4.2437	4.4806
	0.2–0.6	7.4752	6.1228	5.4529	3.5354	2.9294	2.9294	4.8190	1.6418	2.6640	2.6998
	0.6–0.2	11.5042	10.3292	9.8267	7.8618	8.2493	8.2493	8.6847	5.7844	6.5317	6.5312
	0.6–0.6	11.6833	9.8500	9.9648	8.0205	8.3952	8.3952	8.9095	6.3631	7.9200	7.7941
245	0.2–0.2	3.5566	2.9579	2.6457	1.6730	1.6921	1.6921	2.1670	0.7031	1.3296	1.3296
	0.2–0.6	2.3937	1.7929	1.4724	0.5335	0.5237	0.5237	0.9998	0.4486	0.2260	0.2117
	0.6–0.2	1.8170	1.0847	0.8747	0.0990	0.1131	0.1131	0.2955	1.0198	0.1144	0.0392
	0.6–0.6	2.2427	1.6363	1.3049	0.3679	0.5262	0.5262	0.7982	0.5269	0.2555	0.0506
Percentage of Mean Absolute Deviation of Estimated Standard Errors from Empirical Std: $\widehat{\beta}_3$											
49	0.2–0.2	15.3437	12.0550	11.4189	6.5697	6.7268	6.7268	8.1641	1.4487	3.9340	3.8760
	0.2–0.6	16.3313	13.2693	11.9498	8.2776	7.7108	7.7108	32.1175	2.2914	3.4331	3.3202
	0.6–0.2	10.9722	7.5155	6.5279	1.1002	1.7594	1.7594	3.2602	5.0743	3.5857	3.7416
	0.6–0.6	8.8256	5.6368	4.1728	0.4347	0.0070	0.0070	25.8709	7.1486	8.3256	8.6624
98	0.2–0.2	11.0722	9.7050	8.7315	6.4732	6.4711	6.4711	7.6905	4.2422	5.3824	5.3386
	0.2–0.6	9.9533	8.4846	7.7902	5.7620	5.9486	5.9486	7.1853	4.2623	5.1958	5.1924
	0.6–0.2	6.7043	5.3732	4.5893	2.4671	3.1075	3.1075	3.3618	0.0300	1.8412	1.8411
	0.6–0.6	7.9092	6.0482	5.3323	3.3586	3.7662	3.7662	4.6139	2.4716	3.4754	3.4258
245	0.2–0.2	3.1858	2.5893	2.2107	1.2047	1.0076	1.0076	1.5728	0.0898	0.1787	0.1789
	0.2–0.6	3.4818	2.8796	2.4714	1.4894	1.3387	1.3387	1.9373	0.3860	0.7725	0.7452
	0.6–0.2	3.8497	3.4978	3.1065	1.9307	1.7349	1.7349	2.4034	0.6600	0.9899	1.0822
	0.6–0.6	0.1945	0.8092	1.2134	2.2070	2.4420	2.4420	1.8577	3.5661	3.4422	3.4489

Table 4: Percentage Deviations for $\widehat{\lambda}$ and $\widehat{\rho}$

n	$\rho_0 - \lambda_0$	$SHC0$	$SHC1$	$SHC2$	$SHC3$	$SHC4$	$SHC5$	$SHC2^\star$	$SHC3^\star$	$SHC4^\star$	$SHC5^\star$
Percentage of Mean Absolute Deviation of Estimated Standard Errors from Empirical Std: $\widehat{\lambda}$											
49	0.2–0.2	21.2885	19.8722	19.2331	16.4778	18.0530	18.0530	15.1333	9.7803	9.1865	9.2468
	0.2–0.6	29.7537	29.4164	28.4406	28.3134	24.7199	24.7199	26.8728	17.5841	15.2319	15.9147
	0.6–0.2	33.1772	32.6374	30.3112	30.3300	29.5974	29.5974	27.7123	24.9730	26.0197	25.8477
	0.6–0.6	32.4412	31.2069	30.5103	26.9563	31.9509	31.9509	39.2427	24.4933	22.4891	20.4610
98	0.2–0.2	8.4635	7.1472	7.3013	5.3475	6.3845	6.3845	4.4771	1.4815	1.0772	1.5960
	0.2–0.6	29.4856	28.1784	28.7567	27.8222	27.8712	27.8712	27.1963	25.1345	25.0432	25.0399
	0.6–0.2	18.8646	17.4953	17.7758	17.4527	17.3959	17.3959	16.8203	14.6509	14.1312	14.1307
	0.6–0.6	24.1198	22.3079	23.8712	22.3083	22.9481	22.9481	21.4568	18.8493	20.1847	19.8967
245	0.2–0.2	3.1530	2.5446	2.4101	1.7720	1.9965	1.9965	1.3822	0.4144	0.3986	0.3987
	0.2–0.6	9.5314	8.9725	8.9488	8.3791	8.6421	8.6421	8.1292	6.5613	6.7964	6.7252
	0.6–0.2	3.4639	2.3481	2.5628	2.9782	3.1173	3.1173	1.9422	0.8056	1.5711	1.0520
	0.6–0.6	0.6287	0.0291	0.0639	0.7304	0.2896	0.2896	1.0714	2.3414	0.8504	3.6303
Percentage of Mean Absolute Deviation of Estimated Standard Errors from Empirical Std: $\widehat{\rho}$											
49	0.2–0.2	10.4973	11.0453	9.2311	7.3595	7.2145	7.2145	8.0059	2.2600	1.9996	2.1870
	0.2–0.6	2.9319	1.7135	1.4363	0.1059	0.7515	0.7515	2.2544	1.3480	5.5737	5.8609
	0.6–0.2	32.3294	32.1195	29.1765	28.8492	27.2187	27.2187	29.4601	25.6331	26.6675	26.8690
	0.6–0.6	11.6173	11.4964	11.4232	9.1080	9.6350	9.6350	9.2391	6.3381	4.8103	3.6848
98	0.2–0.2	1.2867	2.1325	2.1865	3.5304	2.5797	2.5797	3.5107	5.2994	5.4982	5.3109
	0.2–0.6	1.0778	1.8317	1.2633	1.6829	1.9734	1.9734	2.7480	4.0457	3.7090	3.7330
	0.6–0.2	22.2224	21.7561	22.0248	21.5038	21.1317	21.1317	21.3796	20.2613	18.5219	18.5214
	0.6–0.6	9.9990	9.9521	9.4537	8.0799	9.6315	9.6315	8.6571	5.1320	7.0567	6.7768
245	0.2–0.2	14.9026	15.2531	15.2790	15.6397	15.4993	15.4993	15.9196	16.9738	17.1381	17.1380
	0.2–0.6	14.5739	14.9098	14.9218	15.2879	15.1807	15.1807	15.4848	16.2517	16.1923	16.2098
	0.6–0.2	4.2323	5.3852	4.7011	3.3116	2.8142	2.8142	4.1983	5.1340	4.9666	5.4973
	0.6–0.6	12.8692	13.3165	13.3389	13.7668	12.8752	12.8752	14.9336	14.8308	14.3158	15.8317



(a) $(\rho_0, \lambda_0) = (0.2, 0.2)$

(b) $(\rho_0, \lambda_0) = (0.2, 0.6)$

(c) $(\rho_0, \lambda_0) = (0.6, 0.2)$

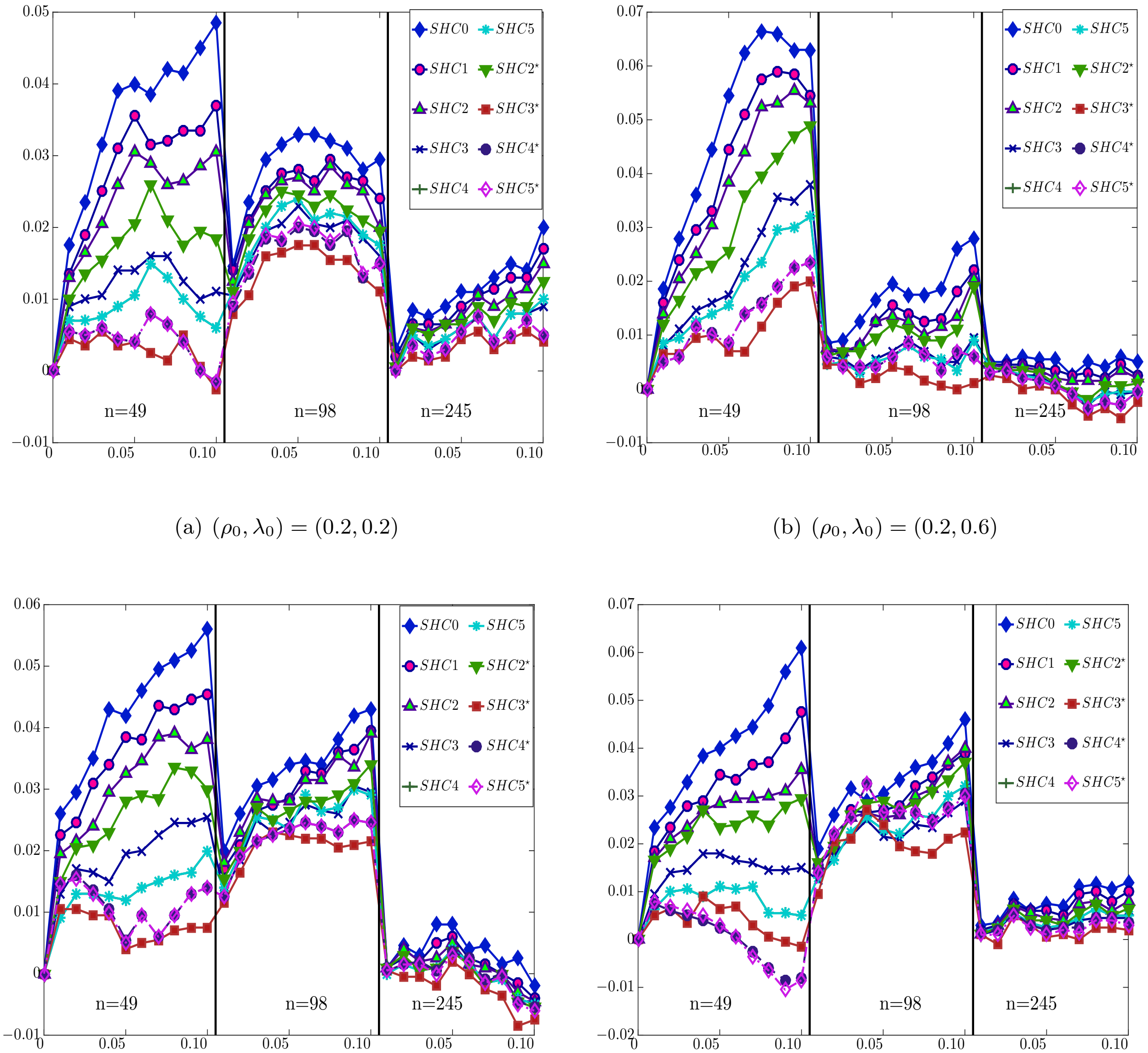
(d) $(\rho_0, \lambda_0) = (0.6, 0.6)$

Figure 1: P value discrepancy plots: $H_0 : \beta_1 = 1$

322 ### 5.3 Simulation Results on Effects Estimates

In this section, we investigate the effect of correction methods on the effects estimates (or marginal effects) of exogenous variables within the context of our spatial model. First, we describe how these marginal effects (impact measures) and their dispersions can be calculated. The marginal effect of a change in X_{kn} is given by the following $n \times n$ matrix:

$$\frac{\partial Y_n}{\partial X_{kn}'} = S_n^{-1}\beta_{k0}, \tag{5.2}$$



(a) $(\rho_0, \lambda_0) = (0.2, 0.2)$

(b) $(\rho_0, \lambda_0) = (0.2, 0.6)$

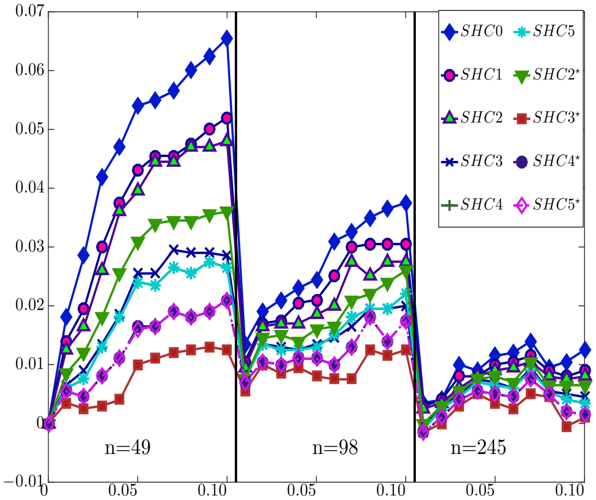
(c) $(\rho_0, \lambda_0) = (0.6, 0.2)$

(d) $(\rho_0, \lambda_0) = (0.6, 0.6)$

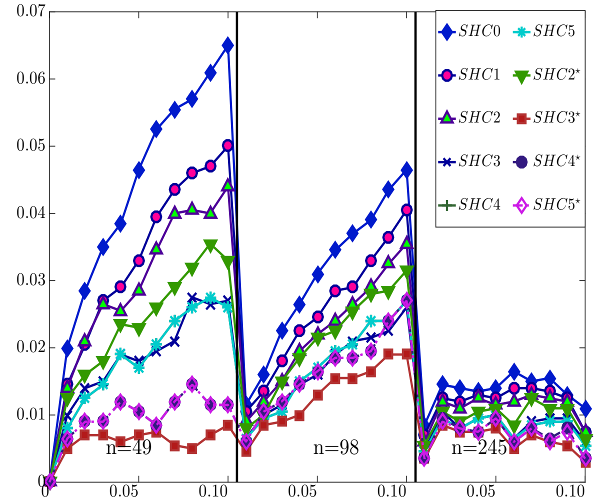
Figure 2: P value discrepancy plots: $H_0 : \beta_2 = -1.2$

where β_{k0} is the kth component of β_0. The diagonal elements of this matrix $\left(\partial Y_{in}/\partial X_{k,in}\right)$ contain the own-partial derivatives, while the off-diagonal elements represent the cross-partial derivatives $\left(\partial Y_{jn}/\partial X_{k,in}\right)$. 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.

