SUPPLEMENT TO "RIDGE REGRESSION REVISITED: DEBIASING, THRESHOLDING AND BOOTSTRAP"

BY YUNYI ZHANG [1] AND DIMITRIS N. POLITIS [2]

[1] *Department of Mathematics, Univ. of California–San Diego, yuz334@ucsd.edu*

[2] *Department of Mathematics and Halicioglu Data Science Institute, Univ. of California–San Diego, dpolitis@ucsd.edu*

APPENDIX A: SOME IMPORTANT LEMMAS

This section introduces three useful lemmas. Lemma A.1 comes from Whittle (1960), which directly contributes to the model selection consistency. Lemma A.2 and A.3 are similar to Chernozhukov, Chetverikov and Kato (2013), they used a joint normal distribution to approximate the distribution of linear combinations of independent random variables.

LEMMA A.1. *Suppose random variables $\epsilon_1, ..., \epsilon_n$ are i.i.d., $\mathbf{E}\epsilon_1 = 0$, and $\exists$ a constant $m > 0$ such that $\mathbf{E}|\epsilon_1|^m < \infty$. In addition suppose the matrix $\Gamma = (\gamma_{ij})_{i=1,2,...,k,j=1,2,...,n}$ satisfies*

$$\max_{i=1,2,...,k}\sum_{j=1}^n \gamma_{ij}^2 \leq D,\ D > 0 \tag{A.1}$$

Then $\exists$ a constant E which only depends on m and $\mathbf{E}|\epsilon_1|^m$ such that for $\forall\delta > 0$,

$$Prob\left(\max_{i=1,2,...,k}|\sum_{j=1}^n \gamma_{ij}\epsilon_j| > \delta\right) \leq \frac{kED^{m/2}}{\delta^m} \tag{A.2}$$

PROOF. From theorem 2 in Whittle (1960), for any $i = 1, 2, ..., k$,

$$Prob\left(|\sum_{j=1}^n \gamma_{ij}\epsilon_j| > \delta\right) \leq \frac{\mathbf{E}|\sum_{j=1}^n \gamma_{ij}\epsilon_j|^m}{\delta^m} \leq \frac{2^m C(m)\mathbf{E}|\epsilon_1|^m(\sum_{j=1}^n \gamma_{ij}^2)^{m/2}}{\delta^m} \leq \frac{2^m C(m)\mathbf{E}|\epsilon_1|^m D^{m/2}}{\delta^m} \tag{A.3}$$

Here $C(m)$ is a constant depending on m. Choose $E = 2^m C(m)\mathbf{E}|\epsilon_1|^m$,

$$Prob\left(\max_{i=1,2,...,k}|\sum_{j=1}^n \gamma_{ij}\epsilon_j| > \delta\right) \leq \sum_{i=1}^k Prob\left(|\sum_{j=1}^n \gamma_{ij}\epsilon_j| > \delta\right) \leq \frac{kED^{m/2}}{\delta^m} \tag{A.4}$$

$\square$

LEMMA A.2. *Suppose $\epsilon = (\epsilon_1, ..., \epsilon_n)^T$ are joint normal random variables with mean $\mathbf{E}\epsilon = 0$, non-singular covariance matrix $\mathbf{E}\epsilon\epsilon^T$, and positive marginal variance $\sigma_i^2 = \mathbf{E}\epsilon_i^2 > 0$, $i = 1, 2, ..., n$. In addition, suppose $\exists$ two constants $0 < c_0 \leq C_0 < \infty$ such that $c_0 \leq \sigma_i \leq C_0$ for $i = 1, 2, ..., n$. Then for any given $\delta > 0$,*

$$\sup_{x\in\mathbf{R}}\left(Prob(\max_{i=1,2,...,n}|\epsilon_i| \leq x + \delta) - Prob(\max_{i=1,2,...,n}|\epsilon_i| \leq x)\right) \leq C\delta(\sqrt{\log(n)} + \sqrt{|\log(\delta)|} + 1) \tag{A.5}$$

C only depends on c_0 and C_0.

PROOF OF LEMMA A.2. First for any $i = 1, 2, ..., n$,

$$|\epsilon_i| = \max(\epsilon_i, -\epsilon_i) \Rightarrow \max_{i=1,...,n} |\epsilon_i| = \max(\max_{i=1,...,n} \epsilon_i, \max_{i=1,...,n} -\epsilon_i) \tag{A.6}$$

Therefore, for any $x \in \mathbf{R}$,

$$\begin{aligned} &Prob(\max_{i=1,2,...,n} |\epsilon_i| \leq x+\delta) - Prob(\max_{i=1,2,...,n} |\epsilon_i| \leq x) \\ &= Prob(0 < \max(\max_{i=1,...,n} \epsilon_i, \max_{i=1,...,n} -\epsilon_i) - x \leq \delta) \\ &\leq Prob(0 < \max_{i=1,...,n} \epsilon_i - x \leq \delta) + Prob(0 < \max_{i=1,...,n} -\epsilon_i - x \leq \delta) \\ &\leq Prob(|\max_{i=1,...,n} \epsilon_i - x| \leq \delta) + Prob(|\max_{i=1,...,n} -\epsilon_i - x| \leq \delta) \end{aligned} \tag{A.7}$$

$-\epsilon$ is also joint normal with mean 0 and marginal variance $\mathbf{E}(-\epsilon_j)^2 = \sigma_j^2$. From theorem 3 and (18), (19) in Chernozhukov, Chetverikov and Kato (2015), by defining $\underline{\sigma} = \min_{i=1,2,...,n} \sigma_i \leq \max_{i=1,2,...,n} \sigma_i = \overline{\sigma}$, we have

(A.8)

$$\begin{aligned} \sup_{x\in\mathbf{R}} Prob\left(|\max_{i=1,2,...,n} \epsilon_i - x| \leq \delta\right) &\leq \frac{\sqrt{2}\delta}{\underline{\sigma}}\left(\sqrt{\log(n)} + \sqrt{\max(1, \log(\underline{\sigma}) - \log(\delta))}\right) \\ &+ \frac{4\sqrt{2}\delta}{\underline{\sigma}} \times \left(\frac{\overline{\sigma}}{\underline{\sigma}}\sqrt{\log(n)} + 2 + \frac{\overline{\sigma}}{\underline{\sigma}}\sqrt{\max(0, \log(\underline{\sigma}) - \log(\delta))}\right) \\ &\leq \frac{\sqrt{2}\delta}{c_0}\left(\sqrt{\log(n)} + \sqrt{1 + |\log(c_0)| + |\log(C_0)|} + \sqrt{|\log(\delta)|}\right) \\ &+ \frac{4\sqrt{2}\delta C_0}{c_0^2}\left(\sqrt{\log(n)} + 2 + \sqrt{|\log(c_0)| + |\log(C_0)|} + \sqrt{|\log(\delta)|}\right) \\ \leq &\left(\frac{\sqrt{2 \times (1 + |\log(c_0)| + |\log(C_0)|)}}{c_0} + \frac{4\sqrt{2}C_0}{c_0^2}(2 + \sqrt{|\log(c_0)| + |\log(C_0)|})\right) \\ &\times\delta\left(\sqrt{\log(n)} + 1 + \sqrt{|\log(\delta)|}\right) \end{aligned}$$

Choose $C = \frac{\sqrt{2\times(1+|\log(c_0)|+|\log(C_0)|)}}{c_0} + \frac{4\sqrt{2}C_0}{c_0^2}(2 + \sqrt{|\log(c_0)| + |\log(C_0)|})$, which only depends on $c_0,\ C_0$. Then

(A.9)

$$\sup_{x\in\mathbf{R}}(Prob(\max_{i=1,2,...,n} |\epsilon_i| \leq x+\delta) - Prob(\max_{i=1,2,...,n} |\epsilon_i| \leq x)) \leq 2C\delta(1 + \sqrt{\log(n)} + \sqrt{|\log(\delta)|})$$

□

LEMMA A.3. *Suppose* $\epsilon = (\epsilon_1, ...\epsilon_n)^T$ *are i.i.d. random variables with* $\mathbf{E}\epsilon_1 = 0$, $\mathbf{E}\epsilon_1^2 = \sigma^2$ *and* $\mathbf{E}|\epsilon_1|^3 < \infty$. $\Gamma = (\gamma_{ij})_{i=1,2,...,n, j=1,2,...,k}$ *is an* $n \times k$ $(1 \leq k \leq n)$ *rank* k *matrix. And* $\exists$ *constants* $0 < c_\Gamma \leq C_\Gamma < \infty$ *such that* $c_\Gamma^2 \leq \sum_{j=1}^n \gamma_{ji}^2 \leq C_\Gamma^2$ *for* $i = 1, 2, ..., k$. $\widehat{\sigma}^2 = \widehat{\sigma}^2(\epsilon)$ *is an estimator of* σ^2 *and random variables* $\epsilon^*|\epsilon = (\epsilon_1^*, ..., \epsilon_n^*)^T|\epsilon$ *are i.i.d. with* ϵ_1^* *having normal distribution* $\mathcal{N}(0, \widehat{\sigma}^2)$. $\frac{\epsilon_i^*}{\widehat{\sigma}}$ *is independent of* ϵ *for* $i = 1, 2, ..., n$. *In addition, suppose one of the following conditions:*

C1. $\exists$ *a constant* $0 < \alpha_\sigma \leq 1/2$ *such that*

(A.10)

$$|\sigma^2 - \widehat{\sigma}^2| = O_p(n^{-\alpha_\sigma}) \text{ and } \max_{j=1,2,...,n,\ i=1,2,...,k} |\gamma_{ji}| = o(\min(n^{(\alpha_\sigma-1)/2} \times \log^{-3/2}(n),\ n^{-1/3} \times \log^{-3/2}(n))$$

C2. $\exists$ *a constant* $0 < \alpha_\sigma < 1/2$ *such that*

(A.11)
$$|\sigma^2 - \widehat{\sigma}^2| = O_p(n^{-\alpha_\sigma}),\ k = o(n^{\alpha_\sigma} \times \log^{-3}(n)),\ \max_{j=1,...,n,i=1,...,k} |\gamma_{ji}| = O(n^{-\alpha_\sigma} \times \log^{-3/2}(n))$$

Then we have

(A.12)
$$\sup_{x\in[0,\infty)} |Prob(\max_{i=1,2,...,k} |\sum_{j=1}^{n} \gamma_{ji}\epsilon_j| \le x) - Prob^*(\max_{i=1,2,...,k} |\sum_{j=1}^{n} \gamma_{ji}\epsilon_j^*| \le x)| = o_P(1)$$

In particular, if $\widehat{\sigma} = \sigma$*, by assuming one of the following conditions,*
$C_1^{'}$.

$$\max_{j=1,2,...,n,i=1,2,...,k} |\gamma_{ji}| = o(n^{-1/3} \times \log^{-3/2}(n)) \tag{A.13}$$

$C_2^{'}$.

$$k \times \max_{j=1,2,...,n,i=1,2,...,k} |\gamma_{ji}| = o(\log^{-9/2}(n)) \tag{A.14}$$

Then we have

$$\sup_{x\in[0,\infty)} |Prob(\max_{i=1,2,...,k} |\sum_{j=1}^{n} \gamma_{ji}\epsilon_j| \le x) - Prob(\max_{i=1,2,...,k} |\sum_{j=1}^{n} \gamma_{ji}\epsilon_j^*| \le x)| = o(1) \tag{A.15}$$

PROOF OF LEMMA A.3. In this proof we define $\Gamma = (\gamma_1, ..., \gamma_k)$ with $\gamma_i = (\gamma_{1i}, \gamma_{2i}, ..., \gamma_{ni})^T \in \mathbf{R}^n$. For $i = 1, 2, ..., k$, $\gamma_i^T\epsilon = \sum_{j=1}^{n} \gamma_{ji}\epsilon_j$. From lemma A.2 and (8) in Chernozhukov, Chetverikov and Kato (2013), and (S1) to (S5) in Xu, Zhang and Wu (2019), for $x = (x_1, ..., x_n)$ and $y, z \in \mathbf{R}$, define

(A.16)
$$F_\beta(x) = \frac{1}{\beta}\log\left(\sum_{i=1}^{n} \exp(\beta x_i)\right),\ g_0(y) = (1 - \min(1, \max(y, 0))^4)^4,\ g_{\psi,z}(y) = g_0(\psi(y - z))$$

Here $\beta, \psi > 0$. Then $g_{\psi,z} \in \mathbf{C}^3$ is nonincreasing function. $g_0 = 1$ with $y \le 0$, 0 with $y \ge 1$, and

$$g_* = \max_{y\in\mathbf{R}}(|g_0^{'}(y)| + |g_0^{''}(y)| + |g_0^{'''}(y)|) < \infty,\ \mathbf{1}_{y\le z} \le g_{\psi,z}(y) \le \mathbf{1}_{y \le z+\psi^{-1}}$$

$$\sup_{y,z\in\mathbf{R}} |g_{\psi,z}^{'}(y)| \le g_*\psi,\ \sup_{y,z\in\mathbf{R}} |g_{\psi,z}^{''}(y)| \le g_*\psi^2,\ \sup_{y,z\in\mathbf{R}} |g_{\psi,z}^{'''}(y)| \le g_*\psi^3$$

$$\begin{gathered} \frac{\partial F_\beta}{\partial x_i} = \frac{\exp(\beta x_i)}{\sum_{j=1}^{n}\exp(\beta x_j)} \Rightarrow \frac{\partial F_\beta}{\partial x_i} \ge 0,\ \sum_{i=1}^{n} \frac{\partial F_\beta}{\partial x_i} = 1 \\ \sum_{i=1}^{n}\sum_{j=1}^{n} |\frac{\partial^2 F_\beta}{\partial x_i \partial x_j}| \le 2\beta,\ \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} |\frac{\partial^3 F_\beta}{\partial x_i \partial x_j \partial x_k}| \le 6\beta^2 \\ F_\beta(x_1, ..., x_n) - \frac{\log(n)}{\beta} \le \max_{i=1,...,n} x_i \le F_\beta(x_1, ..., x_n) \end{gathered} \tag{A.17}$$

For any given $x = (x_1, ..., x_n) \in \mathbf{R}^n$, define function

$G_\beta(x) = \frac{1}{\beta}\log(\sum_{i=1}^n \exp(\beta x_i) + \sum_{i=1}^n \exp(-\beta x_i)) = F_\beta(x_1,...,x_n,-x_1,...,-x_n)$. From (A.17) and (A.6), for $i,j,k = 1,...,n$

