NETWORK IMPUTATION FOR A SPATIAL AUTOREGRESSION MODEL WITH INCOMPLETE DATA

Zhimeng Sun and Hansheng Wang*

Central University of Finance and Economics & Peking University, P. R. China

Supplementary Material

The Supplementary Material contains detailed proofs of Theorem 1, Theorem 2, and Theorem 3.

1. Two Lemmas

To establish the claimed theoretical results in Theorems 1, 2, and 3, we need the following two lemmas. We provide proofs of the two lemmas in this section.

Lemma 1. *For the n dimensional standard normal random vector $\mathbb{Z} = (Z_1, \ldots, Z_n) \in \mathbb{R}^n$, and $n \times n$ dimensional symmetric matrix $M = (m_{ij})$. If there exist some constants $C_1 > 0$ and $C_2 > 0$, such that $\lambda_{\max}(M) \leq C_1$, and $n^{-1} tr(M^2) > C_2$, as $n \to \infty$, where $\lambda_{\max}(M)$ denotes the biggest eigenvalue of M, we have $\Big\{\mathbb{Z}^\top M \mathbb{Z} - tr(M)\Big\}/tr^{\frac{1}{2}}(2M^2) \xrightarrow{d} N(0,1)$.*

Proof of Lemma 1. Denote the eigen values of M by $\lambda_1, \lambda_2, \cdots, \lambda_n$. We find by a spectral decomposition of M as $M = Q^\top \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) Q$ where the column vectors of Q are the eigenvectors of M, and they are orthonormal. Further, if $U = Q\mathbb{Z}$, we know that U is a standard normal random vector, and $R = \mathbb{Z}^\top M \mathbb{Z} = \sum_{i=1}^n \lambda_i U_i^2$ is the sum of these independent variables. From easy calculations, we have $\mathrm{E}(R) = \sum_{i=1}^n \lambda_i = \mathrm{tr}(M)$, and $\mathrm{Var}(R) = 2\sum_{i=1}^n \lambda_i^2 = 2\mathrm{tr}(M^2)$ that is bounded below by $2nC_2$. Meanwhile, we have $\sum_{i=1}^n \mathrm{E}(\lambda_i U_i^2 - \lambda_i)^4 = \mathrm{E}(U_1^2 - 1)^4 \mathrm{tr}(M^4) \leq \mathrm{E}(U_1^2 - 1)^4 \cdot nC_1^4$. This leads to $\sum_{i=1}^n \mathrm{E}(\lambda_i U_i^2 -$

$\lambda_i)^4/[\text{Var}(R)]^2 \le \text{E}(U_1^2-1)^4 \cdot C_1^4/(4nC_2^2) \to 0,$ *as* $n \to \infty$ This verifies the Lyapunov condition, and the Central Limit Theorem immediately leads to Lemma 1.

Lemma 2. (Asymptotic normality of the estimation equation) *Denote* $d\ell(\theta)/d\theta = \Big((d\ell(\theta)/d\beta)^\top, d\ell(\theta)/d\rho, d\ell(\theta)/d\sigma^2\Big)^\top$. *Assuming conditions (C1) and (C2), we have* $n^{-1/2}\cdot d\ell(\theta)/d\theta \xrightarrow{d} N(0,\Lambda)$, *where* Λ *is defined in the following proof of this lemma.*

Proof of Lemma 2. From easy calculations of the first derivative of $\ell(\theta)$, we obtain the estimation equation, with its components specified as

$$\begin{aligned}
\frac{d\ell(\theta)}{d\beta} &= \frac{1}{\sigma^2}\mathbb{X}_1^\top\Omega(\rho)(\mathbb{Y}_1-\mathbb{X}_1\beta),\\
\frac{d\ell(\theta)}{d\rho} &= \frac{1}{2}\text{tr}\Big[\Omega(\rho)^{-1}\dot{\Omega}(\rho)\Big] - \frac{1}{2\sigma^2}(\mathbb{Y}_1-\mathbb{X}_1\beta)^\top\dot{\Omega}(\rho)(\mathbb{Y}_1-\mathbb{X}_1\beta),\\
\frac{d\ell(\theta)}{d\sigma^2} &= \frac{1}{2\sigma^4}(\mathbb{Y}_1-\mathbb{X}_1\beta)^\top\Omega(\rho)(\mathbb{Y}_1-\mathbb{X}_1\beta) - \frac{n}{2\sigma^2}.
\end{aligned}$$

