Quasi maximum likelihood estimation for simultaneous spatial autoregressive models

Wang, Luya and Li, Kunpeng and Wang, Zhengwei

November 2014

Quasi maximum likelihood estimation for simultaneous spatial autoregressive models

Luya Wang* and Kunpeng Li† and Zhengwei Wang‡

Abstract

This paper considers the problem of estimating a simultaneous spatial autoregressive model (SSAR). We propose using the quasi maximum likelihood method to estimate the model. The asymptotic properties of the maximum likelihood estimator including consistency and limiting distribution are investigated. We also run Monte Carlo simulations to examine the finite sample performance of the maximum likelihood estimator.

Key Words: Simultaneous equations model, Spatial autoregressive model, Maximum likelihood estimation, Asymptotic theory.

*The corresponding author. School of Banking and Finance, University of International Business and Economics, Beijing, China, 100029, Email:wangluya3501@163.com

†School of International Economics and Management, Capital University of Economics and Business, Beijing, China,

‡PBC School of Finance, Tsinghua University, Beijing, China, 100083.

1 Introduction

Spatial econometric models provide an effective way to study the spatial interactions among units and are widely used in urban, real estate, regional, public, agricultural, environmental economics. Among the spatial models, spatial autoregressive (SAR) model proposed by Cliff and Ord (1973) has received much attention. Popular methods to estimate SAR models include the maximum likelihood method (Anselin, 1988; Lee, 2004; Baltagi and Bresson, 2011) and the generalized moments method (Kelejian and Prucha, 1998; Kelejian and Prucha, 1999; Baltagi and Liu, 2011). Readers are referred to Lee and Yu (2010) for a survey on a recent development of spatial models.

The spatial econometric literature, so far, focuses mainly on the single equation SAR model. In structural economic models, the endogenous variables, which possibly have spatial effects, are simultaneously determined in equilibrium. As a result, a single equation model may not be appropriate to estimate the structural parameters. This motivates the necessity to extend the single equation model to a multiple equation system. In this paper, we study the estimation and inferential theory of a simultaneous spatial autoregressive (SSAR) model. We establish the asymptotic theory of the maximum likelihood estimator including consistency and limiting distribution, which is new to the spatial econometric literature.

A related work to our paper is Baltagi and Deng (2012) who consider estimating a SSAR model in a random effects panel data framework. They propose a three-stage least squares method to estimate the coefficients, but they do not study the asymptotic properties of the estimator. In this paper we use the quasi maximum likelihood method to estimate a SSAR model under cross-sectional data setup. Although we focus on the cross-sectional data, the result of this paper, with extra efforts, can be extended to deal with a fixed effects panel data model. This would complement Baltagi and Deng's work.

The rest of the paper is organized as follows. Section 2 describes the model and lists the assumptions that are needed for the asymptotic analysis. Section 3 presents the quasi likelihood function and the asymptotic theory of the MLE. Section 4 conducts Monte Carlo simulations to investigate the finite sample performance of the MLE. Section 5 concludes the paper. In appendix A, we give a detailed expressions for two matrices that are important parts of the limiting variance. The technical materials including the proofs of the main results of the paper are delegated to the supplementary appendix B.

2 Model and Assumptions

We consider the following SSAR model

$$Y_1 = \rho_1 W_1 Y_1 + \gamma_1 Y_2 + X_1 \beta_1 + e_1 \tag{1}$$

$$Y_2 = \rho_2 W_2 Y_2 + \gamma_2 Y_1 + X_2 \beta_2 + e_2 \tag{2}$$

where Y_1 and Y_2 are both $N \times 1$ dependent variables. W_1 and W_2 are respective $N \times N$ spatial weights matrices. X_1 is a set of $N \times k_1$ explanatory variables and β_1 is the corresponding k_1-dimensional vector of coefficients. X_2 is a set of $N \times k_2$ explanatory variables and β_2 is the corresponding k_2-dimensional vector of coefficients.

In the SSAR model (1) and (2), we consider a simple case of two equations system. We note that this is just for expositional simplicity. Extension to more equations system involves no fundamentally new contents.

The SSAR model can be rewritten as

$$\underbrace{\begin{bmatrix} I_N - \rho_1 W_1 & -\gamma_1 I_N \\ -\gamma_2 I_N & I_N - \rho_2 W_2 \end{bmatrix}}_{\equiv S(\rho_1,\rho_2,\gamma_1,\gamma_2)} \underbrace{\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}}_{\equiv Y} = \underbrace{\begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}}_{\equiv X} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}}_{\equiv \beta} + \underbrace{\begin{bmatrix} e_1 \\ e_2 \end{bmatrix}}_{\equiv e}. \tag{3}$$

Let $\delta = (\rho_1, \rho_2, \gamma_1, \gamma_2)$, the above model is equivalent to

$$S(\delta)Y = X\beta + e, \tag{4}$$

where Y, X, β and e are defined in (3). Throughout the paper, we use the symbols with asterisk to denote the underlying true values, for example, ρ_1^*, γ_1^*, etc. Let $\delta^* = (\rho_1^*, \rho_2^*, \gamma_1^*, \gamma_2^*)$. We make the following assumptions for the subsequent analysis.

Assumption A: Let e_{1i} and e_{2i} be the disturbances of the two equations corresponding to the ith observation. We assume that e_{1i} and e_{2i} are mutually independent with $E(e_{1i}^{4+\kappa}) \leq C$ and $E(e_{2i}^{4+\kappa}) \leq C$ for some given κ, where C is a generic positive constant. In addition e_{1i} and e_{2i} are both independent and identically distributed over i with the variances σ_1^2 and σ_2^2, respectively.

Assumption B: Matrix $S(\delta)$ is invertible for all $\delta \in \Delta$, where Δ is a compact set, and δ^* is an interior point of Δ.

Assumption C: W_1 and W_2 are $N \times N$ exogenously spatial weights matrices such that the diagonal elements of W_1 and W_2 are all zeros. In addition, both W_1 and W_2 are bounded in absolute value in column and row sums. Furthermore, $S^{-1}(\delta)$ is bounded in absolute value in column and row sums uniformly in $\delta \in \Delta$.

Assumption D: The elements of X are nonrandom and bounded in absolute value by some constant C. In addition, $Q = \lim\limits_{N\to\infty} \frac{1}{N} X'X$ exists and is positively definite.

Assumption E: One of the following two conditions holds:

E1 For any $\delta \neq \delta^*$, $\lim\limits_{N\to\infty} [(P - P^*)S^{-1}(\delta^*)X\beta^*, X]'[(P - P^*)S^{-1}(\delta^*)X\beta^*, X]/N$ is positively definite, where

$$P = \begin{bmatrix} \rho_1 W_1 & \gamma_1 I_N \\ \gamma_2 I_N & \rho_2 W_2 \end{bmatrix}, \qquad P^* = \begin{bmatrix} \rho_1^* W_1 & \gamma_1^* I_N \\ \gamma_2^* I_N & \rho_2^* W_2 \end{bmatrix}.$$

E2 Let $R(\delta) = S(\delta)'S^{-1\prime}(\delta^*)\Sigma_{ee}^* S^{-1}(\delta^*)S(\delta)$ where $\Sigma_{ee}^* = \mathrm{var}(e) = \mathrm{diag}(\sigma_1^{*2} I_N, \sigma_2^{*2} I_N)$, and $R_{11}(\delta)$ and $R_{22}(\delta)$ be the respective left-upper and right-lower $N \times N$ submatrices of $R(\delta)$. Then for any $\delta \neq \delta^*$,

$$\liminf_{N\to\infty} \left(\ln \left| \frac{1}{N} \mathrm{tr}[R_{11}(\delta)] \right| + \ln \left| \frac{1}{N} \mathrm{tr}[R_{22}(\delta)] \right| - \frac{1}{N} \ln \left| R(\delta) \right| \right) > 0.$$

Remark: Assumption A imposes some regularity conditions on the disturbances. It allows for that the disturbances from different equations to have different variances. Assumptions B and D put some conditions on the spatial weights matrices, the explanatory variables and the underlying parameters. These conditions are standard in the spatial econometric models. Similar assumptions are also made in Yu et al. (2008). Assumption D assumes that the explanatory variables are nonrandom. If the explanatory variable is random but independent with the disturbance, then the analysis of this paper can be viewed as conditional on the realizations of explanatory variables. Assumption E is the identification for the parameters δ. Assumption E1 is a local identification condition since it depends on the underlying value β^*. If $\beta^* = 0$, Assumption E1 breaks down. To account for this possibility, we need a stronger condition. This gives Assumption E2, which is a global identification that does not depend on β^*. Assumptions E1 and E2 correspond to Assumptions 8 and 9 in Lee (2004), respectively. But our conditions are more complicated. This is because the transformation matrix $S(\delta)$ has a more complicated form and the errors in (4) are not homoscedastic.

