A Blockwise Network Autoregressive Model with Application for Fraud Detection

Supplementary Material

Bofei Xiao[a], Bo Lei[a], Wei Lan[a] and Bin Guo[a,*]

[a] *School of Statistics and Center of Statistical Research, Southwestern University of Finance and Economics, 555 LiuTai Road, Wenjiang, Chengdu, Sichuan, China, 611130.*

[*] *Corresponding Author. E-mail address: guobin@swufe.edu.cn.*

This supplementary material contains four components. In Section S.1, some notations, basic properties and three technical lemmas are listed. Section S.2 gives eight useful conditions. Theorem 1 and 2 are proved in Sections S.3 and S.4 respectively.

S.1. Three Technical Lemmas

NOTATIONS: The following list summarizes some frequently used notations in the text:

$S(\lambda) = I_n - \Lambda W$;

$\ln L(\theta)$ is the log likelihood of $\theta = (\beta^\top, \lambda^\top, \sigma^2)^\top$;

$\ln L(\lambda)$ is the concentrated log likelihood of λ;

$\mathrm{M} = I_n - X(X^\top X)^{-1}X^\top$.

SOME BASIC PROPERTIES: The following statements summarize some basic properties on network weights matrices and some laws of large numbers and central limit theorems for linear and quadratic forms of Kelejian and Prucha (2001). Let $\mathcal{E} = (\varepsilon_1, \cdots, \varepsilon_n)^\top$, where $\varepsilon_1, \cdots, \varepsilon_n$ are independent and identically distributed random variables with mean 0 and finite variance σ^2.

(1) Let A $= (a_{ij})_{n\times n}$ be an n-dimensional square matrix. Then, $\mathrm{E}(\mathcal{E}^\top A\mathcal{E}) = \sigma^2 tr(A)$ and $\mathrm{var}(\mathcal{E}^\top A\mathcal{E}) = (\mu^4 - 3\sigma^4)\sum\limits_{i=1}^{n} a_{ii}^2 + \sigma^4\left[tr(AA^\top) + tr(A^2)\right]$;

(2) Suppose the elements a_{ij} of the n-dimensional square matrices A are $O(1/h_n)$ uniformly for all i, j. If $n\times n$ matrices B are uniformly bounded in row and column sums, then the elements of AB have the uniform order $O(1/h_n)$. For these cases, $tr(AB) = tr(BA) = O(1/h_n)$. Then $\mathrm{E}(\mathcal{E}^\top A\mathcal{E}) = O(n/h_n)$ and $\mathrm{var}(\mathcal{E}^\top A\mathcal{E}) = O(n/h_n)$. If $\lim\limits_{\mathrm{n}\to\infty} h_n/n = 0$, $(\frac{h_n}{n})\left[\mathcal{E}^\top A\mathcal{E} - \mathrm{E}(\mathcal{E}^\top A\mathcal{E})\right] = o_p(1)$.

Based on the above two results, we can have the following lemmas.

Lemma 1. *Define* $Q_n = \mathcal{E}^\top A\mathcal{E} + b^\top\mathcal{E} - \sigma^2 tr(A)$, *where* $A = (a_{ij})_{n\times n} \in \mathbb{R}^{n\times n}$ *and* $b = (b_1, \cdots, b_n)^\top \in \mathbb{R}^{n\times 1}$. *Suppose the following assumptions are satisfied:*

(1) for $i, j = 1, \cdots, \mathrm{n}$, $a_{ij} = a_{ji}$;

(2) $sup_{n\geq 1}\|A\|_1 < \infty$;

(3) for some $\eta_1 > 0, sup_{n\geq 1} n^{-1}\|b\|_{2+\eta_1}^{2+\eta_1} < \infty$;

(4) for some $\eta_2 > 0, sup_{n\geq 1} E|\varepsilon_i|^{4+\eta_2} < \infty$.

Then, we have $E(Q_n) = 0$ *and*

$$\sigma_{Q_n}^2 := var(Q_n) = 4\sigma^4\sum_{i=1}^{n}\sum_{j=1}^{i-1} a_{ij}^2 + \sigma^2\sum_{i=1}^{n} b_i^2 + \sum_{i=1}^{n}\left[\left(\mu^4 - \sigma^4\right)a_{ii}^2 + 2\mu^3 b_i a_{ii}\right],$$

where $\mu^j = E(\varepsilon_i^j)$ *for j=3,4. Furthermore, suppose*

(5) $n^{-1}\sigma^2_{Q_n} > c$ *for some* $c > 0$.

