INFERENCE ON LOW-RANK DATA MATRICES WITH APPLICATIONS TO MICROARRAY DATA

By Xingdong Feng and Xuming He

University of Illinois at Urbana-Champaign

1. SUPPLEMENTARY MATERIAL.

Lemma 1.1. *Consider an $m\times m$ matrix Σ. Denote the eigenvalues of the matrix as λ_i and the corresponding eigenvectors as $\underline{\phi}_i$, $i=1,\cdots,m$. Assume that $\{\Sigma_n\}$ is a sequence of matrices with $\Sigma_n \to \Sigma$ as $n\to\infty$. Denote the eigenvalues of the matrix Σ_n as λ_{in} and the corresponding eigenvectors as $\underline{\phi}_{in}$, for $i=1,\cdots,m$ and $\lambda_{1n}\ge\lambda_{2n}\ge\lambda_{3n}\ge\cdots\ge\lambda_{mn}\ge 0$. If $\lambda_1>\lambda_2>\cdots>\lambda_l\ge\cdots\ge\lambda_m\ge 0$, where $l\in\{2,3,\cdots,m\}$, then $\underline{\phi}_{in}\to\underline{\phi}_i$ for $i=1,2,\cdots,l-1$.*

Proof of Lemma 1.1 Consider a convergent subsequence of $\{\underline{\phi}_{1n}\}$, denoted as $\{\underline{\phi}_{1n_m}\}$, with the limit as $\underline{\phi}_0$. Obviously, $\lambda_{1n_m}=\underline{\phi}_{1n_m}^T\Sigma_{n_m}\underline{\phi}_{1n_m}$. Hence, $\lambda_{1n_m}\to\underline{\phi}_0^T\Sigma\underline{\phi}_0$. Taking limits on both sides of $\Sigma_{n_m}\underline{\phi}_{1n_m}=\lambda_{1n_m}\underline{\phi}_{1n_m}$, we have $\Sigma\underline{\phi}_0=(\underline{\phi}_0^T\Sigma\underline{\phi}_0)\underline{\phi}_0$. Therefore, $\underline{\phi}_0$ is an eigenvector of Σ.

It then follows that if $\underline{\phi}_0\neq\underline{\phi}_1$ or $\underline{\phi}_0\neq-\underline{\phi}_1$, then $\underline{\phi}_0^T\Sigma\underline{\phi}_0\le\lambda_2$.

Since λ_{1n_m} is the largest eigenvalue of the matrix Σ_{n_m}, if follows that $\underline{\phi}^T\Sigma_{n_m}\underline{\phi}\le\underline{\phi}_{1n_m}^T\Sigma_{n_m}\underline{\phi}_{1n_m}=\lambda_{1n_m}$ for any $\underline{\phi}$ satisfying $\underline{\phi}^T\underline{\phi}=1$. Taking limits, we obtain $\underline{\phi}^T\Sigma\underline{\phi}\le\underline{\phi}_0^T\Sigma\underline{\phi}_0\le\lambda_2<\lambda_1$ for any $\underline{\phi}$ satisfying $\underline{\phi}^T\underline{\phi}=1$, which is a contradiction. Therefore, any convergent subsequence of $\{\underline{\phi}_{1n}\}$ converges to $\underline{\phi}_1$.

Let $\Sigma_n^*=\Sigma_n-\lambda_{1n}\underline{\phi}_{1n}\underline{\phi}_{1n}^T$ and $\Sigma^*=\Sigma-\lambda_1\underline{\phi}_1\underline{\phi}_1^T$, then $\lim_{n\to\infty}\Sigma_n^*=\Sigma^*$. Repeat the above proof on Σ_n^*, we can show that $\underline{\phi}_{2n}\to\underline{\phi}_2$. Recursively, for $k\le l-1<m$, we can show that $\underline{\phi}_{kn}\to\underline{\phi}_k$. $\diamond$