3 Likelihood function and asymptotic theory

Suppose that e_{1i} and e_{2i} are normally distributed. Therefore the log-likelihood function is

$$\mathcal{L}(\theta) = -\frac{1}{2N} \ln |\Sigma_{ee}| + \frac{1}{N} \ln |S(\delta)| - \frac{1}{2N} [S(\delta)Y - X\beta]' \Sigma_{ee}^{-1} [S(\delta)Y - X\beta] \tag{5}$$

where $\Sigma_{ee} = \mathrm{diag}(\sigma_1^2 I_N, \sigma_2^2 I_N)$ and $\theta = (\delta, \beta_1, \beta_2, \sigma_1^2, \sigma_2^2)$. Let Θ be the parameters space that are specified by Assumptions B and C. The quasi maximum likelihood estimator (MLE) is defined as

$$\hat{\theta} = \underset{\theta \in \Theta}{\mathrm{argmax}}\ \mathcal{L}(\theta).$$

The following theorem delivers the limiting distribution of the MLE. The consistency and the rate of convergence are implicitly given by the theorem.

Theorem 1 *Under Assumptions A-E, when $N \to \infty$,*

$$\sqrt{N}(\hat{\theta} - \theta^*) \xrightarrow{d} N(0, \Omega^{\star -1}(\Omega^\star + \Sigma^\star)\Omega^{\star -1}).$$

with $\Omega^\star = \lim_{N \to \infty} \Omega$ and $\Sigma^\star = \lim_{N \to \infty} \Sigma$, where Ω and Σ are defined by (6) and (7) in Appendix.

The above theorem shows that the limiting variance of the MLE has a sandwich expression. This is a well-known result in Quasi-MLE (see Lee (2004)), due to the misspecification of the distribution of the errors. However, if the errors are normally distributed, the distribution of errors is correctly specified. Then the limiting variance has a more elegant expression, which is stated in the following theorem.

Theorem 2 *Under the assumptions of Theorem 1, if e_{1i} and e_{2i} are normally distributed, when $N \to \infty$, we have*

$$\sqrt{N}(\hat{\theta} - \theta^*) \xrightarrow{d} N(0, \Omega^{\star -1}).$$

Theorem 2 follows from Theorem 1 directly by noting that $\Sigma^\star = 0$ when e_{1i} and e_{2i} are normally distributed.

4 Monte Carlo simulations

We run the Monte Carlo simulations to investigate the finite sample performance of the MLE. The data are generated according to

$$Y_1 = \alpha_1 + \rho_1 W_1 Y_1 + \gamma_1 Y_2 + x_1 \zeta_1 + e_1$$
$$Y_2 = \alpha_2 + \rho_2 W_2 Y_2 + \gamma_2 Y_1 + x_2 \zeta_2 + e_2$$

with $\alpha_1 = 1$, $\alpha_2 = 1$, $\rho_1 = 0.3$, $\rho_2 = 0.4$, $\gamma_1 = 0.2$, $\gamma_2 = 0.4$, $\zeta_1 = 1$, $\zeta_2 = 2$, $\sigma_1^2 = 0.5$ and $\sigma_2^2 = 2$. Let $\beta_1 = (\alpha_1, \zeta_1)'$, $\beta_2 = (\alpha_2, \zeta_2)'$, $X_1 = (\mathbf{1}_N, x_1)$ and $X_2 = (\mathbf{1}_N, x_2)$ where $\mathbf{1}_N$ is a N-dimensional vector with all elements equal to 1. Then we have the same expressions as model (1) and (2). As for the spatial weights matrices, W_1 is fixed to be one-ahead-and-one-behind weights matrix and W_2 is fixed to be three-ahead-and-three-behind weights matrix. The "q ahead and q behind" spatial weights matrix is defined the same as in Kelejian and Prucha (1999) and Kapoor et al. (2007). More specifically, all the units are arranged in a circle and each unit is affected only by the q units immediately before it and immediately after it with equal weight. Following Kelejian and Prucha (1999), we normalize the spatial weights matrix by letting the sum of each row be equal to 1 (so the weight is $\frac{1}{2q}$). The spatial weights matrix generated in this way is called "q ahead and q behind". The other terms such as x_{1i}, x_{2i}, e_{1i} and e_{2i} are all generated independently from $N(0, 1)$. The following table presents the performance of the MLE under different sample size, which are obtained by 1000 repetitions. From the table, we see that the MLE performs well in our simulations.

Table 1: The performance of the MLE

	ρ_1		ρ_2		γ_1		γ_1	
N	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
50	-0.0137	0.0849	-0.0602	0.1473	0.0009	0.0408	-0.0068	0.1666
100	-0.0096	0.0571	-0.0300	0.0955	0.0021	0.0269	0.0002	0.1127
150	-0.0036	0.0466	-0.0169	0.0708	-0.0005	0.0210	0.0057	0.0877
200	-0.0042	0.0390	-0.0172	0.0611	-0.0013	0.0195	0.0042	0.0754
	ζ_1		ζ_2		σ_1^2		σ_2^2	
N	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
50	-0.0029	0.1070	-0.0133	0.2202	-0.0365	0.1056	-0.1127	0.4398
100	-0.0010	0.0712	0.0010	0.1494	-0.0218	0.0742	-0.0768	0.3098
150	-0.0021	0.0599	0.0012	0.1289	-0.0124	0.0586	-0.0464	0.2536
200	0.0010	0.0511	-0.0003	0.1110	-0.0119	0.0499	-0.0385	0.2182

5 Conclusion

This paper proposes the simultaneous spatial autoregressive models. We consider the maximum likelihood method to estimate the parameters in the model. The asymptotic properties of the MLE, including consistency and limiting distribution, are investigated. Simulations confirm that the MLE performs well in the finite sample.

Appendix A: The expressions of Ω and Σ

In this appendix, we give the explicit expressions of Ω and Σ. The matrix Ω is defined as

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} & \Omega_{14} & \Omega_{15} & 0 & \Omega_{17} & 0 \\ * & \Omega_{22} & \Omega_{23} & \Omega_{24} & 0 & \Omega_{26} & 0 & \Omega_{28} \\ * & * & \Omega_{33} & \Omega_{34} & \Omega_{35} & 0 & \Omega_{37} & 0 \\ * & * & * & \Omega_{44} & 0 & \Omega_{46} & 0 & \Omega_{48} \\ * & * & * & * & \Omega_{55} & 0 & 0 & 0 \\ * & * & * & * & * & \Omega_{66} & 0 & 0 \\ * & * & * & * & * & * & \frac{1}{2\sigma_1^{*4}} & 0 \\ * & * & * & * & * & * & * & \frac{1}{2\sigma_2^{*4}} \end{bmatrix} \tag{6}$$

with

$$\Omega_{11} = \frac{1}{N\sigma_1^{*2}}\Big[(V_{11}^*X_1\beta_1^* + V_{12}^*X_2\beta_2^*)'W_1'W_1(V_{11}^*X_1\beta_1^* + V_{12}^*X_2\beta_2^*) + \sigma_1^{*2}\mathrm{tr}(W_1V_{11}^*W_1V_{11}^*)$$
$$+ \sigma_1^{*2}\mathrm{tr}(V_{11}^{*\prime}W_1'W_1V_{11}^*) + \sigma_2^{*2}\mathrm{tr}(V_{12}^{*\prime}W_1'W_1V_{12}^*)\Big],$$
$$\Omega_{12} = \frac{1}{N}\mathrm{tr}(W_1V_{12}^*W_2V_{21}^*),$$
$$\Omega_{13} = \frac{1}{N\sigma_1^{*2}}\Big[(V_{11}^*X_1\beta_1^* + V_{12}^*X_2\beta_2^*)'W_1'(V_{21}^*X_1\beta_1^* + V_{22}^*X_2\beta_2^*) + \sigma_1^{*2}\mathrm{tr}(W_1V_{11}^*V_{21}^*)$$
$$+ \sigma_1^{*2}\mathrm{tr}(V_{21}^{*\prime}W_1V_{11}^*) + \sigma_2^{*2}\mathrm{tr}(V_{22}^{*\prime}W_1V_{12}^*)\Big],$$
$$\Omega_{14} = \frac{1}{N}\mathrm{tr}(V_{12}^*V_{11}^*W_1),$$
$$\Omega_{15} = \frac{1}{N\sigma_1^{*2}}(V_{11}^*X_1\beta_1^* + V_{12}^*X_2\beta_2^*)'W_1'X_1,$$
$$\Omega_{17} = \frac{1}{N\sigma_1^{*2}}\mathrm{tr}(W_1V_{11}^*),$$
$$\Omega_{22} = \frac{1}{N\sigma_2^{*2}}\Big[(V_{21}^*X_1\beta_1^* + V_{22}^*X_2\beta_2^*)'W_2'W_2(V_{21}^*X_1\beta_1^* + V_{22}^*X_2\beta_2^*) + \sigma_2^{*2}\mathrm{tr}(W_2V_{22}^*W_2V_{22}^*)$$
$$+ \sigma_2^{*2}\mathrm{tr}(V_{22}^{*\prime}W_2'W_2V_{22}^*) + \sigma_1^{*2}\mathrm{tr}(V_{21}^{*\prime}W_2'W_2V_{21}^*)\Big],$$
$$\Omega_{23} = \frac{1}{N}\mathrm{tr}(W_2V_{21}^*V_{22}^*),$$
$$\Omega_{24} = \frac{1}{N\sigma_2^{*2}}\Big[(V_{21}^*X_1\beta_1^* + V_{22}^*X_2\beta_2^*)'W_2'(V_{11}^*X_1\beta_1^* + V_{12}^*X_2\beta_2^*)$$
$$+ \sigma_2^{*2}\mathrm{tr}(W_2V_{22}^*V_{12}^*) + \sigma_2^{*2}\mathrm{tr}(V_{12}^{*\prime}W_2V_{22}^*) + \sigma_1^{*2}\mathrm{tr}(V_{11}^{*\prime}W_2V_{21}^*)\Big],$$
$$\Omega_{26} = \frac{1}{N\sigma_2^{*2}}(V_{21}^*X_1\beta_1^* + V_{22}^*X_2\beta_2^*)'W_2'X_1,$$
$$\Omega_{28} = \frac{1}{N\sigma_2^{*2}}\mathrm{tr}(W_2V_{22}^*),$$
$$\Omega_{33} = \frac{1}{N\sigma_1^{*2}}\Big[(V_{21}^*X_1\beta_1^* + V_{22}^*X_2\beta_2^*)'(V_{21}^*X_1\beta_1^* + V_{22}^*X_2\beta_2^*)' + \sigma_1^{*2}\mathrm{tr}(V_{21}^*V_{21}^*)$$