The components of the estimation equation are linear or quadratic functions of the multivariate standard normal vector $\Sigma_{11}^{-\frac{1}{2}}(\mathbb{Y}_1-\mathbb{X}_1\beta)$. Under conditions C1 and C2, we derive the asymptotic normal distribution of $n^{-1/2}d\ell(\theta)/d\theta$ from the central limit theorem for linear-quadratic functions. We can see that $\text{E}\Big[d\ell(\theta)/d\beta\Big]=0$. Denoting $\Sigma_{11}^{-\frac{1}{2}}(\mathbb{Y}_1-\mathbb{X}_1\beta)$ by $\mathbb{Z}=(Z_1,\cdots,Z_n)^\top$, we have $\text{E}\Big[d\ell(\theta)/d\rho\Big] = \text{tr}\Big[\Omega(\rho)^{-1}\dot{\Omega}(\rho)\Big]/2 - \text{E}\Big[\mathbb{Z}^\top\Sigma_{11}^{\frac{1}{2}}\dot{\Omega}(\rho)\Sigma_{11}^{\frac{1}{2}}\mathbb{Z}\Big]/(2\sigma^2)=0$ and $\text{E}\Big[d\ell(\theta)/d\sigma^2\Big] = -n/(2\sigma^2)+\text{E}(\mathbb{Z}^\top\mathbb{Z})/(2\sigma^2)=0$. Thus, the estimation equation $n^{-1/2}d\ell(\theta)/d\theta$ is unbiased, and its covariance matrix is $\text{E}\Big[(d\ell(\theta)/d\theta)\cdot(d\ell(\theta)/d\theta^\top)\Big]/n$, which is equal to the $(p+2)\times(p+2)$ symmetric information matrix $\Lambda_n = -n^{-1}\text{E}\Big[\frac{d^2\ell(\theta)}{d\theta d\theta^\top}\Big]$.

To derive the asymptotic covariance matrix of $n^{-1/2}d\ell(\theta)/d\theta$, we first denote H_1 and H_2 as $H_1 = \Big[4\rho(W_{11}^\top W_{12}+W_{21}^\top W_{22})-2(W_{21}^\top+W_{12})\Big]\Upsilon_{22}^{-1}(\rho)$ and $H_2 = \Upsilon_{22}^{-1}(\rho)\Big[-(W_{22}^\top+W_{22})+2\rho(W_{12}^\top W_{12}+W_{22}^\top W_{22})\Big]\Upsilon_{22}^{-1}(\rho)$, respectively. Then, we consider $\Lambda_n = [\Lambda_{n11},\Lambda_{n12},\Lambda_{n13};\Lambda_{n21},\Lambda_{n22},\Lambda_{n23};\Lambda_{n31},\Lambda_{n32},\Lambda_{n33}]$, and calculate each of its components, respectively

$$\Lambda_{n11} = -n^{-1}\mathrm{E}\Big[\frac{d^2\ell(\theta)}{d\beta d\beta^\top}\Big] = (n\sigma^2)^{-1}\mathbb{X}_1^\top\Omega(\rho)\mathbb{X}_1 \;\rightarrow\; \Lambda_{11},$$

$$\Lambda_{n12} = -n^{-1}\mathrm{E}\Big[\frac{d^2\ell(\theta)}{d\beta d\rho}\Big] = (n\sigma^2)^{-1}E\Big[\mathbb{X}_1^\top\dot{\Omega}(\rho)(\mathbb{Y}_1 - \mathbb{X}_1\beta)\Big] \;=\; 0,$$

$$\Lambda_{n13} = -n^{-1}E\Big[\frac{d^2\ell(\theta)}{d\beta d\sigma^2}\Big] \;=\; (n\sigma^4)^{-1}\mathrm{E}\Big[\mathbb{X}_1^\top\Omega(\rho)(\mathbb{Y}_1 - \mathbb{X}_1\beta)\Big] \;=\; 0,$$

$$\Lambda_{n22} = -n^{-1}\mathrm{E}\Big[\frac{d^2\ell(\theta)}{d\rho^2}\Big] \;=\; (2n)^{-1}\mathrm{tr}\Big[(\Omega^{-1}(\rho)\dot{\Omega}(\rho))^2\Big] - (2n)^{-1}\mathrm{tr}\Big[\Omega^{-1}(\rho)\ddot{\Omega}(\rho)\Big]$$

