Supplementary Materials for "Asymptotic Normality and Confidence Intervals for Prediction Risks of the Min-Norm Least Squares Estimator "

Zeng Li [1] **Chuanlong Xie** [2] **Qinwen Wang** [3]

The Supplementary materials contain the proofs of Theorems 4.1–4.5 and some additional simulation experiments.

1. Proof of Theorem 4.1 and Theorem 4.2

Let $\mathbf{X} = \mathbf{Z}\boldsymbol{\Sigma}^{1/2}$. According to the Bai-Yin theorem (Bai & Yin, 2008), the smallest eigenvalue of $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}/n$ is almost surely larger than $(1-\sqrt{c})^2/2$ for sufficiently large n. Thus

$$\lambda_{min}(\frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{X}) \geq c_0\lambda_{\min}(\frac{1}{n}\mathbf{Z}^{\mathrm{T}}\mathbf{Z}) \geq \frac{c_0}{2}(1-\sqrt{c})^2,$$

which implies that the matrix $\mathbf{X}^{\mathrm{T}}\mathbf{X}/n$ is almost surely invertible for large n. By Section 3.2,

$$\begin{aligned} B_{\mathbf{X}}(\hat{\boldsymbol{\beta}},\boldsymbol{\beta}) &= B_{\mathbf{X},\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}},\boldsymbol{\beta}) = 0 \\ V_{\mathbf{X}}(\hat{\boldsymbol{\beta}},\boldsymbol{\beta}) &= V_{\mathbf{X},\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}},\boldsymbol{\beta}). \end{aligned}$$

The first equality holds since $\boldsymbol{\Pi} = \mathbf{0}$. Thus the asymptotic of $R_{\mathbf{X}}(\hat{\boldsymbol{\beta}},\boldsymbol{\beta})$ is same to that of $R_{\mathbf{X},\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}},\boldsymbol{\beta})$. For simplicity, we focus on $R_{\mathbf{X}}(\hat{\boldsymbol{\beta}},\boldsymbol{\beta})$ in the following. Notice that

$$\begin{aligned} V_{\mathbf{X}}(\hat{\boldsymbol{\beta}},\boldsymbol{\beta}) &= \frac{\sigma^2}{n}\operatorname{Tr}(\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}) \\ &= \frac{\sigma^2}{n}\operatorname{Tr}\left(\boldsymbol{\Sigma}^{-1/2}(\frac{\mathbf{Z}^{\mathrm{T}}\mathbf{Z}}{n})^{-1}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}\right) \\ &= \frac{\sigma^2}{n}\sum_{i=1}^{p}\frac{1}{s_i} = \frac{\sigma^2 p}{n}\int\frac{1}{s}dF_{\mathbf{Z}}(s) \end{aligned}$$

where s_i's are eigenvalues of $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}/n$. $F_{\mathbf{Z}}$ is the spectral measure of $\mathbf{Z}^{\mathrm{T}}\mathbf{Z}/n$. According to the convergence of empirical spectral distributions of sample covariance matrices $F_{\mathbf{Z}}$ established in Yin (1986), as $n,p\to\infty$ such that $p/n = c_n \to c \in (0,\infty)$, $F_{\mathbf{Z}}(x)$ weakly converges to the standard Marcenko-Pastur law $F_c(x)$ and

$$V_{\mathbf{X}}(\hat{\boldsymbol{\beta}},\boldsymbol{\beta}) \to \sigma^2 c\int\frac{1}{s}dF_c(s) = \sigma^2\frac{c}{1-c}.$$

[1]First author, Department of Statistics and Data Science, Southern University of Science and Technology, China [2]Noah Ark Lab, Huawei Technology, Hong Kong [3]Corresponding author, School of Data Science, Fudan University, China. Correspondence to: Qinwen Wang <wqw@fudan.edu.cn>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

Here the standard Marcenko-Pastur law $F_c(x)$ has a density function

$$p_c(x) = \begin{cases} \frac{1}{2\pi cx}\sqrt{(b-x)(x-a)}, & \text{if } a\leq x\leq b, \\ 0, & \text{o.w.}, \end{cases}$$

where $a = (1-\sqrt{c})^2$, $b = (1+\sqrt{c})^2$ and $p_c(x)$ has a point mass $1-\frac{1}{c}$ at the origin if $c > 1$. Hence

$$\begin{aligned} & R_{\mathbf{X}}(\hat{\boldsymbol{\beta}},\boldsymbol{\beta}) - \sigma^2\frac{c_n}{1-c_n} \\ =\; & \frac{\sigma^2 p}{n}\int\frac{1}{s}dF_Z(s) - \sigma^2 c_n\int\frac{1}{s}dF_{c_n}(s) \\ =\; & \sigma^2 c_n\int\frac{1}{s}\big(dF_Z(s) - dF_{c_n}(s)\big). \end{aligned}$$

According to Theorem 1.1 of Bai & Silverstein (2004),

$$p\Big(R_{\mathbf{X}}(\hat{\boldsymbol{\beta}},\boldsymbol{\beta}) - \sigma^2\frac{c_n}{1-c_n}\Big) \xrightarrow{d} N(\mu_c,\sigma_c^2), \qquad (1)$$

where

$$\begin{aligned} \mu_c &= -\frac{\sigma^2 c}{2\pi i}\oint_{\mathcal{C}}\frac{1}{z}\frac{c\underline{m}(z)^3(1+\underline{m}(z))^{-3}}{\{1-c\underline{m}(z)^2(1+\underline{m}(z))^{-2}\}^2}dz \qquad (2) \\ &\quad -\frac{\sigma^2 c(\nu_4-3)}{2\pi i}\oint_{\mathcal{C}}\frac{1}{z}\frac{c\underline{m}(z)^3(1+\underline{m}(z))^{-3}}{1-c\underline{m}(z)^2(1+\underline{m}(z))^{-2}}dz, \\ \sigma_c^2 &= -\frac{\sigma^4 c^2}{2\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2}\frac{1}{z_1 z_2}\frac{1}{(\underline{m}(z_1)-\underline{m}(z_2))^2} \qquad (3) \\ &\quad\times\frac{d}{dz_1}\underline{m}(z_1)\frac{d}{dz_2}\underline{m}(z_2)dz_1dz_2 \\ &\quad -\frac{\sigma^4 c^3(\nu_4-3)}{4\pi^2}\oint_{\mathcal{C}_1}\oint_{\mathcal{C}_2}\frac{1}{z_1 z_2}\frac{1}{(1+\underline{m}(z_1))^2} \\ &\quad\times\frac{1}{(1+\underline{m}(z_2))^2}d\underline{m}(z_1)d\underline{m}(z_2). \end{aligned}$$