(A.18)

$$G_\beta(x) - \frac{\log(2n)}{\beta} \le \max_{i=1,...,n} |x_i| \le G_\beta(x)$$

$$\frac{\partial G_\beta}{\partial x_i} = \frac{\partial F_\beta}{\partial x_i} - \frac{\partial F_\beta}{\partial x_{i+n}} \Rightarrow \sum_{i=1}^n |\frac{\partial G_\beta}{\partial x_i}| \le \sum_{i=1}^n \frac{\partial F_\beta}{\partial x_i} + \frac{\partial F_\beta}{\partial x_{i+n}} = 1$$

$$\frac{\partial^2 G_\beta}{\partial x_i \partial x_j} = \frac{\partial^2 F_\beta}{\partial x_i \partial x_j} - \frac{\partial^2 F_\beta}{\partial x_i \partial x_{j+n}} - \frac{\partial^2 F_\beta}{\partial x_{i+n} \partial x_j} + \frac{\partial^2 F_\beta}{\partial x_{i+n} \partial x_{j+n}}$$

$$\Rightarrow \sum_{i=1}^n \sum_{j=1}^n |\frac{\partial^2 G_\beta}{\partial x_i \partial x_j}| \le \sum_{i=1}^{2n} \sum_{j=1}^{2n} |\frac{\partial^2 F_\beta}{\partial x_i \partial x_j}| \le 2\beta$$

$$\frac{\partial^3 G_\beta}{\partial x_i \partial x_j \partial x_k} = \frac{\partial^3 F_\beta}{\partial x_i \partial x_j \partial x_k} - \frac{\partial^3 F_\beta}{\partial x_i \partial x_j \partial x_{k+n}} - \frac{\partial^3 F_\beta}{\partial x_i \partial x_{j+n} \partial x_k} + \frac{\partial^3 F_\beta}{\partial x_i \partial x_{j+n} \partial x_{k+n}} - \frac{\partial^3 F_\beta}{\partial x_{i+n} \partial x_j \partial x_k}$$

$$+ \frac{\partial^3 F_\beta}{\partial x_{i+n} \partial x_j \partial x_{k+n}} + \frac{\partial^3 F_\beta}{\partial x_{i+n} \partial x_{j+n} \partial x_k} - \frac{\partial^3 F_\beta}{\partial x_{i+n} \partial x_{j+n} \partial x_{k+n}}$$

$$\Rightarrow \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n |\frac{\partial^3 G_\beta}{\partial x_i \partial x_j \partial x_k}| \le \sum_{i=1}^{2n} \sum_{j=1}^{2n} \sum_{k=1}^{2n} |\frac{\partial^3 F_\beta}{\partial x_i \partial x_j \partial x_k}| \le 6\beta^2$$

Define $h_{\beta,\psi,x}(x_1,...,x_n) = g_{\psi,x}(G_\beta(x_1,...,x_n))$. Direct calculation shows $\frac{\partial h_{\beta,\psi,x}(x_1,...,x_n)}{\partial x_i} = g'_{\psi,x}(G_\beta(x_1,...,x_n))\frac{\partial G_\beta}{\partial x_i} \Rightarrow \sum_{i=1}^n |\frac{\partial h_{\beta,\psi,x}(x_1,...,x_n)}{\partial x_i}| \le g_*\psi$;

(A.19)

$$\frac{\partial^2 h_{\beta,\psi,x}(x_1,...,x_n)}{\partial x_i \partial x_j} = g''_{\psi,x}(G_\beta(x_1,...,x_n))\frac{\partial G_\beta}{\partial x_i}\frac{\partial G_\beta}{\partial x_j} + g'_{\psi,x}(G_\beta(x_1,...,x_n))\frac{\partial^2 G_\beta}{\partial x_i \partial x_j}$$

$$\Rightarrow \sum_{i=1}^n \sum_{j=1}^n |\frac{\partial^2 h_{\beta,\psi,x}(x_1,...,x_n)}{\partial x_i \partial x_j}| \le g_*\psi^2 \left(\sum_{i=1}^n |\frac{\partial G_\beta}{\partial x_i}|\right)^2 + g_*\psi \sum_{i=1}^n \sum_{j=1}^n |\frac{\partial^2 G_\beta}{\partial x_i \partial x_j}| \le g_*\psi^2 + 2g_*\psi\beta$$

$$\text{and } \frac{\partial^3 h_{\beta,\psi,x}(x_1,...,x_n)}{\partial x_i \partial x_j \partial x_k} = g'''_{\psi,x}(G_\beta(x_1,...,x_n))\frac{\partial G_\beta}{\partial x_i}\frac{\partial G_\beta}{\partial x_j}\frac{\partial G_\beta}{\partial x_k} + g''_{\psi,x}(G_\beta(x_1,...,x_n))\frac{\partial^2 G_\beta}{\partial x_i \partial x_k}\frac{\partial G_\beta}{\partial x_j}$$

$$+ g''_{\psi,x}(G_\beta(x_1,...,x_n))\frac{\partial G_\beta}{\partial x_i}\frac{\partial^2 G_\beta}{\partial x_j \partial x_k} + g''_{\psi,x}(G_\beta(x_1,...,x_n))\frac{\partial^2 G_\beta}{\partial x_i \partial x_j}\frac{\partial G_\beta}{\partial x_k} + g'_{\psi,x}(G_\beta(x_1,...,x_n))\frac{\partial^3 G_\beta}{\partial x_i \partial x_j \partial x_k}$$

$$\Rightarrow \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n |\frac{\partial^3 h_{\beta,\psi,x}(x_1,...,x_n)}{\partial x_i \partial x_j \partial x_k}| \le g_*\psi^3 \left(\sum_{i=1}^n |\frac{\partial G_\beta}{\partial x_i}|\right)^3 + 3g_*\psi^2 \left(\sum_{i=1}^n \sum_{j=1}^n |\frac{\partial^2 G_\beta}{\partial x_i \partial x_j}|\right) \times \left(\sum_{k=1}^n |\frac{\partial G_\beta}{\partial x_k}|\right)$$

$$+ g_*\psi \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n |\frac{\partial^3 G_\beta}{\partial x_i \partial x_j \partial x_k}| \le g_*\psi^3 + 6g_*\psi^2\beta + 6g_*\psi\beta^2$$

Define $\xi = (\xi_1,...,\xi_n)$ as i.i.d. random variables with the same marginal distribution as ϵ_1, and is independent of ϵ, ϵ^*. Therefore, $Prob(\max_{i=1,2,...,k} |\gamma_i^T \epsilon| \le x) = Prob^*(\max_{i=1,2,...,k} |\gamma_i^T \xi| \le x)$ for any x. Since $c_\Gamma^2 \le \mathbf{E}^* \left(\sum_{l=1}^n \frac{\gamma_{il}\epsilon_l^*}{\widehat{\sigma}}\right)^2 = \sum_{l=1}^n \gamma_{il}^2 \le C_\Gamma^2$ for $i = 1,2,...,k$. According to (A.6), (A.18) and lemma A.2, $\exists$ a constant C which only depends on c_Γ and C_Γ such that for any given $\psi, \beta, \widehat{\sigma} > 0$,

(A.20)

$$\sup_{x\in\mathbf{R}}\left(Prob^*\left(\max_{i=1,2,...,k}|\gamma_i^T\epsilon^*|\le x+\frac{1}{\psi}+\frac{\log(2k)}{\beta}\right)-Prob^*\left(\max_{i=1,2,...,k}|\gamma_i^T\epsilon^*|\le x\right)\right)$$

$$=\sup_{x\in\mathbf{R}}\left(Prob^*\left(\max_{i=1,2,...,k}|\frac{\gamma_i^T\epsilon^*}{\widehat{\sigma}}|\le \frac{x}{\widehat{\sigma}}+\frac{1}{\psi\widehat{\sigma}}+\frac{\log(2k)}{\beta\widehat{\sigma}}\right)-Prob^*\left(\max_{i=1,2,...,k}|\frac{\gamma_i^T\epsilon^*}{\widehat{\sigma}}|\le \frac{x}{\widehat{\sigma}}\right)\right)$$

$$\le C\times\left(\frac{1}{\psi\widehat{\sigma}}+\frac{\log(2k)}{\beta\widehat{\sigma}}\right)\times\left(1+\sqrt{\log(k)}+\sqrt{|\log\left(\frac{1}{\psi\widehat{\sigma}}+\frac{\log(2k)}{\beta\widehat{\sigma}}\right)|}\right)$$

Define $z=C\times\left(\frac{1}{\psi\widehat{\sigma}}+\frac{\log(2k)}{\beta\widehat{\sigma}}\right)\times\left(1+\sqrt{\log(k)}+\sqrt{|\log\left(\frac{1}{\psi\widehat{\sigma}}+\frac{\log(2k)}{\beta\widehat{\sigma}}\right)|}\right)$. For any $x\ge 0$,

(A.21)

$$Prob(\max_{i=1,2,...,k}|\gamma_i^T\epsilon|\le x)-Prob^*(\max_{i=1,2,...,k}|\gamma_i^T\epsilon^*|\le x)$$

$$\le Prob^*(\max_{i=1,2,...,k}|\gamma_i^T\xi|\le x)-Prob^*(\max_{i=1,2,...,k}|\gamma_i^T\epsilon^*|\le x+\frac{1}{\psi}+\frac{\log(2k)}{\beta})+z$$

$$\le Prob^*(G_\beta(\gamma_1^T\xi,...,\gamma_k^T\xi)\le x+\frac{\log(2k)}{\beta})-Prob^*(G_\beta(\gamma_1^T\epsilon^*,...,\gamma_k^T\epsilon^*)\le x+\frac{1}{\psi}+\frac{\log(2k)}{\beta})+z$$

$$\le \mathbf{E}^*h_{\beta,\psi,x+\frac{\log(2k)}{\beta}}(\gamma_1^T\xi,...,\gamma_k^T\xi)-\mathbf{E}^*h_{\beta,\psi,x+\frac{\log(2k)}{\beta}}(\gamma_1^T\epsilon^*,...,\gamma_k^T\epsilon^*)+z$$

$$Prob(\max_{i=1,2,...,k}|\gamma_i^T\epsilon|\le x)-Prob^*(\max_{i=1,2,...,k}|\gamma_i^T\epsilon^*|\le x)$$

$$\ge Prob^*(\max_{i=1,2,...,k}|\gamma_i^T\xi|\le x)-Prob^*(\max_{i=1,2,...,k}|\gamma_i^T\epsilon^*|\le x-\frac{1}{\psi}-\frac{\log(2k)}{\beta})-z$$

$$\ge Prob^*(G_\beta(\gamma_1^T\xi,...,\gamma_k^T\xi)\le x)-Prob^*(G_\beta(\gamma_1^T\epsilon^*,...,\gamma_k^T\epsilon^*)\le x-\frac{1}{\psi})-z$$

$$\ge \mathbf{E}^*h_{\beta,\psi,x-\frac{1}{\psi}}(\gamma_1^T\xi,...,\gamma_k^T\xi)-\mathbf{E}^*h_{\beta,\psi,x-\frac{1}{\psi}}(\gamma_1^T\epsilon^*,...,\gamma_k^T\epsilon^*)-z$$

Therefore, we have

(A.22)

$$\begin{aligned}&\sup_{x\in[0,\infty)}|Prob(\max_{i=1,2,...,k}|\gamma_i^T\epsilon|\le x)-Prob^*(\max_{i=1,2,...,k}|\gamma_i^T\epsilon^*|\le x)|\\&\le z+\sup_{x\in\mathbf{R}}|\mathbf{E}^*h_{\beta,\psi,x}(\gamma_1^T\xi,...,\gamma_k^T\xi)-\mathbf{E}^*h_{\beta,\psi,x}(\gamma_1^T\epsilon^*,...,\gamma_k^T\epsilon^*)|\end{aligned}$$

For any $i=1,2,...,k, j=1,2,...,n$, define $H_{ij}=\sum_{s=1}^{j-1}\gamma_{si}\xi_s+\sum_{s=j+1}^{n}\gamma_{si}\epsilon_s^*$, $m_{ij}=\gamma_{ji}\xi_j$ and $m_{ij}^*=\gamma_{ji}\epsilon_j^*$, we have $H_{ij}+m_{ij}=H_{ij+1}+m_{ij+1}^*$, and

(A.23)

$$\sup_{x\in\mathbf{R}}|\mathbf{E}^*h_{\beta,\psi,x}(\gamma_1^T\xi,...,\gamma_k^T\xi)-\mathbf{E}^*h_{\beta,\psi,x}(\gamma_1^T\epsilon^*,...,\gamma_k^T\epsilon^*)|$$

$$=\sup_{x\in\mathbf{R}}|\sum_{s=1}^n\mathbf{E}^*h_{\beta,\psi,x}(H_{1s}+m_{1s},...,H_{ks}+m_{ks})-\mathbf{E}^*h_{\beta,\psi,x}(H_{1s}+m_{1s}^*,...,H_{ks}+m_{ks}^*)|$$

$$\le\sum_{s=1}^n\sup_{x\in\mathbf{R}}|\mathbf{E}^*h_{\beta,\psi,x}(H_{1s}+m_{1s},...,H_{ks}+m_{ks})-\mathbf{E}^*h_{\beta,\psi,x}(H_{1s}+m_{1s}^*,...,H_{ks}+m_{ks}^*)|$$

Since $\mathbf{E}(\xi_s|\epsilon,\xi_b,\epsilon_b^*,b\neq s)=\mathbf{E}(\epsilon_s^*|\epsilon,\xi_b,\epsilon_b^*,b\neq s)=0$, $\mathbf{E}(\xi_s^2-\epsilon_s^{*2}|\epsilon,\xi_b,\epsilon_b^*,b\neq s)=\sigma^2-\widehat{\sigma}^2$, from multivariate Taylor's theorem and (A.19), for any $s=1,2,...,n$ and $x\in\mathbf{R}$,
(A.24)

$$
\begin{aligned}
&|\mathbf{E}\left(h_{\beta,\psi,x}(H_{1s}+m_{1s},...,H_{ks}+m_{ks})-h_{\beta,\psi,x}(H_{1s}+m_{1s}^*,...,H_{ks}+m_{ks}^*)\right)\Big|\epsilon,\xi_b,\epsilon_b^*,b\neq s| \\
&\leq |\sum_{i=1}^{k}\frac{\partial h_{\beta,\psi,x}(H_{1s},...,H_{ks})}{\partial x_i}\gamma_{si}\mathbf{E}(\xi_s-\epsilon_s^*|\epsilon,\xi_b,\epsilon_b^*,b\neq s)| \\
&+\frac{1}{2}|\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\partial^2 h_{\beta,\psi,x}(H_{1s},...,H_{ks})}{\partial x_i\partial x_j}\gamma_{si}\gamma_{sj}\mathbf{E}(\xi_s^2-\epsilon_s^{*2}|\epsilon,\xi_b,\epsilon_b^*,b\neq s)| \\
&+(g_*\psi^3+g_*\psi^2\beta+g_*\psi\beta^2)\max_{i=1,2,...,k}|\gamma_{si}|^3\times(\mathbf{E}|\epsilon_1|^3+\widehat{\sigma}^3D)
\end{aligned}
$$

$$
\begin{aligned}
&\Rightarrow \sup_{x\in\mathbf{R}}|\mathbf{E}h_{\beta,\psi,x}(H_{1s}+m_{1s},...,H_{ks}+m_{ks})-h_{\beta,\psi,x}(H_{1s}+m_{1s}^*,...,H_{ks}+m_{ks}^*)|\epsilon,\xi_b,\epsilon_b^*,b\neq s| \\
&\leq g_*(\psi^2+\psi\beta)|\sigma^2-\widehat{\sigma}^2|\times\max_{i=1,...,k}\gamma_{si}^2+(\mathbf{E}|\epsilon_1|^3+\widehat{\sigma}^3D)\times g_*(\psi^3+\psi^2\beta+\psi\beta^2)\times\max_{i=1,...,k}|\gamma_{si}|^3
\end{aligned}
$$

Here $D=\mathbf{E}|Y|^3$ with Y having normal distribution with mean 0 and variance 1. Then

$$
\begin{aligned}
&\sup_{x\in[0,\infty)}|Prob(\max_{i=1,2,...,k}|\gamma_i^T\epsilon|\leq x)-Prob^*(\max_{i=1,2,...,k}|\gamma_i^T\epsilon^*|\leq x)| \\
&\leq z+(g_*\psi^2+g_*\psi\beta)|\sigma^2-\widehat{\sigma}^2|\times\sum_{s=1}^{n}\max_{i=1,...,k}\gamma_{si}^2 \qquad \text{(A.25)} \\
&+(\mathbf{E}|\epsilon_1|^3+\widehat{\sigma}^3D)\times g_*(\psi^3+\psi^2\beta+\psi\beta^2)\times\sum_{s=1}^{n}\max_{i=1,...,k}|\gamma_{si}|^3
\end{aligned}
$$

In particular, for any given $\delta>0$, choose $\psi=\beta=\log^{3/2}(n)/\delta^{1/4}$ and suppose $\frac{3\sigma}{2}>\widehat{\sigma}>\frac{\sigma}{2}$. For sufficiently large n we have $\frac{1}{\psi\widehat{\sigma}}+\frac{\log(2k)}{\beta\widehat{\sigma}}\leq\frac{4\log(n)}{\psi\sigma}\leq\frac{4\delta^{1/4}}{\sigma\sqrt{\log(n)}}<1$ and
(A.26)

$$
z\leq\frac{4C\log(n)}{\psi\sigma}\times\left(2\sqrt{\log(n)}+\sqrt{\log(\psi\widehat{\sigma})}\right)\leq\frac{4C\delta^{1/4}}{\sigma}\left(2+\sqrt{\frac{\frac{3}{2}\log(\log(n))+\log(3\sigma/2\delta^{1/4})}{\log(n)}}\right)\leq C'\delta^{1/4}
$$

Here $C'=\frac{12C}{\sigma}$.