$$+\,(2n\sigma^2)^{-1}\mathrm{E}\Big[(\mathbb{Y}_1 - \mathbb{X}_1\beta)^\top\ddot{\Omega}(\rho)(\mathbb{Y}_1 - \mathbb{X}_1\beta)\Big] \;=\; (2n)^{-1}\mathrm{tr}\Big[(\Omega^{-1}(\rho)\dot{\Omega}(\rho))^2\Big] \rightarrow\; \Lambda_{22},$$

$$\Lambda_{n23} = -n^{-1}\mathrm{E}\Big[\frac{d^2\ell(\theta)}{d\sigma^2 d\rho}\Big] = -(2n\sigma^4)^{-1}\mathrm{E}\Big[(\mathbb{Y}_1 - \mathbb{X}_1\beta)^\top\dot{\Omega}(\rho)(\mathbb{Y}_1 - \mathbb{X}_1\beta)\Big] \;\rightarrow\; \Lambda_{23},$$

$$\Lambda_{n33} = -n^{-1}\mathrm{E}\Big[\frac{d^2\ell(\theta)}{d\sigma^2 d\sigma^2}\Big] = -(2\sigma^4)^{-1} + (n\sigma^4)^{-1}\mathrm{E}\Big[(\mathbb{Y}_1 - \mathbb{X}_1\beta)^\top\Sigma_{11}^{-1}(\mathbb{Y}_1 - \mathbb{X}_1\beta)\Big] = (2\sigma^4)^{-1}$$

where $\ddot{\Omega}(\rho)$ is defined as

$$\begin{aligned}
\ddot{\Omega}(\rho) \;&=\; H_1\Big[2\rho(W_{12}^\top W_{12} + W_{22}^\top W_{22})\Upsilon_{22}^{-1}(\rho)\Upsilon_{21}(\rho) - (W_{22}^\top + W_{22})\Upsilon_{22}^{-1}(\rho)\Upsilon_{21}(\rho)\Big] \\
&+\; H_1\Big[(W_{12}^\top + W_{21}) - \rho(W_{12}^\top W_{11} + W_{22}^\top W_{21}) - \rho(W_{11}^\top W_{11} + W_{12}^\top W_{21})\Big] \\
&+\; 2(W_{11}^\top W_{11} + W_{21}^\top W_{21}) - 2(W_{11}^\top W_{12} + W_{21}^\top W_{22})\Upsilon_{22}^{-1}(\rho)\Upsilon_{21}(\rho) \\
&-\; \Big[\rho^2(W_{11}^\top W_{12} + W_{21}^\top W_{22}) - \rho(W_{21}^\top + W_{12})\Big]D(\rho).
\end{aligned}$$

The quantity $D(\rho)$ in $\ddot{\Omega}(\rho)$ is specified as

$$
\begin{aligned}
D(\rho) = & -2\rho\Upsilon_{22}^{-1}(\rho)(W_{12}^{\top}W_{12} + W_{22}^{\top}W_{21})\Upsilon_{22}^{-1}(\rho)\Big[2\rho(W_{12}^{\top}W_{11} + W_{22}^{\top}W_{21}) - (W_{12}^{\top} + W_{21})\Big] \\
& + 2\rho\Upsilon_{22}^{-1}(\rho)(W_{12}^{\top}W_{12} + W_{22}^{\top}W_{21})H_2\Upsilon_{21}(\rho) - \Upsilon_{22}^{-1}(\rho)(W_{22}^{\top} + W_{22})H_2\Upsilon_{21}(\rho) \\
& - \Upsilon_{22}^{-1}(\rho)(W_{22}^{\top} + W_{22})\Upsilon_{22}^{-1}(\rho)\Big[2\rho(W_{12}^{\top}W_{11} + W_{22}^{\top}W_{21}) - (W_{12}^{\top} + W_{21})\Big] \\
& + H_2\Big[2\rho(W_{12}^{\top}W_{12} + W_{22}^{\top}W_{22}) - (W_{22}^{\top} + W_{22})\Upsilon_{22}^{-1}(\rho)\Upsilon_{21}(\rho)\Big] \\
& + H_2\Big[(W_{12}^{\top} + W_{21}) - 2\rho(W_{11}^{\top}W_{11} + W_{12}^{\top}W_{21})\Big].
\end{aligned}
$$

Moreover, note that $\Lambda_{n31} = \Lambda_{n13}$ and $\Lambda_{n32} = \Lambda_{n23}$. This leads to the specified asymptotic variance matrix of $n^{-1/2}d\ell(\theta)/d\theta$ as

$$
\Lambda = \lim_{n\to\infty} \Lambda_n = \begin{pmatrix} \Lambda_{11} & 0 & 0 \\ 0 & \Lambda_{22} & \Lambda_{23} \\ 0 & \Lambda_{23} & \Lambda_{33} \end{pmatrix}.
$$

This completes the proof of Lemma 1.