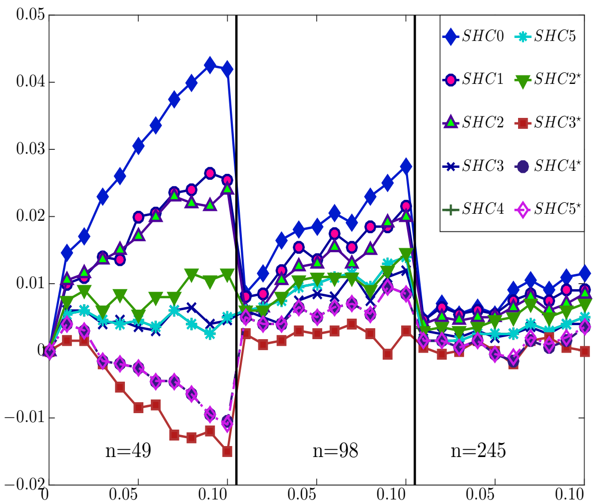
We consider the Delta method for the calculation of dispersions of these impact measures (Debarsy et al., 2015; Taspinar et al., 2016). The result in (5.2) indicates that the estimator



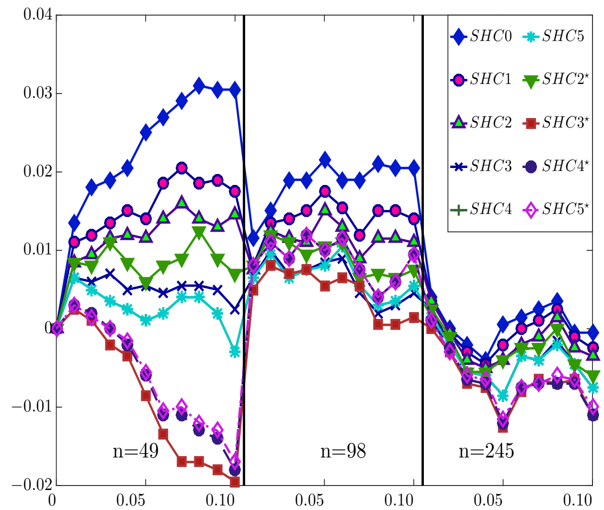
(a) $(\rho_0, \lambda_0) = (0.2, 0.2)$



(b) $(\rho_0, \lambda_0) = (0.2, 0.6)$



(c) $(\rho_0, \lambda_0) = (0.6, 0.2)$



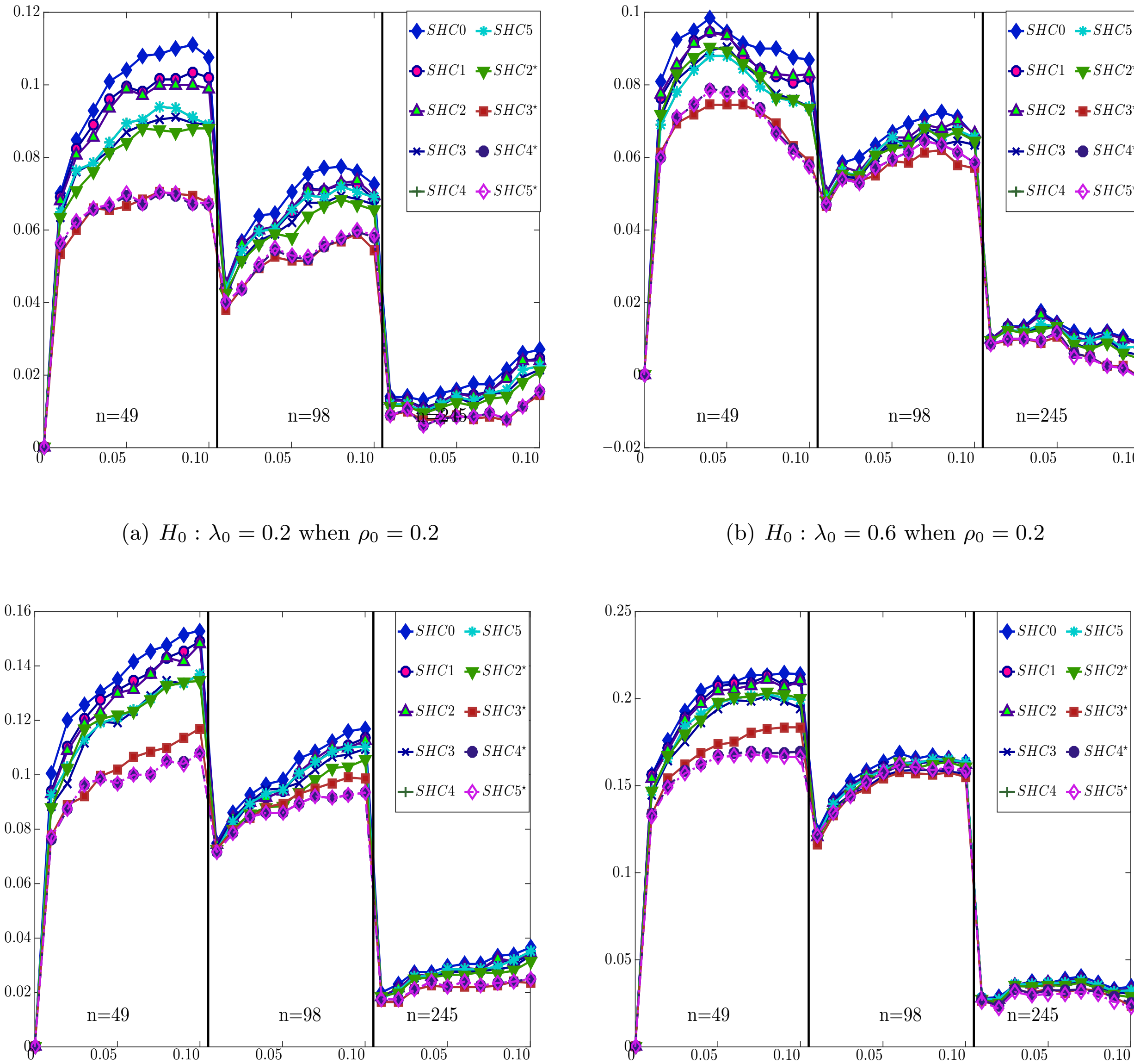
(d) $(\rho_0, \lambda_0) = (0.6, 0.6)$

Figure 3: P value discrepancy plots: $H_0 : \beta_2 = -0.2$

of direct effect is $\frac{1}{n}\mathrm{tr}\big(S_n^{-1}(\widehat{\lambda}_n)\widehat{\beta}_{kn}\big)$. By the mean value theorem,

$$\frac{1}{\sqrt{n}}\big[\mathrm{tr}\big(S_n^{-1}(\widehat{\lambda}_n)\widehat{\beta}_{kn}\big) - \mathrm{tr}\big(S_n^{-1}\beta_{k0}\big)\big] = A_{1n} \times \sqrt{n}\big(\widehat{\lambda}_n - \lambda_0, \widehat{\beta}_{kn} - \beta_{k0}\big)^{'} + o_p(1)$$
$$\xrightarrow{d} N\big(0, \lim_{n\to\infty} A_{1n}B_nA_{1n}^{'}\big), \tag{5.3}$$

where $A_{1n} = \big[\frac{1}{n}\mathrm{tr}\left(S_n^{-1}G_n\beta_{k0}\right), \frac{1}{n}\mathrm{tr}\left(S_n^{-1}\right)\big]$, and B_n is the asymptotic covariance of $\sqrt{n}\big(\widehat{\lambda}_n - \lambda_0, \widehat{\beta}_{kn} - \beta_{k0}\big)^{'}$. The result in (5.3) indicates that the asymptotic variance of direct effects can be



(a) $H_0 : \lambda_0 = 0.2$ when $\rho_0 = 0.2$

(b) $H_0 : \lambda_0 = 0.6$ when $\rho_0 = 0.2$

(c) $H_0 : \lambda_0 = 0.2$ when $\rho_0 = 0.6$

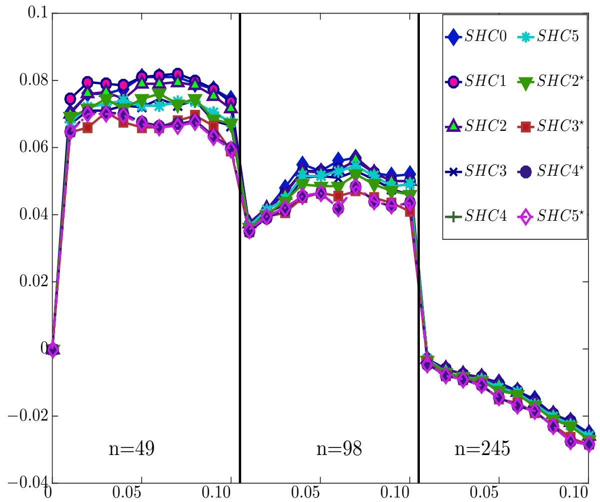
(d) $H_0 : \lambda_0 = 0.6$ when $\rho_0 = 0.6$

Figure 4: P value discrepancy plots

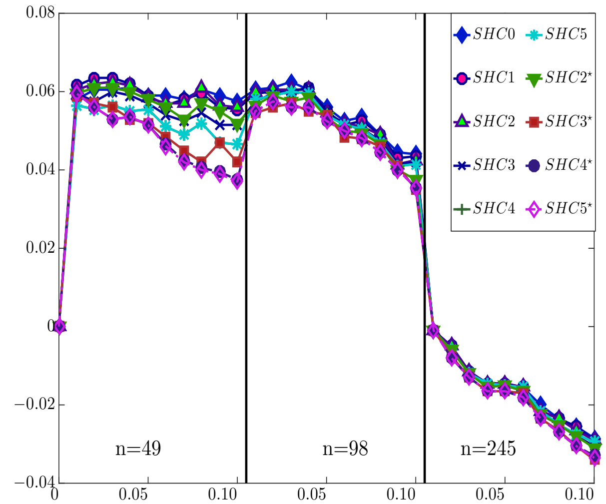
estimated by $\frac{1}{n}\widehat{A}_{1n}\widehat{B}_n\widehat{A}_{1n}^{'}$, where $\widehat{A}_{1n} = \left[\frac{1}{n}\mathrm{tr}\big(S_n^{-1}(\widehat{\lambda}_n)G_n(\widehat{\lambda}_n)\widehat{\beta}_{kn}\big),\ \frac{1}{n}\mathrm{tr}\big(S_n^{-1}(\widehat{\lambda}_n)\big)\right]$, and $\widehat{B}_n$ is the estimated asymptotic covariance of $\sqrt{n}\big(\widehat{\lambda}_n - \lambda_0,\, \widehat{\beta}_{kn} - \beta_{k0}\big)^{'}$.