Here the contours in (2) and (3) are closed and taken in the positive direction in the complex plane, enclosing the support of $F^{c,H}$. The Stieltjes transform $\underline{m}(z)$ satisfies the equation

$$z = -\frac{1}{\underline{m}} + \frac{c}{1+\underline{m}}.$$

To further simplify the integrations in μ_c and σ_c, let $z = 1+\sqrt{c}(r\xi+\frac{1}{r\xi})+c$ and perform change of variables, then we have

$$\begin{aligned}\underline{m}(z) &= -\frac{1}{1+\sqrt{c}r\xi}, \\ dz &= \sqrt{c}(r-\frac{1}{r\xi^2})d\xi, \\ d\underline{m} &= \frac{\sqrt{c}r}{(1+\sqrt{c}r\xi)^2}d\xi\end{aligned}$$

and when ξ moves along the unit circle $|\xi|=1$ on the complex plane, z will orbit around the center point $1+c$ along an ellipse which enclosing the support of $F^{c,H}$. Thus

$$\begin{aligned}\mu_c &= -\frac{\sigma^2 c}{2\pi i}\oint_{|\xi|=1}\frac{c\underline{m}^3(1+\underline{m})}{z\left\{(1+\underline{m})^2-c\underline{m}^2\right\}^2} \\ &\quad \times\sqrt{c}(r-\frac{1}{r\xi^2})d\xi \\ &= \frac{\sigma^2 c}{2\pi i}\oint_{|\xi|=1}\frac{1}{r(\sqrt{c}+r\xi)(1+\sqrt{c}r\xi)} \\ &\quad \times\frac{1}{(\xi-\frac{1}{r})(\xi+\frac{1}{r})}d\xi \\ &\quad +\frac{\sigma^2 c(\nu_4-3)}{2\pi i}\oint_{|\xi|=1}\frac{1}{r\xi^2(\sqrt{c}+r\xi)} \\ &\quad \times\frac{1}{(1+\sqrt{c}r\xi)}d\xi \\ &= \frac{\sigma^2c^2}{(c-1)^2}+\frac{\sigma^2c^2(\nu_4-3)}{1-c}.\end{aligned}$$

As for σ_c^2, note that

$$\begin{aligned}&\frac{1}{2\pi i}\oint_{\mathcal{C}_1}\frac{1}{z_1(\underline{m}_1-\underline{m}_2)^2}d\underline{m}_1 \\ =\ &\frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{1}{1+\sqrt{c}(r_1\xi_1+\frac{1}{r_1\xi_1})+c} \\ &\times\frac{\sqrt{c}\,r_1}{(\underline{m}_2+\frac{1}{1+\sqrt{c}r_1\xi_1})^2(1+\sqrt{c}r_1\xi_1)^2}d\xi_1 \\ =\ &\frac{1}{2\pi i}\oint_{|\xi_1|=1}\frac{\sqrt{c}\,r_1\xi_1}{(\xi_1+\frac{\sqrt{c}}{r_1})(r_1\xi_1\sqrt{c}+1)} \\ &\times\frac{1}{\left((r_1\xi_1\sqrt{c}+1)\underline{m}_2+1\right)^2}d\xi_1 \\ =\ &\frac{c}{(c-1)\left((c-1)\underline{m}_2-1\right)^2},\end{aligned}$$

therefore

$$\begin{aligned}&-\frac{\sigma^4c^2}{2\pi^2}\oiint\frac{1}{z_1z_2(\underline{m}_1-\underline{m}_2)^2}d\underline{m}_1d\underline{m}_2 \\ =\ &\frac{2\sigma^4c^2}{2\pi i}\oint_{|\xi_2|=1}\frac{c}{z_2(c-1)\left\{(c-1)\underline{m}_2-1\right\}^2}d\underline{m}_2 \\ =\ &\frac{2\sigma^4c^2}{2\pi i}\oint_{|\xi_2|=1}\frac{\sqrt{c}\,r_2^2\xi_2}{(c-1)(1+\sqrt{c}\,r_2\xi_2)(\sqrt{c}+r_2\xi_2)^3}d\xi_2 \\ =\ &\frac{2c^3\sigma^4}{(c-1)^4}.\end{aligned}$$

Meanwhile,

$$\begin{aligned}&\frac{1}{2\pi i}\oint_{\mathcal{C}_1}\frac{1}{z_1}\frac{1}{(1+\underline{m}(z_1))^2}d\underline{m}(z_1) \\ =\ &\frac{1}{2\pi i}\oint_{|\xi|=1}\frac{1}{\sqrt{c}\xi(1+\sqrt{c}r\xi)(\sqrt{c}+r\xi)}d\xi \\ =\ &\frac{1}{c-1},\end{aligned}$$

hence

$$\begin{aligned}&-\frac{\sigma^4c^3(\nu_4-3)}{4\pi^2}\Big(\oint_{\mathcal{C}_1}\frac{1}{z_1}\frac{1}{(1+\underline{m}(z_1))^2}d\underline{m}(z_1)\Big)^2 \\ =\ &\sigma^4c^3(\nu_4-3)\times\frac{1}{(1-c)^2}.\end{aligned}$$

Then we have,

$$\sigma_c^2=\frac{2c^3\sigma^4}{(c-1)^4}+\frac{\sigma^4c^3(\nu_4-3)}{(1-c)^2}.$$

Let

$$T_n=\frac{p}{\sigma_c}\Big(R_{\mathbf{X}}(\hat{\boldsymbol{\beta}},\boldsymbol{\beta})-\sigma^2\frac{c_n}{1-c_n}-\frac{\mu_c}{p}\Big).$$

According to (1), we have

$$\begin{aligned}&P(L_{\alpha,c}\le R_{\mathbf{X},\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}},\boldsymbol{\beta})\le U_{\alpha,c}) \\ =\ &P(-Z_{\alpha/2}\le T_n\le Z_{\alpha/2}) \\ \to\ &1-\alpha,\end{aligned}$$

where

$$\begin{aligned}L_{\alpha,c} &= \sigma^2\frac{c_n}{1-c_n}+\frac{1}{p}(\mu_c-Z_{\alpha/2}\sigma_c), \\ U_{\alpha,c} &= \sigma^2\frac{c_n}{1-c_n}+\frac{1}{p}(\mu_c+Z_{\alpha/2}\sigma_c).\end{aligned}$$

□

2. Proof of Theorem 4.3

Notice that

$$\begin{aligned}B_{\mathbf{X}}(\hat{\boldsymbol{\beta}},\boldsymbol{\beta}) &= \boldsymbol{\beta}^{\mathrm{T}}(\boldsymbol{I}_p-\hat{\boldsymbol{\Sigma}}^{+}\hat{\boldsymbol{\Sigma}})\boldsymbol{\beta} \\ &= \lim_{z\to0^+}\boldsymbol{\beta}^{\mathrm{T}}\big(\boldsymbol{I}_p-(\hat{\boldsymbol{\Sigma}}+z\boldsymbol{I}_p)^{-1}\hat{\boldsymbol{\Sigma}}\big)\boldsymbol{\beta} \\ &= \lim_{z\to0^+}z\boldsymbol{\beta}^{\mathrm{T}}(\hat{\boldsymbol{\Sigma}}+z\boldsymbol{I}_p)^{-1}\boldsymbol{\beta}.\end{aligned}$$