2. Proof of Theorem 1

Our proof of Theorem 1 includes two steps. We show the $\sqrt{n}-$ consistency of $\hat{\theta}$ in step 1 and then prove the limiting normal distribution of the estimator in step 2.

Step 1. To show the $\sqrt{n}-$consistency of $\hat{\theta}$, we first show that there exists some constant $C > 0$, such that

$$
\sup_{\|t\|=C} \ell(\theta + n^{-1/2}t) \; < \; \ell(\theta), \tag{1}
$$

with a probability tending to one, as $n \to \infty$, and $t \in \mathbb{R}^{p+2}$. We obtain (1) by Taylor's expansion of $\ell(\theta+n^{-1/2}t)$, which leads to

$$R_n(\theta) \;=\; \ell(\theta + n^{-1/2}t) - \ell(\theta) = n^{-1/2}t^\top \frac{d\ell(\theta)}{d\theta} + (2n)^{-1}t^\top \frac{d^2\ell(\theta)}{d\theta d\theta^\top}t + o_p(1). \tag{2}$$

From Lemma 1, we know that $n^{-1/2}d\ell(\theta)/d\theta = O_p(1)$. On the other hand, by the law of large numbers, we obtain $n^{-1}\frac{d^2\ell(\theta)}{d\theta d\theta^\top} = -\Lambda_n + o_p(1) \to -\Lambda$, which is a negative definite matrix. Thus, the second term of (2), which is quadratic and negative, would dominate the first term, which is linear, for a sufficiently large C. We have (1) with a probability tending to one, as $n \to \infty$. From the convexity of $\ell(\theta)$, this leads to $\sup_{\|t\| \geq C} \ell(\theta + n^{-1/2}t) < \ell(\theta)$ with a probability tending to one, as $n \to \infty$. Note that $\ell(\theta)$ is maximized at $\hat{\theta}$, implying that $\hat{\theta}$ lies in the ball $\{\theta + n^{-1/2}t :\| u \|\leq C\}$; that is, $\| \hat{\theta} \|= O_p(n^{-1/2})$.

Step 2. We next prove Theorem 1 through a routine Taylor expansion of the estimation equation $d\ell(\hat{\theta})/d\theta = 0$ at the true value of θ, which easily leads to

$$\sqrt{n}(\hat{\theta} - \theta) \;=\; \Big[-\frac{1}{n}\frac{d^2\ell(\breve{\theta})}{d\theta d\theta^\top}\Big]^{-1} \frac{1}{\sqrt{n}}\frac{d\ell(\theta)}{d\theta},$$

with $\breve{\theta}$ lying between $\hat{\theta}$ and θ. To derive Theorem 1, we have only to show that

$$\frac{1}{n}\frac{d^2\ell(\breve{\theta})}{d\theta d\theta^\top} \;=\; \frac{1}{n}\frac{d^2\ell(\theta)}{d\theta d\theta^\top} + o_p(1). \tag{3}$$

Finally, we consider each block of the two related matrices, respectively. First, we show that

$$\frac{1}{n}\frac{d^2\ell(\breve{\theta})}{d\rho^2} \;=\; \frac{1}{n}\frac{d^2\ell(\theta)}{d\rho^2} + o_p(1). \tag{4}$$