Applying the mean value theorem to the estimator of total effects $\frac{1}{n}\widehat{\beta}_{kn}l_n^{'}S_n^{-1}(\widehat{\lambda}_n)l_n$ yields

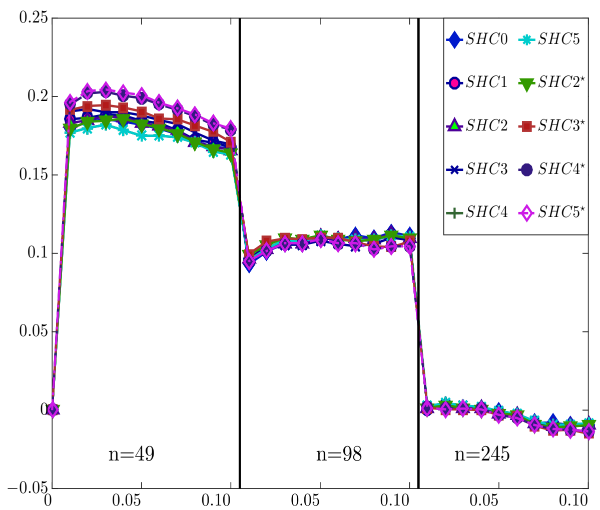
$$\begin{aligned}\frac{1}{\sqrt{n}}\left[\widehat{\beta}_{kn}l_n^{'}S_n^{-1}(\widehat{\lambda}_n)l_n - \beta_{k0}l_n^{'}S_n^{-1}l_n\right] &= A_{2n} \times \sqrt{n}\big(\widehat{\lambda}_n - \lambda_0, \widehat{\beta}_{kn} - \beta_{k0}\big)^{'} + o_p(1) \\ &\xrightarrow{d} N\big(0, \lim_{n\to\infty} A_{2n}B_nA_{2n}^{'}\big), \qquad (5.4)\end{aligned}$$



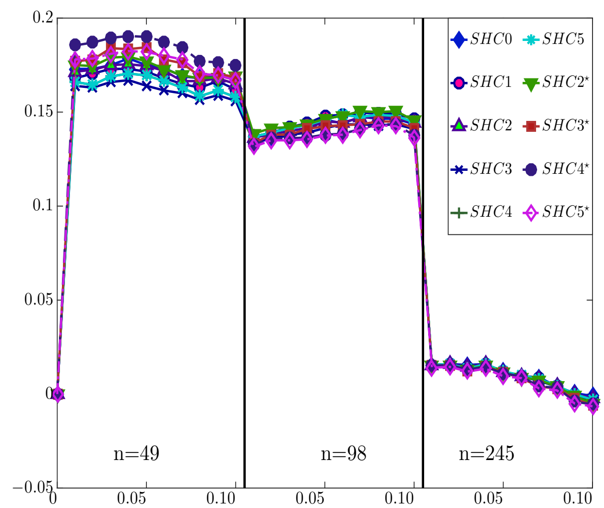
(a) $H_0 : \rho_0 = 0.2$ when $\lambda_0 = 0.2$



(b) $H_0 : \rho_0 = 0.2$ when $\lambda_0 = 0.6$



(c) $H_0 : \rho_0 = 0.6$ when $\lambda_0 = 0.2$



(d) $H_0 : \rho_0 = 0.6$ when $\lambda_0 = 0.6$

Figure 5: P value discrepancy plots

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, $\mathrm{Var}\left(\frac{1}{n}\widehat{\beta}_{kn} l_n' S_n^{-1}(\widehat{\lambda}_n) l_n\right)$ can be estimated by $\frac{1}{n}\widehat{A}_{2n}\widehat{B}_n\widehat{A}_{2n}'$, where $\widehat{A}_{2n} = \left[\frac{1}{n}\widehat{\beta}_{kn} l_n' S_n^{-1}(\widehat{\lambda}_n) G_n(\widehat{\lambda}_n) l_n,\ \frac{1}{n} l_n' S_n^{-1}(\widehat{\lambda}) l_n\right]$.

The estimate of indirect effects is given by $\frac{1}{n}\left[\widehat{\beta}_{kn} l_n' S_n^{-1}(\widehat{\lambda}_n) l_n - \mathrm{tr}\left(S_n^{-1}(\widehat{\lambda}_n)\widehat{\beta}_{kn}\right)\right]$. The results in (5.3) and (5.4) implies that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\left[\left(\widehat{\beta}_{kn} l_n' S_n^{-1}(\widehat{\lambda}_n) l_n - \mathrm{tr}\left(S_n^{-1}(\widehat{\lambda}_n)\widehat{\beta}_{kn}\right)\right) - \left(\beta_{k0} l_n' S_n^{-1} l_n - \mathrm{tr}\left(S_n^{-1}\right)\beta_{k0}\right)\right] \qquad (5.5)\\
&= \left(A_{2n} - A_{1n}\right) \times \sqrt{n}\left(\widehat{\lambda}_n - \lambda_0, \widehat{\beta}_{kn} - \beta_{k0}\right)' + o_p(1) \xrightarrow{d} N\left(0, \lim_{n\to\infty}\left(A_{2n} - A_{1n}\right) B_n \left(A_{2n} - A_{1n}\right)'\right).
\end{aligned}$$

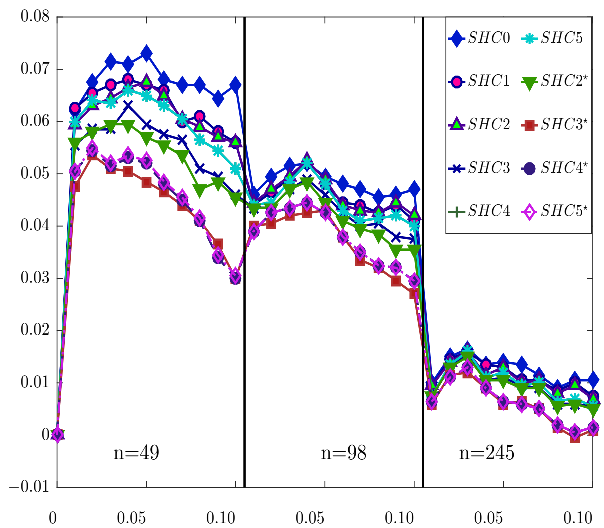
Hence, an estimate of $\mathrm{Var}\big(\frac{1}{n}\big[\widehat{\beta}_{kn}l_n' S_n^{-1}(\widehat{\lambda}_n)l_n - \mathrm{tr}\big(S_n^{-1}(\widehat{\lambda}_n)\widehat{\beta}_{kn}\big)\big]\big)$ is given by $\frac{1}{n}\big(\widehat{A}_{2n}-\widehat{A}_{1n}\big)\widehat{B}_n\big(\widehat{A}_{2n}-\widehat{A}_{1n}\big)'$.

Table 5: Bias Properties of Total Effects

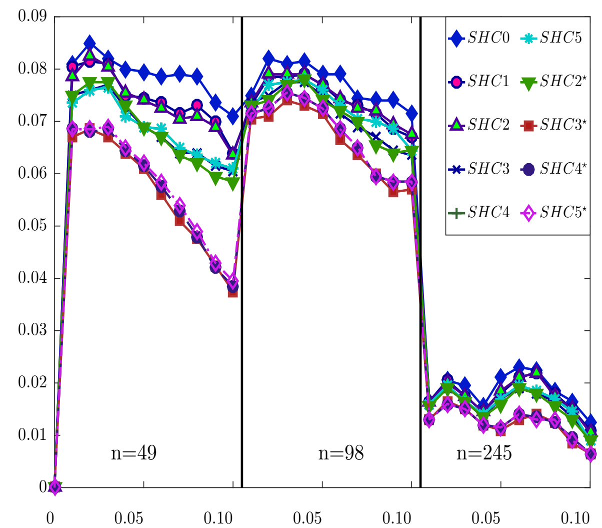
n	$\rho_0-\lambda_0$	$SHC0$	$SHC1$	$SHC2$	$SHC3$	$SHC4$	$SHC5$	$SHC2\star$	$SHC3\star$	$SHC4\star$	$SHC5\star$
Bias on Total Effects: X_1											
49	0.2–0.2	0.0889	0.0875	0.0888	0.0882	0.0889	0.0889	0.0885	0.0810	0.0818	0.0818
	0.2–0.6	-0.0147	-0.0267	-0.0301	-0.0309	-0.0275	-0.0275	-0.0301	-0.0238	-0.0140	-0.0140
	0.6–0.2	-0.0339	-0.0370	-0.0317	-0.0290	-0.0176	-0.0176	-0.0285	-0.0373	-0.0642	-0.0673
	0.6–0.6	0.0380	0.0209	0.0446	0.0591	0.0453	0.0453	0.0097	-0.0346	-0.0425	-0.0334
98	0.2–0.2	-0.0019	-0.0024	-0.0027	-0.0027	-0.0027	-0.0027	-0.0026	-0.0037	-0.0040	-0.0040
	0.2–0.6	-0.0690	-0.0677	-0.0690	-0.0694	-0.0643	-0.0643	-0.0663	-0.0656	-0.0698	-0.0680
	0.6–0.2	-0.0051	-0.0047	-0.0058	-0.0047	-0.0059	-0.0059	-0.0075	-0.0101	-0.0031	-0.0031
	0.6–0.6	-0.0792	-0.0830	-0.0775	-0.0869	-0.0993	-0.0993	-0.0923	-0.0759	-0.0802	-0.0750
245	0.2–0.2	-0.0039	-0.0039	-0.0039	-0.0040	-0.0040	-0.0040	-0.0039	-0.0040	-0.0038	-0.0038
	0.2–0.6	-0.0275	-0.0273	-0.0271	-0.0280	-0.0277	-0.0277	-0.0270	-0.0275	-0.0265	-0.0265
	0.6–0.2	0.0181	0.0176	0.0176	0.0168	0.0168	0.0168	0.0178	0.0168	0.0163	0.0166
	0.6–0.6	0.0177	0.0181	0.0181	0.0190	0.0176	0.0176	0.0183	0.0179	0.0177	0.0180
Bias on Total Effects: X_2											
49	0.2–0.2	-0.0788	-0.0764	-0.0779	-0.0796	-0.0777	-0.0777	-0.0772	-0.0756	-0.0743	-0.0743
	0.2–0.6	0.1074	0.1130	0.1158	0.1156	0.1080	0.1080	0.1097	0.1016	0.0923	0.0923
	0.6–0.2	0.0127	0.0187	0.0115	0.0112	0.0121	0.0121	0.0029	0.0175	0.0514	0.0538
	0.6–0.6	-0.0609	-0.0368	-0.0648	-0.0781	-0.0484	-0.0484	-0.0375	-0.0052	0.0787	0.0646
98	0.2–0.2	-0.0252	-0.0238	-0.0241	-0.0241	-0.0270	-0.0270	-0.0253	-0.0240	-0.0267	-0.0267
	0.2–0.6	0.0754	0.0799	0.0802	0.0757	0.0744	0.0744	0.0785	0.0796	0.0804	0.0787
	0.6–0.2	0.0038	0.0068	0.0045	0.0018	0.0042	0.0042	0.0076	0.0067	-0.0000	-0.0000
	0.6–0.6	0.0184	0.0332	0.0098	0.0356	0.0623	0.0623	0.0466	0.0219	0.0143	0.0066
245	0.2–0.2	0.0009	0.0010	0.0010	0.0011	0.0010	0.0010	0.0010	0.0011	0.0011	0.0011
	0.2–0.6	0.0221	0.0222	0.0222	0.0224	0.0223	0.0223	0.0222	0.0222	0.0220	0.0220
	0.6–0.2	-0.0211	-0.0216	-0.0216	-0.0204	-0.0210	-0.0210	-0.0217	-0.0207	-0.0212	-0.0215
	0.6–0.6	-0.0183	-0.0167	-0.0167	-0.0169	-0.0155	-0.0155	-0.0166	-0.0156	-0.0162	-0.0180
Bias on Total Effects: X_3											
49	0.2–0.2	0.0077	0.0083	0.0078	0.0102	0.0099	0.0099	0.0089	0.0098	0.0111	0.0107
	0.2–0.6	0.0452	0.0465	0.0437	0.0505	0.0456	0.0456	0.0453	0.0452	0.0466	0.0461
	0.6–0.2	0.0296	0.0285	0.0292	0.0299	0.0260	0.0260	0.0285	0.0332	0.0395	0.0401
	0.6–0.6	0.0849	0.0842	0.0862	0.0854	0.0783	0.0783	0.0916	0.0988	0.1198	0.1105
98	0.2–0.2	0.0090	0.0091	0.0091	0.0093	0.0091	0.0091	0.0089	0.0092	0.0093	0.0093
	0.2–0.6	0.0554	0.0553	0.0554	0.0555	0.0556	0.0556	0.0549	0.0559	0.0569	0.0569
	0.6–0.2	0.0118	0.0119	0.0122	0.0137	0.0136	0.0136	0.0135	0.0134	0.0128	0.0128
	0.6–0.6	0.0783	0.0793	0.0792	0.0790	0.0802	0.0802	0.0824	0.0778	0.0803	0.0800
245	0.2–0.2	0.0029	0.0029	0.0029	0.0029	0.0029	0.0029	0.0029	0.0029	0.0029	0.0029
	0.2–0.6	0.0185	0.0185	0.0185	0.0185	0.0186	0.0186	0.0185	0.0185	0.0185	0.0185
	0.6–0.2	-0.0004	-0.0005	-0.0004	-0.0003	-0.0004	-0.0004	-0.0004	-0.0004	-0.0001	-0.0002
	0.6–0.6	0.0233	0.0234	0.0234	0.0235	0.0235	0.0235	0.0234	0.0236	0.0236	0.0233

We use the same Monte Carlo set up described in Section 5.1 to evaluate the finite sample properties of these effects estimators. We report the simulation results only for the total effects estimator. The results for the finite sample bias properties of the estimator are reported in Table 5. The total effects estimator reports similar bias across different methods in all cases, and the bias is relatively larger when $n = 49$. The bias becomes negligible when $n = 245$ across all methods. The results in Table 5 indicate that the total effects estimator of marginal effect of X_3 has relatively smaller bias. Overall, it seems that the estimators impose relatively large bias on the impact measures when there is strong spatial dependence both in the dependent variable and the disturbance term.