Suppose condition C1. For any $1>\delta>0$, $\exists D_\delta>0$ such that for sufficiently large n,
(A.27)

$$
\begin{aligned}
&Prob\left(|\sigma^2-\widehat{\sigma}^2|\leq D_\delta\times n^{-\alpha_\sigma}\right)>1-\delta, \quad \max_{j=1,2,...,n,i=1,2,...,k}|\gamma_{ji}|<\delta\times n^{(\alpha_\sigma-1)/2}\times\log^{-3/2}(n), \\
&\text{and} \max_{j=1,2,...,n,i=1,2,...,k}|\gamma_{ji}|<\delta\times n^{-1/3}\times\log^{-3/2}(n)
\end{aligned}
$$

Choose $\psi = \beta = \log^{3/2}(n)/\delta^{1/4}$. According to (A.25), for sufficiently large n, (A.27) happens and $\frac{1}{2}\sigma < \widehat{\sigma} < \frac{3}{2}\sigma$ with probability $1-\delta$. If (A.27) happens,

(A.28)
$$\sup_{x\in[0,\infty)} |Prob(\max_{i=1,2,\dots,k} |\gamma_i^T\epsilon| \le x) - Prob^*(\max_{i=1,2,\dots,k} |\gamma_i^T\epsilon^*| \le x)|$$
$$\le C^{'}\delta^{1/4} + 2g_*\psi^2 \times D_\delta \times n^{-\alpha_\sigma} \times \frac{\delta^2 \times n^{\alpha_\sigma}}{\log^3(n)} + (\mathbf{E}|\epsilon_1|^3 + \frac{27D}{8}\sigma^3) \times 3g_*\psi^3 \times \delta^3 \times n \times \frac{1}{n\log^{9/2}(n)}$$
$$= C^{'}\delta^{1/4} + 2g_* D_\delta \delta^{3/2} + 3g_*(\mathbf{E}|\epsilon_1|^3 + \frac{27D}{8}\sigma^3) \times \delta^{9/4}$$

For $\delta > 0$ can be arbitrarily small, we prove (A.12).

Suppose condition C2. For any $\delta > 0$, there exists $D_\delta > 0$ such that for sufficiently large n

(A.29)
$$Prob\left(|\sigma^2 - \widehat{\sigma}^2| \le D_\delta \times n^{-\alpha_\sigma}\right) \ge 1-\delta, \quad k \le \frac{\delta n^{\alpha_\sigma}}{\log^3(n)}, \quad \max_{i=1,2,\dots,k} \sum_{j=1}^n \gamma_{ji}^2 \le D_\delta$$
$$\text{and} \max_{j=1,2,\dots,n,i=1,2,\dots,k} |\gamma_{ji}| \le \frac{D_\delta \times n^{-\alpha_\sigma}}{\log^{3/2}(n)}$$

Since

(A.30)
$$\sum_{j=1}^n \max_{i=1,\dots,k} \gamma_{ji}^2 \le \sum_{j=1}^n \sum_{i=1}^k \gamma_{ji}^2 \le kD_\delta$$
$$\sum_{j=1}^n \max_{i=1,\dots,k} \gamma_{ji}^3 \le \max_{j=1,2,\dots,n,i=1,2,\dots,k} |\gamma_{ji}| \times \sum_{j=1}^n \max_{i=1,\dots,k} \gamma_{ji}^2 \le kD_\delta \times \max_{j=1,2,\dots,n,i=1,2,\dots,k} |\gamma_{ji}|$$

If (A.29) happens, by choosing $\psi = \beta = \log^{3/2}(n)/\delta^{1/4}$

(A.31)
$$\sup_{x\in[0,\infty)} |Prob(\max_{i=1,2,\dots,k} |\gamma_i^T\epsilon| \le x) - Prob^*(\max_{i=1,2,\dots,k} |\gamma_i^T\epsilon^*| \le x)|$$
$$\le C^{'}\delta^{1/4} + 2g_*\psi^2 D_\delta n^{-\alpha_\sigma} \times kD_\delta + (\mathbf{E}|\epsilon_1|^3 + \frac{27D}{8}\sigma^3) \times 3g_*\psi^3 \times kD_\delta \max_{j=1,2,\dots,n,i=1,2,\dots,k} |\gamma_{ji}|$$
$$\le C^{'}\delta^{1/4} + 2g_* D_\delta^2 \times \frac{\log^3(n)}{\delta^{1/2}} \times \frac{\delta n^{\alpha_\sigma}}{\log^3(n)} \times n^{-\alpha_\sigma}$$
$$+3(\mathbf{E}|\epsilon_1|^3 + \frac{27D}{8}\sigma^3) g_* D_\delta^2 \times \frac{\log^{9/2}(n)}{\delta^{3/4}} \times \frac{\delta n^{\alpha_\sigma}}{\log^3(n)} \times \frac{n^{-\alpha_\sigma}}{\log^{3/2}(n)}$$
$$= C^{'}\delta^{1/4} + 2g_* D_\delta^2 \delta^{1/2} + 3(\mathbf{E}|\epsilon_1|^3 + \frac{27D}{8}\sigma^3) g_* D_\delta^2 \times \delta^{1/4}$$

and we prove (A.12).

If $\widehat{\sigma} = \sigma$. We choose $\psi = \beta = \log^{3/2}(n)/\delta^{1/4}$, (A.25) can be modified to

(A.32)
$$\sup_{x\in[0,\infty)} |Prob(\max_{i=1,2,\dots,k} |\gamma_i^T\epsilon| \le x) - Prob(\max_{i=1,2,\dots,k} |\gamma_i^T\epsilon^*| \le x)|$$
$$\le C^{'}\delta^{1/4} + (\mathbf{E}|\epsilon_1|^3 + D\sigma^3) g_*\psi(\psi^2 + \psi\beta + \beta^2) \sum_{s=1}^n \max_{i=1,\dots,k} |\gamma_{si}|^3$$

Suppose condition $C1^{'}$. For any $\delta > 0$ and sufficiently large n, $\max_{j=1,2,\ldots,n,i=1,2,\ldots,k} |\gamma_{ji}| \leq \delta \times n^{-1/3} \log^{-3/2}(n)$,

$$
\begin{aligned}
\text{(A.33)} \quad & \sup_{x\in[0,\infty)} |Prob(\max_{i=1,2,\ldots,k} |\sum_{j=1}^{n} \gamma_{ji}\epsilon_j| \leq x) - Prob(\max_{i=1,2,\ldots,k} |\sum_{j=1}^{n} \gamma_{ji}\epsilon_j^*| \leq x)| \\
& \leq C^{'}\delta^{1/4} + 3(\mathbf{E}|\epsilon_1|^3 + D\sigma^3)g_* \times \delta^{9/4}
\end{aligned}
$$

and we prove (A.15).

Suppose condition $C2^{'}$. For any $\delta > 0$ and sufficiently large n, $k \times \max_{j=1,2,\ldots,n,i=1,2,\ldots,k} |\gamma_{ji}| \leq \delta \log^{-9/2}(n)$. According to (A.30), for sufficiently large n we have

$$
\begin{aligned}
\text{(A.34)} \quad & \sup_{x\in[0,\infty)} |Prob(\max_{i=1,2,\ldots,k} |\sum_{j=1}^{n} \gamma_{ji}\epsilon_j| \leq x) - Prob(\max_{i=1,2,\ldots,k} |\sum_{j=1}^{n} \gamma_{ji}\epsilon_j^*| \leq x)| \\
& \leq C^{'}\delta^{1/4} + 3(\mathbf{E}|\epsilon_1|^3 + D\sigma^3)g_* D_\delta \times \delta^{1/4}
\end{aligned}
$$

and we prove (A.15). □

Condition C1 implies $C1^{'}$, and condition C2 implies $C2^{'}$. The additional proportions in C1 and C2 accommodate the error introduced in estimating errors' variance σ^2. Condition C1 is designed for the situation when the number of linear combinations k is as large as the sample size n; and condition C2 can be used when k is significantly smaller than n.

The difference between lemma A.3 and the classical central limit theorem is that k *can grow as* n *increases*. The maximum $\max_{i=1,2,\ldots,k} |\sum_{j=1}^{n} \gamma_{ji}\epsilon_j|$ does not have an asymptotic distribution if $k \to \infty$. However, if the random variables are mixed well, approximating the distribution of $\max_{i=1,2,\ldots,k} |\sum_{j=1}^{n} \gamma_{ji}\epsilon_j|$ by the distribution of the maximum of normal random variables is still applicable. With the help of lemma A.3, we can establish the normal approximation theorem and construct the simultaneous confidence region for $\widehat{\gamma}$(defined in (17)).

APPENDIX B: PROOFS OF THEOREMS IN SECTION 3

This section applies notations in section 2.

PROOF OF THEOREM 1. From (16),

(B.1)

$$
Prob\left(\widehat{\mathcal{N}}_{b_n} \neq \mathcal{N}_{b_n}\right) \leq Prob\left(\min_{i\in\mathcal{N}_{b_n}} |\widetilde{\theta}_i| \leq b_n\right) + Prob\left(\max_{i\notin\mathcal{N}_{b_n}} |\widetilde{\theta}_i| > b_n\right)
$$

$$
\leq Prob\left(\min_{i\in\mathcal{N}_{b_n}} |\theta_i| - \max_{i\in\mathcal{N}_{b_n}} \rho_n^2 |\sum_{j=1}^{r} \frac{q_{ij}\zeta_j}{(\lambda_j^2+\rho_n)^2}| - \max_{i\in\mathcal{N}_{b_n}} |\sum_{j=1}^{r} q_{ij}\left(\frac{\lambda_j}{\lambda_j^2+\rho_n} + \frac{\rho_n\lambda_j}{(\lambda_j^2+\rho_n)^2}\right)\sum_{l=1}^{n} p_{lj}\epsilon_l| \leq b_n\right)
$$

$$
+Prob\left(\max_{i\notin\mathcal{N}_{b_n}} |\theta_i| + \max_{i\notin\mathcal{N}_{b_n}} \rho_n^2 |\sum_{j=1}^{r} \frac{q_{ij}\zeta_j}{(\lambda_j^2+\rho_n)^2}| + \max_{i\notin\mathcal{N}_{b_n}} |\sum_{j=1}^{r} q_{ij}\left(\frac{\lambda_j}{\lambda_j^2+\rho_n} + \frac{\rho_n\lambda_j}{(\lambda_j^2+\rho_n)^2}\right)\sum_{l=1}^{n} p_{lj}\epsilon_l| > b_n\right)
$$

From Cauchy inequality,
(B.2)

$$\max_{i=1,2,\ldots,p}\rho_n^2|\sum_{j=1}^{r}\frac{q_{ij}\zeta_j}{(\lambda_j^2+\rho_n)^2}|\leq\max_{i=1,2,\ldots,p}\rho_n^2\sqrt{\sum_{j=1}^{r}q_{ij}^2}\times\sqrt{\sum_{j=1}^{r}\frac{\zeta_j^2}{(\lambda_j^2+\rho_n)^4}}=O(n^{\alpha_\theta-2\delta})$$

$$\max_{i=1,2,\ldots,p}\sum_{l=1}^{n}\left(\sum_{j=1}^{r}q_{ij}\left(\frac{\lambda_j}{\lambda_j^2+\rho_n}+\frac{\rho_n\lambda_j}{(\lambda_j^2+\rho_n)^2}\right)p_{lj}\right)^2=\max_{i=1,2,\ldots,p}\sum_{j=1}^{r}q_{ij}^2\left(\frac{\lambda_j}{\lambda_j^2+\rho_n}+\frac{\rho_n\lambda_j}{(\lambda_j^2+\rho_n)^2}\right)^2$$
$$\leq\max_{i=1,2,\ldots,p}\frac{4\sum_{j=1}^{r}q_{ij}^2}{\lambda_r^2}$$

Therefore, for sufficiently large n, from assumption 4 and lemma A.1
(B.3)

$$\min_{i\in\mathcal{N}_{b_n}}|\theta_i|-\max_{i\in\mathcal{N}_{b_n}}\rho_n^2|\sum_{j=1}^{r}\frac{q_{ij}\zeta_j}{(\lambda_j^2+\rho_n)^2}|-b_n>\frac{1}{2}(\frac{1}{c_b}-1)b_n$$

$$b_n-\max_{i\notin\mathcal{N}_{b_n}}|\theta_i|-\max_{i\in\mathcal{N}_{b_n}}\rho_n^2|\sum_{j=1}^{r}\frac{q_{ij}\zeta_j}{(\lambda_j^2+\rho_n)^2}|>\frac{1}{2}(1-c_b)b_n$$

$$\Rightarrow Prob\left(\widehat{\mathcal{N}}_{b_n}\neq\mathcal{N}_{b_n}\right)\leq\frac{|\mathcal{N}_{b_n}|\times E\times 2^m}{\lambda_r^m\times(\frac{1}{2}(\frac{1}{c_b}-1)b_n)^m}+\frac{(p-|\mathcal{N}_{b_n}|)\times E\times 2^m}{\lambda_r^m\times(\frac{1}{2}(1-c_b)b_n)^m}=O(n^{\alpha_p+m\nu_b-m\eta})$$

and we prove (21).

