Robust High-dimensional Tuning Free Multiple Testing[*]

Jianqing Fan   Zhipeng Lou   Mengxin Yu

November 22, 2022

Abstract

A stylized feature of high-dimensional data is that many variables have heavy tails, and robust statistical inference is critical for valid large-scale statistical inference. Yet, the existing developments such as Winsorization, Huberization and median of means require the bounded second moments and involve variable-dependent tuning parameters, which hamper their fidelity in applications to large-scale problems. To liberate these constraints, this paper revisits the celebrated Hodges-Lehmann (HL) estimator for estimating location parameters in both the one- and two-sample problems, from a non-asymptotic perspective. Our study develops Berry-Esseen inequality and Cramér type moderate deviation for the HL estimator based on newly developed non-asymptotic Bahadur representation, and builds data-driven confidence intervals via a weighted bootstrap approach. These results allow us to extend the HL estimator to large-scale studies and propose *tuning-free* and *moment-free* high-dimensional inference procedures for testing global null and for large-scale multiple testing with false discovery proportion control. It is convincingly shown that the resulting tuning-free and moment-free methods control false discovery proportion at a prescribed level. The simulation studies lend further support to our developed theory.

1 Introduction

Large-scale, high-dimensional data with rich structures have been widely collected in almost all scientific disciplines and humanities, thanks to the advancements of modern technologies. Massive developments have been made in statistics over the past two decades on extracting valuable information from these high dimensional data; see Bühlmann and Van De Geer (2011); Hastie et al. (2009, 2015); Wainwright (2019); Fan et al. (2020a); Chen et al. (2021) for a detailed account and references therein.

Despite convenience for theoretical analysis, the sub-Gaussian tails condition is not realistic in many applications that involve high-dimensional variables. For instance, it is well known that heavy-tailed distributions is a stylized feature for financial returns and macroeconomic variables (Cont, 2001; Stock and Watson, 2002; Fan and Yao, 2017; Fan et al., 2021). Therefore, tools designed for sub-Gaussian data can lead to erroneous scientific conclusions. Asking thousands of gene expressions to have all sub-Gaussian tails is a mathematical dream, not a reality that data scientists face. For example, comparing gene expression profiles between various cell sub-populations,

---

especially after treatments and therapies, is an essential statistical task (Nagalakshmi et al., 2008; Shendure and Ji, 2008; Wang et al., 2009; Li et al., 2012; Li and Tibshirani, 2013; Gupta et al., 2014; Finotello and Di Camillo, 2015). However, it is unrealistic to hope all thousands of gene expressions have sub-Gaussian distributions: outliers in non-sub-Gaussian distributions can have a significant impact on nonrobust procedures and lead to many false positives and negatives (Gupta et al., 2014; Wang et al., 2015). The situation also arises for inferences using functional magnetic resonance imaging (fMRI) data since the data do not conform to the assumed Gaussian distribution (Eklund et al., 2016). These practical challenges demand for developing efficient and reliable robust inference methods.

Recently, robust statistical methods have gained popularity as a mean of resolving outliers and heavy-tailed noises. Many preceding arts have taken a significant stride toward effective statistical estimation under heavy-tailed distributions. For instance, aiming at dealing with the heavy-tailed noise contamination, the Huber regression is proposed (Huber, 1973), and subsequent publications along these lines include Yohai and Maronna (1979); Mammen (1989); He and Shao (1996, 2000), where the asymptotic properties of the Huber estimator have been thoroughly investigated. From a non-asymptotic perspective, the Huber-type estimator was recently revisited by Sun et al. (2020), in which the authors propose an adaptive Huber regression method and establish its non-asymptotic deviation bounds by only requiring finite $(1+\delta)$-th moment of the noise with any $\delta > 0$. Moreover, using a similar idea for making the correspondent M-estimator insensitive to extreme values, Catoni (2012) developed a novel approach through minimizing a robust empirical loss. It is demonstrated that the estimator has exponential concentration around the true mean and enjoys the same statistical rate as the sample average for sub-Gaussian distributions when the population only has a bounded second moment. Brownlees et al. (2015) further investigates empirical risk minimization based on the robust estimator proposed in Catoni (2012). Additionally, the so-called *median of means* strategy, which can be traced back to Nemirovsky and Yudin (1983), is another successful method for handling heavy-tailed distributions. By only requiring bounded second moment, it achieves the sub-Gaussian type of concentration around the population mean parameter. Minsker (2015) and Hsu and Sabato (2016) further generalize this idea to multivariate cases. Moreover, there also exists a series of works that focus on solving the issue caused by heavy-tailed noises using quantile-based robust estimation; see Arcones (1995); Koenker and Hallock (2001); Belloni and Chernozhukov (2011); Fan et al. (2014); Zheng et al. (2015) for more details. Furthermore, recently, in Fan et al. (2021); Yang et al. (2017); Fan et al. (2022c), under heavy-tailed contamination, they proposed a novel principle by simply truncating or shrinking the response variables appropriately to achieve sub-Gaussian rates, and they only require bounded second moment of the measurements. Additionally, the aforementioned methodologies can also be applied to a wide range of problems, such as matrix sensing, matrix completion, robust PCA, factor analysis, and neural networks. For interested readers, we refer to Minsker (2015); Hsu and Sabato (2016); Fan et al. (2017); Loh (2017); Minsker (2018); Goldstein et al. (2018); Wang et al. (2020); Fan et al. (2022b); Wang and Fan (2022); Fan et al. (2022a) for more details.

