Supplementary Material

A Sample splitting

Our procedure described in Sections 3.2 and 3.3 consists of two parts, the calibration of a factor model (i.e. estimating $\mathbf{B}$ in equation (1)) and multiple inference. The construction of the test statistics, or equivalently, the P-values, relies on a "fine" estimate of $\bar{\boldsymbol{f}}$ based on the linear model in (25). In practice, $\boldsymbol{b}_j$'s are replaced by the fitted loadings $\widehat{\boldsymbol{b}}_j$'s using the methods in Section 3.2.

To avoid mathematical challenges caused by the reuse of the sample, we resort to the simple idea of sample splitting (Hartigan, 1969; Cox, 1975): half the data are used for calibrating a factor model and the other half are used for multiple inference. We refer to Rinaldo et al. (2016) for a modern look at inference based on sample splitting. Specifically, the steps are summarized below.

(1) Split the data $\mathcal{X} = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$ into two halves $\mathcal{X}_1$ and $\mathcal{X}_2$. For simplicity, we assume that the data are divided into two groups of equal size $m = n/2$.

(2) We use $\mathcal{X}_1$ to estimate $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_p$ using either the U-type method (Section 3.2.1) or the adaptive Huber method (Section 3.2.2). For simplicity, we focus on the latter and denote the estimators by $\widehat{\boldsymbol{b}}_1(\mathcal{X}_1), \ldots, \widehat{\boldsymbol{b}}_p(\mathcal{X}_1)$.

(3) Proceed with the remain steps in the FarmTest procedure using the data in $\mathcal{X}_2$. Denote the resulting test statistics by $T_1, \ldots, T_p$. For a given threshold $z \geq 0$, the corresponding FDP and its asymptotic expression are defined as

$$\mathrm{FDP}_{\mathrm{sp}}(z) = V(z)/R(z) \quad \text{and} \quad \mathrm{AFDP}_{\mathrm{sp}}(z) = 2p\Phi(-z)/R(z),$$

respectively, where $V(z) = \sum_{j \in \mathcal{H}_0} I(|T_j| \geq z)$, $R(z) = \sum_{1 \leq j \leq p} I(|T_j| \geq z)$ and the subscript "sp" stands for sample splitting.

The purpose of sample splitting employed in the above procedure is to facilitate the theoretical analysis. The following result shows that the asymptotic FDP $\mathrm{AFDP}_{\mathrm{sp}}(z)$, constructed via sample splitting, provides a consistent estimate of $\mathrm{FDP}(z)$.

Theorem A.1. Suppose that Assumptions 1 (i)–(iv), Assumptions 2–4 hold. Let $\tau_j = a_j\omega_{n,p}$, $\tau_{jj} = a_{jj}\omega_{n,p}$ with $a_j \geq \sigma_{jj}^{1/2}$, $a_{jj} \geq \mathrm{var}(X_j^2)^{1/2}$ for $j = 1,\ldots,p$, and let $\gamma = \gamma_0\{p/\log(np)\}^{1/2}$ with $\gamma_0 \geq \overline{\sigma}_\varepsilon$. Then, for any $z \geq 0$, $|\mathrm{AFDP}_{\mathrm{sp}}(z) - \mathrm{FDP}_{\mathrm{sp}}(z)| = o_{\mathbb{P}}(1)$ as $n, p \to \infty$.

B Derivation of (6)

For any t and $a_j \geq \sigma_{jj}^{1/2}$, Lemma C.3 in Section C.1 shows that, conditionally on $\boldsymbol{f}_i$'s, the rescaled robust estimator $\sqrt{n}\,\widehat{\mu}_j$ with $\tau_j = a_j(n/t)^{1/2}$ satisfies

$$\sqrt{n}\,(\widehat{\mu}_j - \mu_j - \boldsymbol{b}_j^{\mathrm{T}}\bar{\boldsymbol{f}}) = \left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ell'_{\tau_j}(\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}_i + \varepsilon_{ij}) - \sqrt{n}\,\boldsymbol{b}_j^{\mathrm{T}}\bar{\boldsymbol{f}}\right\} + R_{1j}, \tag{B.1}$$

where the remainder R_{1j} satisfies $\mathbb{P}(|R_{1j}| \lesssim a_j n^{-1/2}t) \geq 1 - 3e^{-t}$. The stochastic term $n^{-1/2}\sum_{i=1}^{n}\ell'_{\tau_j}(\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}_i + \varepsilon_{ij}) - \sqrt{n}\,\boldsymbol{b}_j^{\mathrm{T}}\bar{\boldsymbol{f}}$ in (6) can be decomposed as

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\ell'_{\tau_j}(\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}_i + \varepsilon_{ij}) - \sqrt{n}\,\boldsymbol{b}_j^{\mathrm{T}}\bar{\boldsymbol{f}} &= \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\ell'_{\tau_j}(\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}_i + \varepsilon_{ij}) - \mathbb{E}_{\boldsymbol{f}_i}\ell'_{\tau_j}(\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}_i + \varepsilon_{ij})\}}_{S_j} \\ &\quad + \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\mathbb{E}_{\boldsymbol{f}_i}\ell'_{\tau_j}(\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}_i + \varepsilon_{ij}) - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}_i\}}_{R_{2j}},\end{aligned} \tag{B.2}$$

where $\bar{\boldsymbol{f}} = n^{-1}\sum_{i=1}^{n}\boldsymbol{f}_i$ and $\mathbb{E}_{\boldsymbol{f}_i}(\cdot) = \mathbb{E}(\cdot|\boldsymbol{f}_i)$ denotes the conditional expectation given $\boldsymbol{f}_i$. Under the finite fourth moment condition $\upsilon_j := (\mathbb{E}\varepsilon_j^4)^{1/4} < \infty$, it follows from Lemma C.4 that as long as $n \geq 4a_j^{-2}\max_{1\leq i\leq n}(\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}_i)^2 t$,

$$|R_{2j}| \leq 8a_j^{-3}\upsilon_j^4\, n^{-1}t^{3/2}. \tag{B.3}$$

Given $\{\boldsymbol{f}_i\}_{i=1}^n$, S_j in (B.2) is a sum of (conditionally) independent random variables with (conditional) mean zero. In addition, we note from (C.4) in Lemma C.4 that the (conditional) variance of $\ell'_{\tau_j}(\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}_i + \varepsilon_{ij})$ given $\boldsymbol{f}_i$ satisfies $|\mathrm{var}_{\boldsymbol{f}_i}\{\ell'_{\tau_j}(\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}_i + \varepsilon_{ij})\} - \sigma_{\varepsilon,jj}| \lesssim n^{-1}t$. Therefore, by the central limit theorem, the conditional distribution of S_j given $\{\boldsymbol{f}_i\}_{i=1}^n$ is asymptotically normal with mean zero and variance $\sigma_{\varepsilon,jj}$ as long as $t = t(n,p) = o(n)$.

This, together with (B.3) implies that, conditioning on $\{\boldsymbol{f}_i\}_{i=1}^n$, the distribution of $\sqrt{n}\,\widehat{\mu}_j$ is close to a normal distribution with mean $\sqrt{n}\,(\mu_j + \boldsymbol{b}_j^{\mathrm{T}}\bar{\boldsymbol{f}})$ and variance $\sigma_{\varepsilon,jj}$. Under the identifiability condition (2), $\sigma_{\varepsilon,jj} = \sigma_{jj} - \|\boldsymbol{b}_j\|_2^2$ for $j = 1, \ldots, p$. We complete the proof.

C Proofs of main results

In this section, we present the proofs for Theorems 1–5 and Theorem A.1, starting with some preliminary results whose proofs can be found in Section D. Recall that

$$w_{n,p} = \sqrt{\frac{n}{\log(np)}},$$

which will be frequently used in the sequel. Also, we use $c_1, c_2, \ldots$ and $C_1, C_2, \ldots$ to denote constants that are independent of (n,p), which may take different values at each occurrence.

C.1 Preliminaries

For each $1 \le j \le p$, define the zero-mean error variable $\xi_j = X_j - \mu_j$ and let $\mu_{j,\tau} = \operatorname{argmin}_{\theta \in \mathbb{R}} \mathbb{E}\ell_\tau(X_j - \theta)$ be the approximate mean, where $\ell_\tau(\cdot)$ is the Huber loss given in (5). Throughout, we use ψ_τ to denote the derivative of ℓ_τ, that is,

$$\psi_\tau(u) = \ell_\tau'(u) = \min(|u|, \tau)\,\mathrm{sign}(u), \quad u \in \mathbb{R}.$$

Lemma C.1 provides an upper bound on the approximation bias $|\mu_j - \mu_{j,\tau}|$, whose proof is given in Section D.3.

Lemma C.1. Let $1 \le j \le p$ and assume that $\upsilon_{\kappa,j} = \mathbb{E}(|\xi_j|^\kappa) < \infty$ for some $\kappa \ge 2$. Then, as long as $\tau > \sigma_{jj}^{1/2}$, we have

$$|\mu_{j,\tau} - \mu_j| \le (1 - \sigma_{jj}\tau^{-2})^{-1}\upsilon_{\kappa,j}\tau^{1-\kappa}. \tag{C.1}$$

The following concentration inequality is from Theorem 5 in Fan *et al.* (2017), showing that $\widehat{\mu}_j$ with a properly chosen robustification parameter τ exhibits sub-Gaussian tails when the underlying distribution has heavy tails with only finite second moment.

Lemma C.2. For every $1 \le j \le p$ and $t > 0$, the estimator $\widehat{\mu}_j$ in (5) with $\tau = a(n/t)^{1/2}$ for $a \ge \sigma_{jj}^{1/2}$ satisfies $\mathbb{P}\{|\widehat{\mu}_j - \mu_j| \ge 4a(t/n)^{1/2}\} \le 2e^{-t}$ as long as $n \ge 8t$.

The next result provides a nonasymptotic Bahadur representation for $\widehat{\mu}_j$. In particular, we show that when the second moment is finite, the remainder of the Bahadur representation for $\widehat{\mu}_j$ exhibits sub-exponential tails. The proof of Lemmas C.3–C.6 can be found respectively in Sections D.4–D.7.

Lemma C.3. For every $1 \le j \le p$ and for any $t \ge 1$, the estimator $\widehat{\mu}_j$ in (5) with $\tau = a(n/t)^{1/2}$ and $a \ge \sigma_{jj}^{1/2}$ satisfies that as long as $n \ge 8t$,

$$\left| \sqrt{n}\,(\widehat{\mu}_j - \mu_j) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\tau(\xi_{ij}) \right| \le C \frac{at}{\sqrt{n}} \tag{C.2}$$

with probability greater than $1 - 3e^{-t}$, where $\xi_{ij} = X_{ij} - \mu_j$ and $C > 0$ is an absolute constant.

Under factor model (1), note that $\xi_j = \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f} + \varepsilon_j$ for every j. The following conclusion reveals that the differences between the first two (conditional) moments of ξ_j and $\psi_\tau(\xi_j)$ given $\boldsymbol{f}$ vanish faster if higher moments of ε_j exist.

Lemma C.4. Assume that $\mathbb{E}(|\varepsilon_j|^\kappa) < \infty$ for some $1 \le j \le p$ and $\kappa \ge 2$.

(1) On the event $\mathcal{G}_j := \{|\boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}| < \tau\}$,

$$|\mathbb{E}_{\boldsymbol{f}} \psi_\tau(\xi_j) - \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}| \le \min\left\{ \frac{\sigma_{\varepsilon,jj}}{\tau - |\boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}|}, \frac{\mathbb{E}|\varepsilon_j|^\kappa}{(\tau - |\boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}|)^{\kappa-1}} \right\} \tag{C.3}$$

almost surely. In addition, if $\kappa > 2$, we have

$$\sigma_{\varepsilon,jj} - \frac{\mathbb{E}(|\varepsilon_j|^\kappa)}{(\tau - |\boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}|)^{\kappa-2}} \left\{ \frac{2}{\kappa - 2} + \frac{\mathbb{E}(|\varepsilon_j|^\kappa)}{(\tau - |\boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}|)^{\kappa}} \right\} \le \mathrm{var}_{\boldsymbol{f}}\{\psi_\tau(\xi_j)\} \le \sigma_{\varepsilon,jj} \tag{C.4}$$

almost surely on $\mathcal{G}_j$.

(2) Assume that $\upsilon_{jk} := \mathbb{E}(|\varepsilon_j|^\kappa) \vee \mathbb{E}(|\varepsilon_k|^\kappa) < \infty$ for some $1 \le j \ne k \le p$ and $\kappa > 2$. Then

$$|\mathrm{cov}_{\boldsymbol{f}}(\psi_\tau(\xi_j), \psi_\tau(\xi_k)) - \mathrm{cov}(\varepsilon_j, \varepsilon_k)| \le C \max(\upsilon_{jk} \tau^{2-\kappa}, \upsilon_{jk}^2 \tau^{2-2\kappa}) \tag{C.5}$$

almost surely on $\mathcal{G}_{jk} := \{|\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}| \vee |\boldsymbol{b}_k^{\mathrm{T}}\boldsymbol{f}| \le \tau/2\}$, where $C > 0$ is an absolute constant.

Lemma C.5. Suppose that Assumption 1 holds. Then, for any $t > 0$,

$$\mathbb{P}\{\|\sqrt{n}\bar{\boldsymbol{f}}\|_2 > C_1 A_f (K+t)^{1/2}\} \le e^{-t}, \tag{C.6}$$

$$\mathbb{P}\bigg\{\max_{1\le i\le n} \|\boldsymbol{f}_i\|_2 > C_1 A_f (K + \log n + t)^{1/2}\bigg\} \le e^{-t}, \tag{C.7}$$

$$\text{and} \quad \mathbb{P}[\|\widehat{\boldsymbol{\Sigma}}_f - \mathbf{I}_K\|_2 > C_2 \max\{A_f^2 n^{-1/2}(K+t)^{1/2}, A_f^4 n^{-1}(K+t)\}] \le 2e^{-t}, \tag{C.8}$$

where $\bar{\boldsymbol{f}} = n^{-1}\sum_{i=1}^n \boldsymbol{f}_i$, $\widehat{\boldsymbol{\Sigma}}_f = n^{-1}\sum_{i=1}^n \boldsymbol{f}_i \boldsymbol{f}_i^{\mathrm{T}}$ and $C_1, C_2 > 0$ are absolute constants.

The following lemma provides an ℓ_∞-error bound for estimating the eigenvectors $\overline{\boldsymbol{v}}_\ell$'s of $\mathbf{B}^{\mathrm{T}}\mathbf{B}$. The proof is based on an ℓ_∞ eigenvector perturbation bound developed in Fan *et al.* (2018) and is given in Appendix D.