Define $\widehat{\gamma}=M\widehat{\theta}=(\widehat{\gamma}_1,...,\widehat{\gamma}_{p_1})^T$ and $\gamma=M\beta=(\gamma_1,...,\gamma_{p_1})^T$. For $\beta=\theta+\theta_\perp$, if $\widehat{\mathcal{N}}_{b_n}=\mathcal{N}_{b_n}$, (16) and (4) imply
(B.4)

$$\max_{i=1,2,\ldots,p_1}|\widehat{\gamma}_i-\gamma_i|=\max_{i=1,2,\ldots,p_1}|\sum_{j\in\mathcal{N}_{b_n}}m_{ij}\widetilde{\theta}_j-\sum_{j\in\mathcal{N}_{b_n}}m_{ij}\theta_j-\sum_{j\notin\mathcal{N}_{b_n}}m_{ij}\theta_j-\sum_{j=1}^{p}m_{ij}\theta_{\perp,j}|$$
$$\leq\max_{i=1,2,\ldots,p_1}\rho_n^2|\sum_{k=1}^{r}\frac{c_{ik}\zeta_k}{(\lambda_k^2+\rho_n)^2}|+\max_{i=1,2,\ldots,p_1}|\sum_{k=1}^{r}c_{ik}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda^2+\rho_n)^2}\right)\sum_{l=1}^{n}p_{lk}\epsilon_l|$$
$$+\max_{i=1,2,\ldots,p_1}|\sum_{j\notin\mathcal{N}_{b_n}}m_{ij}\theta_j|+\max_{i=1,2,\ldots,p_1}|\sum_{j=1}^{p}m_{ij}\theta_{\perp,j}|$$

From (4) and assumption 5, if $i \notin \mathcal{M}$, then $c_{ik} = 0$ for $k = 1, 2, ..., r$, so from Cauchy inequality and lemma A.1,
(B.5)

$$\max_{i=1,2,\ldots,p_1} \rho_n^2 \Big| \sum_{k=1}^r \frac{c_{ik}\zeta_k}{(\lambda_k^2+\rho_n)^2} \Big| \le \max_{i\in\mathcal{M}} \rho_n^2 \sqrt{\sum_{k=1}^r c_{ik}^2} \times \sqrt{\sum_{k=1}^r \frac{\zeta_k^2}{(\lambda_k^2+\rho_n)^4}} \le \sqrt{C_{\mathcal{M}}}\rho_n^2 \times \frac{\|\theta\|_2}{\lambda_r^4} = O(n^{\alpha_\theta - 2\delta})$$

$$\max_{i\in\mathcal{M}} \sum_{l=1}^n \left(\sum_{k=1}^r c_{ik} p_{lk} \left(\frac{\lambda_k}{\lambda_k^2+\rho_n} + \frac{\rho_n \lambda_k}{(\lambda_k^2+\rho_n)^2} \right) \right)^2 = \max_{i\in\mathcal{M}} \sum_{k=1}^r c_{ik}^2 \left(\frac{\lambda_k}{\lambda_k^2+\rho_n} + \frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2} \right)^2 \le \frac{4C_{\mathcal{M}}}{\lambda_r^2}$$

$$\Rightarrow Prob\left(\max_{i=1,2,\ldots,p_1} \Big| \sum_{k=1}^r c_{ik} \left(\frac{\lambda_k}{\lambda_k^2+\rho_n} + \frac{\rho_n\lambda_k}{(\lambda^2+\rho_n)^2} \right) \sum_{l=1}^n p_{lk}\epsilon_l \Big| > \delta \right) \le \frac{|\mathcal{M}| \times E \times 2^m C_{\mathcal{M}}^{m/2}}{\lambda_r^m \delta^m} \quad \text{for } \forall \delta > 0$$

$$\Rightarrow \max_{i=1,2,\ldots,p_1} \Big| \sum_{k=1}^r c_{ik} \left(\frac{\lambda_k}{\lambda_k^2+\rho_n} + \frac{\rho_n\lambda_k}{(\lambda^2+\rho_n)^2} \right) \sum_{l=1}^n p_{lk}\epsilon_l \Big| = O_p(|\mathcal{M}|^{1/m} \times n^{-\eta})$$

Here E is the constant defined in lemma A.1. Combine with assumption 2, assumption 5, and (B.3), we prove (22).

If $\widehat{\mathcal{N}}_{b_n} = \mathcal{N}_{b_n}$, since $X\beta = X\theta$, we have
(B.6)

$$\widehat{\sigma}^2 - \sigma^2 = \frac{1}{n}\sum_{i=1}^n \left(\epsilon_i - \sum_{j\in\mathcal{N}_{b_n}} x_{ij}(\widetilde{\theta}_j - \theta_j) + \sum_{j\notin\mathcal{N}_{b_n}} x_{ij}\theta_j \right)^2 - \sigma^2$$

$$= \frac{1}{n}\sum_{i=1}^n \epsilon_i^2 - \sigma^2 + \frac{1}{n}\sum_{i=1}^n \left(\sum_{j\in\mathcal{N}_{b_n}} x_{ij}(\widetilde{\theta}_j - \theta_j) \right)^2 + \frac{1}{n}\sum_{i=1}^n \left(\sum_{j\notin\mathcal{N}_{b_n}} x_{ij}\theta_j \right)^2 - \frac{2}{n}\sum_{i=1}^n \sum_{j\in\mathcal{N}_{b_n}} \epsilon_i x_{ij}(\widetilde{\theta}_j - \theta_j)$$

$$+ \frac{2}{n}\sum_{i=1}^n \sum_{j\notin\mathcal{N}_{b_n}} \epsilon_i x_{ij}\theta_j - \frac{2}{n}\sum_{i=1}^n \left(\sum_{j\in\mathcal{N}_{b_n}} x_{ij}(\widetilde{\theta}_j - \theta_j) \right) \times \left(\sum_{j\notin\mathcal{N}_{b_n}} x_{ij}\theta_j \right)$$

From assumption 3, $\mathbf{E}\left(\frac{1}{n}\sum_{i=1}^n \epsilon_i^2 - \sigma^2\right)^2 \le \frac{2}{n}(\mathbf{E}\epsilon_1^4 + \sigma^4) = O(1/n) \Rightarrow \frac{1}{n}\sum_{i=1}^n \epsilon_i^2 - \sigma^2 = O_p(1/\sqrt{n})$. For the second term, from assumption 1 and (B.2),
(B.7)

$$\frac{1}{n}\sum_{i=1}^n \left(\sum_{j\in\mathcal{N}_{b_n}} x_{ij}(\widetilde{\theta}_j - \theta_j) \right)^2 \le C_\lambda^2 \sum_{j\in\mathcal{N}_{b_n}} (\widetilde{\theta}_j - \theta_j)^2$$

$$\le 2C_\lambda^2 \sum_{j\in\mathcal{N}_{b_n}} \left(\rho_n^4 \left(\sum_{k=1}^r \frac{q_{jk}\zeta_k}{(\lambda_k^2+\rho_n)^2} \right)^2 + \left(\sum_{k=1}^r q_{jk} \left(\frac{\lambda_k}{\lambda_k^2+\rho_n} + \frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2} \right) \sum_{l=1}^n p_{lk}\epsilon_l \right)^2 \right)$$

$$= O(|\mathcal{N}_{b_n}| \times n^{2\alpha_\theta - 4\delta}) + 2C_\lambda^2 \sum_{j\in\mathcal{N}_{b_n}} \left(\sum_{k=1}^r q_{jk} \left(\frac{\lambda_k}{\lambda_k^2+\rho_n} + \frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2} \right) \sum_{l=1}^n p_{lk}\epsilon_l \right)^2$$

Since

$$
\begin{aligned}
&\mathbf{E}\sum_{j\in\mathcal{N}_{b_n}}\left(\sum_{k=1}^{r} q_{jk}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)\sum_{l=1}^{n} p_{lk}\epsilon_l\right)^2 \\
&=\sigma^2\sum_{j\in\mathcal{N}_{b_n}}\sum_{l=1}^{n}\left(\sum_{k=1}^{r} q_{jk}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)p_{lk}\right)^2 \qquad \text{(B.8)}\\
&=\sigma^2\sum_{j\in\mathcal{N}_{b_n}}\sum_{k=1}^{r} q_{jk}^2\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)^2 \le \frac{4\sigma^2|\mathcal{N}_{b_n}|}{\lambda_r^2}
\end{aligned}
$$

We have $\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j\in\mathcal{N}_{b_n}} x_{ij}(\widetilde{\theta}_j-\theta_j)\right)^2 = O_p(|\mathcal{N}_{b_n}|\times n^{2\alpha_\theta-4\delta}+|\mathcal{N}_{b_n}|\times n^{-2\eta})$. For the third term, from assumption 6 we have

$$
\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j\notin\mathcal{N}_{b_n}} x_{ij}\theta_j\right)^2 \le C_\lambda^2\sum_{j\notin\mathcal{N}_{b_n}}\theta_j^2 \le C_\lambda^2\times b_n\sum_{j\notin\mathcal{N}_{b_n}}|\theta_j| = O(n^{-\alpha_\sigma}) \qquad \text{(B.9)}
$$

For the fourth term, from Cauchy inequality and (B.7),
(B.10)

$$
\begin{aligned}
&\mathbf{E}\frac{1}{n}|\sum_{i=1}^{n}\sum_{j\in\mathcal{N}_{b_n}}\epsilon_i x_{ij}(\widetilde{\theta}_j-\theta_j)| \le \frac{1}{n}\mathbf{E}\sqrt{\sum_{i=1}^{n}\epsilon_i^2}\times\sqrt{\sum_{i=1}^{n}(\sum_{j\in\mathcal{N}_{b_n}} x_{ij}(\widetilde{\theta}_j-\theta_j))^2} \\
&\le\sqrt{\frac{\mathbf{E}\sum_{i=1}^{n}\epsilon_i^2}{n}}\times\sqrt{\frac{1}{n}\mathbf{E}\sum_{i=1}^{n}(\sum_{j\in\mathcal{N}_{b_n}} x_{ij}(\widetilde{\theta}_j-\theta_j))^2} = \sigma\times O(\sqrt{|\mathcal{N}_{b_n}|\times n^{2\alpha_\theta-4\delta}+|\mathcal{N}_{b_n}|\times n^{-2\eta}}) \\
&\Rightarrow \frac{1}{n}|\sum_{i=1}^{n}\sum_{j\in\mathcal{N}_{b_n}}\epsilon_i x_{ij}(\widetilde{\theta}_j-\theta_j)| = O_p(\sqrt{|\mathcal{N}_{b_n}|}\times n^{\alpha_\theta-2\delta}+\sqrt{|\mathcal{N}_{b_n}|}\times n^{-\eta})
\end{aligned}
$$

For the fifth term,

$$
\begin{aligned}
&\mathbf{E}|\frac{1}{n}\sum_{i=1}^{n}\sum_{j\notin\mathcal{N}_{b_n}}\epsilon_i x_{ij}\theta_j|^2 = \frac{\sigma^2}{n^2}\sum_{i=1}^{n}\left(\sum_{j\notin\mathcal{N}_{b_n}} x_{ij}\theta_j\right)^2 \le \frac{\sigma^2 C_\lambda^2}{n}\sum_{j\notin\mathcal{N}_{b_n}}\theta_j^2 \\
&\Rightarrow \frac{1}{n}\sum_{i=1}^{n}\sum_{j\notin\mathcal{N}_{b_n}}\epsilon_i x_{ij}\theta_j = O_p(n^{-(1+\alpha_\sigma)/2})
\end{aligned} \qquad \text{(B.11)}
$$

For the last term,
(B.12)

$$
\begin{aligned}
&\frac{1}{n}|\sum_{i=1}^{n}\left(\sum_{j\in\mathcal{N}_{b_n}} x_{ij}(\widetilde{\theta}_j-\theta_j)\right)\times\left(\sum_{j\notin\mathcal{N}_{b_n}} x_{ij}\theta_j\right)| \le C_\lambda^2\sqrt{\sum_{j\in\mathcal{N}_{b_n}}(\widetilde{\theta}_j-\theta_j)^2}\times\sqrt{\sum_{j\notin\mathcal{N}_{b_n}}\theta_j^2} \\
&= O_p(\sqrt{|\mathcal{N}_{b_n}|}\times n^{\alpha_\theta-2\delta-\alpha_\sigma/2}+\sqrt{|\mathcal{N}_{b_n}|}\times n^{-\eta-\alpha_\sigma/2})
\end{aligned}
$$

From (21), $Prob(\widehat{\mathcal{N}}_{b_n}\neq\mathcal{N}_{b_n})\to 0$. So we have

$$
|\widehat{\sigma}^2-\sigma^2| = O_p\left(\frac{1}{\sqrt{n}}+\sqrt{|\mathcal{N}_{b_n}|}\times n^{\alpha_\theta-2\delta}+\sqrt{|\mathcal{N}_{b_n}|}\times n^{-\eta}+n^{-\alpha_\sigma}\right) \qquad \text{(B.13)}
$$

From assumption 2 and 6, we prove the second result. □

Define $T=(c_{ik})_{i\in\mathcal{M},k=1,2,...,r}$. From assumption 7, since the matrix $\left(\frac{1}{\tau_i}c_{ik}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)\right)_{i\in\mathcal{M},j=1,2,...,r}=D_1TD_2$ with $D_1=diag(1/\tau_i,i\in\mathcal{M})$ and $D_2=diag\left(\frac{\lambda_1}{\lambda_1^2+\rho_n}+\frac{\rho_n\lambda_1}{(\lambda_1^2+\rho_n)^2},...,\frac{\lambda_r}{\lambda_r^2+\rho_n}+\frac{\rho_n\lambda_r}{(\lambda_r^2+\rho_n)^2}\right)$, the matrix $\left(\frac{1}{\tau_i}c_{ik}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)\right)_{i\in\mathcal{M},j=1,2,...,r}$ also has rank $|\mathcal{M}|$. The proof of theorem 2 uses this result.