Proof of Theorem 2.1 Under Model (2), we have

$$\begin{aligned}
n^{-1}\sum_{i=1}^n\underline{y}_i\underline{y}_i^T &= n^{-1}||\underline{\theta}_1^{(0)}||^2\underline{\phi}_1^{(0)}\underline{\phi}_1^{(0)T}+n^{-1}||\underline{\theta}_2^{(0)}||^2\underline{\phi}_2^{(0)}\underline{\phi}_2^{(0)T}\\
&+ n^{-1}\underline{\phi}_1^{(0)}(\underline{\theta}_1^{(0)T}\mathbf{E})+n^{-1}\underline{\phi}_2^{(0)}(\underline{\theta}_2^{(0)T}\mathbf{E})\\
&+ n^{-1}(\underline{\theta}_1^{(0)T}\mathbf{E})^T\underline{\phi}_1^{(0)T}+n^{-1}(\underline{\theta}_2^{(0)T}\mathbf{E})^T\underline{\phi}_2^{(0)T}+n^{-1}\mathbf{E}^T\mathbf{E}\\
&+ n^{-1}\underline{\theta}_1^{(0)T}\underline{\theta}_2^{(0)}(\underline{\phi}_1^{(0)}\underline{\phi}_2^{(0)T}+\underline{\phi}_2^{(0)}\underline{\phi}_1^{(0)T}),
\end{aligned}$$

where $\mathbf{E} = (\underline{\varepsilon}_1, \cdots, \underline{\varepsilon}_n)^T$.

Therefore, $n^{-1}\sum_{i=1}^n \underline{y}_i\underline{y}_i^T \xrightarrow{a.s.} \Gamma$, where Γ is given in (11), from the Kolmogorov strong law of large numbers under Assumptions (M1)–(M6). It follows from Lemma 1.1 that $\underline{\hat{\phi}}_1 \xrightarrow{a.s.} \underline{\phi}_1^{(0)}$ and $\underline{\hat{\phi}}_2 \xrightarrow{a.s.} \underline{\phi}_2^{(0)}$, because $\underline{\phi}_1^{(0)}$ and $\underline{\phi}_2^{(0)}$ are the eigenvectors of Γ corresponding to the largest eigenvalues $\mu_1^2 + \sigma_1^2 + \sigma^2$ and $\mu_2^2 + \sigma_2^2 + \sigma^2$, respectively.$\diamond$

Proof of Theorem 2.2

It is clear that $\Gamma_n \to \Gamma$ as $n \to \infty$, where Γ_n is given in (8). Also, Γ_n has the same eigenspace as Γ does, and $E(n^{-1}\sum_{i=1}^n \underline{y}_i\underline{y}_i^T) = \Gamma_n$.

The gradient vector of $\rho(\underline{y}_i;\varphi) - (\underline{\phi}_1^T\underline{y}_i\underline{y}_i^T\underline{\phi}_1)(1-\underline{\phi}_1^T\underline{\phi}_1) - (\underline{\phi}_2^T\underline{y}_i\underline{y}_i^T\underline{\phi}_2)(1-\underline{\phi}_2^T\underline{\phi}_2)$ is

$$\psi(\underline{y}_i;\varphi) = \begin{pmatrix} -2\underline{y}_i\underline{y}_i^T\underline{\phi}_1 + 2(\underline{\phi}_1^T\underline{y}_i\underline{y}_i^T\underline{\phi}_1)\underline{\phi}_1 \\ -2\underline{y}_i\underline{y}_i^T\underline{\phi}_2 + 2(\underline{\phi}_2^T\underline{y}_i\underline{y}_i^T\underline{\phi}_2)\underline{\phi}_2 \end{pmatrix},$$

and the expectation of $\sum_{i=1}^n \psi(\underline{y}_i;\varphi)$ is

$$\Lambda_n(\varphi) = \begin{pmatrix} -n(2\Gamma_n\underline{\phi}_1 - 2(\underline{\phi}_1^T\Gamma_n\underline{\phi}_1)\underline{\phi}_1) \\ -n(2\Gamma_n\underline{\phi}_2 - 2(\underline{\phi}_2^T\Gamma_n\underline{\phi}_2)\underline{\phi}_2) \end{pmatrix},$$

where $||\underline{\phi}_1|| = ||\underline{\phi}_2|| = 1$.