$$+\sigma_1^{*2}\mathrm{tr}(V_{21}^{*\prime}V_{21}^*)+\sigma_2^{*2}\mathrm{tr}(V_{22}^{*\prime}V_{22}^*),$$

$$\Omega_{34}=\frac{1}{N}\mathrm{tr}(V_{11}^*V_{22}^*),$$

$$\Omega_{35}=\frac{1}{N\sigma_1^{*2}}(V_{21}^*X_1\beta_1^*+V_{22}^*X_2\beta_2^*)'X_1,$$

$$\Omega_{37}=\frac{1}{N\sigma_1^{*2}}\mathrm{tr}(V_{21}^*),$$

$$\Omega_{44}=\frac{1}{N\sigma_2^{*2}}\Big[(V_{11}^*X_1\beta_1^*+V_{12}^*X_2\beta_2^*)'(V_{11}^*X_1\beta_1^*+V_{12}^*X_2\beta_2^*)'+\sigma_2^{*2}\mathrm{tr}(V_{12}^*V_{12}^*)$$

$$+\sigma_2^{*2}\mathrm{tr}(V_{12}^{*\prime}V_{12}^*)+\sigma_1^{*2}\mathrm{tr}(V_{11}^{*\prime}V_{11}^*),$$

$$\Omega_{46}=\frac{1}{N\sigma_2^{*2}}(V_{11}^*X_1\beta_1^*+V_{12}^*X_2\beta_2^*)'X_2,$$

$$\Omega_{48}=\frac{1}{N\sigma_2^{*2}}\mathrm{tr}(V_{12}^*),$$

$$\Omega_{55}=\frac{1}{N\sigma_1^{*2}}X_1'X_1,$$

$$\Omega_{66}=\frac{1}{N\sigma_2^{*2}}X_2'X_2,$$

and the matrix Σ is defined as

$$\Sigma=\begin{bmatrix}\Sigma_{11}&0&\Sigma_{13}&0&\Sigma_{15}&0&\Sigma_{17}&0\\ *&\Sigma_{22}&0&\Sigma_{24}&0&\Sigma_{26}&0&\Sigma_{28}\\ *&*&\Sigma_{33}&0&\Sigma_{35}&0&\Sigma_{37}&0\\ *&*&*&\Sigma_{44}&0&\Sigma_{46}&0&\Sigma_{48}\\ *&*&*&*&0&0&\Sigma_{57}&0\\ *&*&*&*&*&0&0&\Sigma_{68}\\ *&*&*&*&*&*&\Sigma_{77}&0\\ *&*&*&*&*&*&*&\Sigma_{88}\end{bmatrix}\tag{7}$$

with

$$\Sigma_{11}=\frac{2\kappa_3}{N\sigma_1^{*4}}\sum_{i=1}^N W_{1,i*}V_{11,*i}^*W_{1,i*}(V_{11}^*X_1\beta_1^*+V_{12}^*X_2\beta_2^*)+\frac{\kappa_4-3\sigma_1^{*4}}{N\sigma_1^{*4}}\mathrm{tr}[(W_1V_{11}^*)\circ(W_1V_{11}^*)],$$

$$\Sigma_{13}=\frac{1}{N\sigma_1^{*4}}\Big\{\kappa_3\sum_{i=1}^N W_{1,i*}\Big[(V_{11}^*X_1\beta_1^*+V_{12}^*X_2\beta_2^*)V_{21,ii}^*+V_{11,*i}^*(V_{21,i*}^*X_1\beta_1^*+V_{22,i*}^*X_2\beta_2^*)\Big]$$

$$+(\kappa_4-3\sigma_1^{*4})\mathrm{tr}[(W_1V_{11}^*)\circ V_{21}^*]\Big\},$$

$$\Sigma_{15}=\frac{\kappa_3}{N\sigma_1^{*4}}\sum_{i=1}^N W_{1,i*}V_{11,*i}^*X_{1,i},$$

$$\Sigma_{17}=\frac{1}{2N\sigma_1^{*6}}\Big[\kappa_3\sum_{i=1}^N W_{1,i*}(V_{11}^*X_1\beta_1^*+V_{12}^*X_2\beta_2^*)+(\kappa_4-3\sigma_1^{*4})\mathrm{tr}(W_1V_{11}^*)\Big],$$

$$\Sigma_{22}=\frac{2\mu_3}{N\sigma_2^{*4}}\sum_{i=1}^N W_{2,i*}V_{22,*i}^*W_{2,i*}(V_{21}^*X_1\beta_1^*+V_{22}^*X_2\beta_2^*)+\frac{\mu_4-3\sigma_2^{*4}}{N\sigma_2^{*4}}\mathrm{tr}[(W_2V_{22}^*)\circ(W_2V_{22}^*)],$$

$$\Sigma_{24} = \frac{1}{N\sigma_2^{*4}}\Big\{\mu_3 \sum_{i=1}^{N} W_{2,i*}\Big[(V_{21}^* X_1\beta_1^* + V_{22}^* X_2\beta_2^*)V_{12,ii}^* + V_{22,*i}^*(V_{11,i*}^* X_1\beta_1^* + V_{12,i*}^* X_2\beta_2^*)\Big]$$
$$+ (\mu_4 - 3\sigma_2^{*4})\mathrm{tr}[(W_2V_{22}^*) \circ V_{12}^*]\Big\},$$

$$\Sigma_{26} = \frac{\mu_3}{N\sigma_2^{*4}} \sum_{i=1}^{N} W_{2,i*}V_{22,*i}^* X_{2,i},$$

$$\Sigma_{28} = \frac{1}{2N\sigma_2^{*6}}\Big[\mu_3 \sum_{i=1}^{N} W_{2,i*}(V_{21}^* X_1\beta_1^* + V_{22}^* X_2\beta_2^*) + (\mu_4 - 3\sigma_2^{*4})\mathrm{tr}(W_2V_{22}^*)\Big],$$

$$\Sigma_{33} = \frac{1}{N\sigma_1^{*4}}\Big[2\kappa_3 \sum_{i=1}^{N}(V_{21,i*}^* X_1\beta_1^* + V_{22,i*}^* X_2\beta_2^*)V_{21,ii}^* + (\kappa_4 - 3\sigma_1^{*4})\mathrm{tr}(V_{21}^* \circ V_{21}^*)\Big],$$

$$\Sigma_{35} = \frac{1}{N\sigma_1^{*4}}\kappa_3 \sum_{i=1}^{N} V_{21,ii}^* X_{1,i},$$

$$\Sigma_{37} = \frac{1}{2N\sigma_1^{*6}}\Big[\kappa_3 \sum_{i=1}^{N}(V_{21,i*}^* X_1\beta_1^* + V_{22,i*}^* X_2\beta_2^*) + (\kappa_4 - 3\sigma_1^{*4})\mathrm{tr}(V_{21}^*)\Big],$$

$$\Sigma_{44} = \frac{1}{N\sigma_2^{*4}}\Big[2\mu_3 \sum_{i=1}^{N}(V_{11,i*}^* X_1\beta_1^* + V_{12,i*}^* X_2\beta_2^*)V_{12,ii}^* + (\mu_4 - 3\sigma_2^{*4})\mathrm{tr}(V_{12}^* \circ V_{12}^*)\Big],$$