Since $\boldsymbol{\beta}$ is a constant vector, we can make use of the results in Theorem 3 in Bai et al. (2007) and Theorem 1.3 in Pan & Zhou (2008) regarding eigenvectors. Their works investigate the sample covariance matrix

$$\mathbf{A}_p = \boldsymbol{T}_p^{1/2}\mathbf{X}_p^{\mathrm{T}}\mathbf{X}_p\boldsymbol{T}_p^{1/2}/n,$$

where $\boldsymbol{T}_p$ is an $p \times p$ nonnegative definite Hermitian matrix with a square root $\boldsymbol{T}_p^{1/2}$ and $\mathbf{X}_p$ is an $n \times p$ matrix with i.i.d. entries $(\mathrm{x}_{ij})_{n\times p}$. Let $\boldsymbol{U}_p\boldsymbol{\Lambda}_p\boldsymbol{U}_p^{\mathrm{T}}$ denote the spectral decomposition of $\mathbf{A}_p$ where $\boldsymbol{\Lambda}_p = \mathrm{diag}(\lambda_1, \cdots, \lambda_p)$ and $\boldsymbol{U}_p$ is a unitary matrix consisting of the orthonormal eigenvectors of $\mathbf{A}_p$. Assume that $\boldsymbol{x}_p$ is an arbitrary nonrandom unit vector and $\boldsymbol{y} = (y_1, y_2, \cdots, y_p)^{\mathrm{T}} = \boldsymbol{U}_p^{\mathrm{T}}\boldsymbol{x}_p$, two empirical distribution functions based on eigenvectors and eigenvalues are defined as

$$\begin{aligned} F_1^{\mathbf{A}_p}(x) &= \sum_{i=1}^{p} |y_i|^2 \mathbb{1}(\lambda_i \le x), \\ F^{\mathbf{A}_p}(x) &= \frac{1}{p}\sum_{i=1}^{p} \mathbb{1}(\lambda_i \le x). \end{aligned}$$

Then for a bounded continuous function $g(x)$, we have

$$\begin{aligned} &\sum_{j=1}^{p} |\mathrm{y}_j|^2 g(\lambda_j) - \frac{1}{p}\sum_{j=1}^{p} g(\lambda_j) \\ =& \int g(x)dF_1^{\boldsymbol{A}_p}(x) - \int g(x)dF^{\boldsymbol{A}_p}(x). \end{aligned}$$

The results in Bai et al. (2007) and Pan & Zhou (2008) are summarized in the following lemma.

Lemma 2.1. (Theorem 3 (Bai et al., 2007) and Theorem 1.3 (Pan & Zhou, 2008)) *Suppose that*

(1) x_{ij}*'s are i.i.d. satisfying* $\mathbb{E}(\mathrm{x}_{ij}) = 0$, $\mathbb{E}(|\mathrm{x}_{ij}|^2) = 1$ *and* $\mathbb{E}(|\mathrm{x}_{ij}|^4) < \infty$*;*

(2) $\boldsymbol{x}_p \in \mathbb{C}^p$, $\|\boldsymbol{x}_p\| = 1$, $\lim_{n,p\to\infty} p/n = c \in (0,\infty)$*;*

(3) $\boldsymbol{T}_p$ *is nonrandom Hermitian non-negative definite with with its spectral norm bounded in* p*, with* $H_p = F^{\boldsymbol{T}_p} \xrightarrow{d} H$ *a proper distribution function and* $\boldsymbol{x}_p^{\mathrm{T}}(\boldsymbol{T}_p - z\boldsymbol{I}_p)^{-1}\boldsymbol{x}_p \to m_{F^H}(z)$*, where* $m_{F^H}(z)$ *denotes the Stieltjes transform of* $H(t)$*;*

(4) $g_1, \cdots, g_k$ *are analytic functions on an open region of the complex plain which contains the real interval*

$$\Big[\liminf_p \lambda_{min}(\boldsymbol{T}_p)\mathbb{1}_{(0,1)}(c)(1-\sqrt{c})^2, \limsup_p \lambda_{max}(\boldsymbol{T}_p)\mathbb{1}_{(0,1)}(c)(1+\sqrt{c})^2\Big];$$

(5) as $n, p \to \infty$*,*

$$\sup_z \sqrt{n}\Big\|\boldsymbol{x}_p^{\mathrm{T}}\big(\underline{m}_{F^{c_n,H_p}}(z)\boldsymbol{T}_p - \boldsymbol{I}_p\big)^{-1}\boldsymbol{x}_p - \int \frac{1}{1+t\underline{m}_{F^{c_n,H_p}}(z)}dH_n(t)\Big\| \to 0.$$

Define $G_p(x) = \sqrt{n}(F_1^{\boldsymbol{A}_p}(x) - F^{\boldsymbol{A}_p}(x))$*, then the random vectors*

$$\Big(\int g_1(x)dG_p(x), \cdots, \int g_k(x)dG_p(x)\Big)$$

forms a tight sequence and converges weakly to a Gaussian vector $\mathrm{x}_{g_1}, \cdots, \mathrm{x}_{g_k}$ *with mean zero and covariance function*

$$\begin{aligned} \mathrm{Cov}(\mathrm{x}_{g_1}, \mathrm{x}_{g_2}) =& -\frac{1}{2\pi^2}\int_{\mathcal{C}_1}\int_{\mathcal{C}_2} g_1(z_1)g_2(z_2) \\ &\times \frac{(z_2\underline{m}_2 - z_1\underline{m}_1)^2}{c^2 z_1 z_2 (z_2 - z_1)(\underline{m}_2 - \underline{m}_1)}dz_1dz_2. \end{aligned}$$

The contours $\mathcal{C}_1$*,* $\mathcal{C}_2$ *are disjoint, both contained in the analytic region for the functions* $(g_1, \cdots, g_k)$ *and enclose the support of* F^{c_n,H_p} *for all large* p*.*

(6) If $H(x)$ *satisfies*

$$\begin{aligned} &\int \frac{dH(t)}{(1+t\underline{m}(z_1))(1+t\underline{m}(z_2))} \\ =& \int \frac{dH(t)}{1+t\underline{m}(z_1)}\int \frac{dH(t)}{1+t\underline{m}(z_2)}, \end{aligned}$$

then the covariance function can be further simplified to

$$\begin{aligned} &\mathrm{Cov}(\mathrm{x}_{g_1}, \mathrm{x}_{g_2}) \\ =& \frac{2}{c}\Big(\int g_1(x)g_2(x)dF^{c,H}(x) \\ &- \int g_1(x)dF^{c,H}(x)\int g_2(x)dF^{c,H}(x)\Big). \end{aligned}$$

Recall that $B_{\mathbf{X}}(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}) = \lim_{z\to 0^+} z\boldsymbol{\beta}^{\mathrm{T}}(\hat{\boldsymbol{\Sigma}} + z\boldsymbol{I}_p)^{-1}\boldsymbol{\beta}$. Let $g(x) = 1/(x+z)$ and $\boldsymbol{x}_p = \boldsymbol{\beta}/r$. Then we have

$$\begin{aligned} &\int g(x)dG_n(x) \\ =& \sqrt{n}\Big(\frac{1}{r^2}\boldsymbol{\beta}^{\mathrm{T}}(\hat{\boldsymbol{\Sigma}} + z\boldsymbol{I}_p)^{-1}\boldsymbol{\beta} - \int g(x)dF_{c_n}(x)\Big), \end{aligned}$$

where $F_{c_n}(x)$ is the standard Marcenko-Pastur law with parameter c_n. It is not difficult to check that under Assumptions (A1), (B1) and (C1), all the conditions *(1)-(6)* in Lemma 2.1 are satisfied.