To give the Bahadur representation of $\underline{\hat{\phi}}_1$ and $\underline{\hat{\phi}}_2$, we shall use Corollary 2.2 of He and Shao (1996). The main steps are to verify conditions (B3) and (B4) among four conditions of He and Shao (1996) because (B1) is obvious and (B2) is shown in Lemma 2.1.

Consider

$$\psi(\underline{y}_i;\varphi_1)-\psi(\underline{y}_i;\varphi_2) = \begin{pmatrix} -2\underline{y}_i\underline{y}_i^T(\underline{\phi}_1 - \underline{\nu}_1) + 2[(\underline{\phi}_1^T\underline{y}_i\underline{y}_i^T\underline{\phi}_1)\underline{\phi}_1 - (\underline{\nu}_1^T\underline{y}_i\underline{y}_i^T\underline{\nu}_1)\underline{\nu}_1] \\ -2\underline{y}_i\underline{y}_i^T(\underline{\phi}_2 - \underline{\nu}_2) + 2[(\underline{\phi}_2^T\underline{y}_i\underline{y}_i^T\underline{\phi}_2)\underline{\phi}_2 - (\underline{\nu}_2^T\underline{y}_i\underline{y}_i^T\underline{\nu}_2)\underline{\nu}_2] \end{pmatrix},$$

where $\varphi_1 = (\underline{\phi}_1^T, \underline{\phi}_2^T)^T$ and $\varphi_2 = (\underline{\nu}_1^T, \underline{\nu}_2^T)^T$, and $||\underline{\phi}_1|| = ||\underline{\phi}_2|| = ||\underline{\nu}_1|| = ||\underline{\nu}_2|| = 1$. Note that

$$\begin{aligned} & 2\underline{y}_i\underline{y}_i^T(\underline{\phi}_j - \underline{\nu}_j) - 2[(\underline{\phi}_j^T\underline{y}_i\underline{y}_i^T\underline{\phi}_j)\underline{\phi}_j - (\underline{\nu}_j^T\underline{y}_i\underline{y}_i^T\underline{\nu}_j)\underline{\nu}_j] \\ = \ & 2\underline{y}_i\underline{y}_i^T(\underline{\phi}_j - \underline{\nu}_j) - 2[(\underline{\phi}_j - \underline{\nu}_j)^T\underline{y}_i\underline{y}_i^T(\underline{\phi}_j - \underline{\nu}_j) + 2\underline{\nu}_j^T\underline{y}_i\underline{y}_i^T(\underline{\phi}_j - \underline{\nu}_j)]\underline{\nu}_j, \end{aligned}$$

for $j = 1, 2$. Under the assumption that $\underline{\theta}_1^{(0)}$, $\underline{\theta}_2^{(0)}$ and $\underline{\varepsilon}_i$ have finite fourth moments, we have

$$\begin{aligned} & E\{\sup_{|\underline{\phi}_j - \underline{\nu}_j| \le d} |2\underline{y}_i\underline{y}_i^T(\underline{\phi}_j - \underline{\nu}_j) - 2[(\underline{\phi}_j^T\underline{y}_i\underline{y}_i^T\underline{\phi}_j)\underline{\phi}_j - (\underline{\nu}_j^T\underline{y}_i\underline{y}_i^T\underline{\nu}_j)\underline{\nu}_j]|\}^2 \\ \le \ & E(||\underline{y}_i||^4)m^2[1 + 2(d+2)m]d^2, \end{aligned}$$

for $i = 1, \cdots, n$ and $j = 1, 2$. Hence, if Assumption (M6) holds, the conditions (B3) and (B4) in He and Shao (1996) are satisfied with $r = 2$.