Lemma C.6. Suppose Assumption 2 holds. Then we have

$$\max_{1\le \ell\le K} |\widetilde{\lambda}_\ell - \overline{\lambda}_\ell| \le p\|\widehat{\boldsymbol{\Sigma}}_{\mathrm{H}} - \boldsymbol{\Sigma}\|_{\max} + \|\boldsymbol{\Sigma}_\varepsilon\| \tag{C.9}$$

$$\text{and} \quad \max_{1\le \ell\le K} \|\widehat{\boldsymbol{v}}_\ell - \overline{\boldsymbol{v}}_\ell\|_\infty \le C(p^{-1/2}\|\widehat{\boldsymbol{\Sigma}}_{\mathrm{H}} - \boldsymbol{\Sigma}\|_{\max} + p^{-1}\|\boldsymbol{\Sigma}_\varepsilon\|), \tag{C.10}$$

where $C > 0$ is a constant independent of (n, p).

C.2 Proof of Theorem 1

To prove (15) and (16), we will derive the following stronger results that

$$p_0^{-1}V(z) = 2\Phi(-z) + O_{\mathbb{P}}\{p^{-\kappa_1/2} + w_{n,p}^{-1/2} + n^{-1/2}\log(np)\} \tag{C.11}$$

$$\text{and} \quad p^{-1}R(z) = \frac{1}{p}\sum_{j=1}^p \bigg\{\Phi\bigg(-z + \frac{\sqrt{n}\mu_j}{\sqrt{\sigma_{\varepsilon,jj}}}\bigg) + \Phi\bigg(-z - \frac{\sqrt{n}\mu_j}{\sqrt{\sigma_{\varepsilon,jj}}}\bigg)\bigg\} + O_{\mathbb{P}}\{p^{-\kappa_1/2} + w_{n,p}^{-1/2} + n^{-1/2}\log(np)\} \tag{C.12}$$

uniformly over $z \ge 0$ as $n, p \to \infty$, where $w_{n,p} = \sqrt{n/\log(np)}$.

For $1 \le j \le p$ and $t \ge 1$, set $\tau_j = a_j\sqrt{n/t}$ with $a_j \ge \sigma_{jj}^{1/2}$. By Lemma C.3, for every

$j \in \mathcal{H}_0$ so that $\mu_j = 0$,

$$|T_j^\circ - \sigma_{\varepsilon,jj}^{-1/2}(S_j + R_{2j})| \leq c_1 \frac{a_j t}{\sqrt{\sigma_{\varepsilon,jj} n}} \tag{C.13}$$

with probability greater than $1 - 3e^{-t}$ as long as $n \geq 8t$, where

$$S_j = \frac{1}{\sqrt{n}} \sum_{i=1}^n S_{ij} \text{ with } S_{ij} := \psi_{\tau_j}(\boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}_i + \varepsilon_{ij}) - \mathbb{E}_{\boldsymbol{f}_i} \psi_{\tau_j}(\boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}_i + \varepsilon_{ij}), \tag{C.14}$$

$R_{2j} = n^{-1/2} \sum_{i=1}^n \{\mathbb{E}_{\boldsymbol{f}_i} \psi_{\tau_j}(\boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}_i + \varepsilon_{ij}) - \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}_i\}$. For $j = 1, \ldots, p$, denote by $\mathcal{E}_{1j}(t)$ the event that (C.13) holds. Define $\mathcal{E}_1(t) = \bigcap_{j=1}^p \mathcal{E}_{1j}(t)$, on which it holds

$$\sum_{j \in \mathcal{H}_0} I\left(|T_{0j}| \geq z + \frac{c_1 a_j t}{\sqrt{\sigma_{\varepsilon,jj} n}}\right) \leq V(z) \leq \sum_{j \in \mathcal{H}_0} I\left(|T_{0j}| \geq z - \frac{c_1 a_j t}{\sqrt{\sigma_{\varepsilon,jj} n}}\right), \tag{C.15}$$

where $T_{0j} := \sigma_{\varepsilon,jj}^{-1/2}(S_j + R_{2j})$. Next, let $\mathcal{E}_2(t)$ denote the event on which the following hold:

$$\|\sqrt{n}\bar{\boldsymbol{f}}\|_2 \leq C_1 A_f (K + t)^{1/2}, \quad \max_{1 \leq i \leq n} \|\boldsymbol{f}_i\|_2 \leq C_1 A_f (K + \log n + t)^{1/2},$$
$$\text{and} \quad \|\widehat{\boldsymbol{\Sigma}}_f - \mathbf{I}_K\|_2 \leq C_2 \max\{A_f^2 n^{-1/2}(K + t)^{1/2}, A_f^4 n^{-1}(K + t)\}.$$

From Lemmas C.3, C.5 and the union bound, it follows that

$$\mathbb{P}\{\mathcal{E}_1(t)^{\mathrm{c}}\} \leq p e^{-t} \quad \text{and} \quad \mathbb{P}\{\mathcal{E}_2(t)^{\mathrm{c}}\} \leq 4e^{-t}.$$

With the above preparations, we are ready to prove (C.11). The proof of (C.12) follows the same argument and therefore is omitted. Note that, on the event $\mathcal{E}_2(t)$,

$$\max_{1 \leq i \leq n} |\boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}_i| \leq C_1 A_f \|\boldsymbol{b}_j\|_2 (K + \log n + t)^{1/2} \text{ for all } 1 \leq j \leq p.$$

By the definition of τ_j's,

$$\max_{1 \leq i \leq n} |\boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}_i| \leq \tau_j / 2 \text{ for all } j = 1, \ldots, p, \tag{C.16}$$

as long as $n \geq 4(C_1 A_f)^2 (K + \log n + t)t$. This, together with Lemma C.5, implies $|R_{2j}| \leq$

$8a_j^{-3}v_j^4 n^{-1}t^{3/2}$ almost surely on $\mathcal{E}_2(t)$ for all sufficiently large n. Moreover, taking (C.15) into account we obtain that, almost surely on the event $\mathcal{E}_1(t) \cap \mathcal{E}_2(t)$,

$$\sum_{j \in \mathcal{H}_0} I(|\sigma_{\varepsilon,jj}^{-1/2} S_j| \geq z + c_2 n^{-1/2} t) \leq V(z) \leq \sum_{j \in \mathcal{H}_0} I(|\sigma_{\varepsilon,jj}^{-1/2} S_j| \geq z - c_2 n^{-1/2} t) \tag{C.17}$$

as long as $n \gtrsim (K + t)t$. For $x \in \mathbb{R}$, define

$$\widetilde{V}_+(x) = \sum_{j \in \mathcal{H}_0} I(\sigma_{\varepsilon,jj}^{-1/2} S_j \geq x) \quad \text{and} \quad \widetilde{V}_-(x) = \sum_{j \in \mathcal{H}_0} I(\sigma_{\varepsilon,jj}^{-1/2} S_j \leq -x),$$

so that (C.17) can be written as

$$\begin{aligned} & p_0^{-1}\{\widetilde{V}_+(z + c_2 n^{-1/2} t) + \widetilde{V}_-(z + c_2 n^{-1/2} t)\} \\ & \leq p_0^{-1} V(z) \leq p_0^{-1}\{\widetilde{V}_+(z - c_2 n^{-1/2} t) + \widetilde{V}_-(z - c_2 n^{-1/2} t)\}. \end{aligned} \tag{C.18}$$

Therefore, to prove (C.11) it suffices to focus on $\widetilde{V}_+$ and $\widetilde{V}_-$.

Observe that, conditional on $\mathcal{F}_n := \{\boldsymbol{f}_1, \ldots, \boldsymbol{f}_n\}$, $I(\sigma_{\varepsilon,11}^{-1/2} S_1 \geq z), \ldots, I(\sigma_{\varepsilon,pp}^{-1/2} S_p \geq z)$ are weakly correlated random variables. Define

$$Y_j = I(\sigma_{\varepsilon,jj}^{-1/2} S_j \geq z) \quad \text{and} \quad P_j = \mathbb{P}(\sigma_{\varepsilon,jj}^{-1/2} S_j \geq z | \mathcal{F}_n)$$

for $j = 1, \ldots, p$, and note that

$$\begin{aligned} \operatorname{var}\left(\frac{1}{p_0} \sum_{j \in \mathcal{H}_0} Y_j \middle| \mathcal{F}_n \right) &= \frac{1}{p_0^2} \sum_{j \in \mathcal{H}_0} \operatorname{var}(Y_j | \mathcal{F}_n) + \frac{1}{p_0^2} \sum_{j,k \in \mathcal{H}_0 : j \neq k} \operatorname{cov}(Y_j, Y_k | \mathcal{F}_n) \\ &\leq \frac{1}{4p_0} + \frac{1}{p_0^2} \sum_{j,k \in \mathcal{H}_0 : j \neq k} \{\mathbb{E}(Y_j Y_k | \mathcal{F}_n) - P_j P_k\} \end{aligned} \tag{C.19}$$

almost surely. In the following, we will study P_j and $\mathbb{E}(Y_j Y_k | \mathcal{F}_n)$ separately, starting with the former. Conditional on $\mathcal{F}_n$, S_j is a sum of independent zero-mean random variables with conditional variance $s_j^2 := \operatorname{var}(S_j | \mathcal{F}_n) = n^{-1} \sum_{i=1}^n s_{ij}^2$ where $s_{ij}^2 := \operatorname{var}(S_{ij} | \mathcal{F}_n)$. Let $G \sim \mathcal{N}(0, 1)$ be a standard normal random variable independent of the data. By the

Berry-Esseen inequality,

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}|\mathbb{P}(\sigma_{\varepsilon,jj}^{-1/2}S_j\le x|\mathcal{F}_n)-\mathbb{P}(s_j\sigma_{\varepsilon,jj}^{-1/2}G\le x|\mathcal{F}_n)|\\
&\lesssim\frac{1}{(ns_j)^{3/2}}\sum_{i=1}^n\mathbb{E}_{\boldsymbol{f}_i}|\psi_{\tau_j}(\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}_i+\varepsilon_{ij})|^3\lesssim\frac{1}{(ns_j)^{3/2}}\sum_{i=1}^n(|\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}_i|^3+\mathbb{E}|\varepsilon_{ij}|^3)
\end{aligned}\tag{C.20}$$

almost surely, where conditional on $\mathcal{F}_n$, $s_j\sigma_{\varepsilon,jj}^{-1/2}G\sim\mathcal{N}(0,s_j^2\sigma_{\varepsilon,jj}^{-1})$. Since $\max_{1\le i\le n}|\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}_i|\le\tau_j/2$ for all $1\le j\le p$ on $\mathcal{E}_2(t)$, applying Lemma C.4 with $\kappa=4$ yields

$$\sigma_{\varepsilon,jj}-4a_j^{-2}\upsilon_j^4(1+16a_j^{-4}\upsilon_j^4n^{-2}t^2)n^{-1}t\le s_j^2\le\sigma_{\varepsilon,jj}\tag{C.21}$$

almost surely on the event $\mathcal{E}_2(t)$. Using (C.21) and Lemma A.7 in the supplement of Spokoiny and Zhilova (2015), we get

$$\sup_{x\in\mathbb{R}}|\mathbb{P}(s_j\sigma_{\varepsilon,jj}^{-1/2}G\le x|\mathcal{F}_n)-\Phi(x)|\lesssim a_j^{-2}\upsilon_j^4n^{-1}t\tag{C.22}$$

almost surely on $\mathcal{E}_2(t)$ as long as $n\gtrsim(K+t)t$. Putting (C.20) and (C.22) together we conclude that, almost surely on $\mathcal{E}_2(t)$,

$$\max_{1\le j\le p}|P_j-\Phi(-z)|\lesssim n^{-1/2}(K+\log n+t)^{1/2}\tag{C.23}$$

uniformly for all $z\ge0$ as long as $n\gtrsim(K+t)t$.

Next we consider the joint probability $\mathbb{E}(Y_jY_k|\mathcal{F}_n)=\mathbb{P}(\sigma_{\varepsilon,jj}^{-1/2}S_j\ge z,\sigma_{\varepsilon,kk}^{-1/2}S_k\ge z|\mathcal{F}_n)$ for a fixed pair (j,k) satisfying $1\le j\ne k\le p$. Define bivariate random vectors $\boldsymbol{\xi}_i=(s_j^{-1}S_{ij},s_k^{-1}S_{ik})^{\mathrm{T}}$ for $i=1,\ldots,n$. Observe that $\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_n$ are conditionally independent random vectors given $\mathcal{F}_n$. Denote by $\mathbf{A}=(a_{uv})_{1\le u,v\le2}$ the covariance matrix of $n^{-1/2}\sum_{i=1}^n\boldsymbol{\xi}_i=(s_j^{-1}S_j,s_k^{-1}S_k)^{\mathrm{T}}$ given $\mathcal{F}_n$ such that

$$a_{11}=a_{22}=1\quad\text{and}\quad a_{12}=a_{21}=\frac{1}{ns_js_k}\sum_{i=1}^n\mathrm{cov}_{\boldsymbol{f}_i}(S_{ij},S_{ik}).$$

By Lemma C.4 and (C.21), we have $|a_{12}-r_{\varepsilon,jk}|\lesssim n^{-1}t$ almost surely on $\mathcal{E}_2(t)$. Therefore, the matrix $\mathbf{A}$ is positive definite almost surely on $\mathcal{E}_2(t)$ whenever $n\gtrsim t$. Let $\boldsymbol{G}=(G_1,G_2)^{\mathrm{T}}$

be a Gaussian random vector with $\mathbb{E}(\boldsymbol{G}) = \mathbf{0}$ and $\text{cov}(\boldsymbol{G}) = \mathbf{A}$. Then, applying Theorem 1.1 in Bentkus (2005), a multivariate Berry-Esseen bound, to $n^{-1/2}\sum_{i=1}^n \boldsymbol{\xi}_i$ gives

$$\sup_{x,y\in\mathbb{R}} |\mathbb{P}(s_j^{-1}S_j \geq x, s_k^{-1}S_k \geq y|\mathcal{F}_n) - \mathbb{P}(G_1 \geq x, G_2 \geq y)|$$
$$\lesssim \frac{1}{n^{3/2}}\sum_{i=1}^n \mathbb{E}(\|\mathbf{A}^{-1/2}\boldsymbol{\xi}_i\|_2^3) \lesssim \frac{1}{\sqrt{n}} + \frac{1}{n^{3/2}}\sum_{i=1}^n (|\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}_i|^3 + |\boldsymbol{b}_k^{\mathrm{T}}\boldsymbol{f}_i|^3)$$

almost surely on $\mathcal{E}_2(t)$. Taking $x = s_j^{-1}\sigma_{\varepsilon,jj}^{1/2}\, z$ and $y = s_k^{-1}\sigma_{\varepsilon,kk}^{1/2}\, z$ implies

$$|\mathbb{E}(Y_jY_k|\mathcal{F}_n) - \mathbb{P}(G_1 \geq s_j^{-1}\sigma_{\varepsilon,jj}^{1/2}\, z, G_2 \geq s_k^{-1}\sigma_{\varepsilon,kk}^{1/2}\, z|\mathcal{F}_n)|$$
$$\lesssim \frac{1}{\sqrt{n}} + \frac{1}{n^{3/2}}\sum_{i=1}^n (|\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}_i|^3 + |\boldsymbol{b}_k^{\mathrm{T}}\boldsymbol{f}_i|^3). \tag{C.24}$$

For the bivariate Gaussian random vector $(G_1, G_2)^{\mathrm{T}}$ with $a_{12} = \text{cov}(G_1, G_2)$, it follows from Corollary 2.1 in Li and Shao (2002) that, for any $x, y \in \mathbb{R}$,

$$|\mathbb{P}(G_1 \geq x, G_2 \geq y) - \{1-\Phi(x)\}\{1-\Phi(y)\}| \leq \frac{|a_{12}|}{4}\exp\left\{-\frac{x^2+y^2}{2(1+|a_{12}|)}\right\} \leq \frac{|a_{12}|}{4}.$$

This, together with the Gaussian comparison inequality (C.22) gives

$$|\mathbb{P}(G_1 \geq s_j^{-1}\sigma_{\varepsilon,jj}^{1/2}\, z, G_2 \geq s_k^{-1}\sigma_{\varepsilon,kk}^{1/2}\, z|\mathcal{F}_n) - \Phi(-z)^2| \lesssim |r_{\varepsilon,jk}| + n^{-1}t \tag{C.25}$$

almost surely on $\mathcal{E}_2(t)$ as long as $n \gtrsim (K+t)t$.