Then, we obtain

$$Q_n/\sigma_{Q_n} \xrightarrow{d} N(0,1).$$

This result is directly from Theorem 1 of Kelejian and Prucha (2001).

Lemma 2. *Under Conditions (C1)-(C8) in Section S.2 below, we have that, as* $n \to \infty$,

(i) $\mathrm{I}_n(\theta_0) = -H_n(\theta_0) \equiv -\frac{1}{n}\frac{\partial^2 \ln L(\theta_0)}{\partial\theta\partial\theta^\top} \xrightarrow{p} -\mathrm{H}(\theta_0) = I(\theta_0)$,

(ii) $\frac{1}{\sqrt{n}}\frac{\partial \ln L(\theta_0)}{\partial\theta} \xrightarrow{d} N(0, I(\theta_0) + \Omega(\theta_0, \mu^3, \mu^4))$.

This result is similar to Lemma 3 in Appendix of Zou et al. (2021) and can be obtained via the results of Lemma 1.

Lemma 3. *Under Conditions (C1)-(C8) in Section S.2 below, we have that, as* $n \to \infty$, $\hat{\theta} - \theta_0 \xrightarrow{p} 0$.

This result is directly from Theorem 3.2 of Lee (2004).

S.2. Eight Useful Conditions

(C1) The $\{\varepsilon_i\}, i = 1, \cdots, n$, in $\mathcal{E} = (\varepsilon_1, \cdots, \varepsilon_n)^\top$ are *i.i.d.* with mean 0 and variance σ^2. Its moment $E\left(|\varepsilon|^{4+\gamma}\right)$ for some $\gamma > 0$ exists.

(C2) The elements ω_{ij} of W are at most of order h_n^{-1}, denoted by $O(1/h_n)$, uniformly in all i, j, where the rate sequence $\{h_n\}$ can be bounded or divergent. As a normalization, ω_{ij}=0 for all i.

(C3) The ratio $\frac{h_n}{n} \to 0$ *as* n goes to infinity.

(C4) The matrix $S(\lambda)$ is nonsingular.

(C5) The sequences of matrices $\{W_n\}$ are uniformly bounded in both row and column sums.

(C6) The elements of X are uniformly bounded constants for all n. The $\lim_{n\to\infty} X^\top X/n$ exists and is nonsingular.

(C7) $S^{-1}(\lambda)$ are uniformly bounded in either row or column sums, uniformly in λ in a compact parameter space $\mathcal{B}$. The true λ_0 is in the interior of $\mathcal{B}$.

(C8) Assume $H(\theta_0)$ is nonsingular and continuous in the interiors of $\mathcal{B}$, $I(\theta_0) = -\mathrm{H}(\theta_0) = -\lim_{n\to\infty} H_n(\theta_0)$, and $\Omega(\theta_0, \mu^3, \mu^4) = \lim_{n\to\infty} \Omega_n(\theta_0, \mu^3, \mu^4)$.

The boundedness of the moment of ε is assumed in Condition (C1), which is looser than commonly used normal distribution assumption. Conditions (C2) and (C3) are directly derived from in Lee (2004). Like the condition in Lee and Liu (2010), Condition (C4) ensures that $S(\lambda)$ is invertible. Conditions (C5)-(C8) are traditional conditions for establishing the convergence of the Fisher information matrix and the variance of the score function and carefully studied in Lee (2004) and Zou et al. (2021). For instance, Conditions (C6) being a common assumption in linear regression analysis guarantees that $n^{-1}\sigma^{-2}X^\top X$ convergences a positive matrix.

S.3. Proof of Theorem 1

By the mean value theorem, we have that

$$\frac{\partial \ln L\left(\hat{\theta}\right)}{\partial \theta} = 0 = \frac{\partial \ln L\left(\theta_0\right)}{\partial \theta} + \frac{\partial^2 \ln L\left(\bar{\theta}\right)}{\partial \theta \partial \theta^{\top}}\left(\hat{\theta} - \theta_0\right),$$

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) = -\left(\frac{1}{n}\frac{\partial^2 \ln L\left(\bar{\theta}\right)}{\partial \theta \partial \theta^{\top}}\right)^{-1} \frac{1}{\sqrt{n}}\frac{\partial \ln L\left(\theta_0\right)}{\partial \theta}, \tag{1}$$

where, $\bar{\theta} = a\hat{\theta} + (1-a)\,\theta_0$ for $a \in [0,1]$, this, together with Lemma 3 $\hat{\theta} - \theta_0 \xrightarrow{p} 0$, we have

$$\left\|\bar{\theta} - \theta_0\right\| = \left\|a(\hat{\theta} - \theta_0)\right\| \leq \left\|\hat{\theta} - \theta_0\right\| \xrightarrow{p} 0, \tag{2}$$

so $\bar{\theta}$ is a consistent estimator of θ_0.