The derivative of $\Lambda_n(\vartheta)$ is:

$$D_n(\varphi) \triangleq n \begin{pmatrix} -2\Gamma_n + 2(\underline{\phi}_1^T \Gamma_n \underline{\phi}_1 I_m + 2\underline{\phi}_1 \underline{\phi}_1^T \Gamma_n) & \mathbf{0} \\ \mathbf{0} & -2\Gamma_n + 2(\underline{\phi}_2^T \Gamma_n \underline{\phi}_2 I_m + 2\underline{\phi}_2 \underline{\phi}_2^T \Gamma_n) \end{pmatrix}.$$

It then follows that

$$D_n(\varphi^{(0)}) = n \begin{pmatrix} D_{1n} & \mathbf{0} \\ \mathbf{0} & D_{2n} \end{pmatrix},$$

where D_{jn} are given in (10).

When n is sufficiently large, $\underline{\phi}_1^{(0)}$ and $\underline{\phi}_2^{(0)}$ are the eigenvectors of both Γ_n and Γ, corresponding to the largest and the second largest eigenvalues, respectively. Let $\lambda_{jn} = n^{-1}||\underline{\mu}_j||^2 + \sigma_j^2 + \sigma^2$, for $j = 1, 2$ and $\lambda_j = \sigma^2$ for $j = 3, ..., m$. It can be easily verified that the eigenvectors of Γ_n are the same as the eigenvectors of D_{2n}. When n is sufficiently large, the eigenvalues of D_{2n} corresponding to $\underline{\phi}_1^{(0)}$ and $\underline{\phi}_2^{(0)}$ are $2(\lambda_{2n} - \lambda_{1n})$ and $4\lambda_{2n}$, respectively. The other eigenvalues of the matrix D_{2n} are $2(\lambda_{2n} - \lambda_j)$ for $j = 3, ..., m$. Similarly, we can verify that $\underline{\phi}_1^{(0)}$ and $\underline{\phi}_2^{(0)}$ are the eigenvectors of D_{1n}, and the corresponding eigenvalues are $4\lambda_{1n}$ and $2(\lambda_{1n} - \lambda_{2n})$. The other eigenvalues are $2(\lambda_{1n} - \lambda_j)$, $j = 3, ..., m$. Hence, $D_n(\varphi^{(0)})$ is a nonsingular matrix if the three largest eigenvalues of Γ_n are strictly ordered, which is obviously true when n is sufficiently large and Assumption (M5) holds.

Since

$$\begin{aligned} & [-2\Gamma_n + 2(\underline{\phi}_j^T \Gamma_n \underline{\phi}_j I_m + 2\underline{\phi}_j \underline{\phi}_j^T \Gamma_n)] - D_{jn} \\ = \;& 2(\underline{\phi}_j^T \Gamma_n \underline{\phi}_j - \underline{\phi}_j^{(0)T} \Gamma_n \underline{\phi}_j^{(0)}) I_m + 4(\underline{\phi}_j \underline{\phi}_j^T - \underline{\phi}_j^{(0)} \underline{\phi}_j^{(0)T}) \Gamma_n \\ = \;& 2(\underline{\phi}_j - \underline{\phi}_j^{(0)})^T \Gamma_n (\underline{\phi}_j + \underline{\phi}_j^{(0)}) I_m + 4[(\underline{\phi}_j - \underline{\phi}_j^{(0)}) \underline{\phi}_j^T + \underline{\phi}_j^{(0)T} (\underline{\phi}_j - \underline{\phi}_j^{(0)})] \Gamma_n, \end{aligned}$$

and $||\underline{\phi}_j|| = ||\underline{\phi}_j^{(0)}|| = 1$ and $\Gamma_n \to \Gamma$, for $j = 1, 2$, it then follows that there exists a constant κ_0 such that $|D_n(\varphi) - D_n(\varphi^{(0)})| \leq \kappa_0 n |\varphi - \varphi^{(0)}|$ in a neighborhood of $\varphi^{(0)}$ when n is sufficiently large, where $|\cdot|$ is taken to be the sup norm as in He and Shao (1996). Hence, from Corollary 2.2 of He and Shao (1996), we have (10).$\diamond$