While many effective solutions have been developed to address the problem of heavy-tailedness, these solutions still have some potential shortcomings. In specific, the developments call for the

second moments to be bounded and are primarily based on shrinkage data, Huber-type of loss, median of means (Huber, 1973; Nemirovsky and Yudin, 1983; Catoni, 2012; Fan et al., 2021). More critically, Huberization, Winsorization and sample splitting introduce additional tuning parameters and these tuning paprameters should be variable-dependent that makes large-scale applications difficult and damages the fidelity of empirical results . Although, the quantile estimators (Koenker and Hallock, 2001) such as the median and the Hodge-Lehmann (HL) estimator (Hodges and Lehmann, 1963) can eliminate the restriction on moment conditions and tuning parameters selection, the empirical median is often less efficient and requires stronger distribution assumptions. In addition, the existing literature on the HL estimator focuses mainly on low-dimensional asymptotic analysis and can not be applied to large-scale inferences.

In this paper, we revisit the celebrated HL estimator (Hodges and Lehmann, 1963) and conduct non-asymptotic and large-scale theoretical studies for both one-sample and two-sample problems. For one-sample location estimation in the univariate case, we let $X_1, \ldots, X_n \in \mathbb{R}$ be independent and identically distributed (i.i.d.) random variables with

$$X_i = \theta + \xi_i, \ \ i = 1, \ldots, n, \tag{1.1}$$

where θ represents the location parameter of interest and $\xi_1, \ldots, \xi_n$ are i.i.d. random variables drawn from some unknown distribution. In this scenario, the HL estimator of θ is defined by

$$\widehat{\theta} = \text{median} \left\{ \frac{X_i + X_j}{2} : 1 \leq i < j \leq n \right\}. \tag{1.2}$$

By assuming the pseudomedian of ξ_1 to be zero, we derive *non-asymptotic* Bahadur representation of $\widehat{\theta}$. To the best of our knowledge, this is the first study of its kind on the non-asymptotic expansion of the HL estimator. From there, we also establish the Berry-Essen bound and moderate deviation for $\widehat{\theta}$ in the widest range. Furthermore, as there are multiple unknowns in the asymptotic distribution of $\widehat{\theta}$, including the density function of ξ_1 and the unknown location parameter θ, we then propose a weighted bootstrap approach to construct confidence intervals for θ based on data. These results and methods are essential for the large-scale inference.

In addition to the study of one-sample problem, two-sample location shift problems arise frequently in many scientific studies, including choosing genes that are expressed differently in normal and injured spinal cord, determining the effects of treatment between treated and control groups, finding change points, etc. To this end, we let $Y_1, \ldots, Y_m \in \mathbb{R}$ be another independent sample of i.i.d. random variables satisfying

$$Y_j = \theta^\circ + \varepsilon_j, \ \ j = 1, \ldots, m. \tag{1.3}$$

The primary goal is to conduct statistical inference for $\Theta = \theta - \theta^\circ$. In the sequel, following Hodges and Lehmann (1963), the two-sample HL estimator for Θ is given by

$$\widehat{\Theta} = \text{median}\{X_i - Y_j : i = 1, \ldots, n; j = 1, \ldots, m\}. \tag{1.4}$$

Instead of assuming the noises are generated from the same distribution (Hodges and Lehmann, 1963), we only require median$(\xi_1 - \varepsilon_1) = 0$, which is more general and allows random noises to have

different distributions. In a similar vein, we establish the non-asymptotic expansion of $\widehat{\Theta}$, investigate its asymptotic distributions, and calculate the confidence interval via bootstrap techniques. Again, the techniques and results developed can be applied to large-scale multiple testing problems.