PROOF OF THEOREM 2. From Cauchy inequality and assumption 2, suppose $\delta=\frac{\eta+\alpha_\theta+\delta_1}{2}$ with $\delta_1>0$. For $i\in\mathcal{M}$,

$$\text{(B.14)}\quad \begin{aligned}|\sum_{k=1}^{r}\frac{c_{ik}\zeta_k}{(\lambda_k^2+\rho_n)^2}|&\leq\sqrt{\sum_{k=1}^{r}\frac{c_{ik}^2\lambda_k^2}{(\lambda_k^2+\rho_n)^2}}\times\sqrt{\sum_{k=1}^{r}\frac{\zeta_k^2}{\lambda_k^2(\lambda_k^2+\rho_n)^2}}\leq\tau_i\times\frac{\|\theta\|_2}{\lambda_r^3}\\ &\Rightarrow\max_{i\in\mathcal{M}}\frac{\rho_n^2}{\tau_i}|\sum_{k=1}^{r}\frac{c_{ik}\zeta_k}{(\lambda_k^2+\rho_n)^2}|=O(n^{-\delta_1})\end{aligned}$$

Define $t_{il}=\frac{1}{\tau_i}\times\sum_{k=1}^{r}c_{ik}p_{lk}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)$ for $i\in\mathcal{M}$ and $l=1,2,...,n$. From (16), (5), (B.4) and assumption 5, if $\widehat{\mathcal{N}}_{b_n}=\mathcal{N}_{b_n}$, we have $\widehat{\tau}_i=\tau_i\geq 1/\sqrt{n}$ and $\exists$ a constant $C>0$, for any $a>0$ and sufficiently large n,

(B.15)

$$\begin{aligned}\max_{i=1,2,...,p_1}\frac{|\widehat{\gamma}_i-\gamma_i|}{\widehat{\tau}_i}&\leq\max_{i\in\mathcal{M}}\frac{\rho_n^2}{\tau_i}|\sum_{k=1}^{r}\frac{c_{ik}\zeta_k}{(\lambda_k^2+\rho_n)^2}|+\max_{i\in\mathcal{M}}|\sum_{l=1}^{n}t_{il}\epsilon_l|+\max_{i=1,2,...,p_1}\frac{|\sum_{j\notin\mathcal{N}_{b_n}}m_{ij}\theta_j|}{\tau_i}\\&\quad+\max_{i=1,2,...,p_1}\frac{|\sum_{j=1}^{p}m_{ij}\theta_{\perp,j}|}{\tau_i}\\&\leq\max_{i\in\mathcal{M}}|\sum_{l=1}^{n}t_{il}\epsilon_l|+Cn^{-\delta_1}+\frac{a}{\sqrt{\log(n)}}\\ \max_{i=1,2,...,p_1}\frac{|\widehat{\gamma}_i-\gamma_i|}{\widehat{\tau}_i}&\geq\max_{i\in\mathcal{M}}|\sum_{l=1}^{n}t_{il}\epsilon_l|-\max_{i\in\mathcal{M}}\frac{\rho_n^2}{\tau_i}|\sum_{k=1}^{r}\frac{c_{ik}\zeta_k}{(\lambda_k^2+\rho_n)^2}|-\max_{i=1,2,...,p_1}\frac{|\sum_{j\notin\mathcal{N}_{b_n}}m_{ij}\theta_j|}{\tau_i}\\&\quad-\max_{i=1,2,...,p_1}\frac{|\sum_{j=1}^{p}m_{ij}\theta_{\perp,j}|}{\tau_i}\\&\geq\max_{i\in\mathcal{M}}|\sum_{l=1}^{n}t_{il}\epsilon_l|-Cn^{-\delta_1}-\frac{a}{\sqrt{\log(n)}}\end{aligned}$$

According to theorem 1, $\exists$ a constant C and for any given $a > 0$, for sufficiently large n and any $x \geq 0$,
(B.16)

$$
\begin{aligned}
Prob\left(\max_{i=1,2,\dots,p_1}\frac{|\widehat{\gamma}_i-\gamma_i|}{\widehat{\tau}_i}\leq x\right) &\leq Prob\left(\max_{i=1,2,\dots,p_1}\frac{|\widehat{\gamma}_i-\gamma_i|}{\widehat{\tau}_i}\leq x\cap\widehat{\mathcal{N}}_{b_n}=\mathcal{N}_{b_n}\right)+Prob\left(\widehat{\mathcal{N}}_{b_n}\neq\mathcal{N}_{b_n}\right)\\
&\leq Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^{n}t_{il}\epsilon_l|\leq x+Cn^{-\delta_1}+\frac{a}{\sqrt{\log(n)}}\right)+Cn^{\alpha_p+m\nu_b-m\eta}\\
&\leq Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^{n}t_{il}\epsilon_l^*|\leq x\right)+Cn^{\alpha_p+m\nu_b-m\eta}\\
&+\sup_{x\geq 0}|Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^{n}t_{il}\epsilon_l|\leq x\right)-Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^{n}t_{il}\epsilon_l^*|\leq x\right)|\\
&+\sup_{x\in\mathbf{R}}\left(Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^{n}t_{il}\epsilon_l^*|\leq x+Cn^{-\delta_1}+\frac{a}{\sqrt{\log(n)}}\right)-Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^{n}t_{il}\epsilon_l^*|\leq x\right)\right)\\
Prob\left(\max_{i=1,2,\dots,p_1}\frac{|\widehat{\gamma}_i-\gamma_i|}{\widehat{\tau}_i}\leq x\right) &\geq Prob\left(\max_{i=1,2,\dots,p_1}\frac{|\widehat{\gamma}_i-\gamma_i|}{\widehat{\tau}_i}\leq x\cap\widehat{\mathcal{N}}_{b_n}=\mathcal{N}_{b_n}\right)\\
&\geq Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^{n}t_{il}\epsilon_l|\leq x-Cn^{-\delta_1}-\frac{a}{\sqrt{\log(n)}}\right)-Prob\left(\widehat{\mathcal{N}}_{b_n}\neq\mathcal{N}_{b_n}\right)\\
&\geq Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^{n}t_{il}\epsilon_l^*|\leq x\right)-Cn^{\alpha_p+m\nu_b-m\eta}\\
&-\sup_{x\in\mathbf{R}}\left(Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^{n}t_{il}\epsilon_l^*|\leq x\right)-Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^{n}t_{il}\epsilon_l^*|\leq x-Cn^{-\delta_1}-\frac{a}{\sqrt{\log(n)}}\right)\right)\\
&-|Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^{n}t_{il}\epsilon_l|\leq x-Cn^{-\delta_1}-\frac{a}{\sqrt{\log(n)}}\right)-Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^{n}t_{il}\epsilon_l^*|\leq x-Cn^{-\delta_1}-\frac{a}{\sqrt{\log(n)}}\right)|
\end{aligned}
$$

From assumption 1, 2, 5 and 7, for sufficiently large n we have
(B.17)

$$
\begin{aligned}
\max_{i\in\mathcal{M}}\mathbf{E}\left(\sum_{l=1}^{n}t_{il}\epsilon_l^*\right)^2 &= \sigma^2\max_{i\in\mathcal{M}}\sum_{l=1}^{n}t_{il}^2=\sigma^2\max_{i\in\mathcal{M}}\frac{\sum_{k=1}^{r}c_{ik}^2\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)^2}{\tau_i^2}\leq\sigma^2\\
\min_{i\in\mathcal{M}}\mathbf{E}\left(\sum_{l=1}^{n}t_{il}\epsilon_l^*\right)^2 &= \sigma^2\min_{i\in\mathcal{M}}\frac{1}{1+\frac{1}{n\sum_{k=1}^{r}c_{ik}^2\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)^2}}\\
&\geq\sigma^2\min_{i\in\mathcal{M}}\frac{1}{1+\frac{1}{n\sum_{k=1}^{r}c_{ik}^2\frac{\lambda_k^2}{(\lambda_k^2+\rho_n)^2}}}\geq\frac{\sigma^2}{1+\frac{4C_\lambda^2}{c_{\mathcal{M}}}}>0
\end{aligned}
$$

and $(t_{il})_{i\in\mathcal{M},l=1,2,\dots,n}=D_1TD_2P^T$, here $T=(c_{ik})_{i\in\mathcal{M},k=1,2,\dots,r}$, $D_1=diag(1/\tau_i,i\in\mathcal{M})$, and

$D_2 = diag\left(\frac{\lambda_1}{\lambda_1^2+\rho_n} + \frac{\rho_n\lambda_1}{(\lambda_1^2+\rho_n)^2}, \dots, \frac{\lambda_r}{\lambda_r^2+\rho_n} + \frac{\rho_n\lambda_r}{(\lambda_r^2+\rho_n)^2}\right)$. So $(t_{il})_{i\in\mathcal{M},l=1,2,\dots,n}$ has full rank(rank $|\mathcal{M}|$). From lemma A.2, $\exists$ a constant $C^{'}$ which only depends on $\sigma, c_{\mathcal{M}}, C_\lambda$ such that

(B.18)
$$\sup_{x\in\mathbf{R}}\left(Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^n t_{il}\epsilon_l^*| \le x + Cn^{-\delta_1} + \frac{a}{\sqrt{\log(n)}}\right) - Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^n t_{il}\epsilon_l^*| \le x\right)\right)$$
$$\le C^{'}\left(Cn^{-\delta_1} + \frac{a}{\sqrt{\log(n)}}\right) \times \left(1 + \sqrt{\log(|\mathcal{M}|)} + \sqrt{|\log(Cn^{-\delta_1} + \frac{a}{\sqrt{\log(n)}})|}\right)$$

For sufficiently large n, we have $Cn^{-\delta_1} + \frac{a}{\sqrt{\log(n)}} < 1$ and

(B.19)
$$|\log(Cn^{-\delta_1} + \frac{a}{\sqrt{\log(n)}})| \le \log(\frac{\sqrt{\log(n)}}{a}) = \frac{\log(\log(n))}{2} - \log(a) \le \log(\log(n))$$
$$\Rightarrow \sup_{x\in\mathbf{R}}\left(Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^n t_{il}\epsilon_l^*| \le x + Cn^{-\delta_1} + \frac{a}{\sqrt{\log(n)}}\right) - Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^n t_{il}\epsilon_l^*| \le x\right)\right)$$
$$\le C^{'}\left(Cn^{-\delta_1} + \frac{a}{\sqrt{\log(n)}}\right) \times \left(1 + \sqrt{\log(n)} + \sqrt{\log(\log(n))}\right) \le 6C^{'}a$$

From assumption 7, (B.17) and lemma A.3, we have

(B.20)
$$\sup_{x\ge 0}|Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^n t_{il}\epsilon_l| \le x\right) - Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^n t_{il}\epsilon_l^*| \le x\right)| < a \text{ for sufficiently large } n$$

If $x < Cn^{-\delta_1} + \frac{a}{\sqrt{\log(n)}}$, then $Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^n t_{il}\epsilon_l| \le x - Cn^{-\delta_1} - \frac{a}{\sqrt{\log(n)}}\right) = 0$ and $Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^n t_{il}\epsilon_l^*| \le x - Cn^{-\delta_1} - \frac{a}{\sqrt{\log(n)}}\right) = 0$. Combine with (B.16) to (B.20), we have

(B.21)
$$\sup_{x\ge 0}|Prob\left(\max_{i=1,2,\dots,p_1}\frac{|\widehat{\gamma}_i - \gamma_i|}{\widehat{\tau}_i} \le x\right) - Prob\left(\max_{i\in\mathcal{M}}|\sum_{l=1}^n t_{il}\epsilon_l^*| \le x\right)| \le Cn^{\alpha_p+m\nu_b-m\eta} + 6C^{'}a + a$$

and we prove (27). $\square$

Define $c_{1-\alpha}$ as the $1-\alpha$ quantile of H. The density of a multivariate normal random variable with a full rank covariance matrix is positive for $\forall x \in \mathcal{R}^{|\mathcal{M}|}$. And $\forall x \ge 0,\ \delta > 0$, the set $\{t = (t_i, i\in\mathcal{M}) |\ x < \max_{i\in\mathcal{M}}|t_i| \le x+\delta\}$ has positive Lebesgue measure. Therefore, $H(x)$ is strictly increasing, and for any $0<\alpha<1$, $H(c_{1-\alpha}) = 1-\alpha$. From theorem 2, for any given $0 < \alpha_0 < \alpha_1 < 1$,

(B.22)
$$\sup_{\alpha_0\le\alpha\le\alpha_1}|Prob\left(\max_{i=1,2,\dots,p_1}\frac{|\widehat{\gamma}_i - \gamma_i|}{\widehat{\tau}_i} \le c_{1-\alpha}\right) - (1-\alpha)|$$
$$\le \sup_{x\ge 0}|Prob\left(\max_{i=1,2,\dots,p_1}\frac{|\widehat{\gamma}_i - \gamma_i|}{\widehat{\tau}_i} \le x\right) - H(x)| \to 0$$

as $n\to\infty$.

APPENDIX C: PROOFS OF THEOREMS IN SECTION 4

PROOF OF THEOREM 3. According to theorem 1, $Prob\left(\widehat{\mathcal{N}}_{b_n} \neq \mathcal{N}_{b_n}\right) = O(n^{\alpha_p + m\nu_b - m\eta})$. If $\widehat{\mathcal{N}}_{b_n} = \mathcal{N}_{b_n}$, from (16)

$$\|\widehat{\theta}\|_2^2 = \sum_{i\in\mathcal{N}_{b_n}} \widetilde{\theta}_i^2 \leq 3\sum_{i\in\mathcal{N}_{b_n}} |\theta_i|^2 + 3\rho_n^4 \sum_{i\in\mathcal{N}_{b_n}} \left(\sum_{j=1}^r \frac{q_{ij}\zeta_j}{(\lambda_j^2+\rho_n)^2}\right)^2$$
$$+3\sum_{i\in\mathcal{N}_{b_n}} \left(\sum_{j=1}^r\sum_{l=1}^n q_{ij}\left(\frac{\lambda_j}{\lambda_j^2+\rho_n} + \frac{\rho_n\lambda_j}{(\lambda_j^2+\rho_n)^2}\right) p_{lj}\epsilon_l\right)^2 \tag{C.1}$$

From assumption 2, $\sum_{i\in\mathcal{N}_{b_n}} |\theta_i|^2 \leq \|\theta\|_2^2 = O(n^{2\alpha_\theta})$. Similarly

$$\rho_n^4 \sum_{i\in\mathcal{N}_{b_n}} \left(\sum_{j=1}^r \frac{q_{ij}\zeta_j}{(\lambda_j^2+\rho_n)^2}\right)^2 \leq \frac{\rho_n^4}{\lambda_r^8} \sum_{i\in\mathcal{N}_{b_n}} \sum_{j=1}^r q_{ij}^2 \sum_{j=1}^r \zeta_j^2 = \frac{\rho_n^4 \times |\mathcal{N}_{b_n}| \times \|\theta\|_2^2}{\lambda_r^8} = o(n^{-2\alpha_\sigma}) \tag{C.2}$$

From assumption 6,

$$\mathbf{E}\sum_{i\in\mathcal{N}_{b_n}} \left(\sum_{j=1}^r\sum_{l=1}^n q_{ij}\left(\frac{\lambda_j}{\lambda_j^2+\rho_n} + \frac{\rho_n\lambda_j}{(\lambda_j^2+\rho_n)^2}\right) p_{lj}\epsilon_l\right)^2$$
$$= \sigma^2 \sum_{i\in\mathcal{N}_{b_n}} \sum_{l=1}^n \left(\sum_{j=1}^r q_{ij}\left(\frac{\lambda_j}{\lambda_j^2+\rho_n} + \frac{\rho_n\lambda_j}{(\lambda_j^2+\rho_n)^2}\right) p_{lj}\right)^2$$
$$= \sigma^2 \sum_{i\in\mathcal{N}_{b_n}} \sum_{j=1}^r q_{ij}^2 \left(\frac{\lambda_j}{\lambda_j^2+\rho_n} + \frac{\rho_n\lambda_j}{(\lambda_j^2+\rho_n)^2}\right)^2 \leq \frac{4\sigma^2|\mathcal{N}_{b_n}|}{\lambda_r^2} \tag{C.3}$$

$$\Rightarrow \sum_{i\in\mathcal{N}_{b_n}} \left(\sum_{j=1}^r\sum_{l=1}^n q_{ij}\left(\frac{\lambda_j}{\lambda_j^2+\rho_n} + \frac{\rho_n\lambda_j}{(\lambda_j^2+\rho_n)^2}\right) p_{lj}\epsilon_l\right)^2 = O_p\left(n^{-2\alpha_\sigma}\right)$$

Since $\alpha_\theta, \alpha_\sigma \geq 0$, $\|\widehat{\theta}\|_2 = O_p(n^{\alpha_\theta})$. According to (15) and (16), define $\widehat{\zeta} = Q^T\widehat{\theta}$,

$$\widetilde{\theta}^* - \widehat{\theta} = \left(I_p + \rho_n Q(\Lambda^2+\rho_n I_r)^{-1}Q^T\right) Q(\Lambda^2+\rho_n I_r)^{-1}\left(\Lambda^2 Q^T\widehat{\theta} + \Lambda P^T\epsilon^*\right) + \widehat{\theta}_\perp - QQ^T\widehat{\theta} - Q_\perp Q_\perp^T\widehat{\theta} \tag{C.4}$$
$$\Rightarrow \widetilde{\theta}_i^* - \widehat{\theta}_i = -\rho_n^2 \sum_{j=1}^r \frac{q_{ij}\widehat{\zeta}_j}{(\lambda_j^2+\rho_n)^2} + \sum_{j=1}^r\sum_{l=1}^n q_{ij}\left(\frac{\lambda_j}{\lambda_j^2+\rho_n} + \frac{\rho_n\lambda_j}{(\lambda_j^2+\rho_n)^2}\right) p_{lj}\epsilon_l^*$$

Similar to (B.5), rewrite δ in assumption 2 as $\delta = \frac{\eta+\alpha_\theta+\delta_1}{2}$ with $\delta_1 > 0$, we have

$$\max_{i=1,2,...,p} |\rho_n^2 \sum_{j=1}^r \frac{q_{ij}\widehat{\zeta}_j}{(\lambda_j^2+\rho_n)^2}| \leq \max_{i=1,2,...,p} \frac{\rho_n^2}{\lambda_r^4}\sqrt{\sum_{j=1}^r q_{ij}^2} \times \sqrt{\sum_{j=1}^r \widehat{\zeta}_j^2} \leq \frac{\rho_n^2\|\widehat{\theta}\|_2}{c_\lambda^4 n^{4\eta}} = O_p\left(n^{-\eta-\delta_1}\right) \tag{C.5}$$

$\epsilon_i^*|\epsilon, i = 1,2,...,n$ are normal random variables with mean 0 and variance $\widehat{\sigma}^2$. Therefore $\mathbf{E}^*|\epsilon_1^*|^m = \widehat{\sigma}^m D$, $D = \mathbf{E}|Y|^m$, Y is a normal random variable with mean 0 and variance 1.