Proof of Theorem 2.3

Since

$$\begin{aligned} & [2\underline{y}_i \underline{y}_i^T \underline{\phi}_2^{(0)} - 2(\underline{\phi}_2^{(0)T} \underline{y}_i \underline{y}_i^T \underline{\phi}_2^{(0)}) \underline{\phi}_2^{(0)}]^T [D_{2n}^{-1}]^T \underline{\phi}_1^{(0)} \\ = \;& (\lambda_{2n} - \lambda_{1n})^{-1} (\theta_{1i}^{(0)} \theta_{2i}^{(0)} + \theta_{1i}^{(0)} \underline{\phi}_2^{(0)T} \underline{\varepsilon}_i + \theta_{2i}^{(0)} \underline{\phi}_1^{(0)T} \underline{\varepsilon}_i + \underline{\phi}_2^{(0)T} \underline{\varepsilon}_i \underline{\varepsilon}_i^T \underline{\phi}_1^{(0)}), \end{aligned}$$

and $[2\underline{y}_i\underline{y}_i^T\underline{\phi}_2^{(0)} - 2(\underline{\phi}_2^{(0)T}\underline{y}_i\underline{y}_i^T\underline{\phi}_2^{(0)})\underline{\phi}_2^{(0)}]^T[D_{2n}^{-1}]^T\underline{\phi}_2^{(0)} = 0$, we have

$$(1)\ \hat{\underline{\phi}}_2^T\underline{y}_i = (\lambda_{1n}-\lambda_{2n})^{-1}\theta_{1i}^{(0)}\zeta_n + (\hat{\underline{\phi}}_2 - \underline{\phi}_2^{(0)})^T\underline{\varepsilon}_i + \theta_{2i}^{(0)} + \underline{\varepsilon}_i^T\underline{\phi}_2^{(0)} + o(n^{-1+\epsilon}),$$

where

$$(2)\quad \zeta_n = n^{-1}[\underline{\theta}_1^{(0)T}\underline{\theta}_2^{(0)} + (\underline{\theta}_1^{(0)T}\mathbf{E})\underline{\phi}_2^{(0)} + (\underline{\theta}_2^{(0)T}\mathbf{E})\underline{\phi}_1^{(0)} + \underline{\phi}_2^{(0)T}(\mathbf{E}^T\mathbf{E})\underline{\phi}_1^{(0)}],$$

from Theorem 2.2. Similarly,

$$(3)\ \hat{\underline{\phi}}_1^T\underline{y}_i = (\lambda_{2n}-\lambda_{1n})^{-1}\theta_{2i}^{(0)}\zeta_n + (\hat{\underline{\phi}}_1 - \underline{\phi}_1^{(0)})^T\underline{\varepsilon}_i + \theta_{1i}^{(0)} + \underline{\varepsilon}_i^T\underline{\phi}_1^{(0)} + o(n^{-1+\epsilon}).$$

By the strong law of large numbers, ζ_n converge to 0 almost surely. Hence, we have $\hat{\theta}_{1i} = \hat{\underline{\phi}}_1^T\underline{y}_i \xrightarrow{L} \theta_{1i}^{(0)} + \underline{\varepsilon}_i^T\underline{\phi}_1^{(0)}$ and $\hat{\theta}_{2i} = \hat{\underline{\phi}}_2^T\underline{y}_i \xrightarrow{L} \theta_{2i}^{(0)} + \underline{\varepsilon}_i^T\underline{\phi}_2^{(0)}$.$\diamond$

Proof of Theorem 3.1

By (5) and (7), the Bahadur representation of the test statistic $T_{\underline{a}}$ is given as follows:

$$\begin{aligned}
T_{\underline{a}} = \quad & n^{-1}\underline{a}^T(\mathbf{Y}\hat{\underline{\phi}}_2) \\
= \quad & (\lambda_{1n}-\lambda_{2n})^{-1}\zeta_n(n^{-1}\underline{a}^T\underline{\theta}_1^{(0)}) + n^{-1}\underline{a}^T\underline{\theta}_2^{(0)} + n^{-1}(\underline{a}^T\mathbf{E})^T\underline{\phi}_2^{(0)} + o(n^{-1+\epsilon}) \\
= \quad & (\lambda_{1n}-\lambda_{2n})^{-1}\zeta_n(n^{-1}\underline{a}^T\underline{\mu}_1) + n^{-1}\underline{a}^T\underline{\theta}_2^{(0)} + n^{-1}(\underline{a}^T\mathbf{E})^T\underline{\phi}_2^{(0)} + o_p(n^{-1+\epsilon}) \\
= \quad & n^{-1}\underline{a}^T\underline{\theta}_2^{(0)} + n^{-1}(\underline{a}^T\mathbf{E})^T\underline{\phi}_2^{(0)} + o_p(n^{-1+\epsilon}),
\end{aligned}$$