There is a rich literature on large-scale multiple testing problems for location parameters (Benjamini and Hochberg, 1995; Storey, 2002, 2003; Genovese and Wasserman, 2004; Ferreira and Zwinderman, 2006; Chi, 2007; Blanchard and Roquain, 2009). However, most of these works assume the noise distributions sub-Gaussian. Moving away for sub-Gaussian assumptions, Fan et al. (2019) propose estimating mean vector via minimizing the Huber type loss and perform the false discovery proportion (FDP) control. However, leveraging Huber type estimators necessitates moment limits and introduces tuning parameters, making it hard to be applied in large-scale inference, as the tuning parameters should ideally be variable-dependent. Additionally, while the HL estimator enjoys tuning-free and moment-free qualities in the univariate setting, its behavior in high dimensions is largely unknown. To this end, in this research, we further expand the HL estimator to high-dimensional regimes prompted by the lack of tuning free large-scale multiple testing problems for heavy-tailed distributions.

In specific, we assume $\boldsymbol{X}_i = \boldsymbol{\theta} + \boldsymbol{\xi}_i$, $i \in [n]$ and $\boldsymbol{Y}_j = \boldsymbol{\theta}^\circ + \boldsymbol{\varepsilon}_j$, $j \in [m]$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top$ and $\boldsymbol{\theta}^\circ = (\theta_1^\circ, \ldots, \theta_p^\circ)^\top$ are p-dimensional vectors of unknown parameters and random noises $\boldsymbol{\xi}_i, i \in [n], \boldsymbol{\varepsilon}_j, j \in [m]$. Let $\boldsymbol{\Theta} = \boldsymbol{\theta} - \boldsymbol{\theta}^\circ$. For both one- and two-sample problems, we propose a carefully constructed Gaussian multiplier bootstrap to test global null hypotheses

$$H_0 : \theta_\ell \text{ or } \Theta_\ell = 0 \text{ for all } \ell \in [p] \text{ versus } H_1 : \theta_\ell \text{ or } \Theta_\ell \neq 0 \text{ for some } \ell \in [p], \qquad (1.5)$$

by extending the HL estimator to high-dimensional regimes. When the null hypothesis above is rejected, we then perform multiple testing, allowing weakly dependent measurements, and efficiently control the FDP. Compared with existing literature (Liu and Shao, 2014; Fan et al., 2019), our procedures do not involve any tuning parameters and moment conditions for testing global null and large-scale multiple testing. These theoretical finds are further supported by exhaustive numerical studies.

The main contributions of the paper can be summarized as follows:

- The existing studies on the HL estimator mainly focus on its asymptotic behavior, which is too weak for high-dimensional applications. In practice, however, it is crucial to understand the HL estimator's performance under finite sample, especially in high-dimensional and large-scale experiments. For this purpose, we first derive the non-asymptotic expansions of the HL estimators for both one-sample and two-sample problems.

- With the non-asymptotic expansions of the HL estimators, for both one- and two-sample problems, we derive its Berry-Essen type bounds and Cramér type moderate deviations, with the widest range. To deal with unknown components in the distribution, we further develop the weighted bootstrap to build data-driven confidence intervals. In addition, we also furnish the non-asymptotic analysis of the bootstrap estimator.

- Existing work on large-scale testing with heavy-tail errors typically involves additional tuning parameters and the moment conditions. In order to address these issues, we generalize the HL

estimator to large-scale studies and propose tuning-free and moment-free high-dimensional testing procedures. Additionally, we develop bootstrap methods for calculating critical values for large-scale applications. We show that the resulting false discovery proportion is well controlled.