If $\widehat{\sigma} > 0$, from (B.2) and lemma A.1, $\exists$ a constant E which depends on m and D such that for any $a > 0$,

$$Prob^*\left(\max_{i=1,2,\dots,p}|\sum_{j=1}^{r}\sum_{l=1}^{n}q_{ij}\left(\frac{\lambda_j}{\lambda_j^2+\rho_n}+\frac{\rho_n\lambda_j}{(\lambda_j^2+\rho_n)^2}\right)p_{lj}\frac{\epsilon_l^*}{\widehat{\sigma}}|>\frac{a}{\widehat{\sigma}}\right)\leq\frac{pE\widehat{\sigma}^m}{\lambda_r^m a^m} \tag{C.6}$$

Suppose $\widehat{\mathcal{N}}_{b_n}=\mathcal{N}_{b_n}$, $\frac{\sigma}{2}<\widehat{\sigma}<\frac{3\sigma}{2}$, and $\max_{i=1,2,\dots,p}|\rho_n^2\sum_{j=1}^{r}\frac{q_{ij}\widehat{\zeta}_j}{(\lambda_j^2+\rho_n)^2}|\leq C\times n^{-\eta-\delta_1}$ for a constant C. Since $\widehat{\theta}_i=0$ if $i\notin\widehat{\mathcal{N}}_{b_n}$,

(C.7)

$$Prob^*\left(\widehat{\mathcal{N}}_{b_n}^*\neq\mathcal{N}_{b_n}\right)\leq Prob^*\left(\min_{i\in\mathcal{N}_{b_n}}|\widetilde{\theta}_i^*|\leq b_n\right)+Prob^*\left(\max_{i\notin\mathcal{N}_{b_n}}|\widetilde{\theta}_i^*|>b_n\right)$$

$$\leq Prob^*\left(\min_{i\in\mathcal{N}_{b_n}}|\widehat{\theta}_i|-\max_{i\in\mathcal{N}_{b_n}}|\rho_n^2\sum_{j=1}^{r}\frac{q_{ij}\widehat{\zeta}_j}{(\lambda_j^2+\rho_n)^2}|-b_n\leq\max_{i\in\mathcal{N}_{b_n}}|\sum_{j=1}^{r}\sum_{l=1}^{n}q_{ij}\left(\frac{\lambda_j}{\lambda_j^2+\rho_n}+\frac{\rho_n\lambda_j}{(\lambda_j^2+\rho_n)^2}\right)p_{lj}\epsilon_l^*|\right)$$

$$+Prob^*\left(\max_{i\notin\mathcal{N}_{b_n}}|\sum_{j=1}^{r}\sum_{l=1}^{n}q_{ij}\left(\frac{\lambda_j}{\lambda_j^2+\rho_n}+\frac{\rho_n\lambda_j}{(\lambda_j^2+\rho_n)^2}\right)p_{lj}\epsilon_l^*|>b_n-\rho_n^2\max_{i\notin\mathcal{N}_{b_n}}|\sum_{j=1}^{r}\frac{q_{ij}\widehat{\zeta}_j}{(\lambda_j^2+\rho_n)^2}|\right)$$

From assumption 4, for sufficiently large n,

$$b_n-\rho_n^2\max_{i\notin\mathcal{N}_{b_n}}|\sum_{j=1}^{r}\frac{q_{ij}\widehat{\zeta}_j}{(\lambda_j^2+\rho_n)^2}|\geq C_b n^{-\nu_b}-Cn^{-\eta-\delta_1}\geq\frac{b_n}{2} \tag{C.8}$$

From (B.2), lemma A.1, assumption 1 and 4, we have

$$\max_{i=1,2,\dots,p}|\sum_{j=1}^{r}q_{ij}\left(\frac{\lambda_j}{\lambda_j^2+\rho_n}+\frac{\rho_n\lambda_j}{(\lambda_j^2+\rho_n)^2}\right)\sum_{l=1}^{n}p_{lj}\epsilon_l|=O_p\left(n^{\alpha_p/m-\eta}\right) \tag{C.9}$$

Suppose a constant C such that $\max_{i=1,2,\dots,p}|\sum_{j=1}^{r}q_{ij}\left(\frac{\lambda_j}{\lambda_j^2+\rho_n}+\frac{\rho_n\lambda_j}{(\lambda_j^2+\rho_n)^2}\right)\sum_{l=1}^{n}p_{lj}\epsilon_l|\leq Cn^{\alpha_p/m-\eta}$(from lemma A.1), and (since $\frac{\rho_n^2\|\theta\|_2}{\lambda_r^4}=O(n^{-\eta-\delta_1})$) $\frac{\rho_n^2\|\theta\|_2}{\lambda_r^4}\leq Cn^{-\eta-\delta_1}$. From assumption 4, for sufficiently large n,

(C.10)

$$\min_{i\in\mathcal{N}_{b_n}}|\widehat{\theta}_i|\geq\min_{i\in\mathcal{N}_{b_n}}|\theta_i|-\max_{i\in\mathcal{N}_{b_n}}\rho_n^2|\sum_{j=1}^{r}\frac{q_{ij}\zeta_j}{(\lambda_j^2+\rho_n)^2}|-\max_{i\in\mathcal{N}_{b_n}}|\sum_{j=1}^{r}q_{ij}\left(\frac{\lambda_j}{\lambda_j^2+\rho_n}+\frac{\rho_n\lambda_j}{(\lambda_j^2+\rho_n)^2}\right)\sum_{l=1}^{n}p_{lj}\epsilon_l|$$

$$\geq\frac{b_n}{c_b}-\frac{\rho_n^2\|\theta\|_2}{\lambda_r^4}-Cn^{\alpha_p/m-\eta}\Rightarrow\min_{i\in\mathcal{N}_{b_n}}|\widehat{\theta}_i|-\max_{i\in\mathcal{N}_{b_n}}|\rho_n^2\sum_{j=1}^{r}\frac{q_{ij}\widehat{\zeta}_j}{(\lambda_j^2+\rho_n)^2}|-b_n$$

$$\geq\left(\frac{1}{c_b}-1\right)b_n-Cn^{-\eta-\delta_1}-Cn^{\alpha_p/m-\eta}-Cn^{-\eta-\delta_1}>\frac{b_n}{2}\left(\frac{1}{c_b}-1\right)$$

Correspondingly

(C.11)

$$Prob^*\left(\widehat{\mathcal{N}}_{b_n}^*\neq\mathcal{N}_{b_n}\right)\leq Prob^*\left(\max_{i\in\mathcal{N}_{b_n}}|\sum_{j=1}^{r}\sum_{l=1}^{n}q_{ij}\left(\frac{\lambda_j}{\lambda_j^2+\rho_n}+\frac{\rho_n\lambda_j}{(\lambda_j^2+\rho_n)^2}\right)p_{lj}\epsilon_l^*|>\frac{b_n}{2}(\frac{1}{c_b}-1)\right)$$

$$+Prob^*\left(\max_{i\notin\mathcal{N}_{b_n}}|\sum_{j=1}^{r}\sum_{l=1}^{n}q_{ij}\left(\frac{\lambda_j}{\lambda_j^2+\rho_n}+\frac{\rho_n\lambda_j}{(\lambda_j^2+\rho_n)^2}\right)p_{lj}\epsilon_l^*|>b_n/2\right)\leq\frac{pE\widehat{\sigma}^m}{c_\lambda^m n^{m\eta}b_n^m}\times\left(\frac{2^m}{(1/c_b-1)^m}+2^m\right)$$

which has order $O_p(n^{\alpha_p+m\nu_b-m\eta})$. If $\widehat{\mathcal{N}}^*_{b_n}=\mathcal{N}_{b_n}$, then $\widehat{\tau}^*_i=\tau_i$ for $i=1,2,...,p_1$. Similar to (B.14),

(C.12)
$$\max_{i=1,2,\dots,p_1}\frac{|\widehat{\gamma}^*_i-\widehat{\gamma}_i|}{\widehat{\tau}^*_i}=\max_{i=1,2,\dots,p_1}\frac{|-\rho_n^2\sum_{k=1}^r\frac{c_{ik}\widehat{\zeta}_k}{(\lambda_k^2+\rho_n)^2}+\sum_{l=1}^n\sum_{k=1}^r c_{ik}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)p_{lk}\epsilon^*_l|}{\tau_i}$$
$$\le\max_{i\in\mathcal{M}}\rho_n^2\frac{|\sum_{k=1}^r\frac{c_{ik}\widehat{\zeta}_k}{(\lambda_k^2+\rho_n)^2}|}{\tau_i}+\max_{i\in\mathcal{M}}\frac{|\sum_{l=1}^n\sum_{k=1}^r c_{ik}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)p_{lk}\epsilon^*_l|}{\tau_i}$$
$$\le\frac{\rho_n^2\|\widehat{\theta}\|_2}{\lambda_r^3}+\max_{i\in\mathcal{M}}\frac{|\sum_{l=1}^n\sum_{k=1}^r c_{ik}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)p_{lk}\epsilon^*_l|}{\tau_i}$$
$$\max_{i=1,2,\dots,p_1}\frac{|\widehat{\gamma}^*_i-\widehat{\gamma}_i|}{\widehat{\tau}^*_i}\ge\max_{i\in\mathcal{M}}\frac{|\sum_{l=1}^n\sum_{k=1}^r c_{ik}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)p_{lk}\epsilon^*_l|}{\tau_i}-\frac{\rho_n^2\|\widehat{\theta}\|_2}{\lambda_r^3}$$

From theorem 1, for any $a>0$ and sufficiently large n, $\exists$ constant D_a such that $|\widehat{\sigma}^2-\sigma^2|\le D_a n^{-\alpha_\sigma}$ and $\frac{1}{2}\sigma<\widehat{\sigma}<\frac{3}{2}\sigma$ with probability $1-a$,

$$|\sigma-\widehat{\sigma}|=\frac{|\sigma^2-\widehat{\sigma}^2|}{\sigma+\widehat{\sigma}}\le\frac{D_a n^{-\alpha_\sigma}}{\sigma} \tag{C.13}$$

If $0<x\le n^{\alpha_\sigma/2}$, according to lemma A.2, assumption 7 and (B.17), $\exists$ a constant $C^{'}$ which only depends on $\sigma, c_{\mathcal{M}}, C_\lambda$ such that

(C.14)
$$|Prob^*\left(\max_{i\in\mathcal{M}}\frac{|\sum_{l=1}^n\sum_{k=1}^r c_{ik}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)p_{lk}\epsilon^*_l|}{\tau_i}\le x\right)-H(x)|=|H(\frac{x\sigma}{\widehat{\sigma}})-H(x)|$$
$$\le C^{'}\left(1+\sqrt{\log(|\mathcal{M}|)}+\sqrt{|\log(\frac{x|\sigma-\widehat{\sigma}|}{\widehat{\sigma}})|}\right)\frac{x|\sigma-\widehat{\sigma}|}{\widehat{\sigma}}$$
$$\le\frac{2D_aC^{'}}{\sigma^2}\left(1+\sqrt{\log(n)}\right)n^{-\alpha_\sigma/2}+C^{'}\sqrt{\frac{x|\sigma-\widehat{\sigma}|}{\widehat{\sigma}}|\log(\frac{x|\sigma-\widehat{\sigma}|}{\widehat{\sigma}})|}\times\sqrt{\frac{2D_a}{\sigma^2}}n^{-\alpha_\sigma/4}$$

Function $x\log(x)$ is continuous when $x>0$, $x\log(x)\to 0$ as $x\to 0$, and $\frac{x|\sigma-\widehat{\sigma}|}{\widehat{\sigma}}\le\frac{2D_an^{-\alpha_\sigma/2}}{\sigma^2}\to 0$ as $n\to\infty$. So $\sqrt{\frac{x|\sigma-\widehat{\sigma}|}{\widehat{\sigma}}|\log(\frac{x|\sigma-\widehat{\sigma}|}{\widehat{\sigma}})|}\le\sup_{x\in(0,1]}\sqrt{|x\log(x)|}<\infty$ for sufficiently large n.

