City University of New York (CUNY)

CUNY Academic Works

Economics Working Papers

CUNY Academic Works

2013

GMM Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances

Osman Dogan
CUNY Graduate Center

Süleyman Taşpınar
CUNY Graduate Center

Follow this and additional works at: https://academicworks.cuny.edu/gc_econ_wp

Part of the [Economics Commons](#)

[How does access to this work benefit you? Let us know!](#)

More information about this work at: https://academicworks.cuny.edu/gc_econ_wp/6

Discover additional works at: <https://academicworks.cuny.edu>

This work is made publicly available by the City University of New York (CUNY).
Contact: AcademicWorks@cuny.edu

**GMM Estimation of Spatial Autoregressive Models
with Autoregressive and Heteroskedastic Disturbances**

Osman Dogan

Süleyman Taşpınar

Working Paper 1

Ph.D. Program in Economics
CUNY Graduate Center
365 Fifth Avenue
New York, NY 10016
December 2013

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. We would like to thank Wim Vijverberg for insightful and instructive comments on this work.

© 2013 by Osman Dogan and Süleyman Taşpınar. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

GMM Estimation of Spatial Autoregressive Models with Autoregressive and
Heteroskedastic Disturbances
Osman Dogan and Süleyman Taşpınar
CUNY Graduate Center Working Paper No. 1
December 2013
JEL No. C13, C21, C31

ABSTRACT

We consider a spatial econometric model containing a spatial lag in the dependent variable and the disturbance term with an unknown form of heteroskedasticity in innovations. We first prove that the maximum likelihood (ML) estimator for spatial autoregressive models is generally inconsistent when heteroskedasticity is not taken into account in the estimation. We show that the necessary condition for the consistency of the ML estimator of spatial autoregressive parameters depends on the structure of the spatial weight matrices. Then, we extend the robust generalized method of moment (GMM) estimation approach in Lin and Lee (2010) for the spatial model allowing for a spatial lag not only in the dependent variable but also in the disturbance term. We show the consistency of the robust GMM estimator and determine its asymptotic distribution. Finally, through a comprehensive Monte Carlo simulation, we compare finite sample properties of the robust GMM estimator with other estimators proposed in the literature.

Osman Dogan
Ph.D. Program in Economics
City University of New York
365 Fifth Avenue
New York, NY 10016
ODogan@gc.cuny.edu

Süleyman Taşpınar
Ph.D. Program in Economics
City University of New York
365 Fifth Avenue
New York, NY 10016
STaspinar@gc.cuny.edu

1 Introduction

Spatial econometric models that have a long history in regional science and geography has been receiving attention in economics in recent years. Spatial econometric models allow regression specifications through which spatial dependence among observations can be incorporated in economic analysis and in the estimation of models. The spatial dependence is a special form of cross-sectional dependence among observations determined by locations of observations in space. The estimation of models with spatial dependence requires special estimation techniques. There are three main estimation approaches: (i) the maximum likelihood (ML) estimation method, (ii) the generalized method of moment (GMM/IV) estimation method, and (iii) the Bayesian Markov Chain Monte Carlo (MCMC) estimation method. For many spatial model specifications, the ML estimation has been the most widely used technique and has often been the only technique that is implemented (Anselin, 1988; LeSage and Pace, 2009). However, formal results concerning the asymptotic properties of the (quasi) ML estimator have recently been established in Lee (2004) only for pure spatial and spatial autoregressive models. The ML estimation can involve a significant computational difficulty due to the presence of the determinant of a matrix in the likelihood function, whose dimensions depend on the sample size. (Das, Kelejian, and Prucha, 2003; Kelejian and Prucha, 1998, 2010). Several solutions have been suggested to overcome the computational burden of the ML method (Barry and Pace, 1999; LeSage and Pace, 2004, 2007; Ord, 1975; Pace and Barry, 1997a,b; Smirnov and Anselin, 2001).¹

The GMM and IV estimators have the advantage that they do not require any distributional assumption for the disturbance term and remain to be computationally more feasible than ML estimation. In the literature, different kinds of two stage least squares (2SLS) estimators corresponding to the different set of instrumental variables have been suggested (Anselin, 1988; Kelejian and Prucha, 1998, 2007, 2010; Lee, 2003, 2007a). The spatial structure of regression equations motivate the selection of the instruments which are usually constructed from the exogenous variables and spatial weight matrices. Despite its computational simplicity, the 2SLS estimator is inefficient relative to the ML estimator. The inefficiency arises because the 2SLS estimator focuses only on the deterministic part of the endogenous variable (i.e., the spatial lag term) and the information in the stochastic part is not used in the estimation.

Kelejian and Prucha (1998, 2010) propose a multi-step estimation method that involves a combination of IV and GMM estimation for the spatial model that has a spatial autoregressive process in the dependent variable and disturbance term (for short SARAR(1,1)). This kind of model specification is often referred as the Kelejian-Prucha Model (Elhorst, 2010). In the first step, the initial estimates of the parameters of the exogenous variable and the autoregressive parameter of the spatial lag of the dependent variable are estimated by the 2SLS estimator. In the second step, residuals from the first step are used to estimate the autoregressive parameter of the spatial lag of the disturbance term by the GMM estimator. In the final step, the parameters are re-estimated by the

¹For the Bayesian MCMC approach, see Lesage (1997), Parent and Lesage (2007) and LeSage and Pace (2009).

2SLS estimator after transforming the model via a Cochrane-Orcut type transformation to account for the spatial correlation. However, the estimation approach in Kelejian and Prucha (1998) is inefficient relative to the ML estimation (Prucha, Forthcoming 2012). The extensive Monte Carlo results in Das, Kelejian, and Prucha (2003) demonstrate that the difference between finite sample efficiency, measured with root mean squared errors (RMSE), between the ML and the GMM and IV estimators of Kelejian and Prucha (1998, 1999) is very small. Drukker, Egger, and Prucha (2012) consider the specification SARAR(1,1) where they allow for endogenous regressors in addition to spatial lag of the dependent variable. The estimation approach involves several steps and is an extension of GMM/IV estimation method of Kelejian and Prucha (1998, 1999).

To increase the efficiency of the GMM estimator, Lee (2007a, 2007b), Lin and Lee (2010), Liu, Lee, and Bollinger (2010), and Lee and Liu (2010b) suggest sets of moment functions that are linear and quadratic in the disturbance term for the GMM estimation. In this approach, the linear moment functions are based on the deterministic part of the spatial lag term and the quadratic moment functions are constructed for exploiting the stochastic part of the spatial lag variable (i.e., the endogenous variable). The quadratic moment functions are chosen in a such way that the GMM estimator is asymptotically equivalent to the ML estimator when disturbances are i.i.d. normal. When disturbances are simply i.i.d., Liu, Lee, and Bollinger (2010) and Lee and Liu (2010b) show that the one step GMM estimator (joint GMM estimator) is more efficient than the quasi ML estimator, respectively for the case of an SARAR(1,1) and an SARAR(p,q).

Most of the estimation methods mentioned above are valid under the assumption that the disturbance terms of the spatial models are i.i.d. In many regression applications, heteroskedasticity might be present.² In the presence of unknown heteroskedasticity, the ML and GMM estimators are generally not consistent. The ML estimator is inconsistent if the heteroskedasticity is not incorporated into the estimation. For an SARAR(1,0), Lin and Lee (2010) shows that the likelihood function is not maximized at the true parameter values in the presence of the unknown heteroskedasticity. The GMM estimators are also inconsistent since the moment functions are often designed under the assumption that disturbances are i.i.d. Hence, the orthogonality conditions for the moment functions might not be satisfied. To handle unknown heteroskedasticity, Kelejian and Prucha (2010) extend their estimation approach by modifying the moment functions for the case of an SARAR(1,1). Badinger and Egger (2011) extend the robust estimation approach in Kelejian and Prucha (2010) to the case of SARAR(p,q). Likewise, Lin and Lee (2010) suggest a one-step robust GMM estimator for the model with only spatial dependence in the dependent variable.³

In the present study, the one-step robust GMM estimation approach suggested by Lin and Lee (2010) is extended to the spatial model with a spatial autoregressive process in both the dependent variable and the disturbance term under the assumption that there is unknown form of heteroskedasticity in the disturbance term. We show that the ML estimator might not be consistent in the presence of the unknown heteroskedasticity, as the probability limits of the first order con-

²For an example, see the empirical application in Lin and Lee (2010).

³For a robust 2SLS estimator of SARAR(1,0), see Anselin (2007).

ditions evaluated at the true parameter values are generally not zero. We show that the necessary condition for the consistency of the ML estimator of spatial autoregressive parameters depends on the structure of the spatial weight matrices. Then, a robust GMM estimator is derived from a set of moment functions that are composed of both linear and quadratic moment functions. The consistency of the estimator is established and its asymptotic distribution is determined. Finite sample properties are compared with that of other estimators through a comprehensive Monte Carlo simulation.

This paper is organized in the following way. In Section 2, the theoretical motivation for the case of an SARAR(1,1) is provided along with the model assumptions and their implications. In Section 3, the GMM estimators that have been suggested in the literature are reviewed. In Section 4.1, we show the inconsistency of the ML estimator in the presence of unknown heteroskedasticity. We determine the asymptotic bias of the parameters of the exogenous variables. In Section 4.2, a robust GMM estimation method is considered for the case of an SARAR(1,1). The identification conditions are determined. The main large sample properties of the robust GMM estimator are stated in three propositions. The Monte Carlo simulations are carried out in Section 5. Section 6 closes with concluding remarks.

2 The Model Specification and Theoretical Motivation

In the literature, spatial dependence in regression specifications is categorized in two broad categories known as spatial lag and spatial error models. The spatial lag model includes functional forms in which a dependent variable at a point in space depends on dependent variables of surrounding locations. The equilibrium outcome of theoretical economic models of interacting spatial units motivates this kind of specification. In spatial error models, cross-sectional correlations among error terms are incorporated into the specification and estimation of models. Measurement error in data usually tends to vary systematically over space, which causes spatial dependence among error terms of a specification.⁴

In this study, the following first order SARAR(1,1) specification is considered:

$$Y_n = \lambda_0 W_n Y_n + X_n \beta_0 + u_n, \quad u_n = \rho_0 M_n u_n + \varepsilon_n, \quad (2.1)$$

where Y_n is $n \times 1$ vector of dependent variable, X_n is $n \times k$ matrix of nonstochastic exogenous variables, W_n and M_n are $n \times n$ spatial weight matrices of known constants with zero diagonal elements, and ε_n is $n \times 1$ vector of disturbances (or innovations). The variables $W_n Y_n$ and $M_n u_n$ are known respectively as spatial lag of the dependent variable and the disturbance term. The spatial effect parameters λ_0 and ρ_0 are known as the spatial autoregressive parameters. The above specification is fairly general in the sense that it allows for spatial spillovers in the dependent variable, exogenous variables and disturbances.⁵ As the spatial data is characterized with triangular arrays,

⁴For the motivation of model specifications, see Anselin (1988); Anselin (2007) and LeSage and Pace (2009)

⁵Elhorst (2010) names the model with spatial spillovers in the dependent variable, exogenous variable and distur-

the variables in (2.1) have subscript n .⁶ Let Θ be the parameter space of the model. In order to distinguish the true parameter vector from other possible values in Θ , the model is stated with the true parameter vector $\theta_0 = (\rho_0, \varsigma_0)'$ with $\varsigma_0 = (\lambda_0, \beta_0)'$.

For the notational simplicity, we denote $S_n(\lambda) = (I_n - \lambda W_n)$, $R_n(\rho) = (I_n - \rho M_n)$, $G_n(\lambda) = W_n S_n^{-1}(\lambda)$ and $H_n(\rho) = M_n R_n^{-1}(\rho)$. Also, at the true parameter values (ρ_0, λ_0) , we denote $S_n(\lambda_0) = S_n$, $R_n(\rho_0) = R_n$, $G_n(\lambda_0) = G_n$, $H_n(\rho_0) = H_n$ and $\bar{G}_n = R_n G_n R_n^{-1}$.

Next, assumptions that are required for the asymptotic properties of estimators are elaborated and then their interpretations are considered for (2.1).

Assumption 1: *The elements ε_{ni} of the disturbance term ε_n are distributed independently with mean zero and variance σ_{ni}^2 , and $E|\varepsilon_{in}|^\nu < \infty$ for some $\nu > 4$ for all n and i .*

This assumption allows independent and heteroskedastic disturbances. The elements of the disturbance term have moments higher than the fourth moment. This condition is specifically required for the application of the central limit theorem for the quadratic form given in Kelejian and Prucha (2010) for the GMM estimator. In addition, the variance of a quadratic form in ε_n exists and is finite when the first four moments are finite.⁷ Finally, Liapunov's inequality guarantees that the moment less than ν are also uniformly bounded for all n and i .

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

In the literature, weight matrices are usually treated as exogenous and fixed. Lee (2004, 2007b) formulate the weight matrix as a function of the sample size. According to this formulation, the sequence of weight matrix $\{W_n\}$ is uniformly bounded in both row and column sums and its elements $w_{n,ij}$ s are $O(\frac{1}{h_n})$. The sequence $\{h_n\}$ can be bounded or divergent with the property that $\lim_{n \rightarrow 0} \frac{h_n}{n} = 0$, which implies that h_n is allowed to diverge only at a rate slower than that of n . This formulation provides an explicit way that describes how the spatial weight matrix W_n is expanding as the sample size increases. For example, assume that an economy consists of r regions and each region is populated by k agents. Then, the total number of observations from this economy is $n = rk$. In addition, in each region each agent is equally affected by other agents of the same region. There is no interaction among regions. Denote the row normalized spatial weight matrix of a region by C_k which is given by $\frac{1}{k-1}(l_k l_k' - I_k)$ where l_k is a k -dimensional vector of ones. Then, the spatial weight matrix W_n for this economy is block diagonal $W_n = I_r \otimes C_k$. Each element in

bance term as the Manski Model. He states that that the parameter estimates cannot be interpreted in a meaningful way for this kind of model since the endogenous and exogenous effects cannot be distinguished from each other. See also Anselin (2007).

⁶See Kelejian and Prucha (2010).

⁷For the variance of the quadratic form in ε_n , see Lemma 2 (3).

a diagonal block is given by $\frac{1}{k-1}$, so that $w_{n,ij} = O(\frac{1}{k-1})$. Then, $\frac{h_n}{n} = \frac{k-1}{k \times r} = O(\frac{1}{r})$. Assume that the increase in n is generated by the increase of both r and k . Then, the fraction $\frac{h_n}{n}$ tends to zero, as h_n diverges to infinity. This kind of spatial weight matrix is used for large group interactions scenarios which have important implications for the convergence rate of estimators (Lee, 2004).⁸

For large group interactions for which $\lim_{n \rightarrow \infty} \frac{h_n}{n} \neq 0$, consistency of estimators might not be available. As an example, Kelejian and Prucha (2002) and Yuzefovich, Kelejian, and Prucha (2006) consider a row normalized spatial weight matrix that has equal weights for all observations. The spatial weight matrix is formulated as $W_n = \frac{1}{n-1} l_n l_n' - \frac{1}{n-1} I_n$ where each off-diagonal element is $\frac{1}{n-1}$. In that case, $w_{n,ij} = O(\frac{1}{n-1})$ and $\lim_{n \rightarrow \infty} \frac{h_n}{n} = \lim_{n \rightarrow \infty} \frac{n-1}{n} = 1$. With this specification, Kelejian and Prucha (2002) show that OLS, 2SLS and ML estimators are inconsistent for spatial autoregressive models. In this study, we assumed that h_n is bounded.

The uniform boundedness of the terms in *Assumption 1* and *2* is motivated to control spatial autocorrelations in the model at a tractable level (Kelejian and Prucha, 1998).⁹ *Assumption 2* also implies that the model in (2.1) represents an equilibrium relation for the dependent variable. By this assumption, the reduced form of the model becomes feasible as $Y_n = S_n^{-1} X_n \beta_0 + S_n^{-1} R_n^{-1} \varepsilon_n$. Finally, the statement of *Assumption 2* is assumed to hold at the true and arbitrary autoregressive parameter vector. The uniform boundedness of $S_n^{-1}(\lambda)$ and $R_n^{-1}(\rho)$ is required for the ML estimator not for the GMM estimator (Liu, Lee, and Bollinger, 2010).

In the literature, the parameter space for spatial autoregressive parameters λ_0 and ρ_0 is restricted to the interval $(-1, 1)$, when spatial weight matrices are row normalized.¹⁰ In that case, matrices S_n and R_n are nonsingular. More general parameter spaces have also been considered in the literature.¹¹ Let ν_{jn} for $j = 1, \dots, n$ be eigenvalues of W_n . The spectral radius of W_n is defined by $\tau_n = \max_{1 \leq j \leq n} |\nu_{jn}|$. Then, S_n is nonsingular for all values of λ_0 in the interval $(\frac{-1}{\tau_n}, \frac{1}{\tau_n})$. However, the computation of eigenvalues involves computational difficulties, and becomes numerically unstable for spatial weight matrices with more than 1000 observations (Smirnov and Anselin, 2001). Another formulation for the parameter space base on the maximum row and column sums of spatial weight matrices is also considered in the literature. Denote R_i and C_j respectively as i th row sum and j th column sum of W_n in absolute value. Let the maximum row sum be given by $R = \max_i \sum_{j=1}^n |w_{ij}| = \max_i^n R_i$. Likewise, the maximum column sum is defined by $C = \max_j \sum_{i=1}^n |w_{ij}| = \max_j^n C_j$. Let $m = \max\{C, R\}$. Then, S_n is nonsingular for all values of λ_0 in the interval $(\frac{-1}{m}, \frac{1}{m})$.¹²

The following assumptions are the usual regularity conditions required for the GMM estimator. Throughout this study, the vector of moment functions considered for the GMM estimator is in the form of $g(\theta_0) = (\varepsilon_n' P_{1n} \varepsilon_n, \dots, \varepsilon_n' P_{mn} \varepsilon_n, \varepsilon_n' Q_n)'$. The moment functions involving $n \times n$

⁸For examples of this kind of weight matrices, see Case (1991, 1992).

⁹For a definition and some properties of uniform boundedness see Kelejian and Prucha (2010).

¹⁰Kelejian and Prucha (2010) states that the interval $(-1, 1)$ is not natural in the sense that equivalent model formulation are possible by applying an arbitrary scale factor to autoregressive parameters and its inverse to weight matrices and therefore the parameter space will depend on the scaling factor.

¹¹Elhorst, Lacombe, and Piras (2012) outline a simple procedure for finding the parameter space for models with multiple spatial weights matrices.

¹²For a proof of this result see Kelejian and Prucha (2007).

constant matrices P_{jn} for $j = 1, \dots, m$ are known as quadratic moment functions. The last moment function $Q'_n \varepsilon_n$ is the linear moment function, where the full column rank matrix Q_n is $n \times k^*$ with $k^* \geq k + 1$. The matrices P_{jn} s and Q_n are chosen in such way that orthogonality conditions of population moment functions are not violated. Let $\mathcal{P}_{1n}$ be the class of $n \times n$ constant matrices with zero trace and $\mathcal{P}_{2n}$ be class of $n \times n$ constant matrices with zero diagonal elements.¹³ The quadratic moment functions involving matrices from these both classes satisfy the orthogonality conditions when disturbance terms are i.i.d. As it will be shown, when disturbance terms are merely independent, matrices from the class $\mathcal{P}_{1n} \setminus \mathcal{P}_{2n}$ can not be used to form quadratic moment functions.¹⁴ *Assumption 4* states regularity conditions for these matrices and the last assumption characterizes the parameter space.

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

Assumption 4: Elements of IV matrix Q_n are uniformly bounded. P_{jn} for $j = 1, \dots, m$ is uniformly bounded in absolute value in row and column sums.

Assumption 5: The parameter space Θ is a compact subset of $\mathbb{R}^{k+2}$ and θ_0 is in the interior of Θ .

3 GMM Estimation of Spatial Autoregressive Models

The GMM estimation approach depends on the moment functions that are derived from the structure of the model. The endogenous variable $W_n Y_n$ on the right hand side of the model is given more explicitly by $W_n Y_n = W_n S_n^{-1} X_n \beta_0 + W_n S_n^{-1} R_n^{-1} \varepsilon_n = G_n X_n \beta_0 + G_n R_n^{-1} \varepsilon_n$ where $G_n = W_n S_n^{-1} = W_n (I_n - \lambda_0 W_n)^{-1}$ exists by *Assumption 2*. Thus, $W_n Y_n$ is a function of a non-stochastic term $G_n X_n \beta_0$ and a stochastic term $G_n R_n^{-1} \varepsilon_n$. Lee (2001a, 2007a), Liu, Lee, and Bollinger (2010), Lee and Liu (2010b) and Lin and Lee (2010) form moment functions based on stochastic and non-stochastic terms. The non-stochastic term is instrumented by $Q_{1n} = (R_n G_n X_n \beta_0, R_n X_n)$, which forms the linear moment function $Q'_{1n} \varepsilon_n$. The linear moment matrix Q_{1n} is constructed from the expectation of $Z_n = (W_n Y_n, X_n)$. Given consistent initial estimates of λ_0 , ρ_0 and β_0 , the IV matrix Q_{1n} becomes available. Lee (2003) shows that the 2SLS estimator with Q_{1n} is best in the sense that its asymptotic variance covariance matrix is the smallest among the class of 2SLS estimators based on linear moment conditions.

The stochastic part $G_n R_n^{-1} \varepsilon_n$ of $W_n Y_n$ is instrumented by $P_{jn} \varepsilon_n$, where $P_{jn} \in \mathcal{P}_{1n}$ and/or $P_{jn} \in \mathcal{P}_{2n}$ for $j = 1, \dots, m$. In this case, the quadratic moment is in the form of $\varepsilon'_n P_{jn} \varepsilon_n$ and the orthogonality (or population moment) condition is satisfied when disturbances are simply i.i.d. In that case, $E(\varepsilon'_n P_{jn} \varepsilon_n) = \text{tr}(P_{jn} E(\varepsilon_n \varepsilon'_n)) = 0$ for P_{jn} s from either $\mathcal{P}_{1n}$ or $\mathcal{P}_{2n}$.¹⁵ For both

¹³Note that $\mathcal{P}_{2n}$ is a subclass of $\mathcal{P}_{1n}$, i.e., $\mathcal{P}_{2n} \subset \mathcal{P}_{1n}$.

¹⁴Here, $\mathcal{P}_{1n} \setminus \mathcal{P}_{2n}$ denotes set-theoretic difference of $\mathcal{P}_{1n}$ and $\mathcal{P}_{2n}$.

¹⁵ $\text{tr}(\cdot)$ returns the sum of the diagonal elements of an input matrix.

stochastic and non-stochastic term, the IVs are constructed in a such way that they are correlated with $W_n Y_n$ but uncorrelated with ε_n .¹⁶

The consistency of the GMM estimator does not depend on a particular P_{jn} but the asymptotic variance-covariance matrix is a function of P_{jn} s. Therefore, for the selection of P_{jn} s, the asymptotic efficiency of estimators needs to be considered. Liu, Lee, and Bollinger (2010) and Lee and Liu (2010b) provide the best selection of $P_{jn} \in \mathcal{P}_{1n}$ in the case of an SARAR (1,1) and SARAR(p,q), respectively.¹⁷ In the case of SARAR (1,1) with i.i.d normal innovations, the best selection is (1) $P_{1n} = (R_n G_n R_n^{-1} - \frac{1}{n} \text{tr}(R_n G_n R_n^{-1}) I_n)$, and (2) $P_{2n} = (H_n - \frac{1}{n} \text{tr}(H_n) I_n)$. Let $g_n(\theta) = (\varepsilon_n'(\theta) P_{1n} \varepsilon_n(\theta), \varepsilon_n'(\theta) P_{2n} \varepsilon_n(\theta), \varepsilon_n'(\theta) Q_{1n})'$ be the set of sample moment functions. Liu, Lee, and Bollinger (2010) show that given the set of moment function $g_n(\theta)$, any other moment functions that can be added to this set is redundant. They also show that the ML estimator is characterized by the set of moment functions $g_n(\theta)$, therefore, the GMM estimator based on these moment functions is asymptotically equivalent to the ML estimator. When the innovations are simply i.i.d, Liu, Lee, and Bollinger (2010) suggest another best set of quadratic moment functions so that the optimal GMM estimator is asymptotically more efficient than the quasi ML estimator.

When disturbance terms are independent and heteroskedastic, some matrices P_{jn} with zero trace property cannot be used in the formation of the quadratic moment functions. Let $\Sigma_n = \text{Diag}(\sigma_{1n}^2, \dots, \sigma_{nn}^2)$ be the diagonal variance matrix of the disturbance terms. If $P_{jn} \in (\mathcal{P}_{1n} \setminus \mathcal{P}_{2n})$ for any $j = 1, \dots, m$, then the covariance $E(\varepsilon_n' P_{jn} \varepsilon_n) = \text{tr}(P_{jn} E(\varepsilon_n \varepsilon_n')) = \text{tr}(P_{jn} \Sigma_n) \neq 0$. On the other hand, P_{jn} with zero diagonal property is still available for the formation of the quadratic moments, since $\text{tr}(P_{jn} E(\varepsilon_n \varepsilon_n')) = \text{tr}(P_{jn} \Sigma_n) = 0$ for any $P_{jn} \in \mathcal{P}_{2n}$. Thus, the class of matrices with zero diagonal elements provides robustness for the heteroskedasticity.

Lin and Lee (2010) extend the GMM estimation method in Lee (2001a, 2007a) to SARAR(1,0) that has an unknown form of heteroskedasticity in innovations. The quadratic moment functions are based on the class $\mathcal{P}_{2n}$. Let $\varsigma_0 = (\lambda_0, \beta_0')'$ be the parameter vector of the model, Lin and Lee (2010) suggest the set of moment functions $g_n(\varsigma) = (\varepsilon_n'(\varsigma) P_{1n} \varepsilon_n(\varsigma), \varepsilon_n'(\varsigma) Q_{2n})'$, where $P_{1n} = (G_n - \text{Diag}(G_n)) \in \mathcal{P}_{2n}$ and $Q_{2n} = (G_n X_n \beta_0, X_n)$.¹⁸ The optimal robust GMM estimator derived from $\min_{\varsigma \in \Theta} g_n'(\varsigma) \hat{\Omega}_n^{-1} g_n(\varsigma)$ is consistent and asymptotically normally distributed. Here, $\hat{\Omega}_n$ is an estimate of $\text{var}(g_n(\varsigma_0)) = \Omega_n$ based on an initial $\sqrt{n}$ -consistent estimator of ς_0 . For the heteroskedastic case, the best selection of P_{nj} is not available. Lin and Lee (2010) suggest that the selection from $\mathcal{P}_{2n}$ for the simply i.i.d case can be used for the case of independently distributed disturbance terms. Thus, the consistent estimates of $(G_n - \text{Diag}(G_n))$ and $(G_n X_n \beta_0, X_n)$ are used in $g_n(\varsigma)$ for the robust optimal GMM estimator.

The computationally simple two-step GMM estimation approach in Kelejian and Prucha (1998, 1999) for the case of an SARAR(1,1) is based on two quadratic moment matrices from $\mathcal{P}_{1n}$: (1) $P_{1n} = v(M_n' M_n - \frac{1}{n} \text{tr}(M_n' M_n) I_n)$ with $v = \frac{1}{1 + (n^{-1} \text{tr}(M_n' M_n))^2}$, and (2) $P_{2n} = M_n$. When the innovations

¹⁶Note that $\text{cov}(Q_n, \varepsilon_n) = 0$ and $\text{cov}(P_{jn} \varepsilon_n, \varepsilon_n) = 0$.

¹⁷Liu, Lee, and Bollinger (2010) also consider the best GMM estimation for the case of an SARAR(1,0) and an SARAR(0,1).

¹⁸ $\text{Diag}(\cdot)$ is an operator that creates a matrix from the diagonal elements of an input matrix.

are heteroskedastic, the orthogonality condition of the quadratic moment function based on P_{1n} is violated, therefore Kelejian and Prucha (2010) consider a quadratic moment matrix from the class $\mathcal{P}_{2n}$. In that case, the first moment is formed with $P_{1n} = (M_n' M_n - \text{Diag}(M_n' M_n))$.

The linear moment conditions in Kelejian and Prucha (1998) are based on the linearly independent columns of the set $Q_{3n} = (X_n, W_n X_n, W_n^2 X_n, \dots, M_n W_n X_n, M_n W_n^2 X_n, \dots)$. The IV matrix Q_{3n} provides an approximation for $E(Z_n)$ and $E(M_n Z_n)$.

For the illustration of two-step GMM estimation approach of Kelejian and Prucha (2010), let $g_n(\rho, \varsigma) = \frac{1}{n}(\varepsilon_n'(\theta)P_{1n}\varepsilon_n(\theta), \varepsilon_n'(\theta)P_{2n}\varepsilon_n(\theta))'$ be the set of sample moment functions, and let $\tilde{\varsigma}_n$ be an initial consistent estimator based on the instrument matrix Q_{3n} . The optimal GMM estimator of ρ_0 is defined as $\hat{\rho}_n = \text{argmin}_{\rho} g_n'(\rho, \tilde{\varsigma}_n) \hat{\Psi}_n^{-1} g_n(\rho, \tilde{\varsigma}_n)$, where $\hat{\Psi}_n$ is an estimator of the variance matrix of the limiting distribution of the normalized sample moment $\sqrt{n}g_n(\rho, \tilde{\varsigma}_n)$.¹⁹ The estimator $\hat{\rho}_n$ is used for the two step GMM estimator of ς_0 , which is based on the linear instrumental matrix Q_{3n} . Let $g_{2n}(\hat{\rho}_n, \varsigma) = \frac{1}{n}Q_{3n}'\varepsilon_n(\hat{\rho}_n, \varsigma)$ be the sample moment function, where $\varepsilon_n(\hat{\rho}_n, \varsigma) = R_n(\hat{\rho}_n)S_n(\lambda)Y_n - R_n(\hat{\rho}_n)X_n\beta$. Then the optimal two-step GMM estimator of ς_0 is defined by $\hat{\varsigma}_n = \text{argmin}_{\varsigma} g_{2n}'(\hat{\rho}_n, \varsigma) \Upsilon_n g_{2n}(\hat{\rho}_n, \varsigma)$, where $\Upsilon_n = (\frac{1}{n}Q_{3n}'Q_{3n})^{-1}$.²⁰

As illustrated, the estimation approach in Kelejian and Prucha (1998), Kelejian and Prucha (2010) and Drukker, Egger, and Prucha (2012) is characterized by a sequential two-step GMM estimation method.²¹ The sequential GMM estimation is motivated by computational simplicity as the ML estimation involves significant computational burden for the large samples. In addition, the Kelejian-Prucha methodology also does not involve the computation of the inverse of the $n \times n$ matrix S_n in the GMM framework. A possible disadvantage of the two-step GMM approach is that the resulting estimators may be inefficient relative to the joint GMM estimator (one step GMM estimator) derived by using the complete set of moment functions with an optimal weight matrix (Lee, 2007b; Lee and Liu, 2010b).²²

4 Estimation Approach under Unknown Heteroskedasticity

In this section, we consider GMM and ML estimation of spatial autoregressive models with heteroskedastic disturbances. In the first subsection, the necessary condition for the consistency of the ML estimator is studied. The results show that the ML estimator of autoregressive parameters is generally inconsistent when heteroskedasticity is not incorporated into estimation. The next subsection covers a robust GMM estimation method for a spatial model with spatial dependence in the dependent variable and in the disturbance term. The results indicate that the robust GMM estimator is consistent and asymptotically normally distributed.