$$\Sigma_{46} = \frac{1}{N\sigma_2^{*4}}\mu_3 \sum_{i=1}^{N} V_{12,ii}^* X_{2,i},$$

$$\Sigma_{48} = \frac{1}{2N\sigma_2^{*6}}\Big[\mu_3 \sum_{i=1}^{N}(V_{11,i*}^* X_1\beta_1^* + V_{12,i*}^* X_2\beta_2^*) + (\mu_4 - 3\sigma_2^{*4})\mathrm{tr}(V_{12}^*)\Big],$$

$$\Sigma_{57} = \frac{1}{2N\sigma_1^{*6}}\kappa_3 \sum_{i=1}^{N} X_{1,i},$$

$$\Sigma_{68} = \frac{1}{2N\sigma_2^{*6}}\mu_3 \sum_{i=1}^{N} X_{2,i},$$

$$\Sigma_{77} = \frac{1}{4\sigma_1^{*8}}(\kappa_4 - 3\sigma_1^{*4}),$$

$$\Sigma_{88} = \frac{1}{4\sigma_2^{*8}}(\mu_4 - 3\sigma_2^{*4}),$$

The symbols appearing in the above expression are defined as follows: $\kappa_3 = E(e_{1i}^3), \mu_3 = E(e_{2i}^3), \kappa_4 = E(e_{1i}^4)$ and $\mu_4 = E(e_{2i}^4)$; $V^* = S^{-1}(\delta^*)$ and $V_{11}^*, V_{12}^*, V_{21}^*, V_{22}^*$ are defined by

$$V^* = \begin{bmatrix} V_{11}^* & V_{12}^* \\ V_{21}^* & V_{22}^* \end{bmatrix}.$$

M_{i*} denotes the ith row of M; M_{*i} denotes the ith column of M and M_{ij} denotes the (i, j)th element of M. In addition, "$\circ$" denotes the Hadamard product.

References

Anselin, L. (1988). Spatial econometrics: methods and models. The Netherlands: Kluwer Academic Publishers.

Baltagi, B. H. and Deng, Y. (2012). EC3SLS estimator for a simultaneous system of spatial autoregressive equations with random effects. *Econometric Review*, Forthcoming.

Baltagi, B. H. and Bresson, G. (2011). Maximum likelihood estimation and lagrange multiplier tests for panel seemingly unrelated regressions with spatial lag and spatial errors: an application to hedonic housing prices in Paris, *Journal of Urban Economics*, 69(1), 24-42.

Baltagi, B. H. and Liu, L. (2011). An improved generalized moments estimator for a spatial moving average error model, *Economics Letters*, 282-284.

Cliff, A. D., and Ord, J. K. (1973) *Spatial autocorrelation,* London: Pion Ltd.

Kapoor, M., Kelejian, H. H., and Prucha, I. R. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics*, 140(1), 97-130.

Kelejian, H. H., and Prucha, I. R. (1998) A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1), 99-121.

Kelejian, H. H., and Prucha, I. R. (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. *International economic review*, 40(2), 509-533.

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

Lee L., and Yu J. (2010). Some recent developments in spatial panel data models. *Regional Science and Urban Economics*, 40(5), 255–271.

Yu, J., de Jong, R., and Lee, L. (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics*, 146(1), 118-134.

Supplementary Appendix B: Quasi maximum likelihood estimation for simultaneous spatial autoregressive models

In this supplementary Appendix B, we provide a detailed proof for Theorem 1. Consider the following likelihood function

$$\begin{aligned}\mathcal{L}(\theta) = & -\frac{1}{2}\ln\sigma_1^2 - \frac{1}{2}\ln\sigma_2^2 + \frac{1}{N}\ln|S(\delta)| - \frac{1}{2N}[S(\delta)Y - X\beta]'\Sigma_{ee}^{-1}[S(\delta)Y - X\beta] \\ & + \frac{1}{2}\ln\sigma_1^{*2} + \frac{1}{2}\ln\sigma_2^{*2} - \frac{1}{N}\ln|S(\delta^*)| + 1\end{aligned}$$

where $\theta = (\delta, \beta_1', \beta_2', \sigma_1^2, \sigma_2^2)'$. The above objective function is only different from the original likelihood function with a constant and will be treated as the objective function in the subsequent analysis. Given $\delta, \sigma_1^2, \sigma_2^2$, it is seen that the objective function is maximized at

$$\beta_1^\star(\delta) = (X_1'X_1)^{-1}X_1'(Y_1 - \rho_1 W_1 Y_1 - \gamma_1 Y_2), \tag{S.1}$$

$$\beta_2^\star(\delta) = (X_2'X_2)^{-1}X_2'(Y_2 - \rho_2 W_2 Y_2 - \gamma_2 Y_1). \tag{S.2}$$

Using the above two equations to concentrate out β_1 and β_2, the objective function is

$$\begin{aligned}\mathcal{L}(\theta) = & -\frac{1}{2}\ln\sigma_1^2 - \frac{1}{2}\ln\sigma_2^2 + \frac{1}{N}\ln|S(\delta)| \\ & - \frac{1}{2N\sigma_1^2}(Y_1 - \rho_1 W_1 Y_1 - \gamma_1 Y_2)'M_{X_1}(Y_1 - \rho_1 W_1 Y_1 - \gamma_1 Y_2) \\ & - \frac{1}{2N\sigma_2^2}(Y_2 - \rho_2 W_2 Y_2 - \gamma_2 Y_1)'M_{X_2}(Y_2 - \rho_2 W_2 Y_2 - \gamma_2 Y_1) \\ & + \frac{1}{2}\ln\sigma_1^{*2} + \frac{1}{2}\ln\sigma_2^{*2} - \frac{1}{N}\ln|S(\delta^*)| + 1\end{aligned}$$

Again, given $\rho_1, \rho_2, \gamma_1, \gamma_2$, it is seen that the above objective function is optimized

$$\sigma_1^{\star 2}(\delta) = \frac{1}{N}(Y_1 - \rho_1 W_1 Y_1 - \gamma_1 Y_2)'M_{X_1}(Y_1 - \rho_1 W_1 Y_1 - \gamma_1 Y_2), \tag{S.3}$$

$$\sigma_2^{\star 2}(\delta) = \frac{1}{N}(Y_2 - \rho_2 W_2 Y_2 - \gamma_2 Y_1)'M_{X_2}(Y_2 - \rho_2 W_2 Y_2 - \gamma_2 Y_1). \tag{S.4}$$

Using the preceding two solutions to further concentrate out σ_1^2 and σ_2^2, the objective function now is

$$\mathcal{L}(\delta) = -\frac{1}{2}\ln\sigma_1^{\star 2}(\delta) - \frac{1}{2}\ln\sigma_2^{\star 2}(\delta) + \frac{1}{N}\ln|S(\delta)| + \frac{1}{2}\ln\sigma_1^{*2} + \frac{1}{2}\ln\sigma_2^{*2} - \frac{1}{N}\ln|S(\delta^*)|.$$

Hereafter, we use S^* to represent $S(\delta^*)$ for notational simplicity. Let

$$P = \begin{bmatrix}\rho_1 W_1 & \gamma_1 I_N \\ \gamma_2 I_N & \rho_2 W_2\end{bmatrix}, \qquad P^* = \begin{bmatrix}\rho_1^* W_1 & \gamma_1^* I_N \\ \gamma_2^* I_N & \rho_2^* W_2\end{bmatrix}.$$

By the definition, we have

$$\begin{bmatrix}Y_1 - \rho_1 W_1 Y_1 - \gamma_1 Y_2 \\ Y_2 - \rho_2 W_2 Y_2 - \gamma_2 Y_1\end{bmatrix} = S(\delta)S^{*-1}(X\beta + e) = X\beta + e - (P - P^*)S^{*-1}(X\beta + e).$$

Let $V^* = S^{*-1}$ and $R(\delta) = S(\delta)S^{*-1}\Sigma_{ee}^* S^{*-1\prime}S(\delta)'$. For ease of exposition, we partition $R(\delta)$ and V^* into

$$R(\delta) = \begin{bmatrix} R_{11}(\delta) & R_{12}(\delta) \\ R_{21}(\delta) & R_{22}(\delta) \end{bmatrix}, \qquad V^* = \begin{bmatrix} V_{11}^* & V_{12}^* \\ V_{21}^* & V_{22}^* \end{bmatrix}.$$