Consequently, it follows from (C.19), (C.23), (C.24), (C.25) and Assumption 1 that

$$\mathbb{E}[\{p_0^{-1}\widetilde{V}_+(z) - \Phi(-z)\}^2|\mathcal{F}_n] \lesssim p^{-\kappa_1} + n^{-1/2}(K + \log n + t)^{1/2} \tag{C.26}$$

almost surely on $\mathcal{E}_2(t)$ as long as $n \gtrsim (K+t)t$. A similar bound can be obtained for $\mathbb{E}[\{p_0^{-1}\widetilde{V}_-(z) - \Phi(-z)\}^2|\mathcal{F}_n]$. Recall that $\mathbb{P}\{\mathcal{E}_1(t) \cap \mathcal{E}_2(t)\} \geq 1-(p+4)e^{-t}$ whenever $n \geq 8t$. Finally, taking $t = \log(np)$ in (C.18) and (C.26) proves (C.11). □

C.3 Proof of Proposition 1

To begin with, observe that

$$\left|\widetilde{T}_j - \sqrt{\frac{n}{\widetilde{\sigma}_{\varepsilon,jj}}}(\widehat{\mu}_j - \boldsymbol{b}_j^{\mathrm{T}}\bar{\boldsymbol{f}}\,)\right| = \sqrt{\frac{n}{\widetilde{\sigma}_{\varepsilon,jj}}}|(\widetilde{\boldsymbol{b}}_j - \boldsymbol{b}_j)^{\mathrm{T}}\bar{\boldsymbol{f}}| \le \sqrt{\frac{n}{\widetilde{\sigma}_{\varepsilon,jj}}}\|\bar{\boldsymbol{f}}\|_2\|\widetilde{\boldsymbol{b}}_j - \boldsymbol{b}_j\|_2,$$
$$\left|\sqrt{\frac{n}{\widetilde{\sigma}_{\varepsilon,jj}}}(\widehat{\mu}_j - \boldsymbol{b}_j^{\mathrm{T}}\bar{\boldsymbol{f}}\,) - T_j^{\circ}\right| \le \left|\frac{1}{\sqrt{\widetilde{\sigma}_{\varepsilon,jj}}} - \frac{1}{\sqrt{\sigma_{\varepsilon,jj}}}\right|(|\sqrt{n}\,\widehat{\mu}_j| + \|\boldsymbol{b}_j\|_2\|\sqrt{n}\bar{\boldsymbol{f}}\|_2).$$

By Lemma C.5, $\|\sqrt{n}\bar{\boldsymbol{f}}\|_2 \lesssim (K+\log n)^{1/2}$ with probability greater than $1-n^{-1}$. Moreover, it follows from Lemma C.2 that $\max_{1\le j\le p}|\widehat{\mu}_j - \mu_j| \lesssim \{\log(np)\}^{1/2}n^{-1/2}$ with probability at least $1-2n^{-1}$. Putting the above calculations together, we conclude that

$$\max_{j\in\mathcal{H}_0}|\widetilde{T}_j - T_j^{\circ}| \lesssim \frac{\log(np)}{\sqrt{n}} + (K+\log n)^{1/2}\max_{1\le j\le p}(\|\widetilde{\boldsymbol{b}}_j - \boldsymbol{b}_j\|_2 + |\widetilde{\sigma}_{jj} - \sigma_{jj}|)$$

with probability at least $1-3n^{-1}$. Combining this with the proof of Theorem 1 and condition (17) implies $p_0^{-1}\widetilde{V}(z) = 2\Phi(-z) + o_{\mathbb{P}}(1)$. Similarly, it can be proved that (C.12) holds with $R(z)$ replaced by $\widetilde{R}(z)$. The conclusion follows immediately. □

C.4 Proof of Theorem 2

We first note that the $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Sigma}}_U$ defined is a U-statistic of order two. For simplicity, let $\mathcal{C}$ denote the set of $\binom{n}{2}$ distinct pairs (i_1, i_2) satisfying $1 \le i_1 < i_2 \le n$. Let $h(\boldsymbol{X}_i, \boldsymbol{X}_j) = 2^{-1}(\boldsymbol{X}_i - \boldsymbol{X}_j)(\boldsymbol{X}_i - \boldsymbol{X}_j)^{\mathrm{T}}$ and $Y_{ij} = \psi_\tau(h(\boldsymbol{X}_i, \boldsymbol{X}_j)) = \tau\psi_1(\tau^{-1}h(\boldsymbol{X}_i, \boldsymbol{X}_j))$, such that

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{\binom{n}{2}}\sum_{(i,j)\in\mathcal{C}} Y_{ij}.$$

We now rewrite the U-statistic $\widehat{\boldsymbol{\Sigma}}$ as an average of dependent averages of identically and independently distributed random matrices. Define $k = [n/2]$, the greatest integer $\le n/2$ and define

$$W_{(1,\ldots,n)} = k^{-1}(Y_{12} + Y_{23} + \ldots + Y_{2k-1,2k}).$$

Let $\mathcal{P}$ denote the class of all $n!$ permutations of $(1,\ldots,n)$ and $\pi = (i_1,\ldots,i_n) : \{1,\ldots,n\} \mapsto$

$\{1,\ldots,n\}$ be a permutation, i.e. $\pi(k)=i_k$ for $k=1,\ldots,n$. Then it can be shown that

$$\widehat{\boldsymbol{\Sigma}}=\frac{1}{n!}\sum_{\pi\in\mathcal{P}}W_\pi.$$

Using the convexity of maximum eigenvalue function $\lambda_{\max}(\cdot)$ along with the convexity of the exponential function, we obtain

$$\exp\{\lambda_{\max}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})/\tau\}\leq\frac{1}{n!}\sum_{\pi\in\mathcal{P}}\exp\{\lambda_{\max}(W_\pi-\boldsymbol{\Sigma})/\tau\}.$$

Combining this with Chebyshev's inequality delivers

$$\begin{aligned}
&\mathbb{P}\{\lambda_{\max}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\geq t/\sqrt{n}\}\\
&=\mathbb{P}\Big[\exp\{\lambda_{\max}(k\widehat{\boldsymbol{\Sigma}}-k\boldsymbol{\Sigma})/\tau\}\geq\exp\{kt/(\tau\sqrt{n}\,)\}\Big]\\
&\leq e^{-kt/(\tau\sqrt{n})}\frac{1}{n!}\sum_{\pi\in\mathcal{P}}\mathbb{E}\exp\{\lambda_{\max}(kW_\pi-k\boldsymbol{\Sigma})/\tau\}\\
&\leq e^{-kt/(\tau\sqrt{n})}\frac{1}{n!}\sum_{\pi\in\mathcal{P}}\mathbb{E}\,\mathrm{tr}\exp\{(kW_\pi-k\boldsymbol{\Sigma})/\tau\},
\end{aligned}$$

where we use the property $e^{\lambda_{\max}(\mathbf{A})}\leq\mathrm{tr}\,e^{\mathbf{A}}$ in the last inequality. For a given permutation $\pi=(i_1,\ldots,i_n)\in\mathcal{P}$, we write $Y_{\pi j}=Y_{i_{2j-1}i_{2j}}$ and $H_{\pi j}=h(\boldsymbol{X}_{i_{2j-1}},\boldsymbol{X}_{i_{2j}})$ with $\mathbb{E}H_{\pi j}=\boldsymbol{\Sigma}$. We then rewrite W_π as $W_\pi=k^{-1}(Y_{\pi 1}+\ldots+Y_{\pi k})$, where $Y_{\pi j}$'s are mutually independent. Before proceeding, we introduce the following lemma whose proof is based on elementary calculations.

Lemma C.7. For any $\tau>0$ and $x\in\mathbb{R}$, we have $\psi_\tau(x)=\tau\psi_1(x/\tau)$ and

$$-\log(1-x+x^2)\leq\psi_1(x)\leq\log(1+x+x^2)\ \text{ for all } x\in\mathbb{R}.$$

From Lemma C.7 we see that the matrix $Y_{\pi j}$ can be bounded as

$$-\log(\mathbf{I}_p-H_{\pi j}/\tau+H_{\pi j}^2/\tau^2)\leq Y_{\pi j}/\tau\leq\log(\mathbf{I}_p+H_{\pi j}/\tau+H_{\pi j}^2/\tau^2).$$

Using this property we can bound $\mathbb{E}\exp\{\mathrm{tr}(kW_\pi - k\mathbf{\Sigma})/\tau\}$ by

$$\begin{aligned}
&\mathbb{E}_{[k-1]}\mathbb{E}_k \,\mathrm{tr}\exp\left\{\sum_{j=1}^{k-1} Y_{\pi j} - (k/\tau)\mathbf{\Sigma} + Y_{\pi k}\right\} \\
&\quad \le \mathbb{E}_{[k-1]}\mathbb{E}_k \,\mathrm{tr}\exp\left\{\sum_{j=1}^{k-1} Y_{\pi j} - (k/\tau)\mathbf{\Sigma} + \log(\mathbf{I}_p + H_{\pi j}/\tau + H_{\pi j}^2/\tau^2)\right\} \qquad \text{(C.27)}
\end{aligned}$$

To further bound the right-hand side of (C.27), we follow a similar argument as in Minsker (2016). The following lemma, which is taken from Lieb (2002), is commonly referred to as the Lieb's concavity theorem.

Lemma C.8. For any symmetric matrix $H \in \mathbb{R}^{d\times d}$, the function

$$f(A) = \mathrm{tr}\exp(H + \log A), \quad A \in \mathbb{R}^{d\times d}$$

is concave over the set of all positive definite matrices.

Applying Lemma C.8 repeatedly along with Jensen's inequality, we obtain

$$\begin{aligned}
\mathbb{E}\{\mathrm{tr}\exp(kW_\pi - k\mathbf{\Sigma})/\tau\} &\le \mathbb{E}\,\mathrm{tr}\exp\left\{\sum_{j=1}^{k-1} Y_{\pi j} - (k/\tau)\mathbf{\Sigma} + \log(\mathbf{I}_p + \mathbb{E}H_{\pi k}/\tau + \mathbb{E}H_{\pi k}^2/\tau^2)\right\} \\
&\le \mathrm{tr}\exp\left\{\sum_{j=1}^{k} \log(\mathbf{I}_p + \mathbb{E}H_{\pi j}/\tau + \mathbb{E}H_{\pi j}^2/\tau^2) - (k/\tau)\mathbf{\Sigma}\right\} \\
&\le \mathrm{tr}\exp\left(\sum_{j=1}^{k} \mathbb{E}H_{\pi k}^2/\tau^2\right),
\end{aligned}$$

where we use the inequality $\log(1+x) \le x$ for $x > -1$ in the last step. The following lemma gives an explicit form for $v^2 := \|\mathbb{E}H_{\pi k}^2\|_2$.

Lemma C.9. We have

$$\|\mathbb{E}h^2(\boldsymbol{X}_1, \boldsymbol{X}_2)\|_2 = \frac{1}{2}\left\|\mathbb{E}\{(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^{\mathrm{T}}\}^2 + \mathrm{tr}(\mathbf{\Sigma})\mathbf{\Sigma} + 2\mathbf{\Sigma}^2\right\|.$$

Proof of Lemma C.9. Write $\boldsymbol{X} = \boldsymbol{X}_1$ and $\boldsymbol{Y} = \boldsymbol{X}_2$. Without loss of generality, assume that

$\mathbb{E}(\boldsymbol{X})=\mathbb{E}(\boldsymbol{Y})=\mathbf{0}$. Let $H_1 = \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}},\ H_2 = \boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}},\ H_{12} = \boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}}$ and $H_{21} = \boldsymbol{Y}\boldsymbol{X}^{\mathrm{T}}$. Then

$$\begin{aligned}
\{(\boldsymbol{X}-\boldsymbol{Y})(\boldsymbol{X}-\boldsymbol{Y})^{\mathrm{T}}\}^2 &= (H_1 + H_2 - H_{12} - H_{21})^2 \\
&= H_1^2 + H_2^2 + H_{12}^2 + H_{21}^2 + H_1H_2 + H_2H_1 + H_{12}H_{21} + H_{21}H_{12} \\
&\quad - H_1H_{12} - H_{12}H_1 - H_1H_{21} - H_{21}H_1 - H_2H_{12} - H_{12}H_2 \\
&\quad - H_2H_{21} - H_{21}H_2,
\end{aligned}$$

which, by symmetry, implies that

$$\mathbb{E}\{(\boldsymbol{X}-\boldsymbol{Y})(\boldsymbol{X}-\boldsymbol{Y})^{\mathrm{T}}\}^2 = 2\mathbb{E}H_1^2 + 2\mathbb{E}H_{12}^2 + 2\mathbb{E}H_1H_2 + 2\mathbb{E}H_{12}H_{21}.$$

In the following we calculate the four expectations on the right-hand side of the above equality separately. For the first term, note that

$$\mathbb{E}H_1^2 = \mathbb{E}(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}).$$

Let $A = (A_{jk}) = H_{12}^2$ and we have

$$\mathbb{E}A_{jk} = \mathbb{E}\biggl(\sum_{\ell=1}^{p} X_\ell Y_\ell X_j Y_k\biggr) = \mathbb{E}\biggl(Y_k \sum_{\ell=1}^{p} X_j X_\ell Y_\ell\biggr) = \sum_{\ell=1}^{p} \sigma_{j\ell}\sigma_{\ell k},$$

where σ_{jk} is the (j,k)-th entry of $\boldsymbol{\Sigma}$. Therefore, we have $\mathbb{E}H_{12}^2 = \boldsymbol{\Sigma}^2$. For $\mathbb{E}H_1H_2$, using independence, we can show that $\mathbb{E}H_1H_2 = \boldsymbol{\Sigma}^2$. For $\mathbb{E}H_{12}H_{21}$, we have

$$\mathbb{E}H_{12}H_{21} = \mathbb{E}(\boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{X}^{\mathrm{T}}) = \mathbb{E}\{\mathbb{E}(\boldsymbol{X}\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}\boldsymbol{X}^{\mathrm{T}}|\boldsymbol{Y})\} = \mathrm{tr}(\boldsymbol{\Sigma})\boldsymbol{\Sigma}.$$