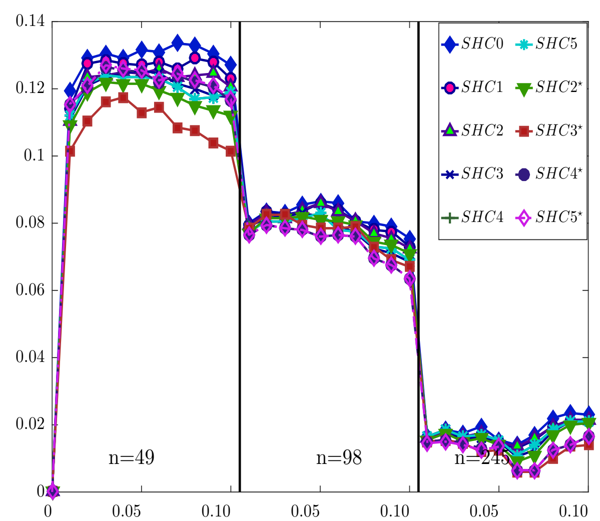
The size properties of standard Wald test for the total effects are illustrated by the P value discrepancy plots presented in Figures 6 through 8. The size distortions presented in Figures 6(a)–6(d) for the total effects of X_1 indicate that the Wald tests based on $SHC0$ produce relatively large discrepancies when $n = 49$ and $n = 98$. The same pattern is also valid in Figures 7 and 8 for the Wald tests of the marginal effects of X_2 and X_3. The size distortions are relatively smaller in the case of $SHC2^{\star} - SHC5^{\star}$, especially when $n = 49$ and $n = 98$. The correction methods $SHC2 - SHC5$, generally, perform better than $SHC0$, but worse than $SHC2^{\star} - SHC5^{\star}$. Figures 6 through 8 also indicate that the difference in size distortions across methods get smaller when there is strong spatial dependence either in the disturbance term or in the dependent variable.



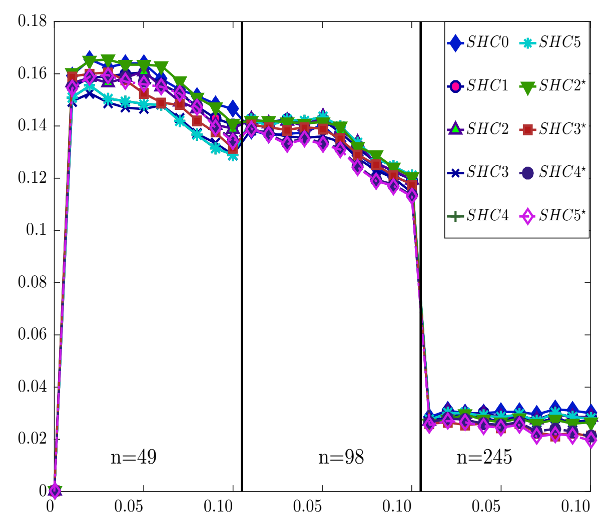
(a) $(\rho_0, \lambda_0) = (0.2, 0.2)$



(b) $(\rho_0, \lambda_0) = (0.2, 0.6)$

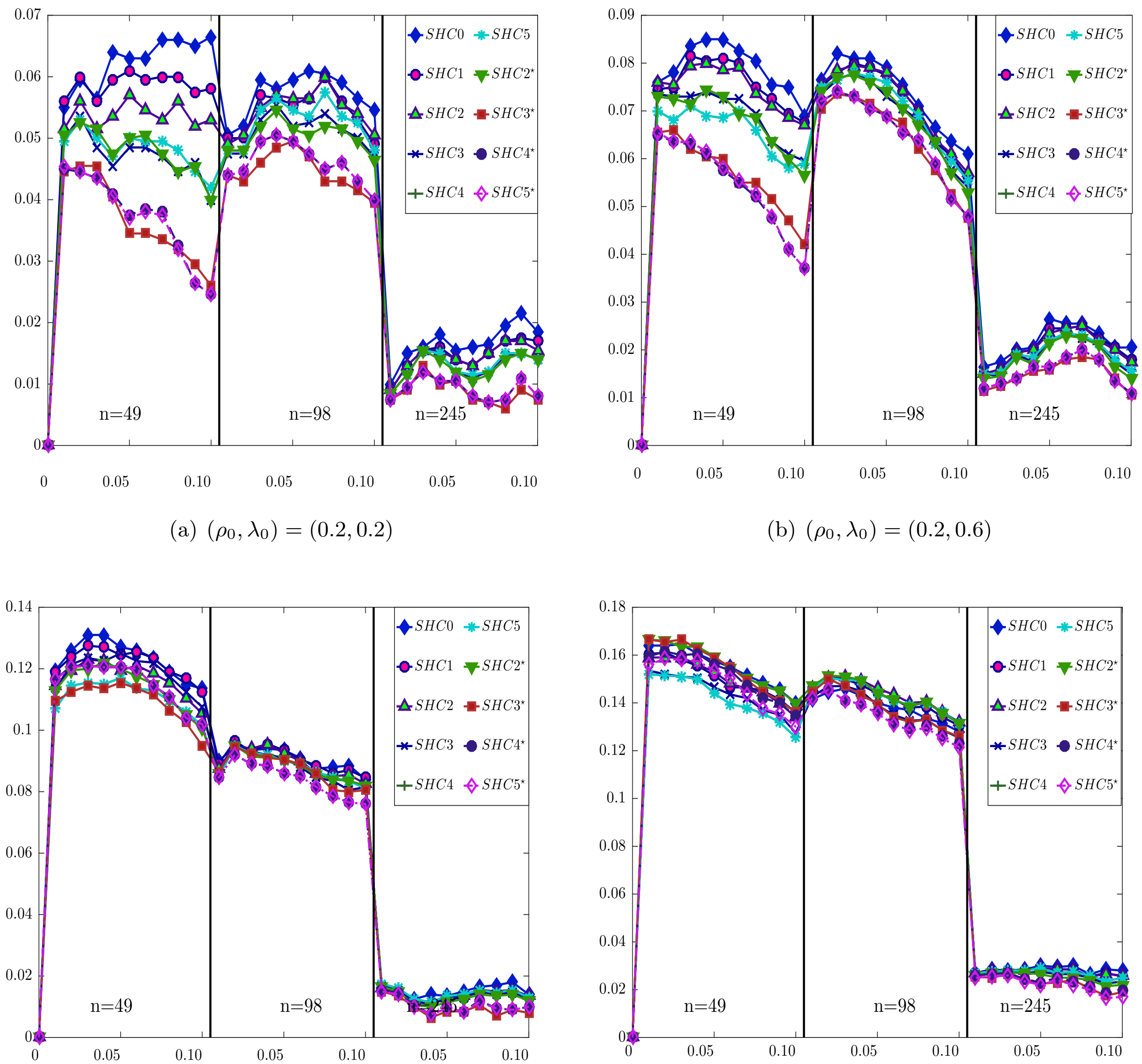


(c) $(\rho_0, \lambda_0) = (0.6, 0.2)$



(d) $(\rho_0, \lambda_0) = (0.6, 0.6)$

Figure 6: P value discrepancy plots for total effects: X_1



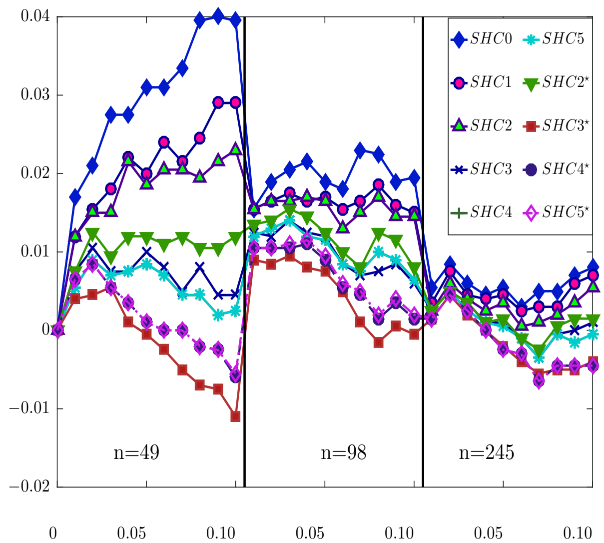
(a) $(\rho_0, \lambda_0) = (0.2, 0.2)$

(b) $(\rho_0, \lambda_0) = (0.2, 0.6)$

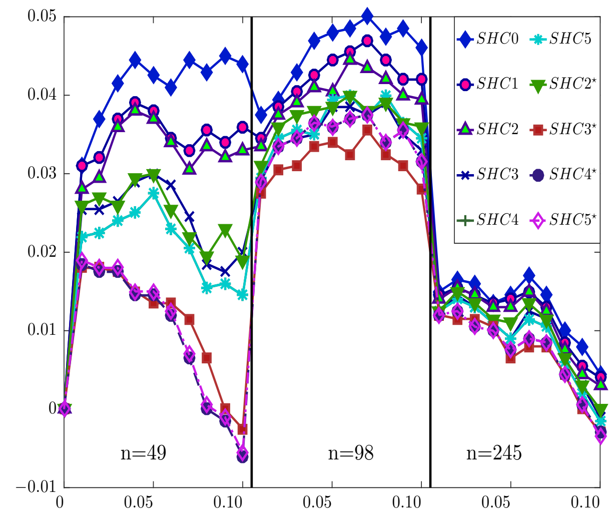
(c) $(\rho_0, \lambda_0) = (0.6, 0.2)$

(d) $(\rho_0, \lambda_0) = (0.6, 0.6)$

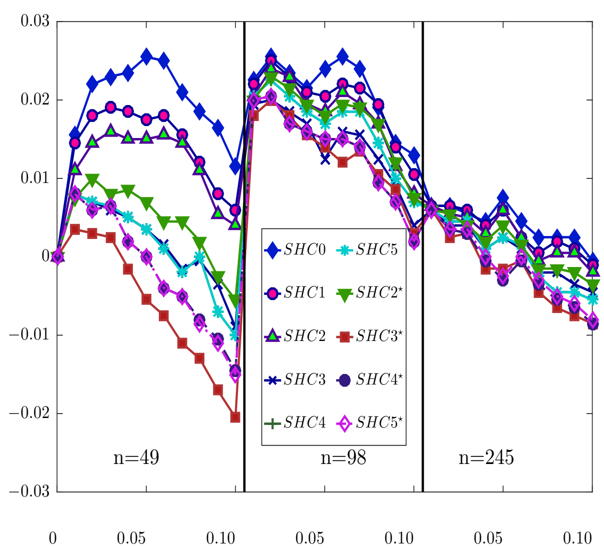
Figure 7: P value discrepancy plots for total effects: X_2



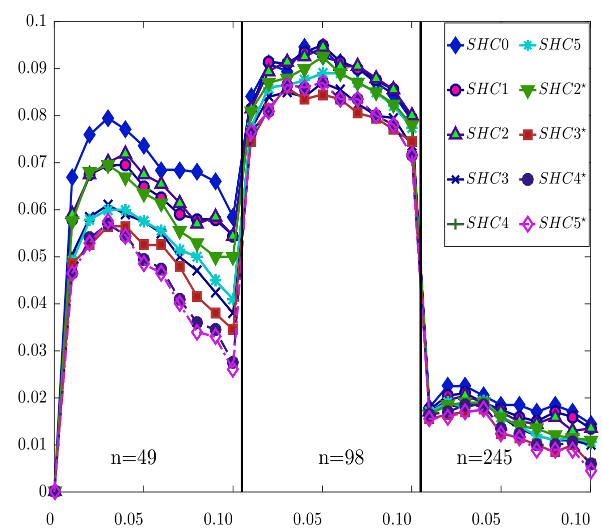
(a) $(\rho_0, \lambda_0) = (0.2, 0.2)$



(b) $(\rho_0, \lambda_0) = (0.2, 0.6)$



(c) $(\rho_0, \lambda_0) = (0.6, 0.2)$



(d) $(\rho_0, \lambda_0) = (0.6, 0.6)$

Figure 8: P value discrepancy plots for total effects: X_3

6 Conclusion

In this study, we investigate the finite sample properties of a robust GMME suggested for a SARAR(1,1) specification that has heteroskedastic disturbance terms. We consider various refinement methods suggested in the non-spatial literature and extend these method for our spatial autoregressive model. We provide a general argument by assuming an arbitrary set of moment functions. To formulate leverage-adjusted residuals within the context of our spatial model, we suggest two (quasi) hat matrices. The first hat matrix is formulated using the first order asymptotic results established for the GMME. The spatial dependence in our context provide a different stochastic dimension which complicates the formulation. We show how this hat matrix can be determined for the spatial autoregressive models. Based on this hat matrix, we formulate the finite sample correction methods $SHC2^{\star} - SHC5^{\star}$. The second hat matrix is ad-hoc in the sense that its formulation is feasible when the autoregressive parameters are known. Based on this particular hat matrix, we formulate the finite sample correction methods $SHC2 - SHC5$.