Then we have

$$\begin{aligned} Y_1 - \rho_1 W_1 Y_1 - \gamma_1 Y_2 = X_1\beta_1^* + e_1 &- \Big[(\rho_1 - \rho_1^*)W_1 V_{11}^* + (\gamma_1 - \gamma_1^*)V_{21}^*\Big](X_1\beta_1^* + e_1) \\ &- \Big[(\rho_1 - \rho_1^*)W_1 V_{12}^* + (\gamma_1 - \gamma_1^*)V_{22}^*\Big](X_2\beta_2^* + e_2) \qquad \text{(S.5)} \\ Y_2 - \rho_2 W_2 Y_2 - \gamma_2 Y_1 = X_2\beta_2^* + e_2 &- \Big[(\rho_2 - \rho_2^*)W_2 V_{21}^* + (\gamma_2 - \gamma_2^*)V_{11}^*\Big](X_1\beta_1^* + e_1) \\ &- \Big[(\rho_2 - \rho_2^*)W_2 V_{22}^* + (\gamma_2 - \gamma_2^*)V_{12}^*\Big](X_2\beta_2^* + e_2) \qquad \text{(S.6)} \end{aligned}$$

Using the above results, we can further rewrite the objective function as

$$\mathcal{L}(\delta) = \mathcal{L}_1(\delta) + \mathcal{L}_2(\delta),$$

with

$$\mathcal{L}_1(\delta) = -\frac{1}{2}\ln\Big|W_1(\delta) + \frac{1}{N}\mathrm{tr}[R_{11}(\delta)]\Big| - \frac{1}{2}\ln\Big|W_1(\delta) + \frac{1}{N}\mathrm{tr}[R_{11}(\delta)]\Big| + \frac{1}{2N}\ln|R(\delta)|$$

and

$$\begin{aligned} \mathcal{L}_2(\delta) = &\ln\Big|W_1(\delta) + \frac{1}{N}\mathrm{tr}[R_{11}(\delta)] + \mathcal{R}_1(\delta)\Big| - \ln\Big|W_1(\delta) + \frac{1}{N}\mathrm{tr}[R_{11}(\delta)]\Big| \\ &+ \ln\Big|W_2(\delta) + \frac{1}{N}\mathrm{tr}[R_{22}(\delta)] + \mathcal{R}_2(\delta)\Big| - \ln\Big|W_1(\delta) + \frac{1}{N}\mathrm{tr}[R_{22}(\delta)]\Big|, \end{aligned}$$

where

$$W_1(\delta) = \frac{1}{N}U_1(\delta)'M_{X_1}U_1(\delta), \qquad W_2(\delta) = \frac{1}{N}U_2(\delta)'M_{X_2}U_2(\delta).$$

In addition,

$$\begin{aligned} \mathcal{R}_1(\delta) = &\frac{2}{N}U_1(\delta)'M_{X_1}U_3(\delta)e_1 - \frac{2}{N}U_1(\delta)'M_{X_1}U_5(\delta)e_2 - \frac{2}{N}e_1'U_3(\delta)'M_{X_1}U_5(\delta)e_2 \\ &-\frac{1}{N}e_1'U_3(\delta)'P_{X_1}U_3(\delta)e_1 + \frac{1}{N}\mathrm{tr}\Big[U_3(\delta)'U_3(\delta)(e_1e_1' - \sigma_1^{*2}I_N)\Big] \\ &-\frac{1}{N}e_2'U_5(\delta)'P_{X_1}U_5(\delta)e_2 + \frac{1}{N}\mathrm{tr}\Big[U_5(\delta)'U_5(\delta)(e_2e_2' - \sigma_2^{*2}I_N)\Big], \end{aligned}$$

and

$$\begin{aligned} \mathcal{R}_2 = &\frac{2}{N}U_2(\delta)'M_{X_2}U_4(\delta)e_2 - \frac{2}{N}U_2(\delta)'M_{X_2}U_6(\delta)e_1 - \frac{2}{N}e_1'U_6(\delta)'M_{X_2}U_4(\delta)e_2 \\ &-\frac{1}{N}e_2'U_4(\delta)'P_{X_2}U_4(\delta)e_2 + \frac{1}{N}\mathrm{tr}\Big[U_4(\delta)'U_4(\delta)(e_2e_2' - \sigma_2^{*2}I_N)\Big] \\ &-\frac{1}{N}e_1'U_6(\delta)'P_{X_2}U_6(\delta)e_1 + \frac{1}{N}\mathrm{tr}\Big[U_6(\delta)'U_6(\delta)(e_1e_1' - \sigma_1^{*2}I_N)\Big]. \end{aligned}$$

The symbols $U_1(\delta), U_2(\delta), \ldots, U_6(\delta)$ are defined as

$$
\begin{aligned}
U_1(\delta) &= (\rho_1-\rho_1^*)W_1(V_{11}^*X_1\beta_1^*+V_{12}^*X_2\beta_2^*)+(\gamma_1-\gamma_1^*)(V_{21}^*X_1\beta_1^*+V_{22}^*X_2\beta_2^*),\\
U_2(\delta) &= (\rho_2-\rho_2^*)W_2(V_{21}^*X_1\beta_1^*+V_{22}^*X_2\beta_2^*)+(\gamma_2-\gamma_2^*)(V_{11}^*X_1\beta_1^*+V_{12}^*X_2\beta_2^*),\\
U_3(\delta) &= I_N-(\rho_1-\rho_1^*)W_1V_{11}^*-(\gamma_1-\gamma_1^*)V_{21}^*,\\
U_4(\delta) &= I_N-(\rho_2-\rho_2^*)W_2V_{22}^*-(\gamma_2-\gamma_2^*)V_{12}^*,\\
U_5(\delta) &= (\rho_1-\rho_1^*)W_1V_{12}^*+(\gamma_1-\gamma_1^*)V_{22}^*,\\
U_6(\delta) &= (\rho_2-\rho_2^*)W_2V_{21}^*+(\gamma_2-\gamma_2^*)V_{11}^*.
\end{aligned}
$$

Under Assumptions A-D, it can be shown that $|\mathcal{R}_1(\delta)|=o_p(1)$ and $|\mathcal{R}_2(\delta)|=o_p(1)$ uniformly on Θ. These two results imply

$$
\sup_{\theta\in\Theta}|\mathcal{L}(\delta)-\mathcal{L}_1(\delta)|=o_p(1). \tag{S.7}
$$

To show the consistency, it suffices to show

$$
\sup_{\theta\in\mathcal{N}_\epsilon^c}(\mathcal{L}_1(\delta)-\mathcal{L}_1(\delta^*))<0.
$$

where $\mathcal{N}_\epsilon^c$ denotes the complement of an open neighborhood of δ^* in Δ of diameter of ϵ. Our normalized objective function gives $\mathcal{L}_1(\delta^*)=0$. As regards to $\mathcal{L}_1(\delta)$, which is equivalent to

$$
\begin{aligned}
\mathcal{L}_1(\delta) = &-\frac{1}{2}\Big[\ln\Big|\frac{1}{N}\mathrm{tr}[R_{11}(\delta)]\Big|-\frac{1}{N}\ln|R_{11}(\delta)|\Big]-\frac{1}{2}\Big[\ln\Big|\frac{1}{N}\mathrm{tr}[R_{22}(\delta)]\Big|-\frac{1}{N}\ln|R_{22}(\delta)|\Big]\\
&-\frac{1}{2}\Big[\ln\big[W_1(\delta)+\frac{1}{N}\mathrm{tr}[R_{11}(\delta)]\big]-\ln\big|\frac{1}{N}\mathrm{tr}[R_{11}(\delta)]\big|\Big]\\
&-\frac{1}{2}\Big[\ln\big[W_2(\delta)+\frac{1}{N}\mathrm{tr}[R_{22}(\delta)]\big]-\ln\big|\frac{1}{N}\mathrm{tr}[R_{22}(\delta)]\big|\Big]\\
&+\frac{1}{2N}\ln\Big|I_N-R_{22}^{-1/2}(\delta)R_{21}(\delta)R_{11}^{-1}(\delta)R_{12}(\delta)R_{22}^{-1/2}(\delta)\Big|.
\end{aligned}
$$

The first two terms are non-positive by the Jensen's inequality. The third and fourth terms are also non-positive since $W_1(\delta)$ and $W_2(\delta)$ are non-negative for all δ. The last term is also negative since all the eigenvalues of $I_N-R_{22}^{-1/2}(\delta)R_{21}(\delta)R_{11}^{-1}(\delta)R_{12}(\delta)R_{22}^{-1/2}(\delta)$ are no greater than 1. Given these results, we have

$$
\mathcal{L}_1(\delta)\le 0.
$$

So we only need to show that for any point $\delta\neq\delta^*$ and $\delta\in\Delta$ such that

$$
\mathcal{L}_1(\delta)<0.
$$

Notice that

$$
W_1(\delta)+W_2(\delta)=\frac{1}{N}\beta^{*\prime}X'S^{*-1}(P-P^*)'M_X(P-P^*)S^{*-1}X\beta^*.
$$