To proceed further, denote $a = (1-\sqrt{c})^2$, $b = (1+\sqrt{c})^2$. If c is replaced by c_n, a and b are denoted by a_n and b_n

respectively. By some algebraic calculations, we have

$$\begin{aligned}
&\int g(x)dF_{c_n}(x)\\
=\ & (1-\frac{1}{c_n})\cdot\frac{1}{z}\\
&+\int_{a_n}^{b_n}\frac{1}{x+z}\cdot\frac{1}{2\pi c_n x}\sqrt{(b_n-x)(x-a_n)}dx\\
=\ & (1-\frac{1}{c_n})\cdot\frac{1}{z}\\
&-\frac{-1+c_n+z-\sqrt{c_n^2+2c_n(z-1)+(1+z)^2}}{2c_n z},
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{Var}(\mathrm{x}_g)\\
=\ & \frac{2}{c}\Big(\int\{g(x)\}^2dF_c(x)-\Big\{\int g(x)dF_c(x)\Big\}^2\Big)\\
=\ & \frac{2}{c}\Big\{(1-\frac{1}{c})\cdot\frac{1}{z^2}\\
&+\int_a^b\frac{1}{(x+z)^2}\cdot\frac{1}{2\pi cx}\sqrt{(b-x)(x-a)}dx\Big\}\\
&-\frac{2}{c}\Big\{(1-\frac{1}{c})\cdot\frac{1}{z}\\
&+\int_a^b\frac{1}{x+z}\cdot\frac{1}{2\pi cx}\sqrt{(b-x)(x-a)}dx\Big\}^2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\lim_{z\to0^+} z\int g(x)dF_{c_n}(x) &= 1-\frac{1}{c_n},\\
\lim_{z\to0^+} z^2\mathrm{Var}(\mathrm{x}_g) &= \frac{2(c-1)}{c^3}.
\end{aligned}$$

Furthermore, as $n,p\to\infty$, $p/n=c_n\to c>1$,

$$\sqrt{n}\Big(B_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)-(1-\frac{1}{c_n})r^2\Big)\xrightarrow{d}N\Big(0,\frac{2(c-1)}{c^3}r^4\Big).$$

This can be rewritten as

$$\sqrt{p}\Big(B_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)-(1-\frac{1}{c_n})r^2\Big)\xrightarrow{d}N\Big(0,\frac{2(c-1)}{c^2}r^4\Big).$$

Next we deal with the variance term $V_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)$. According to the Assumption *(B1)*, the variance term is

$$V_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)=\frac{\sigma^2}{n}\mathrm{Tr}\{\hat{\boldsymbol\Sigma}^+\}=\frac{\sigma^2}{n}\sum_{i=1}^n\frac{1}{s_i},$$

where $s_i,\ i=1,\dots,n$ are the nonzero eigenvalues of $\mathbf{X}^{\mathrm{T}}\mathbf{X}/n$. Let $\{t_i,\ i=1,\dots n\}$ denote the non-zero eigenvalues of $\mathbf{X}\mathbf{X}^{\mathrm{T}}/p$, then we have

$$\begin{aligned}
V_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta) &= \frac{\sigma^2}{p}\sum_{i=1}^n\frac{1}{t_i}\\
&= \frac{\sigma^2 n}{p}\int\frac{1}{t}dF_{\mathbf{X}\mathbf{X}^{\mathrm{T}}/p}(t)\\
&\to \frac{\sigma^2}{c-1}.
\end{aligned}$$

By interchanging the role of p and n, from the result in Theorem 4.1, as $n,p\to\infty$, $p/n=c_n\to c>1$, we have that the term

$$\sum_{i=1}^n\frac{1}{t_i}-\frac{n}{1-c'_n}$$

weakly converges to a normal distribution:

$$N\Big(\frac{c'}{(c'-1)^2}+\frac{c'(\nu_4-3)}{1-c'},\ \frac{2c'}{(c'-1)^4}+\frac{c'(\nu_4-3)}{(1-c')^2}\Big),$$

where $c'_n=n/p=1/c_n$, $c'=1/c$. This result can be rewritten as

$$\begin{aligned}
&\sum_{i=1}^n\frac{1}{t_i}-\frac{p}{c_n-1}\\
\xrightarrow{d}\ & N\Big(\frac{c}{(1-c)^2}+\frac{(\nu_4-3)}{c-1},\ \frac{2c^3}{(1-c)^4}+\frac{c(\nu_4-3)}{(c-1)^2}\Big).
\end{aligned}$$

Hence the CLT of $V_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)$ is given by

$$\begin{aligned}
&p\Big(V_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)-\frac{\sigma^2}{c_n-1}\Big)\\
\xrightarrow{d}\ & N\Big(\frac{c\sigma^2}{(1-c)^2}+\frac{\sigma^2(\nu_4-3)}{c-1},\ \frac{2c^3\sigma^4}{(1-c)^4}+\frac{c\sigma^4(\nu_4-3)}{(c-1)^2}\Big).
\end{aligned}$$

Notice that $\mathrm{Cov}\Big(B_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta),\ V_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)\Big)=0$. According to the consistency rate and the limiting distribution of $B_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)$ and $V_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)$, we know that the bias $B_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)$ is the leading term of $R_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)$. This implies that

$$\sqrt{p}\Big\{R_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)-(1-\frac{1}{c_n})\|\boldsymbol\beta\|_2^2-\frac{\sigma^2}{c_n-1}\Big\}\xrightarrow{d}N\big(0,\sigma_{c,1}^2\big),$$

where $\sigma_{c,1}^2=2(c-1)r^4/c^2$. A practical version of this CLT is given by

$$\sqrt{p}\Big\{R_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)-(1-\frac{1}{c_n})\|\boldsymbol\beta\|_2^2-\frac{\sigma^2}{c_n-1}\Big\}\xrightarrow{d}N\big(\tilde\mu_{c,1},\tilde\sigma_{c,1}^2\big),$$

where

$$\begin{aligned}
\tilde\mu_{c,1} &= \frac{1}{\sqrt{p}}\Big\{\frac{c\sigma^2}{(1-c)^2}+\frac{\sigma^2(\nu_4-3)}{c-1}\Big\},\\
\tilde\sigma_{c,1}^2 &= \frac{2(c-1)}{c^2}r^4+\frac{1}{p}\Big\{\frac{2c^3\sigma^4}{(1-c)^4}+\frac{c\sigma^4(\nu_4-3)}{(c-1)^2}\Big\}.
\end{aligned}$$