For $H\left(\theta\right)$ is continuous in the interiors of $\mathcal{B}$, together with inequality and Lemma 2, we have

$$\begin{aligned} H_n\left(\bar{\theta}\right) - H(\theta_0) &= H_n\left(\bar{\theta}\right) - H\left(\bar{\theta}\right) + H\left(\bar{\theta}\right) - H(\theta_0) \\ &\leq \sup H_n\left(\bar{\theta}\right) - H\left(\bar{\theta}\right) + H\left(\bar{\theta}\right) - H\left(\theta_0\right) \xrightarrow{p} 0. \end{aligned} \tag{3}$$

By the Slutsky's theorem, in conjunction with Lemma 2, implies

$$\begin{aligned} \sqrt{n}\left(\hat{\theta} - \theta_0\right) &= -\left(H_n\left(\bar{\theta}\right)\right)^{-1} \frac{1}{\sqrt{n}}\frac{\partial \ln L\left(\theta_0\right)}{\partial \theta} \\ &\xrightarrow{d} N\left(0, I^{-1}\left(\theta_0\right) + I^{-1}\left(\theta_0\right)\Omega(\theta_0, \mu^3, \mu^4) I^{-1}\left(\theta_0\right)\right), \end{aligned} \tag{4}$$

where $\mathrm{var}\left(-\left(H_n\left(\bar{\theta}\right)\right)^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L(\theta_0)}{\partial \theta}\right) = I^{-1}\left(\theta_0\right)\left[I\left(\theta_0\right) + \Omega(\theta_0, \mu^3, \mu^4)\right] I^{-1}\left(\theta_0\right) = I^{-1}\left(\theta_0\right)$ $+ I^{-1}\left(\theta_0\right)\Omega(\theta_0, \mu^3, \mu^4) I^{-1}\left(\theta_0\right)$.

S.4. Proof of Theorem 2

To facilitate this proof, we slightly arrange the notation $\theta = \left(\beta^\top, \lambda^\top, \sigma^2\right)^\top$ to be $\theta = \left(\sigma^2, \beta^\top, \lambda^\top\right)^\top = \left(\theta_1^\top, \theta_2^\top\right)^\top$, where $\theta_1 = \left(\sigma^2, \beta^\top\right)^\top$. $\theta_2 = \lambda = (\lambda_1, \cdots, \lambda_k)^\top$.

The null hypothesis is equivalent to $\mathrm{H}_0 : R(\theta) = R\theta = r_c$, where $R = \left(0_{(k-1)\times(p+1)}, R_1\right)$,

$$R_1 = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -1 & & 0 & 0 & 0 \\ & \vdots & & \ddots & & \vdots & \\ 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{pmatrix}_{(k-1)\times k},$$

$\mathrm{tr}\,(R) = k-1$, $r_c = 0_{(k-1)\times 1} \in \mathbb{R}^{(k-1)\times 1}$.

We employ the similar techniques to those used for proving (4).

Step I: We get the linear-quadratic forms of $\mathcal{LR} = -2\left[\ln L(\tilde{\theta}) - \ln L(\hat{\theta})\right]$ with $(\tilde{\theta} - \hat{\theta})$. The first-order derivatives condition For unconstrained QMLE $\hat{\theta}$ is

$$\frac{\partial \ln L(\hat{\theta})}{\partial \theta} = 0,$$

and the constrained QMLE $\tilde{\theta} = \arg\max_{\theta\in\mathcal{B}} \left\{\ln L(\theta) + nl^\top \left[r_c - R(\theta)\right]\right\}$, we have

$$\begin{aligned} \frac{\partial \ln L(\tilde{\theta})}{\partial \theta} - nR'(\tilde{\theta})^\top l = 0, \\ R(\tilde{\theta}) - r_c = 0, \end{aligned} \tag{5}$$

where l is the Lagrange multiplier. The second-order Taylor expansion of $\mathcal{LR}$ at $\hat{\theta}$ is

$$\begin{aligned}
\mathcal{LR} &= -2\left[\ln L(\tilde{\theta}) - \ln L(\hat{\theta})\right] \\
&= -2(\ln L(\hat{\theta}) - \ln L(\hat{\theta})) - 2\frac{\partial \ln L(\hat{\theta})}{\partial \theta}(\tilde{\theta} - \hat{\theta}) - (\tilde{\theta} - \hat{\theta})^\top \frac{\partial^2 \ln L(\bar{\theta}_a)}{\partial \theta \partial \theta^\top}(\tilde{\theta} - \hat{\theta}) \\
&= (\tilde{\theta} - \hat{\theta})^\top (-\frac{\partial^2 \ln L(\bar{\theta}_a)}{\partial \theta \partial \theta^\top})(\tilde{\theta} - \hat{\theta}) \\
&= \sqrt{n}(\tilde{\theta} - \hat{\theta})^\top (-H_n(\bar{\theta}_a))\sqrt{n}(\tilde{\theta} - \hat{\theta}),
\end{aligned} \tag{6}$$

where $\bar{\theta}_a$ lies between $\tilde{\theta}$ and $\hat{\theta}$.