Putting the above calculations together completes the proof. $\square$

For any $u > 0$, putting the above calculations together and letting $\tau \geq 2v^2\sqrt{n}/u$ yield

$$
\begin{aligned}
&\mathbb{P}\{\lambda_{\max}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}) \geq u/\sqrt{n}\} \\
&\leq e^{-ku/(\sqrt{n}\tau)} \operatorname{tr}\exp\bigg(\sum_{j=1}^{k} \mathbb{E}H_{\pi k}^2/\tau^2\bigg) \leq p\exp\bigg(-\frac{ku}{\sqrt{n}\tau} + \frac{kv^2}{\tau^2}\bigg) \\
&\leq p\exp\bigg(-\frac{ku^2}{4nv^2}\bigg) \leq p\exp\bigg(-\frac{u^2}{16v^2}\bigg),
\end{aligned}
$$

where we use the fact that $k := [n/2] \geq n/4$ for $n \geq 2$ in the last inequality. On the other hand, it can be similarly shown that

$$
\mathbb{P}\{\lambda_{\min}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}) \leq -u/\sqrt{n}\} \leq p\exp\bigg(-\frac{u^2}{16v^2}\bigg)
$$

Combining the above two inequalities and putting $u = 4v\sqrt{t}$ complete the proof. $\square$

C.5 Proof of Theorem 3

First we bound $\max_{1\leq j\leq p}\|\widehat{\boldsymbol{b}}_j - \boldsymbol{b}_j\|_2$. For any $t > 0$, it follows from Theorem 2 that with probability greater than $1 - 2pe^{-t}$, $\|\widehat{\boldsymbol{\Sigma}}_U - \boldsymbol{\Sigma}\| \leq 4v(t/n)^{1/2}$, where v is as in (19). Define $\widetilde{\boldsymbol{b}}_j = (\overline{\lambda}_1^{1/2}\widehat{v}_{1j}, \ldots, \overline{\lambda}_K^{1/2}\widehat{v}_{Kj})^{\mathrm{T}} \in \mathbb{R}^K$, such that $\|\widehat{\boldsymbol{b}}_j - \boldsymbol{b}_j\|_2 \leq \|\widehat{\boldsymbol{b}}_j - \widetilde{\boldsymbol{b}}_j\|_2 + \|\widetilde{\boldsymbol{b}}_j - \boldsymbol{b}_j\|_2$. By Assumption 2, (20) and (21), we have

$$
\begin{aligned}
|\widehat{\lambda}_\ell^{1/2} - \overline{\lambda}_\ell^{1/2}| = |\widehat{\lambda}_\ell - \overline{\lambda}_\ell|/(\widehat{\lambda}_\ell^{1/2} + \overline{\lambda}_\ell^{1/2}) \lesssim p^{-1/2}(\|\widehat{\boldsymbol{\Sigma}}_U - \boldsymbol{\Sigma}\| + \|\boldsymbol{\Sigma}_\varepsilon\|), \\
\|\overline{\boldsymbol{v}}_\ell\|_\infty = \|\overline{\boldsymbol{b}}_\ell\|_\infty/\|\overline{\boldsymbol{b}}_\ell\|_2 \leq \|\mathbf{B}\|_{\max}/\|\overline{\boldsymbol{b}}_\ell\|_2 \lesssim p^{-1/2} \\
\text{and} \quad \|\widehat{\boldsymbol{v}}_\ell\|_\infty \leq \|\widehat{\boldsymbol{v}}_\ell - \overline{\boldsymbol{v}}_\ell\|_2 + \|\overline{\boldsymbol{v}}_\ell\|_\infty \lesssim p^{-1}\|\widehat{\boldsymbol{\Sigma}}_U - \boldsymbol{\Sigma}\| + p^{-1/2}.
\end{aligned}
$$

On the event $\{\|\widehat{\boldsymbol{\Sigma}}_U - \boldsymbol{\Sigma}\| \leq 4v(t/n)^{1/2}\}$, it follows that

$$
|\widehat{\lambda}_\ell^{1/2} - \overline{\lambda}_\ell^{1/2}| \lesssim v\sqrt{t}\,(np)^{-1/2} + p^{-1/2} \quad \text{and} \quad \|\widehat{\boldsymbol{v}}_\ell\|_\infty \lesssim p^{-1/2} \tag{C.28}
$$

as long as $n \ge v^2 p^{-1} t$. Write $\widehat{\boldsymbol{v}}_\ell = (\widehat{v}_{\ell 1}, \dots \widehat{v}_{\ell p})^{\mathrm{T}}$. It follows that, with probability at least $1 - 2pe^{-t}$,

$$\|\widehat{\boldsymbol{b}}_j - \widetilde{\boldsymbol{b}}_j\|_2 = \left\{ \sum_{\ell=1}^{K} (\widehat{\lambda}_\ell^{1/2} - \overline{\lambda}_\ell^{1/2})^2 \, \widehat{v}_{\ell j}^2 \right\}^{1/2} \lesssim p^{-1}(v\sqrt{t}\, n^{-1/2} + 1)$$

for all $1 \le j \le p$. Similarly,

$$\|\widetilde{\boldsymbol{b}}_j - \boldsymbol{b}_j\|_2 = \left\{ \sum_{\ell=1}^{K} \overline{\lambda}_\ell (\widehat{v}_{\ell j} - \overline{v}_{\ell j})^2 \right\}^{1/2} \le \max_{1 \le \ell \le K} \overline{\lambda}_\ell^{1/2} \cdot \sqrt{K}\, \|\widehat{\boldsymbol{v}}_\ell - \overline{\boldsymbol{v}}_\ell\|_\infty \lesssim v\sqrt{t}\,(np)^{-1/2} + p^{-1/2}.$$

By taking $t = \log(np)$, the previous two displays together imply (22).

Next we consider $\max_{1 \le j \le p} |\widehat{\sigma}_{\varepsilon,jj} - \sigma_{\varepsilon,jj}|$. Note that with probability at least $1 - 4pe^{-t}$, $\max_{1 \le j \le p} |\widehat{\theta}_j - \mathbb{E}(X_j^2)| \lesssim (t/n)^{1/2}$ as long as $n \gtrsim t$. Therefore, it suffices to focus on $\|\widehat{\boldsymbol{b}}_j\|_2^2 - \|\boldsymbol{b}_j\|_2^2$, which can be written as $\sum_{\ell=1}^K (\widehat{\lambda}_\ell - \overline{\lambda}_\ell)\widehat{v}_{\ell j}^2 + \sum_{\ell=1}^K \overline{\lambda}_\ell (\widehat{v}_{\ell j}^2 - \overline{v}_{\ell j}^2)$. Under Assumption 2, it follows from (20) and (21) that on the event $\{\|\widehat{\boldsymbol{\Sigma}}_U - \boldsymbol{\Sigma}\| \le 4v(t/n)^{1/2}\}$,

$$\begin{aligned}
&|\|\widehat{\boldsymbol{b}}_j\|_2^2 - \|\boldsymbol{b}_j\|_2^2| \\
&\le \sum_{\ell=1}^{K} |\widehat{\lambda}_\ell - \overline{\lambda}_\ell| \|\widehat{\boldsymbol{v}}_\ell\|_\infty^2 + \sum_{\ell=1}^{K} \overline{\lambda}_\ell (\|\widehat{\boldsymbol{v}}_\ell\|_\infty + \|\overline{\boldsymbol{v}}_\ell\|_\infty) \|\widehat{\boldsymbol{v}}_\ell - \overline{\boldsymbol{v}}_\ell\|_\infty \\
&\lesssim v\sqrt{t}\,(np)^{-1/2} + p^{-1/2}
\end{aligned}$$

as long as $n \ge v^2 p^{-1} t$, which proves (23) by taking $t = \log(np)$. $\square$

C.6 Proof of Theorem 4

For $\widehat{\mu}_j$'s and $\widehat{\theta}_{jk}$'s with $\tau_j = a_j (n/t_1)^{1/2}$ and $\tau_{jk} = a_{jk}(n/t_2)^{1/2}$, it follows from Lemma C.2 and the union bound that as long as $n \ge 8 \max(t_1, t_2)$,

$$\max_{1 \le j \le p} |\widehat{\mu}_j - \mu_j| \le 4 \max_{1 \le j \le p} a_j \sqrt{\frac{t_1}{n}} \quad \text{and} \quad \max_{1 \le j \le k \le p} |\widehat{\theta}_{jk} - \mathbb{E}(X_j X_k)| \le 4 \max_{1 \le j \le k \le p} a_{jk} \sqrt{\frac{t_2}{n}}$$

with probability at least $1 - 2pe^{-t_1} - (p^2 + p)e^{-t_2}$. In particular, taking $t_1 = \log(np)$ and $t_2 = \log(np^2)$ implies that as long as $n \gtrsim \log(np)$, $\|\widehat{\boldsymbol{\Sigma}}_{\mathrm{H}} - \boldsymbol{\Sigma}\|_{\max} \lesssim w_{n,p}^{-1}$ with probability greater than $1 - 4n^{-1}$.

The rest of the proof is similar to that of Theorem 3, simply with the following modifications. Under Assumption 2, it follows from (C.9) and (C.10) in Lemma C.6 that, with probability at least $1 - 4n^{-1}$,

$$
\begin{aligned}
|\widetilde{\lambda}_\ell^{1/2} - \overline{\lambda}_\ell^{1/2}| = |\widetilde{\lambda}_\ell - \overline{\lambda}_\ell|/(\widetilde{\lambda}_\ell^{1/2} + \overline{\lambda}_\ell^{1/2}) \lesssim \sqrt{p}\,(w_{n,p}^{-1} + p^{-1}), \\
\|\overline{\boldsymbol{v}}_\ell\|_\infty = \|\overline{\boldsymbol{b}}_\ell\|_\infty / \|\overline{\boldsymbol{b}}_\ell\|_2 \le \|\mathbf{B}\|_{\max}/\|\overline{\boldsymbol{b}}_\ell\|_2 \lesssim p^{-1/2}, \\
\|\widetilde{\boldsymbol{v}}_\ell - \overline{\boldsymbol{v}}_\ell\|_\infty \lesssim p^{-1/2} w_{n,p}^{-1} + p^{-1} \quad \text{and} \quad \|\widetilde{\boldsymbol{v}}_\ell\|_\infty \lesssim p^{-1/2}.
\end{aligned}
$$

Plugging the above bounds into the proof of Theorem 3 proves the conclusions. $\square$

C.7 Proof of Theorem 5

The key of the proof is to show that $T_j(\mathbf{B})$ provides a good approximation of T_j° uniformly over $1 \le j \le p$. To begin with, note that the estimator $\widehat{\theta}_j$ with $\tau_{jj} = a_{jj}(n/t)^{1/2}$ for $a_{jj} \ge \mathrm{var}(X_j^2)^{1/2}$ satisfies $\mathbb{P}\{|\widehat{\theta}_j - \theta_j| \ge 4a_{jj}(t/n)^{1/2}\} \le 2e^{-t}$, where $\theta_j = \mathbb{E}(X_j^2)$. Together with the union bound, this yields that with probability greater than $1 - 2pe^{-t}$,

$$
\max_{1\le j\le p} |\widehat{\theta}_j - \theta_j| \le 4 \max_{1 \le j \le p} a_{jj}^{1/2} \sqrt{\frac{t}{n}} \tag{C.29}
$$

as long as $n \ge 8t$. Next, observe that

$$
\left| T_j(\mathbf{B}) - \sqrt{\frac{n}{\widehat{\sigma}_{\varepsilon,jj}}} (\widehat{\mu}_j - \boldsymbol{b}_j^{\mathrm{T}} \bar{\boldsymbol{f}}\,) \right| = \sqrt{\frac{n}{\widehat{\sigma}_{\varepsilon,jj}}} |\boldsymbol{b}_j^{\mathrm{T}} \{ \bar{\boldsymbol{f}} - \widehat{\boldsymbol{f}}(\mathbf{B}) \}| \le \sqrt{\frac{n}{\widehat{\sigma}_{\varepsilon,jj}}} \|\boldsymbol{b}_j\|_2 \|\widehat{\boldsymbol{f}}(\mathbf{B}) - \bar{\boldsymbol{f}}\|_2 \tag{C.30}
$$

and

$$
\left| \sqrt{\frac{n}{\widehat{\sigma}_{\varepsilon,jj}}} (\widehat{\mu}_j - \boldsymbol{b}_j^{\mathrm{T}} \bar{\boldsymbol{f}}\,) - T_j^\circ \right| \le \left| \frac{1}{\sqrt{\widehat{\sigma}_{\varepsilon,jj}}} - \frac{1}{\sqrt{\sigma_{\varepsilon,jj}}} \right| (|\sqrt{n}\,\widehat{\mu}_j| + \|\boldsymbol{b}_j\|_2 \|\sqrt{n} \bar{\boldsymbol{f}}\|_2). \tag{C.31}
$$

Applying Proposition 3 with $t = \log n$ shows that, with probability at least $1 - C_1 n^{-1}$,

$$
\|\widehat{\boldsymbol{f}}(\mathbf{B}) - \bar{\boldsymbol{f}}\|_2 \lesssim (K \log n)^{1/2} p^{-1/2}. \tag{C.32}
$$

Moreover, it follows from Lemma C.2, (C.6) and (C.29) that, with probability greater than $1 - 4pe^{-t_1} - e^{-t_2}$,

$$\max_{1\le j\le p} |\widehat{\mu}_j - \mu_j| \lesssim \sqrt{\frac{t_1}{n}}, \quad \max_{1\le j\le p} \left| \frac{\widehat{\sigma}_{\varepsilon,jj}}{\sigma_{\varepsilon,jj}} - 1 \right| \lesssim \sqrt{\frac{t_1}{n}} \quad \text{and} \quad \|\bar{\boldsymbol{f}}\|_2 \lesssim \sqrt{\frac{K+t_2}{n}}.$$

Taking $t_1 = \log(np)$ and $t_2 = \log n$, we deduce from (C.30)–(C.32) that, with probability at least $1 - C_2 n^{-1}$,

$$\max_{j\in\mathcal{H}_0} |T_j(\mathbf{B}) - T_j^\circ| \lesssim \{K + \log(np)\} n^{-1/2} + (Kn\log n)^{1/2} p^{-1/2}. \tag{C.33}$$

Based on (C.33), the rest of the proof is almost identical to that of Theorem 1 and therefore is omitted. □

C.8 Proof of Theorem A.1

For convenience, we write $\widehat{\boldsymbol{b}}_j = \widehat{\boldsymbol{b}}_j(\mathcal{X}_1)$ for $j = 1, \ldots, p$, which are the estimated loading vectors using the first half of the data. Let $\widehat{\boldsymbol{f}}(\mathcal{X}_2)$ be the estimator of $\bar{\boldsymbol{f}}$ obtained by solving (26) using only the second half of the data and with $\boldsymbol{b}_j$'s replaced by $\widehat{\boldsymbol{b}}_j$'s.