3. Proof of Theorem 4.4

First we consider the bias term $B_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)$. By Assumption (A1), (B1), and (C2),

$$\begin{aligned} B_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta) &= \mathbb{E}[\boldsymbol\beta^{\mathrm T}\Pi\boldsymbol\Sigma\Pi\boldsymbol\beta|\mathbf{X}] = \mathbb{E}[\boldsymbol\beta^{\mathrm T}\Pi\boldsymbol\beta|\mathbf{X}] \\ &= \mathrm{Tr}\left\{(\boldsymbol I_p - \hat{\boldsymbol\Sigma}^{+}\hat{\boldsymbol\Sigma})\mathbb{E}(\boldsymbol\beta\boldsymbol\beta^{\mathrm T}|\mathbf{X})\right\} \\ &= \frac{r^2}{p}\,\mathrm{Tr}\{\boldsymbol I_p - \hat{\boldsymbol\Sigma}^{+}\hat{\boldsymbol\Sigma}\} = r^2(1-n/p). \end{aligned}$$

Alternatively, we can rewrite the bias as

$$\begin{aligned} B_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta) &= \lim_{z\to0^+}\mathbb{E}[\boldsymbol\beta^{\mathrm T}(\boldsymbol I_p-(\hat{\boldsymbol\Sigma}+z\boldsymbol I_p)^{-1}\hat{\boldsymbol\Sigma})\boldsymbol\beta|\mathbf{X}] \\ &= \lim_{z\to0^+}\mathbb{E}[z\boldsymbol\beta^{\mathrm T}(\hat{\boldsymbol\Sigma}+z\boldsymbol I_p)^{-1}\boldsymbol\beta|\mathbf{X}] \\ &= \lim_{z\to0^+} z\frac{r^2}{p}\,\mathrm{Tr}(\hat{\boldsymbol\Sigma}+z\boldsymbol I_p)^{-1}. \end{aligned}$$

Define that $f_n(z) = z\frac{r^2}{p}\,\mathrm{Tr}(\hat{\boldsymbol\Sigma}+z\boldsymbol I_p)^{-1}$. Notice that $|f_n(z)|$ and $|f_n'(z)|$ are bounded above. By the Arzela-Ascoli theorem, we deduce that $f_n(z)$ converges uniformly to its limit. Under Assumption *(C2)*, by the Moore-Osgood theorem, almost surely,

$$\begin{aligned} &\lim_{n,p\to\infty} B_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta) \\ &= \lim_{z\to0^+}\lim_{n,p\to\infty} z\frac{r^2}{p}\,\mathrm{Tr}(\hat{\boldsymbol\Sigma}+z\boldsymbol I_p)^{-1} \\ &= \lim_{z\to0^+}\lim_{n,p\to\infty} z\frac{r^2}{p}\,\mathrm{Tr}\left(\frac{1}{n}\mathbf{X}\mathbf{X}^{\mathrm T}+z\boldsymbol I_n\right)^{-1}, \end{aligned}$$

In fact,

$$\lim_{n,p\to\infty} B_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta) = r^2\lim_{z\to0^+}\lim_{n,p\to\infty} z m_n(-z),$$

where $m_n(z)$ is the Stieltjes transform of empirical spectral distribution of $\hat{\boldsymbol\Sigma}=\mathbf{X}^{\mathrm T}\mathbf{X}/n$. According to Theorem 2.1 in (Zheng et al., 2015) and Lemma 1.1 in (Bai & Silverstein, 2004), the truncated version of $p(m_n(z)-m(z))$ converges weakly to a two-dimensional Gaussian process $M(\cdot)$ satisfying

$$\begin{aligned} \mathbb{E}[M(z)] &= \frac{c\underline{m}^3(1+\underline{m})}{\{(1+\underline{m})^2-c\underline{m}^2\}^2} \\ &\quad+\frac{c(\nu_4-3)\underline{m}^3}{(1+\underline{m})\{(1+\underline{m})^2-c\underline{m}^2\}}, \end{aligned}$$

and

$$\begin{aligned} &\mathrm{Cov}\big(M(z_1),M(z_2)\big) \\ &= 2\Big\{\frac{\underline{m}'(z_1)\underline{m}'(z_2)}{(\underline{m}(z_1)-\underline{m}(z_2))^2}-\frac{1}{(z_1-z_2)^2}\Big\} \\ &\quad+\frac{c(\nu_4-3)\underline{m}'(z_1)\underline{m}'(z_2)}{(1+\underline{m}(z_1))^2(1+\underline{m}(z_2))^2}, \end{aligned}$$

where $\underline{m}=\underline{m}(z)$ represents the Stieltjes transform of limiting spectral distribution of companion matrix $\mathbf{X}\mathbf{X}^{\mathrm T}/n$ satisfying the equation

$$z=-\frac{1}{\underline{m}}+\frac{c}{1+\underline{m}},\quad \underline{m}(z)=-\frac{1-c}{z}+cm(z).$$

When $p>n$, we can actually solve $\underline{m}(z)$ equation and obtain that

$$\begin{aligned} \underline{m}(z) &= \frac{-1+c-z+\sqrt{-4z+(1-c+z)^2}}{2z}, \\ m(z) &= \frac{1-c-z+\sqrt{-4z+(1-c+z)^2}}{2cz}. \end{aligned}$$

Therefore, by some algebraic calculations, we have

$$\begin{aligned} &\lim_{n,p\to\infty} B_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta) \\ &= \lim_{n,p\to\infty} r^2\lim_{z\to0^+} z m_n(-z) \\ &= r^2\lim_{z\to0^+}\Big\{zm(-z)+z(1-\frac{1}{c})\frac{1}{z}\Big\} \\ &= \lim_{n,p\to\infty} r^2\lim_{z\to0^+} z\frac{n}{p}\underline{m}_n(z) \\ &= r^2\frac{1}{c}\lim_{z\to0^+} z\underline{m}(-z) \\ &= r^2(1-\frac{1}{c}). \end{aligned}$$

Moreover,

$$\begin{aligned} \mathrm{Var}\big(M(z)\big) &= \lim_{z_1\to z_2=z}\mathrm{Cov}\big(M(z_1),M(z_2)\big) \\ &= \frac{2\underline{m}'(z)\underline{m}'''(z)-3(\underline{m}''(z))^2}{6(\underline{m}'(z))^2} \\ &\quad+\frac{c(\nu_4-3)(\underline{m}'(z))^2}{(1+\underline{m}(z))^4}. \end{aligned}$$

By substituting of the explicit form of $\underline{m}(z)$, we can easily derive that

$$\lim_{z\to0^+} z\mathbb{E}[M(-z)]=0,\quad \lim_{z\to0^+} z^2\mathrm{Var}(M(-z))=0,$$

which means that the second-order limit of $B_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)$ is still $r^2(1-1/c)$. All in all, $B_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)$ is identical with a constant $r^2(1-1/c)$ in distribution.