¹⁹For the explicit form of $\hat{\Psi}_n$, see Arraiz et al. (2010). Note that $\tilde{\varsigma}_n$ can be updated by using the weight matrix I_n for an initial first step $\hat{\rho}_n$.

²⁰The estimator $\hat{\varsigma}_n$ has been called the feasible generalized spatial two-stage least squares (FGS2SLS) estimator.

²¹For the description of the estimation steps, see Arraiz et al. (2010) and Drukker, Egger, and Prucha (2012).

²²For a different approach of the GMM estimation method, see Conley (1999).

4.1 The Inconsistency of Maximum Likelihood Estimator

Lin and Lee (2010) show that the MLE is inconsistent for the case of an SARAR(1,0). In this section, we show that the ML estimator is also inconsistent for the spatial model in (2.1) when there is an unknown form of heteroskedasticity in the innovation terms. Let $\zeta = (\theta', \sigma^2)'$ with $\theta = (\rho, \lambda, \beta)'$. The log likelihood of the model in (2.1) under the assumption that disturbances are i.i.d. $N(0, \sigma_0^2)$ is given by

$$\begin{aligned} \ln L_n(\zeta) = & -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) + \ln |S_n(\lambda)| + \ln |R_n(\rho)| \\ & - \frac{1}{2\sigma^2} [S_n(\lambda)Y_n - X_n\beta]' R_n'(\rho) R_n(\rho) [S_n(\lambda)Y_n - X_n\beta]. \end{aligned} \quad (4.1)$$

For notational simplicity, let $R_n(\rho)X_n = \bar{X}_n(\rho)$, $\bar{M}_n(\rho) = (I_n - \mathbb{P}_n(\rho))$ with $\mathbb{P}_n(\rho) = \bar{X}_n(\rho)[\bar{X}_n'(\rho)\bar{X}_n(\rho)]^{-1}\bar{X}_n'(\rho)$, and $\delta = (\rho, \lambda)'$. Note that $\bar{X}_n'(\rho)\bar{M}_n(\rho) = 0_{k \times n}$ and $\bar{M}_n(\rho)\bar{X}_n(\rho) = 0_{n \times k}$. The solution of the first order conditions for β and σ^2 yields the following ML estimators.²³

$$\hat{\beta}_n(\delta) = [\bar{X}_n'(\rho)\bar{X}_n(\rho)]^{-1}\bar{X}_n'(\rho)R_n(\rho)S_n(\lambda)Y_n \quad (4.2a)$$

$$\hat{\sigma}_n^2(\delta) = \frac{1}{n}\varepsilon_n'(\theta)\varepsilon_n(\theta) = Y_n'S_n'(\lambda)R_n'(\rho)\bar{M}_n(\rho)R_n(\rho)S_n(\lambda)Y_n. \quad (4.2b)$$

For a given value of δ , the ML estimators $\hat{\beta}_n(\delta)$ and $\hat{\sigma}_n^2(\delta)$ can be seen as OLS estimators from the regression equation $R_n(\rho)S_n(\lambda)Y_n = R_n(\rho)X_n\beta + \varepsilon_n$. Substitution of $R_n(\rho)S_n(\lambda)Y_n = R_n(\rho)X_n\beta + \varepsilon_n$ into $\hat{\sigma}_n^2(\delta)$ yields $\hat{\sigma}_n^2(\delta) = \frac{1}{n}\varepsilon_n'\bar{M}_n(\rho)\varepsilon_n$.

For the asymptotic argument of this section, we modify Assumption 3 in the following way.

Assumption 3': The exogenous variables matrix X_n is an $n \times k$ matrix consisting of constant elements that are uniformly bounded. It has full column rank k . Moreover, $\lim_{n \rightarrow \infty} \frac{1}{n}X_n'X_n$ and $\lim_{n \rightarrow \infty} \frac{1}{n}X_n'R_n'(\rho)R_n(\rho)X_n$ exist and are nonsingular for all values of ρ in Θ .

The compact parameter space contains ρ_0 by Assumption 5, therefore the modified assumption also requires a finite and nonsingular limit for the term $\frac{1}{n}X_n'R_n'(\rho)R_n(\rho)X_n$. With this new assumption, orders of certain terms can be obtained via the asymptotic analysis given in Appendix B.

At δ_0 , the probability limit of $\hat{\sigma}_n^2(\delta_0)$ is

$$\text{plim}_{n \rightarrow \infty} \hat{\sigma}_n^2(\delta_0) = \text{plim}_{n \rightarrow \infty} \frac{1}{n}\varepsilon_n'\varepsilon_n - \text{plim}_{n \rightarrow \infty} \frac{1}{n^2}\varepsilon_n\bar{X}_n[\frac{1}{n}\bar{X}_n'\bar{X}_n]^{-1}\bar{X}_n'\varepsilon_n. \quad (4.3)$$

In (4.3), the first term on the right hand side converges to $\frac{1}{n} \sum_{i=1}^n \sigma_{ni}^2$ by Chebyshev Weak Law of Large numbers. The second term vanishes in probability so that the average of variances of the disturbance terms is asymptotically equivalent to $\hat{\sigma}_n^2(\delta_0)$, namely, $\hat{\sigma}_n^2(\delta_0) = \frac{1}{n} \sum_{i=1}^n \sigma_{ni}^2 + o_p(1)$.²⁴

²³The first order conditions from (4.1) are given in Appendix B.

²⁴For the asymptotic argument see Appendix B.

The concentrated log-likelihood function is obtained by substituting $\hat{\beta}_n(\delta)$ and $\hat{\sigma}_n^2(\delta)$ into (4.1):

$$\ln L_n(\delta) = -\frac{n}{2}(\ln(2\pi) + 1) - \frac{n}{2} \ln(\hat{\sigma}_n^2(\delta)) + \ln |S_n(\lambda)| + \ln |R_n(\rho)|. \quad (4.4)$$

The MLE $\hat{\delta}_n = (\hat{\lambda}_n, \hat{\rho}_n)'$ is the extremum estimator derived from the concentrated log-likelihood function. The first order conditions of the concentrated log-likelihood function with respect to ρ and λ are given by

$$\frac{\partial \ln L_n(\delta)}{\partial \rho} = -\frac{n}{2\hat{\sigma}_n^2(\delta)} \frac{\partial \hat{\sigma}_n^2(\delta)}{\partial \rho} - \text{tr}(H_n(\rho)), \quad (4.5)$$

$$\frac{\partial \ln L_n(\delta)}{\partial \lambda} = -\frac{n}{2\hat{\sigma}_n^2(\delta)} \frac{\partial \hat{\sigma}_n^2(\delta)}{\partial \lambda} - \text{tr}(G_n(\lambda)), \quad (4.6)$$

where $G_n(\lambda) = W_n S_n^{-1}(\lambda)$ and $H_n(\rho) = M_n R_n^{-1}(\rho)$. The consistency of the MLE $\hat{\delta}_n$ requires that the first order conditions evaluated at the true parameter value δ_0 converges in probability to zero i.e., $\text{plim}_{n \rightarrow \infty} \frac{1}{n} \frac{\partial \ln L_n(\delta_0)}{\partial \delta} = 0$. This necessary condition for the consistency of the ML estimator of δ_0 is

$$\frac{1}{n} \frac{\partial \ln L_n(\delta_0)}{\partial \delta} = \begin{pmatrix} \frac{\frac{1}{n} \sum_{i=1}^n H_{n,ii} \sigma_{ni}^2}{\frac{1}{n} \sum_{i=1}^n \sigma_{ni}^2} - \frac{1}{n} \text{tr}(H_n) + o_p(1) \\ \frac{\frac{1}{n} \sum_{i=1}^n \bar{G}_{n,ii} \sigma_{ni}^2}{\frac{1}{n} \sum_{i=1}^n \sigma_{ni}^2} - \frac{1}{n} \text{tr}(G_n) + o_p(1) \end{pmatrix}. \quad (4.7)$$

Denote $\bar{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \sigma_{ni}^2$, $H_n^* = \frac{1}{n} \text{tr}(H_n) = \frac{1}{n} \sum_{i=1}^n H_{n,ii}$, and $\bar{G}_n^* = \frac{1}{n} \text{tr}(\bar{G}_n) = \frac{1}{n} \sum_{i=1}^n \bar{G}_{n,ii}$ where $\bar{G}_n = R_n G_n R_n^{-1}$. Then, (4.7) can be written in a more convenient form²⁵

$$\begin{aligned} \frac{1}{n} \frac{\partial \ln L_n(\delta_0)}{\partial \delta} &= \begin{pmatrix} \frac{\frac{1}{n} \sum_{i=1}^n [H_{n,ii} - H_n^*] [\sigma_{ni}^2 - \bar{\sigma}^2]}{\bar{\sigma}^2} + o_p(1) \\ \frac{\frac{1}{n} \sum_{i=1}^n [\bar{G}_{n,ii} - \bar{G}_n^*] [\sigma_{ni}^2 - \bar{\sigma}^2]}{\bar{\sigma}^2} - \frac{1}{n} \text{tr}(\bar{G}_n - G_n) + o_p(1) \end{pmatrix} \\ &= \begin{pmatrix} \frac{\text{cov}(H_{n,ii}, \sigma_{ni}^2)}{\bar{\sigma}^2} + o_p(1) \\ \frac{\text{cov}(\bar{G}_{n,ii}, \sigma_{ni}^2)}{\bar{\sigma}^2} + o_p(1) \end{pmatrix}. \end{aligned} \quad (4.8)$$

The above equation shows that the ML estimators $\hat{\lambda}_n$ and $\hat{\rho}_n$ are inconsistent unless $\frac{\text{cov}(H_{n,ii}, \sigma_{ni}^2)}{\bar{\sigma}^2} = 0$ and $\frac{\text{cov}(\bar{G}_{n,ii}, \sigma_{ni}^2)}{\bar{\sigma}^2} = 0$. The inconsistency of $\hat{\lambda}_n$ and $\hat{\rho}_n$ depends on the covariance between variances of elements of the disturbance terms and diagonal elements of H_n and $\bar{G}_n$. It is obvious that when ε_n is homoskedastic, $\frac{1}{n} \frac{\partial \ln L_n(\delta_0)}{\partial \delta}$ is $o_p(1)$ as $\sigma_{ni}^2 = \bar{\sigma}^2$ for $i = 1, \dots, n$. This result also holds for the trivial case of $\rho_0 = \lambda_0 = 0$. Intuitively, the result in (4.8) indicates that the concentrated log-likelihood function is not maximized at the true parameter vector when disturbance terms have

²⁵Note that $\bar{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \sigma_{ni}^2$ is the average of the variance of the disturbance terms, and $\frac{1}{n} \text{tr}(\bar{G}_n - G_n) = 0$.

unknown heteroskedasticity.

The ML estimator of β_0 in (4.2a) is also inconsistent, since it is a function of inconsistent estimators $\hat{\lambda}_n$ and $\hat{\rho}_n$. Explicitly,

$$\begin{aligned}\hat{\beta}_n(\hat{\delta}_n) &= \beta_0 + (\lambda_0 - \hat{\lambda}_n)\mathbb{D}_n(\hat{\rho}_n)\bar{X}'_n\bar{G}_n\bar{X}_n\beta_0 + (\lambda_0 - \hat{\lambda}_n)(\rho_0 - \hat{\rho}_n)\mathbb{D}_n(\hat{\rho}_n)\bar{X}'_nH_n^s\bar{G}_n\bar{X}_n\beta_0 \\ &\quad + (\lambda_0 - \hat{\lambda}_n)(\rho_0 - \hat{\rho}_n)^2\mathbb{D}_n(\hat{\rho}_n)\bar{X}'_nH_n'H_n\bar{G}_n\bar{X}_n\beta_0 + o_p(1)\end{aligned}\quad (4.9)$$

where $\mathbb{D}_n(\hat{\rho}_n) = [\bar{X}'_n(\hat{\rho}_n)\bar{X}_n(\hat{\rho}_n)]^{-1}$.²⁶ The above result shows that the asymptotic bias of $\hat{\beta}_n(\hat{\delta}_n)$ depends on weight matrices and the regressors matrix, and is not zero unless autoregressive parameters are consistent. For the special case of $\hat{\lambda}_n = \lambda_0 + o_p(1)$, the inconsistency of $\hat{\rho}_n$ has no effect on the asymptotic bias of $\hat{\beta}_n(\hat{\delta}_n)$, so that $\hat{\beta}_n(\hat{\delta}_n) = \beta_0 + o_p(1)$.

For the spatial autoregressive model, where $\rho_0 = 0$ in (2.1), the result in the second row of (4.8) simplifies to $\frac{1}{n}\frac{\partial \ln L_n(\lambda_0)}{\partial \lambda} = \frac{\text{cov}(G_{n,ii}, \sigma_{ni}^2)}{\bar{\sigma}^2} + o_p(1)$ since $\bar{G}_n = G_n$. The term $\mathbb{D}_n(\hat{\rho}_n)\bar{X}'_n\bar{G}_n\bar{X}_n\beta_0$ in (4.9) simplifies to $(X'_nX_n)^{-1}X'_nG_nX_n\beta_0$ so that $\hat{\beta}_n(\hat{\lambda}_n) = \beta_0 + (\lambda_0 - \hat{\lambda}_n)(X'_nX_n)^{-1}X'_nG_nX_n\beta_0 + o_p(1)$, which is the exact result stated in Lin and Lee (2010).

The concentrated log-likelihood function is nonlinear in δ , which makes it hard to make any general conclusion about the asymptotic bias of the MLE $\hat{\delta}_n = (\hat{\lambda}_n, \hat{\rho}_n)'$. For the spatial autoregressive model, Lin and Lee (2010) investigate the asymptotic bias of $\hat{\beta}_n(\hat{\lambda}_n)$ for a case of group interactions, where W_n is assumed to be a block-diagonal matrix such that each block has different number of units and each unit is equally affected by the other units. Lin and Lee (2010) shows that when covariates are i.i.d with mean zero for all blocks, the asymptotic bias of the intercept is larger than those of other coefficients, and the bias of all coefficients are negatively related to the average block size.

The specification in (2.1) with $\lambda_0 = 0$ is called the special error model (SEM or SARAR (0,1)) in the literature (LeSage and Pace, 2009). For this model, the necessary condition for the consistency of the ML estimator of ρ_0 is not satisfied, since the result in the first row of (4.8) is generally not zero. The MLE of β_0 for the SEM is given by $\hat{\beta}_n(\rho) = \mathbb{D}_n(\rho)\bar{X}'_n(\rho)R_n(\rho)Y_n$ for a given ρ , which is the OLS estimator from the artificial regression $R_n(\rho)Y_n = R_n(\rho)X_n\beta + \varepsilon_n$. Substituting $Y_n = X_n\beta_0 + R_n^{-1}\varepsilon_n$ into $\hat{\beta}_n(\rho)$ yields $\hat{\beta}_n(\rho) = \beta_0 + \mathbb{D}_n(\rho)\bar{X}'_n(\rho)R_n(\rho)R_n^{-1}\varepsilon_n$. Under *Assumption 3'*, it can be shown that $\hat{\beta}_n(\hat{\rho}_n) = \beta_0 + o_p(1)$.²⁷ This result indicates that under unknown form of heteroskedasticity, the MLE $\hat{\beta}_n(\hat{\rho}_n)$ has no asymptotic bias, even when the MLE $\hat{\rho}_n$ is inconsistent.

The spatial model specification with $\beta_0 = 0$ and $\rho_0 = 0$ in (2.1) is known as the pure spatial autoregressive model in the literature. The MLE estimator of λ_0 for this kind of model is also inconsistent under heteroskedastic disturbances. The first order condition of the concentrated log-likelihood function of this model with respect to λ is $\frac{1}{n}\frac{\partial \ln L_n(\lambda_0)}{\partial \lambda} = \frac{1}{n}\frac{1}{\hat{\sigma}_n^2(\lambda)}Y_n'W_n'S_n(\lambda)Y_n - \frac{1}{n}\text{tr}(G_n(\lambda))$, where $\hat{\sigma}_n^2(\lambda) = \frac{1}{n}\varepsilon_n'(\lambda)\varepsilon_n(\lambda)$ with $\varepsilon_n(\lambda) = S_n(\lambda)Y_n$. At λ_0 , $\hat{\sigma}_n^2(\lambda_0) = \frac{1}{n}\sum_{i=1}^n \sigma_{ni}^2 + o_p(1)$. Then, $\frac{1}{n}\frac{\partial \ln L_n(\lambda_0)}{\partial \lambda} = \frac{\text{cov}(G_{n,ii}, \sigma_{ni}^2)}{\bar{\sigma}^2} + o_p(1)$ by the same asymptotic argument applied in the derivation of (4.8). This result is the same as with the one obtained in Lin and Lee (2010) for the case of an

²⁶For the asymptotic argument, see Appendix B.

²⁷For the asymptotic argument, see Appendix B.

SARAR(1,0).

In the special case, where the spatial weight matrices are the same and the true parameter values λ_0 and ρ_0 are equal, the covariance terms in (4.8) are equal.²⁸ In this special case, the result in (4.8) simplifies to $\frac{1}{n} \frac{\partial \ln L_n(\delta_0)}{\partial \delta} = \frac{\text{cov}(G_{n,ii}, \sigma_{ni}^2)}{\bar{\sigma}^2} + o_p(1)$, which is the necessary condition stated in Lin and Lee (2010) for a spatial model with only a spatial lag in the dependent variable. Despite this result, the asymptotic bias of the MLE $\hat{\beta}_n(\hat{\delta}_n)$ will not simplify to the one derived for a spatial model with a spatial lag in the dependent variable.

A natural question is that under what conditions the covariance terms in (4.8) are zero. An obvious case is when both $\bar{G}_n$ and H_n have diagonal elements that are equal. Then, the necessary condition for the consistency of $\hat{\lambda}_n$ and $\hat{\rho}_n$ is not violated, even if the disturbances are heteroskedastic. As an example, consider a circular world weight matrix with equal diagonal elements that relate each unit to the units in front and in back. In that case, both H_n and $\bar{G}_n$ have equal diagonal elements. Another case arises, when the weight matrices W_n and M_n are block-diagonal matrices with an identical submatrix in the diagonal blocks and zeros elsewhere. This is a special case of group interactions example in Lee (2001a) where all group sizes are equal and each neighbor of the same unit has equal weight.

In this section, we have shown that the ML estimators for autoregressive spatial models are generally inconsistent when heteroskedasticity is present in the disturbance terms. Besides its computational burden, the consistency of ML estimator is not ensured.

4.2 Robust GMM Estimation of SARAR(1,1)

In this section, the robust GMM estimation method suggested by Lin and Lee (2010) is extended for the model in (2.1). We consider the set of population moment functions $g_n(\theta_0) = (\varepsilon_n' P_{1n} \varepsilon_n, \dots, \varepsilon_n' P_{mn} \varepsilon_n, \varepsilon_n' Q_n)'$ where $P_{jn} \in \mathcal{P}_{2n}$ for $j = 1, \dots, m$. This set defines the orthogonality conditions that are considered for the estimation. Throughout this section, we assume that the model in (2.1) satisfies *Assumptions* 1 through *Assumptions* 5. First, we discuss the identification of the parameter vector θ_0 in the GMM framework and state conditions for the identification. Then, we determine the large sample properties of the robust GMM estimator.²⁹

The identification of parameters in a GMM framework requires $\lim_{n \rightarrow \infty} \frac{1}{n} E(g_n(\theta_0)) = 0$.³⁰ For any value of the parameter vector $\theta \in \Theta$, consider the expectation of the set of moment functions

²⁸In this case, $W_n = M_n$ and $R_n = S_n$ so that $H_n = W_n S_n^{-1} = G_n$ and $\bar{G}_n = S_n G_n S_n^{-1} = S_n W_n S_n^{-1} S_n^{-1} = G_n$.

²⁹The arguments provided here is general, and issues about the selection of particular P_{jn} and Q_n are presented in the final part of Section 4.2.

³⁰See *Lemma 2.3* in Newey and McFadden (1994).

in (4.10):

$$E(g_n(\theta)) = \begin{pmatrix} E(\varepsilon_n'(\theta)P_{1n}\varepsilon_n(\theta)) \\ E(\varepsilon_n'(\theta)P_{2n}\varepsilon_n(\theta)) \\ \vdots \\ E(\varepsilon_n'(\theta)P_{mn}\varepsilon_n(\theta)) \\ E(Q_n'\varepsilon_n(\theta)) \end{pmatrix}. \quad (4.10)$$

From (2.1), $\varepsilon_n(\theta)$ can be written in terms of the model parameters in the following way:

$$\varepsilon_n(\theta) = R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}\varepsilon_n + R_n(\rho)[S_n(\lambda)S_n^{-1}X_n\beta_0 - X_n\beta] = K_n(\delta)K_n^{-1}\varepsilon_n + R_n(\rho)k_n(\varsigma), \quad (4.11)$$

where $K_n(\delta) = R_n(\rho)S_n(\lambda)$, $K_n = R_nS_n$, $k_n(\varsigma) = [S_n(\lambda)S_n^{-1}X_n\beta_0 - X_n\beta]$, and $\varsigma = (\lambda, \beta)'$ are introduced for notational simplicity. Substituting (4.11) into (4.10) and taking expectation yield

$$E(g_n(\theta)) = \begin{pmatrix} k_n'(\varsigma)R_n'(\rho)P_{1n}R_n(\rho)k_n(\varsigma) + \text{tr}(\Sigma_n K_n'^{-1}K_n'(\delta)P_{1n}K_n(\delta)K_n^{-1}) \\ k_n'(\varsigma)R_n'(\rho)P_{2n}R_n(\rho)k_n(\varsigma) + \text{tr}(\Sigma_n K_n'^{-1}K_n'(\delta)P_{2n}K_n(\delta)K_n^{-1}) \\ \vdots \\ k_n'(\varsigma)R_n'(\rho)P_{mn}R_n(\rho)k_n(\varsigma) + \text{tr}(\Sigma_n K_n'^{-1}K_n'(\delta)P_{mn}K_n(\delta)K_n^{-1}) \\ Q_n'R_n(\rho)k_n(\varsigma) \end{pmatrix}. \quad (4.12)$$

where $\Sigma_n = \text{Diag}(\sigma_{ni}^2, \dots, \sigma_{nn}^2)$. The identification of the parameter vector θ_0 can be verified from $E(g_n(\theta)) = 0$; i.e., θ_0 is identified if θ_0 is the unique solution of $E(g_n(\theta)) = 0$. The term $Q_n'R_n(\rho)k_n(\varsigma)$ in (4.12) can be written more explicitly as

$$Q_n'R_n(\rho)k_n(\varsigma) = Q_n'R_n(\rho)(X_n, G_nX_n\beta_0) \begin{pmatrix} \beta_0 - \beta \\ \lambda_0 - \lambda \end{pmatrix} = 0. \quad (4.13)$$

The unique solution of (4.13) is (β_0, λ_0) if the matrix, $Q_n'R_n(\rho)(X_n, G_nX_n\beta_0) = (Q_n'R_n(\rho)X_n, Q_n'R_n(\rho)G_nX_n\beta_0)$ has full column rank $k+1$ for each possible value of $\rho \in \Theta$ by the virtue of *Lemma 1* of Appendix A. Since the linear IV matrix Q_n has column rank greater than or equal to $k+1$, this rank condition is equivalent to the fact that the matrix $(X_n, G_nX_n\beta_0)$ has full column rank $k+1$.

Under this rank condition, the remaining moment equations in $E(g_n(\theta))$ are for the identification of ρ_0 . To this end, $K_n(\delta)$ is decomposed in the following way³¹

$$\begin{aligned} K_n(\delta) &= R_n(\rho)S_n(\lambda) = (R_n - (\rho - \rho_0)M_n)(S_n - (\lambda - \lambda_0)W_n) \\ &= K_n + (\rho_0 - \rho)M_nS_n + (\lambda_0 - \lambda)R_nW_n + (\rho_0 - \rho)(\lambda_0 - \lambda)M_nW_n. \end{aligned} \quad (4.14)$$

³¹ $K_n(\delta)$ can be decomposed by using identities $S_n(\lambda) = S_n - (\lambda - \lambda_0)W_n$ and $R_n(\rho) = R_n - (\rho - \rho_0)M_n$.

Consider the moment equation with P_{jn} , $k'_n(\varsigma)R'_n(\rho)R_n(\rho)P_{jn}k_n(\varsigma) + \text{tr}(\Sigma_n K_n'^{-1} K'_n(\delta) P_{jn} K_n(\delta) K_n^{-1})$. Since β_0 and λ_0 are identified from the rank condition of the last moment equation, the first term in the j th moment equation is zero and $K_n(\delta)$ term reduces to $K_n + (\rho_0 - \rho)M_n S_n$. The remaining term in the j th moment equation can be explicitly written as

$$\begin{aligned} \text{tr}(\Sigma_n K_n'^{-1} K'_n(\delta) P_{jn} K_n(\delta) K_n^{-1}) &= (\rho_0 - \rho) \text{tr}(P_{jn}^s M_n R_n^{-1} \Sigma_n) \\ &+ (\rho_0 - \rho)^2 \text{tr}(R_n'^{-1} M_n' P_{jn} M_n R_n^{-1} \Sigma_n) = 0, \end{aligned} \quad (4.15)$$

where $P_{jn}^s = P_{jn} + P_{jn}'$. There are two roots for ρ in (4.15). The first root is the true parameter value ρ_0 , and the second root is

$$\rho = \rho_0 + \frac{\text{tr}(P_{jn}^s M_n R_n^{-1} \Sigma_n)}{\text{tr}(R_n'^{-1} M_n' P_{jn} M_n R_n^{-1} \Sigma_n)}. \quad (4.16)$$

There are three cases in which ρ_0 is the unique root. If $\text{tr}(P_{jn}^s M_n R_n^{-1} \Sigma_n) = 0$ and the denominator is not zero, then ρ_0 is the unique root. If the numerator is not zero but the denominator is zero, then the second root is not defined. In both cases, ρ_0 is uniquely identified. If there is more than one matrix for the quadratic moment equations, then there is another case in which ρ_0 can be uniquely identified. The condition for this case is that the fraction in (4.16) must be different for each P_{jn} for $j = 1, \dots, m$ so that the second root does not exist,

$$\frac{\text{tr}(P_{in}^s M_n R_n^{-1} \Sigma_n)}{\text{tr}(R_n'^{-1} M_n' P_{in} M_n R_n^{-1} \Sigma_n)} \neq \frac{\text{tr}(P_{jn}^s M_n R_n^{-1} \Sigma_n)}{\text{tr}(R_n'^{-1} M_n' P_{jn} M_n R_n^{-1} \Sigma_n)} \quad \text{for all } i \neq j. \quad (4.17)$$

When the rank condition for $Q'_n R_n(\rho) k_n(\varsigma) = 0$ fails then β_0 and λ_0 are not identified separately from the last moment equation in $E(g_n(\theta))$. In this case, the column rank of the matrix $(X_n, G_n X_n \beta_0)$ is less than $k + 1$. This implies that there exists a constant vector v such that $X_n v = G_n X_n \beta_0$. Using this relation in (4.13)

$$\begin{aligned} Q'_n R_n(\rho) k_n(\varsigma) &= Q'_n R_n(\rho) [X_n(\beta - \beta_0) + X_n v(\lambda - \lambda_0)] \\ &= Q'_n R_n(\rho) X_n [(\beta - \beta_0) + v(\lambda - \lambda_0)] = 0. \end{aligned} \quad (4.18)$$

The regressors matrix X_n has full column rank k by *Assumption 3*; therefore, the matrix $Q'_n R_n(\rho) X_n$ in the above equation has full column rank k for each $\rho \in \Theta$. This implies that all solutions of (4.18) satisfies the relation $\beta = \beta_0 - v(\lambda - \lambda_0)$ by virtue of *Lemma 1* Appendix A. This indicates that β_0 and λ_0 are not separately identified from this moment equation and that only once λ_0 is identified the identification of β_0 will be feasible. The remaining moment equations in (4.12) are functions of $\delta = (\rho, \lambda)'$. Hence, these moment functions may provide identification for the parameter vector δ_0 . In this case, these moment equations are simplified to $\text{tr}(\Sigma_n K_n'^{-1} K'_n(\delta) P_{jn} K_n(\delta) K_n^{-1}) = 0$ for $j = 1, \dots, m$ (since, $k_n(\varsigma) = X_n[(\beta - \beta_0) + v(\lambda - \lambda_0)] = 0$ at $\beta = \beta_0 - v(\lambda - \lambda_0)$). Lee (2001b)

makes the observation that these remaining moment equations corresponds to the moment equations of the following process:

$$Y_n = \lambda_0 W_n Y_n + u_n, \quad u_n = \rho_0 M_n u_n + \varepsilon_n. \quad (4.19)$$

For the above process, $\varepsilon_n(\theta) = R_n(\rho) S_n(\lambda) Y_n = R_n(\rho) S_n(\lambda) (S_n^{-1} R_n^{-1} \varepsilon_n) = K_n(\delta) K_n^{-1} \varepsilon_n$. Thus, the expectation of the j th quadratic moment is $E(\varepsilon_n'(\theta) P_{jn} \varepsilon_n(\theta)) = \text{tr}(\Sigma_n K_n'^{-1} K_n'(\delta) P_{jn} K_n(\delta) K_n^{-1})$. Therefore, the identification of δ_0 can be investigated from (4.19). When $M_n = W_n$, the reduced form of (4.19) is $Y_n = (\rho_0 + \lambda_0) W_n Y_n - \rho_0 \lambda_0 W_n^2 Y_n + \varepsilon_n$. The identification of δ_0 is not possible from this process since λ_0 and ρ_0 can not be distinguished from each other (Anselin, 1988, p. 88). Thus, only under the condition that $M_n \neq W_n$, the identification issue can be investigated from the equation $\text{tr}(\Sigma_n K_n'^{-1} K_n'(\delta) P_{jn} K_n(\delta) K_n^{-1}) = 0$. This equation can be explicitly written as

$$\begin{aligned} \text{tr}(\Sigma_n K_n'^{-1} K_n'(\delta) P_{jn} K_n(\delta) K_n^{-1}) &= \text{tr}(\Sigma_n (I_n + (\rho_0 - \rho) M_n S_n K_n^{-1} + (\lambda_0 - \lambda) R_n W_n K_n^{-1} \\ &\quad + (\rho_0 - \rho)(\lambda_0 - \lambda) M_n W_n K_n^{-1})' P_{jn} (I_n + (\rho_0 - \rho) M_n S_n K_n^{-1} \\ &\quad + (\lambda_0 - \lambda) R_n W_n K_n^{-1} + (\rho_0 - \rho)(\lambda_0 - \lambda) M_n W_n K_n^{-1})) = 0. \end{aligned} \quad (4.20)$$