Denote $d^3\Omega(\rho)/d\rho^3$ by $\dddot{\Omega}$. By the mean value theorem, we have

$$\ddot{\Omega}(\breve{\rho}) = \ddot{\Omega}(\rho) + \dddot{\Omega}(\bar{\rho})(\breve{\rho} - \rho), \tag{5}$$

$$\Omega^{-1}(\breve{\rho})\ddot{\Omega}(\breve{\rho}) = \Omega^{-1}(\rho)\ddot{\Omega}(\rho) + \Omega^{-1}(\bar{\rho})\Big[\mathrm{I} - \dot{\Omega}(\bar{\rho})\Omega^{-1}(\bar{\rho})\Big]\dddot{\Omega}(\bar{\rho})(\breve{\rho} - \rho), \tag{6}$$

$$\Big[\Omega^{-1}(\breve{\rho})\dot{\Omega}(\breve{\rho})\Big]^2 = \Big[\Omega^{-1}(\rho)\dot{\Omega}(\rho)\Big]^2 + 2\Bigg\{\Omega^{-1}(\bar{\rho})\Big[\mathrm{I} - \dot{\Omega}(\bar{\rho})\Omega^{-1}(\bar{\rho})\Big]\ddot{\Omega}(\bar{\rho})\Bigg\}(\breve{\rho} - \rho), \tag{7}$$

with $\bar{\rho}$ lying between $\breve{\rho}$ and ρ. Then, we obtain

$$\begin{aligned} L_{n1} &= n^{-1}\frac{d^2\ell(\breve{\theta})}{d\rho^2} - n^{-1}\frac{d^2\ell(\theta)}{d\rho^2} = (2n)^{-1}\Bigg[\mathrm{tr}\Big\{\Omega^{-1}(\breve{\rho})\dot{\Omega}(\breve{\rho})\Big\}^2 - \mathrm{tr}\Big\{\Omega^{-1}(\rho)\dot{\Omega}(\rho)\Big\}^2\Bigg] \\ &- (2n)^{-1}\Bigg[\mathrm{tr}\Big\{\Omega^{-1}(\breve{\rho})\ddot{\Omega}(\breve{\rho})\Big\} - \mathrm{tr}\{\Omega^{-1}(\rho)\ddot{\Omega}(\rho)\}\Bigg] \\ &+ (2n)^{-1}(\breve{\sigma}^{-2} - \sigma^{-2})(\mathbb{Y}_1 - \mathbb{X}_1\breve{\beta})^\top\ddot{\Omega}(\breve{\rho})(\mathbb{Y}_1 - \mathbb{X}_1\breve{\beta}) \\ &+ (2n\sigma^2)^{-1}\Bigg[(\mathbb{Y}_1 - \mathbb{X}_1\breve{\beta})^\top\ddot{\Omega}(\breve{\rho})(\mathbb{Y}_1 - \mathbb{X}_1\breve{\beta}) - (\mathbb{Y}_1 - \mathbb{X}_1\beta)^\top\ddot{\Omega}(\rho)(\mathbb{Y}_1 - \mathbb{X}_1\beta)\Bigg] \\ &\doteq L_{n11} + L_{n12} + L_{n13} + L_{n14}. \end{aligned}$$

Because $n^{-1}\Big[\Omega^{-1}(\bar{\rho})\dddot{\Omega}(\bar{\rho}) - \Omega^{-1}(\bar{\rho})\dot{\Omega}(\bar{\rho})\Omega^{-1}(\bar{\rho})\ddot{\Omega}\Big] = O_p(1)$, we know from (7) that $L_{n11} = o_p(1)$. Similarly, we have $L_{n12} = o_p(1)$ from (6).