1.1 Roadmap

In §2, we first set up the model and introduce basic settings. We then derive both one-sample and two-sample non-asymptotic expansions of the HL estimator, its Berry-Esseen bound and moderate deviations. In addition, as the asymptotic distribution of the estimator involves unknown quantities, in §3, we conduct multiplier bootstrap to construct valid data-driven confidence intervals. Moreover, §4 is devoted to extending the HL estimator to large-scale multiple testing problems. §5 contains comprehensive numerical studies to verify theoretical results. Finally we conclude the paper with some discussions in §6. All the proofs are deferred to the appendix.

1.2 Notation

For any integer m, we use $[m]$ to denote the set $[m] = \{1, 2, \ldots, m\}$. For any function $h : \mathbb{R} \to \mathbb{R}$, we denote $\|h\|_\infty = \sup_{z \in \mathbb{R}} |h(z)|$. Throughout this paper, we use $C, C_1, C_2, \ldots$ to denote universal positive constants whose values may vary at different places. We use $\mathbb{I}\{\cdot\}$ to denote the indicator function. For any set A, we use $|A|$ to denote its cardinality. For two positive sequences $\{a_n\}_{n \geq 1}$ and $\{b_n\}_{n \geq 1}$, we write $a_n = \mathcal{O}(b_n)$ or $a_n \lesssim b_n$ if there exists a positive constant C such that $a_n \leq C \cdot b_n$ and we write $a_n = o(b_n)$ if $a_n/b_n \to 0$. In addition, we define the pseudomedian of a distribution F to be the median of the distribution of $(Z_1 + Z_2)/2$, where Z_1 and Z_2 are independent, each with the same distribution F. Moreover, for any distribution F and constant c, we let $c \cdot F$ represent the distribution of the random variable $c \cdot X$, where X is the random variable drawn from F.

2 Estimation and Inference

This section is devoted to studying the non-asymptotic expansions of the Hodges-Lehmann estimator and conducting statistical estimation and inference for population location shift parameters. For both one-sample and two-sample problems, the theoretical properties, which are needed for large-scale inferences, are presented in the following sections.

2.1 One-sample Problem

Let $X_i = \theta + \xi_i$, $i \in [n]$, be i.i.d. real-valued random variables, where $\theta \in \mathbb{R}$ is the unknown location parameter of interest and $\xi_1, \ldots, \xi_n$ are i.i.d. random variables drawn from some unknown distribution. It is assumed that ξ_1 has a pseudomedian (Høyland, 1965) of zero, throughout this section. As a consequence, letting $U(t) = \mathbb{P}\{(X_1 + X_2)/2 \leq t\}$, it holds that $\theta = \inf\{t \in \mathbb{R} : U(t) \geq 1/2\}$. The HL estimator (Hodges and Lehmann, 1963) of θ is given by the median of all *Walsh*

averages of the observations $X_1, \ldots, X_n$, namely,

$$\widehat{\theta} = \mathrm{median}\{(X_i + X_j)/2 : i \neq j \in [n]\}. \tag{2.1}$$

Equivalently, if we define the U-process $U_n(t) = \{n(n-1)\}^{-1} \sum_{i \neq j \in [n]} \mathbb{I}\{(X_i + X_j)/2 \leq t\}$, the HL estimator $\widehat{\theta}$ in (2.1) can also be expressed as the sample median of the process $U_n(t)$, namely,

$$\widehat{\theta} = \inf\{t \in \mathbb{R} : U_n(t) \geq 1/2\}. \tag{2.2}$$

Let $F(t) = \mathbb{P}(\xi_1 \leq t)$ denote the cumulative distribution function of ξ_1 and $f(t) = F'(t)$ be its density function. We then present the non-asymptotic Bahadur representation of $\widehat{\theta}$ in the following theorem.

Theorem 2.1. Assume that there exist positive constants c_0 and κ_0 such that $\inf_{|\delta| \leq c_0} U'(\theta + \delta) \geq \kappa_0$. Then for any $z > 0$, we have

$$\mathbb{P}(|\widehat{\theta} - \theta| > z) \leq 2 \exp\{-n\kappa_0^2 (z \wedge c_0)^2\}. \tag{2.3}$$

Furthermore, assume that $\sup_{z \in \mathbb{R}} |f(z)| < \infty$ and there exist positive constants c_1 and κ_1 such that $\sup_{|\delta| \leq c_1} |U''(\theta + \delta)| \leq \kappa_1$. Then for any $z > 0$ such that $z = o(n)$, we have

$$\mathbb{P}\left\{\left|\widehat{\theta} - \theta - \frac{2}{nU'(\theta)} \sum_{i=1}^{n} \left\{\frac{1}{2} - F(-\xi_i)\right\}\right| > \frac{C_1(z \vee 1)}{n}\right\} \leq C_2 \exp(-z), \tag{2.4}$$

where C_1, C_2 are positive constants depending only on $c_0, \kappa_0, c_1, \kappa_1$ and $\|f\|_\infty$.