In a Monte Carlo study, we investigate the effect of these correction methods on the finite sample properties of the GMME of a SARAR(1,1) specification. In terms of bias properties, our results indicate that the correction methods produce similar point estimates for all parameters. Our results also indicate that the usual estimated standard errors (based on $SHC0$) differ substantially from the empirical standard deviations, which suggests that the asymptotic distribution does not approximate the finite sample distribution well enough. Further, our results show that the Wald tests based on the usual estimated standard errors can have substantial size distortions in small samples. We show that the GMME based on the correction methods $SHC2^{\star} - SHC5^{\star}$ can perform better in terms of finite sample properties. In particular, our results show that the Wald tests based on the correction methods $SHC2^{\star} - SHC5^{\star}$ have relatively smaller size distortions in finite samples. All of these results can be useful for applied researchers who estimate and test spatial models with the GMM estimators.

References

Abadir, Karim M. and Jan R. Magnus (2005). *Matrix Algebra*. New York: Cambridge University Press.

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

Bera, Anil K., Totok Suprayitno, and Gamini Premaratne (2002). "On some heteroskedasticity-robust estimators of variance-covariance matrix of the least-squares estimators". In: *Journal of Statistical Planning and Inference* 108.1âĂŞ2.

Chesher, Andrew (1989). "Hajek Inequalities, Measures of Leverage and the Size of Heteroskedasticity Robust Wald Tests". In: *Econometrica* 57.4, pp. 971–977.

Chesher, Andrew and Gerard Austin (1991). "The finite-sample distributions of heteroskedasticity robust Wald statistics". In: *Journal of Econometrics* 47.1, pp. 153 –173.

Chesher, Andrew and Ian Jewitt (1987). "The Bias of a Heteroskedasticity Consistent Covariance Matrix Estimator". In: *Econometrica* 55.5.

Cribari-Neto, Francisco (2004). "Asymptotic inference under heteroskedasticity of unknown form". In: *Computational Statistics & Data Analysis* 45.2, pp. 215 –233.

Cribari-Neto, Francisco, Tatiene C. Souza, and Klaus L. P. Vasconcellos (2007). "Inference Under Heteroskedasticity and Leveraged Data". In: *Communications in Statistics-Theory and Methods* 36.10.

Debarsy, Nicolas, Fei Jin, and Lung fei Lee (2015). "Large sample properties of the matrix exponential spatial specification with an application to FDI". In: *Journal of Econometrics* 188.1.

Dogan, Osman and Suleyman Taspinar (2013). *GMM Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances*. Working Papers 1. City University of New York Graduate Center, Ph.D. Program in Economics. URL: `http://ideas.repec.org/p/cgc/wpaper/001.html`.

Eicker, Friedhelm (1967). "Limit theorems for regressions with unequal and dependent errors". In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. Berkeley, Calif.: University of California Press, pp. 59–82.

Hinkley, David V. (1977). "Jackknifing in Unbalanced Situations". In: *Technometrics* 19.3.

Horn, Susan D., Roger A. Horn, and David B. Duncan (1975). "Estimating Heteroscedastic Variances in Linear Models". In: *Journal of the American Statistical Association* 70.350, pp. 380–385.

Judge, George G. et al. (1988). *Introduction to the Theory and Practice of Econometrics*. 2nd Edition. Wiley series in probability and mathematical statistics. Applied probability and statistics. Wiley.

Kauermann, Goran and Raymond J. Carroll (2001). "A Note on the Efficiency of Sandwich Covariance Matrix Estimation". In: *Journal of the American Statistical Association* 96.456.

Kelejian, Harry H. and Ingmar R. Prucha (1998). "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances". In: *Journal of Real Estate Finance and Economics* 17.1, pp. 1899–1926.

— (2010). "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances". In: *Journal of Econometrics* 157, pp. 53–67.

Lee, Lung-fei (2003). "Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances". In: *Econometric Reviews* 22.4, pp. 307–335.

— (2007). "GMM and 2SLS estimation of mixed regressive, spatial autoregressive models". In: *Journal of Econometrics* 137.2, pp. 489–514.

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

Lin, Eric S. and Ta-Sheng Chou (2015). "Finite-Sample Refinement of GMM Approach to Nonlinear Models Under Heteroskedasticity of Unknown Form". In: *Econometric Reviews* 0.0, pp. 1–37.

Lin, Xu and Lung-fei Lee (2010). "GMM estimation of spatial autoregressive models with unknown heteroskedasticity". In: *Journal of Econometrics* 157.1, pp. 34–52.

Long, J. Scott and Laurie H. Ervin (2000). "Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model". In: *The American Statistician* 54.3.

MacKinnon, James G. (2013). "Thirty Years of Heteroskedasticity Robust Inference". In: *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*. Ed. by Xiaohong Chen and Norman R. Swanson. Springer New York, pp. 437–461.

MacKinnon, James G and Halbert White (1985). "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties". In: *Journal of Econometrics* 29.3, pp. 305 –325.

Pace, Robert K., James P. LeSage, and Shuang Zhu (2012). "Spatial Dependence in Regressors and its Effect on Performance of Likelihood-Based and Instrumental Variable Estimators". In: ed. by Daniel Millimet Dek Terrell. 30th Anniversary Edition (Advances in Econometrics, Volume 30). Emerald Group Publishing Limited, pp. 257–295.

Taspinar, Suleyman, Osman Dogan, and Wim P.M. Vijverberg (2016). "GMM inference in spatial autoregressive models". In: *Econometric Reviews* Forthcoming.

White, Halbert G. (1980). "A Heteroskedasticity-Consistent Covariance Matrix Estimator a Direct Test for Heteroskedasticity". In: *Econometrica* 48, pp. 817–838.

Appendix

A Some Useful Lemmas

Lemma 1. — Assume that ε_{in}s are i.i.d with mean zero and variance σ_0^2. Let $\mathrm{E}\big(\varepsilon_{in}^3\big) = \mu_3$, $\mathrm{E}\big(\varepsilon_{in}^4\big) = \mu_4$. Let A_n and B_n be $n \times n$ matrices of constants with zero diagonal elements, i.e., $\mathrm{vec}_D(A_n) = \mathrm{vec}_D(B_n) = 0_{n\times 1}$. Then,

$$
\begin{aligned}
&\textit{(1)}\, \mathrm{E}\big(\varepsilon_n^{'} A_n \varepsilon_n\big)^2 = \sigma_0^4 \mathrm{tr}\big(A_n A_n^{(s)}\big),\ \textit{(2)}\, \mathrm{E}\big(\varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_n^{'} B_n \varepsilon_n\big) = \sigma_0^4 \mathrm{tr}\big(A_n B_n^{(s)}\big),\\
&\textit{(3)}\, \mathrm{E}\big(A_n \varepsilon_n \cdot \varepsilon_n^{'} B_n \varepsilon_n\big) = A_n \,\mathrm{vec}_D\,\big(B_n\big)\mu_3 = 0,\ \textit{(4)}\, \mathrm{E}\big(\varepsilon_n^{'} B_n \varepsilon_n \cdot \varepsilon_n^{'} A_n\big) = \mu_3 \,\mathrm{vec}_D^{'}\,\big(B_n\big) A_n = 0,\\
&\textit{(5)}\, \mathrm{tr}\big(A_n B_n\big) = \mathrm{vec}^{'}\,\big(A_n^{'}\big) \cdot \mathrm{vec}\,\big(B_n\big).
\end{aligned}
$$

Lemma 2. — Assume that A_n and B_n are two $n \times n$ non-stochastic matrices with zero diagonal elements. Assume that ε_{in}s are i.i.d with mean zero and variance σ_0^2. Let e_s and e_t be elementary vectors in $\mathbb{R}^n$ for $s = 1, \ldots, n$, $t = 1, \ldots, n$, and $s \neq t$. For notational simplicity, let $A_{is,n}^{(s)} = A_{is,n} + A_{si,n}$, $A_{s\bullet,n}^{(s)} = \big(A_{s\bullet,n} + A_{\bullet s,n}^{'}\big)$, and $A_{\bullet s,n}^{(s)} = \big(A_{s\bullet,n}^{'} + A_{\bullet s,n}\big) = A_{s\bullet,n}^{(s)'}$. Then,
(1) $\mathrm{E}\big(\varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_{sn}^2\big) = 0$, and $\mathrm{E}\big(\varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_{sn}\varepsilon_{tn}\big) = \sigma_0^4\big(A_{ts,n} + A_{st,n}\big)$.
(2) Let Q_n be an $n \times r$ non-stochastic matrix. Then,

$$
\begin{aligned}
&(2.1) \quad \mathrm{E}\big(Q_n^{'} \varepsilon_n \cdot \varepsilon_{sn}^2\big) = \mu_3 Q_{s\bullet,n}^{'},\\
&(2.2) \quad \mathrm{E}\big(Q_n^{'} \varepsilon_n \cdot \varepsilon_{sn}\varepsilon_{tn}\big) = 0_{r\times 1}.
\end{aligned}
$$

(3) The expectation of the (s, s)th element of $\big(\varepsilon_n \cdot \varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_n^{'} B_n \varepsilon_n \cdot \varepsilon_n^{'}\big)$ is given by

$$
\mathrm{E}\big(e_s^{'}\varepsilon_n \cdot \varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_n^{'} B_n \varepsilon_n \cdot \varepsilon_n^{'} e_s\big) = \sigma_0^6 \,\mathrm{vec}^{'}\,\big(A_n^{(s)}\big)\,\mathrm{vec}\,\big(B_n\big) - \big(\sigma_0^6 - \mu_4\sigma_0^2\big) A_{\bullet s,n}^{(s)'} B_{\bullet s,n}^{(s)}.
$$

(4) The expectation of the (s, t)th element of $\big(\varepsilon_n \cdot \varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_n^{'} B_n \varepsilon_n \cdot \varepsilon_n^{'}\big)$ is given by

$$
\mathrm{E}\big(e_s^{'}\varepsilon_n \cdot \varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_n^{'} B_n \varepsilon_n \cdot \varepsilon_n^{'} e_t\big) = \mu_3^2 A_{st,n}^{(s)} B_{st,n}^{(s)}.
$$

(5) Let Q_n be an $n \times r$ non-stochastic matrix. Then,

$$
\begin{aligned}
&(5.1) \quad \mathrm{E}\big(\varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_n^{'} Q_n \cdot \varepsilon_{sn}^2\big) = \sigma_0^2 \mu_3 A_{\bullet s,n}^{(s)'} Q_n,\\
&(5.2) \quad \mathrm{E}\big(\varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_n^{'} Q_n \cdot \varepsilon_{sn}\varepsilon_{tn}\big) = \sigma_0^2 \mu_3 A_{st,n}^{(s)}\big(Q_{s\bullet,n} + Q_{t\bullet,n}\big),\\
&(5.3) \quad \mathrm{E}\big(Q_n^{'} \varepsilon_n \cdot \varepsilon_n^{'} Q_n \cdot \varepsilon_{sn}^2\big) = \sigma_0^4 Q_n^{'} Q_n + \big(\mu_4 - \sigma_0^4\big) Q_{s\bullet,n}^{'} Q_{s\bullet,n},\\
&(5.4) \quad \mathrm{E}\big(Q_n^{'} \varepsilon_n \cdot \varepsilon_n^{'} Q_n \cdot \varepsilon_{sn}\varepsilon_{tn}\big) = \sigma_0^4\big(Q_{s\bullet,n}^{'} Q_{t\bullet,n} + Q_{t\bullet,n}^{'} Q_{s\bullet,n}\big).
\end{aligned}
$$

(6) $\mathrm{E}\big(\varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_n^{'} B_n \varepsilon_n \cdot \varepsilon_{tn}\big) = \sigma_0^2 \mu_3 A_{\bullet t,n}^{(s)'} B_{\bullet t,n}^{(s)}$.
(7) Let Q_n be an $n \times r$ non-stochastic matrix. Then,

$$
\begin{aligned}
&(7.1) \quad \mathrm{E}\big(\varepsilon_n^{'} B_n \varepsilon_n \cdot \varepsilon_n^{'} Q_n \cdot \varepsilon_{tn}\big) = \sigma_0^4 A_{\bullet t,n}^{(s)'} Q_n,\\
&(7.2) \quad \mathrm{E}\big(Q_n^{'} \varepsilon_n \cdot \varepsilon_n^{'} Q_n \cdot \varepsilon_{tn}\big) = \mu_3 Q_{t\bullet,n}^{'} Q_{t\bullet,n}.
\end{aligned}
$$