where ζ_n is given in (18). Therefore, under the null hypothesis H_0, we have $\sqrt{n}T_{\underline{a}} \xrightarrow{L} N(0, \sigma_2^2 + \sigma^2)$, and thus, $\sqrt{n}T_{\underline{a}}/\hat{\sigma} \xrightarrow{L} N(0,1)$, where $\hat{\sigma}$ and $\hat{\theta}_2$. are given in (14).

Under the alternative hypothesis,

$$n^{-1}\underline{a}^T\underline{\theta}_2 - n^{-1}\underline{a}^T\underline{\mu}_2 \xrightarrow{a.s.} 0$$

and

$$\hat{\sigma}^2 \xrightarrow{a.s.} \mu_2^2 + \sigma_2^2 + \sigma^2,$$

so we should expect to observe larger $T_{\underline{a}}$ when $n^{-1}\underline{a}^T\underline{\mu}_2$ does not converge to 0.$\diamond$

Proof of Theorem 3.2

When k is fixed, it follows that

$$\begin{aligned} & n^{-1}\underline{\hat{\theta}}_2^T A^T A \underline{\hat{\theta}}_2 \\ = & \sum_{j=1}^{k} (n^{-1/2} \underline{a}_j^T \underline{\hat{\theta}}_2)^2 \\ = & n^{-1} \sum_{j=1}^{k} [\underline{a}_j^T \underline{\theta}_2^{(0)} + (\underline{a}_j^T \mathbf{E}) \underline{\phi}_2^{(0)}]^2 + o(1) \\ = & n^{-1} \underline{\theta}^T A^T A \underline{\theta} + o(1), \end{aligned}$$

where $\underline{\theta} = (\theta_{21}^{(0)} + \underline{\varepsilon}_1^T \underline{\phi}_2^{(0)}, \theta_{22}^{(0)} + \underline{\varepsilon}_2^T \underline{\phi}_2^{(0)}, \cdots, \theta_{2n}^{(0)} + \underline{\varepsilon}_n^T \underline{\phi}_2^{(0)})^T$. Under the null hypothesis, $cov(A\underline{\theta}) = Acov(\underline{\theta})A^T = (\sigma_2^2 + \sigma^2) I_k$, so $A\underline{\theta}$ is jointly normally distributed with mean 0 and variance-covariance matrix $(\sigma_2^2 + \sigma^2) I_k$, asymptotically. Thus,

$$P\big(n^{-1}\underline{\hat{\theta}}_2^T A^T A \underline{\hat{\theta}}_2 / (\sigma^2 + \sigma_2^2) \le x\big) - F_k(x) \to 0,$$

where F_k is the cumulative distribution function of the χ_k^2 distribution. Hence,

$$P\big(n^{-1}\underline{\hat{\theta}}_2^T A^T A \underline{\hat{\theta}}_2 / \hat{\sigma}^2 \le x\big) - F_k(x) \to 0, \tag{4}$$

where $\hat{\sigma}^2$ is given in (14), under the null hypothesis.$\diamond$

Proof of Theorem 3.3

Firstly, let us show that $M_n = \widetilde{M}_n + o(n^{-1/2+\epsilon})$. Actually, from (5) and (17), we have

$$n^{-1} \underline{a}_j^T \underline{\hat{\theta}}_2 = n^{-1} \underline{a}_j^T \underline{\theta}_2^{(0)} + n^{-1} (\underline{a}_j^T \mathbf{E}) \underline{\phi}_2^{(0)} + n^{-1} (\underline{a}_j^T \mathbf{Y}) \underline{\varsigma}_n, \tag{5}$$