We keep the notation used in Section 3.2.2, but with all the estimators constructed from $\mathcal{X}_1$ instead of the whole data set. Recall that $\widehat{\mathbf{B}} = (\widehat{\boldsymbol{b}}_1, \ldots, \widehat{\boldsymbol{b}}_p)^{\mathrm{T}} = (\widetilde{\lambda}_1^{1/2}\widehat{\boldsymbol{v}}_1, \ldots, \widetilde{\lambda}_K^{1/2}\widehat{\boldsymbol{v}}_K)$. Following the proof of Theorem 4, we see that as long as $n \gtrsim \log(np)$, the event $\mathcal{E}_{\max} := \{\|\widehat{\boldsymbol{\Sigma}}_{\mathrm{H}} - \boldsymbol{\Sigma}\|_{\max} \lesssim w_{n,p}^{-1}\}$ occurs with probability at least $1 - 4n^{-1}$. On $\mathcal{E}_{\max}$, we have

$$\max_{1\le \ell\le K} |\widetilde{\lambda}_\ell^{1/2} - \overline{\lambda}_\ell^{1/2}| \lesssim \sqrt{p}\,(w_{n,p}^{-1} + p^{-1}) \quad \text{and} \quad \max_{1\le\ell\le K} \|\widehat{\boldsymbol{v}}_\ell\|_\infty \lesssim p^{-1/2},$$

which, combined with the pervasiveness assumption $\overline{\lambda}_\ell \asymp p$, implies $\max_{1\le\ell\le K} \widetilde{\lambda}_\ell^{1/2} \lesssim \sqrt{p}$. Moreover, write $\boldsymbol{\delta}_j = \widehat{\boldsymbol{b}}_j - \boldsymbol{b}_j$ for $1 \le j \le p$ and note that

$$\widehat{\mathbf{B}}^{\mathrm{T}}\widehat{\mathbf{B}} - \mathbf{B}^{\mathrm{T}}\mathbf{B} = \sum_{j=1}^p (\widehat{\boldsymbol{b}}_j\widehat{\boldsymbol{b}}_j^{\mathrm{T}} - \boldsymbol{b}_j\boldsymbol{b}_j^{\mathrm{T}}) = \sum_{j=1}^p \boldsymbol{\delta}_j\boldsymbol{\delta}_j^{\mathrm{T}} + 2\sum_{j=1}^p \boldsymbol{\delta}_j\boldsymbol{b}_j^{\mathrm{T}}.$$

It follows that $\|p^{-1}(\widehat{\mathbf{B}}^{\mathrm{T}}\widehat{\mathbf{B}} - \mathbf{B}^{\mathrm{T}}\mathbf{B})\| \le \max_{1\le j\le p}(\|\boldsymbol{\delta}_j\|_2^2 + 2\|\boldsymbol{b}_j\|_2\|\boldsymbol{\delta}_j\|_2)$. Again, from the proof of Theorem 4 we see that on the event $\mathcal{E}_{\max}$, $\|p^{-1}(\widehat{\mathbf{B}}^{\mathrm{T}}\widehat{\mathbf{B}} - \mathbf{B}^{\mathrm{T}}\mathbf{B})\| \lesssim w_{n,p}^{-1} + p^{-1/2}$.

Under Assumption 3, putting the above calculations together yields that with probability greater than $1 - 4n^{-1}$,

$$\lambda_{\min}(p^{-1}\widehat{\mathbf{B}}^{\mathrm{T}}\widehat{\mathbf{B}}) \geq \frac{c_l}{2} \quad \text{and} \quad \|\widehat{\mathbf{B}}\|_{\max} \leq C_1$$

as long as $n \gtrsim \log(np)$. By the independence between $\widehat{\boldsymbol{b}}_j$'s and $\mathcal{X}_2$, the conclusion of Proposition 3 holds for $\widehat{\boldsymbol{f}}(\mathcal{X}_2)$.

Next, recall that

$$T_j = \sqrt{\frac{n}{\widehat{\sigma}_{\varepsilon,jj}}}\{\widehat{\mu}_j - \widehat{\boldsymbol{b}}_j^{\mathrm{T}}\widehat{\boldsymbol{f}}(\mathcal{X}_2)\},$$

where $\widehat{\mu}_j$'s and $\widehat{\sigma}_{\varepsilon,jj}$'s are all constructed from $\mathcal{X}_2$. Note that

$$|\sqrt{n}\{\widehat{\mu}_j - \widehat{\boldsymbol{b}}_j^{\mathrm{T}}\widehat{\boldsymbol{f}}(\mathcal{X}_2)\} - \sqrt{n}\{\widehat{\mu}_j - \boldsymbol{b}_j^{\mathrm{T}}\bar{\boldsymbol{f}}\}| \leq \sqrt{n}\|\widehat{\boldsymbol{b}}_j\|_2\|\widehat{\boldsymbol{f}}(\mathcal{X}_2) - \bar{\boldsymbol{f}}\|_2 + \sqrt{n}\|\bar{\boldsymbol{f}}\|_2\|\widehat{\boldsymbol{b}}_j - \boldsymbol{b}_j\|_2.$$

This, together with (28), Theorem 4 and (C.6), implies that with probability at least $1 - C_2n^{-1}$,

$$\begin{aligned}
&\max_{1\leq j\leq p}|\sqrt{n}\{\widehat{\mu}_j - \widehat{\boldsymbol{b}}_j^{\mathrm{T}}\widehat{\boldsymbol{f}}(\mathcal{X}_2)\} - \sqrt{n}\{\widehat{\mu}_j - \boldsymbol{b}_j^{\mathrm{T}}\bar{\boldsymbol{f}}\}| \\
&\qquad \lesssim (Kn\log n)^{1/2}p^{-1/2} + (K + \log n)^{1/2}(w_{n,p}^{-1} + p^{-1/2}).
\end{aligned}$$

Following the proof of Theorem 5, it can be shown that with probability at least $1 - C_3n^{-1}$,

$$\max_{j\in\mathcal{H}_0}|T_j - T_j^{\circ}| \lesssim (Kn\log n)^{1/2}p^{-1/2} + \{K + \log(np)\}n^{-1/2}.$$

The rest of the proof is almost identical to that of Theorem 1 and therefore is omitted. $\square$

D Additional proofs

In this section, we prove Propositions 2 and 3 in the main text, and Lemmas C.1–C.6 in Section C.

D.1 Proof of Proposition 2

By Weyl's inequality and the decomposition that $\widehat{\mathbf{\Sigma}} = \mathbf{B}\mathbf{B}^{\mathrm{T}} + (\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}) + \mathbf{\Sigma}_{\varepsilon}$, we have

$$\max_{1\le \ell \le K} |\widehat{\lambda}_\ell - \overline{\lambda}_\ell| \le \|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_2 + \|\mathbf{\Sigma}_\varepsilon\|_2 \quad \text{and} \quad \max_{K+1\le \ell \le p} |\widehat{\lambda}_\ell| \le \|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_2 + \|\mathbf{\Sigma}_\varepsilon\|_2,$$

where $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_p$ are the eigenvalues of $\widehat{\mathbf{\Sigma}}$ in a non-increasing order. Thus, (20) follows immediately. Next, applying Corollary 1 in Yu *et al.* (2015) to the pair $(\widehat{\mathbf{\Sigma}}, \mathbf{B}\mathbf{B}^{\mathrm{T}})$ gives that, for every $1 \le \ell \le K$,

$$\|\widehat{\boldsymbol{v}}_\ell - \overline{\boldsymbol{v}}_\ell\|_2 \le \frac{2^{3/2}\|(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}) + \mathbf{\Sigma}_\varepsilon\|_2}{\min(\overline{\lambda}_{\ell-1} - \overline{\lambda}_\ell, \overline{\lambda}_\ell - \overline{\lambda}_{\ell+1})},$$

where we put $\overline{\lambda}_0 = \infty$ and $\overline{\lambda}_{K+1} = 0$. Under Assumption 2, this proves (21). $\square$

D.2 Proof of Proposition 3

To begin with, we introduce the following notation. Define the loss function $L_\gamma(\boldsymbol{w}) = p^{-1}\sum_{j=1}^p \ell_\gamma(\bar{X}_j - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{w})$ for $\boldsymbol{w} \in \mathbb{R}^K$, $\boldsymbol{w}^* = \bar{\boldsymbol{f}}$ and $\widehat{\boldsymbol{w}} = \mathrm{argmin}_{\boldsymbol{w}\in\mathbb{R}^K} L_\gamma(\boldsymbol{w})$. Without loss of generality, we assume $\|\mathbf{B}\|_{\max} \le 1$ for simplicity.

Define an intermediate estimator $\widehat{\boldsymbol{w}}_\eta = \boldsymbol{w}^* + \eta(\widehat{\boldsymbol{w}} - \boldsymbol{w}^*)$ such that $\|\widehat{\boldsymbol{w}}_\eta - \boldsymbol{w}^*\|_2 \le r$ for some $r > 0$ to be specified below (D.7). We take $\eta = 1$ if $\|\widehat{\boldsymbol{w}} - \boldsymbol{w}^*\|_2 \le r$; otherwise, we choose $\eta \in (0, 1)$ so that $\|\widehat{\boldsymbol{w}}_\eta - \boldsymbol{w}^*\|_2 = r$. Then, it follows from Lemma A.1 in Sun *et al.* (2017) that

$$\langle \nabla L_\gamma(\widehat{\boldsymbol{w}}_\eta) - \nabla L_\gamma(\boldsymbol{w}^*), \widehat{\boldsymbol{w}}_\eta - \boldsymbol{w}^*\rangle \le \eta\langle \nabla L_\gamma(\widehat{\boldsymbol{w}}) - \nabla L_\gamma(\boldsymbol{w}^*), \widehat{\boldsymbol{w}} - \boldsymbol{w}^*\rangle, \tag{D.1}$$

where $\nabla L_\gamma(\widehat{\boldsymbol{w}}) = \mathbf{0}$ according to the Karush-Kuhn-Tucker condition. By the mean value theorem for vector-valued functions, we have

$$\nabla L_\gamma(\widehat{\boldsymbol{w}}_\eta) - \nabla L_\gamma(\boldsymbol{w}^*) = \int_0^1 \nabla^2 L_\gamma((1-t)\boldsymbol{w}^* + t\widehat{\boldsymbol{w}}_\eta)\, dt\, (\widehat{\boldsymbol{w}}_\eta - \boldsymbol{w}^*).$$

If, there exists some constant $a_{\min} > 0$ such that

$$\min_{\boldsymbol{w}\in\mathbb{R}^K:\|\boldsymbol{w}-\boldsymbol{w}^*\|_2\le r} \lambda_{\min}(\nabla^2 L_\gamma(\boldsymbol{w})) \ge a_{\min}, \tag{D.2}$$

then it follows $a_{\min}\|\widehat{\boldsymbol{w}}_\eta - \boldsymbol{w}^*\|_2^2 \le -\eta\langle \nabla L_\gamma(\boldsymbol{w}^*), \widehat{\boldsymbol{w}} - \boldsymbol{w}^*\rangle \le \|\nabla L_\gamma(\boldsymbol{w}^*)\|_2\|\widehat{\boldsymbol{w}}_\eta - \boldsymbol{w}^*\|_2$, or equivalently,

$$a_{\min}\|\widehat{\boldsymbol{w}}_\eta - \boldsymbol{w}^*\|_2 \le \|\nabla L_\gamma(\boldsymbol{w}^*)\|_2, \tag{D.3}$$

where $\nabla L_\gamma(\boldsymbol{w}^*) = -p^{-1}\sum_{j=1}^p \psi_\gamma(\mu_j + \bar{\varepsilon}_j)\boldsymbol{b}_j$.

First we verify (D.2). Write $\mathbf{S} = p^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{B}$ and note that

$$\nabla^2 L_\gamma(\boldsymbol{w}) = \frac{1}{p}\sum_{j=1}^p \boldsymbol{b}_j\boldsymbol{b}_j^{\mathrm{T}} I(|\bar{X}_j - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{w}| \le \gamma),$$

where $\bar{X}_j - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{w} = \boldsymbol{b}_j^{\mathrm{T}}(\boldsymbol{w}^* - \boldsymbol{w}) + \mu_j + \bar{\varepsilon}_j$. Then, for any $\boldsymbol{u} \in \mathbb{S}^{K-1}$ and $\boldsymbol{w} \in \mathbb{R}^K$ satisfying $\|\boldsymbol{w} - \boldsymbol{w}^*\|_2 \le r$,

$$\begin{aligned}
&\boldsymbol{u}^{\mathrm{T}}\nabla^2 L_\gamma(\boldsymbol{w})\boldsymbol{u} \\
&\ge \boldsymbol{u}^{\mathrm{T}}\mathbf{S}\boldsymbol{u} - \frac{1}{p}\sum_{j=1}^p(\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{u})^2 I(|\bar{\varepsilon}_j + \mu_j| > \gamma/2) - \frac{1}{p}\sum_{j=1}^p(\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{u})^2 I\{|\boldsymbol{b}_j^{\mathrm{T}}(\boldsymbol{w}^* - \boldsymbol{w})| > \gamma/2\} \\
&\ge \boldsymbol{u}^{\mathrm{T}}\mathbf{S}\boldsymbol{u} - \max_{1\le j\le p}\|\boldsymbol{b}_j\|_2^2\left\{\frac{1}{p}\sum_{j=1}^p I(|\bar{\varepsilon}_j + \mu_j| > \gamma/2) + \frac{4}{\gamma^2}\|\boldsymbol{w} - \boldsymbol{w}^*\|_2^2\,\boldsymbol{u}^{\mathrm{T}}\mathbf{S}\boldsymbol{u}\right\}.
\end{aligned}$$

By Assumption 3, $\lambda_{\min}(\mathbf{S}) \ge c_l$ for some constant $c_l > 0$ and $\max_{1\le j\le p}\|\boldsymbol{b}_j\|_2^2 \le K$. Therefore, as long as $\gamma > 2r\sqrt{K}$ we have

$$\min_{\boldsymbol{w}\in\mathbb{R}^K:\|\boldsymbol{w}-\boldsymbol{w}^*\|_2\le r} \lambda_{\min}(\nabla^2 L_\gamma(\boldsymbol{w})) \ge (1 - 4\gamma^{-2}r^2K)c_l - \frac{K}{p}\sum_{j=1}^p I(|\bar{\varepsilon}_j + \mu_j| > \gamma/2), \tag{D.4}$$

To bound the last term on the right-hand side of (D.4), it follows from Hoeffding's inequality that for any $t > 0$,

$$\frac{1}{p}\sum_{j=1}^p I(|\bar{\varepsilon}_j + \mu_j| > \gamma/2) \le \frac{1}{p}\sum_{j=1}^p \mathbb{P}(|\bar{\varepsilon}_j + \mu_j| > \gamma/2) + \sqrt{\frac{t}{2p}}$$

with probability at least $1-e^{-t}$. This, together with (D.4) and the inequality

$$\frac{1}{p}\sum_{j=1}^{p}\mathbb{P}(|\bar{\varepsilon}_j+\mu_j|>\gamma/2)\le\frac{4}{\gamma^2 p}\sum_{j=1}^{p}(\mu_j^2+\mathbb{E}\bar{\varepsilon}_j^2)=4\gamma^{-2}(p^{-1}\|\boldsymbol{\mu}\|_2^2+n^{-1}\overline{\sigma}_\varepsilon^2)$$

implies that, with probability greater than $1-e^{-t}$,

$$\min_{\boldsymbol{w}\in\mathbb{R}^K:\|\boldsymbol{w}-\boldsymbol{w}^*\|_2\le r}\lambda_{\min}(\nabla^2 L_\gamma(\boldsymbol{w}))\ge\frac{3}{4}c_l-K\sqrt{\frac{t}{2p}}-\frac{4K}{\gamma^2}\left(\frac{\|\boldsymbol{\mu}\|_2^2}{p}+\frac{\overline{\sigma}_\varepsilon^2}{n}\right)\tag{D.5}$$

as long as $\gamma\ge 4r\sqrt{K}$.