Lemma 3. — Let A_n, B_n and C_n be $n \times n$ matrices with ijth elements respectively denoted by $A_{ij,n}$, $B_{ij,n}$ and $C_{ij,n}$. Assume that A_n and B_n have zero diagonal elements, and C_n has uniformly bounded row and column sums in absolute value. Let q_n be $n \times 1$ vector with uniformly bounded elements in absolute value. Assume that ε_n satisfies Assumption 1 with covariance matrix denoted by Σ_n=D$\left(\sigma_{1n}^2, \ldots, \sigma_{nn}^2\right)$. Then,

$$
\begin{aligned}
&(1) \quad \mathrm{E}\big(\varepsilon_n' A_n \varepsilon_n \cdot \varepsilon_n' B_n \varepsilon_n\big) = \sum_{i=1}^n \sum_{j=1}^n A_{ij,n}\big(B_{ij,n} + B_{ji,n}\big)\sigma_{in}^2 \sigma_{jn}^2 = \mathrm{tr}\big(\Sigma_n A_n \big(B_n' \Sigma_n + \Sigma_n B_n\big)\big),\\
&(2) \quad \mathrm{E}\big(\varepsilon_n C_n \varepsilon_n\big)^2 = \sum_{i=1}^n C_{ii,n}^2\big[\mathrm{E}\big(\varepsilon_{in}^4\big) - 3\sigma_{in}^4\big] + \mathrm{tr}^2\big(\Sigma_n C_n\big) + \mathrm{tr}\big(\Sigma_n C_n C_n' \Sigma_n + \Sigma_n C_n \Sigma_n C_n\big),\\
&(3) \quad \mathrm{Var}\big(\varepsilon_n C_n \varepsilon_n\big) = \sum_{i=1}^n C_{ii,n}^2\big[\mathrm{E}\big(\varepsilon_{in}^4\big) - 3\sigma_{in}^4\big] + \sum_{i=1}^n \sum_{j=1}^n C_{ij,n}\big(C_{ij,n} + C_{ji,n}\big)\sigma_{in}^2 \sigma_{jn}^2\\
&\qquad\qquad\qquad\quad\;\; = \sum_{i=1}^n C_{ii,n}^2\big[\mathrm{E}\big(\varepsilon_{in}^4\big) - 3\sigma_{in}^4\big] + \mathrm{tr}\big(\Sigma_n C_n C_n' \Sigma_n + \Sigma_n C_n \Sigma_n C_n\big),\\
&(4) \quad \mathrm{E}\big(\varepsilon_n' C_n \varepsilon_n\big) = O(n),\ \mathrm{Var}\big(\varepsilon_n' C_n \varepsilon_n\big) = O(n),\ \varepsilon_n' C_n \varepsilon_n = O_p(n),\\
&(5) \quad \mathrm{E}\big(C_n \varepsilon_n\big) = 0,\ \mathrm{Var}\big(C_n \varepsilon_n\big) = O(n),\ C_n \varepsilon_n = O_p(n),\ \mathrm{Var}\big(q_n' C_n \varepsilon_n\big) = O(n),\ q_n' C_n \varepsilon_n = O_p(n).
\end{aligned}
$$

Lemma 4. — Let A_n, B_n and C_n be $n \times n$ three matrices. Assume that A_n has zero diagonal elements, i.e., $\mathrm{D}(A_n) = 0_{n\times n}$, and C_n is a diagonal matrix, i.e., $\mathrm{D}\left(C_n\right) \neq 0_{n\times n}$. Then,

$$
\begin{aligned}
&(1) \quad \mathrm{tr}\big(A_n^{(s)} B_n\big) = \frac{1}{2}\mathrm{tr}\big(A_n^{(s)} B_n^{(s)}\big) = \frac{1}{2}\,\mathrm{vec}'\left(A_n^{(s)}\right)\mathrm{vec}\left(B_n^{(s)}\right).\\
&(2) \quad \mathrm{tr}\big(A_n^{(s)} B_n\big) = \frac{1}{2}\mathrm{tr}\big(A_n^{(s)}\big[B_n - \mathrm{D}\left(B_n\right)\big]^{(s)}\big) = \mathrm{vec}'\left(\big[B_n - \mathrm{D}\left(B_n\right)\big]^{(s)}\right)\mathrm{vec}\left(A_n^{(s)}\right).\\
&(3) \quad \mathrm{vec}'\left(\big[B_n - \mathrm{D}(B_n)\big]^{(s)}\right)\mathrm{vec}\left(C_n A_n^{(s)}\right) = \mathrm{vec}'\left(\big[B_n - \mathrm{D}\left(B_n\right)\big]^{(s)}\right)\mathrm{vec}\left(\big(C_n A_n\big)^{(s)}\right).
\end{aligned}
$$

B Proofs of Lemmas

Proof of Lemma 1. For (1), (2), (3) and (4), see Lee, (2007). For (5), see Abadir and Magnus, (2005, p. 283) . Using (5), (1) and (2) can also be written as

$$
\begin{aligned}
&\mathrm{E}\big(\varepsilon_n' A_n \varepsilon_n\big)^2 = \sigma_0^4\,\mathrm{vec}'\left(A_n'\right)\mathrm{vec}\left(A_n^{(s)}\right) = \sigma_0^4\,\mathrm{vec}'\left(A_n^{(s)}\right)\mathrm{vec}\left(A_n\right),\\
&\mathrm{E}\big(\varepsilon_n' A_n \varepsilon_n \cdot \varepsilon_n' B_n \varepsilon_n\big) = \sigma_0^4\,\mathrm{vec}'\left(A_n'\right)\mathrm{vec}(B_n^{(s)}) = \sigma_0^4\,\mathrm{vec}'\left(B_n^{(s)}\right)\mathrm{vec}\left(A_n\right).
\end{aligned}
$$

□

Proof of Lemma 2. (1). $\mathrm{E}\big(\varepsilon_n' A_n \varepsilon_n \cdot \varepsilon_{sn}^2\big) = \sum_{i=1}^n \sum_{j=1}^n A_{ij,n}\mathrm{E}\big(\varepsilon_{in}\varepsilon_{jn}\varepsilon_{sn}^2\big) = \mu_4 A_{ss,n} = 0$, since $A_{ss,n} = 0\ \forall s$. $\mathrm{E}\big(\varepsilon_n' A_n \varepsilon_n \cdot \varepsilon_{sn}\varepsilon_{tn}\big) = \sum_{i=1}^n \sum_{j=1}^n A_{ij,n}\mathrm{E}\big(\varepsilon_{in}\varepsilon_{jn}\varepsilon_{sn}\varepsilon_{tn}\big) = \sigma_0^4\big(A_{ts,n} + A_{st,n}\big)$, since $A_{ij,n}\mathrm{E}\big(\varepsilon_{in}\varepsilon_{jn}\varepsilon_{sn}\varepsilon_{tn}\big)$ is not zero only if (1) $(i = t) \neq (j = s)$, and (2) $(i = s) \neq (j = t)$.

(2.1) $\mathrm{E}\big(Q_n' \varepsilon_n \cdot \varepsilon_{sn}^2\big) = \sum_{i=1}^n Q_{i\bullet,n}' \mathrm{E}\big(\varepsilon_{in}\varepsilon_{sn}^2\big) = \mu_3 Q_{s\bullet,n}'$, since $\mathrm{E}\big(\varepsilon_{in}\varepsilon_{sn}^2\big)$ is not zero only if $(i = s)$.

(2.2) $\mathrm{E}\big(Q_n' \varepsilon_n \cdot \varepsilon_{sn}\varepsilon_{tn}\big) = \sum_{i=1}^n Q_{i\bullet,n}' \mathrm{E}\big(\varepsilon_{in}\varepsilon_{sn}\varepsilon_{tn}\big) = 0_{r\times 1}$ since ε_{in}s are independent.

(3). $\mathrm{E}\big(e_s' \varepsilon_n \cdot \varepsilon_n' A_n \varepsilon_n \cdot \varepsilon_n' B_n \varepsilon_n \cdot \varepsilon_n' e_s\big) = \mathrm{E}\big(\mathrm{tr}\big(\varepsilon_n' A_n \varepsilon_n \cdot \varepsilon_n' B_n \varepsilon_n \cdot \varepsilon_n' e_s e_s' \varepsilon_n\big)\big) = \mathrm{E}\big(\mathrm{tr}\big(\varepsilon_n' A_n \varepsilon_n \cdot \varepsilon_n' B_n \varepsilon_n \cdot$

$\varepsilon^2_{sn})\big) = \mathrm{E}\big(\varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_n^{'} B_n \varepsilon_n \cdot \varepsilon^2_{sn}\big)$. Hence,

$$\mathrm{E}\big(\varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_n^{'} B_n \varepsilon_n \cdot \varepsilon^2_{sn}\big) = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n A_{ij,n} B_{kl,n} \mathrm{E}\big(\varepsilon_{in}\varepsilon_{jn}\varepsilon_{kn}\varepsilon_{ln}\varepsilon^2_{sn}\big).$$

For a given s value, we need to consider (1) $(i = k \neq s) \neq (j = l \neq s)$, (2) $(i = l \neq s) \neq (j = k \neq s)$, (3) $(i = k = s) \neq (j = l)$, (4) $(i = k) \neq (j = l = s)$, (5) $(i = l = s) \neq (j = k)$, and (6) $(i = l) \neq (j = k = s)$. Hence,

$$\begin{aligned}
&\mathrm{E}\big(\varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_n^{'} B_n \varepsilon_n \cdot \varepsilon^2_{sn}\big) = \sigma_0^6 \sum_{i\neq s}\sum_{j \neq s} A_{ij,n} B_{ij,n} + \sigma_0^6 \sum_{i\neq s}\sum_{j \neq s} A_{ij,n} B_{ji,n} + \mu_4 \sigma_0^2 \sum_{i=1}^n A_{si,n} B_{si,n} \\
&+ \mu_4 \sigma_0^2 \sum_{i=1}^n A_{is,n} B_{is,n} + \mu_4 \sigma_0^2 \sum_{i=1}^n A_{si,n} B_{is,n} + \mu_4 \sigma_0^2 \sum_{i=1}^n A_{is,n} B_{si,n} \\
&= \sigma_0^6 \bigg(\sum_{i=1}^n \sum_{j=1}^n A_{ij,n} B_{ij,n} - \sum_{i=1}^n A_{si,n} B_{si,n} - \sum_{i=1}^n A_{is,n} B_{is,n} \bigg) \\
&\quad + \sigma_0^6 \bigg(\sum_{i=1}^n \sum_{j=1}^n A_{ij,n} B_{ji,n} - \sum_{i=1}^n A_{si,n} B_{is,n} - \sum_{i=1}^n A_{is,n} B_{si,n} \bigg) \\
&\quad + \mu_4 \sigma_0^2 \sum_{i=1}^n \left(A_{si,n} + A_{is,n}\right)\left(B_{si,n} + B_{is,n}\right) \\
&= \mathrm{tr}\big(A_n^{(s)} B_n\big) - \sigma_0^6 \big(A_{s\bullet,n} + A^{'}_{\bullet s,n}\big) B^{'}_{s\bullet,n} - \sigma_0^6 \big(A_{s\bullet,n} + A^{'}_{\bullet s,n}\big) B_{\bullet s,n} + \mu_4 \sigma_0^2 A^{(s)'}_{\bullet s} B^{(s)}_{\bullet s,n}.
\end{aligned}$$

We also have $\mathrm{tr}\big(C_n D_n\big) = \mathrm{vec}^{'}\left(C_n^{'}\right) \mathrm{vec}\left(D_n\right)$ for any conformable matrices C_n and D_n. Hence,

$$\begin{aligned}
&\mathrm{E}\big(\varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_n^{'} B_n \varepsilon_n \cdot \varepsilon^2_{sn}\big) = \mathrm{vec}^{'}\left(A_n^{(s)}\right) \mathrm{vec}\left(B_n\right) - \sigma_0^6 \big(A_{s\bullet,n} + A^{'}_{\bullet s,n}\big)\big(B^{'}_{s\bullet,n} + B_{\bullet s,n}\big) \\
&+ \mu_4 \sigma_0^2 A^{(s)'}_{\bullet s} B^{(s)}_{s\bullet,n} = \mathrm{vec}^{'}\left(A_n^{(s)}\right) \mathrm{vec}\left(B_n\right) - \sigma_0^6 A^{(s)'}_{\bullet s,n} B^{(s)}_{\bullet s,n} + \mu_4 \sigma_0^2 A^{(s)'}_{\bullet s} B^{(s)}_{\bullet s,n} \\
&= \mathrm{vec}^{'}\left(A_n^{(s)}\right) \mathrm{vec}\left(B_n\right) - \big(\sigma_0^6 - \mu_4 \sigma_0^2\big) A^{(s)'}_{\bullet s,n} B^{(s)}_{\bullet s,n}.
\end{aligned}$$