In order to simplify the notation, let us introduce the following variables:

$\alpha_{\rho,j} = \text{tr}(\Sigma_n P_{jn}^s H_n)$, $\alpha_{\lambda,j} = \text{tr}(\Sigma_n P_{jn}^s \bar{G}_n)$, $\alpha_{\rho^2,j} = \text{tr}(\Sigma_n H_n' P_{jn} H_n)$, $\alpha_{\lambda^2,j} = \text{tr}(\Sigma_n \bar{G}_n' P_{jn} \bar{G}_n)$, $\alpha_{\rho\lambda,j} = \text{tr}(\Sigma_n P_{jn}^s H_n \bar{G}_n + \Sigma_n \bar{G}_n' P_{jn}^s H_n)$, $\alpha_{\rho^2\lambda,j} = \text{tr}(\Sigma_n \bar{G}_n' H_n' P_{jn}^s H_n)$, $\alpha_{\rho\lambda^2,j} = \text{tr}(\Sigma_n \bar{G}_n' H_n' P_{jn}^s \bar{G}_n)$ and $\alpha_{\rho^2\lambda^2,j} = \text{tr}(\Sigma_n \bar{G}_n' H_n' P_{jn} H_n \bar{G}_n)$. Using these variables, the equation (4.20) simplifies to

$$\begin{aligned} \text{tr}(\Sigma_n K_n'^{-1} K_n'(\delta) P_{jn} K_n(\delta) K_n^{-1}) &= \alpha_{\rho,j}(\rho_0 - \rho) + \alpha_{\lambda,j}(\lambda_0 - \lambda) + \alpha_{\rho^2,j}(\rho_0 - \rho)^2 + \alpha_{\lambda^2,j}(\lambda_0 - \lambda)^2 \\ &\quad + \alpha_{\rho\lambda,j}(\rho_0 - \rho)(\lambda_0 - \lambda) + \alpha_{\rho^2\lambda,j}(\rho_0 - \rho)^2(\lambda_0 - \lambda) \\ &\quad + \alpha_{\rho\lambda^2,j}(\rho_0 - \rho)(\lambda_0 - \lambda)^2 + \alpha_{\rho^2\lambda^2,j}(\rho_0 - \rho)^2(\lambda_0 - \lambda)^2 \\ &= 0 \quad \text{for } j = 1, \dots, m. \end{aligned} \quad (4.21)$$

The above system of equations can be written in matrix form in the following way

$$\begin{pmatrix} \alpha_{\rho,1} & \alpha_{\lambda,1} & \alpha_{\rho^2,1} & \alpha_{\lambda^2,1} & \alpha_{\rho\lambda,1} & \alpha_{\rho^2\lambda,1} & \alpha_{\rho\lambda^2,1} & \alpha_{\rho^2\lambda^2,1} \\ \alpha_{\rho,2} & \alpha_{\lambda,2} & \alpha_{\rho^2,2} & \alpha_{\lambda^2,2} & \alpha_{\rho\lambda,2} & \alpha_{\rho^2\lambda,2} & \alpha_{\rho\lambda^2,2} & \alpha_{\rho^2\lambda^2,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_{\rho,m} & \alpha_{\lambda,m} & \alpha_{\rho^2,m} & \alpha_{\lambda^2,m} & \alpha_{\rho\lambda,m} & \alpha_{\rho^2\lambda,m} & \alpha_{\rho\lambda^2,m} & \alpha_{\rho^2\lambda^2,m} \end{pmatrix} \times \begin{pmatrix} \rho_0 - \rho \\ \lambda_0 - \lambda \\ (\rho_0 - \rho)^2 \\ (\lambda_0 - \lambda)^2 \\ (\rho_0 - \rho)(\lambda_0 - \lambda) \\ (\rho_0 - \rho)^2(\lambda_0 - \lambda) \\ (\rho_0 - \rho)(\lambda_0 - \lambda)^2 \\ (\rho_0 - \rho)^2(\lambda_0 - \lambda)^2 \end{pmatrix} = 0. \quad (4.22)$$

By *Lemma 1* in Appendix A, the system in (4.22) has a unique solution at δ_0 if columns of the above matrix do not have a linear combination with nonlinear non-zero constant coefficients of the form

$$\alpha_\rho c_1 + \alpha_\lambda c_2 + \alpha_{\rho^2} c_1^2 + \alpha_{\lambda^2} c_2^2 + \alpha_{\rho\lambda} c_1 c_2 + \alpha_{\rho^2\lambda} c_1^2 c_2 + \alpha_{\rho\lambda^2} c_1 c_2^2 + \alpha_{\rho^2\lambda^2} c_1^2 c_2^2 = 0, \quad (4.23)$$

where α s represent the column vectors of the above matrix and c_1 and c_2 are arbitrary nonzero constant coefficients. With this condition, ρ_0 and λ_0 are uniquely identified from the system in (4.22). Once λ_0 is identified, the identification of β_0 follows from the last moment function in (4.12).

Assumption 6 summarizes conditions for the identification of the parameter vector θ_0 from the set of moment functions in $g_n(\theta)$ for sufficient large n . The similarity of this assumption with *Assumption 5* in Liu, Lee, and Bollinger (2010) is revealing: the main difference is that the identification conditions now involve covariance matrix Σ_n .³²

Assumption 6: For the identification of the parameter vector $\theta_0 \in \Theta$, one of the following cases is assumed.

Case(1):

- (i) The limiting matrix $\lim_{n \rightarrow \infty} \frac{1}{n} Q_n' R_n(X_n, G_n X_n \beta_0)$ has full column rank $k + 1$ for each $\rho \in \Theta$,
- (ii) The limiting value $\lim_{n \rightarrow \infty} \frac{1}{n} \text{tr}(P_{jn}^s H_n \Sigma_n) \neq 0$ for some j , and the limiting vector $[\lim_{n \rightarrow \infty} \frac{1}{n} \text{tr}(P_{1n}^s H_n \Sigma_n), \dots, \lim_{n \rightarrow \infty} \frac{1}{n} \text{tr}(P_{mn}^s H_n \Sigma_n)]'$ is linearly independent of the limiting vector

$$[\lim_{n \rightarrow \infty} \frac{1}{n} \text{tr}(H_n' P_{1n} H_n \Sigma_n), \dots, \lim_{n \rightarrow \infty} \frac{1}{n} \text{tr}(H_n' P_{mn} H_n \Sigma_n)]'.$$

Case(2):

- (i) The limiting matrix $\lim_{n \rightarrow \infty} \frac{1}{n} Q_n' R_n X_n$ has full column rank k for each $\rho \in \Theta$,
- (ii) $W_n \neq M_n$,
- (iii) The vector α s defined above do not have a linear combination with some nonlinear non-zero constant coefficients c_1 and c_2 in the form of $\alpha_\rho c_1 + \alpha_\lambda c_2 + \alpha_{\rho^2} c_1^2 + \alpha_{\lambda^2} c_2^2 + \alpha_{\rho\lambda} c_1 c_2 + \alpha_{\rho^2\lambda} c_1^2 c_2 + \alpha_{\rho\lambda^2} c_1 c_2^2 + \alpha_{\rho^2\lambda^2} c_1^2 c_2^2 = 0$.

The first condition in *Case(1)* ensures the identification of β_0 and λ_0 from the linear moment function. The second condition in *Case(1)* provides the identification for ρ_0 from the quadratic moment functions. In *Case(2)*, the quadratic moment functions ensures the identification of ρ_0 and λ_0 under the condition of $W_n \neq M_n$. Once λ_0 is identified, the identification of β_0 follows from the first condition in *Case(2)*.

³²See also the identification assumptions in Lee and Liu (2010b) and Lin and Lee (2010), which have a similar structure.

Let $\Omega_n = E[g_n(\theta_0)g_n'(\theta_0)]$. By using *Lemma 2* in Appendix A, the variance covariance matrix of the set of moment functions can be obtained.

$$\Omega_n = \begin{pmatrix} \text{tr}(\Sigma_n P_{1n}(P'_{1n}\Sigma_n + \Sigma_n P_{1n})) & \cdots & \text{tr}(\Sigma_n P_{1n}(P'_{mn}\Sigma_n + \Sigma_n P_{mn})) & 0 \\ \text{tr}(\Sigma_n P_{2n}(P'_{1n}\Sigma_n + \Sigma_n P_{1n})) & \cdots & \text{tr}(\Sigma_n P_{2n}(P'_{mn}\Sigma_n + \Sigma_n P_{mn})) & 0 \\ \vdots & & \vdots & \vdots \\ \text{tr}(\Sigma_n P_{mn}(P'_{1n}\Sigma_n + \Sigma_n P_{1n})) & \cdots & \text{tr}(\Sigma_n P_{mn}(P'_{mn}\Sigma_n + \Sigma_n P_{mn})) & 0 \\ 0 & \cdots & 0 & Q'_n \Sigma_n Q_n \end{pmatrix}. \quad (4.24)$$

The variance-covariance matrix Ω_n has the same structure as the one in Lin and Lee (2010). Let $\Gamma_n = -E\left(\frac{\partial g_n(\theta_0)}{\partial \theta'}\right)$. A straightforward application of matrix calculus yields (4.25). Elements of Γ_n are functions of matrices that are uniformly bounded in absolute value in row and column sums so that the order of elements is either $O(n)$ or $O(1)$, which in turn implies that $\frac{1}{n}\Gamma_n$ is bounded.

$$\Gamma_n = \begin{pmatrix} \text{tr}(\Sigma_n H'_n P_{1n}^s) & \text{tr}(\Sigma_n \bar{G}'_n P_{1n}^s) & 0 \\ \text{tr}(\Sigma_n H'_n P_{2n}^s) & \text{tr}(\Sigma_n \bar{G}'_n P_{2n}^s) & 0 \\ \vdots & \vdots & \vdots \\ \text{tr}(\Sigma_n H'_n P_{mn}^s) & \text{tr}(\Sigma_n \bar{G}'_n P_{mn}^s) & 0 \\ 0 & Q'_n R_n G_n X_n \beta_0 & Q'_n R_n X_n \end{pmatrix}. \quad (4.25)$$

Let $\Psi'_n \Psi_n$ be an arbitrary non-stochastic weighting matrix for the GMM objective function. The weighting matrix plays the role of a metric by which the sample moment functions are made as close as possible to zero. Assume that Ψ_n converges to a constant matrix Ψ_0 that has full rank, and $\lim_{n \rightarrow \infty} \frac{1}{n} \Psi_n \Gamma_n$ exists and has full rank (Hansen, 1982).³³ The following proposition shows that the generic GMM estimator based on the set of moment functions $g_n(\theta_0) = (\varepsilon'_n P_{1n} \varepsilon_n, \dots, \varepsilon'_n P_{mn} \varepsilon_n, \varepsilon'_n Q_n)'$ with general P_{jn} s and Q_n is consistent and has an asymptotic normal distribution.³⁴

Proposition 1. *Suppose $P_{jn} \in \mathcal{P}_{2n}$ for $j = 1, \dots, m$ and Q_n is linear IV matrix. Under Assumptions 1-6, the estimator $\hat{\theta}_n$ derived from the objective function $\min_{\theta \in \Theta} g'_n(\theta) \Psi'_n \Psi_n g_n(\theta)$ is a consistent robust GMM estimator (RGMME) of θ_0 . It has an asymptotic normal distribution, namely*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \Upsilon), \quad (4.26)$$

where $\Upsilon = \lim_{n \rightarrow \infty} [\frac{1}{n} \Gamma'_n \Psi'_n \Psi_n \frac{1}{n} \Gamma_n]^{-1} \frac{1}{n} \Gamma'_n \Psi'_n \Psi_n \frac{1}{n} \Omega_n \Psi'_n \Psi_n \frac{1}{n} \Gamma_n [\frac{1}{n} \Gamma'_n \Psi'_n \Psi_n \frac{1}{n} \Gamma_n]^{-1}$.

³³For our case, matrices $\{\Psi_n\}$ have dimensions equal or bigger than $k+2$. Let Q_n be $n \times (k+1)$ linear IV matrix. In that case, $g : \mathbb{R}^n \times \mathbb{R}^{k+2} \rightarrow \mathbb{R}^{m+k+1}$, where $(m+k+1)$ is the number of orthogonality conditions. Then, matrices $\{\Psi_n\}$ are dimensioned $(m+k+1) \times (m+k+1)$. See Assumption 2.5 in Hansen (1982, p.1033).

³⁴The details of the proofs for all propositions are given in Appendix C.

The estimator in *Proposition 1* is the generic GMM estimator considered in Hansen (1982).³⁵ The variance-covariance matrix of the RGMME of *Proposition 1* is a function of unknown terms Γ_n and Ω_n . As usual, consistent estimates of these terms can be obtained from an initial consistent estimator of θ_0 . In the following proposition, consistent estimators for Γ_n and Ω_n are given.

Proposition 2. *Let $\hat{\varepsilon}_{ni}$ be the residual of the model based on consistent initial estimates of θ_0 and denote $\hat{\Sigma}_n = \text{Diag}(\hat{\varepsilon}_{n1}^2, \hat{\varepsilon}_{n2}^2, \dots, \hat{\varepsilon}_{nn}^2)$. Then, under the assumed regularity conditions,*

$$(1) \quad \frac{1}{n}\hat{\Omega}_n - \frac{1}{n}\Omega_n = o_p(1),$$

$$(2) \quad \frac{1}{n}\hat{\Gamma}_n - \frac{1}{n}\Gamma_n = o_p(1).$$

The proof of *Proposition 2* utilizes the facts that quadratic moment matrices are uniformly bounded in absolute value in row and column sums and disturbance terms have uniformly bounded fourth moments. These two properties ensure that the elements involving the trace operator in $\frac{1}{n}\hat{\Omega}_n$ and $\frac{1}{n}\hat{\Gamma}_n$ converge in probability to the corresponding elements of $\frac{1}{n}\Omega_n$ and $\frac{1}{n}\Gamma_n$. The remaining element in $\frac{1}{n}\hat{\Omega}_n$ is $\frac{1}{n}Q_n'\hat{\Sigma}_nQ_n$. The asymptotic argument for this term is in line with that of White (1980). Under certain regularity conditions, White (1980) shows that $\frac{1}{n}X_n'\hat{\Sigma}_nX_n$ converges almost surely to $\frac{1}{n}X_n'\Sigma_nX_n$, where $\hat{\varepsilon}_{ni}$ is a consistent estimate of ε_{ni} .

In *Proposition 1*, the GMM estimator is derived from the objective function with an arbitrary weighting matrix. It is clear that different choices of weighting matrices give rise to GMM estimators with different asymptotic covariance matrices. The optimal estimator is the one that has an asymptotic covariance matrix at least as small as that of any other GMM estimator. Hansen (1982) shows that the optimal GMM estimator is based on the weighting matrix $\Psi_n'\Psi_n = \Omega_n^{-1}$. This matrix plays a prominent role for the optimal GMM estimator under the following regularity condition.

Assumption 7: *The limiting matrix $\lim_{n \rightarrow \infty} \frac{1}{n}\Omega_n$ exists and is nonsingular.*

In (4.24), notice that the terms in Ω_n are functions of matrices that are uniformly bounded in absolute value in row and column sums. For example, a generic term is $\text{tr}(\Sigma_n P_{in}(P_{jn}'\Sigma_n + \Sigma_n P_{jn}))$ which has an order of $O(n)$. Therefore, $\frac{1}{n}\Omega_n$ is order of $O(1)$ which implies that $\frac{1}{n}\Omega_n$ is bounded. *Proposition 2* yields a consistent estimator $\hat{\Omega}_n$ for this optimal weighting matrix. The next proposition shows that the optimal GMM estimator based on the weighting matrix $\hat{\Omega}_n$ is consistent and asymptotically normal.

Proposition 3. *Under Proposition 2 and Assumption 1-7, the optimal robust GMM estimator*

³⁵The structure of the asymptotic variance-covariance matrix Υ is also generic and it has the same structure with the one given in Theorem 3.1 of Hansen (1982, see p. 1042).

derived from $\min_{\theta \in \Theta} g_n'(\theta) \hat{\Omega}_n^{-1} g_n(\theta)$ has the asymptotic distribution given by

$$\sqrt{n}(\hat{\theta}_{o,n} - \theta_0) \xrightarrow{d} N(0, (\lim_{n \rightarrow \infty} \frac{1}{n} \Gamma_n' \Omega_n^{-1} \Gamma_n)^{-1}). \quad (4.27)$$

An estimator of the asymptotic variance covariance matrix of $\sqrt{n}(\hat{\theta}_{o,n} - \theta_0)$ is needed to make asymptotically valid inferences and construct asymptotically correct confidence regions. *Proposition 2* guarantees that the consistent estimator for the asymptotic variance covariance matrix in *Proposition 3* is $(\frac{1}{n} \hat{\Gamma}_n' \hat{\Omega}_n^{-1} \hat{\Gamma}_n)^{-1}$, where Γ_n and Ω_n are evaluated at $\hat{\theta}_{o,n}$.

The remaining issue is about the selection of Q_n and the selection of the possible best P_{jns} from the class $\mathcal{P}_{2n}$. The asymptotic variance-covariance matrix of the GMM estimator depends on P_{jns} and Q_n . By using the generalized Schwartz inequality, Lee (2001b) show that the best selection of P_{jns} from the class $\mathcal{P}_{2n}$ are given by (i) $(H_n - \text{Diag}(H_n))$ and (ii) $(\bar{G}_n - \text{Diag}(\bar{G}_n))$. With a similar argument, Lee (2003) shows that the best IV matrix is $Q_n = (R_n G_n X_n \beta_0, R_n X_n)$. However, the arguments given in Lee (2001b, 2003) are based on the assumption that the disturbance terms are i.i.d. In case of unknown heteroskedasticity, the application of the generalized Schwartz inequality to the variance covariance matrix in the equation (4.27) might not provide the best selection of P_{jns} and Q_n , since it involves unknown matrix Σ_n . Hence, Lin and Lee (2010) state that the consistently estimated P_{jns} and Q_n for the i.i.d. disturbances case may still be desirable. Therefore, the optimal robust GMM estimator in *Proposition 3* is considered with consistently estimated quadratic moment matrices: (i) $(H_n - \text{Diag}(H_n))$ and (ii) $(\bar{G}_n - \text{Diag}(\bar{G}_n))$ and linear IV matrices $Q_n = (R_n G_n X_n \beta_0, R_n X_n)$. An initial consistent estimator of θ_0 can be obtained from an initial GMM estimation with quadratic moment matrices (i) $M_n' M_n - \text{Diag}(M_n' M_n)$, (ii) $W_n M_n - \text{Diag}(W_n M_n)$, and linear moment matrix $Q_n = (M_n W_n X_n, W_n X_n, X_n)$.³⁶

5 Monte Carlo Experiments

5.1 Design

In order to study the finite sample properties of various robust and non-robust estimators, we design several Monte Carlo experiments. The specifications that are used to generate 1000 replications of each Monte Carlo experiment are described below. For three different values of the sample size, n : 100, 500, and 1,000, the data generating process for the model is:

$$Y_n = \lambda_0 W_n Y_n + X_n \beta_0 + u_n, \quad u_n = \rho_0 M_n u_n + \varepsilon_n. \quad (5.1)$$

There are three regressors and no intercept term such that $X_n = [x_{n,1}, x_{n,2}, x_{n,3}]$ and $\beta_0 = (\beta_{10}, \beta_{20}, \beta_{30})'$, where $x_{n,1}$, $x_{n,2}$, and $x_{n,3}$ are $n \times 1$ independent random vectors that are generated from a Normal(0,1). We let $W_n = M_n$ and set $\beta_{10} = 0.7$, $\beta_{20} = 0.4$ and $\beta_{30} = 1.2$ for all experiments. For the spatial autoregressive parameters (λ_0, ρ_0) , we employ combinations of

³⁶For other candidates, see Section 5.

$\mathcal{D} = (-0.8, -0.3, 0, 0.3, 0.8)$ to allow for weak and strong spatial interactions. We consider two specifications for the innovation vector ε_n . To generate the heteroskedastic errors, we follow Lin and Lee (2010) and consider small group interactions structure for the spatial weight matrix (block diagonal weight matrix). For each sample size n , we generate random groups where the size of each group is drawn from Uniform(3,20) distribution.³⁷ For each group, if the group size is greater than 10, we set the variance equal to the group size; otherwise we set the variance to the square of the inverse of the group size. Lin and Lee (2010) also consider creating heteroskedastic errors by simply letting them equal to inverse of the group sizes. We do not consider the latter case in our experiments. This small group interaction scenario is similar to the one in the Monte Carlo design of Kelejian and Prucha (2007), where they focus on a circular world in which the first and the last one third of a sample observations have 5 neighbors in front and 5 in back, while the middle third only has 1 neighbor in front and 1 in back. Figure 1 illustrates weight matrices and variance processes for a sample of $n = 100$. As figure shows, Lin and Lee (2010) small group interactions set-up yields a richer design for heteroskedasticity. We let the i -th element of the innovation vector ε_n be

$$\varepsilon_{ni} = \sigma_{ni}\xi_{ni}, \quad (5.2)$$

where σ_{ni} is the standard error for the i -th observation and ξ_{ni} 's are i.i.d. Normal(0,1).

Also, in order to evaluate each estimator's relative performance under heteroskedasticity, we consider a corresponding homoskedastic case in which disturbances ε_{ni} s are i.i.d. Normal(0, σ_0^2), where $\sigma_0^2 = 1$.

In all experiments, the following estimators are considered: (1) Gaussian maximum likelihood estimator (MLE), (2) Generalized spatial two-stage least squares estimator (GS2SLSE) in Kelejian and Prucha (1998)³⁸, (3) Best two-stage least squares estimator (B2SLS) in Lee (2003) with IV set $Q_n = (\hat{R}_n \hat{G}_n X_n \hat{\beta}_n \hat{R}_n X_n)$ based on initial estimates from the GS2SLSE, and ρ_0 is estimated by the MOM in Kelejian and Prucha (1998), (4) Best generalized method of moments estimator (BGMME) in Liu, Lee, and Bollinger (2010) based on initial estimates from the GS2SLSE, (5) Robust generalized spatial two-stage least squares estimator (RGS2SLSE) in Kelejian and Prucha (2010)³⁹, (6) Robust best two-stage least squares estimator (BR2SLSE) in Lee (2003) with IV set $Q_n = (\hat{R}_n \hat{G}_n X_n \hat{\beta}_n \hat{R}_n X_n)$ based on initial estimates from the RGS2SLSE, and ρ_0 is estimated by the GMME in Kelejian and Prucha (2010), (7) Robust generalized method of moment estimator of *Proposition 3* (RGMME₁) based on the initial estimates from the RGS2SLSE, and (8) Robust generalized method of moment estimator of *Proposition 3* (RGMME₂) based on the initial estimates from the BR2SLSE.⁴⁰

³⁷The weight matrices are row normalized.

³⁸In Kelejian and Prucha (1998), ρ_0 is estimated by the method of moment, and $\delta_0 = (\beta'_0, \lambda_0)'$ by the GS2SLSE. For short, we call both estimators simply as the GS2SLSE.

³⁹In Kelejian and Prucha (2010), ρ_0 is estimated by the GMM method and $\delta_0 = (\beta'_0, \lambda_0)'$ by the GS2SLSE. For short, we call both estimators simply as the RGS2SLSE.

⁴⁰Matlab routines for estimation are available on request.

Figure 1: Weight Matrices

(a) Small Group Interactions Scenario

(b) Circular World

(c) Variance processes

5.2 Simulation Results

In Tables 1–5, the empirical mean (mean), the bias (Bias), the empirical standard error (SD) and the root mean square error (RMSE) are reported for the estimates of each parameter. We do not present the results for all 8 estimators and all 25 combinations of (λ_0, ρ_0) due to space limitation.⁴¹ The focus will be on the MLE, the RGMME₁, and the RGMME₂. In each table for these three estimators, the results for the homoskedastic and heteroskedastic disturbances are presented next to each other for easy comparison. Before we evaluate the results in each table, a couple of general points need to be stated. First, as Arraiz et al. (2010) and Das, Kelejian, and Prucha (2003) point out, if λ_0 is large in absolute value, it results in larger variances of the elements of the disturbance vector, which deteriorates the estimation precision. Yet, at the same time, the variation in the explanatory variable $W_n Y_n$ is also larger, which tends to improve the estimation precision. The net of these opposing effects determines the estimation precision. A similar argument applies to the magnitude (in absolute value) of ρ_0 . Second, as expected regardless of variance structure of the disturbances, all estimators improve in terms of corresponding bias, SD, and RMSE as the sample size increases. Third, for all sample sizes and non-zero combinations of (λ_0, ρ_0) the MLE under heteroskedasticity is inconsistent and impose severe bias on all parameters. Fourth, the results for the sample size of 100 for all estimators ought to be interpreted with caution. We need to emphasize the fact that sample size of 100 is intentionally chosen to observe the behavior of the estimators when the sample size is extremely small.⁴²

Table 1 presents the results for the specification which employs a strong spatial dependence in the dependent variable and a weak spatial dependence in the disturbances. For $N=100$, the MLE performs poorly even when the disturbances are homoskedastic. It imposes significant bias on all parameters with much higher SDs, thus with much higher RMSEs. On the other hand, both RGMMEs impose quite smaller bias on both λ_0 and ρ_0 relative to MLE, and almost no bias on β_0 under both homoskedasticity and heteroskedasticity. As the sample size increases, the RGMMEs improve faster relative to the MLE in terms of the bias and the SD. For $N=500$ and $N=1000$, the RGMMEs under heteroskedasticity impose trivial bias on all parameters. The MLE imposes trivial bias on all parameters under homoskedasticity only for $N=1000$. Although the negative bias on both λ and ρ is less now, it is still significant under heteroskedasticity. Also, the RGMMEs are as efficient as the MLE for $N=1000$ under homoskedasticity. Table 2 presents the results for the specification which employs a weak spatial dependence both in the dependent variable and the disturbances. We have similar findings in terms of biases, but all estimators are more precise compared to Table 1, which confirms the first general point stated above. Under heteroskedasticity, the MLE imposes significant bias on both λ_0 and ρ_0 regardless of the sample size but surprisingly not on β_0 for $N=500$ and $N=1000$. As expected, RGMMEs perform better under heteroskedasticity for all samples sizes. For $N=500$ and $N=1000$ under homoskedasticity, all estimators impose small trivial bias on both

⁴¹The experiments that are not presented here are available by request.

⁴²Lin and Lee (2010) employ number of groups of 100 and 200, where the size of each group is drawn from $U(3, 20)$ and rounded to the closest integer. This set up yields two intervals from which the sample size is drawn: [300,2000] and [600,4000]. Arraiz et al. (2010) choose sample sizes of 486 and 945.

λ_0 and ρ_0 and the RGMMEs are as efficient as the MLE.

Table 3 presents the results for the specification in which there is no special dependence in both the dependent variable and the disturbances. For this case, our large sample result for the MLE suggests that the necessary condition for the consistency of autoregressive parameters is not violated. That is, the case of $\rho_0 = \lambda_0 = 0$ implies $\frac{1}{n} \frac{\partial \ln L_n(\delta_0)}{\partial \delta} = o_p(1)$ even if the innovation terms are heteroskedastic. These observation also suggest that the MLE of the parameters of the exogenous variable is also consistent. For $N = 100$, the MLE has significant biases, but as N increases both the biases and the RMSEs decreases significantly. This pattern is consistent with the aforementioned large sample results. For $N=1000$ under heteroskedasticity, the MLE imposes trivial bias on all parameters and is as precise as the RGMMEs except for ρ_0 . The RGMMEs perform as good as the MLE under homoskedasticity for all samples sizes.

Table 4 presents the results for the specification which employs a weak spatial dependence in the dependent variable and a strong special dependence in the disturbances. For $N=100$, the RGMMEs relatively perform better than the MLE especially under heteroskedasticity as expected. The MLE with heteroskedastic disturbances seems to be inconsistent and imposes significant bias on all parameters with larger SDs and RMSEs, and it does not improve in the larger samples. Under homoskedasticity, all estimators impose trivial biases in the larger samples and the RGMMEs are as efficient as the MLE.

Table 5 presents the results for the specification which employs a strong spatial dependence in the dependent variable and a weak special dependence in the disturbance terms. For all sample sizes under homoskedasticity, all estimators result in trivial bias on all parameters. Under heteroskedasticity, the MLE imposes significant bias on all parameters especially on ρ_0 for $N=100$. However, as the sample size increases the MLE improves and surprisingly imposes a small bias on all parameters. The RGMMEs are robust to heteroskedasticity and do not seem to be affected by the sample size. Under homoskedasticity, the RGMMEs are again as efficient as the MLE.

Overall, significant biases and high RMSEs for the MLE in the heteroskedastic case are suggesting inconsistency. The results in Table 1-5 indicate that the relative size of the bias and the RMSE for the MLE depends on the true values of autoregressive parameters. Since the concentrated log-likelihood function is nonlinear in these parameter, it is hard to make any general conclusion about the asymptotic biases of these parameters. However, as seen in Table 1-5, when the true values of the autoregressive parameters are large, the MLE imposes larger biases in general. As a result, in Table 1 and Table 5, the biases and RMSEs of the MLE of λ_0 is higher than that of the MLE of ρ_0 , and this situation is in the reverse direction in Table 4. The results in Table 1-5 indicates that the biases of the MLE of β_{10} , β_{20} and β_{30} is higher in general when the biases of the autoregressive parameters are higher under heteroskedastic cases.

We compare the performance of the RGMME₁ with other estimators suggested in the literature. We only present the estimation results for $N = 500$ and $N = 1000$ and for only two combinations of (λ_0, ρ_0) due to space limitation. Table 6 and 7 present the estimation results for both homoskedastic and heteroskedastic cases. For the homoskedastic case, all estimators perform better and are almost

unbiased. The estimation results for the GS2SLSE and the B2SLSE indicates small biases for all parameters under homoskedasticity in both tables. In particular, the RMSEs of the GS2SLSE and the B2SLSE for λ_0 , β_{10} , β_{20} and β_{30} are almost identical, which suggest that the set of linear instruments suggested in Kelejian and Prucha (1998) provides a reasonable approximation to the optimal instruments. The same pattern repeats itself in the comparison of the RGS2SLSE and the RB2SLSE for the biases and RMSEs under homoskedasticity. This results confirm the conclusion that the efficiency gain based on the set of the optimal instruments in Lee (2003) is limited (Das, Kelejian, and Prucha, 2003).