Next we bound $\|\nabla L_\gamma(\boldsymbol{w}^*)\|_2$. For every $1\le\ell\le K$, we write $\Psi_\ell=p^{-1}\sum_{j=1}^p\psi_{j\ell}:=p^{-1}\sum_{j=1}^p\gamma^{-1}\psi_\gamma(\mu_j+\bar{\varepsilon}_j)b_{j\ell}$, such that $\|\nabla L_\gamma(\boldsymbol{w}^*)\|_2\le\sqrt{K}\,\|\nabla L_\gamma(\boldsymbol{w}^*)\|_\infty=\gamma\sqrt{K}\max_{1\le\ell\le d}|\Psi_\ell|$. Recall that, for any $u\in\mathbb{R}$, $-\log(1-u+u^2)\le\gamma^{-1}\psi_\gamma(\gamma u)\le\log(1+u+u^2)$. After some simple algebra, we obtain that

$$\begin{aligned}e^{\psi_{j\ell}}&\le\{1+\gamma^{-1}(\mu_j+\bar{\varepsilon}_j)+\gamma^{-2}(\mu_j+\bar{\varepsilon}_j)^2\}^{b_{j\ell}I(b_{j\ell}\ge 0)}\\&\qquad+\{1-\gamma^{-1}(\mu_j+\bar{\varepsilon}_j)+\gamma^{-2}(\mu_j+\bar{\varepsilon}_j)^2\}^{-b_{j\ell}I(b_{j\ell}<0)}\\&\le 1+\gamma^{-1}(\mu_j+\bar{\varepsilon}_j)b_{j\ell}+\gamma^{-2}(\mu_j+\bar{\varepsilon}_j)^2.\end{aligned}$$

Taking expectation on both sides gives

$$\mathbb{E}(e^{\psi_{j\ell}})\le 1+\gamma^{-1}|\mu_j|+\gamma^{-2}(\mu_j^2+n^{-1}\sigma_{\varepsilon,jj}).$$

Moreover, by independence and the inequality $1+t\le e^t$, we get

$$\begin{aligned}\mathbb{E}(e^{p\Psi_\ell})=\prod_{j=1}^{p}\mathbb{E}(e^{\psi_{j\ell}})&\le\exp\left\{\frac{1}{\gamma}\sum_{j=1}^{p}|\mu_j|+\frac{1}{\gamma^2}\sum_{j=1}^{p}\left(\mu_j^2+\frac{\sigma_{\varepsilon,jj}}{n}\right)\right\}\\&\le\exp\left(\frac{\|\boldsymbol{\mu}\|_1}{\gamma}+\frac{\|\boldsymbol{\mu}\|_2^2}{\gamma^2}+\frac{\overline{\sigma}_\varepsilon^2 p}{\gamma^2 n}\right).\end{aligned}$$

For any $t>0$, it follows from Markov's inequality that

$$\mathbb{P}(p\Psi_j\ge 2t)\le e^{-2t}\mathbb{E}(e^{p\Psi_\ell})\le\exp\left\{\frac{\|\boldsymbol{\mu}\|_1}{\gamma}+\frac{\|\boldsymbol{\mu}\|_2^2}{\gamma^2}+\frac{\overline{\sigma}_\varepsilon^2 p}{\gamma^2 n}-2t\right\}\le\exp(1-t)$$

provided

$$\gamma \geq \max\left\{\|\boldsymbol{\mu}\|_1, \overline{\sigma}_\varepsilon \sqrt{\frac{\|\boldsymbol{\mu}\|_2^2/\overline{\sigma}_\varepsilon^2 + p/n}{t}}\right\}. \tag{D.6}$$

Under the constraint (D.6), it can be similarly shown that $\mathbb{P}(-p\Psi_j \geq 2t) \leq e^{1-t}$. Putting the above calculations together, we conclude that

$$\begin{aligned} &\mathbb{P}\bigg\{\|\nabla L_\gamma(\boldsymbol{w}^*)\|_2 \geq \sqrt{K}\,\frac{2\gamma t}{p}\bigg\} \\ &\leq \mathbb{P}\bigg\{\|\nabla L_\gamma(\boldsymbol{w}^*)\|_\infty \geq \frac{2\gamma t}{p}\bigg\} \leq \sum_{\ell=1}^K \mathbb{P}(|p\Psi_\ell| \geq 2t) \leq 2eK\exp(-t). \end{aligned} \tag{D.7}$$

With the above preparations, now we are ready to prove the final conclusion. It follows from (D.5) that with probability greater than $1-e^{-t}$, (D.2) holds with $a_{\min} = c_l/4$, provided that $\gamma \geq 4\sqrt{K}\max\{r, c_l^{-1/2}(\|\boldsymbol{\mu}\|_2^2/p + \overline{\sigma}_\varepsilon^2/n)^{1/2}\}$ and $p \geq 8c_l^{-2}K^2t$. Hence, combining (D.3) and (D.7) with $r = \frac{\gamma}{4\sqrt{K}}$ yields that, with probability at least $1-(1+2eK)e^{-t}$, $\|\widehat{\boldsymbol{w}}_\eta - \boldsymbol{w}^*\|_2 \leq 8c_l^{-1}\sqrt{K}\,p^{-1}\gamma t < r$ as long as $p > 32c_l^{-1}Kt$. By the definition of $\widehat{\boldsymbol{w}}_\eta$, we must have $\eta = 1$ and thus $\widehat{\boldsymbol{w}} = \widehat{\boldsymbol{w}}_\eta$. □

D.3 Proof of Lemma C.1

Let $1 \leq j \leq p$ be fixed and define the function $h(\theta) = \mathbb{E}\{\ell_\tau(X_j - \theta)\}$, $\theta \in \mathbb{R}$. By the optimality of $\mu_{j,\tau}$ and the mean value theorem, we have $h'(\mu_{j,\tau}) = 0$ and

$$h''(\widetilde{\mu}_{j,\tau})(\mu_j - \mu_{j,\tau}) = h'(\mu_j) - h'(\mu_{j,\tau}) = h'(\mu_j) = -\mathbb{E}\{\psi_\tau(\xi_j)\}, \tag{D.8}$$

where $\widetilde{\mu}_{j,\tau} = \lambda\mu_j + (1-\lambda)\mu_{j,\tau}$ for some $0 \leq \lambda \leq 1$. Since $\mathbb{E}(\xi_j) = 0$, we have $-\mathbb{E}\{\psi_\tau(\xi_j)\} = \mathbb{E}\{\xi_j I(|\xi_j| > \tau) - \tau I(\xi_j > \tau) + \tau I(\xi_j < -\tau)\}$, which implies

$$|\mathbb{E}\{\psi_\tau(\xi_j)\}| \leq \tau^{1-\kappa}\upsilon_{\kappa,j}. \tag{D.9}$$

Next we consider $h''(\widetilde{\mu}_{j,\tau}) = \mathbb{P}(|X_j - \widetilde{\mu}_{j,\tau}| \leq \tau)$. Since h is a convex function that is minimized at $\mu_{j,\tau}$, $h(\widetilde{\mu}_{j,\tau}) \leq \lambda h(\mu_j) + (1-\lambda)h(\mu_{j,\tau}) \leq h(\mu_j) \leq \sigma_{jj}/2$. On the other hand, note that $h(\theta) \geq \mathbb{E}\{(\tau|X_j - \theta| - \tau^2/2)1(|X_j - \theta| > \tau)\}$ for all $\theta \in \mathbb{R}$. Combining these upper and lower bounds on $h(\widetilde{\mu}_{j,\tau})$ with Markov's inequality gives

$$\begin{aligned}
&\tau\mathbb{E}\{|X_j-\widetilde{\mu}_{j,\tau}|I(|X_j-\widetilde{\mu}_{j,\tau}|>\tau)\} \\
&\leq \frac{1}{2}\tau^2\mathbb{P}(|X_j-\widetilde{\mu}_{j,\tau}|>\tau)+\frac{1}{2}\sigma_{jj} \leq \frac{1}{2}\tau\,\mathbb{E}\{|X_j-\widetilde{\mu}_{j,\tau}|I(|X_j-\widetilde{\mu}_{j,\tau}|>\tau)\}+\frac{1}{2}\sigma_{jj},
\end{aligned}$$

which further implies that for every $0 \leq \lambda \leq 1$,

$$\mathbb{P}(|X_j-\widetilde{\mu}_{j,\tau}|>\tau) \leq \tau^{-1}\mathbb{E}\{|X_j-\widetilde{\mu}_{j,\tau}|1(|X_j-\widetilde{\mu}_{j,\tau}|>\tau)\} \leq \sigma_{jj}\tau^{-2}.$$

Together with (D.8) and (D.9), this proves (C.1). □

D.4 Proof of Lemma C.3

Throughout the proof, we let $1 \leq j \leq p$, $a \geq \sigma_{jj}^{1/2}$, $t \geq 1$ be fixed and write $\tau = a(n/t)^{1/2}$ with $n \geq 8t$. The dependence of τ on (a, n, t) will be assumed without displaying. First we introduce the following notations. Define functions $L(\theta) = -\sum_{i=1}^n \ell_\tau(X_{ij}-\theta)$, $\zeta(\theta) = L(\theta) - \mathbb{E}L(\theta)$ and $w^2(\theta) = -\frac{d^2}{d\theta^2}\mathbb{E}L(\theta)$, such that $\widehat{\mu}_j = \operatorname{argmax}_{\theta\in\mathbb{R}} L(\theta)$. Moreover, we write

$$w_0^2 := w^2(\mu_j) = \alpha_\tau n \quad \text{with} \quad \alpha_\tau = \mathbb{P}(|X_j-\mu_j| \leq \tau). \tag{D.10}$$

For every $r > 0$, define the parameter set

$$\Theta_0(r) = \{\theta \in \mathbb{R} : |w_0(\theta-\mu_j)| \leq r\}. \tag{D.11}$$

Then, it follows from Lemma C.2 that

$$\mathbb{P}\{\widehat{\mu}_j \in \Theta_0(r_0)\} \geq 1 - 2\exp(-t), \tag{D.12}$$

where $r_0 = 4a(\alpha_\tau t)^{1/2}$. Based on this result, we only need to focus on the local neighborhood $\Theta_0(r_0)$ of μ_j. The rest of the proof is based on Proposition 3.1 in Spokoiny (2013). To this end, we need to check Conditions $(\mathcal{L}_0)$ and (ED_2) there.

Condition ($\mathcal{L}_0$): Note that, for every $\theta \in \Theta_0(r)$,

$$|w_0^{-1} w^2(\theta) w_0^{-1} - 1| = |\alpha_\tau^{-1} - 1 - \alpha_\tau^{-1} \mathbb{P}(|X_j - \theta| > \tau)|$$
$$\leq \alpha_\tau^{-1} \max[1 - \alpha_\tau, \{\sigma_{jj} + (\theta - \mu_j)^2\}\tau^{-2}].$$

By Chebyshev's inequality, we have $1 \geq \alpha_\tau \geq 1 - \sigma_{jj}\tau^{-2} \geq 7/8$. Therefore,

$$|w_0^{-1} w^2(\theta) w_0^{-1} - 1| \leq \alpha_\tau^{-1}\{\sigma_{jj} + (\alpha_\tau n)^{-1} r^2\}\tau^{-2}.$$

This verifies Condition ($\mathcal{L}_0$) by taking

$$\delta(r) = \alpha_\tau^{-1}\sigma_{jj}\tau^{-2} + \alpha_\tau^{-2}\tau^{-2} n^{-1} r^2, \quad r > 0.$$

Condition (ED_2): Note that $\zeta''(\theta) = -\sum_{i=1}^n \{1(|X_{ij} - \theta| \leq \tau) - \mathbb{P}(|X_{ij} - \theta| \leq \tau)\}$. For every $\lambda \in \mathbb{R}$ satisfying $|\lambda| \leq \alpha_\tau \sqrt{n}$, using the inequalities $1+u \leq e^u$ and $e^u \leq 1+u+u^2 e^{u\vee 0}/2$ we deduce that

$$\begin{aligned} \mathbb{E}\exp\{\lambda\sqrt{n}\zeta''(\theta)/w_0^2\} &= \prod_{i=1}^n \mathbb{E}\exp[-\lambda w_0^{-2}\sqrt{n}\{I(|X_{ij} - \theta| \leq \tau) - \mathbb{P}(|X_{ij} - \theta| \leq \tau)\}] \\ &\leq \prod_{i=1}^n \{1 + (1/2)\lambda^2 w_0^{-4} n \exp(|\lambda| w_0^{-2}\sqrt{n})\} \\ &\leq \prod_{i=1}^n \{1 + (e/2)\alpha_\tau^{-2}\lambda^2 n^{-1}\} \leq \exp\{(e/2)\alpha_\tau^{-2}\lambda^2\}. \end{aligned}$$

This verifies Condition (ED_2) by taking $\omega = n^{-1/2}$, $\nu_0 = e^{1/2}\alpha_\tau^{-1}$ and $\mathbf{g}(r) = \alpha_\tau\sqrt{n}$, $r > 0$. Now, using Proposition 3.1 in Spokoiny (2013) we obtain that as long as $\alpha_\tau^2 n \geq 4 + 2t$,

$$\sup_{\theta \in \Theta_0(r)} |\alpha_\tau \sqrt{n}(\theta - \mu_j) + n^{-1/2}\{L'(\theta) - L'(\mu_j)\}|$$
$$\leq \alpha_\tau^{1/2}\delta(r) r + 6\alpha_\tau^{-1/2} e^{1/2}(2t+4)^{1/2} n^{-1/2} r$$

with probability greater than $1 - e^{-t}$. Under the conditions that $n \geq 8t$ and $t \geq 1$, it is easy to see that $\alpha_\tau^2 n \geq (7/8)^2 \cdot 8t \geq 6t \geq 4 + 2t$. Moreover, observe that

$$\sup_{\boldsymbol{\theta} \in \Theta_0(r)} |(\alpha_\tau - 1)\sqrt{n}(\theta - \mu_j)| \leq \alpha_\tau^{-1/2} \sigma_{jj} \tau^{-2} r.$$

The last two displays, together with (D.12) and the fact that $L'(\widehat{\mu}_j) = 0$ prove (C.2) by taking $r = r_0$. The proof of Lemma C.3 is then complete. □

D.5 Proof of Lemma C.4

Under model (1), we have $\xi_j = \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f} + \varepsilon_j$, where $\mathbb{E}(\varepsilon_j) = 0$ and ε_j and $\boldsymbol{f}$ are independent. Therefore,

$$\begin{aligned} &\mathbb{E}_{\boldsymbol{f}} \psi_\tau(\xi_j) - \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f} \\ &= -\mathbb{E}_{\boldsymbol{f}}(\varepsilon_j + \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f} - \tau) I(\varepsilon_j > \tau - \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}) + \mathbb{E}_{\boldsymbol{f}}(-\varepsilon_j - \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f} - \tau) I(\varepsilon_j < -\tau - \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}). \end{aligned}$$

Therefore, as long as $\tau > |\boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}|$, we have for any $q \in [2, \kappa]$ that

$$|\mathbb{E}_{\boldsymbol{f}} \psi_\tau(\xi_j) - \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}| \leq \mathbb{E}_{\boldsymbol{f}}\{|\varepsilon_j| I(|\varepsilon_j| > \tau - |\boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}|)\} \leq (\tau - |\boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}|)^{1-q} \, \mathbb{E}(|\varepsilon_j|^q)$$

almost surely. This proves (C.3) by taking q to be 2 and κ.