(4) $\mathrm{E}\big(e_s^{'} \varepsilon_n \cdot \varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_n^{'} B_n \varepsilon_n \cdot \varepsilon_n^{'} e_t\big) = \mathrm{E}\big(\mathrm{tr}\big(\varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_n^{'} B_n \varepsilon_n \cdot \varepsilon_n^{'} e_t e_s^{'} \varepsilon_n\big)\big) = \mathrm{E}\big(\varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_n^{'} B_n \varepsilon_n \cdot \varepsilon_{sn} \varepsilon_{tn}\big)$. Hence,

$$\mathrm{E}\big(\varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_n^{'} B_n \varepsilon_n \cdot \varepsilon_{sn}\varepsilon_{tn}\big) = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n A_{ij,n} B_{kl,n} \mathrm{E}\big(\varepsilon_{in}\varepsilon_{jn}\varepsilon_{kn}\varepsilon_{ln}\varepsilon_{sn}\varepsilon_{tn}\big).$$

There are four cases that we need to consider: (1) $(i = k = s) \neq (j = l = t)$, (2) $(i = k = t) \neq (j = l = s)$, (3) $(i = l = s) \neq (j = k = t)$, and (4) $(i = l = t) = (j = k = s)$. Hence,

$$\begin{aligned}
\mathrm{E}\big(\varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_n^{'} B_n \varepsilon_n \cdot \varepsilon_{sn}\varepsilon_{tn}\big) &= \mu_3^2 A_{st,n} B_{st,n} + \mu_3^2 A_{ts,n} B_{ts,n} + \mu_3^2 A_{st,n} B_{ts,n} + \mu_3^2 A_{ts,n} B_{st,n} \\
&= \mu_3^2 \big(A_{st,n} + A_{ts,n}\big)\big(B_{st,n} + B_{ts,n}\big) = \mu_3^2 A^{(s)}_{st,n} B^{(s)}_{st,n}.
\end{aligned}$$

(5.1) $\mathrm{E}\big(\varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_n^{'} Q_n \cdot \varepsilon^2_{sn}\big) = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n A_{ij,n} Q_{k\bullet,n} \mathrm{E}\big(\varepsilon_{in}\varepsilon_{jn}\varepsilon_{kn}\varepsilon^2_{sn}\big)$. Here, we need to consider (1) $(i=k)\neq(j=s)$ and (2) $(i=s)\neq(j=k)$. Hence

$$\mathrm{E}\big(\varepsilon_n^{'}A_n\varepsilon_n\cdot\varepsilon_n^{'}Q_n\cdot\varepsilon_{sn}^2\big)=\sigma_0^2\mu_3\sum_{i=1}^n\big(A_{is,n}+A_{si,n}\big)Q_{i\bullet,n}=\sigma_0^2\mu_3A_{\bullet sn}^{(s)'}Q_n.$$

(5.2) $\mathrm{E}\big(\varepsilon_n^{'}A_n\varepsilon_n\cdot\varepsilon_n^{'}Q_n\cdot\varepsilon_{sn}\varepsilon_{tn}\big)=\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^nA_{ij,n}Q_{k\bullet,n}\mathrm{E}\big(\varepsilon_{in}\varepsilon_{jn}\varepsilon_{kn}\varepsilon_{sn}\varepsilon_{tn}\big)$. Here, we need to consider (1) $(i=k=s)\neq(j=t)$, (2) $(i=k=t)\neq(j=s)$, (3) $(i=s)\neq(j=k=t)$ and (4) $(i=t)\neq(j=k=s)$. Hence,

$$\mathrm{E}\big(\varepsilon_n^{'}A_n\varepsilon_n\cdot\varepsilon_n^{'}Q_n\cdot\varepsilon_{sn}\varepsilon_{tn}\big)=\sigma_0^2\mu_3A_{st,n}^{(s)}\big(Q_{s\bullet,n}+Q_{t\bullet,n}\big).$$

(5.3) $\mathrm{E}\big(Q_n^{'}\varepsilon_n\cdot\varepsilon_n^{'}Q_n\cdot\varepsilon_{sn}^2\big)=\sum_{i=1}^n\sum_{j=1}^nQ_{i\bullet,n}^{'}Q_{j\bullet,n}\mathrm{E}\big(\varepsilon_{in}\varepsilon_{jn}\varepsilon_{sn}^2\big)$. We need to consider two case where $\mathrm{E}\big(\varepsilon_{in}\varepsilon_{jn}\varepsilon_{sn}^2\big)$ is not zero: (i) $(i=j=s)$ and (ii) $(i=j)\neq s$. Hence,

$$\begin{aligned}\sum_{i=1}^n\sum_{j=1}^nQ_{i\bullet,n}^{'}Q_{j\bullet,n}\mathrm{E}\big(\varepsilon_{in}\varepsilon_{jn}\varepsilon_{sn}^2\big)&=\mu_4Q_{s\bullet,n}^{'}Q_{s\bullet,n}+\sigma_0^4\sum_{i\neq s}Q_{i\bullet,n}^{'}Q_{i\bullet,n}\\&=\mu_4Q_{s\bullet,n}^{'}Q_{s\bullet,n}+\sigma_0^4\sum_{i=1}^nQ_{i\bullet,n}^{'}Q_{i\bullet,n}-\sigma_0^4Q_{s\bullet,n}^{'}Q_{s\bullet,n}\\&=\sigma_0^4Q_n^{'}Q_n+\big(\mu_4-\sigma_0^4\big)\sigma_0^4Q_{s\bullet,n}^{'}Q_{s\bullet,n}.\end{aligned}$$

(5.4) $\mathrm{E}\big(Q_n^{'}\varepsilon_n\cdot\varepsilon_n^{'}Q_n\cdot\varepsilon_{sn}\varepsilon_{tn}\big)=\sum_{i=1}^n\sum_{j=1}^nQ_{i\bullet,n}^{'}Q_{j\bullet,n}\mathrm{E}\big(\varepsilon_{in}\varepsilon_{jn}\varepsilon_{sn}\varepsilon_{tn}\big)$. Here, we need to consider (1) $(i=s)\neq(j=t)$ and (2) $(i=t)\neq(j=s)$. Hence,

$$\mathrm{E}\big(Q_n^{'}\varepsilon_n\cdot\varepsilon_n^{'}Q_n\cdot\varepsilon_{sn}\varepsilon_{tn}\big)=\sigma_0^4\big(Q_{s\bullet,n}^{'}Q_{t\bullet,n}+Q_{t\bullet,n}^{'}Q_{s\bullet,n}\big)$$

(6) $\mathrm{E}\big(\varepsilon_n^{'}A_n\varepsilon_n\cdot\varepsilon_n^{'}B_n\varepsilon_n\cdot\varepsilon_{tn}\big)=\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^nA_{ij,n}B_{kl,n}\mathrm{E}\big(\varepsilon_{in}\varepsilon_{jn}\varepsilon_{kn}\varepsilon_{ln}\varepsilon_{tn}\big)$. There are four cases to consider: (1) $(i=k)\neq(j=l=t)$, (2) $(i=k=t)\neq(j=l)$, (3) $(i=l=t)\neq(j=k)$ and (4) $(i=l)\neq(j=k=t)$. Hence,

$$\begin{aligned}\mathrm{E}\big(\varepsilon_n^{'}A_n\varepsilon_n\cdot\varepsilon_n^{'}B_n\varepsilon_n\cdot\varepsilon_{tn}\big)&=\sigma_0^2\mu_3\sum_{i=1}^n\big(A_{it,n}+A_{ti,n}\big)\big(B_{it,n}+B_{ti,n}\big)=\sigma_0^2\mu_3\sum_{i=1}^nA_{it,n}^{(s)}B_{it,n}^{(s)}\\&=\sigma_0^2\mu_3A_{\bullet t,n}^{(s)'}B_{\bullet t,n}^{(s)}.\end{aligned}$$

(7.1) $\mathrm{E}\big(\varepsilon_n^{'}A_n\varepsilon_n\cdot\varepsilon_n^{'}Q_n\cdot\varepsilon_{tn}\big)=\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^nA_{ij,n}Q_{k\bullet,n}\mathrm{E}\big(\varepsilon_{in}\varepsilon_{jn}\varepsilon_{kn}\varepsilon_{tn}\big)$. Here, we need to consider: (1) $(i=k)\neq(j=t)$ and (2) $(i=t)\neq(j=k)$. Hence

$$\mathrm{E}\big(\varepsilon_n^{'}A_n\varepsilon_n\cdot\varepsilon_n^{'}Q_n\cdot\varepsilon_{tn}\big)=\sigma_0^4\sum_{i=1}^nA_{it,n}^{(s)}Q_{i\bullet,n}=\sigma_0^4A_{\bullet t,n}^{(s)'}Q_n.$$

(7.2) $\mathrm{E}\big(Q_n^{'}\varepsilon_n\cdot\varepsilon_n^{'}Q_n\cdot\varepsilon_{tn}\big)=\sum_{i=1}^n\sum_{j=1}^nQ_{i\bullet,n}^{'}Q_{j\bullet,n}\mathrm{E}\big(\varepsilon_{in}\varepsilon_{jn}\varepsilon_{tn}\big)=\mu_3Q_{t\bullet,n}^{'}Q_{t\bullet,n}$, since $\mathrm{E}\big(\varepsilon_{in}\varepsilon_{jn}\varepsilon_{tn}\big)$ is not zero only if $(i=j=t)$. □

Proof Lemma 3. The proofs for (1), (2) and (3) are given in Lin and Lee, (2010). For (4) and (5), see Dogan and Taspinar, (2013). □

Proof of Lemma 4. (1) $\frac{1}{2}\mathrm{tr}\big(A_n^{(s)}B_n^{(s)}\big)=\frac{1}{2}\mathrm{tr}\big(A_n^{(s)}B_n+A_n^{(s)}B_n^{'}\big)=\frac{1}{2}\mathrm{tr}\big(A_n^{(s)}B_n\big)+\frac{1}{2}\big(A_n^{(s)}B_n^{'}\big)=$

$\frac{1}{2}\mathrm{tr}\big(A_n^{(s)}B_n\big)+\frac{1}{2}\mathrm{tr}\big(A_n^{(s)}B_n\big)=\mathrm{tr}\big(A_n^{(s)}B_n\big)$. Then, by Lemma 1(5), we have

$$\mathrm{tr}\big(A_n^{(s)}B_n\big)=\frac{1}{2}\mathrm{tr}\big(A_n^{(s)}B_n^{(s)}\big)=\frac{1}{2}\,\mathrm{vec}'\left(A_n^{(s)}\right)\mathrm{vec}\left(B_n^{(s)}\right)$$

(2) $\frac{1}{2}\mathrm{tr}\big(A_n^{(s)}\left[B_n-\mathrm{D}\left(B_n\right)\right]^{(s)}\big)=\frac{1}{2}\mathrm{tr}\big(A_n^{(s)}\left[B_n-\mathrm{D}\left(B_n\right)\right]+A_n^{(s)}\left[B_n-\mathrm{D}\left(B_n\right)\right]'\big)=\frac{1}{2}\big[\mathrm{tr}\big(A_n^{(s)}B_n\big)-\mathrm{tr}\big(A_n^{(s)}\mathrm{D}\left(B_n\right)\big)+\mathrm{tr}\big(A_n^{(s)}B_n'\big)-\mathrm{tr}\big(A_n^{(s)}\mathrm{D}\left(B_n\right)\big)\big]=\mathrm{tr}\big(A_n^{(s)}B_n\big)$, since $\mathrm{tr}\big(A_n^{(s)}B_n\big)=\mathrm{tr}\big(A_n^{(s)}B_n'\big)$ and $\mathrm{tr}\big(A_n^{(s)}\mathrm{D}\left(B_n\right)\big)=0$. The last equality in this part simply follows from Lemma 1(5).