For L_{n13}, we have

$$\begin{aligned}
L_{n13} &= (2n)^{-1}(\breve{\sigma}^{-2}-\sigma^{-2})(\mathbb{Y}_1-\mathbb{X}_1\beta)^\top\ddot{\Omega}(\rho)(\mathbb{Y}_1-\mathbb{X}_1\beta) \\
&+ (2n)^{-1}(\breve{\sigma}^{-2}-\sigma^{-2})(\breve{\rho}-\rho)(\mathbb{Y}_1-\mathbb{X}_1\beta)^\top\dddot{\Omega}(\bar{\rho})(\mathbb{Y}_1-\mathbb{X}_1\beta) \\
&- (2n)^{-1}(\breve{\sigma}^{-2}-\sigma^{-2})(\breve{\beta}-\beta)^\top\mathbb{X}_1^\top\ddot{\Omega}(\rho)(\mathbb{Y}_1-\mathbb{X}_1\beta) \\
&- \frac{1}{2n}(\breve{\sigma}^{-2}-\sigma^{-2})(\breve{\rho}-\rho)(\breve{\beta}-\beta)^\top\mathbb{X}_1^\top\dddot{\Omega}(\bar{\rho})(\mathbb{Y}_1-\mathbb{X}_1\beta) \\
&- (2n)^{-1}(\breve{\sigma}^{-2}-\sigma^{-2})(\mathbb{Y}_1-\mathbb{X}_1\beta)^\top\ddot{\Omega}(\rho)(\mathbb{Y}_1-\mathbb{X}_1\beta) \\
&- (2n)^{-1}(\breve{\sigma}^{-2}-\sigma^{-2})(\breve{\rho}-\rho)(\mathbb{Y}_1-\mathbb{X}_1\beta)^\top\dddot{\Omega}(\bar{\rho})(\mathbb{Y}_1-\mathbb{X}_1\beta) \\
&+ (2n)^{-1}(\breve{\sigma}^{-2}-\sigma^{-2})(\breve{\beta}-\beta)^\top\mathbb{X}_1^\top\ddot{\Omega}(\rho)\mathbb{X}_1(\breve{\beta}-\beta) \\
&+ (2n)^{-1}(\breve{\sigma}^{-2}-\sigma^{-2})(\breve{\rho}-\rho)(\breve{\beta}-\beta)^\top\mathbb{X}_1^\top\dddot{\Omega}(\bar{\rho})\mathbb{X}_1(\breve{\beta}-\beta).
\end{aligned}$$

L_{n13} is $o_p(1)$ because $n^{-1}(\mathbb{Y}_1-\mathbb{X}_1\beta)^\top\ddot{\Omega}(\rho)(\mathbb{Y}_1-\mathbb{X}_1\beta)=O_p(1)$, $n^{-1}(\mathbb{Y}_1-\mathbb{X}_1\beta)^\top\dddot{\Omega}(\bar{\rho})(\mathbb{Y}_1-\mathbb{X}_1\beta)=O_p(1)$, $n^{-1}\mathbb{X}_1^\top\ddot{\Omega}(\rho)\mathbb{X}_1=O_p(1)$, and $n^{-1}\mathbb{X}_1^\top\dddot{\Omega}(\bar{\rho})\mathbb{X}_1=O_p(1)$. A similar calculation implies that $L_{n14}=o_p(1)$. Then, $L_{n1}=o_p(1)$ and other blocks of $n^{-1}\dfrac{d^2\ell(\breve{\theta})}{d\theta d\theta^\top}-n^{-1}\dfrac{d^2\ell(\theta)}{d\theta d\theta^\top}$ are also $o_p(1)$. Thus, (3) holds, and this with Lemma 2 leads to Theorem 1.

3. Proof of Theorem 2.

We first denote the lth row of $\Upsilon_{22}^{-1}(\rho)$ by $\Upsilon_{22,l}^{-1}(\rho)$, for $l=1,\cdots,N-n$. By a routine calculation, we divide the difference $\hat{Y}_{n+l}-Y_{n+l}$ as follows:

$$
\begin{aligned}
&\hat{Y}_{n+l} - Y_{n+l} \\
&= X_{n+l}^{\top}(\hat{\beta}-\beta) - \Upsilon_{22,l}^{-1}(\hat{\rho})\Upsilon_{21}(\hat{\rho})(\mathbb{Y}_1 - \mathbb{X}_1\beta) + \Upsilon_{22,l}^{-1}(\hat{\rho})\Upsilon_{21}(\hat{\rho})\mathbb{X}_1(\hat{\beta}-\beta) - V_{n+l} \\
&= E(V_{n+l}|\mathbb{V}_1) - V_{n+l} + \Big[X_{n+l}^{\top} + \Upsilon_{22,l}^{-1}(\rho)\Upsilon_{21}(\rho)\mathbb{X}_1\Big](\hat{\beta}-\beta) \\
&\quad + \Big[\Upsilon_{22,l}^{-1}(\hat{\rho})\Upsilon_{21}(\hat{\rho}) - \Upsilon_{22,l}^{-1}(\rho)\Upsilon_{21}(\rho)\Big]\mathbb{X}_1(\hat{\beta}-\beta) - \Big[\Upsilon_{22,l}^{-1}(\hat{\rho})\Upsilon_{21}(\hat{\rho}) - \Upsilon_{22,l}^{-1}(\rho)\Upsilon_{21}(\rho)\Big]\mathbb{V}_1 \\
&= E(V_{n+l}|\mathbb{V}_1) - V_{n+l} + \Big[X_{n+l}^{\top} + \Upsilon_{22,l}^{-1}(\rho)\Upsilon_{21}(\rho)\mathbb{X}_1\Big](\hat{\beta}-\beta) \\
&\quad + \Big[d\Upsilon_{22,l}^{-1}(\breve{\rho})/d\rho \cdot \Upsilon_{21}(\breve{\rho}) + \Upsilon_{22,l}^{-1}(\breve{\rho})d\Upsilon_{21}(\rho)/d\rho\Big]\mathbb{X}_1(\hat{\beta}-\beta)(\hat{\rho}-\rho) \\
&\quad - \Big[d\Upsilon_{22,l}^{-1}(\breve{\rho})/d\rho \cdot \Upsilon_{21}(\breve{\rho}) + \Upsilon_{22,l}^{-1}(\breve{\rho})d\Upsilon_{21}(\rho)/d\rho\Big]\mathbb{V}_1(\hat{\rho}-\rho) \\
&\doteq E(V_{n+l}|\mathbb{V}_1) - V_{n+l} + L_{n21} + L_{n22} + L_{n23}.
\end{aligned}
$$