By Assumption E1, for any $\delta\neq\delta^*$, the above term is strictly greater than 0. This implies that either $W_1(\delta)$ or $W_2(\delta)$ or both $W_1(\delta)$ and $W_2(\delta)$ are greater than 0, which further imply $\mathcal{L}_1(\delta)<0$. Also notice that the expression

$$
-\frac{1}{2}\Big[\ln\Big|\frac{1}{N}\mathrm{tr}[R_{11}(\delta)]\Big|-\frac{1}{N}\ln|R_{11}(\delta)|\Big]-\frac{1}{2}\Big[\ln\Big|\frac{1}{N}\mathrm{tr}[R_{22}(\delta)]\Big|-\frac{1}{N}\ln|R_{22}(\delta)|\Big]
$$

$$+\frac{1}{2N}\ln\Big|I_N - R_{22}^{-1/2}(\delta)R_{21}(\delta)R_{11}^{-1}(\delta)R_{12}(\delta)R_{22}^{-1/2}(\delta)\Big|$$

is equivalent to

$$-\frac{1}{2}\Big[\ln\Big|\frac{1}{N}\mathrm{tr}[R_{11}(\delta)]\Big| + \ln\Big|\frac{1}{N}\mathrm{tr}[R_{22}(\delta)]\Big|\Big] + \frac{1}{2N}\ln|R(\delta)|.$$

If Assumption E2 holds, i.e., for any $\delta \neq \delta^*$,

$$\ln\Big|\frac{1}{N}\mathrm{tr}[R_{11}(\delta)]\Big| + \ln\Big|\frac{1}{N}\mathrm{tr}[R_{22}(\delta)]\Big| - \frac{1}{N}\ln|R(\delta)| \neq 0,$$

we immediately obtain that $\mathcal{L}_1(\delta) < 0$.

Now we show that $\mathcal{L}_1(\delta)$ can identify δ^*. This result, together with the uniform convergence result (S.7), gives the consistency of $\hat{\delta}$. Given the consistency of $\hat{\delta}$, by (S.1)-(S.4) and Assumption D, we obtain the consistency of $\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$. This completes the proof of consistency.

Given the consistency, we now derive the limiting distribution. By the definition of $\hat{\theta}$, we have $\frac{\partial \mathcal{L}(\hat{\theta})}{\partial \theta} = 0$. By the Taylor expansion, it follows

$$0 = \frac{\partial \mathcal{L}(\hat{\theta})}{\partial \theta} = \frac{\partial \mathcal{L}(\theta^*)}{\partial \theta} + \frac{\partial^2 \mathcal{L}(\tilde{\theta})}{\partial \theta \partial \theta'}(\hat{\theta} - \theta^*),$$

where $\tilde{\theta}$ is some point between $\hat{\theta}$ and θ^*. The above result implies

$$\hat{\theta} - \theta^* = -\Big[\frac{\partial^2 \mathcal{L}(\tilde{\theta})}{\partial \theta \partial \theta'}\Big]^{-1}\Big[\frac{\partial \mathcal{L}(\theta^*)}{\partial \theta}\Big].$$

The first order conditions for $\rho_1, \rho_2, \gamma_1, \gamma_2, \beta_1, \beta_2, \sigma_1^2$ and σ_2^2 are

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \rho_1} &= \frac{1}{N}\Big[\frac{1}{\sigma_1^2}(Y_1 - \rho_1 W_1 Y_1 - \gamma_1 Y_2 - X_1\beta_1)' W_1 Y_1 - \mathrm{tr}[S^{-1}(\delta)A_1]\Big] \\
\frac{\partial \mathcal{L}}{\partial \rho_2} &= \frac{1}{N}\Big[\frac{1}{\sigma_2^2}(Y_2 - \rho_2 W_2 Y_2 - \gamma_2 Y_2 - X_2\beta_2)' W_2 Y_2 - \mathrm{tr}[S^{-1}(\delta)A_2]\Big] \\
\frac{\partial \mathcal{L}}{\partial \gamma_1} &= \frac{1}{N}\Big[\frac{1}{\sigma_1^2}(Y_1 - \rho_1 W_1 Y_1 - \gamma_1 Y_2 - X_1\beta_1)' Y_2 - \mathrm{tr}[S^{-1}(\delta)A_3]\Big] \\
\frac{\partial \mathcal{L}}{\partial \gamma_2} &= \frac{1}{N}\Big[\frac{1}{\sigma_2^2}(Y_2 - \rho_2 W_2 Y_2 - \gamma_2 Y_2 - X_2\beta_2)' Y_1 - \mathrm{tr}[S^{-1}(\delta)A_4]\Big] \\
\frac{\partial \mathcal{L}}{\partial \beta_1} &= \frac{1}{N\sigma_1^2}X_1'(Y_1 - \rho_1 W_1 Y_1 - \gamma_1 Y_2 - X_1\beta_1) \\
\frac{\partial \mathcal{L}}{\partial \beta_2} &= \frac{1}{N\sigma_2^2}X_2'(Y_2 - \rho_2 W_2 Y_2 - \gamma_2 Y_1 - X_2\beta_2) \\
\frac{\partial \mathcal{L}}{\partial \sigma_1^2} &= \frac{1}{2N\sigma_1^4}\Big[(Y_1 - \rho_1 W_1 Y_1 - \gamma_1 Y_2 - X_1\beta_1)'(Y_1 - \rho_1 W_1 Y_1 - \gamma_1 Y_2 - X_1\beta_1) - N\sigma_1^2\Big] \\
\frac{\partial \mathcal{L}}{\partial \sigma_2^2} &= \frac{1}{2N\sigma_1^4}\Big[(Y_2 - \rho_2 W_2 Y_2 - \gamma_2 Y_1 - X_2\beta_2)'(Y_2 - \rho_2 W_2 Y_2 - \gamma_2 Y_1 - X_2\beta_2) - N\sigma_2^2\Big]
\end{aligned}$$

Notice that

$$Y = S^{-1}(\delta^*)(X\beta^* + e) = V^*(X\beta^* + e)$$

implying

$$Y_1 = V_{11}^* X_1 \beta_1^* + V_{12}^* X_2 \beta_2^* + V_{11}^* e_1 + V_{12}^* e_2 \tag{S.8}$$

$$Y_2 = V_{21}^* X_1 \beta_1^* + V_{22}^* X_2 \beta_2^* + V_{21}^* e_1 + V_{22}^* e_2 \tag{S.9}$$

Given the above results, we have

$$\frac{\partial \mathcal{L}(\theta^*)}{\partial \theta} = \frac{1}{N} \begin{bmatrix} [e_1' W_1 (V_{11}^* X_1 \beta_1^* + V_{12}^* X_2 \beta_2^*) + e_1' W_1 V_{11}^* e_1 - \sigma_1^{*2} \mathrm{tr}(W_1 V_{11}^*) + e_1' W_1 V_{12}^* e_2]/\sigma_1^{*2} \\ [e_2' W_2 (V_{21}^* X_1 \beta_1^* + V_{22}^* X_2 \beta_2^*) + e_2' W_2 V_{21}^* e_1 + e_2' W_2 V_{22}^* e_2 - \sigma_2^{*2} \mathrm{tr}(W_2 V_{22}^*)]/\sigma_2^{*2} \\ [e_1' V_{21}^* X_1 \beta_1^* + e_1' V_{22}^* X_2 \beta_2^* + e_1' V_{21}^* e_1 - \sigma_1^{*2} \mathrm{tr}(V_{21}^*) + e_1' V_{22}^* e_2]/\sigma_1^{*2} \\ [e_2' V_{11}^* X_1 \beta_1^* + e_2' V_{12}^* X_2 \beta_2^* + e_2' V_{11}^* e_1 + e_2' V_{12}^* e_2 - \sigma_2^{*2} \mathrm{tr}(V_{12}^*)]/\sigma_2^{*2} \\ X_1' e_1/\sigma_1^{*2} \\ X_2' e_2/\sigma_2^{*2} \\ (e_1' e_1 - N\sigma_1^{*2})/(2\sigma_1^{*4}) \\ (e_2' e_2 - N\sigma_2^{*2})/(2\sigma_2^{*4}) \end{bmatrix}$$

With the existence of high-order moments of e_{1i} and e_{2i} in Assumption 1, we can apply the central limit theorem for linear-quadratic forms of Kelejian and Prucha (2001) to the above expression. This gives

$$\sqrt{N} \frac{\partial \mathcal{L}(\theta^*)}{\partial \theta} \xrightarrow{d} N(0, \Omega^\star + \Sigma^\star). \tag{S.10}$$

where $\Omega^\star$ and $\Sigma^\star$ are defined in Theorem 1.