For the conditional variance, observe that

$$\mathbb{E}_{\boldsymbol{f}}\{\psi_\tau(\xi_j) - \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}\}^2 = \mathrm{var}_{\boldsymbol{f}}\{\psi_\tau(\xi_j)\} + \{\mathbb{E}_{\boldsymbol{f}} \psi_\tau(\xi_j) - \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}\}^2 \tag{D.13}$$

and that $\psi_\tau(\xi_j) - \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}$ can be written as

$$\varepsilon_j I(|\boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f} + \varepsilon_j| \leq \tau) + (\tau - \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}) I(\boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f} + \varepsilon_j > \tau) - (\tau + \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}) I(\boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f} + \varepsilon_j < -\tau),$$

which further implies

$$\begin{aligned} &\{\psi_\tau(\xi_j) - \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}\}^2 \\ &= \varepsilon_j^2 I(|\boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f} + \varepsilon_j| \leq \tau) + (\tau - \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f})^2 I(\boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f} + \varepsilon_j > \tau) + (\tau + \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f})^2 I(\boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f} + \varepsilon_j < -\tau). \end{aligned}$$

Taking conditional expectation on both sides yields

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{f}}\{\psi_\tau(\xi_j) - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}\}^2 \\
&= \mathbb{E}(\varepsilon_j^2) - \mathbb{E}_{\boldsymbol{f}}\{\varepsilon_j^2 I(|\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f} + \varepsilon_j| > \tau)\} \\
&\quad + (\tau - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f})^2 \mathbb{P}_{\boldsymbol{f}}(\varepsilon_j > \tau - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}) + (\tau + \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f})^2 \mathbb{P}_{\boldsymbol{f}}(\varepsilon_j < -\tau - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}).
\end{aligned}$$

Using the equality $u^2 = 2\int_0^u t\,dt$ for $u > 0$ we deduce that as long as $\tau > |\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}|$,

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{f}}\varepsilon_j^2 I(\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f} + \varepsilon_j > \tau) \\
&= 2\mathbb{E}_{\boldsymbol{f}} \int_0^\infty I(\varepsilon_j > t) I(\varepsilon_j > \tau - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}) t\,dt \\
&= 2\mathbb{E}_{\boldsymbol{f}} \int_0^{\tau - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}} I(\varepsilon_j > \tau - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}) t\,dt + 2\mathbb{E}_{\boldsymbol{f}} \int_{\tau - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}}^\infty I(\varepsilon_j > t) t\,dt \\
&= (\tau - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f})^2 \mathbb{P}_{\boldsymbol{f}}(\varepsilon_j > \tau - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}) + 2\int_{\tau - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}}^\infty \mathbb{P}(\varepsilon_j > t) t\,dt.
\end{aligned}$$

Analogously, it can be shown that

$$\mathbb{E}_{\boldsymbol{f}}\{\varepsilon_j^2 I(\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f} + \varepsilon_j < -\tau)\} = (\tau + \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f})^2 \mathbb{P}_{\boldsymbol{f}}(\varepsilon_j < -\tau - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}) + 2\int_{\tau + \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}}^\infty \mathbb{P}(-\varepsilon_j > t) t\,dt.$$

Together, the last three displays imply

$$\begin{aligned}
&0 \geq \mathbb{E}_{\boldsymbol{f}}\{\psi_\tau(\xi_j) - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}\}^2 - \mathbb{E}(\varepsilon_j^2) \\
&\geq -2\int_{\tau - |\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}|}^\infty \mathbb{P}(|\varepsilon_j| > t) t\,dt \geq -2\mathbb{E}(|\varepsilon_j|^\kappa) \int_{\tau - |\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}|}^\infty t^{1-\kappa}\,dt = -\frac{2}{\kappa - 2} \frac{\mathbb{E}(|\varepsilon_j|^\kappa)}{(\tau - |\boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}|)^{\kappa - 2}}.
\end{aligned}$$

Combining this with (D.13) and (C.3) proves (C.4).

Finally, we study the covariance $\mathrm{cov}_{\boldsymbol{f}}(\psi_\tau(\xi_j), \psi_\tau(\xi_k))$ for $j \neq k$. By definition,

$$\begin{aligned}
&\mathrm{cov}_{\boldsymbol{f}}(\psi_\tau(\xi_j), \psi_\tau(\xi_k)) \\
&= \mathbb{E}_{\boldsymbol{f}}\{\psi_\tau(\xi_j) - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f} + \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f} - \mathbb{E}_{\boldsymbol{f}}\psi_\tau(\xi_j)\}\{\psi_\tau(\xi_k) - \boldsymbol{b}_k^{\mathrm{T}}\boldsymbol{f} + \boldsymbol{b}_k^{\mathrm{T}}\boldsymbol{f} - \mathbb{E}_{\boldsymbol{f}}\psi_\tau(\xi_k)\} \\
&= \underbrace{\mathbb{E}_{\boldsymbol{f}}\{\psi_\tau(\xi_j) - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}\}\{\psi_\tau(\xi_k) - \boldsymbol{b}_k^{\mathrm{T}}\boldsymbol{f}\}}_{\Pi_1} - \underbrace{\{\mathbb{E}_{\boldsymbol{f}}\psi_\tau(\xi_j) - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}\}\{\mathbb{E}_{\boldsymbol{f}}\psi_\tau(\xi_k) - \boldsymbol{b}_k^{\mathrm{T}}\boldsymbol{f}\}}_{\Pi_2}.
\end{aligned}$$

Recall that $\psi_\tau(\xi_j) - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f} = \varepsilon_j I(|\xi_j| \leq \tau) + (\tau - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}) I(\xi_j > \tau) - (\tau + \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}) I(\xi_j < -\tau)$.

Hence,

$$
\begin{aligned}
\Pi_1 = {} & \mathbb{E}_{\boldsymbol{f}} \varepsilon_j \varepsilon_k I(|\xi_j| \le \tau, |\xi_k| \le \tau) + (\tau - \boldsymbol{b}_k^{\mathrm{T}} \boldsymbol{f}) \mathbb{E}_{\boldsymbol{f}} \varepsilon_j I(|\xi_j| \le \tau, \xi_k > \tau) \\
& - (\tau + \boldsymbol{b}_k^{\mathrm{T}} \boldsymbol{f}) \mathbb{E}_{\boldsymbol{f}} \varepsilon_j I(|\xi_j| \le \tau, \xi_k < -\tau) + (\tau - \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}) \mathbb{E}_{\boldsymbol{f}} \varepsilon_k I(\xi_j > \tau, |\xi_k| \le \tau) \\
& + (\tau - \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f})(\tau - \boldsymbol{b}_k^{\mathrm{T}} \boldsymbol{f}) \mathbb{E}_{\boldsymbol{f}} I(\xi_j > \tau, \xi_k > \tau) - (\tau - \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f})(\tau + \boldsymbol{b}_k^{\mathrm{T}} \boldsymbol{f}) \mathbb{E}_{\boldsymbol{f}} I(\xi_j > \tau, \xi_k < -\tau) \\
& - (\tau + \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}) \mathbb{E}_{\boldsymbol{f}} \varepsilon_k I(\xi_j < -\tau, |\xi_k| \le \tau) - (\tau + \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f})(\tau - \boldsymbol{b}_k^{\mathrm{T}} \boldsymbol{f}) \mathbb{E}_{\boldsymbol{f}} I(\xi_j < -\tau, \xi_k > \tau) \\
& + (\tau + \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f})(\tau + \boldsymbol{b}_k^{\mathrm{T}} \boldsymbol{f}) \mathbb{E}_{\boldsymbol{f}} I(\xi_j < -\tau, \xi_k < -\tau). \qquad \text{(D.14)}
\end{aligned}
$$

Note that the first term on the right-hand side of (D.14) can be written as

$$
\begin{aligned}
& \mathbb{E}_{\boldsymbol{f}} \varepsilon_j \varepsilon_k I(|\xi_j| \le \tau, |\xi_k| \le \tau) \\
& = \mathrm{cov}(\varepsilon_j, \varepsilon_k) - \mathbb{E}_{\boldsymbol{f}} \varepsilon_j \varepsilon_k I(|\xi_j| > \tau) - \mathbb{E}_{\boldsymbol{f}} \varepsilon_j \varepsilon_k I(|\xi_k| > \tau) + \mathbb{E}_{\boldsymbol{f}} \varepsilon_j \varepsilon_k I(|\xi_j| > \tau, |\xi_k| > \tau),
\end{aligned}
$$

where

$$
|\mathbb{E}_{\boldsymbol{f}} \varepsilon_j \varepsilon_k I(|\xi_j| > \tau)| \le |\tau - \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}|^{2-\kappa} \mathbb{E}(|\varepsilon_j|^{\kappa-1} |\varepsilon_k|) \le 2^{\kappa-2} \tau^{2-\kappa} (\mathbb{E}|\varepsilon_j|^{\kappa})^{(\kappa-1)/\kappa} (\mathbb{E}|\varepsilon_k|^{\kappa})^{1/\kappa}
$$

and

$$
\begin{aligned}
& |\mathbb{E}_{\boldsymbol{f}} \varepsilon_j \varepsilon_k I(|\xi_j| > \tau, |\xi_k| > \tau)| \\
& \le |\tau - \boldsymbol{b}_j^{\mathrm{T}} \boldsymbol{f}|^{2-\kappa} \mathbb{E}(|\varepsilon_j|^{\kappa/2} |\varepsilon_k|^{\kappa/2}) \le 2^{\kappa-2} \tau^{2-\kappa} (\mathbb{E}|\varepsilon_j|^{\kappa})^{1/2} (\mathbb{E}|\varepsilon_k|^{\kappa})^{1/2}
\end{aligned}
$$

almost surely on $\mathcal{G}_{jk}$. The previous three displays together imply

$$
|\mathbb{E}_{\boldsymbol{f}} \varepsilon_j \varepsilon_k I(|\xi_j| \le \tau, |\xi_k| \le \tau) - \mathrm{cov}(\varepsilon_j, \varepsilon_k)| \lesssim \tau^{2-\kappa}
$$

almost surely on $\mathcal{G}_{jk}$. For the remaining terms on the right-hand side of (D.14), it can be similarly obtained that, almost surely on $\mathcal{G}_{jk}$,

$$
\begin{aligned}
|\mathbb{E}_{\boldsymbol{f}}\varepsilon_j I(|\xi_j| \le \tau, \xi_k > \tau)| &\le |\tau - \boldsymbol{b}_k^{\mathrm{T}}\boldsymbol{f}|^{1-\kappa}\mathbb{E}(|\varepsilon_j||\varepsilon_k|^{\kappa-1}), \\
|\mathbb{E}_{\boldsymbol{f}}\varepsilon_j I(|\xi_j| \le \tau, \xi_k < -\tau)| &\le |\tau + \boldsymbol{b}_k^{\mathrm{T}}\boldsymbol{f}|^{1-\kappa}\mathbb{E}(|\varepsilon_j||\varepsilon_k|^{\kappa-1}), \\
\text{and} \quad \mathbb{E}_{\boldsymbol{f}} I(\xi_j > \tau, \xi_k < -\tau) &\le |\tau - \boldsymbol{b}_j^{\mathrm{T}}\boldsymbol{f}|^{-\kappa/2}|\tau + \boldsymbol{b}_k^{\mathrm{T}}\boldsymbol{f}|^{-\kappa/2}\mathbb{E}(|\varepsilon_j\varepsilon_k|^{\kappa/2}).
\end{aligned}
$$

Putting together the pieces, we get $|\Pi_1 - \mathrm{cov}(\varepsilon_j, \varepsilon_k)| \lesssim \upsilon_{jk}\tau^{2-\kappa}$ almost surely on $\mathcal{G}_{jk}$. For Π_2, it follows directly from (C.3) that $|\Pi_2| \lesssim \upsilon_{jk}^2\tau^{2-2\kappa}$ almost surely on $\mathcal{G}_{jk}$. These bounds, combined with the fact that $\mathrm{cov}_{\boldsymbol{f}}(\psi_\tau(\xi_j), \psi_\tau(\xi_k)) = \Pi_1 - \Pi_2$, yield (C.5). $\square$

D.6 Proof of Lemma C.5

For any $\boldsymbol{u} \in \mathbb{R}^K$, by independence we have

$$
\begin{aligned}
\mathbb{E}\exp(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{f}_i) &\le \exp(C_1\|\boldsymbol{f}\|_{\psi_2}^2\|\boldsymbol{u}\|_2^2) \quad \text{for all } i = 1, \ldots, n, \qquad \text{(D.15)} \\
\text{and} \quad \mathbb{E}\exp(\sqrt{n}\,\boldsymbol{u}^{\mathrm{T}}\bar{\boldsymbol{f}}) &= \prod_{i=1}^n \mathbb{E}\exp(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{f}_i/\sqrt{n}) \le \exp(C_1\|\boldsymbol{f}\|_{\psi_2}^2\|\boldsymbol{u}\|_2^2),
\end{aligned}
$$

where $C_1 > 0$ is an absolute constant. From Theorem 2.1 in Hsu *et al.* (2012) we see that, for any $t > 0$,

$$
\begin{aligned}
\mathbb{P}\{\|\sqrt{n}\bar{\boldsymbol{f}}\|_2^2 > 2C_1\|\boldsymbol{f}\|_{\psi_2}^2(K + 2\sqrt{Kt} + 2t)\} &\le e^{-t} \\
\text{and} \quad \mathbb{P}\{\|\boldsymbol{f}_i\|_2^2 > 2C_1\|\boldsymbol{f}\|_{\psi_2}^2(K + 2\sqrt{Kt} + 2t)\} &\le e^{-t}, \quad i = 1, \ldots, n.
\end{aligned}
$$

This proves (C.6) and (C.7) by the union bound.