From Theorem 1, we have

$$\hat{\beta} - \beta = O_p(n^{-\frac{1}{2}}), \tag{8}$$

and

$$\hat{\rho} - \rho = O_p(n^{-\frac{1}{2}}). \tag{9}$$

Moreover, note that their corresponding coefficients in L_{n21}, L_{n22} , and L_{n23} are all $O_p(1)$, we have $L_{n2i} = O_p(1), i = 1, 2, 3$. This leads to $\hat{Y}_{n+l} - Y_{n+l} \to_d \mathrm{E}(V_{n+l}|\mathbb{V}_1) - V_{n+l}$. Additionally, $\mathbb{V}_1$ and V_{n+l} are normally distributed, and hence $\mathrm{E}(V_{n+l}|\mathbb{V}_1) - V_{n+l}$ is normally distributed. Theorem 2 follows by stating $\Phi = \mathrm{E}\Big[\mathrm{E}(V_{n+l}|\mathbb{V}_1) - V_{n+l}\Big]^2$. This completes the proof.

4. Proof of Theorem 3

By a routine calculation, we have

$$\begin{aligned}
\sqrt{N}(\hat{\mu}_N-\mu) &= N^{-\frac{1}{2}}1_N^\top(\mathbb{X}\beta-\mu)+N^{-\frac{1}{2}}1_N^\top\mathbb{V}+N^{-\frac{1}{2}}1_{N-n}^\top\mathbb{X}_2(\hat{\beta}-\beta)\\
&- N^{-\frac{1}{2}}1_{N-n}^\top\Big[\Upsilon_{22}^{-1}(\hat{\rho})\Upsilon_{21}(\hat{\rho})-\Upsilon_{22}^{-1}(\rho)\Upsilon_{21}(\rho)\Big]\mathbb{V}_1\\
&- N^{-\frac{1}{2}}1_{N-n}^\top\Upsilon_{22}^{-1}(\rho)\Upsilon_{21}(\rho)\mathbb{V}_1-N^{-\frac{1}{2}}1_{N-n}^\top\mathbb{V}_2\\
&+ N^{-\frac{1}{2}}1_{N-n}^\top\Big[\Upsilon_{22}^{-1}(\hat{\rho})\Upsilon_{21}(\hat{\rho})-\Upsilon_{22}^{-1}(\rho)\Upsilon_{21}(\rho)\Big]\mathbb{X}_1(\hat{\beta}-\beta)\\
&+ N^{-\frac{1}{2}}1_{N-n}^\top\Upsilon_{22}^{-1}(\rho)\Upsilon_{21}(\rho)\mathbb{X}_1(\hat{\beta}-\beta)
\end{aligned}$$