Step II: We explore the relation between $\sqrt{n}(\tilde{\theta} - \hat{\theta})$ and $\sqrt{n}l$. By (5) and first-order Taylor expansion of $\frac{\partial \ln L(\tilde{\theta})}{\partial \theta}$ at $\hat{\theta}$, we have

$$\begin{aligned}
\frac{\partial \ln L(\hat{\theta})}{\partial \theta} + \frac{\partial^2 \ln L(\bar{\theta}_b)}{\partial \theta \partial \theta^\top}(\tilde{\theta} - \hat{\theta}) - nR'(\tilde{\theta})^\top l = 0, \\
H_n(\bar{\theta}_b)(\tilde{\theta} - \hat{\theta}) - R'(\tilde{\theta})^\top l = 0,
\end{aligned}$$

where $\bar{\theta}_b$ lies between $\tilde{\theta}$ and $\hat{\theta}$. As n goes to infinity, we have

$$\sqrt{n}(\tilde{\theta} - \hat{\theta}) = \ H_n^{-1}(\bar{\theta}_b)R'(\tilde{\theta})^\top \sqrt{n}l. \tag{7}$$

Step III: We get the asymptotic distribution of $\sqrt{n}l$. By (5) and first-order Taylor expansion of $\frac{\partial \ln L(\tilde{\theta})}{\partial \theta}$ at θ_0, we have

$$R'(\tilde{\theta})^\top l = \frac{1}{n}\frac{\partial \ln L(\tilde{\theta})}{\partial \theta} = \frac{1}{n}\frac{\partial \ln L(\theta_0)}{\partial \theta} + \frac{1}{n}\frac{\partial^2 \ln L(\bar{\theta}_c)}{\partial \theta \partial \theta^\top}(\tilde{\theta} - \theta_0),$$

where $\bar{\theta}_c$ lies between $\tilde{\theta}$ and θ_0. As n goes to infinity, we have

$$H_n^{-1}(\bar{\theta}_c)R'(\tilde{\theta})^\top\sqrt{n}l = H_n^{-1}\left(\bar{\theta}_c\right)\frac{1}{\sqrt{n}}\frac{\partial \ln L\left(\theta_0\right)}{\partial\theta} + \sqrt{n}\left(\tilde{\theta}-\theta_0\right) \tag{8}$$

Applying the Taylor expansion of $R(\tilde{\theta})-r_c = 0$ at θ_0, together with $\mathrm{H}_0 : R(\theta_0)-r_c = 0$, we have

$$\begin{aligned} (R(\theta_0)-r_c) + R'(\bar{\theta}_d)(\tilde{\theta}-\theta_0) &= 0, \\ R'(\bar{\theta}_d)\sqrt{n}(\tilde{\theta}-\theta_0) &= 0, \end{aligned} \tag{9}$$

where $\bar{\theta}_d$ lies between $\tilde{\theta}$ and θ_0, By (8) and (9), we obtain

$$\begin{aligned} R'(\bar{\theta}_d)H_n^{-1}(\bar{\theta}_c)R'(\tilde{\theta})^\top\sqrt{n}l &= R'(\bar{\theta}_d)H_n^{-1}(\bar{\theta}_c)\frac{1}{\sqrt{n}}\frac{\partial \ln L(\theta_0)}{\partial\theta} + R'(\bar{\theta}_d)\sqrt{n}(\tilde{\theta}-\theta_0) \\ &= R'(\bar{\theta}_d)H_n^{-1}(\bar{\theta}_c)\frac{1}{\sqrt{n}}\frac{\partial \ln L(\theta_0)}{\partial\theta}. \end{aligned}$$

Then employing similar techniques to those used for proving (3), we obtain that $\bar{\theta}_a, \bar{\theta}_b, \bar{\theta}_c, \bar{\theta}_d$ are consistent estimator of θ_0, and $H_n(\bar{\theta}_a), H_n(\bar{\theta}_b), H_n(\bar{\theta}_c)$ are consistent estimator of $\mathrm{H}(\theta)$.