We proceed to consider the second order derivatives. By some tedious but straightforward computation, we have

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial \rho_1 \partial \rho_1} &= -\frac{1}{N}\Big(\mathrm{tr}[S^{-1}(\delta) A_1 S^{-1}(\delta) A_1] + Y' A_1' A_1 Y/\sigma_1^2\Big), \\
\frac{\partial^2 \mathcal{L}}{\partial \rho_1 \partial \rho_2} &= -\frac{1}{N}\mathrm{tr}[S^{-1}(\delta) A_2 S^{-1}(\delta) A_1], \\
\frac{\partial^2 \mathcal{L}}{\partial \rho_1 \partial \gamma_1} &= -\frac{1}{N}\Big(\mathrm{tr}[S(\delta)^{-1} A_3 S(\delta)^{-1} A_1] + Y' A_3' A_1 Y/\sigma_1^2\Big), \\
\frac{\partial^2 \mathcal{L}}{\partial \rho_1 \partial \gamma_2} &= -\frac{1}{N}\mathrm{tr}[S^{-1}(\delta) A_4 S^{-1}(\delta) A_1], \\
\frac{\partial^2 \mathcal{L}}{\partial \rho_1 \partial \beta_1} &= -\frac{1}{N\sigma_1^2} X_1' W_1 Y_1, \\
\frac{\partial^2 \mathcal{L}}{\partial \rho_1 \partial \beta_2} &= 0, \\
\frac{\partial^2 \mathcal{L}}{\partial \rho_1 \partial \sigma_1^2} &= -\frac{1}{N\sigma_1^4}(Y_1 - \rho_1 W_1 Y_1 - \gamma_1 Y_2 - X_1 \beta_1)' W_1 Y_1, \\
\frac{\partial^2 \mathcal{L}}{\partial \rho_1 \partial \sigma_2^2} &= 0, \\
\frac{\partial^2 \mathcal{L}}{\partial \rho_2 \partial \rho_2} &= -\frac{1}{N}\Big(\mathrm{tr}[S^{-1}(\delta) A_2 S^{-1}(\delta) A_2] + Y' W_2' W_2 Y\Big),
\end{aligned}$$

$$\frac{\partial^2\mathcal{L}}{\partial\rho_2\partial\gamma_1} = -\frac{1}{N}\mathrm{tr}[S^{-1}(\delta)A_3S^{-1}(\delta)A_2],$$

$$\frac{\partial^2\mathcal{L}}{\partial\rho_2\partial\gamma_2} = -\frac{1}{N}\Big(\mathrm{tr}[S^{-1}(\delta)A_4S^{-1}(\delta)A_2] + Y'A_4'A_2Y/\sigma_2^2\Big),$$

$$\frac{\partial^2\mathcal{L}}{\partial\rho_2\partial\beta_1} = 0,$$

$$\frac{\partial^2\mathcal{L}}{\partial\rho_2\partial\beta_2} = -\frac{1}{N\sigma_2^2}X_2'W_2Y_2,$$

$$\frac{\partial^2\mathcal{L}}{\partial\rho_2\partial\sigma_1^2} = 0,$$

$$\frac{\partial^2\mathcal{L}}{\partial\rho_2\partial\sigma_2^2} = -\frac{1}{N\sigma_2^4}(Y_2 - \rho_2W_2Y_2 - \gamma_2Y_1 - X_2\beta_2)'W_2Y_2,$$

$$\frac{\partial^2\mathcal{L}}{\partial\gamma_1\partial\gamma_1} = -\frac{1}{N}\Big(\mathrm{tr}[S^{-1}(\delta)A_3S^{-1}(\delta)A_3] + Y'A_3'A_3Y/\sigma_1^2\Big),$$

$$\frac{\partial^2\mathcal{L}}{\partial\gamma_1\partial\gamma_2} = -\frac{1}{N}\mathrm{tr}[S^{-1}(\delta)A_4S^{-1}(\delta)A_3],$$

$$\frac{\partial^2\mathcal{L}}{\partial\gamma_1\partial\beta_1} = -\frac{1}{N\sigma_1^2}X_1'Y_2,$$

$$\frac{\partial^2\mathcal{L}}{\partial\gamma_1\partial\beta_2} = 0,$$

$$\frac{\partial^2\mathcal{L}}{\partial\gamma_1\partial\sigma_1^2} = -\frac{1}{N\sigma_1^4}(Y_1 - \rho_1W_1Y_1 - \gamma_1Y_2 - X_1\beta_1)'Y_2,$$

$$\frac{\partial^2\mathcal{L}}{\partial\gamma_1\partial\sigma_2^2} = 0,$$

$$\frac{\partial^2\mathcal{L}}{\partial\gamma_2\partial\gamma_2} = -\frac{1}{N}\Big(\mathrm{tr}[S^{-1}(\delta)A_4S^{-1}(\delta)A_4] + Y'A_4'A_4Y/\sigma_2^2\Big),$$

$$\frac{\partial^2\mathcal{L}}{\partial\gamma_2\partial\beta_1} = 0,$$

$$\frac{\partial^2\mathcal{L}}{\partial\gamma_2\partial\beta_2} = -\frac{1}{N\sigma_2^2}X_2'Y_1,$$

$$\frac{\partial^2\mathcal{L}}{\partial\gamma_2\partial\sigma_1^2} = 0,$$

$$\frac{\partial^2\mathcal{L}}{\partial\gamma_1\partial\sigma_2^2} = -\frac{1}{N\sigma_2^4}(Y_2 - \rho_2W_2Y_2 - \gamma_2Y_1 - X_2\beta_2)'Y_1,$$

$$\frac{\partial^2\mathcal{L}}{\partial\beta_1\partial\beta_1'} = -\frac{1}{N\sigma_1^2}X_1'X_1,$$

$$\frac{\partial^2\mathcal{L}}{\partial\beta_1\partial\beta_2'} = 0,$$

$$\frac{\partial^2\mathcal{L}}{\partial\beta_1\partial\sigma_1^2} = -\frac{1}{N\sigma_1^4}X_1'(Y_1 - \rho_1W_1Y_1 - \gamma_1Y_2 - X_1\beta_1),$$

$$\frac{\partial^2\mathcal{L}}{\partial\beta_1\partial\sigma_2^2} = 0,$$

$$\frac{\partial^2 \mathcal{L}}{\partial \beta_2 \partial \beta_2'} = -\frac{1}{N\sigma_2^2} X_2' X_2,$$
$$\frac{\partial^2 \mathcal{L}}{\partial \beta_2 \partial \sigma_1^2} = 0,$$
$$\frac{\partial^2 \mathcal{L}}{\partial \beta_2 \partial \sigma_2^2} = -\frac{1}{N\sigma_2^4} X_2'(Y_2 - \rho_2 W_2 Y_2 - \gamma_2 Y_1 - X_2\beta_2),$$
$$\frac{\partial^2 \mathcal{L}}{\partial \sigma_1^2 \partial \sigma_1^2} = \frac{1}{2\sigma_1^4} - \frac{1}{N\sigma_1^6}(Y_1 - \rho_1 W_1 Y_1 - \gamma_1 Y_2 - X_1\beta_1)'(Y_1 - \rho_1 W_1 Y_1 - \gamma_1 Y_2 - X_1\beta_1),$$
$$\frac{\partial^2 \mathcal{L}}{\partial \sigma_1^2 \partial \sigma_2^2} = 0,$$
$$\frac{\partial^2 \mathcal{L}}{\partial \sigma_2^2 \partial \sigma_2^2} = \frac{1}{2\sigma_2^4} - \frac{1}{N\sigma_2^6}(Y_2 - \rho_2 W_2 Y_2 - \gamma_2 Y_1 - X_2\beta_2)'(Y_2 - \rho_2 W_2 Y_2 - \gamma_2 Y_1 - X_2\beta_2).$$

where

$$A_1 = \begin{bmatrix} W_1 & 0 \\ 0 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 0 \\ 0 & W_2 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 0 & I_N \\ 0 & 0 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 0 & 0 \\ I_N & 0 \end{bmatrix}.$$

Now we show that

$$-\frac{\partial^2 \mathcal{L}(\tilde{\theta})}{\partial \theta \partial \theta'} \xrightarrow{p} \Omega^\star. \tag{S.11}$$

where $\tilde{\theta}$ is some point between $\hat{\theta}$ and θ^* and $\Omega^\star$ is defined in Theorem 1. We note that there are 36 different second-order derivatives in $\frac{\partial^2 \mathcal{L}(\theta)}{\partial \theta \partial \theta'}$, as listed above. These 36 derivatives can be classified into five categories according to their expressions. The first category includes the 6th, 8th, 12th, 14th, 19th, 21th, 23th, 25th, 28th, 30th, 32th and 35th derivatives, which are equal to zero exactly. The second category includes the first four derivatives, 9th-11th, 16th, 17th and 22th derivatives. The third category includes the 5th, 13th, 18th, 24th, 27th and 31th derivatives. The fourth category includes 7th, 15th, 20th, 26th, 29th and 33th derivatives. The last category includes the 34th and 36th derivatives. Since the derivations in each category are similar, we only choose one from each category to illustrate. We start our illustration with the second category. Consider the first derivative,