As expected, the GS2SLSE and the B2SLSE impose significant biases on ρ_0 under heteroskedastic cases. As stated, the GS2SLSE of ρ_0 is inconsistent as the orthogonality conditions of moments in Kelejian and Prucha (1998) do not hold in the presence of heteroskedastic disturbances. The results in Table 6 and 7 for the GS2SLSE and the B2SLSE of ρ_0 are consistent with this asymptotic argument. The estimation approach in Kelejian and Prucha (1998) is a two-step GMM estimation method as a result of which the inconsistency of the estimator of ρ_0 affects the estimates of other parameters.

Overall, the estimation results in Table 6 and 7 indicates that the GS2SLSE, the MLE, the B2SLSE and the BGMME are inconsistent under heteroskedasticity.

Finally, the performance of the RGS2LSE and the RB2SLSE is compared with that of the RGMME₁ under heteroskedasticity. In Table 6, both RGS2LSE and RB2SLSE impose significant biases on the autoregressive parameters. In Table 7, both RGS2LSE and RB2SLSE impose small biases on the autoregressive parameters and they are as efficient as RGMME₁.

In general, our Monte Carlo results are consistent with our analytical large-sample results, namely that the RGMME of *Proposition 3* is consistent, and the ML estimator of autoregressive parameters is generally inconsistent when there is heteroskedasticity in the model.

6 Conclusion

Heteroskedasticity of unknown form has important consequences for the estimation of spatial econometric models. Asymptotic properties of estimators for spatial models are significantly affected in the presence of unknown heteroskedasticity. Therefore, heteroskedasticity should be accounted in the design of any estimation approach.

If heteroskedasticity is not accounted in estimation, the ML estimator for spatial autoregressive models (including the SAR model, the Kelejian-Prucha model, and the SEM) is generally inconsistent. We show that the probability limit of derivative of the concentrated log-likelihood function evaluated at the true parameter vector is generally not zero for the spatial autoregressive models—the concentrated log-likelihood function is not maximized at the true parameter vector. This necessary condition for the consistency of the ML estimator of the autoregressive parameters depends on the structure of the spatial weight matrices. We also show that the ML estimator of the parameters of the exogenous variable in the SAR and the Kelejian-Prucha model is inconsistent,

and we state the expressions of the corresponding asymptotic biases. For the SEM, we show that the MLE of the parameters of the exogenous variable is consistent, despite the fact that the MLE of the autoregressive parameter of the spatial lag of the disturbance terms is inconsistent. Thus, besides its computational burden, the consistency of ML estimators for the autoregressive spatial models is not ensured in the presence of unknown form of heteroskedasticity.

In GMM estimation framework, heteroskedasticity of unknown form can be incorporated into estimation through formation of the moment functions. We extend the robust GMM approach in Lin and Lee (2010) for the spatial model that has a spatial lag both in the dependent variable and the disturbance term. For the GMM estimator, the quadratic moment matrices are constructed from the spatial weight matrices in the way that the orthogonality conditions of the quadratic moment functions are not violated under the unknown form of heteroskedasticity. These quadratic moment functions are combined with the linear moment function for the GMM estimation. In particular, we show that the robust GMME is consistent and has a properly centered asymptotic normal distribution.

The small sample properties of the RGMM estimator along with the ML and other estimators are studied. In general, our Monte Carlo results are consistent with our analytical large-sample results, namely that the RGMME of *Proposition 3* is consistent, and the ML estimator of autoregressive parameters of spatial models is generally inconsistent when there is unknown form of heteroskedasticity in the model. The RGMME of *Proposition 3* has desirable finite sample properties under both cases of heteroskedasticity and homoskedasticity.

As Monte Carlo experiments clearly indicate, researchers ought to be careful in interpreting the estimation results if the sample size is smaller than 500 and heteroskedastic errors might be present. The MLE clearly performs very poorly under these circumstances regardless of heteroskedasticity. It is quite convenient for researchers to estimate spatial econometric models with spatial dependence both in the dependent variable and the disturbance term via ML method, as the spatial econometrics toolbox by James LeSage provides the routine. However, in our opinion a more rigorous approach is (i) to estimate the model with the RGMMEs for the sample sizes less than 500, and (ii) to estimate the model with the RGMMEs along with the MLE for the larger samples to compare the parameter estimates.

A Some Useful Lemmas

Lemma 1. *Let $A \in \mathbb{R}^{m \times n}$ be matrix of coefficients and $x \in \mathbb{R}^n$ be vector of unknowns. Consider the homogeneous equation $Ax = 0$. Then,*

- (1) *There always exists a solution of $Ax = 0$.*
- (2) *If $\text{rank}(A) < n$, then infinitely many solutions exist.*
- (3) *$Ax = 0$ has only the trivial solution $x = 0$ if and only if $\text{rank}(A) = n$.*

Proof. (1) Obviously, $x = 0$ satisfies the equation. This solution is the trivial one.

(2) First, we will show that there exist a non-trivial solution when $\text{rank}(A) < n$. Let $a_1, a_2, \dots, a_n$ be the column vectors of A . If $\text{rank}(A) < n$, then columns of A are linearly dependent. There exist real number $x_1, x_2, \dots, x_n$ (not all zero) such that $a_1x_1 + a_2x_2 + \dots + a_nx_n = 0$. This implies that $x = (x_1, x_2, \dots, x_n)'$ satisfies $Ax = 0$. Therefore there exist a nontrivial solution x . Secondly, we will show that there exist infinitely many solution when $\text{rank}(A) < n$. Let $x \neq 0$ be the non-trivial solution, then cx is also a solution for any $c \in \mathbb{R}$.

(3) Assume that $\text{rank}(A) = n$. Consider the column and the null space (or kernel) of A . In set notation, $\text{col}(A) = \{y \in \mathbb{R}^m : y = Ax \text{ for some } x \in \mathbb{R}^n\}$ and $\text{ker}(A) = \{x \in \mathbb{R}^n : Ax = 0\}$. It can be shown that the sum of the dimension of $\text{col}(A)$ and the dimension of $\text{ker}(A)$ is n .⁴³ That is, $\dim(\text{col}(A)) + \dim(\text{ker}(A)) = n$. Since $\dim(\text{col}(A)) = \text{rank}(A)$, the dimension of the kernel is $n - \text{rank}(A)$. When $\text{rank}(A) = n$, the dimension of the kernel is zero. This means that there are no linearly independent vectors satisfying $Ax = 0$. Thus, $x = 0$ is the only element of $\text{ker}(A)$. On the other hand, if $x = 0$ is the only solution of $Ax = 0$, then the columns of A are linearly independent, which implies $\text{rank}(A) = n$. $\square$

Lemma 2. *Let A_n , B_n and C_n be $n \times n$ matrices with ij th elements respectively denoted by $a_{n,ij}$, $b_{n,ij}$ and $c_{n,ij}$. 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*

⁴³For a proof see Exercise 4.4 in Abadir and Magnus (2005, page 77).

by $\Sigma_n = \text{Diag}\{\sigma_{n1}^2, \dots, \sigma_{nn}^2\}$. Then,

$$\begin{aligned}
(1) \quad E(\varepsilon_n' A_n \varepsilon_n \cdot \varepsilon_n' B_n \varepsilon_n) &= \sum_{i=1}^n \sum_{j=1}^n a_{n,ij} (b_{n,ij} + b_{n,ji}) \sigma_{ni}^2 \sigma_{nj}^2 \\
&= \text{tr}(\Sigma_n A_n (B_n' \Sigma_n + \Sigma_n B_n)) \quad \text{where } \Sigma_n = \text{Diag}(\sigma_{n1}^2, \dots, \sigma_{nn}^2). \\
(2) \quad E(\varepsilon_n C_n \varepsilon_n)^2 &= \sum_{i=1}^n c_{n,ii}^2 [E(\varepsilon_{ni}^4) - 3\sigma_{ni}^4] + \left(\sum_{i=1}^n c_{n,ii} \sigma_{ni}^2\right)^2 + \sum_{i=1}^n \sum_{j=1}^n c_{n,ij} (c_{n,ij} + c_{n,ji}) \sigma_{ni}^2 \sigma_{nj}^2 \\
&= \sum_{i=1}^n c_{n,ii}^2 [E(\varepsilon_{ni}^4) - 3\sigma_{ni}^4] + \text{tr}^2(\Sigma_n C_n) + \text{tr}(\Sigma_n C_n C_n' \Sigma_n + \Sigma_n C_n \Sigma_n C_n), \\
(3) \quad \text{var}(\varepsilon_n C_n \varepsilon_n) &= \sum_{i=1}^n c_{n,ii}^2 [E(\varepsilon_{ni}^4) - 3\sigma_{ni}^4] + \sum_{i=1}^n \sum_{j=1}^n c_{n,ij} (c_{n,ij} + c_{n,ji}) \sigma_{ni}^2 \sigma_{nj}^2 \\
&= \sum_{i=1}^n c_{n,ii}^2 [E(\varepsilon_{ni}^4) - 3\sigma_{ni}^4] + \text{tr}(\Sigma_n C_n C_n' \Sigma_n + \Sigma_n C_n \Sigma_n C_n). \\
(4) \quad E(\varepsilon_n' C_n \varepsilon_n) &= O(n), \quad \text{var}(\varepsilon_n' C_n \varepsilon_n) = O(n), \quad \varepsilon_n' C_n \varepsilon_n = O_p(n). \\
(5) \quad E(C_n \varepsilon_n) &= 0, \quad \text{var}(C_n \varepsilon_n) = O(n), \quad C_n \varepsilon_n = O_p(n), \quad \text{var}(q_n' C_n \varepsilon_n) = O(n), \quad q_n' C_n \varepsilon_n = O_p(n).
\end{aligned}$$

Proof. For (1), (2) and (3) see Lin and Lee (2010). For the rest of proof, let c_1, c_2, c_3, c_4 and m be positive constant real numbers.

(4) $E(\varepsilon_n' C_n \varepsilon_n) = \text{tr}(C_n E(\varepsilon_n \varepsilon_n')) = \text{tr}(C_n \Sigma_n) = \sum_{i=1}^n c_{n,ii} \sigma_{ni}^2$. By hypothesis and *Assumption 1*, $c_{n,ii}$ s and σ_{ni}^2 s are uniformly bounded. Then, $E(\varepsilon_n' C_n \varepsilon_n) = O(n)$. The order of $\text{var}(\varepsilon_n C_n \varepsilon_n)$ can be obtained from (3). The first term in (3) is $\sum_{i=1}^n c_{n,ii}^2 [E(\varepsilon_{ni}^4) - 3\sigma_{ni}^4] = O(n)$, since $\sigma_{ni}^2, E(\varepsilon_{ni}^4)$ and $c_{n,ii}$ are uniformly bounded $\forall i$. The second term in (3) is $\text{tr}(\Sigma_n C_n C_n' \Sigma_n + \Sigma_n C_n \Sigma_n C_n) = O(n)$, since $\Sigma_n C_n C_n' \Sigma_n$ and $\Sigma_n C_n \Sigma_n C_n$ are uniformly bounded in both row and column sums. Thus, $\text{var}(\varepsilon_n C_n \varepsilon_n) = O(n)$. The next result can be obtained by Markov's inequality: $P(|\varepsilon_n' C_n \varepsilon_n| > m) \leq \frac{1}{m} E(\varepsilon_n' C_n \varepsilon_n) = \frac{1}{m} O(n)$ and hence, $\varepsilon_n' C_n \varepsilon_n = O_p(n)$.

(5) By *Assumption 1*, $E(C_n \varepsilon_n) = 0$ and $\text{var}(C_n \varepsilon_n) = C_n \Sigma_n C_n' = \sum_{i=1}^n \sigma_{ni}^2 c_{n,i} c_{n,i}'$, where $c_{n,i}$ is the i th column of C_n . By hypothesis and *Assumption 1*, there are constants c_1 and c_2 such that $|\sigma_{ni}^2| \leq c_1$ and $\|c_{n,i}\| \leq c_2 \forall i, n$. Hence, $\|\sum_{i=1}^n \sigma_{ni}^2 c_{n,i} c_{n,i}'\| \leq \sum_{i=1}^n |\sigma_{ni}^2| \times \|c_{n,i}\|^2 \leq n c_1 c_2^2 = O(n)$. The next result follows from Chebyshev's inequality: $P(\|C_n \varepsilon_n - E(C_n \varepsilon_n)\| > m) \leq \frac{1}{m^2} \text{var}(C_n \varepsilon_n) = \frac{1}{m^2} O(n)$. Hence, $C_n \varepsilon_n = O_p(n)$. Next, $\text{var}(q_n' C_n \varepsilon_n) = q_n' C_n \Sigma_n C_n' q_n = \sum_{i=1}^n \sigma_{ni}^2 k_{ni}^2$, where k_{ni} is the i th element of $q_n' C_n$. By hypothesis, there are constants c_3 and c_4 such that $|q_{ni}| \leq c_3$ and $|\sum_{j=1}^n c_{n,ji}| \leq c_4 \forall i, n$. Then, $|k_{ni}| = |\sum_{j=1}^n q_{nj} c_{n,ji}| \leq c_3 c_4 \forall i, n$. Thus, $\text{var}(q_n' C_n \varepsilon_n) = \sum_{i=1}^n \sigma_{ni}^2 k_{ni}^2 \leq n c_1 (c_3 c_4)^2 = O(n)$. The last result follows from Chebyshev's inequality: $P(|q_n' C_n \varepsilon_n - E(q_n' C_n \varepsilon_n)| > m) \leq \frac{1}{m^2} \text{var}(q_n' C_n \varepsilon_n) = \frac{1}{m^2} O(n)$ and hence, $q_n' C_n \varepsilon_n = O_p(n)$. $\square$

Lemma 3. Assume that $n \times k$ matrix X_n has uniformly bounded elements in absolute value, and $\lim_{n \rightarrow \infty} \frac{1}{n} X_n' R_n' R_n X_n$ exists and is nonsingular. Let $\bar{\mathbb{M}}_n = (I_n - \mathbb{P}_n)$, where $\mathbb{P}_n = R_n X_n (X_n' R_n' R_n X_n)^{-1} X_n' R_n'$. Assume that ε_n satisfies *Assumption 1* with covariance matrix de-

noted by $\Sigma_n = \text{Diag}\{\sigma_{n1}^2, \dots, \sigma_{nn}^2\}$. Then,

- (1) $\bar{\mathbb{M}}_n$ and $\mathbb{P}_n$ are uniformly bounded in absolute value in both row and column sums.
- (2) $\text{var}(\mathbb{P}_n \varepsilon_n) = O(\frac{1}{n})$, $\mathbb{P}_n \varepsilon_n = o_p(1)$, $\text{var}(\varepsilon_n \mathbb{P}_n \varepsilon_n) = O(\frac{1}{n})$, $\varepsilon_n \mathbb{P}_n \varepsilon_n = O_p(1)$.
- (3) Elements of $\mathbb{P}_n$ are $O(\frac{1}{n})$.

Proof. (1) Let $K_n = (\frac{1}{n} X_n' R_n' R_n X_n)^{-1}$. By hypothesis, K_n has finite limit so that there exist constant c_1 such that $|k_{n,ij}| \leq c_1$ for all i, j and n , where $k_{n,ij}$ is the (i, j) th element of K_n . Let $\bar{X}_n = R_n X_n$. Elements of $\bar{X}_n$ are uniformly bounded since both X_n and R_n are uniformly bounded in absolute value in row and column sums. Denote (i, j) th element of $\bar{X}_n$ by $\bar{x}_{n,ij}$, then there exists a constant c_2 such that $|\bar{x}_{n,ij}| \leq c_2$ for all i, j and n . Then, $\mathbb{P}_n = \frac{1}{n} \bar{X}_n (\frac{1}{n} X_n' R_n' R_n X_n)^{-1} \bar{X}_n' = \frac{1}{n} \bar{X}_n K_n \bar{X}_n' = \frac{1}{n} \sum_{s=1}^k \sum_{r=1}^k k_{n,rs} \bar{x}_{n,s} \bar{x}_{n,r}'$, where $\bar{x}_{n,r}$ and $\bar{x}_{n,s}$ are respectively the r th and the s th columns of $\bar{X}_n$. Denote (i, j) th element of $\mathbb{P}_n$ by $p_{n,ij}$, then $\sum_{j=1}^n |p_{n,ij}| = \frac{1}{n} \sum_{j=1}^n |\sum_{s=1}^k \sum_{r=1}^k k_{n,rs} \bar{x}_{n,s} \bar{x}_{n,r}'| \leq k^2 c_1 c_2^2$ for all i and n . Likewise, $\sum_{i=1}^n |p_{n,ij}| = \frac{1}{n} \sum_{i=1}^n |\sum_{s=1}^k \sum_{r=1}^k k_{n,rs} \bar{x}_{n,s} \bar{x}_{n,r}'| \leq k^2 c_1 c_2^2$ for all j and n . These results show that $\mathbb{P}_n$ is uniformly bounded in absolute value in row and column sums, which implies that $\bar{\mathbb{M}}_n = (I_n - \mathbb{P}_n)$ is also uniformly bounded in absolute value in row and column sums. (2) These results directly follow from *Lemma 2 (4) and (5)*. (3) The (i, j) th element of $\mathbb{P}_n$ is $|p_{n,ij}| = \frac{1}{n} |\sum_{s=1}^k \sum_{r=1}^k k_{n,rs} \bar{x}_{n,s} \bar{x}_{n,r}'| \leq \frac{k^2 c_1 c_2^2}{n} = O(\frac{1}{n})$. $\square$

B The Inconsistency of the ML Estimator

The first order conditions of the log-likelihood function with respect to β and σ^2 are given by

$$\frac{\partial \ln L_n(\zeta)}{\partial \beta} = \frac{1}{\sigma^2} X_n' R_n'(\rho) R_n(\rho) [S_n(\lambda) Y_n - X_n \beta], \quad (\text{B.1a})$$

$$\frac{\partial \ln L_n(\zeta)}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4} \varepsilon_n'(\theta) \varepsilon_n(\theta). \quad (\text{B.1b})$$

The solutions of the above first order conditions yield the ML estimators for β and σ^2 :

$$\hat{\beta}_n(\delta) = [\bar{X}_n'(\rho) \bar{X}_n(\rho)]^{-1} \bar{X}_n'(\rho) R_n(\rho) S_n(\lambda) Y_n, \quad (\text{B.2a})$$

$$\hat{\sigma}_n^2(\delta) = \frac{1}{n} \varepsilon_n'(\theta) \varepsilon_n(\theta). \quad (\text{B.2b})$$

where, $\bar{X}_n(\rho) = R_n(\rho) X_n$ and $\varepsilon_n(\theta) = R_n(\rho) [S_n(\lambda) Y_n - X_n \hat{\beta}_n(\delta)]$. Explicitly, the MLE $\hat{\sigma}_n^2(\delta)$ can be written as

$$\hat{\sigma}_n^2(\delta) = \frac{1}{n} Y_n' S_n'(\lambda) R_n'(\rho) \bar{\mathbb{M}}_n(\rho) R_n(\rho) S_n(\lambda) Y_n, \quad (\text{B.3})$$

where $\bar{\mathbb{M}}_n(\rho) = (I_n - \mathbb{P}_n(\rho))$ with $\mathbb{P}_n(\rho) = \bar{X}_n(\rho) \left(\bar{X}_n'(\rho) \bar{X}_n(\rho) \right)^{-1} \bar{X}_n'(\rho)$ is a projection type matrix. Note that $\bar{X}_n'(\rho) \bar{\mathbb{M}}_n(\rho) = 0_{k \times n}$ and $\bar{\mathbb{M}}_n(\rho) \bar{X}_n(\rho) = 0_{n \times k}$. Substitution of $R_n(\rho) S_n(\lambda) Y_n = R_n(\rho) X_n \beta_n + \varepsilon_n$ into $\hat{\sigma}_n^2(\delta)$ yields $\hat{\sigma}_n^2(\delta) = \frac{1}{n} \varepsilon_n' \bar{\mathbb{M}}_n(\rho) \varepsilon_n$.

At δ_0 , the probability limit of $\hat{\sigma}_n^2(\delta_0)$ is

$$\text{plim}_{n \rightarrow \infty} \hat{\sigma}_n^2(\delta_0) = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \varepsilon_n' \varepsilon_n - \text{plim}_{n \rightarrow \infty} \frac{1}{n^2} \varepsilon_n \bar{X}_n \left[\frac{1}{n} \bar{X}_n' \bar{X}_n \right]^{-1} \bar{X}_n' \varepsilon_n. \quad (\text{B.4})$$

The first term on the right hand side converges to $\frac{1}{n} \sum_{i=1}^n \sigma_{ni}^2$ by Chebyshev Weak Law of Large numbers. The second term vanishes by the virtue of *Lemma 2 (4)* and *Lemma 3*. Therefore, we have

$$\hat{\sigma}_n^2(\delta_0) = \frac{1}{n} \sum_{i=1}^n \sigma_{ni}^2 + o_p(1). \quad (\text{B.5})$$

Concentrating out β and σ^2 from the log-likelihood function yield

$$\ln L_n(\delta) = -\frac{n}{2} (\ln(2\pi) + 1) - \frac{n}{2} \ln \hat{\sigma}_n^2(\delta) + \ln |S_n(\lambda)| + \ln |R_n(\rho)|. \quad (\text{B.6})$$

For the first order conditions with respect to ρ and λ , the partial derivatives $\frac{\partial \hat{\sigma}_n^2(\delta)}{\partial \rho}$ and $\frac{\partial \hat{\sigma}_n^2(\delta)}{\partial \lambda}$ are required. These terms are given by

$$\begin{aligned} \frac{\partial \hat{\sigma}_n^2(\delta)}{\partial \rho} &= -\frac{2}{n} Y_n' S_n'(\lambda) R_n'(\rho) \bar{\mathbb{M}}_n(\rho) M_n S_n(\lambda) Y_n \\ &\quad + \frac{2}{n} Y_n' S_n'(\lambda) R_n'(\rho) \mathbb{P}_n(\rho) H_n'(\rho) \bar{\mathbb{M}}_n(\rho) R_n(\rho) S_n(\lambda) Y_n. \end{aligned} \quad (\text{B.7})$$

$$\frac{\partial \hat{\sigma}_n^2(\delta)}{\partial \lambda} = -\frac{2}{n} Y_n' S_n'(\lambda) R_n'(\rho) \bar{\mathbb{M}}_n(\rho) R_n(\rho) W_n Y_n. \quad (\text{B.8})$$

The first order conditions of the concentrated log-likelihood function with respect to ρ and λ are given by

$$\frac{\partial \ln L_n(\delta)}{\partial \rho} = -\frac{n}{2 \hat{\sigma}_n^2(\delta)} \frac{\partial \hat{\sigma}_n^2(\delta)}{\partial \rho} - \text{tr}(H_n(\rho)), \quad (\text{B.9a})$$

$$\frac{\partial \ln L_n(\delta)}{\partial \lambda} = -\frac{n}{2 \hat{\sigma}_n^2(\delta)} \frac{\partial \hat{\sigma}_n^2(\delta)}{\partial \lambda} - \text{tr}(G_n(\lambda)), \quad (\text{B.9b})$$

where $G_n(\lambda) = W_n S_n^{-1}(\lambda)$ and $H_n(\rho) = M_n R_n^{-1}(\rho)$. For the consistency of the ML estimators $\hat{\lambda}_n$ and $\hat{\rho}_n$, the necessary condition is $\text{plim}_{n \rightarrow \infty} \frac{1}{n} \frac{\partial \ln L_n(\delta_0)}{\partial \delta} = 0$. More explicitly, the probability limit of the following equation must be zero.

$$\frac{1}{n} \frac{\partial \ln L_n(\delta_0)}{\partial \delta} = \begin{pmatrix} \frac{1}{n} \left(-\frac{n}{\frac{2}{n} \varepsilon_n' \bar{\mathbb{M}}_n \varepsilon_n} \frac{\partial \hat{\sigma}_n^2(\delta_0)}{\partial \rho} \right) - \frac{1}{n} \text{tr}(H_n) \\ \frac{1}{n} \left(-\frac{n}{\frac{2}{n} \varepsilon_n' \bar{\mathbb{M}}_n \varepsilon_n} \frac{\partial \hat{\sigma}_n^2(\delta_0)}{\partial \lambda} \right) - \frac{1}{n} \text{tr}(G_n) \end{pmatrix}. \quad (\text{B.10})$$

The probability limit of $\frac{\partial \hat{\sigma}_n^2(\delta_0)}{\partial \rho}$ and $\frac{\partial \hat{\sigma}_n^2(\delta_0)}{\partial \lambda}$ are required for the above equation, which can be found by using the derivative expressions in (B.7) and (B.8).

The probability limit of the first term in the first row of (B.10) is written as

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \frac{1}{n} \left(-\frac{n}{\frac{2}{n} \varepsilon_n' \bar{\mathbb{M}}_n \varepsilon_n} \frac{\partial \hat{\sigma}_n^2(\delta_0)}{\partial \rho} \right) &= \text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n} Y_n' S_n' R_n' \bar{\mathbb{M}}_n M_n S_n Y_n}{\frac{1}{n} \varepsilon_n' \bar{\mathbb{M}}_n \varepsilon_n} \\ &\quad - \text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n} Y_n' S_n' R_n' \mathbb{P}_n H_n' \bar{\mathbb{M}}_n R_n S_n Y_n}{\frac{1}{n} \varepsilon_n' \bar{\mathbb{M}}_n \varepsilon_n}. \end{aligned} \quad (\text{B.11})$$

Each term is handled separately below by using equalities $R_n S_n Y_n = R_n X_n \beta_0 + \varepsilon_n$ and $S_n Y_n = X_n \beta_0 + R_n^{-1} \varepsilon_n$. Note that $\bar{X}_n' \bar{\mathbb{M}}_n = 0_{k \times n}$ and $\bar{\mathbb{M}}_n \bar{X}_n = 0_{n \times k}$. The first term on the r.h.s. of (B.11) can be written as $\text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n} Y_n' S_n' R_n' \bar{\mathbb{M}}_n M_n S_n Y_n}{\frac{1}{n} \varepsilon_n' \bar{\mathbb{M}}_n \varepsilon_n} = \text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n} \varepsilon_n' \bar{\mathbb{M}}_n M_n X_n \beta_0}{\frac{1}{n} \varepsilon_n' \bar{\mathbb{M}}_n \varepsilon_n} + \text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n} \varepsilon_n' \bar{\mathbb{M}}_n M_n R_n^{-1} \varepsilon_n}{\frac{1}{n} \varepsilon_n' \bar{\mathbb{M}}_n \varepsilon_n}$. Substitution of $\bar{\mathbb{M}}_n = (I_n - \bar{X}_n [\bar{X}_n' \bar{X}_n]^{-1} \bar{X}_n')$ into r.h.s. yields

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n} Y_n' S_n' R_n' \bar{\mathbb{M}}_n M_n S_n Y_n}{\frac{1}{n} \varepsilon_n' \bar{\mathbb{M}}_n \varepsilon_n} &= \text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n} \varepsilon_n' H_n \varepsilon_n}{\frac{1}{n} \varepsilon_n' \bar{\mathbb{M}}_n \varepsilon_n} + \text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n} \varepsilon_n' \bar{\mathbb{M}}_n M_n X_n \beta_0}{\frac{1}{n} \varepsilon_n' \bar{\mathbb{M}}_n \varepsilon_n} \\ &\quad - \text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n^2} \varepsilon_n' \bar{X}_n [\frac{1}{n} \bar{X}_n' \bar{X}_n]^{-1} \bar{X}_n' M_n R_n^{-1} \varepsilon_n}{\frac{1}{n} \varepsilon_n' \bar{\mathbb{M}}_n \varepsilon_n}. \end{aligned} \quad (\text{B.12})$$

By *Lemma 2 (5)* and (B.5), the second term on the r.h.s of (B.12) vanishes. The third term vanishes by *Lemma 2 (4)* and (B.5). The probability limit of the first term on the r.h.s. of (B.12) can be found by Chebyshev inequality. By *Lemma 2 (4)*, the variance of the term $\frac{1}{n} \varepsilon_n' H_n \varepsilon_n$ is $O(\frac{1}{n}) = o(1)$. Hence, $\text{plim}_{n \rightarrow \infty} (\frac{1}{n} \varepsilon_n' H_n \varepsilon_n - E(\frac{1}{n} \varepsilon_n' H_n \varepsilon_n)) = \text{plim}_{n \rightarrow \infty} (\frac{1}{n} \varepsilon_n' H_n \varepsilon_n - \frac{1}{n} \sum_{i=1}^n H_{n,ii} \sigma_{ni}^2) = 0$. Combining this result with the result in (B.5), we get

$$\text{plim}_{n \rightarrow \infty} \left(\frac{\frac{1}{n} Y_n' S_n' R_n' \bar{\mathbb{M}}_n M_n S_n Y_n}{\frac{1}{n} \varepsilon_n' \bar{\mathbb{M}}_n \varepsilon_n} - \frac{\frac{1}{n} \sum_{i=1}^n H_{n,ii} \sigma_{ni}^2}{\frac{1}{n} \sum_{i=1}^n \sigma_{ni}^2} \right) = 0. \quad (\text{B.13})$$

Now, we return to the second term on the r.h.s of (B.11): $\text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n} Y_n' S_n' R_n' \mathbb{P}_n H_n' \bar{\mathbb{M}}_n R_n S_n Y_n}{\frac{1}{n} \varepsilon_n' \bar{\mathbb{M}}_n \varepsilon_n} = \text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n} (X_n \beta_0)' \mathbb{P}_n M_n' \bar{\mathbb{M}}_n \varepsilon_n}{\frac{1}{n} \varepsilon_n' \bar{\mathbb{M}}_n \varepsilon_n} + \text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n} \varepsilon_n' \bar{X}_n [\bar{X}_n' \bar{X}_n]^{-1} \bar{X}_n' H_n' \bar{\mathbb{M}}_n \varepsilon_n}{\frac{1}{n} \varepsilon_n' \bar{\mathbb{M}}_n \varepsilon_n}$. The first term on the r.h.s. converges in probability to zero by *Lemma 2 (5)* and (B.5). The second term on the r.h.s. converge in

probability to zero by the virtue of *Lemma 2 (4)*, (B.5) and *Lemma 3*. Hence,

$$\text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} \left(-\frac{n}{\frac{2}{n} \varepsilon'_n \bar{\mathbb{M}}_n \varepsilon_n} \frac{\partial \hat{\sigma}_n^2(\delta_0)}{\partial \rho} \right) - \frac{\frac{1}{n} \sum_{i=1}^n H_{n,ii} \sigma_{ni}^2}{\frac{1}{n} \sum_{i=1}^n \sigma_{ni}^2} \right) = 0. \quad (\text{B.14})$$

Now, the probability limit in the second row of (B.10) is evaluated.