On the other hand, if $x>n^{\alpha_\sigma/2}$, then $\frac{x\sigma}{\widehat{\sigma}}>\frac{2n^{\alpha_\sigma/2}}{3}$. From lemma A.1, we may choose sufficiently large m_1 such that $m_1\alpha_\sigma/2>2$, since $\mathbf{E}|\xi_1|^{m_1}<\infty$(Here ξ_1 is a normal random variable with mean 0 and variance σ^2) is a constant for given m_1 and

$\max_{i\in\mathcal{M}}\sum_{k=1}^{r}\frac{1}{\tau_i^2}c_{ik}^2\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)^2\leq 1$, we have

$$Prob\left(\max_{i\in\mathcal{M}}\frac{|\sum_{k=1}^{r}c_{ik}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)\xi_k|}{\tau_i}>\frac{2n^{\alpha_\sigma/2}}{3}\right)\leq\frac{3^{m_1}|\mathcal{M}|\times E}{2^{m_1}n^{m_1\alpha_\sigma/2}}$$

$$\begin{aligned}
&\Rightarrow|Prob^*\left(\max_{i\in\mathcal{M}}\frac{|\sum_{l=1}^{n}\sum_{k=1}^{r}c_{ik}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)p_{lk}\epsilon_l^*|}{\tau_i}\leq x\right)-H(x)| \\
&\leq Prob\left(\max_{i\in\mathcal{M}}\frac{|\sum_{k=1}^{r}c_{ik}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)\xi_k|}{\tau_i}>\frac{2n^{\alpha_\sigma/2}}{3}\right) \\
&+Prob\left(\max_{i\in\mathcal{M}}\frac{|\sum_{k=1}^{r}c_{ik}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)\xi_k|}{\tau_i}>n^{\alpha_\sigma/2}\right) \\
&\leq 2\times\frac{3^{m_1}|\mathcal{M}|\times E}{2^{m_1}n^{m_1\alpha_\sigma/2}}
\end{aligned}\tag{C.15}$$

Since $H(0)=0$, from (C.14) and (C.15), for any given $a>0$ and sufficiently large n,

$$\sup_{x\geq 0}|Prob^*\left(\max_{i\in\mathcal{M}}\frac{|\sum_{l=1}^{n}\sum_{k=1}^{r}c_{ik}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)p_{lk}\epsilon_l^*|}{\tau_i}\leq x\right)-H(x)|<a \tag{C.16}$$

As a summary, for any given $a>0$, $\exists$ a constant D_a such that for sufficiently large n, the event $|\widehat{\sigma}^2-\sigma^2|\leq D_a n^{-\alpha_\sigma}$, $\frac{1}{2}\sigma<\widehat{\sigma}<\frac{3}{2}\sigma$, $\widehat{\mathcal{N}}_{b_n}=\mathcal{N}_{b_n}$, $\|\widehat{\theta}\|_2\leq D_a\times n^{\alpha_\theta}\Rightarrow\frac{\rho_n^2\|\widehat{\theta}\|_2}{\lambda_r^3}\leq D_a^{'}n^{-\delta_1}$ for constant $D_a^{'}$ and $\max_{i=1,2,\ldots,p}|\rho_n^2\sum_{j=1}^{r}\frac{q_{ij}\widehat{\zeta}_j}{(\lambda_j^2+\rho_n)^2}|\leq D_a\times n^{-\eta-\delta_1}$ happen with probability $1-a$. From (C.12), assumption 5 and lemma A.2, we have for any $x\geq 0$, there exists a constant $C^{'}$ such that

(C.17)

$$Prob^*\left(\max_{i=1,2,\ldots,p_1}\frac{|\widehat{\gamma}_i^*-\widehat{\gamma}_i|}{\widehat{\tau}_i^*}\le x\right)-H(x)\le Prob^*\left(\widehat{\mathcal{N}}_{b_n}^*\ne\mathcal{N}_{b_n}\right)$$

$$+C^{'}\frac{\rho_n^2\|\widehat{\theta}\|_2}{\lambda_r^3}\times\left(1+\sqrt{\log(|\mathcal{M}|)}+\sqrt{|\log(\frac{\rho_n^2\|\widehat{\theta}\|_2}{\lambda_r^3})|}\right)$$

$$+Prob^*\left(\max_{i\in\mathcal{M}}\frac{|\sum_{l=1}^n\sum_{k=1}^r c_{ik}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)p_{lk}\epsilon_l^*|}{\tau_i}\le x+\frac{\rho_n^2\|\widehat{\theta}\|_2}{\lambda_r^3}\right)-H(x+\frac{\rho_n^2\|\widehat{\theta}\|_2}{\lambda_r^3})$$

$$\le a+C^{'}D_a^{'}(1+\sqrt{\log(n)})n^{-\delta_1}+C^{'}\sqrt{D_a^{'}}n^{-\delta_1/2}\sqrt{|\log(\frac{\rho_n^2\|\widehat{\theta}\|_2}{\lambda_r^3})\times\frac{\rho_n^2\|\widehat{\theta}\|_2}{\lambda_r^3}|}+Prob^*\left(\widehat{\mathcal{N}}_{b_n}^*\ne\mathcal{N}_{b_n}\right)$$

$$Prob^*\left(\max_{i=1,2,\ldots,p_1}\frac{|\widehat{\gamma}_i^*-\widehat{\gamma}_i|}{\widehat{\tau}_i^*}\le x\right)-H(x)\ge Prob^*\left(\max_{i=1,2,\ldots,p_1}\frac{|\widehat{\gamma}_i^*-\widehat{\gamma}_i|}{\widehat{\tau}_i^*}\le x\cap\widehat{\mathcal{N}}_{b_n}^*=\mathcal{N}_{b_n}\right)-H(x)$$

$$\ge Prob^*\left(\max_{i\in\mathcal{M}}\frac{|\sum_{l=1}^n\sum_{k=1}^r c_{ik}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)p_{lk}\epsilon_l^*|}{\tau_i}\le x-\frac{\rho_n^2\|\widehat{\theta}\|_2}{\lambda_r^3}\right)-H(x-\frac{\rho_n^2\|\widehat{\theta}\|_2}{\lambda_r^3})$$

$$-Prob^*\left(\widehat{\mathcal{N}}_{b_n}^*\ne\mathcal{N}_{b_n}\right)-C^{'}D_a^{'}(1+\sqrt{\log(n)})n^{-\delta_1}-C^{'}\sqrt{D_a^{'}}n^{-\delta_1/2}\sqrt{|\log(\frac{\rho_n^2\|\widehat{\theta}\|_2}{\lambda_r^3})\times\frac{\rho_n^2\|\widehat{\theta}\|_2}{\lambda_r^3}|}$$

If $0\le x\le\frac{\rho_n^2\|\widehat{\theta}\|_2}{\lambda_r^3}$, then $Prob^*\left(\max_{i\in\mathcal{M}}\frac{|\sum_{l=1}^n\sum_{k=1}^r c_{ik}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)p_{lk}\epsilon_l^*|}{\tau_i}\le x-\frac{\rho_n^2\|\widehat{\theta}\|_2}{\lambda_r^3}\right)=$ $H(x-\frac{\rho_n^2\|\widehat{\theta}\|_2}{\lambda_r^3})=0$. Therefore, for sufficiently large n, from (C.16) and (C.11), $\exists$ a constant C such that

(C.18)

$$\sup_{x\ge0}|Prob^*\left(\max_{i=1,2,\ldots,p_1}\frac{|\widehat{\gamma}_i^*-\widehat{\gamma}_i|}{\widehat{\tau}_i^*}\le x\right)-H(x)|\le\frac{pE\widehat{\sigma}^m}{c_\lambda^m n^{m\eta}b_n^m}\times\left(2^m+\frac{2^m}{(1/c_b-1)^m}\right)+a$$

$$+C^{'}D_a^{'}(1+\sqrt{\log(n)})n^{-\delta_1}+C^{'}\sqrt{D_a^{'}}n^{-\delta_1/2}\sqrt{\sup_{x\in(0,1]}|x\log(x)|}\le Cn^{m(\nu_b+\alpha_p/m-\eta)}+2a$$

and we prove (32).

For any given $a>0$, from the first result, for sufficiently large n, we have

$$Prob\left(\sup_{x\ge0}|Prob^*\left(\max_{i=1,2,\ldots,p_1}\frac{|\widehat{\gamma}_i^*-\widehat{\gamma}_i|}{\widehat{\tau}_i^*}\le x\right)-H(x)|\le a\right)>1-a \tag{C.19}$$

Choose sufficiently small a such that $0<1-\alpha-2a<1-\alpha+2a<1$. If (C.19) happens, for any $1>\alpha>0$, define $c_{1-\alpha}$ as the $1-\alpha$ quantile of $H(x)$,

$$\begin{aligned}&Prob^*\left(\max_{i=1,2,\ldots,p_1}\frac{|\widehat{\gamma}_i^*-\widehat{\gamma}_i|}{\widehat{\tau}_i^*}\le c_{1-\alpha+2a}\right)-(1-\alpha+2a)\ge-a\Rightarrow c_{1-\alpha}^*\le c_{1-\alpha+2a}\\&Prob^*\left(\max_{i=1,2,\ldots,p_1}\frac{|\widehat{\gamma}_i^*-\widehat{\gamma}_i|}{\widehat{\tau}_i^*}\le c_{1-\alpha-2a}\right)-(1-\alpha-2a)\le a\Rightarrow c_{1-\alpha}^*>c_{1-\alpha-2a}\end{aligned} \tag{C.20}$$

From theorem 2, we have for sufficiently large n,
(C.21)

$$Prob\left(\max_{i=1,2,...,p_1}\frac{|\widehat{\gamma}_i-\gamma_i|}{\widehat{\tau}_i}\le c^*_{1-\alpha}\right)$$

$$\le Prob\left(\sup_{x\ge 0}\left|Prob^*\left(\max_{i=1,2,...,p_1}\frac{|\widehat{\gamma}^*_i-\widehat{\gamma}_i|}{\widehat{\tau}^*_i}\le x\right)-H(x)\right|>a\right)+Prob\left(\max_{i=1,2,...,p_1}\frac{|\widehat{\gamma}_i-\gamma_i|}{\widehat{\tau}_i}\le c_{1-\alpha+2a}\right)$$

$$\le a+(H(c_{1-\alpha+2a})+a)=1-\alpha+4a$$

$$Prob\left(\max_{i=1,2,...,p_1}\frac{|\widehat{\gamma}_i-\gamma_i|}{\widehat{\tau}_i}\le c^*_{1-\alpha}\right)$$

$$\ge Prob\left(\max_{i=1,2,...,p_1}\frac{|\widehat{\gamma}_i-\gamma_i|}{\widehat{\tau}_i}\le c^*_{1-\alpha}\cap\sup_{x\ge 0}\left|Prob^*\left(\max_{i=1,2,...,p_1}\frac{|\widehat{\gamma}^*_i-\widehat{\gamma}_i|}{\widehat{\tau}^*_i}\le x\right)-H(x)\right|\le a\right)$$

$$\ge Prob\left(\max_{i=1,2,...,p_1}\frac{|\widehat{\gamma}_i-\gamma_i|}{\widehat{\tau}_i}\le c_{1-\alpha-2a}\right)-Prob\left(\sup_{x\ge 0}\left|Prob^*\left(\max_{i=1,2,...,p_1}\frac{|\widehat{\gamma}^*_i-\widehat{\gamma}_i|}{\widehat{\tau}^*_i}\le x\right)-H(x)\right|>a\right)$$

$$\ge (H(c_{1-\alpha-2a})-a)-a=1-\alpha-4a\Rightarrow\left|Prob\left(\max_{i=1,2,...,p_1}\frac{|\widehat{\gamma}_i-\gamma_i|}{\widehat{\tau}_i}\le c^*_{1-\alpha}\right)-(1-\alpha)\right|\le 4a$$

For $a>0$ can be arbitrarily small, we prove (31). $\square$

APPENDIX D: PROOFS OF THEOREMS IN SECTION 5

PROOF OF LEMMA 1. Define the design matrix $X=(x_{ij})_{i=1,...,n,j=1,...,p}$, $\overline{x}_j=\frac{1}{n}\sum_{i=1}^n x_{ij}$, and $x^{'}_{ij}=x_{ij}-\overline{x}_j$. If $\widehat{\mathcal{N}}_{b_n}=\mathcal{N}_{b_n}$, for $i=1,2,...,n$,
(D.1)

$$\widehat{\epsilon}^{'}_i=\epsilon_i+\sum_{j\notin\mathcal{N}_{b_n}}x_{ij}\theta_j-\sum_{j\in\mathcal{N}_{b_n}}x_{ij}(\widetilde{\theta}_j-\theta_j)\Rightarrow\widehat{\epsilon}_i=\epsilon_i-\frac{1}{n}\sum_{i=1}^n\epsilon_i+\sum_{j\notin\mathcal{N}_{b_n}}x^{'}_{ij}\theta_j-\sum_{j\in\mathcal{N}_{b_n}}x^{'}_{ij}(\widetilde{\theta}_j-\theta_j)$$

Define $\widetilde{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\epsilon_i \le x},\ x \in \mathbf{R}$. From (A.17), for any given $\psi > 0$,

(D.2)

$$\begin{aligned}
\widehat{F}(x) - F(x) &= \left(\widehat{F}(x) - \widetilde{F}(x+1/\psi)\right) + \left(\widetilde{F}(x+1/\psi) - F(x+1/\psi)\right) + (F(x+1/\psi) - F(x)) \\
&\le \frac{1}{n}\sum_{i=1}^{n}(g_{\psi,x}(\widehat{\epsilon}_i) - g_{\psi,x}(\epsilon_i)) + \sup_{x\in\mathbf{R}}|\widetilde{F}(x) - F(x)| + (F(x+1/\psi) - F(x)) \\
&\le g_*\psi\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{\epsilon}_i - \epsilon_i)^2} + \sup_{x\in\mathbf{R}}|\widetilde{F}(x) - F(x)| + (F(x+1/\psi) - F(x)) \\
\widehat{F}(x) - F(x) &= \left(\widehat{F}(x) - \widetilde{F}(x-1/\psi)\right) + \left(\widetilde{F}(x-1/\psi) - F(x-1/\psi)\right) - (F(x) - F(x-1/\psi)) \\
&\ge \frac{1}{n}\sum_{i=1}^{n}(g_{\psi,x-1/\psi}(\widehat{\epsilon}_i) - g_{\psi,x-1/\psi}(\epsilon_i)) - \sup_{x\in\mathbf{R}}|\widetilde{F}(x) - F(x)| - (F(x) - F(x-1/\psi)) \\
&\ge -g_*\psi\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{\epsilon}_i - \epsilon_i)^2} - \sup_{x\in\mathbf{R}}|\widetilde{F}(x) - F(x)| - (F(x) - F(x-1/\psi)) \\
\Rightarrow \sup_{x\in\mathbf{R}}|\widehat{F}(x) - F(x)| &\le g_*\psi\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{\epsilon}_i - \epsilon_i)^2} + \sup_{x\in\mathbf{R}}|\widetilde{F}(x) - F(x)| + \sup_{x\in\mathbf{R}}|F(x+1/\psi) - F(x)|
\end{aligned}$$

Suppose assumption 1 to 6. From (B.7), (B.8), (B.9) and $\frac{1}{n}\sum_{i=1}^n \epsilon_i = O_p(1/\sqrt{n})$, for any $0 < a < 1$, $\exists$ a constant C_a such that with probability at least $1-a$

(D.3)
$$\frac{1}{n}\sum_{i=1}^n (\widehat{\epsilon}_i - \epsilon_i)^2 = \frac{1}{n}\sum_{i=1}^n \left(\sum_{j\notin\mathcal{N}_{b_n}} x_{ij}^{'}\theta_j - \sum_{j\in\mathcal{N}_{b_n}} x_{ij}^{'}(\widetilde{\theta}_j - \theta_j) - \frac{1}{n}\sum_{j=1}^n \epsilon_j \right)^2$$
$$\leq \frac{3}{n}\sum_{i=1}^n \left(\sum_{j\notin\mathcal{N}_{b_n}} x_{ij}^{'}\theta_j \right)^2 + \frac{3}{n}\sum_{i=1}^n \left(\sum_{j\in\mathcal{N}_{b_n}} x_{ij}^{'}(\widetilde{\theta}_j - \theta_j) \right)^2 + 3\left(\frac{1}{n}\sum_{j=1}^n \epsilon_j \right)^2$$
$$\leq \frac{6}{n}\sum_{i=1}^n \left(\sum_{j\notin\mathcal{N}_{b_n}} x_{ij}\theta_j \right)^2 + 6\left(\sum_{j\notin\mathcal{N}_{b_n}} \overline{x}_j\theta_j \right)^2 + \frac{6}{n}\sum_{i=1}^n \left(\sum_{j\in\mathcal{N}_{b_n}} x_{ij}(\widetilde{\theta}_j - \theta_j) \right)^2$$
$$+6\left(\sum_{j\in\mathcal{N}_{b_n}} \overline{x}_j(\widetilde{\theta}_j - \theta_j) \right)^2 + 3\left(\frac{1}{n}\sum_{j=1}^n \epsilon_j \right)^2$$
$$\leq C_a n^{-\alpha_\sigma} + \frac{6}{n^2}\left(\sum_{i=1}^n \sum_{j\notin\mathcal{N}_{b_n}} x_{ij}\theta_j \right)^2 + C_a|\mathcal{N}_{b_n}|(n^{2\alpha_\theta - 4\delta} + n^{-2\eta}) + \frac{6}{n^2}\left(\sum_{i=1}^n \sum_{j\in\mathcal{N}_{b_n}} x_{ij}(\widetilde{\theta}_j - \theta_j) \right)^2 + \frac{C_a}{n}$$
$$\leq C_a n^{-\alpha_\sigma} + \frac{6}{n}\sum_{i=1}^n \left(\sum_{j\notin\mathcal{N}_{b_n}} x_{ij}\theta_j \right)^2 + C_a|\mathcal{N}_{b_n}|(n^{2\alpha_\theta - 4\delta} + n^{-2\eta}) + \frac{6}{n}\sum_{i=1}^n \left(\sum_{j\in\mathcal{N}_{b_n}} x_{ij}(\widetilde{\theta}_j - \theta_j) \right)^2 + \frac{C_a}{n}$$
$$\Rightarrow \sqrt{\frac{1}{n}\sum_{i=1}^n (\widehat{\epsilon}_i - \epsilon_i)^2} = O_p(n^{-\alpha_\sigma/2})$$