$$\begin{aligned}
-\frac{\partial^2 \mathcal{L}(\tilde{\theta})}{\partial \rho_1 \partial \rho_1} &= \frac{1}{N}\Big(\mathrm{tr}[S^{-1}(\tilde{\delta})A_1 S^{-1}(\tilde{\delta})A_1] + Y'A_1'A_1Y/\tilde{\sigma}_1^2\Big) \\
&= \frac{1}{N}\Big(\mathrm{tr}[S^{*-1}A_1 S^{*-1}A_1] + Y'A_1'A_1Y/\sigma_1^{*2}\Big) + \frac{2}{N}\mathrm{tr}[(S^{-1}(\delta^\dagger)A_1)^3](\tilde{\rho}_1 - \rho_1^*) \\
&+ \frac{2}{N}\mathrm{tr}[S^{-1}(\delta^\dagger)A_2(S^{-1}(\delta^\dagger)A_1)^2](\tilde{\rho}_2 - \rho_2^*) + \frac{2}{N}\mathrm{tr}[S^{-1}(\delta^\dagger)A_3(S^{-1}(\delta^\dagger)A_1)^2](\tilde{\gamma}_1 - \gamma_1^*) \\
&+ \frac{2}{N}\mathrm{tr}[S^{-1}(\delta^\dagger)A_4(S^{-1}(\delta^\dagger)A_1)^2](\tilde{\gamma}_2 - \gamma_2^*) + \Big[\frac{1}{\tilde{\sigma}_1^2} - \frac{1}{\sigma_i^{*2}}\Big]\frac{1}{N}Y'A_1'A_1Y
\end{aligned}$$

where $\delta^\dagger$ is some point between $\tilde{\delta}$ and δ^*. Since $S^{-1}(\delta^\dagger)$, A_1, A_2, A_3 and A_4 are all bounded in absolute value in column and row sums by Assumption C, we have $(S^{-1}(\delta^\dagger)A_1)^3$, $S^{-1}(\delta^\dagger)A_2(S^{-1}(\delta^\dagger)A_1)^2$, $S^{-1}(\delta^\dagger)A_3(S^{-1}(\delta^\dagger)A_1)^2$ and $S^{-1}(\delta^\dagger)A_4(S^{-1}(\delta^\dagger)A_1)^2$ are all bounded in absolute value in column and row sums, implying that the 2th-5th terms are all $o_p(1)$ by the consistency of $\hat{\delta}$. The last term is also $o_p(1)$ by $\frac{1}{N}Y'A_1'A_1Y = O_p(1)$ and $\hat{\sigma}_1^2 =$

$\sigma_1^{*2} + o_p(1)$. Next consider the first term. Notice that

$$\frac{1}{N\sigma_1^{*2}} Y' A_1' A_1 Y = \frac{1}{N\sigma_1^{*2}} Y_1' W_1' W_1 Y_1.$$

By (S.8), we have

$$\frac{1}{N\sigma_1^{*2}} Y_1' W_1' W_1 Y_1 = \frac{1}{N\sigma_1^{*2}} \Big[(V_{11}^* X_1 \beta_1^* + V_{12}^* X_2 \beta_2^*)' W_1' W_1 (V_{11}^* X_1 \beta_1^* + V_{12}^* X_2 \beta_2^*)$$
$$+ \sigma_1^2 \mathrm{tr}(V_{11}^{*\prime} W_1' W_1 V_{11}^*) + \sigma_2^2 \mathrm{tr}(V_{12}^{*\prime} W_1' W_1 V_{12}^*) \Big] + \frac{1}{N\sigma_1^{*2}} \mathrm{tr}\Big[V_{11}^{*\prime} W_1' W_1 V_{11}^* (e_1 e_1' - \sigma_1^{*2} I_N) \Big]$$
$$+ \frac{1}{N\sigma_1^{*2}} \mathrm{tr}\Big[V_{12}^{*\prime} W_1' W_1 V_{12}^* (e_2 e_2' - \sigma_2^{*2} I_N) \Big] + \frac{2}{N\sigma_1^{*2}} e_1' V_{11}^{*\prime} W_1' W_1 V_{12}^* e_2$$
$$+ \frac{2}{N\sigma_1^{*2}} (V_{11}^* X_1 \beta_1^* + V_{12}^* X_2 \beta_2^*)' W_1' W_1 (V_{11}^* e_1 + V_{12}^* e_2).$$

It can be shown the last four terms are all $o_p(1)$. Given this result, we have

$$\frac{1}{N\sigma_1^{*2}} Y_1' W_1' W_1 Y_1 = \frac{1}{N\sigma_1^{*2}} \Big[(V_{11}^* X_1 \beta_1^* + V_{12}^* X_2 \beta_2^*)' W_1' W_1 (V_{11}^* X_1 \beta_1^* + V_{12}^* X_2 \beta_2^*)$$
$$+ \sigma_1^{*2} \mathrm{tr}(V_{11}^{*\prime} W_1' W_1 V_{11}^*) + \sigma_2^{*2} \mathrm{tr}(V_{12}^{*\prime} W_1' W_1 V_{12}^*) \Big] + o_p(1).$$

The above results, together with $\frac{1}{N}\mathrm{tr}[S^{*-1} A_1 S^{*-1} A_1] = \frac{1}{N}\mathrm{tr}[W_1 V_{11}^* W_1 V_{11}^*]$, gives

$$-\frac{\partial^2 \mathcal{L}(\tilde{\theta})}{\partial \rho_1 \partial \rho_1} = \Omega_{11} + o_p(1).$$

Consider the third category. We choose the 5th derivative as the representative one to prove. The 5th derivative is

$$-\frac{\partial^2 \mathcal{L}(\tilde{\theta})}{\partial \rho_1 \partial \beta_1} = \frac{1}{N\tilde{\sigma}_1^2} X_1' W_1 Y_1.$$

The right hand side can be written as

$$\Big[\frac{1}{\tilde{\sigma}_1^2} - \frac{1}{\sigma_1^{*2}}\Big] \frac{1}{N} X_1' W_1 Y_1 + \frac{1}{N\sigma_1^{*2}} X_1' W_1 Y_1.$$

The first term is $o_p(1)$ by $N^{-1} X_1' W_1 Y_1 = O_p(1)$ and the consistency of $\hat{\sigma}_1^2$. By (S.8), it can be shown that

$$\frac{1}{N\sigma_1^{*2}} X_1' W_1 Y_1 = \frac{1}{N\sigma_1^{*2}} X_1' W_1 (V_{11}^* X_1 \beta_1^* + V_{12}^* X_2 \beta_2^*) + o_p(1).$$

Given this result, we have

$$-\frac{\partial^2 \mathcal{L}(\tilde{\theta})}{\partial \rho_1 \partial \beta_1} = \Omega_{15} + o_p(1).$$

Consider the fourth category. We choose the 7th derivative as the representative one to prove. The 7th derivation is

$$-\frac{\partial^2 \mathcal{L}(\tilde{\theta})}{\partial \rho_1 \partial \sigma_1^2} = \frac{1}{N\tilde{\sigma}_1^2} (Y_1 - \tilde{\rho}_1 W_1 Y_1 - \tilde{\gamma}_1 Y_2 - X_1 \tilde{\beta}_1)' W_1 Y_1.$$

The right hand side can be written as

$$-\frac{1}{N\tilde{\sigma}_1^2}(\tilde{\rho}_1-\rho_1^*)Y_1'W_1'W_1Y_1-\frac{1}{N\tilde{\sigma}_1^2}(\tilde{\gamma}_1-\gamma_1^*)Y_2'W_1Y_1-\frac{1}{N\tilde{\sigma}_1^2}(\tilde{\beta}_1-\beta_1^*)'X_1'W_1Y_1$$
$$+\Big[\frac{1}{\tilde{\sigma}_1^2}-\frac{1}{\sigma_1^{*2}}\Big]\frac{1}{N}e_1'W_1Y_1+\frac{1}{N\sigma_1^{*2}}e_1'W_1Y_1.$$

Notice that $N^{-1}Y_1'W_1'W_1Y_1, N^{-1}Y_2'W_1Y_1, N^{-1}X_1'W_1Y_1$ and $N^{-1}e_1'W_1Y_1$ are all $O_p(1)$. Given this result, together with consistency of $\tilde{\sigma}_1^2, \tilde{\rho}_1, \tilde{\gamma}_1$ and $\tilde{\beta}_1$, we have that the first four terms are all $o_p(1)$. By (S.8), it can be shown that

$$\frac{1}{N\sigma_1^{*2}}e_1'W_1Y_1=\frac{1}{N}\mathrm{tr}(W_1V_{11}^*)+o_p(1).$$

Given the above results, we prove

$$-\frac{\partial^2\mathcal{L}(\tilde{\theta})}{\partial\rho_1\partial\sigma_1^2}=\Omega_{17}+o_p(1).$$

Consider the last category. The last category has two derivatives. The proofs of the consistencies of these two derivatives are very similar as the preceding one and we hence omit it. Now we have proved (S.11). Given (S.10) and (S.11), by the Slutsky's theorem, we obtain the same result as stated in Theorem 1. This completes the proof of Theorem 1.