$$\begin{aligned}
&= N^{-\frac{1}{2}}1_N^\top(\mathbb{X}\beta-\mu)+N^{-\frac{1}{2}}1_{N-n}^\top\Big[\mathbb{X}_2+\Upsilon_{22}^{-1}(\rho)\Upsilon_{21}(\rho)\mathbb{X}_1\Big](\hat{\beta}-\beta)\\
&+ N^{-\frac{1}{2}}\Big[1_n^\top-1_{N-n}^\top\Upsilon_{22}^{-1}(\rho)\Upsilon_{21}(\rho)\Big]\mathbb{V}_1\\
&- N^{-\frac{1}{2}}1_{N-n}^\top\Big[d\Upsilon_{22}^{-1}(\breve{\rho})/d\rho\cdot\Upsilon_{21}(\breve{\rho})+\Upsilon_{22}^{-1}(\breve{\rho})d\Upsilon_{21}(\rho)/d\rho\Big]\mathbb{V}_1(\hat{\rho}-\rho) \qquad (10)\\
&+ N^{-\frac{1}{2}}1_{N-n}^\top\Big[d\Upsilon_{22}^{-1}(\breve{\rho})/d\rho\cdot\Upsilon_{21}(\breve{\rho})+\Upsilon_{22}^{-1}(\breve{\rho})d\Upsilon_{21}(\rho)/d\rho\Big]\mathbb{X}_1(\hat{\beta}-\beta)(\hat{\rho}-\rho)\\
&\doteq \sum_{i=1}^{5}L_{3ni}.
\end{aligned}$$

Moreover, from Theorem 1, we have $\hat{\beta}-\beta = \Big[\mathbb{X}_1^\top\Omega(\rho)\mathbb{X}_1\Big]^{-1}\mathbb{X}_1^\top\Omega(\rho)\mathbb{V}_1+o_p(1)$, and $\hat{\rho}-\rho$ is independent of $\Sigma_{11}^{-\frac{1}{2}}\mathbb{V}_1$. Thus, we obtain

$$L_{3n2} = \frac{1}{\sqrt{N}}1_{N-n}^\top\Big[\mathbb{X}_2+\Upsilon_{22}^{-1}(\rho)\Upsilon_{21}(\rho)\mathbb{X}_1\Big]\Big[\mathbb{X}_1^\top\Omega(\rho)\mathbb{X}_1\Big]^{-1}\mathbb{X}_1^\top\Omega(\rho)\mathbb{V}_1+o_p(1), \qquad (11)$$

With (8) and (9), we have

$$L_{3ni} \;=\; o_p(1), \tag{12}$$

for $i = 4$ and 5. Combining (10),(11), and (12), we get

$$\begin{aligned}
\sqrt{N}(\hat{\mu}_N - \mu) \quad &= \quad \frac{1}{\sqrt{N}} 1_N^\top (\mathbb{X}\beta - \mu) + \frac{1}{\sqrt{N}} \Big[1_n^\top - 1_{N-n}^\top \Upsilon_{22}^{-1}(\rho) \Upsilon_{21}(\rho) \Big] \mathbb{V}_1 \\
&\quad + \frac{1}{\sqrt{N}} 1_{N-n}^\top \Big[\mathbb{X}_2 + \Upsilon_{22}^{-1}(\rho) \Upsilon_{21}(\rho) \mathbb{X}_1 \Big] \Big[\mathbb{X}_1^\top \Omega(\rho) \mathbb{X}_1 \Big]^{-1} \mathbb{X}_1^\top \Omega(\rho) \mathbb{V}_1 + o_p(1) \\
&= \quad \frac{1}{\sqrt{N}} 1_N^\top (\mathbb{X}\beta - \mu) + \frac{1}{\sqrt{n}} \sqrt{r} \Big[1_n^\top - 1_{N-n}^\top \Upsilon_{22}^{-1}(\rho) \Upsilon_{21}(\rho) \Big] \mathbb{V}_1 \\
&\quad + \frac{1}{\sqrt{n}} \sqrt{r} 1_{N-n}^\top \Big[\mathbb{X}_2 + \Upsilon_{22}^{-1}(\rho) \Upsilon_{21}(\rho) \mathbb{X}_1 \Big] \Big[\mathbb{X}_1^\top \Omega(\rho) \mathbb{X}_1 \Big]^{-1} \mathbb{X}_1^\top \Omega(\rho) \mathbb{V}_1 + o_p(1) \\
&= \quad \frac{1}{\sqrt{N}} 1_N^\top (\mathbb{X}\beta - \mu) + \frac{1}{\sqrt{n}} \sqrt{r} \Big[1_n^\top - 1_{N-n}^\top \Upsilon_{22}^{-1}(\rho) \Upsilon_{21}(\rho) + B \Big] \mathbb{V}_1 + o_p(1) \\
&\xrightarrow{d} \quad N(0, \phi).
\end{aligned} \tag{13}$$

Thus, Theorem 3 follows from (13). This completes the proof.