(3) The proof is as follows:

$$\begin{aligned}
&\frac{1}{2}\,\mathrm{vec}'\left(\left[B_n-\mathrm{D}(B_n)\right]^{(s)}\right)\mathrm{vec}\left(C_nP_{jn}^{(s)}\right)=\frac{1}{2}\mathrm{tr}\big(C_nP_{jn}^{(s)}\left[B_n-\mathrm{D}(B_n)\right]^{(s)}\big)\\
&=\frac{1}{2}\mathrm{tr}\big(C_nP_{jn}^{(s)}B_n^{(s)}\big)-\frac{1}{2}\mathrm{tr}\big(C_nP_{jn}^{(s)}\left(\mathrm{D}(B_n)\right)^{(s)}\big)=\frac{1}{2}\mathrm{tr}\big(C_nP_{jn}^{(s)}B_n^{(s)}\big)\\
&=\frac{1}{2}\mathrm{tr}\big(C_n\big(P_{jn}+P_{jn}'\big)B_n^{(s)}\big)=\frac{1}{2}\mathrm{tr}\big(C_nP_{jn}B_n^{(s)}\big)+\frac{1}{2}\mathrm{tr}\big(B_n^{(s)}P_{jn}'C_n\big)\\
&=\mathrm{tr}\big(C_nP_{jn}B_n^{(s)}\big)=\mathrm{tr}\big(C_nP_{jn}B_n\big)+\mathrm{tr}\big(C_nP_{jn}B_n'\big)\\
&=\mathrm{tr}\big(C_nP_{jn}B_n\big)+\mathrm{tr}\big(B_n'C_nP_{jn}\big)=\mathrm{tr}\big(C_nP_{jn}B_n\big)+\mathrm{tr}\big(P_{jn}'C_nB_n\big)\\
&=\mathrm{tr}\big(\left[C_nP_{jn}+P_{jn}'C_n\right]B_n\big)=\mathrm{tr}\big(\left(C_nP_{jn}\right)^{(s)}B_n\big)=\frac{1}{2}\mathrm{tr}\big(\left(C_nP_{jn}\right)^{(s)}B_n^{(s)}\big)\\
&=\frac{1}{2}\mathrm{tr}\big(\left(C_nP_{jn}\right)^{(s)}\left[B_n-\mathrm{D}(B_n)\right]^{(s)}\big)=\frac{1}{2}\,\mathrm{vec}'\left(\left[B_n-\mathrm{D}(B_n)\right]^{(s)}\right)\mathrm{vec}\left(\left(C_nP_{jn}\right)^{(s)}\right).
\end{aligned}$$

□

C Best Quadratic Moments Matrices

Lemma 3 in Appendix A can be used to derive Ω_n and Φ_n.

$$\Omega_n=\begin{bmatrix}
\mathrm{tr}\big(\Sigma_nP_{1n}(\Sigma_nP_{1n})^{(s)}\big) & \cdots & \mathrm{tr}\big(\Sigma_nP_{1n}(\Sigma_nP_{mn})^{(s)}\big) & 0_{1\times r}\\
\vdots & \vdots & \vdots & \vdots\\
\mathrm{tr}\big(\Sigma_nP_{mn}(\Sigma_nP_{1n})^{(s)}\big) & \cdots & \mathrm{tr}\big(\Sigma_nP_{mn}(\Sigma_nP_{mn})^{(s)}\big) & 0_{1\times r}\\
0_{r\times 1} & \cdots & 0_{r\times 1} & Q_n'\Sigma_nQ_n
\end{bmatrix}$$

$$\Phi_n=-\begin{bmatrix}
\mathrm{tr}\big(\Sigma_nH_n'P_{1n}^{(s)}\big) & \mathrm{tr}\big(\Sigma_n\overline{G}_n'P_{1n}^{(s)}\big) & 0_{1\times k}\\
\vdots & \vdots & \\
\mathrm{tr}\big(\Sigma_nH_n'P_{mn}^{(s)}\big) & \mathrm{tr}\big(\Sigma_n\overline{G}_n'P_{mn}^{(s)}\big) & 0_{1\times k}\\
0_{r\times 1} & Q_n'\overline{G}_n\overline{X}_n\beta_0 & Q_n'\overline{X}_n
\end{bmatrix}$$

Proof of Claim 1. Let $\mathcal{C}_{1mn}=\big[\mathrm{tr}\big(\Sigma_nH_n'P_{1n}^{(s)}\big),\ldots,\mathrm{tr}\big(\Sigma_nH_n'P_{mn}^{(s)}\big)\big]$ and $\mathcal{C}_{2mn}=\big[\mathrm{tr}\big(\Sigma_n\overline{G}_n'P_{1n}^{(s)}\big),\ldots,\mathrm{tr}\big(\Sigma_n\overline{G}_n'P_{mn}^{(s)}\big)\big]$. We will investigate an upper bound for $\mathcal{B}$ and $\mathcal{G}$. By

Lemma 4, when $P_{jn} \in \mathcal{P}_n$, a generic term in $\mathcal{C}_{1mn}$ can be written as

$$\begin{aligned}\operatorname{tr}\big(\Sigma_n H_n' P_{jn}^{(s)}\big) = \operatorname{tr}\big(\Sigma_n P_{jn}^{(s)} H_n\big) &= \frac{1}{2}\operatorname{tr}\big(\Sigma_n P_{jn}^{(s)} \left[H_n - \mathrm{D}(H_n)\right]^{(s)}\big) \\ &= \frac{1}{2}\operatorname{vec}'\left(\left[H_n - \mathrm{D}(H_n)\right]^{(s)}\right)\operatorname{vec}\left(\Sigma_n P_{jn}^{(s)}\right).\end{aligned}$$

Thus, $\mathcal{C}_{1mn} = \frac{1}{2}\operatorname{vec}'\left(\left[H_n - \mathrm{D}(H_n)\right]^{(s)}\right)\left[\operatorname{vec}\left(\Sigma_n P_{1n}^{(s)}\right)\cdots\operatorname{vec}\left(\Sigma_n P_{mn}^{(s)}\right)\right]$. The above same argument also applies to $\mathcal{C}_{2mn}$. Hence, $\mathcal{C}_{2mn} = \frac{1}{2}\operatorname{vec}'\left(\left[\overline{G}_n - \mathrm{D}\left(\overline{G}_n\right)\right]^{(s)}\right)\left[\operatorname{vec}\left(\Sigma_n P_{1n}^{(s)}\right)\cdots\operatorname{vec}\left(\Sigma_n P_{mn}^{(s)}\right)\right]$. By Lemma 4 (3), we can also write a generic term of $\mathcal{C}_{1mn}$ in the following way:

$$\frac{1}{2}\operatorname{vec}'\left(\left[H_n - \mathrm{D}\left(H_n\right)\right]^{(s)}\right)\operatorname{vec}\left(\Sigma_n P_{jn}^{(s)}\right) = \frac{1}{2}\operatorname{vec}'\left(\left[H_n - \mathrm{D}\left(H_n\right)\right]^{(s)}\right)\operatorname{vec}\left(\left(\Sigma_n P_{jn}\right)^{(s)}\right).$$

Hence, $\mathcal{C}_{1mn}$ and $\mathcal{C}_{2mn}$ can be written as

$$\begin{aligned}\mathcal{C}_{1mn} &= \frac{1}{2}\operatorname{vec}'\left(\left[H_n - \mathrm{D}\left(H_n\right)\right]^{(s)}\right)\left[\operatorname{vec}\left(\left(\Sigma_n P_{1n}\right)^{(s)}\right),\ldots,\operatorname{vec}\left(\left(\Sigma_n P_{mn}\right)^{(s)}\right)\right], \\ \mathcal{C}_{2mn} &= \frac{1}{2}\operatorname{vec}'\left(\left[\overline{G}_n - \mathrm{D}\left(\overline{G}_n\right)\right]^{(s)}\right)\left[\operatorname{vec}\left(\left(\Sigma_n P_{1n}\right)^{(s)}\right),\ldots,\operatorname{vec}\left(\left(\Sigma_n P_{mn}\right)^{(s)}\right)\right].\end{aligned}$$

First, we investigate an upper bound for $\mathcal{B}$ by using the Schwartz inequality:

$$\begin{aligned}|\mathcal{B}| &= \left|\mathcal{C}_{1mn}\mathcal{A}_n^{-1}\mathcal{C}_{1mn}'\right| \leq \left\|\mathcal{A}_n^{-1}\mathcal{C}_{1mn}'\right\| \times \left\|\mathcal{C}_{1mn}\right\| \leq \left\|\mathcal{A}_n^{-1}\right\| \times \left\|\mathcal{C}_{1mn}'\right\| \times \left\|\mathcal{C}_{1mn}\right\| \\ &= \left\|\left(\left[\operatorname{vec}\left((\Sigma_n P_{1n})^{(s)}\right),\ldots,\operatorname{vec}\left((\Sigma_n P_{mn})^{(s)}\right)\right]'\left[\operatorname{vec}\left((\Sigma_n P_{1n})^{(s)}\right),\ldots,\operatorname{vec}\left((\Sigma_n P_{mn})^{(s)}\right)\right]\right)^{-1}\right\| \\ &\times \left\|\frac{1}{2}\operatorname{vec}'\left([H_n - \mathrm{D}(H_n)]^{(s)}\right)\left[\operatorname{vec}\left((\Sigma_n P_{1n})^{(s)}\right),\ldots,\operatorname{vec}\left((\Sigma_n P_{mn})^{(s)}\right)\right]\right\| \\ &\times \left\|\left[\operatorname{vec}\left((\Sigma_n P_{1n})^{(s)}\right),\ldots,\operatorname{vec}\left((\Sigma_n P_{mn})^{(s)}\right)\right]'\operatorname{vec}\left([H_n - \mathrm{D}(H_n)]^{(s)}\right)\right\| \\ &\leq \left\|\left(\left[\operatorname{vec}\left((\Sigma_n P_{1n})^{(s)}\right),\ldots,\operatorname{vec}\left((\Sigma_n P_{mn})^{(s)}\right)\right]'\left[\operatorname{vec}\left((\Sigma_n P_{1n})^{(s)}\right),\ldots,\operatorname{vec}\left((\Sigma_n P_{mn})^{(s)}\right)\right]\right)^{-1}\right\| \\ &\times \left\|\operatorname{vec}\left([H_n - \mathrm{D}(H_n)]^{(s)}\right)\right\| \times \left\|\left[\operatorname{vec}\left((\Sigma_n P_{1n})^{(s)}\right),\ldots,\operatorname{vec}\left((\Sigma_n P_{mn})^{(s)}\right)\right]\right\| \\ &\times \frac{1}{2}\left\|\operatorname{vec}\left([H_n - \mathrm{D}(H_n)]^{(s)}\right)\right\| \times \left\|\left[\operatorname{vec}\left((\Sigma_n P_{1n})^{(s)}\right),\ldots,\operatorname{vec}\left(\left(\Sigma_n P_{mn}^{(s)}\right)\right)\right]\right\| \\ &= \frac{1}{2}\left\|\operatorname{vec}\left([H_n - \mathrm{D}(H_n)]^{(s)}\right)\right\| \times \left\|\operatorname{vec}\left([H_n - \mathrm{D}(H_n)]^{(s)}\right)\right\|.\end{aligned}$$

Hence, we obtain

$$\begin{aligned}|\mathcal{B}| \leq & \left\|\operatorname{vec}\left([H_n - \mathrm{D}(H_n)]^{(s)}\right)\right\| \times \left\|\frac{1}{2}\operatorname{vec}\left([H_n - \mathrm{D}(H_n)]^{(s)}\right)\right\| \\ &= \frac{1}{2}\left\|\operatorname{vec}'\left([H_n - \mathrm{D}(H_n)]^{(s)}\right) \times \operatorname{vec}\left([H_n - \mathrm{D}(H_n)]^{(s)}\right)\right\| \\ &= \operatorname{tr}\left(H_n\left[H_n - \mathrm{D}(H_n)\right]^{(s)}\right) = \operatorname{tr}\left(\Sigma_n^{-1}\left[H_n - \mathrm{D}(H_n)\right]^{(s)} H_n \Sigma_n\right).\end{aligned} \tag{C.1}$$

The argument above also applies to $\mathcal{G}$. That is,

$$
\begin{aligned}
|\mathcal{G}| \leq & \frac{1}{2}\left\| \operatorname{vec}\left(\left[\overline{G}_n - \mathrm{D}(\overline{G}_n)\right]^{(s)}\right) \right\| \times \left\| \operatorname{vec}\left(\left[\overline{G}_n - \mathrm{D}(\overline{G}_n)\right]^{(s)}\right) \right\| \\
&= \operatorname{tr}\left(\overline{G}_n \left[\overline{G}_n - \mathrm{D}(\overline{G}_n)\right]^{(s)}\right) = \operatorname{tr}\left(\Sigma_n^{-1} \left[\overline{G}_n - \mathrm{D}(\overline{G}_n)\right]^{(s)} \overline{G}_n \Sigma_n\right).
\end{aligned} \tag{C.2}
$$

The same argument for $\mathcal{B}$ indicates that

$$
|\mathcal{D}| \leq \left\| \operatorname{vec}\left(\left[\overline{G}_n - \mathrm{D}(\overline{G}_n)\right]^{(s)}\right) \right\| \times \left\| \frac{1}{2} \operatorname{vec}\left(\left[H_n - \mathrm{D}(H_n)\right]^{(s)}\right) \right\|. \tag{C.3}
$$

The results in (C.1), (C.2) and (C.3) indicates that $\Sigma_n^{-1}\left[H_n - \mathrm{D}(H_n)\right]$ and $\Sigma_n^{-1}\left[\overline{G}_n - \mathrm{D}(\overline{G}_n)\right]$ provide the best matrices for the quadratic moment functions.

$\square$