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \frac{1}{n} \left(-\frac{n}{\frac{2}{n} \varepsilon'_n \bar{\mathbb{M}}_n(\rho_0) \varepsilon_n} \frac{\partial \hat{\sigma}_n^2(\delta_0)}{\partial \lambda} \right) &= \text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n} \varepsilon'_n \bar{\mathbb{M}}_n R_n G_n R_n^{-1} \varepsilon_n}{\frac{1}{n} \varepsilon'_n \bar{\mathbb{M}}_n \varepsilon_n} + \text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n} \varepsilon'_n \bar{\mathbb{M}}_n R_n G_n X_n \beta_0}{\frac{1}{n} \varepsilon'_n \bar{\mathbb{M}}_n \varepsilon_n} \\ &+ \text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n} \beta'_0 X'_n R'_n \bar{\mathbb{M}}_n R_n G_n X_n \beta_0}{\frac{1}{n} \varepsilon'_n \bar{\mathbb{M}}_n \varepsilon_n} + \text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n} \beta'_0 X'_n R'_n \bar{\mathbb{M}}_n R_n G_n R_n^{-1} \varepsilon_n}{\frac{1}{n} \varepsilon'_n \bar{\mathbb{M}}_n \varepsilon_n}. \end{aligned} \quad (\text{B.15})$$

The first term on the r.h.s. of (B.15) is handled separately later. First, the third and the fourth term on the r.h.s. are zero since $X'_n R'_n \bar{\mathbb{M}}_n = \bar{X}'_n \bar{\mathbb{M}}_n = 0_{k \times n}$. The second term on the r.h.s. vanishes in probability, since

$$\text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n} \varepsilon'_n \bar{\mathbb{M}}_n R_n G_n X_n \beta_0}{\frac{1}{n} \varepsilon'_n \bar{\mathbb{M}}_n \varepsilon_n} = \text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n} \varepsilon'_n R_n G_n X_n \beta_0}{\frac{1}{n} \varepsilon'_n \bar{\mathbb{M}}_n \varepsilon_n} - \text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n^2} \varepsilon'_n \bar{X}_n [\frac{1}{n} \bar{X}'_n \bar{X}_n]^{-1} \bar{X}'_n R_n G_n X_n \beta_0}{\frac{1}{n} \varepsilon'_n \bar{\mathbb{M}}_n \varepsilon_n} \quad (\text{B.16})$$

The numerators on the r.h.s. of (B.16) converge in probability to zero by *Lemma 2 (5)* and *Lemma 3*, and the term in the denominator converges to $\frac{1}{n} \sum_{i=1}^n \sigma_{ni}^2$ as shown in (B.5). The overall result is zero since $\frac{1}{n} \sum_{i=1}^n \sigma_{ni}^2$ is uniformly bounded for all n and i by *Assumption 1*. As for the first term on the r.h.s. of (B.15), let $\bar{G}_n = R_n G_n R_n^{-1}$. Then,

$$\text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n} \varepsilon'_n \bar{\mathbb{M}}_n \bar{G}_n \varepsilon_n}{\frac{1}{n} \varepsilon'_n \bar{\mathbb{M}}_n \varepsilon_n} = \text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n} \varepsilon_n \bar{G}_n \varepsilon_n}{\frac{1}{n} \varepsilon'_n \bar{\mathbb{M}}_n \varepsilon_n} - \text{plim}_{n \rightarrow \infty} \frac{\frac{1}{n} \varepsilon_n \bar{X}_n [\bar{X}'_n \bar{X}_n]^{-1} \bar{X}'_n \bar{G}_n \varepsilon_n}{\frac{1}{n} \varepsilon'_n \bar{\mathbb{M}}_n \varepsilon_n}. \quad (\text{B.17})$$

The numerator of the last term on the r.h.s. of (B.17) is $O_p(\frac{1}{n})$ by *Lemma 2 (4)* and *Lemma 3*. Hence, as n goes to infinity the numerator converges in probability to zero. The denominator converges to the uniformly bounded sum in (B.5). Hence, this term vanishes. Now, we can return to the first term on the r.h.s. of (B.17). By *Lemma 2 (4)*, the variance of $\frac{1}{n} \varepsilon'_n \bar{G}_n \varepsilon_n$ is $O(\frac{1}{n}) = o(1)$. Then, Chebyshev inequality implies that $\text{plim}_{n \rightarrow \infty} (\frac{1}{n} \varepsilon'_n \bar{G}_n \varepsilon_n - E(\frac{1}{n} \varepsilon'_n \bar{G}_n \varepsilon_n)) = \text{plim}_{n \rightarrow \infty} (\frac{1}{n} \varepsilon'_n \bar{G}_n \varepsilon_n - \frac{1}{n} \sum_{i=1}^n \bar{G}_{n,ii} \sigma_{ni}^2) = 0$. Therefore,

$$\text{plim}_{n \rightarrow \infty} \left(\frac{\frac{1}{n} \varepsilon_n \bar{G}_n \varepsilon_n}{\frac{1}{n} \varepsilon'_n \bar{\mathbb{M}}_n \varepsilon_n} - \frac{\frac{1}{n} \sum_{i=1}^n \bar{G}_{n,ii} \sigma_{ni}^2}{\frac{1}{n} \sum_{i=1}^n \sigma_{ni}^2} \right) = 0. \quad (\text{B.18})$$

These results imply that the probability limit in (B.15) is given by

$$\frac{1}{n} \left(-\frac{n}{\frac{2}{n} \varepsilon'_n \bar{\mathbb{M}}_n \varepsilon_n} \frac{\partial \hat{\sigma}_n^2(\delta_0)}{\partial \lambda} \right) = \frac{\frac{1}{n} \sum_{i=1}^n \bar{G}_{n,ii} \sigma_{ni}^2}{\frac{1}{n} \sum_{i=1}^n \sigma_{ni}^2} + o_p(1). \quad (\text{B.19})$$

Combining (B.19) and (B.14), we obtain:

$$\frac{1}{n} \frac{\partial \ln L_n(\delta_0)}{\partial \delta} = \begin{pmatrix} \frac{\frac{1}{n} \sum_{i=1}^n H_{n,ii} \sigma_{ni}^2}{\frac{1}{n} \sum_{i=1}^n \sigma_{ni}^2} - \frac{1}{n} \text{tr}(H_n) + o_p(1) \\ \frac{\frac{1}{n} \sum_{i=1}^n \bar{G}_{n,ii} \sigma_{ni}^2}{\frac{1}{n} \sum_{i=1}^n \sigma_{ni}^2} - \frac{1}{n} \text{tr}(G_n) + o_p(1) \end{pmatrix}. \quad (\text{B.20})$$

The result in (4.8) of the main text follows from (B.20).

The asymptotic bias of the MLE $\hat{\beta}_n(\delta)$ can be determined from (B.2a). Let $\mathbb{D}_n(\rho) = [\bar{X}'_n(\rho) \bar{X}_n(\rho)]^{-1}$. Then, the MLE $\hat{\beta}_n(\delta)$ can be written as

$$\begin{aligned} \hat{\beta}_n(\delta) &= [\bar{X}'_n(\rho) \bar{X}_n(\rho)]^{-1} \bar{X}'_n(\rho) R_n(\rho) S_n(\lambda) Y_n \\ &= \beta_0 + \mathbb{D}_n(\rho) \bar{X}'_n(\rho) R_n(\rho) R_n^{-1} \varepsilon_n + (\lambda_0 - \lambda) \mathbb{D}_n(\rho) \bar{X}'_n(\rho) R_n(\rho) G_n X_n \beta_0 \\ &\quad + (\lambda_0 - \lambda) \mathbb{D}_n(\rho) \bar{X}'_n(\rho) R_n(\rho) G_n R_n^{-1} \varepsilon_n, \end{aligned} \quad (\text{B.21})$$

where we use $S_n(\lambda) = S_n + (\lambda_0 - \lambda) W_n$. Substitution of $R_n(\rho) = R_n + (\rho_0 - \rho) M_n$ into the MLE $\hat{\beta}_n(\delta)$ yields

$$\begin{aligned} \hat{\beta}_n(\delta) &= \beta_0 + \mathbb{D}_n(\rho) \bar{X}'_n \varepsilon_n + (\rho_0 - \rho) \mathbb{D}_n(\rho) \bar{X}'_n M'_n \varepsilon_n + (\rho_0 - \rho) \mathbb{D}_n(\rho) \bar{X}'_n H_n \varepsilon_n \\ &\quad + (\rho_0 - \rho)^2 \mathbb{D}_n(\rho) \bar{X}'_n M'_n H_n \varepsilon_n + (\lambda_0 - \lambda) \mathbb{D}_n(\rho) \bar{X}'_n \bar{G}_n \bar{X}_n \beta_0 + (\lambda_0 - \lambda)(\rho_0 - \rho) \mathbb{D}_n(\rho) \bar{X}'_n H_n^s \bar{G}_n \bar{X}_n \beta_0 \\ &\quad + (\lambda_0 - \lambda)(\rho_0 - \rho)^2 \mathbb{D}_n(\rho) \bar{X}'_n H'_n H_n \bar{G}_n \bar{X}_n \beta_0 + (\lambda_0 - \lambda) \mathbb{D}_n(\rho) \bar{X}'_n \bar{G}_n \varepsilon_n \\ &\quad + (\lambda_0 - \lambda)(\rho_0 - \rho) \mathbb{D}_n(\rho) \bar{X}'_n M'_n \bar{G}_n \varepsilon_n + (\lambda_0 - \lambda)(\rho_0 - \rho) \mathbb{D}_n(\rho) \bar{X}'_n H_n \bar{G}_n \varepsilon_n \\ &\quad + (\lambda_0 - \lambda)(\rho_0 - \rho)^2 \mathbb{D}_n(\rho) \bar{X}'_n H'_n H_n \bar{G}_n \varepsilon_n \end{aligned} \quad (\text{B.22})$$

Assumption 3' along with *Lemma 3* implies that $\mathbb{D}_n(\rho)$ is uniformly bounded in absolute value in row and column sums. Then, using *Lemma 2 (5)*, terms with ε_n vanish in probability in the MLE $\hat{\beta}_n(\delta)$. Thus,

$$\begin{aligned} \hat{\beta}_n(\delta) &= \beta_0 + (\lambda_0 - \lambda) \mathbb{D}_n(\rho) \bar{X}'_n \bar{G}_n \bar{X}_n \beta_0 + (\lambda_0 - \lambda)(\rho_0 - \rho) \mathbb{D}_n(\rho) \bar{X}'_n H_n^s \bar{G}_n \bar{X}_n \beta_0 \\ &\quad + (\lambda_0 - \lambda)(\rho_0 - \rho)^2 \mathbb{D}_n(\rho) \bar{X}'_n H'_n H_n \bar{G}_n \bar{X}_n \beta_0 + o_p(1) \end{aligned}$$

The asymptotic bias of $\hat{\beta}_n(\delta_n)$ follows from the above equation.

The specification with $\lambda_0 = 0$ in (2.1) is called spatial error model (SEM). For the SEM, the log-likelihood function simplifies to $\ln L_n(\zeta) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) + \ln |R_n(\rho)| - \frac{1}{2\sigma^2} [Y_n - X_n \beta]' R'_n(\rho) R_n(\rho) [Y_n - X_n \beta]$, where $\zeta = (\theta', \sigma^2)'$ with $\theta = (\rho, \beta)'$. The first order conditions yield $\hat{\beta}_n(\rho) = \mathbb{D}_n(\rho) \bar{X}'_n(\rho) R_n(\rho) Y_n$ and $\hat{\sigma}_n^2(\rho) = \frac{1}{n} \varepsilon'_n(\theta) \varepsilon_n(\theta)$. The necessary condition for the consistency the MLE $\hat{\rho}_n$ can be obtained from (B.20). From the first row of (B.20), we get $\frac{1}{n} \frac{\partial \ln L_n(\rho_0)}{\partial \rho} = \frac{\text{cov}(H_{n,ii}, \sigma_{ni}^2)}{\bar{\sigma}^2} + o_p(1)$, which implies that the MLE $\hat{\rho}_n$ is inconsistent. Substitution of

$Y_n = X_n\beta_0 + R_n^{-1}\varepsilon_n$ into $\hat{\beta}_n(\rho)$ yields

$$\begin{aligned}\hat{\beta}_n(\rho) &= \beta_0 + \mathbb{D}_n(\rho)\bar{X}_n'(\rho)R_n(\rho)R_n^{-1}\varepsilon_n = \beta_0 + \mathbb{D}_n(\rho)X_n'R_n'\varepsilon_n + (\rho_0 - \rho)\mathbb{D}_n(\rho)X_n'M_n'\varepsilon_n \\ &\quad + (\rho_0 - \rho)\mathbb{D}_n(\rho)X_n'R_n'H_n\varepsilon_n + (\rho_0 - \rho)^2\mathbb{D}_n(\rho)X_n'M_n'H_n\varepsilon_n.\end{aligned}\tag{B.23}$$

Assumption 3' along with *Lemma 3* implies that $\mathbb{D}_n(\rho)$ is uniformly bounded in absolute value in row and column sums. By *Lemma 2 (5)*, terms with ε_n have variances of $O(\frac{1}{n})$. Then, Chebyshev's inequality implies that $\hat{\beta}_n(\rho) = \beta_0 + o_p(1)$ so that $\hat{\beta}_n(\hat{\rho}_n)$ has no asymptotic bias.

C Proof of Main Propositions

Proof of Proposition 1. The GMM estimator is an extremum estimator. The conditions for the consistency of extremum estimators are established in Theorem 4.1.1 in Amemiya (1985, see p. 106-107). Let $\mathbb{L}_n(\theta)$ be the objective function of the GMM estimator. The GMM estimator $\hat{\theta}_n = \operatorname{argmin}_{\theta \in \Theta} \mathbb{L}_n(\theta) = \operatorname{argmin}_{\theta \in \Theta} g_n'(\theta)\Psi_n'\Psi_n g_n(\theta)$ is consistent under the following conditions: (1) The parameter space is a compact subset of $\mathbb{R}^{k+2}$ and $\theta_0 \in \Theta$, (2) $\mathbb{L}_n(\theta)$ is continuous in θ , (3) $\frac{1}{n}\mathbb{L}_n(\theta)$ converges to the non-stochastic function $\frac{1}{n}E(\mathbb{L}_n(\theta))$ in probability uniformly in $\theta \in \Theta$ as $n \rightarrow \infty$, and (4) $\mathbb{L}(\theta) = \lim_{n \rightarrow \infty} \frac{1}{n}E(\mathbb{L}_n(\theta))$ attains a unique global maximum at θ_0 (i.e. the identification conditions given in Assumption 6 are satisfied).

The conditions (1), (2) and (4) are satisfied under our assumptions. For condition (3), it is enough to show that $\frac{1}{n}\Psi_n g_n(\theta)$ converges to its limit $\frac{1}{n}E(\Psi_n g_n(\theta))$ uniformly in $\theta \in \Theta$.

Let $\Psi_n = (\Psi_{n1}, \dots, \Psi_{nm}, \Psi_{nx})$, where Ψ_{nj} is the j th column and Ψ_{nx} is a submatrix. Also, let $\Psi_{i,n}$ be the i th row of the matrix Ψ_n such that $\Psi_{i,n} = (\Psi_{i,n1}, \dots, \Psi_{i,nm}, \Psi_{i,nx})$ where $\Psi_{i,nj}$, $j = 1, \dots, m$, are scalars and $\Psi_{i,nx}$ is a row subvector with its dimension k^* as the number of rows of Q_n . It is sufficient to show the uniform convergence of $\Psi_{i,n} g_n(\theta)$ for each i .

More explicitly, $\Psi_{i,n} g_n(\theta) = \varepsilon_n'(\theta) \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn} \right) \varepsilon_n(\theta) + \Psi_{i,nx} Q_n' \varepsilon_n(\theta)$. Since $\varepsilon_n(\theta) = R_n(\rho) [S_n(\lambda)Y_n - X_n\beta]$, $S_n(\lambda) = S_n + (\lambda_0 - \lambda)W_n$, and $R_n(\rho) = R_n + (\rho_0 - \rho)M_n$, then $\varepsilon_n(\theta) = (R_n + (\rho_0 - \rho)M_n) [h_n(\varsigma) + R_n^{-1}\varepsilon_n + (\lambda_0 - \lambda)G_n R_n^{-1}\varepsilon_n]$ where $h_n(\varsigma) = X_n(\beta_0 - \beta) + (\lambda_0 - \lambda)G_n X_n \beta_0$ and $\varsigma = (\lambda, \beta)'$.

More explicitly, $\varepsilon_n(\theta) = R_n h_n(\varsigma) + (\rho_0 - \rho)M_n h_n(\varsigma) + \varepsilon_n + (\lambda_0 - \lambda)\bar{G}_n \varepsilon_n + (\rho_0 - \rho)H_n \varepsilon_n + (\rho_0 -$

$\rho)(\lambda_0 - \lambda)M_n G_n R_n^{-1} \varepsilon_n$, where $H_n = M_n R_n^{-1}$ and $\bar{G}_n = R_n G_n R_n^{-1}$. Hence,

$$\begin{aligned} \varepsilon_n'(\theta) \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn} \right) \varepsilon_n(\theta) &= \left(h_n'(\varsigma) R_n' + (\rho_0 - \rho) h_n'(\varsigma) M_n' \right) \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn} \right) (R_n h_n(\varsigma) + (\rho_0 - \rho) \\ &\quad \times M_n h_n(\varsigma)) + \left(h_n'(\varsigma) R_n' + (\rho_0 - \rho) h_n'(\varsigma) M_n' \right) \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn} \right) (\varepsilon_n + (\lambda_0 - \lambda) \bar{G}_n \varepsilon_n + (\rho_0 - \rho) H_n \varepsilon_n \\ &\quad + (\rho_0 - \rho)(\lambda_0 - \lambda) M_n G_n R_n^{-1} \varepsilon_n) + (\varepsilon_n' + (\lambda_0 - \lambda) \varepsilon_n' \bar{G}_n' + (\rho_0 - \rho) \varepsilon_n' H_n' + (\rho_0 - \rho)(\lambda_0 - \lambda) \\ &\quad \times \varepsilon_n' R_n'^{-1} G_n' M_n') \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn} \right) (R_n h_n(\varsigma) + (\rho_0 - \rho) M_n h_n(\varsigma)) + (\varepsilon_n' + (\lambda_0 - \lambda) \varepsilon_n' \bar{G}_n' + (\rho_0 - \rho) \\ &\quad \times \varepsilon_n' H_n' + (\rho_0 - \rho)(\lambda_0 - \lambda) \varepsilon_n' R_n'^{-1} G_n' M_n') \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn} \right) (\varepsilon_n + (\lambda_0 - \lambda) \bar{G}_n \varepsilon_n + (\rho_0 - \rho) H_n \varepsilon_n \\ &\quad + (\rho_0 - \rho)(\lambda_0 - \lambda) M_n G_n R_n^{-1} \varepsilon_n). \end{aligned}$$

For notational simplifications define $l_n(\theta)$ and $q_n(\theta)$ in the following way.

$$\begin{aligned} l_n(\theta) &= \left(h_n'(\varsigma) R_n' + (\rho_0 - \rho) h_n'(\varsigma) M_n' \right) \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) (\varepsilon_n + (\lambda_0 - \lambda) \bar{G}_n \varepsilon_n + (\rho_0 - \rho) H_n \varepsilon_n \\ &\quad + (\rho_0 - \rho)(\lambda_0 - \lambda) M_n G_n R_n^{-1} \varepsilon_n), \end{aligned}$$

where $P_{jn}^s = P_{jn} + P_{jn}'$, and

$$\begin{aligned} q_n(\theta) &= \left(\varepsilon_n' + (\lambda_0 - \lambda) \varepsilon_n' \bar{G}_n' + (\rho_0 - \rho) \varepsilon_n' H_n' + (\rho_0 - \rho)(\lambda_0 - \lambda) \varepsilon_n' R_n'^{-1} G_n' M_n' \right) \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn} \right) \\ &\quad \times (\varepsilon_n + (\lambda_0 - \lambda) \bar{G}_n \varepsilon_n + (\rho_0 - \rho) H_n \varepsilon_n + (\rho_0 - \rho)(\lambda_0 - \lambda) M_n G_n R_n^{-1} \varepsilon_n). \end{aligned}$$

More compactly,

$$\begin{aligned} \varepsilon_n'(\theta) \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn} \right) \varepsilon_n(\theta) &= \left(h_n'(\varsigma) R_n' + (\rho_0 - \rho) h_n'(\varsigma) M_n' \right) \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn} \right) (R_n h_n(\varsigma) \\ &\quad + (\rho_0 - \rho) M_n h_n(\varsigma)) + l_n(\theta) + q_n(\theta). \end{aligned}$$

Notice that $h_n'(\varsigma) R_n' = (\beta_0 - \beta)' X_n' R_n' + (\lambda_0 - \lambda)(X_n \beta_0)' G_n' R_n'$ and $h_n'(\varsigma) M_n' = (\beta_0 - \beta)' X_n' M_n' +$

$(\lambda_0 - \lambda)(X_n \beta_0)' G_n' M_n'$. By expansion,

$$\begin{aligned}
\frac{1}{n} l_n(\theta) &= (\lambda_0 - \lambda) \frac{1}{n} (X_n \beta_0)' G_n' R_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) \varepsilon_n + (\beta_0 - \beta)' \frac{1}{n} X_n' R_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) \varepsilon_n \\
&+ (\rho_0 - \rho)(\lambda_0 - \lambda) \frac{1}{n} (X_n \beta_0)' G_n' M_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) \varepsilon_n + (\rho_0 - \rho)(\beta_0 - \beta)' \frac{1}{n} X_n' M_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) \varepsilon_n \\
&+ (\lambda_0 - \lambda)^2 \frac{1}{n} (X_n \beta_0)' G_n' R_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) \bar{G}_n \varepsilon_n + (\lambda_0 - \lambda)(\beta_0 - \beta)' \frac{1}{n} X_n' R_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) \bar{G}_n \varepsilon_n \\
&+ (\rho_0 - \rho)(\lambda_0 - \lambda)^2 \frac{1}{n} (X_n \beta_0)' G_n' M_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) \bar{G}_n \varepsilon_n + (\rho_0 - \rho)(\lambda_0 - \lambda)(\beta_0 - \beta)' \frac{1}{n} X_n' M_n' \\
&\times \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) \bar{G}_n \varepsilon_n + (\rho_0 - \rho)(\lambda_0 - \lambda) \frac{1}{n} (X_n \beta_0)' G_n' R_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) H_n \varepsilon_n + (\rho_0 - \rho)(\beta_0 - \beta)' \\
&\times \frac{1}{n} X_n' R_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) H_n \varepsilon_n + (\rho_0 - \rho)^2 (\lambda_0 - \lambda) \frac{1}{n} (X_n \beta_0)' G_n' M_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) H_n \varepsilon_n \\
&+ (\rho_0 - \rho)^2 (\beta_0 - \beta)' \frac{1}{n} X_n' M_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) H_n \varepsilon_n + (\rho_0 - \rho)(\lambda_0 - \lambda)^2 \frac{1}{n} (X_n \beta_0)' G_n' R_n' \\
&\times \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) M_n G_n R_n^{-1} \varepsilon_n + (\rho_0 - \rho)(\lambda_0 - \lambda)(\beta_0 - \beta)' \frac{1}{n} X_n' R_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) M_n G_n R_n^{-1} \varepsilon_n \\
&+ (\rho_0 - \rho)^2 (\lambda_0 - \lambda)^2 \frac{1}{n} (X_n \beta_0)' G_n' M_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) M_n G_n R_n^{-1} \varepsilon_n + (\rho_0 - \rho)^2 (\lambda_0 - \lambda)(\beta_0 - \beta)' \frac{1}{n} \\
&\times X_n' M_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) M_n G_n R_n^{-1} \varepsilon_n. \tag{C.1}
\end{aligned}$$

Each matrix in the above expansion is uniformly bounded. Thus, applying *Lemma 2 (5)*, all the terms on the r.h.s. of (C.1) converge in probability to zero.

Hence, $\frac{1}{n}l_n(\theta) = o_p(1)$ uniformly in $\theta \in \Theta$.⁴⁴ Similarly,

$$\begin{aligned}
\frac{1}{n}q_n(\theta) &= \frac{1}{n}\varepsilon_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn} \right) \varepsilon_n + (\lambda_0 - \lambda) \frac{1}{n} \varepsilon_n' \bar{G}_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) \varepsilon_n + (\rho_0 - \rho) \frac{1}{n} \varepsilon_n' H_n' \\
&\times \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) \varepsilon_n + (\rho_0 - \rho)(\lambda_0 - \lambda) \frac{1}{n} \varepsilon_n' R_n'^{-1} G_n' M_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) \varepsilon_n + (\lambda_0 - \lambda)(\rho_0 - \rho) \\
&\times \frac{1}{n} \varepsilon_n' \bar{G}_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) H_n \varepsilon_n + (\lambda_0 - \lambda)^2 (\rho_0 - \rho) \frac{1}{n} \varepsilon_n' \bar{G}_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) M_n G_n R_n^{-1} \varepsilon_n \\
&+ (\rho_0 - \rho)^2 (\lambda_0 - \lambda) \frac{1}{n} \varepsilon_n' H_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) M_n G_n R_n^{-1} \varepsilon_n + (\lambda_0 - \lambda)^2 \frac{1}{n} \varepsilon_n' \bar{G}_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn} \right) \\
&\times \bar{G}_n \varepsilon_n + (\rho_0 - \rho)^2 \frac{1}{n} \varepsilon_n' H_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn} \right) H_n \varepsilon_n + (\lambda_0 - \lambda)^2 (\rho_0 - \rho)^2 \varepsilon_n' R_n'^{-1} G_n' M_n' \\
&\times \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn} \right) M_n G_n R_n^{-1} \varepsilon_n. \tag{C.2}
\end{aligned}$$

By *Lemma 2 (4)*, the variance of the first term on the r.h.s. of (C.2) has order $O(\frac{1}{n})$. Then, by generalized Chebyshev inequality, this term converges in probability to zero since $\frac{1}{n} \sum_{j=1}^m \Psi_{i,nj} E[\varepsilon_n' P_{jn} \varepsilon_n] = \frac{1}{n} \sum_{j=1}^m \Psi_{i,nj} \text{tr}(\Sigma_n P_{jn}) = 0$ for $P_{jn} \in \mathcal{P}_2$, $j = 1, \dots, m$. All others term has the same structure. Therefore the same asymptotic argument applies. Here, we just provide the asymptotic argument for a generic term. Again by *Lemma 2 (4)*, the variance of all the remaining terms have order $O(\frac{1}{n})$. Consider the second term of the r.h.s. of (C.2) and denote it with $\zeta(\theta)$. Taking expectation yields

$$\begin{aligned}
E[\zeta(\theta)] &= E \left[(\lambda_0 - \lambda) \frac{1}{n} \varepsilon_n' \bar{G}_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) \varepsilon_n \right] = (\lambda_0 - \lambda) \frac{1}{n} \left(\sum_{j=1}^m \Psi_{i,nj} \right) E \left[\text{tr} \left(\varepsilon_n' \bar{G}_n' P_{jn}^s \varepsilon_n \right) \right] \\
&= (\lambda_0 - \lambda) \frac{1}{n} \left(\sum_{j=1}^m \Psi_{i,nj} \right) \text{tr} \left(\Sigma_n \bar{G}_n' P_{jn}^s \right).
\end{aligned}$$

Let $\epsilon > 0$ be any real number and $P(\cdot)$ be the probability measure. Then, by the generalized Chebyshev inequality, we have

$$P \left(\left| (\lambda_0 - \lambda) \frac{1}{n} \varepsilon_n' \bar{G}_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) \varepsilon_n - E \left[(\lambda_0 - \lambda) \frac{1}{n} \varepsilon_n' \bar{G}_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) \varepsilon_n \right] \right| > \epsilon \right) \leq \frac{\text{var}(\zeta(\theta))}{\epsilon^2}.$$

By *Lemma 2 (4)*, $\text{var}(\zeta(\theta))$ is $O(\frac{1}{n})$. As $n \rightarrow \infty$, the r.h.s of the above equation converges to zero. Thus, we have

$$(\lambda_0 - \lambda) \frac{1}{n} \varepsilon_n' \bar{G}_n' \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn}^s \right) \varepsilon_n = (\lambda_0 - \lambda) \frac{1}{n} \left(\sum_{j=1}^m \Psi_{i,nj} \right) \text{tr}(\Sigma_n \bar{G}_n' P_{jn}^s) + o_p(1).$$

⁴⁴The uniform convergence in θ follows since Θ is a compact set and $l_n(\theta)$ is a quadratic and continuous function in θ . Thus, $l_n(\theta)$ has a bounded range. From this observation, the uniform convergence follows from Lemma 2.4 of Newey and McFadden (1994, see p. 2129). The same argument can also be seen in the proof of Proposition 1 in Lee (2007a)

Applying the same asymptotic argument to the remaining terms, the equation in (C.2) simplifies to

$$\begin{aligned}
\frac{1}{n}q_n(\theta) &= (\lambda_0 - \lambda) \frac{1}{n} \sum_{j=1}^m \Psi_{i,nj} \text{tr}(\Sigma_n \bar{G}'_n P_{jn}^s) + (\rho_0 - \rho) \frac{1}{n} \sum_{j=1}^m \Psi_{i,nj} \text{tr}(\Sigma_n H'_n P_{jn}^s) + (\rho_0 - \rho)(\lambda_0 - \lambda) \\
&\times \frac{1}{n} \sum_{j=1}^m \Psi_{i,nj} \text{tr}(\Sigma_n P_{jn}^s M_n G_n R_n^{-1}) + (\lambda_0 - \lambda)(\rho_0 - \rho) \frac{1}{n} \sum_{j=1}^m \Psi_{i,nj} \text{tr}(\Sigma_n \bar{G}'_n P_{jn}^s H_n) + (\lambda_0 - \lambda)^2(\rho_0 - \rho) \\
&\times \frac{1}{n} \sum_{j=1}^m \Psi_{i,nj} \text{tr}(\Sigma_n \bar{G}'_n P_{jn}^s M_n G_n R_n^{-1}) + (\lambda_0 - \lambda)(\rho_0 - \rho)^2 \frac{1}{n} \sum_{j=1}^m \Psi_{i,nj} \text{tr}(\Sigma_n H'_n P_{jn}^s M_n G_n R_n^{-1}) \\
&+ (\lambda_0 - \lambda)^2 \frac{1}{n} \sum_{j=1}^m \Psi_{i,nj} \text{tr}(\Sigma_n \bar{G}'_n P_{jn}^s \bar{G}_n) + (\rho_0 - \rho)^2 \frac{1}{n} \sum_{j=1}^m \Psi_{i,nj} \text{tr}(\Sigma_n H'_n P_{jn}^s H_n) + (\lambda_0 - \lambda)^2(\rho_0 - \rho)^2 \\
&\times \frac{1}{n} \sum_{j=1}^m \Psi_{i,nj} \text{tr}(\Sigma_n R_n'^{-1} G'_n M_n' P_{jn} M_n G_n R_n^{-1}) + o_p(1), \tag{C.3}
\end{aligned}$$

uniformly in $\theta \in \Theta$. The uniform convergence in θ follows since Θ is a compact set and $q_n(\theta)$ is quadratic and continuous function in θ . Hence,

$$\begin{aligned}
\frac{1}{n} \varepsilon'_n(\theta) \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn} \right) \varepsilon_n(\theta) &= \frac{1}{n} \left(h'_n(\zeta) R'_n + (\rho_0 - \rho) h'_n(\zeta) M_n' \right) \left(\sum_{j=1}^m \Psi_{i,nj} P_{jn} \right) (R_n h_n(\zeta) \\
&+ (\rho_0 - \rho) M_n h_n(\zeta)) + (\lambda_0 - \lambda) \frac{1}{n} \sum_{j=1}^m \Psi_{i,nj} \text{tr}(\Sigma_n P_{jn}^s \bar{G}_n) + (\rho_0 - \rho) \frac{1}{n} \sum_{j=1}^m \Psi_{i,nj} \text{tr}(\Sigma_n P_{jn}^s H_n) \\
&+ (\rho_0 - \rho)(\lambda_0 - \lambda) \frac{1}{n} \sum_{j=1}^m \Psi_{i,nj} \text{tr}(\Sigma_n P_{jn}^s M_n G_n R_n^{-1}) + (\lambda_0 - \lambda)(\rho_0 - \rho) \frac{1}{n} \sum_{j=1}^m \Psi_{i,nj} \text{tr}(\Sigma_n H'_n P_{jn}^s \bar{G}_n) \\
&+ (\lambda_0 - \lambda)^2(\rho_0 - \rho) \frac{1}{n} \sum_{j=1}^m \Psi_{i,nj} \text{tr}(\Sigma_n \bar{G}'_n P_{jn}^s M_n G_n R_n^{-1}) + (\rho_0 - \rho)^2(\lambda_0 - \lambda) \frac{1}{n} \sum_{j=1}^m \Psi_{i,nj} \\
&\times \text{tr}(\Sigma_n H'_n P_{jn}^s M_n G_n R_n^{-1}) + (\lambda_0 - \lambda)^2 \frac{1}{n} \sum_{j=1}^m \Psi_{i,nj} \text{tr}(\Sigma_n \bar{G}'_n P_{jn}^s \bar{G}_n) + (\rho_0 - \rho)^2 \frac{1}{n} \sum_{j=1}^m \Psi_{i,nj} \\
&\times \text{tr}(\Sigma_n H'_n P_{jn}^s H_n) + (\lambda_0 - \lambda)^2(\rho_0 - \rho)^2 \frac{1}{n} \sum_{j=1}^m \Psi_{i,nj} \text{tr}(\Sigma_n R_n'^{-1} G'_n M_n' P_{jn} M_n G_n R_n^{-1}) + o_p(1), \tag{C.4}
\end{aligned}$$

uniformly in $\theta \in \Theta$. The r.h.s of the above equation is simply expectation of the term in the l.h.s. The above relation holds for all i . Therefore, $\frac{1}{n} \Psi_n g_n(\theta)$ converges to $\frac{1}{n} E(\Psi_n g_n(\theta))$ uniformly in $\theta \in \Theta$. By the identification condition (*Assumption 6*) and the above uniform convergence result, GMM estimator $\hat{\theta}_n$ is consistent.