According to Gilvenko-Cantelli lemma, $\sup_{x\in\mathbf{R}} |\widetilde{F}(x) - F(x)| \to 0$ almost surely. Therefore, for any $a > 0$ and sufficiently large n, $Prob\left(\sup_{x\in\mathbf{R}} |\widetilde{F}(x) - F(x)| \leq a\right) > 1 - a$. Choose sufficiently small a and $\psi = 1/a$, from assumption 8 and (D.3), we prove (35). $\square$

PROOF OF THEOREM 4. Define $X_f = (x_{f,ij})_{i=1,\ldots,p_1, j=1,\ldots,p}$. From theorem 1, since $p_1 = O(1)$,

$$\max_{i=1,2,\ldots,p_1} |\sum_{j=1}^p x_{f,ij}\widehat{\theta}_j - \sum_{j=1}^p x_{f,ij}\beta_j| = O_p(n^{-\eta}) \tag{D.4}$$

For any given $0 < a < 1$, choose a constant C_a such that $Prob\left(\max_{i=1,2,\ldots,p_1} |\sum_{j=1}^p x_{f,ij}\widehat{\theta}_j - \sum_{j=1}^p x_{f,ij}\beta_j| \leq C_a n^{-\eta}\right) \geq 1 - a$ for any $n = 1, 2, \ldots$. Define $F^-(x) = \lim_{y<x, y\to x} F(y)$ for any $x \in \mathbf{R}$, and $G(x) = Prob\left(\max_{i=1,2,\ldots,p_1} |\epsilon_{f,i}| \leq x\right) = (F(x) - F^-(-x))^{p_1}$ for $x \geq 0$. G is continuous according to assumption 8. With probability at least $1 - a$

$$\sup_{x\geq 0}|Prob^*\left(\max_{i=1,2,\ldots,p_1}|y_{f,i}-\sum_{j=1}^{p}x_{f,ij}\widehat{\theta}_j|\leq x\right)-G(x)|$$

$$\tag{D.5}\begin{aligned}&\leq \sup_{x\geq 0}|Prob^*\left(\max_{i=1,2,\ldots,p_1}|\epsilon_{f,i}|\leq x+\max_{i=1,2,\ldots,p_1}|\sum_{j=1}^{p}x_{f,ij}(\beta_j-\widehat{\theta}_j)|\right)-G(x)|\\&+\sup_{x\geq 0}|Prob^*\left(\max_{i=1,2,\ldots,p_1}|\epsilon_{f,i}|\leq x-\max_{i=1,2,\ldots,p_1}|\sum_{j=1}^{p}x_{f,ij}(\beta_j-\widehat{\theta}_j)|\right)-G(x)|\\&\leq \sup_{x\geq 0}|G(x+C_an^{-\eta})-G(x)|+\sup_{x\geq 0}|G(x)-G(\max(0,x-C_an^{-\eta}))|\end{aligned}$$

For any $\delta>0$ and any $x\geq 0$

$$\tag{D.6}\begin{aligned}&G(x+\delta)-G(x)\\&=\sum_{i=1}^{p_1}(F(x+\delta)-F(-x-\delta))^{i-1}\times(F(x)-F(-x))^{p_1-i}\times(F(x+\delta)-F(-x-\delta)-F(x)+F(-x))\\&\leq 2p_1\times\sup_{x\in\mathbf{R}}(F(x+\delta)-F(x))\Rightarrow\sup_{x\geq 0}(G(x+\delta)-G(x))\leq 2p_1\times\sup_{x\in\mathbf{R}}(F(x+\delta)-F(x))\end{aligned}$$

From (D.5) and assumption 8

$$\tag{D.7}\sup_{x\geq 0}|Prob^*\left(\max_{i=1,2,\ldots,p_1}|y_{f,i}-\sum_{j=1}^{p}x_{f,ij}\widehat{\theta}_j|\leq x\right)-G(x)|=o_p(1)$$

If $\widehat{\mathcal{N}}_{b_n}=\mathcal{N}_{b_n}$, $\frac{\sigma}{2}<\widehat{\sigma}<\frac{3\sigma}{2}$, and $\|\widehat{\theta}\|_2\leq C\times n^{\alpha_\theta}$ (see (C.1) to (C.3)), then

$$\tag{D.8}\begin{aligned}&\max_{i=1,2,\ldots,p}|\rho_n^2\sum_{j=1}^{r}\frac{q_{ij}\widehat{\zeta}_j}{(\lambda_j^2+\rho_n)^2}|\leq Cn^{-\eta-\delta_1}\\&\max_{i=1,2,\ldots,p}|\sum_{j=1}^{r}q_{ij}\left(\frac{\lambda_j}{\lambda_j^2+\rho_n}+\frac{\rho_n\lambda_j}{(\lambda_j^2+\rho_n)^2}\right)\sum_{l=1}^{n}p_{lj}\epsilon_l|\leq Cn^{\alpha_p/m-\eta}\end{aligned}$$

for some constant C with probability at least $1-a$. Here $2\delta=\eta+\alpha_\theta+\delta_1$. From (C.11), $\exists$ a constant E such that

$$\tag{D.9}Prob^*\left(\widehat{\mathcal{N}}^*_{b_n}\neq\mathcal{N}_{b_n}\right)\leq\frac{Ep}{n^{m\eta}b_n^m}$$

If $\widehat{\mathcal{N}}^*_{b_n}=\mathcal{N}_{b_n}$,

$$\tag{D.10}\begin{aligned}&|\sum_{j=1}^{p}x_{f,ij}\widehat{\theta}^*_j-\sum_{j=1}^{p}x_{f,ij}\widehat{\theta}_j|=|\sum_{j\in\mathcal{N}_{b_n}}x_{f,ij}(\widetilde{\theta}^*_j-\widehat{\theta}_j)|\\&\leq\rho_n^2|\sum_{k=1}^{r}\frac{c_{ik}\widetilde{\zeta}_k}{(\lambda_k^2+\rho_n)^2}|+|\sum_{k=1}^{r}\sum_{l=1}^{n}c_{ik}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)p_{lk}\epsilon_l^*|\\&\leq\rho_n^2\frac{\sqrt{C_{\mathcal{M}}}\|\widehat{\theta}\|_2}{\lambda_r^4}+|\sum_{k=1}^{r}\sum_{l=1}^{n}c_{ik}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)p_{lk}\epsilon_l^*|\end{aligned}$$

Form (B.5) and lemma A.1, $\exists$ a constant E which only depends on m, and for any $1 > a > 0$ with a sufficiently large $C_a > 0$

(D.11)

$$Prob^*\left(\max_{i=1,2,\ldots,p_1} |\sum_{k=1}^{r}\sum_{l=1}^{n} c_{ik}\left(\frac{\lambda_k}{\lambda_k^2+\rho_n}+\frac{\rho_n\lambda_k}{(\lambda_k^2+\rho_n)^2}\right)p_{lk}\frac{\epsilon_l^*}{\widehat{\sigma}}| > \frac{C_a n^{-\eta}}{\widehat{\sigma}}\right) \le \frac{p_1 E\widehat{\sigma}^m}{n^{m\eta}C_a^m n^{-m\eta}} < a$$

Combine with (D.9), there exists a constant C_a, with conditional probability at least $1-a$

(D.12)

$$\max_{i=1,2,\ldots,p_1} |\sum_{j=1}^{p} x_{f,ij}\widehat{\theta}_j^* - \sum_{j=1}^{p} x_{f,ij}\widehat{\theta}_j| \le C_a n^{-\eta}$$

$$\Rightarrow Prob^*\left(\max_{i=1,2,\ldots,p_1} |y_{f,i}^* - \widehat{y}_{f,i}^*| \le x\right) - G(x) \le a + Prob^*\left(\max_{i=1,2,\ldots,p_1} |\epsilon_{f,i}^*| \le x + C_a n^{-\eta}\right) - G(x)$$

$$\le a + \sup_{x\ge 0} |Prob^*\left(\max_{i=1,2,\ldots,p_1} |\epsilon_{f,i}^*| \le x\right) - G(x)| + 2p_1 \sup_{x\in\mathbf{R}}(F(x+C_a n^{-\eta}) - F(x))$$

$$Prob^*\left(\max_{i=1,2,\ldots,p_1} |y_{f,i}^* - \widehat{y}_{f,i}^*| \le x\right) - G(x) \ge -a + Prob^*\left(\max_{i=1,2,\ldots,p_1} |\epsilon_{f,i}^*| \le x - C_a n^{-\eta}\right) - G(x)$$

$$\ge -a + Prob^*\left(\max_{i=1,2,\ldots,p_1} |\epsilon_{f,i}^*| \le x - C_a n^{-\eta}\right) - G(x - C_a n^{-\eta}) - 2p_1 \sup_{x\in\mathbf{R}}(F(x+C_a n^{-\eta}) - F(x))$$

Since $G(x) = 0$ and $Prob^*\left(\max_{i=1,2,\ldots,p_1} |\epsilon_{f,i}^*| \le x\right) = 0$ if $x < 0$, we have

(D.13)

$$\sup_{x\ge 0} |Prob^*\left(\max_{i=1,2,\ldots,p_1} |y_{f,i}^* - \widehat{y}_{f,i}^*| \le x\right) - G(x)|$$

$$\le a + \sup_{x\ge 0} |Prob^*\left(\max_{i=1,2,\ldots,p_1} |\epsilon_{f,i}^*| \le x\right) - G(x)| + 2p_1 \sup_{x\in\mathbf{R}}(F(x+C_a n^{-\eta}) - F(x))$$

From lemma 1, for any $x \ge 0$,

(D.14)

$$|Prob^*\left(\max_{i=1,2,\ldots,p_1} |\epsilon_{f,i}^*| \le x\right) - G(x)| = |\left(\widehat{F}(x) - \widehat{F}^-(-x)\right)^{p_1} - (F(x) - F(-x))^{p_1}|$$

$$\le \sum_{i=1}^{p_1} |\widehat{F}(x) - \widehat{F}^-(-x)|^{i-1} \times |F(x) - F(-x)|^{p_1-i} \times \left(|\widehat{F}(x) - F(x)| + |\widehat{F}^-(-x) - F^-(-x)|\right)$$

$$\le 2p_1 \sup_{x\in\mathbf{R}} |\widehat{F}(x) - F(x)| \to_p 0$$

as $n \to \infty$. From theorem 1 and (C.1) to (C.3), for any $1 > a > 0$, with probability at least $1 - a$ $\exists$ a constant $C_a > 0$ such that for sufficiently large n, (D.8) happens with $C = C_a$ and $\sup_{x\ge 0} |Prob^*\left(\max_{i=1,2,\ldots,p_1} |\epsilon_{f,i}^*| \le x\right) - G(x)| < a$. Correspondingly for sufficiently large n, with probability at least $1-a$,

(D.15)

$$\sup_{x\geq 0}|Prob^*\left(\max_{i=1,2,\dots,p_1}|y^*_{f,i}-\widehat{y}^*_{f,i}|\leq x\right)-Prob^*\left(\max_{i=1,2,\dots,p_1}|y_{f,i}-\widehat{y}_{f,i}|\leq x\right)|$$

$$\leq \sup_{x\geq 0}|Prob^*\left(\max_{i=1,2,\dots,p_1}|y^*_{f,i}-\widehat{y}^*_{f,i}|\leq x\right)-G(x)|+\sup_{x\geq 0}|Prob^*\left(\max_{i=1,2,\dots,p_1}|y_{f,i}-\widehat{y}_{f,i}|\leq x\right)-G(x)|$$

$$\leq a+\sup_{x\geq 0}|Prob^*\left(\max_{i=1,2,\dots,p_1}|\epsilon^*_{f,i}|\leq x\right)-G(x)|+2p_1\sup_{x\in\mathbf{R}}(F(x+C_a n^{-\eta})-F(x))+a\leq 4a$$

and we prove (37).

For given $0<\alpha<1$ and sufficiently small $a>0$ such that $0<1-\alpha-a<1-\alpha+a<1$, define $c_{1-\alpha}$ as the $1-\alpha$ quantile of $G(x)$. For $G(x)$ is continuous, $G(c_{1-\alpha})=1-\alpha$. With probability at least $1-a$, $\sup_{x\geq 0}|Prob^*\left(\max_{i=1,2,\dots,p_1}|y^*_{f,i}-\widehat{y}^*_{f,i}|\leq x\right)-G(x)|<a/2$. Correspondingly with probability at least $1-a$

(D.16)

$$Prob^*\left(\max_{i=1,2,\dots,p_1}|y^*_{f,i}-\widehat{y}^*_{f,i}|\leq c_{1-\alpha+a}\right)\geq 1-\alpha+a/2\Rightarrow c^*_{1-\alpha}\leq c_{1-\alpha+a}$$

$$Prob^*\left(\max_{i=1,2,\dots,p_1}|y^*_{f,i}-\widehat{y}^*_{f,i}|\leq c_{1-\alpha-a}\right)\leq 1-\alpha-a/2\Rightarrow c^*_{1-\alpha}\geq c_{1-\alpha-a}$$

From (D.7), for sufficiently large n, with probability at least $1-a$

(D.17)

$$Prob^*\left(\max_{i=1,2,\dots,p_1}|y_{f,i}-\widehat{y}_{f,i}|\leq c^*_{1-\alpha}\right)\leq Prob^*\left(\max_{i=1,2,\dots,p_1}|y_{f,i}-\widehat{y}_{f,i}|\leq c_{1-\alpha+a}\right)\leq 1-\alpha+2a$$

$$Prob^*\left(\max_{i=1,2,\dots,p_1}|y_{f,i}-\widehat{y}_{f,i}|\leq c^*_{1-\alpha}\right)\geq Prob^*\left(\max_{i=1,2,\dots,p_1}|y_{f,i}-\widehat{y}_{f,i}|\leq c_{1-\alpha-a}\right)\geq 1-\alpha-2a$$

For $a>0$ can be arbitrarily small, we prove (38). □

REFERENCES

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819.

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47-70.

WHITTLE, P. (1960). Bounds for the Moments of Linear and Quadratic Forms in Independent Variables. *Theory Probab. Appl.* **5** 302-305.

XU, M., ZHANG, D. and WU, W. B. (2019). Pearson's chi-squared statistics: approximation theory and beyond. *Biometrika* **106** 716-723.