On the other hand, by Assumption (B1),

$$V_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)=\frac{\sigma^2}{n}\,\mathrm{Tr}\{\hat{\boldsymbol\Sigma}^{+}\}=\frac{\sigma^2}{n}\sum_{i=1}^{n}\frac{1}{s_i},$$

where s_i, $i=1,\dots,n$ are the nonzero eigenvalues of $\mathbf{X}^{\mathrm T}\mathbf{X}/n$. Similar to the proof of Theorem 4.3, the CLT of $V_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)$ is given by

$$\begin{aligned} &p\Big(V_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)-\frac{\sigma^2}{c_n-1}\Big) \\ &\xrightarrow{d} N\Big(\frac{c\sigma^2}{(1-c)^2}+\frac{\sigma^2(\nu_4-3)}{c-1},\ \frac{2c^3\sigma^4}{(1-c)^4}+\frac{c\sigma^4(\nu_4-3)}{(c-1)^2}\Big). \end{aligned}$$

Combining the results of $B_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)$ and $V_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)$, we have

$$p\Big\{R_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)-r^2(1-\frac{1}{c_n})-\frac{\sigma^2}{c_n-1}\Big\}\xrightarrow{d}N(\mu_{c,2},\sigma_{c,2}^2),$$

where

$$\begin{aligned}\mu_{c,2}&=\frac{c\sigma^2}{(1-c)^2}+\frac{\sigma^2(\nu_4-3)}{c-1},\\ \sigma_{c,2}^2&=\frac{2c^3\sigma^4}{(1-c)^4}+\frac{c\sigma^4(\nu_4-3)}{(c-1)^2}.\end{aligned}$$

4. Proof of Theorem 4.5

Note that under Assumption *(B1)* and *(C2)*, $B_{\mathbf{X},\boldsymbol\beta}(\hat{\boldsymbol\beta},\boldsymbol\beta)=\boldsymbol\beta^{\mathrm T}\Pi\boldsymbol\beta=\boldsymbol\beta^{\mathrm T}(\boldsymbol I_p-\hat{\boldsymbol\Sigma}^+\hat{\boldsymbol\Sigma})\boldsymbol\beta$. If we directly consider $\boldsymbol\beta^{\mathrm T}(\boldsymbol I_p-\hat{\boldsymbol\Sigma}^+\hat{\boldsymbol\Sigma})\boldsymbol\beta$, we can make use of the asymptotic results for quadratic forms, see Theorem 7.2 in Bai & Yao (2008), which is stated as follows.

Lemma 4.1. (Theorem 7.2 in Bai & Yao (2008)) *Let $\{\boldsymbol A_n=[a_{ij}(n)]\}$ be a sequence of $n\times n$ real symmetric matrices, $\{\mathbf{x}_i\}_{i\in\mathbb N}$ be a sequence of i.i.d. K dimensional real random vectors, with $\mathbb E(\mathbf{x}_i)=0$, $\mathbb E(\mathbf{x}_i\mathbf{x}_i^{\mathrm T})=(\gamma_{ij})_{K\times K}$ and $\mathbb E[\|\mathbf{x}_i\|^4]<\infty$. Denote*

$$\begin{aligned}&\mathbf{x}_i=(\mathrm{x}_{\ell i})_{K\times 1},\quad \mathbf{X}(\ell)=(\mathrm{x}_{\ell 1},\cdots,\mathrm{x}_{\ell n})^{\mathrm T},\\ &\ell=1,\cdots,K,\ i=1,\cdots,n,\end{aligned}$$

assume the following limits exist

$$\omega=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n a_{ii}^2(n),\quad \theta=\lim_{n\to\infty}\frac{1}{n}\operatorname{Tr}\boldsymbol A_n^2.$$

Define a K-dimensional random vectors,

$$\mathbf{z}_n=(\mathrm{z}_{n,\ell})_{K\times 1},$$

where, for $1\le\ell\le K$,

$$\mathrm{z}_{n,\ell}=\frac{1}{\sqrt n}\big(\mathbf{X}(\ell)^{\mathrm T}\boldsymbol A_n\mathbf{X}(\ell)-\gamma_{\ell\ell}\operatorname{Tr}\{\boldsymbol A_n\}\big).$$

Then $\mathbf{z}_n$ converges weakly to a zero-mean Gaussian vector with covariance matrix $\boldsymbol D=\boldsymbol D_1+\boldsymbol D_2$, where for any $1\le\ell,\ell'\le K$,

$$[\boldsymbol D_1]_{\ell\ell'}=\omega\left(\mathbb E(x_{\ell 1}^2x_{\ell' 1}^2)-\gamma_{\ell\ell}\gamma_{\ell'\ell'}\right),$$

and

$$[\boldsymbol D_2]_{\ell\ell'}=(\theta-\omega)(\gamma_{\ell\ell'}\gamma_{\ell'\ell}+\gamma_{\ell\ell'}^2)$$

According to the results in Lemma 4.1, let $\boldsymbol A_n=\Pi=\boldsymbol I_p-\hat{\boldsymbol\Sigma}^+\hat{\boldsymbol\Sigma}$, then we have, as $p\to\infty$,

$$\sqrt p\Big\{\boldsymbol\beta^{\mathrm T}\Pi\boldsymbol\beta-\frac{r^2}{p}\operatorname{Tr}(\Pi)\Big\}\xrightarrow{d}N(0,d^2=d_1^2+d_2^2),$$

where

$$\omega=\lim_{p\to\infty}\frac1p\sum_{i=1}^p\Pi_{ii}^2,\quad \theta=\lim_{p\to\infty}\frac1p\operatorname{Tr}(\Pi^2)=1-\frac1c,$$

and

$$\begin{aligned}d_1^2&=\omega\left\{\mathbb E(x_{\ell 1}^2x_{\ell 1}^2)-\gamma_{\ell\ell}^2\right\}=\omega\big(\frac{p^2}{r^4}\,\mathbb E(\beta_i^4)-1\big)r^4,\\ d_2^2&=(\theta-\omega)(\gamma_{\ell\ell}^2+\gamma_{\ell\ell}^2)=2(\theta-\omega)r^4.\end{aligned}$$

Since in the proof of Theorem 4.3, we have already shown that

$$\frac{r^2}{p}\operatorname{Tr}(\Pi)=r^2(1-\frac np).$$

In particular, if $\boldsymbol\beta$ follows multivariate Gaussian distribution, i.e. $\boldsymbol\beta\sim N_p(0,\frac{r^2}{p}\boldsymbol I_p)$, then as $p\to\infty$,

$$\sqrt p\Big\{B_{\mathbf{X},\boldsymbol\beta}(\hat{\boldsymbol\beta},\boldsymbol\beta)-r^2(1-\frac np)\Big\}\xrightarrow{d}N\Big(0,2(1-\frac1c)r^4\Big).$$

Moreover, $V_{\mathbf{X},\boldsymbol\beta}(\hat{\boldsymbol\beta},\boldsymbol\beta)=V_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)$, we have already proved in Theorem 4.3 that

$$\begin{aligned}&p(V_{\mathbf{X},\boldsymbol\beta}(\hat{\boldsymbol\beta},\boldsymbol\beta)-\frac{\sigma^2}{c_n-1})\\ \xrightarrow{d}\ &N\Big(\frac{c\sigma^2}{(1-c)^2}+\frac{\sigma^2(\nu_4-3)}{c-1},\ \frac{2c^3\sigma^4}{(1-c)^4}+\frac{c\sigma^4(\nu_4-3)}{(c-1)^2}\Big).\end{aligned}$$