Next, we show the asymptotic normality of the GMM estimator $\hat{\theta}_n$. The first order condition implies that $\frac{\partial g_n'(\hat{\theta}_n)}{\partial \theta} \Psi_n' \Psi_n g_n(\hat{\theta}_n) = 0$. By the mean value theorem at $\bar{\theta}_n$, we have

$$\begin{aligned}
\hat{\theta}_n &= \theta_0 - \left(\frac{\partial g_n'(\hat{\theta}_n)}{\partial \theta} \Psi_n' \Psi_n \frac{\partial g_n(\bar{\theta}_n)}{\partial \theta'} \right)^{-1} \frac{\partial g_n'(\hat{\theta}_n)}{\partial \theta} \Psi_n' \Psi_n g_n(\theta_0) \\
\sqrt{n}(\hat{\theta}_n - \theta_0) &= - \left(\frac{1}{n} \frac{\partial g_n'(\hat{\theta}_n)}{\partial \theta} \Psi_n' \Psi_n \frac{1}{n} \frac{\partial g_n(\bar{\theta}_n)}{\partial \theta'} \right)^{-1} \frac{1}{n} \frac{\partial g_n'(\hat{\theta}_n)}{\partial \theta} \Psi_n' \frac{1}{\sqrt{n}} \Psi_n g_n(\theta_0), \tag{C.5}
\end{aligned}$$

where

$$\frac{1}{n} \frac{\partial g_n(\theta)}{\partial \theta'} = \frac{1}{n} \begin{pmatrix} \varepsilon_n'(\theta) P_{1n}^s \frac{\partial \varepsilon_n(\theta)}{\partial \theta'} \\ \varepsilon_n'(\theta) P_{2n}^s \frac{\partial \varepsilon_n(\theta)}{\partial \theta'} \\ \vdots \\ \varepsilon_n'(\theta) P_{mn}^s \frac{\partial \varepsilon_n(\theta)}{\partial \theta'} \\ Q_n' \frac{\partial \varepsilon_n(\theta)}{\partial \theta'} \end{pmatrix}. \quad (\text{C.6})$$

We will show that $\frac{1}{n} \frac{\partial g_n(\theta)}{\partial \theta'}$ converges to $\frac{1}{n} E\left(\frac{\partial g_n(\theta)}{\partial \theta'}\right)$ uniformly in $\theta \in \Theta$. Given this result and the fact that $\frac{1}{n} E\left(\frac{\partial g_n(\theta)}{\partial \theta'}\right)$ is continuous in θ and $\text{plim}_{n \rightarrow \infty} \hat{\theta}_n = \theta_0$, we have $\frac{1}{n} \frac{\partial g_n(\hat{\theta}_n)}{\partial \theta'} = \frac{1}{n} E\left(\frac{\partial g_n(\theta_0)}{\partial \theta'}\right) + o_p(1) = -\frac{1}{n} \Gamma_n + o_p(1)$ (Amemiya, 1985, Theorem 4.1.5, p.113).

Since $\frac{\partial \varepsilon_n(\theta)}{\partial \theta'} = -(M_n S_n(\lambda) Y_n - M_n X_n \beta, R_n(\rho) W_n Y_n, R_n(\rho) X_n)$, the gradient in (C.6) can be written as

$$\frac{1}{n} \frac{\partial g_n(\theta)}{\partial \theta'} = -\frac{1}{n} \begin{pmatrix} \varepsilon_n'(\theta) P_{1n}^s (M_n S_n(\lambda) Y_n - M_n X_n \beta) & \varepsilon_n'(\theta) P_{1n}^s R_n(\rho) W_n Y_n & \varepsilon_n'(\theta) P_{1n}^s R_n(\rho) X_n \\ \varepsilon_n'(\theta) P_{2n}^s (M_n S_n(\lambda) Y_n - M_n X_n \beta) & \varepsilon_n'(\theta) P_{2n}^s R_n(\rho) W_n Y_n & \varepsilon_n'(\theta) P_{2n}^s R_n(\rho) X_n \\ \vdots & \vdots & \vdots \\ \varepsilon_n'(\theta) P_{mn}^s (M_n S_n(\lambda) Y_n - M_n X_n \beta) & \varepsilon_n'(\theta) P_{mn}^s R_n(\rho) W_n Y_n & \varepsilon_n'(\theta) P_{mn}^s R_n(\rho) X_n \\ Q_n' (M_n S_n(\lambda) Y_n - M_n X_n \beta) & Q_n' R_n(\rho) W_n Y_n & Q_n' R_n(\rho) X_n \end{pmatrix}.$$

The probability limit of the above gradient is evaluated below. By using $S_n(\lambda) = S_n + (\lambda_0 - \lambda) W_n$ and $R_n(\rho) = R_n + (\rho_0 - \rho) M_n$ equalities, the rows of the above gradient involving P_{jn} is given as

$$\begin{aligned} \frac{1}{n} \left(\frac{\partial g_n(\theta)}{\partial \theta'} \right)_{j,n} &= -\frac{1}{n} (\varepsilon_n'(\theta) P_{jn}^s M_n S_n Y_n + (\lambda_0 - \lambda) \varepsilon_n'(\theta) P_{jn}^s M_n W_n Y_n - \varepsilon_n'(\theta) P_{jn}^s M_n X_n \beta, \\ &\quad \varepsilon_n'(\theta) P_{jn}^s R_n W_n Y_n + (\rho_0 - \rho) \varepsilon_n'(\theta) P_{jn}^s M_n W_n Y_n, \varepsilon_n'(\theta) P_{jn}^s R_n X_n + (\rho_0 - \rho) \varepsilon_n'(\theta) P_{jn}^s M_n X_n). \end{aligned}$$

Each term of the above j th row is evaluated separately. One of the terms is $\frac{1}{n} \varepsilon_n'(\theta) P_{jn}^s R_n W_n Y_n = \frac{1}{n} \varepsilon_n'(\theta) P_{jn}^s R_n G_n X_n \beta_0 + \frac{1}{n} \varepsilon_n'(\theta) P_{jn}^s \bar{G}_n \varepsilon_n$. More explicitly, by substituting the expansion of $\varepsilon_n(\theta)$ into this term:

$$\begin{aligned} \frac{1}{n} \varepsilon_n'(\theta) P_{jn}^s R_n G_n X_n \beta_0 &= \frac{1}{n} (R_n h_n(\varsigma) + (\rho_0 - \rho) M_n h_n(\varsigma) + \varepsilon_n + (\lambda_0 - \lambda) \bar{G}_n \varepsilon_n + (\rho_0 - \rho) H_n \varepsilon_n + (\rho_0 - \rho) \\ &\quad \times (\lambda_0 - \lambda) M_n G_n R_n^{-1} \varepsilon_n)' P_{jn}^s R_n G_n X_n \beta_0 = \frac{1}{n} h_n'(\varsigma) R_n' P_{jn}^s R_n G_n X_n \beta_0 + \frac{1}{n} (\rho_0 - \rho) h_n'(\varsigma) M_n' P_{jn}^s \\ &\quad \times R_n G_n X_n \beta_0 + \frac{1}{n} \varepsilon_n' P_{jn}^s R_n G_n X_n \beta_0 + \frac{1}{n} (\lambda_0 - \lambda) \varepsilon_n' \bar{G}_n' P_{jn}^s R_n G_n X_n \beta_0 + \frac{1}{n} (\rho_0 - \rho) \varepsilon_n' H_n' P_{jn}^s R_n G_n X_n \beta_0 \\ &\quad + \frac{1}{n} (\rho_0 - \rho) (\lambda_0 - \lambda) \varepsilon_n' R_n^{-1} G_n' M_n' P_{jn}^s R_n G_n X_n \beta_0. \end{aligned} \quad (\text{C.7})$$

Notice that all terms except first two elements have the same structure; therefore, they are subject to the same asymptotic argument. By *Lemma 2 (5)* all terms except the first two terms vanish. Thus, $\frac{1}{n} \varepsilon_n'(\theta) P_{jn}^s R_n G_n X_n \beta_0 = \frac{1}{n} h_n'(\varsigma) R_n' P_{jn}^s R_n G_n X_n \beta_0 + \frac{1}{n} (\rho_0 - \rho) h_n'(\varsigma) M_n' P_{jn}^s R_n G_n X_n \beta_0 + o_p(1)$.

Similarly, by *Lemmas 2 (4)* and *(5)*, we have

$$\begin{aligned}
\frac{1}{n}\varepsilon_n'(\theta)P_{jn}^s\bar{G}_n\varepsilon_n &= \frac{1}{n}(R_nh_n(\varsigma) + (\rho_0 - \rho)M_nh_n(\varsigma) + \varepsilon_n + (\lambda_0 - \lambda)\bar{G}_n\varepsilon_n + (\rho_0 - \rho)H_n\varepsilon_n + (\rho_0 - \rho) \\
&\times (\lambda_0 - \lambda)M_nG_nR_n^{-1}\varepsilon_n)'P_{jn}^s\bar{G}_n\varepsilon_n = \frac{1}{n}h_n'(\varsigma)R_n'P_{jn}^s\bar{G}_n\varepsilon_n + \frac{1}{n}(\rho_0 - \rho)h_n'(\varsigma)M_n'P_{jn}^s\bar{G}_n\varepsilon_n + \frac{1}{n}\varepsilon_n'P_{jn}^s\bar{G}_n\varepsilon_n \\
&+ \frac{1}{n}(\lambda_0 - \lambda)\varepsilon_n'\bar{G}_n'P_{jn}^s\bar{G}_n\varepsilon_n + \frac{1}{n}(\rho_0 - \rho)\varepsilon_n'H_n'P_{jn}^s\bar{G}_n\varepsilon_n + \frac{1}{n}(\rho_0 - \rho)(\lambda_0 - \lambda)\varepsilon_n'R_n^{-1}G_n'M_n'P_{jn}^s\bar{G}_n\varepsilon_n \\
&= \frac{1}{n}tr(\Sigma_nP_{jn}^s\bar{G}_n) + \frac{1}{n}(\lambda_0 - \lambda)tr(\Sigma_n\bar{G}_n'P_{jn}^s\bar{G}_n) + \frac{1}{n}(\rho_0 - \rho)tr(\Sigma_nH_n'P_{jn}^s\bar{G}_n) + \frac{1}{n}(\rho_0 - \rho)(\lambda_0 - \lambda) \\
&\times tr(\Sigma_n\bar{G}_n'P_{jn}^sM_nG_nR_n^{-1}) + o_p(1),
\end{aligned} \tag{C.8}$$

uniformly in $\theta \in \Theta$. Combining these results, we get

$$\begin{aligned}
\frac{1}{n}\varepsilon_n'(\theta)P_{jn}^sR_nW_nY_n &= \frac{1}{n}h_n'(\varsigma)R_n'P_{jn}^sR_nG_nX_n\beta_0 + \frac{1}{n}(\rho_0 - \rho)h_n'(\varsigma)M_n'P_{jn}^sR_nG_nX_n\beta_0 \\
&+ \frac{1}{n}tr(\Sigma_nP_{jn}^s\bar{G}_n) + \frac{1}{n}(\lambda_0 - \lambda)tr(\Sigma_n\bar{G}_n'P_{jn}^s\bar{G}_n) + \frac{1}{n}(\rho_0 - \rho)tr(\Sigma_nH_n'P_{jn}^s\bar{G}_n) \\
&+ \frac{1}{n}(\rho_0 - \rho)(\lambda_0 - \lambda)tr(\Sigma_n\bar{G}_n'P_{jn}^sM_nG_nR_n^{-1}) + o_p(1),
\end{aligned} \tag{C.9}$$

uniformly in $\theta \in \Theta$. Since $h_n(\varsigma_0) = 0$ at θ_0 , $\frac{1}{n}\varepsilon_n'(\theta_0)P_{jn}^sR_nW_nY_n = \frac{1}{n}tr(\Sigma_nP_{jn}^s\bar{G}_n) + o_p(1)$.

Now we turn to another element in the j th row. A similar analysis applies to $\frac{1}{n}\varepsilon_n'(\theta)P_{jn}^sM_nS_nY_n = \frac{1}{n}\varepsilon_n'(\theta)P_{jn}^sM_nX_n\beta_0 + \frac{1}{n}\varepsilon_n'(\theta)P_{jn}^sM_nR_n^{-1}\varepsilon_n$. By substituting the expansion of $\varepsilon_n(\theta)$ in this terms, we get

$$\begin{aligned}
\frac{1}{n}\varepsilon_n'(\theta)P_{jn}^sM_nX_n\beta_0 &= \frac{1}{n}(R_nh_n(\varsigma) + (\rho_0 - \rho)M_nh_n(\varsigma) + \varepsilon_n + (\lambda_0 - \lambda)\bar{G}_n\varepsilon_n + (\rho_0 - \rho)H_n\varepsilon_n + (\rho_0 - \rho) \\
&\times (\lambda_0 - \lambda)M_nG_nR_n^{-1}\varepsilon_n)'P_{jn}^sM_nX_n\beta_0 = \frac{1}{n}h_n'(\varsigma)R_n'P_{jn}^sM_nX_n\beta_0 + \frac{1}{n}(\rho_0 - \rho)h_n'(\varsigma)M_n'P_{jn}^sM_nX_n\beta_0 \\
&+ \frac{1}{n}\varepsilon_n'P_{jn}^sM_nX_n\beta_0 + \frac{1}{n}(\lambda_0 - \lambda)\varepsilon_n'\bar{G}_n'P_{jn}^sM_nX_n\beta_0 + \frac{1}{n}(\rho_0 - \rho)\varepsilon_n'H_n'P_{jn}^sM_nX_n\beta_0 + \frac{1}{n}(\rho_0 - \rho) \\
&\times (\lambda_0 - \lambda)\varepsilon_n'R_n^{-1}G_n'M_n'P_{jn}^sM_nX_n\beta_0 = \frac{1}{n}h_n'(\varsigma)R_n'P_{jn}^sM_nX_n\beta_0 + \frac{1}{n}(\rho_0 - \rho)h_n'(\varsigma)M_n'P_{jn}^sM_nX_n\beta_0 \\
&+ o_p(1),
\end{aligned} \tag{C.10}$$

uniformly in $\theta \in \Theta$ by *Lemmas 2 (4)* and *(5)*. Similarly,

$$\begin{aligned}
\frac{1}{n}\varepsilon_n'(\theta)P_{jn}^sM_nR_n^{-1}\varepsilon_n &= \frac{1}{n}(R_nh_n(\varsigma) + (\rho_0 - \rho)M_nh_n(\varsigma) + \varepsilon_n + (\lambda_0 - \lambda)\bar{G}_n\varepsilon_n + (\rho_0 - \rho)H_n\varepsilon_n \\
&+ (\rho_0 - \rho)(\lambda_0 - \lambda)M_nG_nR_n^{-1}\varepsilon_n)'P_{jn}^sM_nR_n^{-1}\varepsilon_n = \frac{1}{n}h_n'(\varsigma)R_n'P_{jn}^sM_nR_n^{-1}\varepsilon_n + \frac{1}{n}(\rho_0 - \rho)h_n'(\varsigma) \\
&\times M_n'P_{jn}^sM_nR_n^{-1}\varepsilon_n + \frac{1}{n}\varepsilon_n'P_{jn}^sM_nR_n^{-1}\varepsilon_n + \frac{1}{n}(\lambda_0 - \lambda)\varepsilon_n'\bar{G}_n'P_{jn}^sM_nR_n^{-1}\varepsilon_n + \frac{1}{n}(\rho_0 - \rho)\varepsilon_n'H_n'P_{jn}^s \\
&\times M_nR_n^{-1}\varepsilon_n + \frac{1}{n}(\rho_0 - \rho)(\lambda_0 - \lambda)\varepsilon_n'R_n^{-1}G_n'M_n'P_{jn}^sM_nR_n^{-1}\varepsilon_n = \frac{1}{n}tr(\Sigma_nP_{jn}^sM_nR_n^{-1}) \\
&+ \frac{1}{n}(\lambda_0 - \lambda)tr(\Sigma_n\bar{G}_n'P_{jn}^sM_nR_n^{-1}) + \frac{1}{n}(\rho_0 - \rho)tr(\Sigma_nH_n'P_{jn}^sM_nR_n^{-1}) + \frac{1}{n}(\rho_0 - \rho)(\lambda_0 - \lambda) \\
&\times tr(\Sigma_nR_n^{-1}G_n'M_n'P_{jn}^sM_nR_n^{-1}) + o_p(1),
\end{aligned} \tag{C.11}$$

uniformly in $\theta \in \Theta$. Since $H_n = M_n R_n^{-1}$,

$$\begin{aligned} \frac{1}{n} \varepsilon_n'(\theta) P_{jn}^s M_n R_n^{-1} \varepsilon_n &= \frac{1}{n} \text{tr}(\Sigma_n P_{jn}^s H_n) + \frac{1}{n} (\lambda_0 - \lambda) \text{tr}(\Sigma_n \bar{G}_n' P_{jn}^s H_n) + \frac{1}{n} (\rho_0 - \rho) \\ &\quad \times \text{tr}(\Sigma_n H_n' P_{jn}^s H_n) + \frac{1}{n} (\rho_0 - \rho) (\lambda_0 - \lambda) \text{tr}(\Sigma_n H_n' P_{jn}^s M_n G_n R_n^{-1}) + o_p(1). \end{aligned} \quad (\text{C.12})$$

Then, combining the above results, we get

$$\begin{aligned} \frac{1}{n} \varepsilon_n'(\theta) P_{jn}^s M_n S_n Y_n &= \frac{1}{n} h_n'(\varsigma) R_n' P_{jn}^s M_n X_n \beta_0 + \frac{1}{n} (\rho_0 - \rho) h_n'(\varsigma) M_n' P_{jn}^s M_n X_n \beta_0 \\ &\quad + \frac{1}{n} \text{tr}(\Sigma_n P_{jn}^s H_n) + \frac{1}{n} (\lambda_0 - \lambda) \text{tr}(\Sigma_n \bar{G}_n' P_{jn}^s H_n) + \frac{1}{n} (\rho_0 - \rho) \text{tr}(\Sigma_n H_n' P_{jn}^s H_n) + \frac{1}{n} (\rho_0 - \rho) \\ &\quad \times (\lambda_0 - \lambda) \text{tr}(\Sigma_n H_n' P_{jn}^s M_n G_n R_n^{-1}) + o_p(1), \end{aligned} \quad (\text{C.13})$$

uniformly in $\theta \in \Theta$. Since $h_n(\varsigma_0) = 0$ at θ_0 , $\frac{1}{n} \varepsilon_n'(\theta_0) P_{jn}^s M_n S_n Y_n = \frac{1}{n} \text{tr}(\Sigma_n P_{jn}^s H_n) + o_p(1)$.

With the same line of argument, at θ_0 , we have $\frac{1}{n} \varepsilon_n'(\theta_0) P_{jn}^s R_n X_n = o_p(1)$ and $\frac{1}{n} \varepsilon_n'(\theta_0) P_{jn}^s M_n X_n \beta_0 = o_p(1)$. All the remaining terms in the j th row vanishes when evaluated at the true parameter value.

Now, we return $\frac{1}{n} Q_n' \frac{\partial \varepsilon_n(\theta)}{\partial \theta'}$ in (C.6). This term can be written as

$$\begin{aligned} \frac{1}{n} Q_n' \frac{\partial \varepsilon_n(\theta)}{\partial \theta'} &= \frac{1}{n} (Q_n' M_n S_n Y_n + (\lambda_0 - \lambda) Q_n' M_n W_n Y_n - Q_n' M_n X_n \beta, Q_n' R_n W_n Y_n \\ &\quad + (\rho_0 - \rho) Q_n' M_n W_n Y_n, Q_n' R_n X_n + (\rho_0 - \rho) Q_n' M_n X_n). \end{aligned}$$

The first term in the r.h.s of the above equation vanishes when evaluated at the true parameter θ_0 . For the second term, we have $\frac{1}{n} Q_n' R_n W_n Y_n = \frac{1}{n} Q_n' R_n G_n X_n \beta_0 + \frac{1}{n} Q_n' \bar{G}_n \varepsilon_n = \frac{1}{n} Q_n' R_n G_n X_n \beta_0 + o_p(1)$. Likewise the last term converges to $Q_n' R_n X_n$.

Combining all the previous results, we get the relation $\frac{1}{n} \frac{\partial g_n(\hat{\theta}_n)}{\partial \theta'} = -\frac{1}{n} \Gamma_n + o_p(1)$ uniformly in θ , where Γ_n is given in (4.25). By CLT in *Theorem 1* of Kelejian and Prucha (2001), $\frac{1}{\sqrt{n}} \Psi_n g_n(\theta_0) = \frac{1}{\sqrt{n}} [\varepsilon_n' (\sum_j^m \Psi_{nj} P_{jn}) \varepsilon_n + \Psi_{nx} Q_n' \varepsilon_n] \xrightarrow{d} N(0, \lim_{n \rightarrow \infty} \frac{1}{n} \Psi_n \Omega_n \Psi_n')$. The asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ in (4.26) now follows from (C.5) by the Slutsky theorem. $\square$

Proof of Proposition 2. We first show the consistency of $\frac{1}{n} \hat{\Omega}_n$ by showing that each element in $\frac{1}{n} (\hat{\Omega}_n - \Omega_n)$ is $o_p(1)$. Notice that some of the elements $\frac{1}{n} \Omega_n$ are of the form: $\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \sigma_{ni}^2 \sigma_{nj}^2$, where $P_{\Delta n, ij} = P_{an, ij} (P_{bn, ij} + P_{bn, ji})$ by *Lemma 2 (1)*. Also, notice that $P_{\Delta n, ii} = 0$. Following the same steps of Lin and Lee (2010), we first show $\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \varepsilon_{ni}^2 \varepsilon_{nj}^2 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \sigma_{ni}^2 \sigma_{nj}^2 + o_p(1)$. Then, we show that this relation still holds, when $\hat{\varepsilon}_{ni}$ replaces ε_{ni} . As an initial step, we need to establish the uniform boundedness of $P_{\Delta n}$ in both the row and column sum norms. P_{bn} is uniformly bounded in both row and column sum norms and therefore its elements are uniformly bounded by *Assumption 4*. Hence, there exists a constant c such that $|P_{bn, ij} + P_{bn, ji}| \leq c$, for all i, j and n . This implies $|P_{\Delta n, ij}| \leq c |P_{an, ij}|$. Since P_{an} is bounded in both row and column sum norms, $P_{\Delta n}$ is uniformly bounded in both row and

the column sum norms.

By expansion, $\varepsilon_{ni}^2 \varepsilon_{nj}^2 - \sigma_{ni}^2 \sigma_{nj}^2 = (\varepsilon_{ni}^2 - \sigma_{ni}^2)(\varepsilon_{nj}^2 - \sigma_{nj}^2) + \sigma_{ni}^2(\varepsilon_{nj}^2 - \sigma_{nj}^2) + \sigma_{nj}^2(\varepsilon_{ni}^2 - \sigma_{ni}^2)$. Hence,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} (\varepsilon_{ni}^2 \varepsilon_{nj}^2 - \sigma_{ni}^2 \sigma_{nj}^2) &= \underbrace{\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} (\varepsilon_{ni}^2 - \sigma_{ni}^2)(\varepsilon_{nj}^2 - \sigma_{nj}^2)}_{A_n} \\ &+ \underbrace{\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \sigma_{ni}^2 (\varepsilon_{nj}^2 - \sigma_{nj}^2)}_{B_n} + \underbrace{\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \sigma_{nj}^2 (\varepsilon_{ni}^2 - \sigma_{ni}^2)}_{C_n}. \end{aligned} \quad (C.14)$$

First, we express A_n , B_n and C_n in terms of quadratic forms for notational simplification. To this end, let $u_n = (u_{n1}, \dots, u_{nn})'$ such that $u_{ni} = \varepsilon_{ni}^2 - \sigma_{ni}^2$ for $i = 1, \dots, n$ and let $\Sigma_{\sigma_n} = (\sigma_{n1}^2, \dots, \sigma_{nn}^2)$. Then, $A_n = \frac{1}{n} u_n' P_{\Delta n} u_n$, $B_n = \frac{1}{n} u_n' P_{\Delta n} \Sigma_{\sigma_n}'$, and $C_n = \frac{1}{n} \Sigma_{\sigma_n} P_{\Delta n} u_n$.

As $E(u_n' P_{\Delta n} u_n) = \text{tr}(P_{\Delta n} \Lambda_n)$ where $\Lambda_n = E(u_n u_n') = \text{diag}\{\mu_{n1}^{(4)} - \sigma_{n1}^4, \dots, \mu_{nn}^{(4)} - \sigma_{nn}^4\}$, where $\mu_{ni}^{(4)} = E(\varepsilon_{ni}^4)$. This implies $E(u_n' P_{\Delta n} u_n) = \text{tr}(P_{\Delta n} \Lambda_n) = 0$ since $P_{\Delta n, ii} = 0 \forall i$. By Lemma 2 (4), $\text{plim}_{n \rightarrow \infty} A_n = 0$. By Lemma 2 (5) and Assumption 1, $\text{plim}_{n \rightarrow \infty} B_n = 0$ and $\text{plim}_{n \rightarrow \infty} C_n = 0$. Hence, $\text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \varepsilon_{ni}^2 \varepsilon_{nj}^2 - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \sigma_{ni}^2 \sigma_{nj}^2 \right) = 0$.

Next, we will show that $\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \hat{\varepsilon}_{ni}^2 \hat{\varepsilon}_{nj}^2 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \varepsilon_{ni}^2 \varepsilon_{nj}^2 + o_p(1)$. By expansion, $\hat{\varepsilon}_{ni}^2 \hat{\varepsilon}_{nj}^2 - \varepsilon_{ni}^2 \varepsilon_{nj}^2 = (\hat{\varepsilon}_{ni}^2 - \varepsilon_{ni}^2)(\hat{\varepsilon}_{nj}^2 - \varepsilon_{nj}^2) + \varepsilon_{nj}^2(\hat{\varepsilon}_{ni}^2 - \varepsilon_{ni}^2) + \varepsilon_{ni}^2(\hat{\varepsilon}_{nj}^2 - \varepsilon_{nj}^2)$. Then,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} (\hat{\varepsilon}_{ni}^2 \hat{\varepsilon}_{nj}^2 - \varepsilon_{ni}^2 \varepsilon_{nj}^2) &= \underbrace{\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} (\hat{\varepsilon}_{ni}^2 - \varepsilon_{ni}^2)(\hat{\varepsilon}_{nj}^2 - \varepsilon_{nj}^2)}_{\vartheta_{n1}} \\ &+ \underbrace{\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \varepsilon_{nj}^2 (\hat{\varepsilon}_{ni}^2 - \varepsilon_{ni}^2)}_{\vartheta_{n2}} + \underbrace{\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \varepsilon_{ni}^2 (\hat{\varepsilon}_{nj}^2 - \varepsilon_{nj}^2)}_{\vartheta_{n3}}. \end{aligned} \quad (C.15)$$

From the model, we have $\hat{\varepsilon}_n = R_n(\hat{\rho})(S_n(\hat{\lambda})Y_n - X_n\hat{\beta})$. By using the relations $R_n(\hat{\rho}) = R_n + (\rho_0 - \hat{\rho})M_n$ and $S(\hat{\lambda}) = S_n + (\lambda_0 - \hat{\lambda})W_n$, we get

$$\begin{aligned} \hat{\varepsilon}_n &= [R_n + (\rho - \hat{\rho})M_n] [S_n Y_n + (\lambda_0 - \hat{\lambda})W_n Y_n - X_n \hat{\beta}] \\ &= [R_n + (\rho - \hat{\rho})M_n] [X_n(\beta_0 - \hat{\beta}) + (\lambda_0 - \hat{\lambda})G_n X_n \beta_0 + R_n^{-1} \varepsilon_n + (\lambda_0 - \hat{\lambda})G_n R_n^{-1} \varepsilon_n]. \end{aligned}$$

Let $h_n(\hat{\varsigma}) = X_n(\beta_0 - \hat{\beta}) + (\lambda_0 - \hat{\lambda})G_n X_n \beta_0$ where $\hat{\varsigma} = (\hat{\lambda}, \hat{\beta})'$. Hence,

$$\hat{\varepsilon}_n = \varepsilon_n + (R_n + (\rho - \hat{\rho})M_n)h_n(\hat{\varsigma}) + (\lambda_0 - \hat{\lambda})\bar{G}_n \varepsilon_n + (\rho_0 - \hat{\rho})H_n \varepsilon_n + (\lambda_0 - \hat{\lambda})(\rho_0 - \hat{\rho})M_n G_n R_n^{-1} \varepsilon_n.$$

Let $e_{i,n}$ be the i -th row of the $n \times n$ identity matrix. Then, in scalar form, $\hat{\varepsilon}_{ni} = \varepsilon_{ni} + a_{ni} + b_{ni} + c_{ni} + f_{ni}$, where $a_{ni} = e_{i,n} R_n h_n(\hat{\varsigma}) + (\rho_0 - \hat{\rho})e_{i,n} M_n h_n(\hat{\varsigma})$, $b_{ni} = (\lambda_0 - \hat{\lambda})(e_{i,n} \bar{G}_n \varepsilon_n)$, $c_{ni} = (\rho_0 -$

$\hat{\rho})(e_{i,n}H_n\varepsilon_n)$, and $f_{ni} = (\lambda_0 - \hat{\lambda})(\rho_0 - \hat{\rho})e_{i,n}M_nG_nR_n^{-1}\varepsilon_n$. Then, $\hat{\varepsilon}_{ni}^2 = (\varepsilon_{ni} + a_{ni} + b_{ni} + c_{ni} + f_{ni})^2 = \varepsilon_{ni}^2 + a_{ni}^2 + b_{ni}^2 + c_{ni}^2 + f_{ni}^2 + 2\varepsilon_{ni}a_{ni} + 2\varepsilon_{ni}b_{ni} + 2\varepsilon_{ni}c_{ni} + 2\varepsilon_{ni}f_{ni} + 2a_{ni}b_{ni} + 2a_{ni}c_{ni} + 2a_{ni}f_{ni} + 2b_{ni}c_{ni} + 2b_{ni}f_{ni} + 2c_{ni}f_{ni}$.