where $\underline{\varsigma}_n = o(n^{-1+\epsilon})$ uniformly in A. Since

$$\max_{1 \le j \le n-1} \left| \left(n^{-1} \underline{a}_j^T \mathbf{Y} \right) \underline{\varsigma}_n \right| \le |\underline{\varsigma}_n| \max_{1 \le j \le n-1} \left| n^{-1} \underline{a}_j^T \mathbf{Y} \right|$$

and

$$\begin{aligned} \left| n^{-1} \underline{a}_j^T \mathbf{Y} \right| & \le \max_{1 \le k \le m} n^{-1} \sum_{i=1}^{n} |a_{ji} y_{ik}| \\ & \le \max_{1 \le k \le m} (2n)^{-1} (||\underline{a}_j||^2 + ||\underline{y}_k||^2) = \max_{1 \le k \le m} [2^{-1} + (2n)^{-1} ||\underline{y}_k||^2], \end{aligned}$$

it then follows that

$$\max_{1\le j\le k-1} n^{-1}\underline{a}_j^T\hat{\underline{\theta}}_2 = \max_{1\le j\le k-1} n^{-1}[\underline{a}_j^T\underline{\theta}_2^{(0)} + (\underline{a}_j^T\mathbf{E})\underline{\phi}_2^{(0)}] + o_p(n^{-1+\epsilon}). \tag{6}$$

Secondly,

$$cov\Big(n^{-1/2}\underline{a}_d^T\underline{\theta}, n^{-1/2}\underline{a}_t^T\underline{\theta}\Big) = n^{-1}\underline{a}_d^T cov(\underline{\theta})\underline{a}_t,$$

where $\underline{\theta} = (\theta_{21}^{(0)}+\underline{\varepsilon}_1^T\underline{\phi}_2^{(0)}, \cdots, \theta_{2n}^{(0)}+\underline{\varepsilon}_n^T\underline{\phi}_2^{(0)})^T$ and $d, t \in \{1, 2, \cdots, n-1\}$. Since $cov(\underline{\theta}) = (\sigma_2^2+\sigma^2)I_n$ under the null hypothesis, we have $\underline{a}_d^T cov(\underline{\theta})\underline{a}_t = 0$ when $d \neq t$. It then follows that

$$P\Big(c_n(\widetilde{M}_n/\sqrt{\sigma_2^2+\sigma^2} - b_n) \le x\Big) \to e^{-e^{-x}}$$

under the null hypothesis by Berman (1964).

Finally, since $\hat{\sigma}^2 = \sigma_2^2 + \sigma^2 + o_p(n^{-1/2+\epsilon})$, it follows from the Slutsky theorem that

$$\begin{aligned} & c_n(M_n/\hat{\sigma} - b_n) - c_n\Big(\widetilde{M}_n/\sqrt{\sigma_2^2+\sigma^2} - b_n\Big) \\ = \quad & \frac{c_n(M_n - \widetilde{M}_n)}{\hat{\sigma}} + c_n\Big(\frac{\widetilde{M}_n}{\sqrt{\sigma_2^2+\sigma^2}} - b_n\Big)\Big(\frac{\sqrt{\sigma_2^2+\sigma^2}}{\hat{\sigma}} - 1\Big) + c_n b_n \frac{\sqrt{\sigma_2^2+\sigma^2} - \hat{\sigma}}{\hat{\sigma}} \xrightarrow{p} 0. \end{aligned}$$

Hence,

$$P\big(c_n(M_n/\hat{\sigma} - b_n) \le x\big) \to e^{-e^{-x}}.\diamond$$

References.

Berman, S. (1964). Limit Theorems for The Maximum Term in Stationary Sequences. *The Annals of Mathematical Statistics* **35**, 502–516.

He, X. and Shao, Q. (1996). A General Bahadur Representation of M-Estimators and Its Application to Linear Regression With Nonstochastic Designs. *The Annals of Statistics* **24**, 2608–2630.

E-mail: xfeng2@uiuc.edu

E-mail: x-he@uiuc.edu