Note that $\operatorname{Cov}(B_{\mathbf{X},\boldsymbol\beta}(\hat{\boldsymbol\beta},\boldsymbol\beta),\ V_{\mathbf{X},\boldsymbol\beta}(\hat{\boldsymbol\beta},\boldsymbol\beta))=0$. According to the consistency rate of $B_{\mathbf{X},\boldsymbol\beta}(\hat{\boldsymbol\beta},\boldsymbol\beta)$ and $V_{\mathbf{X},\boldsymbol\beta}(\hat{\boldsymbol\beta},\boldsymbol\beta)$, we know that the bias $B_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)$ is the leading term of $R_{\mathbf{X},\boldsymbol\beta}(\hat{\boldsymbol\beta},\boldsymbol\beta)$. This implies that

$$\sqrt p\Big\{R_{\mathbf{X},\boldsymbol\beta}(\hat{\boldsymbol\beta},\boldsymbol\beta)-r^2(1-\frac{1}{c_n})-\frac{\sigma^2}{c_n-1}\Big\}\xrightarrow{d}N(0,\sigma_{c,3}^2),$$

where $\sigma_{c,3}^2=2r^4(1-1/c)$. A practical version of this CLT is given by

$$\sqrt p\Big\{R_{\mathbf{X},\boldsymbol\beta}(\hat{\boldsymbol\beta},\boldsymbol\beta)-r^2(1-\frac{1}{c_n})-\frac{\sigma^2}{c_n-1}\Big\}\xrightarrow{d}N(\tilde\mu_{c,3},\tilde\sigma_{c,3}^2),$$

where

$$\begin{aligned}\tilde\mu_{c,3}&=\frac{1}{\sqrt p}\Big\{\frac{c\sigma^2}{(1-c)^2}+\frac{\sigma^2(\nu_4-3)}{c-1}\Big\},\\ \tilde\sigma_{c,3}^2&=2(1-\frac1c)r^4+\frac1p\Big\{\frac{2c^3\sigma^4}{(1-c)^4}+\frac{c\sigma^4(\nu_4-3)}{(c-1)^2}\Big\}.\end{aligned}$$

5. More experiments

5.1. More results of Example 1

In this example, we consider the anisotropic case that the covariance matrix $\boldsymbol\Sigma$ is not an identity matrix. We checks Theorem 4.1 and define a statistic

$$T_n=\frac{p}{\sigma_c}\Big(R_{\mathbf{X}}(\hat{\boldsymbol\beta},\boldsymbol\beta)-\sigma^2\frac{c_n}{1-c_n}\Big)-\frac{\mu_c}{\sigma_c}.$$

According to Theorem 4.1, T_n weakly converges to the standard normal distribution as $n, p \to \infty$. In this example, we take $c = 1/2$ and $p = 50, 100, 200$. To make sure the assumption (A) holds, the generative distribution $P_{\mathbf{x}}$ is taken to be the standard normal distribution, the centered gamma with shape 4.0 and scale 0.5, and the normalized Student-t distribution with 6.0 degree of freedom. The covariance matrix $\boldsymbol{\Sigma}$ is taken to be

$$\boldsymbol{\Sigma} = 0.7\boldsymbol{I}_p + 0.3\mathbf{1}_p\mathbf{1}_p^{\mathrm{T}}.$$

The finite-sample distribution of T_n is estimated by the histogram of T_n under 1000 repetitions. The results are presented in Figure 1. One can find that the finite-sample distribution of T_n tends to the standard normal distribution as $n, p \to +\infty$. When $\alpha = 0.05$, the empirical cover rates of the 95%-confidence interval are reported in Figure 2.

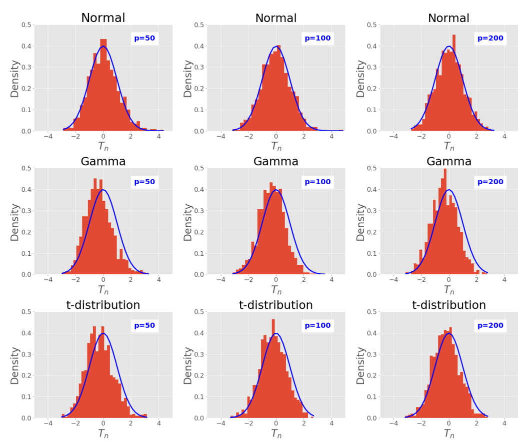

Figure 1. The histogram of T_n. The solid line is the density of the standard normal distribution.

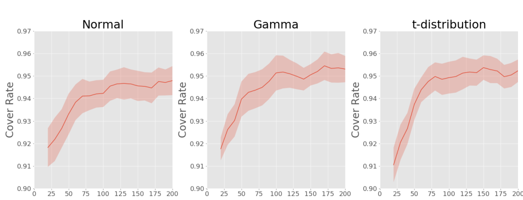

Figure 2. The cover rate of the confidence interval as p creases. The confidence level is 95%.

5.2. Example 3

This example checks Theorem 4.3. To proceed further, we denote two statistics:

$$T_{n,2} = \frac{\sqrt{p}}{\sigma_{c,1}}\Big\{R_{\mathbf{X}}(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}) - (1 - \frac{1}{c_n})r^2 - \frac{\sigma^2}{c_n - 1}\Big\} - \frac{\mu_{c,1}}{\sigma_{c,1}},$$

$$T_{n,3} = \frac{\sqrt{p}}{\tilde{\sigma}_{c,1}}\Big\{R_{\mathbf{X}}(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}) - (1 - \frac{1}{c_n})r^2 - \frac{\sigma^2}{c_n - 1}\Big\} - \frac{\tilde{\mu}_{c,1}}{\tilde{\sigma}_{c,1}}.$$

According to the central limit theorem (8) and its practical version, both $T_{n,2}$ and $T_{n,3}$ weakly converge to the standard normal distribution as $n, p \to +\infty$. We take $c = 2$ and $p = 100, 200, 400$. The finite-sample distributions of $T_{n,2}$ and $T_{n,3}$ are estimated by the histogram of $T_{n,2}$ and $T_{n,3}$ under 1000 repetitions. The results are presented at Figure 3 and Figure 4. One can see that the finite-sample distributions of $T_{n,2}$ and $T_{n,3}$ are close to the standard normal distribution, and the finite-sample performance of $T_{n,3}$ is better than that of $T_{n,2}$. When $\alpha = 0.05$, the empirical cover rates of the 95%-confidence interval (9) are reported in Figure 5.

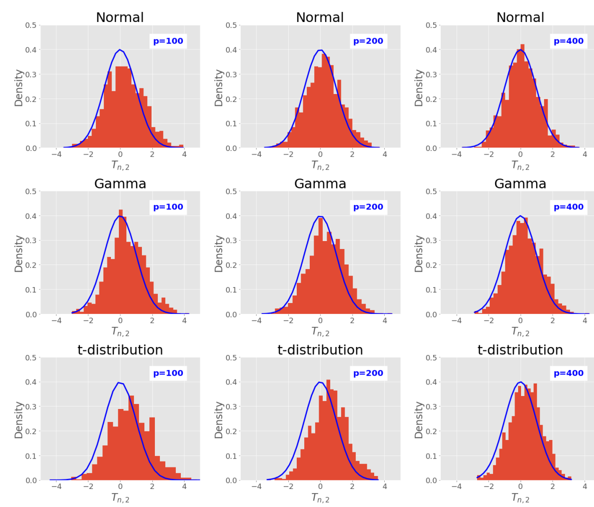

Figure 3. The histogram of $T_{n,2}$. The solid line is the density of the standard normal distribution.