Next, we will evaluate all three terms ϑ_{nl} , $l = 1, 2, 3$ and show that they converge in probability to zero. First, consider ϑ_{n2} :

$$\vartheta_{n2} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \varepsilon_{nj}^2 (a_{ni}^2 + b_{ni}^2 + c_{ni}^2 + f_{ni}^2 + 2\varepsilon_{ni}a_{ni} + 2\varepsilon_{ni}b_{ni} + 2\varepsilon_{ni}c_{ni} + 2\varepsilon_{ni}f_{ni} + 2a_{ni}b_{ni} + 2a_{ni}c_{ni} + 2a_{ni}f_{ni} + 2b_{ni}c_{ni} + 2b_{ni}f_{ni} + 2c_{ni}f_{ni}).$$

We focus on terms with the higher orders in ε s. Consider $\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \varepsilon_{nj}^2 \varepsilon_{ni} b_{ni} = (\lambda_0 - \hat{\lambda}) \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n P_{\Delta n, ij} \bar{G}_{n, il} \varepsilon_{ni} \varepsilon_{nj}^2 \varepsilon_{nl}$.

By Cauchy-Schwartz inequality, $E|\varepsilon_{ni}\varepsilon_{nl}\varepsilon_{nj}^2| \leq [E(\varepsilon_{ni}\varepsilon_{nl})^2]^{\frac{1}{2}} [E(\varepsilon_{nj}^4)]^{\frac{1}{2}} \leq [E(\varepsilon_{ni}^4)]^{\frac{1}{4}} [E(\varepsilon_{nl}^4)]^{\frac{1}{4}} [E(\varepsilon_{nj}^4)]^{\frac{1}{4}} \leq c$, for some constant c , for all i, j, l , and n since $\{\mu_{ni}^{(4)}\}$ is a bounded sequence by *Assumption 1*. This implies

$$E \left| \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n P_{\Delta n, ij} \bar{G}_{n, il} \varepsilon_{ni} \varepsilon_{nj}^2 \varepsilon_{nl} \right| \leq c \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^n |P_{\Delta n, ij}| \right) \left(\sum_{l=1}^n |\bar{G}_{n, il}| \right) = O(1),$$

since $P_{\Delta n}$ and $\bar{G}_n$ are uniformly bounded in row and column sums. By the Markov inequality, $\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n P_{\Delta n, ij} \bar{G}_{n, il} \varepsilon_{ni} \varepsilon_{nj}^2 \varepsilon_{nl} = O_p(1)$, i.e., stochastically bounded. Since $\lambda_0 - \hat{\lambda} = o_p(1)$, $(\lambda_0 - \hat{\lambda}) \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n P_{\Delta n, ij} \bar{G}_{n, il} \varepsilon_{ni} \varepsilon_{nj}^2 \varepsilon_{nl}$ converges in probability to zero.

Another term with high order ε s is $\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \varepsilon_{nj}^2 f_{ni}^2 = (\lambda_0 - \hat{\lambda})^2 (\rho_0 - \hat{\rho})^2 \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n P_{\Delta n, ij} M_{n, ik} G_{n, ik} R_{n, ik}^{-1} M_{n, il} G_{n, il} R_{n, il}^{-1} \varepsilon_{nj}^2 \varepsilon_{nk} \varepsilon_{nl}$. From the proof of the previous term, it follows that

$$\begin{aligned} E \left| \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n P_{\Delta n, ij} M_{n, ik} G_{n, ik} R_{n, ik}^{-1} M_{n, il} G_{n, il} R_{n, il}^{-1} \varepsilon_{nj}^2 \varepsilon_{nk} \varepsilon_{nl} \right| \\ \leq c \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^n |P_{\Delta n, ij}| \right) \left(\sum_{k=1}^n |M_{n, ik} G_{n, ik} R_{n, ik}^{-1}| \right) \left(\sum_{l=1}^n |M_{n, il} G_{n, il} R_{n, il}^{-1}| \right) = O(1), \end{aligned}$$

since $P_{\Delta n}$, $M_n G_n R_n^{-1}$ are uniformly bounded in row and column sums. An application of the Markov inequality provides that $\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n P_{\Delta n, ij} M_{n, ik} G_{n, ik} R_{n, ik}^{-1} M_{n, il} G_{n, il} R_{n, il}^{-1} \varepsilon_{nj}^2 \varepsilon_{nk} \varepsilon_{nl} = O_p(1)$. Since $\lambda_0 - \hat{\lambda} = o_p(1)$ and $\rho_0 - \hat{\rho} = o_p(1)$, $\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \varepsilon_{nj}^2 f_{ni}^2$ converges in probability to zero. The remaining terms in ϑ_{n2} are either of the same order or less in ε 's. A similar analysis with Markov inequality can be applied to each of the remaining terms, which yields $\vartheta_{n2} = o_p(1)$. The structure of ϑ_{n3} is the same as that of ϑ_{n2} , i.e., i 's replaced by j 's and vice versa. Hence, ϑ_{n3} converges to zero in probability.

Now we turn to the first term ϑ_{n1} .

$$\begin{aligned} \vartheta_{n1} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} (a_{ni}^2 + b_{ni}^2 + c_{ni}^2 + f_{ni}^2 + 2\varepsilon_{ni}a_{ni} + 2\varepsilon_{ni}b_{ni} + 2\varepsilon_{ni}c_{ni} + 2\varepsilon_{ni}f_{ni} + 2a_{ni}b_{ni} \\ + 2a_{ni}c_{ni} + 2a_{ni}f_{ni} + 2b_{ni}c_{ni} + 2b_{ni}f_{ni} + 2c_{ni}f_{ni}) (a_{nj}^2 + b_{nj}^2 + c_{nj}^2 + f_{nj}^2 + 2\varepsilon_{nj}a_{nj} + 2\varepsilon_{nj}b_{nj} \\ + 2\varepsilon_{nj}c_{nj} + 2\varepsilon_{nj}f_{nj} + 2a_{nj}b_{nj} + 2a_{nj}c_{nj} + 2a_{nj}f_{nj} + 2b_{nj}c_{nj} + 2b_{nj}f_{nj} + 2c_{nj}f_{nj}), \end{aligned}$$

We will again focus on those terms with highest order in ε 's. These terms are $\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} p_{ni}^2 q_{nj}^2$, where $p, q = \{b, c, f\}$. Let $p = q = b$ for exposition. Then,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} b_{ni}^2 b_{nj}^2 &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} (e_{i,n} \bar{G}_n \varepsilon_n)^2 (e_{j,n} \bar{G}_n \varepsilon_n)^2 (\lambda_0 - \hat{\lambda})^4 \\ &= (\lambda_0 - \hat{\lambda})^4 \underbrace{\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{k_1=1}^n \sum_{k_2=1}^n \sum_{l_1=1}^n \sum_{l_2=1}^n P_{\Delta n, ij} \bar{G}_{n, ik_1} \bar{G}_{n, ik_2} \bar{G}_{n, jl_1} \bar{G}_{n, jl_2} \varepsilon_{nk_1} \varepsilon_{nk_2} \varepsilon_{nl_1} \varepsilon_{nl_2}}_{\mathfrak{L}_n} \end{aligned}$$

Applying Cauchy-Schwartz inequality to the following term yields $E|\varepsilon_{nk_1} \varepsilon_{nk_2} \varepsilon_{nl_1} \varepsilon_{nl_2}| \leq (E(\varepsilon_{nk_1}^2 \varepsilon_{nk_2}^2))^{\frac{1}{2}} (E(\varepsilon_{nl_1}^2 \varepsilon_{nl_2}^2))^{\frac{1}{2}} \leq (E(\varepsilon_{nk_1}^4))^{\frac{1}{4}} (E(\varepsilon_{nk_2}^4))^{\frac{1}{4}} (E(\varepsilon_{nl_1}^4))^{\frac{1}{4}} (E(\varepsilon_{nl_2}^4))^{\frac{1}{4}} \leq c$, for some c for all n since $\mu_{nk_1}^{(4)}, \mu_{nk_2}^{(4)}, \mu_{nl_1}^{(4)}$, and $\mu_{nl_2}^{(4)}$ are bounded by Assumption 1. Note that $\mathfrak{L}_n$ is stochastically bounded, since

$$\begin{aligned} E \left| \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{k_1=1}^n \sum_{k_2=1}^n \sum_{l_1=1}^n \sum_{l_2=1}^n P_{\Delta n, ij} \bar{G}_{n, ik_1} \bar{G}_{n, ik_2} \bar{G}_{n, jl_1} \bar{G}_{n, jl_2} \varepsilon_{nk_1} \varepsilon_{nk_2} \varepsilon_{nl_1} \varepsilon_{nl_2} \right| \\ \leq c \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^n |P_{\Delta n, ij}| \right) \left(\sum_{k_1=1}^n |\bar{G}_{n, ik_1}| \right) \left(\sum_{k_2=1}^n |\bar{G}_{n, ik_2}| \right) \left(\sum_{l_1=1}^n |\bar{G}_{n, jl_1}| \right) \left(\sum_{l_2=1}^n |\bar{G}_{n, jl_2}| \right) \\ = O(1), \end{aligned}$$

and by the Markov inequality, $\mathfrak{L}_n$ is stochastically bounded, i.e., $\mathfrak{L}_n = O_p(1)$. Since $\lambda_0 - \hat{\lambda} = o_p(1)$, $\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} b_{ni}^2 b_{nj}^2$ converges in probability to zero.

A similar analysis with an application of the Markov inequality ensures that each of the remaining combinations $p, q = \{b, c, f\}$ in $\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} p_{ni}^2 q_{nj}^2$ is $o_p(1)$. The rest of the terms in ϑ_{n1} are of smaller order in ε s and can easily verified to be stochastically convergent to zero. Hence, ϑ_{n1} converges in probability to zero.

Then, $\vartheta_{n1} = \vartheta_{n2} = \vartheta_{n3} = o_p(1)$ implies $\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \hat{\varepsilon}_{ni}^2 \hat{\varepsilon}_{nj}^2 - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \varepsilon_{ni}^2 \varepsilon_{nj}^2 = o_p(1)$. Combining with the previous result that $\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \varepsilon_{ni}^2 \varepsilon_{nj}^2 - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \sigma_{ni}^2 \sigma_{nj}^2 = o_p(1)$ yields

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \hat{\varepsilon}_{ni}^2 \hat{\varepsilon}_{nj}^2 - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \sigma_{ni}^2 \sigma_{nj}^2 = o_p(1).$$

The remaining term left in $\frac{1}{n} \Omega_n$ is $\frac{1}{n} Q_n' \Sigma_n Q_n = \frac{1}{n} \sum_{i=1}^n \sigma_{ni}^2 q_{i,n}' q_{i,n}$, where $q_{i,n}$ is the i th row of Q_n . The previous discussion applied to $\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n P_{\Delta n, ij} \sigma_{ni}^2 \sigma_{nj}^2$ ensures that $\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{ni}^2 q_{i,n}' q_{i,n} - \frac{1}{n} \sum_{i=1}^n \sigma_{ni}^2 q_{i,n}' q_{i,n} = o_p(1)$. Then, it follows that $\frac{1}{n} \hat{\Omega}_n - \frac{1}{n} \Omega_n = o_p(1)$.

Next, we show the consistency of $\frac{1}{n} \hat{\Gamma}_n$. One type of the elements with ε s in $\frac{1}{n} \Gamma_n$ is $\frac{1}{n} \sum_{i=1}^n \left(H_n' P_{jn}^s \right)_{ii} \sigma_{ni}^2$. Since P_n s and H_n s are all uniformly bounded in both row and column sums, so are matrices $H_n P_{jn}^s$ s. Hence, it follows from the same argument in the proof of the consistency of

$\frac{1}{n}\hat{\Omega}_n$ that $\frac{1}{n}\sum_{i=1}^n \left(H'_n P_{jn}^s\right)_{ii} \hat{\varepsilon}_{ni}^2 - \frac{1}{n}\sum_{i=1}^n \left(H'_n P_{jn}^s\right)_{ii} \sigma_{ni}^2 = o_p(1)$. The other type of elements with ε_s in $\frac{1}{n}\Gamma_n$ is $\frac{1}{n}\sum_{i=1}^n \left(\bar{G}'_n P_{jn}^s\right)_{ii} \sigma_{ni}^2$. By *Assumption 2*, R_n , G_n , and R_n^{-1} are all uniformly bounded in both row and column sums. Hence, the matrices $\bar{G}'_n P_{jn}^s = O(1)$. By the same argument from the proof of the consistency of $\frac{1}{n}\hat{\Omega}_n$, it follows that $\frac{1}{n}\sum_{i=1}^n \left(\bar{G}'_n P_{jn}^s\right)_{ii} \hat{\varepsilon}_{ni}^2 - \frac{1}{n}\sum_{i=1}^n \left(\bar{G}'_n P_{jn}^s\right)_{ii} \sigma_{ni}^2 = o_p(1)$. Then, $\frac{1}{n}\hat{\Gamma}_n - \frac{1}{n}\Gamma_n = o_p(1)$. $\square$

Proof of Proposition 3. The proof follows in parallel to the proof of *Proposition 3* in Lin and Lee (2010). By generalized Schwartz inequality, the optimal weighting matrix in *Proposition 1* is $(\frac{1}{n}\Omega_n)^{-1}$. First, we show that $\frac{1}{n}g'_n(\theta)\hat{\Omega}_n^{-1}g_n(\theta) - \frac{1}{n}g'_n(\theta)\Omega_n^{-1}g_n(\theta) = o_p(1)$.

Consider $\frac{1}{n}g'_n(\theta)\hat{\Omega}_n^{-1}g_n(\theta) = \frac{1}{n}g'_n(\theta)\Omega_n^{-1}g_n(\theta) + \frac{1}{n}g'_n(\theta)(\hat{\Omega}_n^{-1} - \Omega_n^{-1})g_n(\theta)$. Letting $\Psi_n = (\frac{1}{n}\Omega_n)^{-\frac{1}{2}}$ in *Proposition 1*, *Assumption 7* implies that $\Psi_0 = (\lim_{n \rightarrow \infty} \frac{1}{n}\Omega_n)^{-\frac{1}{2}}$ exists. Because Ψ_0 is nonsingular, θ_0 corresponds to the unique root of $\lim_{n \rightarrow \infty} E(\frac{1}{n}g_n(\theta)) = 0$ at θ_0 , which is satisfied by *Assumption 6*. A similar argument in the proof of *Proposition 1* ensures that $\frac{1}{n}g'_n(\theta)\Omega_n^{-1}g_n(\theta)$ converges in probability to a well defined limit uniformly in $\theta \in \Theta$. Now, we show that $\frac{1}{n}g'_n(\theta)(\hat{\Omega}_n^{-1} - \Omega_n^{-1})g_n(\theta) = o_p(1)$ uniformly in $\theta \in \Theta$. Let $\|\cdot\|$ be the maximum row sum norm for vectors and matrices. Then, by the submultiplicative property of a matrix norm, $\|\frac{1}{n}g'_n(\theta)(\hat{\Omega}_n^{-1} - \Omega_n^{-1})g_n(\theta)\| \leq (\frac{1}{n}\|g_n(\theta)\|)^2 \|(\hat{\Omega}_n^{-1} - \Omega_n^{-1})\|$. From the proof of *Proposition 1*, $\frac{1}{n}(g_n(\theta) - E(g_n(\theta))) = o_p(1)$. Also, from the proof of *Proposition 1*,

$$\begin{aligned} \frac{1}{n}E(\varepsilon'_n(\theta)P_{jn}\varepsilon_n(\theta)) &= h'_n(\varsigma)R'_n(\rho)P_{jn}R_n(\rho)h_n(\varsigma) + (\lambda_0 - \lambda)\frac{1}{n}\text{tr}(\Sigma_n P_{jn}^s \bar{G}_n) + (\rho_0 - \rho)\frac{1}{n} \\ &\quad \times \text{tr}(\Sigma_n P_{jn}^s H_n) + (\rho_0 - \rho)(\lambda_0 - \lambda)\frac{1}{n}\text{tr}(\Sigma_n P_{jn}^s M_n G_n R_n^{-1}) + (\lambda_0 - \lambda)(\rho_0 - \rho)\frac{1}{n}\text{tr}(\Sigma_n H'_n P_{jn}^s \bar{G}_n) \\ &\quad + (\lambda_0 - \lambda)^2(\rho_0 - \rho)\frac{1}{n}\text{tr}(\Sigma_n \bar{G}_n P_{jn}^s M_n G_n R_n^{-1}) + (\rho_0 - \rho)^2(\lambda_0 - \lambda)\frac{1}{n}\text{tr}(\Sigma_n H'_n P_{jn}^s M_n G_n R_n^{-1}) \\ &\quad + (\lambda_0 - \lambda)^2\frac{1}{n}\text{tr}(\Sigma_n \bar{G}'_n P_{jn} \bar{G}_n) + (\rho_0 - \rho)^2\frac{1}{n}\text{tr}(\Sigma_n H'_n P_{jn} H_n) + (\rho_0 - \rho)^2(\lambda_0 - \lambda)^2\frac{1}{n} \\ &\quad \times \text{tr}(\Sigma_n R_n'^{-1} G'_n M'_n P_{jn} M_n G_n R_n^{-1}) = O(1), \end{aligned}$$

uniformly in $\theta \in \Theta$, as

$$\begin{aligned} \frac{1}{n}h'_n(\varsigma)R'_n(\rho)P_{jn}R_n(\rho)h_n(\varsigma) &= (\lambda_0 - \lambda)^2\frac{1}{n}(X_n\beta_0)'G'_n(R_n + (\rho_0 - \rho)M_n)'P_{jn}(R_n + (\rho_0 - \rho)M_n) \\ &\quad \times G_n(X_n\beta_0) + (\lambda_0 - \lambda)\frac{1}{n}(X_n\beta_0)'G'_n(R_n + (\rho_0 - \rho)M_n)'P_{jn}^s(R_n + (\rho_0 - \rho)M_n)X_n(\beta_0 - \beta) \\ &\quad + (\beta_0 - \beta)\frac{1}{n}X'_n(R_n + (\rho_0 - \rho)M_n)'P_{jn}(R_n + (\rho_0 - \rho)M_n)X_n(\beta_0 - \beta) = O_p(1), \end{aligned}$$

uniformly in $\theta \in \Theta$. Similarly, $\frac{1}{n}E(Q'_n\varepsilon_n(\theta)) = \frac{1}{n}Q'_nR_n(\rho)h_n(\varsigma) = (\lambda_0 - \lambda)\frac{1}{n}Q'_n(R_n + (\rho_0 - \rho)M_n)G_nX_n\beta_0 + \frac{1}{n}Q'_n(R_n + (\rho_0 - \rho)M_n)X_n(\beta_0 - \beta) = O(1)$ uniformly in $\theta \in \Theta$. Hence, $\|\frac{1}{n}E(g_n(\theta))\| = O(1)$ uniformly in $\theta \in \Theta$. Then, $\frac{1}{n}\|g_n(\theta)\| = O_p(1)$ uniformly in $\theta \in \Theta$ by the Markov inequality. These imply that $\|\frac{1}{n}g'_n(\theta)(\hat{\Omega}_n^{-1} - \Omega_n^{-1})g_n(\theta)\| = o_p(1)$ uniformly in $\theta \in \Theta$. This result shows the consistency of the optimal robust GMME.

From the proof of *Proposition 1*, we have $\frac{1}{n}\frac{\partial g_n(\hat{\theta}_n)}{\partial \theta'} = -\frac{1}{n}\Gamma_n + o_p(1)$ uniformly in θ . To find the

limiting distribution, by (C.5)

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_{o,n} - \theta_0) &= - \left(\frac{1}{n} \frac{\partial g'_n(\hat{\theta}_n)}{\partial \theta} \left(\frac{\hat{\Omega}_n}{n} \right)^{-1} \frac{1}{n} \frac{\partial g_n(\bar{\theta}_n)}{\partial \theta'} \right)^{-1} \frac{1}{n} \frac{\partial g'_n(\hat{\theta}_n)}{\partial \theta} \left(\frac{\hat{\Omega}_n}{n} \right)^{-1} \frac{1}{\sqrt{n}} g_n(\theta_0) \\
&= \left(\frac{\Gamma'_n}{n} \left(\frac{\Omega_n}{n} \right)^{-1} \frac{\Gamma_n}{n} \right)^{-1} \frac{\Gamma'_n}{n} \left(\frac{\Omega_n}{n} \right)^{-1} \frac{1}{\sqrt{n}} g_n(\theta_0) + o_p(1).
\end{aligned} \tag{C.16}$$

Hence, the limiting distribution of $\sqrt{n}(\hat{\theta}_{o,n} - \theta_0)$ follows immediately from (C.16) by the CLT in *Theorem 1* of Kelejian and Prucha (2001) and the Slutsky theorem. $\square$

D Simulation Results

Table 1: True parameter vector: $(\lambda_0, \beta_{10}, \beta_{20}, \beta_{30}, \rho_0) = (-0.8, 0.7, 0.4, 1.2, -0.3)$.

N			Homoskedasticity				Heteroskedasticity			
			Mean	Bias	SD	RMSE	Mean	Bias	SD	RMSE
100	MLE	λ	-1.309	-0.509	1.105	1.217	-1.220	-0.420	3.330	3.356
		β_1	0.643	-0.057	0.129	0.141	0.396	-0.304	0.389	0.494
		β_2	0.364	-0.036	0.171	0.121	0.199	-0.201	0.311	0.370
		β_3	1.105	-0.095	0.116	0.195	0.671	-0.529	0.526	0.746
		ρ	-0.651	-0.351	1.621	1.659	-4.886	-4.586	5.054	6.825
	RGMME ₁	λ	-0.853	-0.053	0.363	0.367	-0.876	-0.076	0.875	0.878
		β_1	0.696	-0.004	0.105	0.106	0.698	-0.002	0.340	0.340
		β_2	0.394	-0.006	0.102	0.102	0.369	-0.031	0.336	0.338
		β_3	1.194	-0.006	0.102	0.102	1.188	-0.012	0.344	0.344
		ρ	-0.356	-0.056	0.492	0.496	-0.498	-0.198	0.888	0.910
	RGMME ₂	λ	-0.862	-0.062	0.351	0.357	-0.693	0.107	1.003	1.009
		β_1	0.696	-0.004	0.105	0.105	0.710	0.010	0.348	0.348
		β_2	0.394	-0.006	0.101	0.102	0.371	-0.029	0.346	0.347
		β_3	1.193	-0.007	0.102	0.102	1.204	0.004	0.352	0.352
		ρ	-0.348	-0.048	0.497	0.499	-0.552	-0.252	0.812	0.850
500	MLE	λ	-1.028	-0.228	0.624	0.665	-2.275	-1.475	3.748	4.028
		β_1	0.680	-0.020	0.064	0.067	0.475	-0.225	0.337	0.405
		β_2	0.390	-0.010	0.052	0.053	0.271	-0.129	0.225	0.260
		β_3	1.168	-0.032	0.086	0.091	0.817	-0.383	0.547	0.668
		ρ	-0.227	0.073	0.363	0.370	-1.227	-0.927	3.183	3.315
	RGMME ₁	λ	-0.811	-0.011	0.124	0.124	-0.819	-0.019	0.288	0.289
		β_1	0.697	-0.003	0.046	0.046	0.692	-0.008	0.150	0.150
		β_2	0.399	-0.001	0.046	0.046	0.397	-0.003	0.149	0.149
		β_3	1.196	-0.004	0.044	0.045	1.184	-0.016	0.151	0.152
		ρ	-0.312	-0.012	0.169	0.169	-0.313	-0.013	0.276	0.277
	RGMME ₂	λ	-0.813	-0.013	0.124	0.124	-0.815	-0.015	0.269	0.269
		β_1	0.697	-0.003	0.046	0.046	0.693	-0.007	0.150	0.150
		β_2	0.399	-0.001	0.046	0.046	0.397	-0.003	0.149	0.149
		β_3	1.196	-0.004	0.044	0.045	1.185	-0.015	0.150	0.151
		ρ	-0.310	-0.010	0.169	0.169	-0.313	-0.013	0.267	0.268
1000	MLE	λ	-0.890	-0.090	0.408	0.418	-2.235	-1.435	3.666	3.937
		β_1	0.694	-0.006	0.041	0.042	0.546	-0.154	0.305	0.342
		β_2	0.397	-0.003	0.035	0.035	0.314	-0.086	0.193	0.212
		β_3	1.189	-0.011	0.056	0.057	0.931	-0.269	0.506	0.573
		ρ	-0.271	0.029	0.226	0.228	-0.461	-0.161	1.791	1.798
	RGMME ₁	λ	-0.802	-0.002	0.089	0.089	-0.802	-0.002	0.224	0.224
		β_1	0.700	0.000	0.031	0.031	0.698	-0.002	0.105	0.105
		β_2	0.400	0.000	0.032	0.032	0.403	0.003	0.108	0.108
		β_3	1.200	0.000	0.031	0.031	1.195	-0.005	0.107	0.107
		ρ	-0.304	-0.004	0.119	0.119	-0.315	-0.015	0.214	0.214
	RGMME ₂	λ	-0.803	-0.003	0.089	0.089	-0.803	-0.003	0.221	0.221
		β_1	0.700	0.000	0.031	0.031	0.698	-0.002	0.105	0.105
		β_2	0.400	0.000	0.032	0.032	0.403	0.003	0.108	0.108
		β_3	1.200	0.000	0.031	0.031	1.195	-0.005	0.107	0.107
		ρ	-0.304	-0.004	0.118	0.118	-0.313	-0.013	0.211	0.211

Table 2: True parameter vector: $(\lambda_0, \beta_{10}, \beta_{20}, \beta_{30}, \rho_0) = (-0.3, 0.7, 0.4, 1.2, 0.3)$.

<u>N</u>			Homoskedasticity				Heteroskedasticity			
			<u>Mean</u>	<u>Bias</u>	<u>SD</u>	<u>RMSE</u>	<u>Mean</u>	<u>Bias</u>	<u>SD</u>	<u>RMSE</u>
100	MLE	λ	-0.459	-0.159	0.771	0.787	0.207	0.507	1.349	1.441
		β_1	0.685	-0.015	0.118	0.119	0.490	-0.210	0.428	0.477
		β_2	0.389	-0.011	0.103	.104	0.262	-0.138	0.331	0.359
		β_3	1.175	-0.025	0.140	0.142	0.821	-0.379	0.593	0.704
		ρ	0.146	-0.154	0.730	0.746	-4.421	-4.721	5.238	7.052
	RGMME ₁	λ	-0.370	-0.070	0.435	0.440	-0.156	0.144	1.242	1.250
		β_1	0.694	-0.006	0.106	0.106	0.718	0.018	0.358	0.359
		β_2	0.394	-0.006	0.100	0.100	0.380	-0.020	0.353	0.353
		β_3	1.192	-0.008	0.107	0.107	1.211	0.011	0.366	0.366
		ρ	0.253	-0.047	0.361	0.364	0.069	-0.231	0.593	0.637
	RGMME ₂	λ	-0.342	-0.042	0.413	0.415	0.276	0.576	1.394	1.509
		β_1	0.696	-0.004	0.105	0.105	0.740	0.040	0.386	0.388
		β_2	0.395	-0.005	0.100	0.100	0.392	-0.008	0.367	0.367
		β_3	1.195	-0.005	0.105	0.105	1.243	0.043	0.420	0.422
		ρ	0.240	-0.060	0.350	0.355	-0.084	-0.384	0.677	0.779
500	MLE	λ	-0.315	-0.015	0.153	0.154	-0.101	0.199	0.872	0.894
		β_1	0.697	-0.003	0.047	0.047	0.697	-0.003	0.176	0.176
		β_2	0.400	0.000	0.046	0.046	0.399	-0.001	0.162	0.162
		β_3	1.195	-0.005	0.046	0.046	1.189	-0.011	0.207	0.207
		ρ	0.282	-0.018	0.123	0.124	0.021	-0.279	0.776	0.825
	RGMME ₁	λ	-0.323	-0.023	0.161	0.162	-0.397	-0.097	0.468	0.478
		β_1	0.697	-0.003	0.047	0.047	0.686	-0.014	0.156	0.157
		β_2	0.400	0.000	0.046	0.046	0.395	-0.005	0.154	0.154
		β_3	1.195	-0.005	0.046	0.047	1.173	-0.027	0.160	0.162
		ρ	0.297	-0.003	0.126	0.126	0.281	-0.019	0.250	0.250
	RGMME ₂	λ	-0.320	-0.020	0.158	0.159	-0.282	0.018	0.478	0.479
		β_1	0.697	-0.003	0.047	0.047	0.694	-0.006	0.159	0.159
		β_2	0.400	0.000	0.046	0.046	0.399	-0.001	0.156	0.156
		β_3	1.195	-0.005	0.046	0.046	1.186	-0.014	0.165	0.166
		ρ	0.295	-0.005	0.125	0.125	0.242	-0.058	0.231	0.238
1000	MLE	λ	-0.302	-0.002	0.113	0.113	-0.067	0.233	0.699	0.737
		β_1	0.700	0.000	0.031	0.031	0.706	0.006	0.120	0.120
		β_2	0.400	0.000	0.032	0.032	0.407	0.007	0.114	0.114
		β_3	1.200	0.000	0.032	0.032	1.206	0.006	0.141	0.141
		ρ	0.289	-0.011	0.088	0.088	0.082	-0.218	0.086	0.234
	RGMME ₁	λ	-0.306	-0.006	0.115	0.115	-0.335	-0.035	0.326	0.328
		β_1	0.700	0.000	0.031	0.031	0.696	-0.004	0.108	0.108
		β_2	0.400	0.000	0.032	0.032	0.402	0.002	0.110	0.110
		β_3	1.199	-0.001	0.032	.032	1.191	-0.009	0.112	0.113
		ρ	0.298	-0.002	0.089	0.089	0.280	-0.020	0.186	0.187
	RGMME ₂	λ	-0.304	-0.004	0.114	0.114	-0.291	0.009	0.258	0.258
		β_1	0.700	0.000	0.031	0.031	0.699	-0.001	0.108	0.108
		β_2	0.400	0.000	0.032	0.032	0.403	0.003	0.110	0.110
		β_3	1.199	-0.001	0.032	0.032	1.196	-0.004	0.112	0.112
		ρ	0.297	-0.003	0.089	0.089	0.265	-0.035	0.172	0.175

Table 3: True parameter vector: $(\lambda_0, \beta_{10}, \beta_{20}, \beta_{30}, \rho_0) = (0, 0.7, 0.4, 1.2, 0)$.