For $\widehat{\boldsymbol{\Sigma}}_f$, applying Theorem 5.39 in Vershynin (2012) yields that, with probability at least $1 - 2e^{-t}$, $\|\widehat{\boldsymbol{\Sigma}}_f - \mathbf{I}_K\| \le \max(\delta, \delta^2)$, where $\delta = C_2\|\boldsymbol{f}\|_{\psi_2}^2 n^{-1/2}(K+t)^{1/2}$ and $C_2 > 0$ is an absolute constant. Conclusion (C.8) then follows immediately. $\square$

D.7 Proof of Lemma C.6

For each $1 \le \ell \le K$, as $\overline{\lambda}_\ell > 0$ and by Weyl's inequality, we have $|\widetilde{\lambda}_\ell - \overline{\lambda}_\ell| \le |\lambda_\ell(\widehat{\mathbf{\Sigma}}_{\mathrm{H}}) - \overline{\lambda}_\ell| \le \|\widehat{\mathbf{\Sigma}}_{\mathrm{H}} - \mathbf{\Sigma}\| + \|\mathbf{\Sigma}_\varepsilon\|$. Moreover, note that for any matrix $\mathbf{E} \in \mathbb{R}^{d_1 \times d_2}$,

$$\|\mathbf{E}\|^2 \le \|\mathbf{E}\|_1 \vee \|\mathbf{E}\|_\infty \le (d_1 \vee d_2)\|\mathbf{E}\|_{\max}.$$

Putting the above calculations together proves (C.9).

Next, note that

$$\widehat{\mathbf{\Sigma}}_{\mathrm{H}} = \widehat{\mathbf{\Sigma}}_{\mathrm{H}} - \mathbf{\Sigma} + \mathbf{B}\mathbf{B}^{\mathrm{T}} + \mathbf{\Sigma}_\varepsilon = \sum_{\ell=1}^{K} \overline{\lambda}_\ell \overline{\boldsymbol{v}}_\ell \overline{\boldsymbol{v}}_\ell^{\mathrm{T}} + \widehat{\mathbf{\Sigma}}_{\mathrm{H}} - \mathbf{\Sigma} + \mathbf{\Sigma}_\varepsilon.$$

Under Assumption 2, it follows from Theorem 3 and Proposition 3 in Fan *et al.* (2018) that

$$\max_{1 \le \ell \le K} \|\widehat{\boldsymbol{v}}_\ell - \overline{\boldsymbol{v}}_\ell\|_\infty \le \frac{C}{p^{3/2}}(\|\widehat{\mathbf{\Sigma}}_{\mathrm{H}} - \mathbf{\Sigma}\|_\infty + \|\mathbf{\Sigma}_\varepsilon\|_\infty) \le C(p^{-1/2}\|\widehat{\mathbf{\Sigma}}_{\mathrm{H}} - \mathbf{\Sigma}\|_{\max} + p^{-1}\|\mathbf{\Sigma}_\varepsilon\|),$$

where we use the inequalities $\|\widehat{\mathbf{\Sigma}}_{\mathrm{H}} - \mathbf{\Sigma}\|_\infty \le p\|\widehat{\mathbf{\Sigma}}_{\mathrm{H}} - \mathbf{\Sigma}\|_{\max}$ and $\|\mathbf{\Sigma}_\varepsilon\|_\infty \le p^{1/2}\|\mathbf{\Sigma}_\varepsilon\|$ in the last step and $C > 0$ is a constant independent of (n, p). This proves (C.10) . □

E Additional numerical results on FDP/FDR control

In the end, we present some additional simulation results that complement Section 4.5. Under Models 2 and 3 defined in Section 4.2, we compare the numerical performance of the five tests regarding FDP/FDR control. We take $\alpha = 0.05$, $p = 500$ and let n gradually increase from 100 to 200. The empirical FDP is defined as the average false discovery proportion based on 200 simulations. The simulation results are presented in Figures E.1 and E.2, respectively.

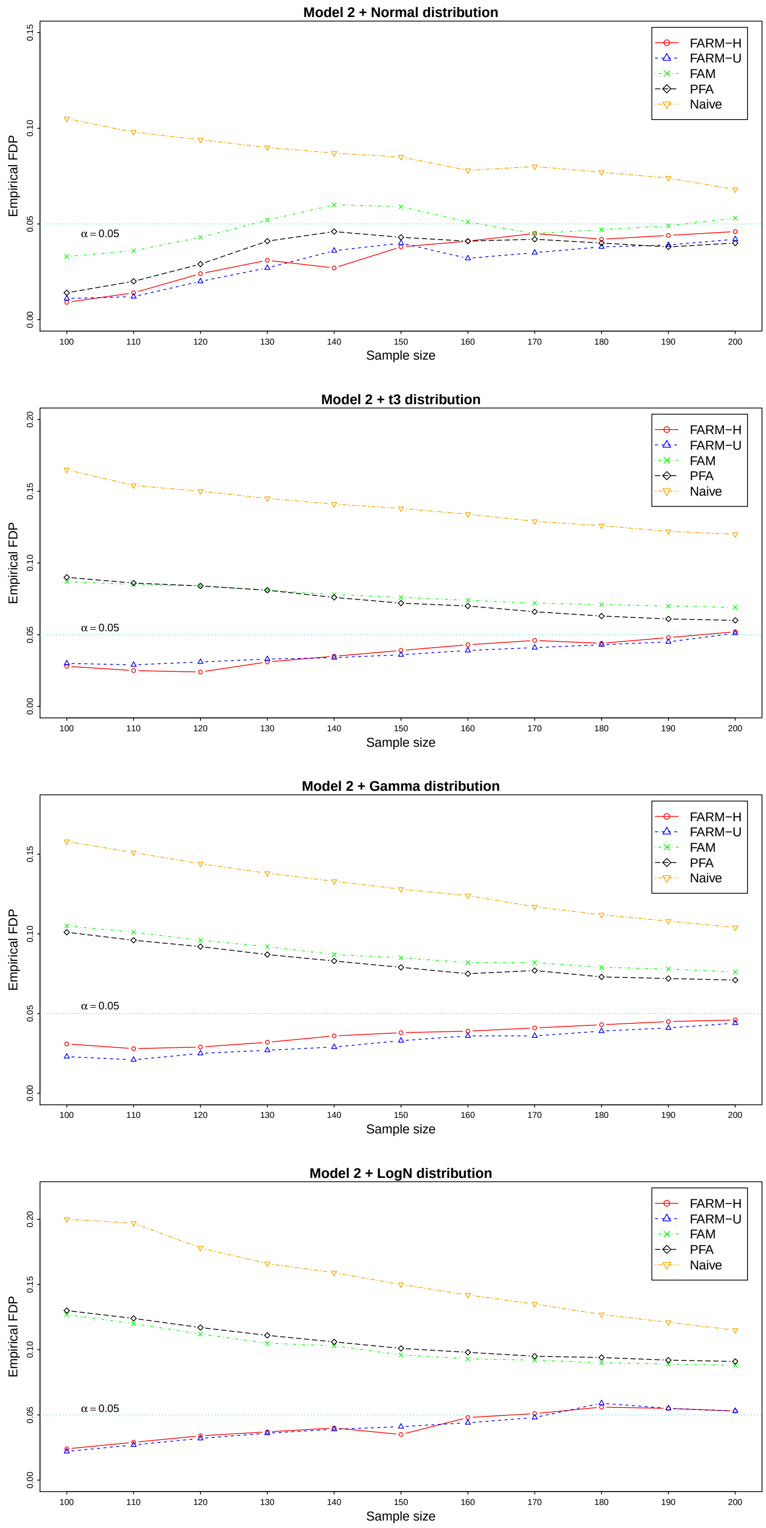

Figure E.1: Empirical FDP versus sample size for the five tests at level $\alpha = 0.05$. The data are generated from Model 2 with $p = 500$ and sample size n ranging from 100 to 200 with a step size of 10. The panels from top to bottom correspond to the four error distributions in Section 4.2.

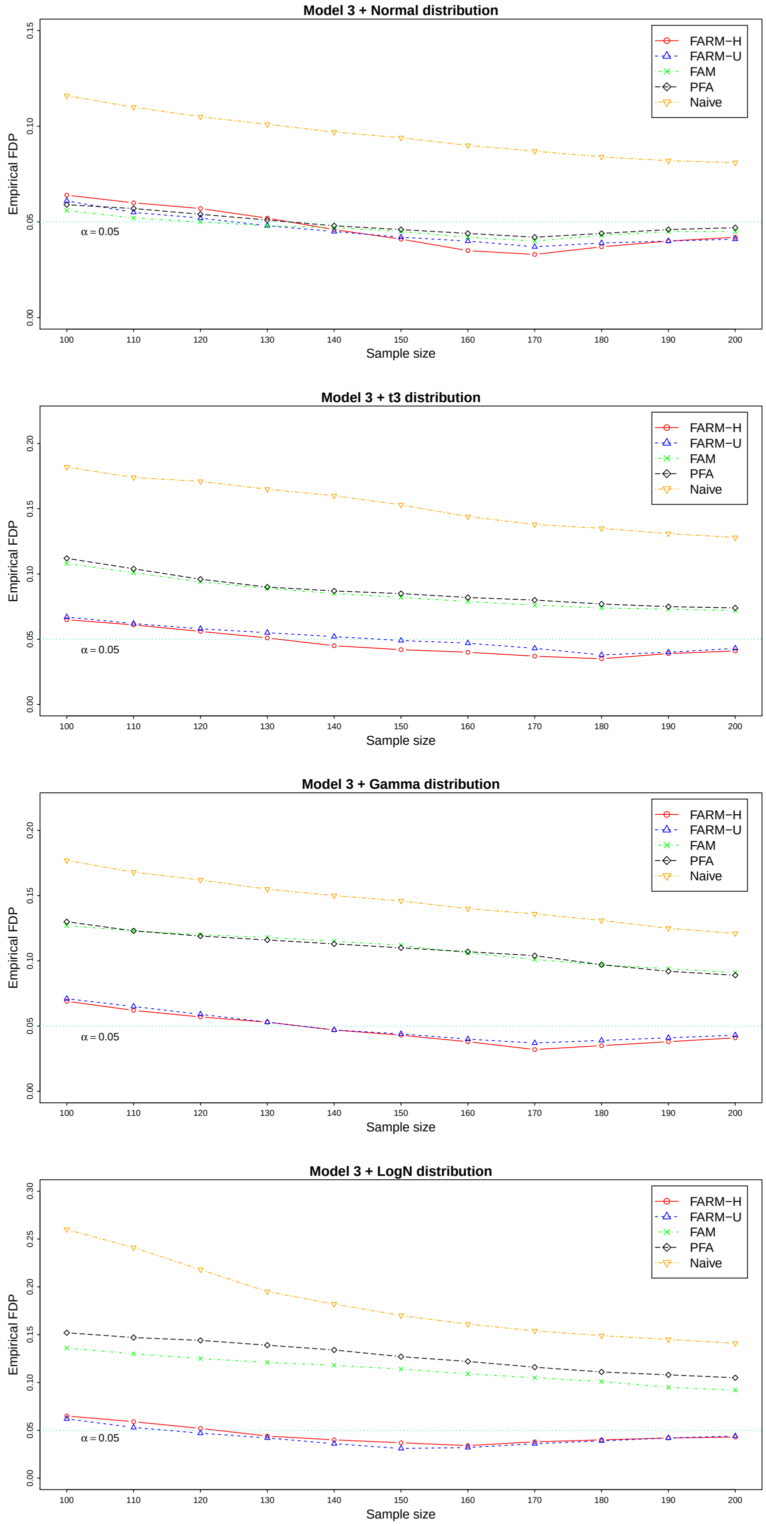

Figure E.2: Empirical FDP versus sample size for the five tests at level $\alpha = 0.05$. The data are generated from Model 3 with $p = 500$ and sample size n ranging from 100 to 200 with a step size of 10. The panels from top to bottom correspond to the four error distributions in Section 4.2.

References

BENTKUS, V. (2005). A Lyapunov-type bound in R^d. *Theory of Probability & Its Applications*, **49**, 311–323.

COX, D. R. (1975). A note on data-splitting for the evaluation of significance levels. *Biometrika*, **62**, 441–444.

FAN, J., LI, Q. and WANG, Y. (2017). Estimation of high dimensional mean regression in the absence of symmetry and light tail assumptions. *Journal of the Royal Statistical Society,* Series B, **79**, 247–265.

FAN, J., WANG, W. and ZHONG, Y. (2018). An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *Journal of Machine Learning Research*, **18**(207), 1–42.

HARTIGAN, J. A. (1969). Using subsample values as typical values. *Journal of the American Statistical Association*, **64**, 1303–1317.

HSU, D., KAKADE, S. M. and ZHANG, T. (2012). A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, **17**(52), 1–6.

LI, W. V. and SHAO, Q.-M. (2002). A normal comparison inequality and its applications. *Probability Theory and Related Fields*, **122**, 494–508.

LIEB, E. H. (1973). Convex trace functions and the Wigner-Yanase-Dyson conjecture. *Advances in Mathematics*, **11**, 267–288.

MINSKER, S. (2016). Sub-Gaussian estimators of the mean of a random matrix with heavy-tailed entries. Available at *arXiv:1605.0712.*

RINALDO A., WASSERMAN, L., G'SNELL, M., LEI, J. and TIBSHIRANI, R. (2016). Bootstrapping and sample splitting for high-dimensional, assumption-free inference. Available at *arXiv:1611.05401.*

SPOKOINY, V. (2013). Bernstein-von Mises theorem for growing parameter dimension. Available at *arXiv:1302.3430.*

SPOKOINY, V. and ZHILOVA, M. (2015). Bootstrap confidence sets under model misspecification. *The Annals of Statistics*, **43**, 2653–2675.

SUN, Q., ZHOU, W.-X. and FAN, J. (2017). Adaptive Huber regression: Optimality and phase transition. Available at *arXiv:1706.06991*.

Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* (eds Y. C. Eldar and G. Kutyniok), pp. 210–268. Cambridge: Cambridge University Press.

Yu, Y., Wang, T. and Samworth, R. J. (2015). A useful variant of the Davis-Kahan theorem for statisticians. *Biometrika*, **102**, 315–323.