For the sake of simplicity, denote $\Sigma = I(\theta_0) + \Omega(\theta_0, \mu^3, \mu^4)$. By Lemma 2 and the Slustsky theorem, we have

$$\begin{aligned} \sqrt{n}l &= (R'(\bar{\theta}_d)H_n^{-1}(\bar{\theta}_c)R'(\tilde{\theta})^\top)^{-1}R'(\bar{\theta}_d)H_n^{-1}(\bar{\theta}_c)\frac{1}{\sqrt{n}}\frac{\partial \ln L(\theta_0)}{\partial\theta} \\ &= (R'(\tilde{\theta})^\top)^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L(\theta_0)}{\partial\theta} \\ &\xrightarrow{d} N\left(0, \left[R'(\theta_0)\Sigma^{-1}R'(\theta_0)^\top\right]^{-1}\right), \end{aligned} \tag{10}$$

This, in conjunction with (6) and (7), leads to

$$
\begin{aligned}
-H_n^{1/2}(\bar{\theta}_a)\sqrt{n}(\tilde{\theta}-\hat{\theta}) &= -H_n^{1/2}(\bar{\theta}_a)H_n^{-1}(\bar{\theta}_b)R'(\tilde{\theta})^\top\sqrt{n}l \\
&\xrightarrow{d} N\left(0,\Pi^2\right),
\end{aligned} \tag{11}
$$

where $\Pi^2 = I(\theta_0)^{-1/2}R'(\theta_0)^\top\left[R'(\theta_0)\Sigma^{-1}R'(\theta_0)^\top\right]^{-1}R'(\theta_0)I(\theta_0)^{-1/2}$. Under the null hypothesis H_0, $\mathrm{tr}(R)=k-1$, then we get $\mathrm{tr}(\Pi)=k-1$.

$$
\begin{aligned}
\mathcal{LR} &= \sqrt{n}(\tilde{\theta}-\hat{\theta})^\top(-H_n^{1/2}(\bar{\theta}_a))(-H_n^{1/2}(\bar{\theta}_b))\sqrt{n}(\tilde{\theta}-\hat{\theta}) \\
&= \left(\sqrt{n}(\tilde{\theta}-\hat{\theta})^\top(-H_n^{1/2}(\bar{\theta}_a))\Pi^{-1}\right)\Pi^2\left(\Pi^{-1}(-H_n^{1/2}(\bar{\theta}_b))\sqrt{n}(\tilde{\theta}-\hat{\theta})\right).
\end{aligned}
$$

Let $\lambda_1(\theta_0,\mu^3,\mu^4),\cdots,\lambda_{k-1}(\theta_0,\mu^3,\mu^4)$ be the eigenvalues of Π^2. The above results, together with the continuous mapping theorem and Slutskys theorem, imply that $\mathcal{LR}$ follows a weighted chi-square distribution $\sum_{r=1}^{k-1}\lambda_r(\theta_0,\mu^3,\mu^4)\mathcal{X}_r^2(1)$ asymptotically. This completes the first part of the proof.

Under the normal assumption of $\mathcal{E}$, the matrix $\Omega_n(\theta_0,\mu^3,\mu^4)$ defined above Theorem 1 is 0. By Condition (C8), we have $\Omega(\theta_0,\mu^3,\mu^4)=0$, which leads to $\Sigma = I(\theta_0)$. Then $\Pi^2 = I(\theta_0)^{-1/2}R'(\theta_0)^\top\left[R'(\theta_0)I(\theta_0)^{-1}R'(\theta_0)^\top\right]^{-1}R'(\theta_0)I(\theta_0)^{-1/2}$ is dempotent. The above results, together with the normality assumption, we obtain

$$
\begin{aligned}
\mathcal{LR} &= \sqrt{n}(\tilde{\theta}-\hat{\theta})^\top(-H_n^{1/2}(\bar{\theta}_a))(-H_n^{1/2}(\bar{\theta}_b))\sqrt{n}(\tilde{\theta}-\hat{\theta}) \\
&\xrightarrow{d} \mathcal{X}^2(k-1),
\end{aligned}
$$

which completes the entire proof.

REFERENCES

Kelejian, H. H. and Prucha, I. R. (2001) "On the asymptotic distribution of the moran i test statistic with applications", *Journal of Econometrics*, 104, 219–257.

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

Lee, L. F., and Liu, X. (2010) "Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances", *Econometric Theory*, 187–230.

Zou, T., Luo, R., Lan, W., and Tsai, C. L. (2021) "Network influence analysis", *Statistica Sinica*, 31, 1727–1748.