<u>N</u>			Homoskedasticity				Heteroskedasticity			
			<u>Mean</u>	<u>Bias</u>	<u>SD</u>	<u>RMSE</u>	<u>Mean</u>	<u>Bias</u>	<u>SD</u>	<u>RMSE</u>
100	MLE	λ	-0.063	-0.063	0.443	0.447	0.359	0.359	0.991	1.054
		β_1	0.685	-0.015	0.121	0.122	0.444	-0.256	0.403	0.478
		β_2	0.388	-0.012	0.109	0.109	0.235	-0.165	0.313	0.354
		β_3	1.175	-0.025	0.146	0.148	0.758	-0.442	0.581	0.730
		ρ	-0.335	-0.335	1.248	1.292	-5.490	-5.490	5.502	7.772
	RGMME ₁	λ	-0.067	-0.067	0.306	0.313	-0.095	-0.095	1.001	1.006
		β_1	0.696	-0.004	0.105	0.105	0.703	0.003	0.347	0.347
		β_2	0.395	-0.005	0.102	0.102	0.370	-0.030	0.347	0.349
		β_3	1.194	-0.006	0.101	0.101	1.189	-0.011	0.348	0.348
		ρ	-0.058	-0.058	0.422	0.426	-0.147	-0.147	0.769	0.873
	RGMME ₂	λ	-0.062	-0.062	0.284	0.290	0.220	0.220	1.145	1.166
		β_1	0.696	-0.004	0.104	0.104	0.716	0.016	0.361	0.361
		β_2	0.395	-0.005	0.101	0.101	0.376	-0.024	0.357	0.358
		β_3	1.195	-0.005	0.100	0.100	1.209	0.009	0.394	0.394
		ρ	-0.064	-0.064	0.418	0.423	-0.276	-0.276	0.828	0.873
500	MLE	λ	-0.011	-0.011	0.100	0.100	0.030	0.030	0.206	0.208
		β_1	0.698	-0.002	0.046	0.046	0.674	-0.026	0.193	0.195
		β_2	0.400	0.000	0.046	0.046	0.385	-0.015	0.166	0.167
		β_3	1.196	-0.004	0.045	0.045	1.150	-0.050	0.253	0.257
		ρ	-0.023	-0.023	0.145	0.147	-0.493	-0.493	2.244	2.298
	RGMME ₁	λ	-0.013	-0.013	0.101	0.102	0.070	0.070	0.335	0.342
		β_1	0.698	-0.002	0.046	0.046	0.690	-0.010	0.154	0.154
		β_2	0.400	0.000	0.046	0.046	0.396	-0.004	0.152	0.152
		β_3	1.196	-0.004	0.045	0.045	1.180	-0.020	0.156	0.157
		ρ	-0.008	-0.008	0.148	0.148	-0.002	-0.002	0.282	0.282
	RGMME ₂	λ	-0.013	-0.013	0.100	0.101	-0.034	-0.034	0.241	0.244
		β_1	0.698	-0.002	0.046	0.046	0.692	-0.008	0.154	0.154
		β_2	0.400	0.000	0.046	0.046	0.397	-0.003	0.153	0.153
		β_3	1.196	-0.004	0.045	0.045	1.183	-0.017	0.154	0.155
		ρ	-0.008	-0.008	0.147	0.148	-0.017	-0.017	0.263	0.264
1000	MLE	λ	-0.002	-0.002	0.071	0.071	0.003	0.003	0.082	0.082
		β_1	0.700	0.000	0.031	0.031	0.699	-0.001	0.113	0.113
		β_2	0.400	0.000	0.032	0.032	0.403	0.003	0.111	0.111
		β_3	1.200	0.000	0.031	0.031	1.195	-0.005	0.126	0.126
		ρ	-0.013	-0.013	0.104	0.104	-0.056	-0.056	0.712	0.715
	RGMME ₁	λ	-0.003	-0.003	0.071	0.072	-0.025	-0.025	0.205	0.206
		β_1	0.700	0.000	0.031	0.031	0.698	-0.002	0.107	0.107
		β_2	0.400	0.000	0.032	0.032	0.403	0.003	0.110	0.110
		β_3	1.200	0.000	0.031	0.031	1.194	-0.006	0.109	0.109
		ρ	-0.004	-0.004	0.104	0.104	-0.009	-0.009	0.216	0.216
	RGMME ₂	λ	-0.003	-0.003	0.071	0.071	-0.020	-0.020	0.195	0.196
		β_1	0.700	0.000	0.031	0.031	0.698	-0.002	0.107	0.107
		β_2	0.400	0.000	0.032	0.032	0.403	0.003	0.110	0.110
		β_3	1.200	0.000	0.031	0.031	1.194	-0.006	0.109	0.109
		ρ	-0.004	-0.004	0.104	0.104	-0.012	-0.012	0.211	0.211

Table 4: True parameter vector: $(\lambda_0, \beta_{10}, \beta_{20}, \beta_{30}, \rho_0) = (0.3, 0.7, 0.4, 1.2, -0.8)$.

N			Homoskedasticity				Heteroskedasticity			
			Mean	Bias	SD	RMSE	Mean	Bias	SD	RMSE
100	MLE	λ	0.359	0.059	0.232	0.239	0.603	0.303	0.924	0.972
		β_1	0.630	-0.070	0.159	0.174	0.322	-0.378	0.345	0.512
		β_2	0.354	-0.046	0.131	0.139	0.171	-0.229	0.262	0.348
		β_3	1.080	-0.120	0.224	0.254	0.552	-0.648	0.507	0.823
		ρ	-2.560	-1.760	2.237	2.846	-8.711	-7.911	4.898	9.304
	RGMME ₁	λ	0.276	-0.024	0.146	0.148	0.108	-0.192	0.756	0.780
		β_1	0.700	0.000	0.700	0.105	0.708	0.008	0.343	0.343
		β_2	0.396	-0.004	0.105	0.101	0.376	-0.024	0.333	0.334
		β_3	1.201	0.001	0.101	0.104	1.203	0.003	0.363	0.363
		ρ	-0.935	-0.135	0.589	0.604	-0.861	-0.061	1.275	1.276
	RGMME ₂	λ	0.273	-0.027	0.145	0.147	0.226	-0.074	0.461	0.467
		β_1	0.701	0.001	0.105	0.105	0.711	0.011	0.343	0.343
		β_2	0.397	-0.003	0.101	0.101	0.378	-0.022	0.335	0.335
		β_3	1.202	0.002	0.104	0.104	1.205	0.005	0.365	0.366
		ρ	-0.934	-0.134	0.587	0.602	-0.951	-0.151	1.268	1.277
500	MLE	λ	0.357	0.057	0.120	0.133	0.601	0.301	0.437	0.531
		β_1	0.648	-0.052	0.099	0.112	0.329	-0.371	0.305	0.480
		β_2	0.369	-0.031	0.068	0.075	0.187	-0.213	0.193	0.288
		β_3	1.112	-0.088	0.156	0.179	0.563	-0.637	0.501	0.811
		ρ	-0.816	-0.016	1.447	1.682	-7.626	-6.826	5.399	8.703
	RGMME ₁	λ	0.297	-0.003	0.045	0.045	0.281	-0.019	0.127	0.128
		β_1	0.698	-0.002	0.046	0.046	0.696	-0.004	0.147	0.147
		β_2	0.399	-0.001	0.045	0.045	0.398	-0.002	0.141	0.141
		β_3	1.198	-0.002	0.046	0.046	1.192	-0.008	0.159	0.159
		ρ	-0.825	-0.025	0.186	0.188	-0.771	0.029	0.366	0.368
	RGMME ₂	λ	0.297	-0.003	0.045	0.045	0.277	-0.023	0.126	0.128
		β_1	0.698	-0.002	0.046	0.046	0.697	-0.003	0.147	0.147
		β_2	0.399	-0.001	0.045	0.045	0.399	-0.001	0.141	0.141
		β_3	1.198	-0.002	0.046	0.046	1.193	-0.007	0.159	0.159
		ρ	-0.824	-0.024	0.186	0.187	-0.765	0.035	0.362	0.364
1000	MLE	λ	0.336	0.036	0.096	0.102	0.572	0.272	0.446	0.522
		β_1	0.670	-0.030	0.077	0.083	0.346	-0.354	0.310	0.471
		β_2	0.383	-0.017	0.052	0.055	0.203	-0.197	0.192	0.275
		β_3	1.148	-0.052	0.127	0.137	0.594	-0.606	0.523	0.801
		ρ	-1.312	-0.512	1.193	1.299	-7.258	-6.458	5.640	8.574
	RGMME ₁	λ	0.300	0.000	0.032	0.032	0.291	-0.009	0.095	0.096
		β_1	0.700	0.000	0.031	0.031	0.699	-0.001	0.105	0.105
		β_2	0.401	0.001	0.032	0.032	0.405	0.005	0.105	0.105
		β_3	1.200	0.000	0.032	0.032	1.197	-0.003	0.115	0.115
		ρ	-0.810	-0.010	0.132	0.132	-0.791	0.009	0.272	0.272
	RGMME ₂	λ	0.299	-0.001	0.032	0.032	0.289	-0.011	0.095	0.096
		β_1	0.700	0.000	0.031	0.031	0.700	0.000	0.105	0.105
		β_2	0.401	0.001	0.032	0.032	0.405	0.005	0.105	0.105
		β_3	1.200	0.000	0.032	0.032	1.198	-0.002	0.115	0.115
		ρ	-0.809	-0.009	0.132	0.132	-0.787	0.013	0.271	0.271

Table 5: True parameter vector: $(\lambda_0, \beta_{10}, \beta_{20}, \beta_{30}, \rho_0) = (0.8, 0.7, 0.4, 1.2, 0.3)$.

<u>N</u>			Homoskedasticity				Heteroskedasticity			
			<u>Mean</u>	<u>Bias</u>	<u>SD</u>	<u>RMSE</u>	<u>Mean</u>	<u>Bias</u>	<u>SD</u>	<u>RMSE</u>
100	MLE	λ	0.769	-0.031	0.197	0.199	0.827	0.027	0.187	0.189
		β_1	0.693	-0.007	0.111	0.111	0.451	-0.249	0.429	0.496
		β_2	0.393	-0.007	0.103	0.103	0.240	-0.160	0.322	0.360
		β_3	1.190	-0.010	0.124	0.124	0.769	-0.431	0.631	0.764
		ρ	0.083	-0.217	0.884	0.910	-5.214	-5.514	5.978	8.133
	RGMME ₁	λ	0.745	-0.055	0.336	0.341	0.728	-0.072	0.732	0.735
		β_1	0.697	-0.003	0.106	0.106	0.706	0.006	0.346	0.346
		β_2	0.396	-0.004	0.101	0.101	0.379	-0.021	0.347	0.348
		β_3	1.197	-0.003	0.104	0.104	1.197	-0.003	0.349	0.349
		ρ	0.255	-0.045	0.357	0.360	0.154	-0.146	0.629	0.646
	RGMME ₂	λ	0.764	-0.036	0.181	0.185	0.875	0.075	0.458	0.464
		β_1	0.698	-0.002	0.104	0.104	0.709	0.009	0.353	0.353
		β_2	0.397	-0.003	0.101	0.101	0.377	-0.023	0.350	0.351
		β_3	1.199	-0.001	0.100	0.100	1.194	-0.006	0.358	0.358
		ρ	0.235	-0.065	0.350	0.356	0.024	-0.276	0.693	0.746
500	MLE	λ	0.797	-0.003	0.033	0.033	0.730	-0.070	0.112	0.132
		β_1	0.698	-0.002	0.046	0.046	0.679	-0.021	0.183	0.185
		β_2	0.400	0.000	0.046	0.046	0.389	-0.011	0.163	0.164
		β_3	1.197	-0.003	0.045	0.045	1.158	-0.042	0.236	0.240
		ρ	0.271	-0.029	0.131	0.134	0.020	-0.280	1.935	1.955
	RGMME ₁	λ	0.794	-0.006	0.033	0.034	0.752	-0.048	0.483	0.486
		β_1	0.698	-0.002	0.046	0.046	0.693	-0.007	0.157	0.157
		β_2	0.400	0.000	0.046	0.046	0.398	-0.002	0.155	0.155
		β_3	1.197	-0.003	0.044	0.045	1.185	-0.015	0.165	0.165
		ρ	0.296	-0.004	0.128	0.128	0.307	0.007	0.286	0.286
	RGMME ₂	λ	0.794	-0.006	0.033	0.034	0.774	-0.026	0.271	0.273
		β_1	0.698	-0.002	0.046	0.046	0.691	-0.009	0.161	0.161
		β_2	0.400	0.000	0.046	0.046	0.397	-0.003	0.156	0.156
		β_3	1.197	-0.003	0.044	0.045	1.181	-0.019	0.171	0.172
		ρ	0.295	-0.005	0.127	0.127	0.247	-0.053	0.734	0.736
1000	MLE	λ	0.800	0.000	0.023	0.023	0.742	-0.058	0.098	0.114
		β_1	0.700	0.000	0.031	0.031	0.697	-0.003	0.108	0.108
		β_2	0.400	0.000	0.032	0.032	0.401	0.001	0.109	0.109
		β_3	1.200	0.000	0.031	0.031	1.193	-0.007	0.114	0.115
		ρ	0.284	-0.016	0.093	0.094	0.299	-0.001	0.471	0.471
	RGMME ₁	λ	0.798	-0.002	0.022	0.023	0.777	-0.023	0.116	0.118
		β_1	0.700	0.000	0.031	0.031	0.698	-0.002	0.106	0.106
		β_2	0.400	0.000	0.032	0.032	0.403	0.003	0.109	0.110
		β_3	1.200	0.000	0.031	0.031	1.195	-0.005	0.109	0.109
		ρ	0.297	-0.003	0.090	0.090	0.295	-0.005	0.218	0.218
	RGMME ₂	λ	0.799	-0.001	0.022	0.022	0.780	-0.020	0.098	0.100
		β_1	0.700	0.000	0.031	0.031	0.698	-0.002	0.106	0.106
		β_2	0.400	0.000	0.032	0.032	0.403	0.003	0.109	0.109
		β_3	1.200	0.000	0.031	0.031	1.195	-0.005	0.110	0.110
		ρ	0.296	-0.004	0.090	0.090	0.290	-0.010	0.214	0.215

Table 6: $(\lambda_0, \beta_{10}, \beta_{20}, \beta_{30}, \rho_0) = (0.3, 0.7, 0.4, 1.2, 0.3)$ and $N=500$.

		Homoskedasticity				Heteroskedasticity			
		<u>Mean</u>	<u>Bias</u>	<u>SD</u>	<u>RMSE</u>	<u>Mean</u>	<u>Bias</u>	<u>SD</u>	<u>RMSE</u>
<u>GS2SLSE</u>	λ	0.302	0.002	0.100	0.100	0.442	0.142	0.223	0.265
	β_1	0.698	-0.002	0.046	0.046	0.681	-0.019	0.158	0.159
	β_2	0.400	0.000	0.046	0.046	0.390	-0.010	0.155	0.155
	β_3	1.197	-0.003	0.045	0.045	1.166	-0.034	0.162	0.166
	ρ	0.263	-0.037	0.125	0.125	-0.094	-0.394	0.355	0.530
<u>MLE</u>	λ	0.285	-0.015	0.103	0.104	0.299	-0.001	0.133	0.133
	β_1	0.698	-0.002	0.046	0.046	0.686	-0.014	0.175	0.176
	β_2	0.400	0.000	0.046	0.046	0.394	-0.006	0.160	0.160
	β_3	1.196	-0.004	0.044	0.045	1.170	-0.030	0.210	0.212
	ρ	0.282	-0.018	0.129	0.130	0.023	-0.277	1.551	1.576
<u>B2SLSE</u>	λ	0.282	-0.018	0.106	0.107	0.039	-0.261	3.329	3.340
	β_1	0.697	-0.003	0.046	0.046	0.667	-0.033	0.822	0.823
	β_2	0.400	0.000	0.046	0.046	0.428	0.028	1.039	1.040
	β_3	1.196	-0.004	0.045	0.045	1.158	-0.042	0.369	0.371
	ρ	0.263	-0.037	0.125	0.131	-0.094	-0.394	0.355	0.530
<u>BGMME</u>	λ	0.255	-0.045	0.218	0.222	-1.412	-1.712	3.420	3.825
	β_1	0.696	-0.004	0.049	0.049	0.344	-0.356	0.305	0.469
	β_2	0.398	-0.002	0.047	0.047	0.193	-0.207	0.206	0.292
	β_3	1.193	-0.007	0.051	0.051	0.581	-0.619	0.482	0.785
	ρ	0.330	0.030	0.153	0.156	-2.393	-2.693	3.791	4.650
<u>RGS2SLSE</u>	λ	0.301	0.001	0.100	0.100	0.437	0.137	0.222	0.261
	β_1	0.698	-0.002	0.046	0.049	0.692	-0.008	0.156	0.156
	β_2	0.400	0.000	0.046	0.047	0.398	-0.002	0.154	0.154
	β_3	1.197	-0.003	0.044	0.051	1.181	-0.019	0.159	0.160
	ρ	0.329	0.029	0.130	0.155	0.120	-0.180	0.255	0.312
<u>RB2SLSE</u>	λ	0.281	-0.019	0.106	0.100	0.113	-0.187	2.026	2.035
	β_1	0.697	-0.003	0.046	0.046	0.684	-0.016	0.295	0.295
	β_2	0.400	0.000	0.046	0.046	0.405	0.005	0.295	0.295
	β_3	1.197	-0.003	0.045	0.045	1.166	-0.034	0.279	0.281
	ρ	0.279	-0.021	0.130	0.132	0.120	-0.180	0.255	0.312
<u>RGMME₁</u>	λ	0.281	-0.019	0.106	0.107	0.270	-0.030	0.399	0.400
	β_1	0.697	-0.003	0.046	0.046	0.691	-0.009	0.158	0.158
	β_2	0.400	0.000	0.046	0.046	0.397	-0.003	0.155	0.155
	β_3	1.196	-0.003	0.045	0.045	1.181	-0.019	0.165	0.166
	ρ	0.297	-0.003	0.128	0.128	0.255	-0.045	0.382	0.385

Table 7: $(\lambda_0, \beta_{10}, \beta_{20}, \beta_{30}, \rho_0) = (0.8, 0.7, 0.4, 1.2, -0.3)$ and $N=1000$.

		Homoskedasticity				Heteroskedasticity			
		<u>Mean</u>	<u>Bias</u>	<u>SD</u>	<u>RMSE</u>	<u>Mean</u>	<u>Bias</u>	<u>SD</u>	<u>RMSE</u>
<u>GS2SLSE</u>	λ	0.801	0.001	0.012	0.012	0.806	0.006	0.038	0.038
	β_1	0.700	0.000	0.031	0.031	0.692	-0.008	0.110	0.110
	β_2	0.400	0.000	0.032	0.032	0.401	0.001	0.109	0.109
	β_3	1.199	-0.001	0.033	0.033	1.185	-0.015	0.118	0.118
	ρ	-0.319	-0.019	0.116	0.117	-0.554	-0.254	0.290	0.386
<u>MLE</u>	λ	0.801	0.001	0.012	0.012	0.639	-0.161	0.115	0.197
	β_1	0.700	0.000	0.031	0.031	0.670	-0.030	0.206	0.208
	β_2	0.400	0.000	0.032	0.032	0.384	-0.016	0.149	0.150
	β_3	1.199	-0.001	0.032	0.032	1.143	-0.057	0.322	0.327
	ρ	-0.322	-0.022	0.115	0.117	-0.705	-0.405	3.363	3.387
<u>B2SLSE</u>	λ	0.800	0.000	0.012	0.012	0.795	-0.005	0.042	0.042
	β_1	0.701	0.001	0.031	0.031	0.700	0.000	0.109	0.109
	β_2	0.400	0.000	0.032	0.032	0.404	0.004	0.110	0.110
	β_3	1.200	0.000	0.033	0.033	1.197	-0.003	0.115	0.115
	ρ	-0.319	-0.019	0.116	0.117	-0.554	-0.254	0.290	0.386
<u>BGMME</u>	λ	0.801	0.001	0.013	0.013	0.951	0.151	0.055	0.161
	β_1	0.700	0.000	0.031	0.031	0.259	-0.441	0.295	0.530
	β_2	0.400	0.000	0.033	0.033	0.149	-0.251	0.186	0.313
	β_3	1.199	-0.001	0.033	0.033	0.438	-0.762	0.489	0.905
	ρ	-0.302	-0.002	0.116	0.116	-6.161	-5.861	3.681	6.921
<u>RGS2SLSE</u>	λ	0.801	0.001	0.012	0.012	0.806	0.006	0.037	0.038
	β_1	0.700	0.000	0.031	0.031	0.694	-0.006	0.109	0.109
	β_2	0.400	0.000	0.032	0.032	0.401	0.001	0.110	0.110
	β_3	1.199	-0.001	0.032	0.033	1.188	-0.012	0.116	0.117
	ρ	-0.313	-0.013	0.115	0.115	0.362	-0.062	0.233	0.241
<u>RB2SLSE</u>	λ	0.800	0.000	0.012	0.012	0.794	-0.006	0.041	0.042
	β_1	0.700	0.000	0.031	0.031	0.700	0.000	0.108	0.108
	β_2	0.400	0.000	0.032	0.032	0.404	0.004	0.110	0.110
	β_3	1.199	-0.001	0.033	0.033	1.197	-0.003	0.114	0.114
	ρ	-0.313	-0.013	0.115	0.115	-0.362	-0.062	0.233	0.241
<u>RGMME₁</u>	λ	0.800	0.000	0.012	0.012	0.795	-0.005	0.038	0.039
	β_1	0.700	0.000	0.031	0.031	0.699	-0.001	0.108	0.108
	β_2	0.400	0.000	0.032	0.032	0.404	0.004	0.110	0.110
	β_3	1.200	0.000	0.032	0.032	1.197	-0.003	0.115	0.116
	ρ	-0.306	-0.006	0.115	0.115	-0.301	-0.001	0.247	0.247

References

- Abadir, Karim M. and Jan R. Magnus (2005). *Matrix Algebra*. Econometric Exercises. New York: Cambridge University Press. ISBN: 9780521822893.
- Amemiya, Takeshi (1985). *Advanced Econometrics*. Cambridge: Harvard University Press.
- Anselin, Luc (1988). *Spatial econometrics: Methods and Models*. New York: Springer.
- Anselin, Luc (2007). “Spatial Econometrics”. In: *Palgrave Handbook of Econometrics: Volume 1, Econometric Theory*. Ed. by Kerry Patterson and Terence C. Mills. Palgrave Macmillan.
- Arraiz, Irani et al. (2010). “A Spatial Cliff-Ord-Type Model with Heteroskedastic Innovations: Small and Large Sample Results”. In: *Journal of Regional Science* 50.2, pp. 592–614.
- Badinger, Harald and Peter Egger (2011). “Estimation of higher-order spatial autoregressive cross-section models with heteroscedastic disturbances”. In: *Papers in Regional Science* 90.1, pp. 213–235. ISSN: 1435-5957.
- Barry, R. and R.K. Pace (1999). “Monte Carlo estimates of the log determinant of large sparse matrices”. In: *Linear Algebra Appl.* 289, pp. 41–54.
- Case, Anne C. (1991). “Spatial Patterns in Household Demand”. In: *Econometrica* 59.4, pp. 953–965. URL: <http://www.jstor.org/stable/2938168>.
- (1992). “Neighborhood influence and technological change”. In: *Regional Science* 22, pp. 491–508.
- Conley, Timothy G. (1999). “GMM estimation with cross sectional dependence”. In: *Journal of Econometrics* 92, pp. 1–45.
- Das, Debabrata, Harry H. Kelejian, and Ingmar R. Prucha (2003). “Small Sample Properties of Estimators of Spatial Autoregressive Models with Autoregressive Disturbances”. In: *Papers in Regional Science* 82, pp. 1–26.
- Drukker, David M., Peter Egger, and Ingmar R. Prucha (2012). “On Two-step Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances and Endogenous Regressors, forthcoming”. In: *Econometric Reviews*.
- Elhorst, J. Paul, Donald J. Lacombe, and Gianfranco Piras (2012). “On model specification and parameter space definitions in higher order spatial econometric models”. In: *Regional Science and Urban Economics* 42.1 - 2, pp. 211 –220. URL: <http://www.sciencedirect.com/science/article/pii/S0166046211001104>.
- Elhorst, J. Paul (2010). “Applied Spatial Econometrics: Raising the Bar”. In: *Spatial Economic Analysis* 5.1, pp. 9–28.
- Hansen, Lars Peter (1982). “Large Sample Properties of Generalized Method of Moments Estimators”. In: *Econometrica* 50.4, pp. 1029–1054.
- 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.
- (1999). “A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model”. In: *International Economic Review* 40.2, pp. 509–533.

- Kelejian, Harry H. and Ingmar R. Prucha (2001). "On the Asymptotic Distribution of the Moran I Test Statistic with Applications". In: *Journal of Econometrics* 104.2, pp. 219–257.
- (2002). "2SLS and OLS in a spatial autoregressive model with equal spatial weights". In: *Regional Science and Urban Economics* 32, pp. 691–707.
- (2007). "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances, Department of Economics, University of Maryland, July."
- (2010). "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances". In: *Journal of Econometrics* 157, pp. 53–67.
- Lee, Lung-fei (2001a). "Generalized Method of Moments Estimation of Spatial Autoregressive Processes Manuscript, Department of Economics, Ohio State University". URL: <http://economics.sbs.ohio-state.edu/pdf/lflee/gmm-sar-proc\%281\%29.pdf>.
- (2001b). "GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models Manuscript, Department of Economics, Ohio State University". URL: <http://www.econ.ohio-state.edu/pdf/lflee/gmm-mrsar\%281\%29.pdf>.
- (2003). "Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances". In: *Econometric Reviews* 22.4, pp. 307–335.
- (2004). "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models". In: *Econometrica* 72.6, pp. 1899–1925.
- (2007a). "GMM and 2SLS estimation of mixed regressive, spatial autoregressive models". In: *Journal of Econometrics* 137.2, pp. 489–514.
- (2007b). "The Method of Elimination and Substitution in the GMM Estimation of Mixed Regressive, Spatial Autoregressive Models". In: *Journal of Econometrics* 140, pp. 155–189. URL: <http://www.sciencedirect.com/science/article/pii/S0304407606002272>.
- Lee, Lung-fei and Xiaodong Liu (2010b). "Efficient GMM Estimation Of High Order Spatial Autoregressive Models With Autoregressive Disturbances". In: *Econometric Theory* 26.01, pp. 187–230.
- LeSage, James and Robert K. Pace (2004). "Chebyshev approximation of log-determinants of spatial weight matrices". In: *Computational Statistics & Data Analysis* 45, pp. 179–196.
- (2007). "A matrix exponential spatial specification". In: *Journal of Econometrics* 140, pp. 190–214.
- (2009). *Introduction to Spatial Econometrics (Statistics: A Series of Textbooks and Monographs)*. London: Chapman and Hall/CRC.
- Lesage, James P. (1997). "Bayesian Estimation of Spatial Autoregressive Models". In: *International Regional Science Review* 20.1-2, pp. 113–129. DOI: doi:10.1177/016001769702000107.
- 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.
- Liu, Xiaodong, Lung-fei Lee, and Christopher R. Bollinger (2010). "An efficient GMM estimator of spatial autoregressive models". In: *Journal of Econometrics* 159.2, pp. 303–319.

- Newey, Whitney K. and Daniel McFadden (1994). “Chapter 36 Large sample estimation and hypothesis testing”. In: ed. by Robert F. Engle and Daniel L. McFadden. Vol. 4. *Handbook of Econometrics*. Elsevier, pp. 2111–2245.
- Ord, Keith (1975). “Estimation Methods for Models of Spatial Interaction”. In: *Journal of the American Statistical Association* 70.349, pp. 120–126. URL: <http://www.jstor.org/stable/2285387>.
- Pace, R.K. and R. Barry (1997a). “Quick computation of spatial autoregressive estimators”. In: *Geogr. Anal.* 29, pp. 232–246.
- (1997b). “Sparse spatial autoregressions”. In: *Statist. Probab. Lett.* 33, pp. 291–297.
- Parent, Olivier and James P. Lesage (2007). “Bayesian Model Averaging for Spatial Econometric Models”. In: *Geographical Analysis* 39, pp. 241–267. DOI: 10.1111/j.1538-4632.2007.00703.x.
- Prucha, Ingmar R. (Forthcoming 2012). “Generalized Method of Moments Estimation of Spatial Models”. In: *Handbook of Regional Science, Spatial Econometrics*. URL: http://econweb.umd.edu/~prucha/Papers/WP_IRP_HRS_2012.pdf.
- Smirnov, Oleg and Luc Anselin (2001). “Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach”. In: *Computational Statistics & Data Analysis* 35.3, pp. 301–319. URL: <http://www.sciencedirect.com/science/article/pii/S0167947300000189>.
- White, Halbert G. (1980). “A Heteroskedasticity-Consistent Covariance Matrix Estimator a Direct Test for Heteroskedasticity”. In: *Econometrica* 48, pp. 817–838.
- Yuzefovich, Yevgeny, Harry H. Kelejian, and Ingmar R. Prucha (2006). “Estimation Problems in Models with Spatial Weighting Matrices Which have Blocks of Equal Elements”. In: *Journal of Regional Science* 46.3, pp. 507–515.