5.3. An anisotropic example for Remark 4.2

In the over-parameterized case, the bias term $B_{\mathbf{X}}(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}) = \boldsymbol{\beta}^{\mathrm{T}}\Pi\boldsymbol{\Sigma}\Pi\boldsymbol{\beta}$ is non-zero while the variance term $V_{\mathbf{X}}(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$ remains the same as under-parameterized case. Therefore in this section, we conduct a small simulation to examine the fluctuation of the bias $B_{\mathbf{X}}$ for both isotropic and anisotropic $\boldsymbol{\Sigma}$ in the over-parameterized case with non-random $\boldsymbol{\beta}$ satisfying Assumption (C1). In particular, in the following we set $r = 1$.

We consider both localized and delocalized $\boldsymbol{\beta}$ such that

1. Localized case: $\boldsymbol{\beta}_1 = (1, 0, \cdots, 0)$;

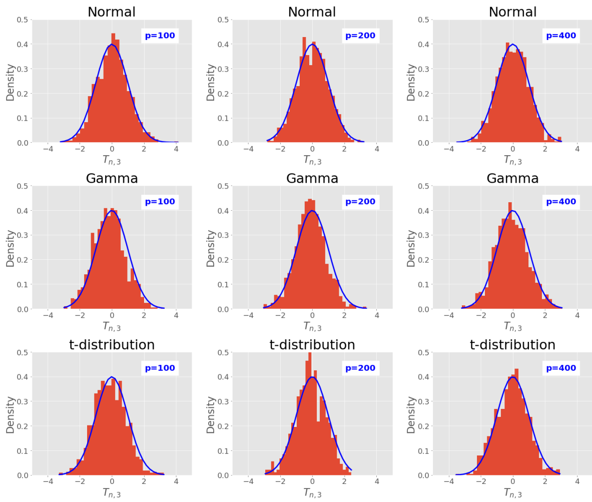

Figure 4. The histogram of $T_{n,3}$. The solid line is the density of the standard normal distribution.

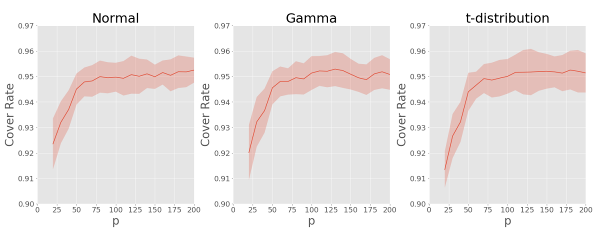

Figure 5. The coverage of confidence interval (9) as p increases. The confidence level is 95%.

2. Delocalized case: $\boldsymbol{\beta}_2 = \frac{1}{\sqrt{p}}(1, \cdots, 1)$;

and both the isotropic and anisotropic $\boldsymbol{\Sigma}$

3. Identity case: $\boldsymbol{\Sigma}_1 = \boldsymbol{I}_p$;
4. Compound symmetric case: $\boldsymbol{\Sigma}_2 = 0.5\boldsymbol{I}_p + 0.5\mathbf{1}_p\mathbf{1}_p^{\mathrm{T}}$.

Then we fix $p/n = 2$ and let p vary from 10 to 300, we present in Figure 6 the empirical variance of $\sqrt{p} * B_{\mathbf{X}}$ and $p * B_{\mathbf{X}}$ under various combinations of $\boldsymbol{\Sigma}$ and $\boldsymbol{\beta}$ with 1000 replications.

From the plot on the top left panel in Figure 6, we can see that the variance of $\sqrt{p} * B_{\mathbf{X}}$ for both $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ remain constant as p grows, which indicates that the convergence rate of $B_{\mathbf{X}}$ is $1/\sqrt{p}$ under the isotropic case regardless of localized or delocalized $\boldsymbol{\beta}$. As for the anisotropic case on the top right corner, the variance of $\sqrt{p} * B_{\mathbf{X}}$ stabilizes for $\boldsymbol{\beta}_1$, while decays for $\boldsymbol{\beta}_2$, which indicates that convergence rate of $B_{\mathbf{X}}$ under $(\boldsymbol{\Sigma}_2, \boldsymbol{\beta}_2)$ and $(\boldsymbol{\Sigma}_2, \boldsymbol{\beta}_1)$ are different.

This simulation result further confirms our conjecture that in the over-parameterized case, there is no universal CLT for the prediction risk $R_{\mathbf{X}}(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta})$ under the anisotropic setting for non-random $\boldsymbol{\beta}$.

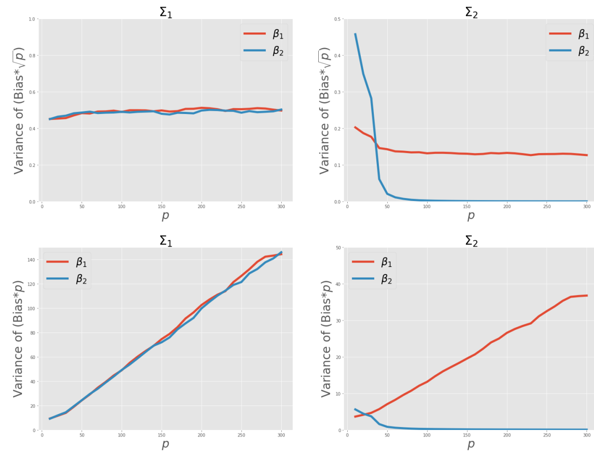

Figure 6. The upper panels are the empirical variances of $\sqrt{p} * B_{\mathbf{X}}$, the lower panels are for $p * B_{\mathbf{X}}$.

References

Bai, Z. and Silverstein, J. CLT for linear spectral statistics of large-dimensional sample covariance matrices. *The Annals of Probability*, 32:553–605, 2004.

Bai, Z. and Yao, J. Central limit theorems for eigenvalues in a spiked population model. *Annales de l'IHP Probabilités et statistiques*, 44(3):447–474, 2008.

Bai, Z. and Yin, Y. Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. In *Advances In Statistics*, pp. 108–127. World Scientific, 2008.

Bai, Z., Miao, B., and Pan, G. On asymptotics of eigenvectors of large sample covariance matrix. *The Annals of Probability*, 35(4):1532–1572, 2007.

Pan, G. and Zhou, W. Central limit theorem for signal-to-interference ratio of reduced rank linear receiver. *The Annals of Applied Probability*, 18(3):1232–1270, 2008.

Yin, Y. Limiting spectral distribution for a class of random matrices. *Journal of multivariate analysis*, 20(1):50–68, 1986.

Zheng, S., Bai, Z., and Yao, J. Substitution principle for clt of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *The Annals of Statistics*, 43(2):546–591, 2015.