We note that existing works mainly study the asymptotic distribution of quantiles of U-statistics instead of non-asymptotic ones (Arcones, 1996). Asymptotic theory, however, is frequently less effective for theoretical studies in high-dimensional statistics (Wainwright, 2019). To fill in the blank, in Theorem 2.1, we present both the non-asymptotic deviation bound and linear approximation of the HL estimator $\widehat{\theta}$. It is worth mentioning that the HL estimator $\widehat{\theta}$ has sub-Gaussian tails without any moment constraints imposed on the noise ξ_1, whereas the Huber-type or winsorized estimator requires the existence of the second moment. Moreover, in contrast to Huber regression (Zhou et al., 2018; Sun et al., 2020) or truncation (Fan et al., 2021), which both require additional tuning parameters, HL-type estimation is tuning-free and thus more scalable.

Moreover, when the distribution of ξ_1 is symmetric around zero, θ reduces to the median of the distribution of X. In this scenario, the sample median $\widehat{\theta}_{\mathrm{med}} = \mathrm{median}\{X_1, \ldots, X_n\}$ serves as a plausible alternative robust estimator for θ. Under similar regularity conditions on the density function $f(t)$, the classical Bahadur representation for $\widehat{\theta}_{\mathrm{med}}$ reveals that

$$\mathbb{P}\left\{\left|\widehat{\theta}_{\mathrm{med}} - \theta - \frac{1}{nf(0)} \sum_{i=1}^{n} \left(\frac{1}{2} - \mathbb{I}\{\xi_i \leq 0\}\right)\right| > \frac{C \log n}{n^{3/4}}\right\} \leq Cn^{-c} \tag{2.5}$$

for any constant $c > 0$. Compared with (2.4) (by taking $z = \mathcal{O}(\log n)$), the linear approximation of HL estimator is much more accurate than that of the quantile estimator (Arcones, 1996).

2.1.1 Asymptotic Distribution

In addition to estimation, statistical inference is also essential in real-world applications. To this end, with the developed non-asymptotic expansion at hand, we next present the asymptotic distribution of the HL estimator $\widehat{\theta}$ in this section.

Let $\Phi(\cdot)$ denote the cumulative distribution function of standard normal random variable. The following theorem establishes a Berry-Esseen theorem for $\widehat{\theta}$.

Theorem 2.2. Under the conditions of Theorem 2.1, we have

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}\left\{\frac{\sqrt{n}(\widehat{\theta}-\theta)}{\sigma_\theta}\leq z\right\}-\Phi(z)\right|\leq\frac{C\log n}{\sqrt{n}},$$

where $C<\infty$ is a positive constant independent of n and

$$\sigma_\theta^2=\frac{4\operatorname{Var}\{F(-\xi_1)\}}{\{U'(\theta)\}^2}. \tag{2.6}$$

Theorem 2.2 establishes the asymptotic normality of $\widehat{\theta}$. When the distribution of ξ_1 is symmetric around zero, the asymptotic variance above reduces to $\sigma_\theta^2=1/[3\{U'(\theta)\}^2]$. Consequently, in view of (2.5) and (2.6), the asymptotic relative efficiency (ARE) between the HL estimator $\widehat{\theta}$ and the sample median $\widehat{\theta}_{\mathrm{med}}$ is $3\{U'(\theta)\}^2/4\{f(0)\}^2$ (Hodges and Lehmann, 1963). A concrete example is given in Table 1, where we summarize the ARE between $\widehat{\theta}$ and $\widehat{\theta}_{\mathrm{med}}$ for $\xi_1\sim t_\nu$. In particular,

ν	1	2	4	8	16	∞
ARE	0.75	1.04	1.25	1.37	1.43	1.50

Table 1: Asymptotic relative efficiency between the HL estimator and the sample median

when $\nu\geq2$, the HL estimator has a strictly smaller asymptotic variance than the sample median. The above example illustrates the effectiveness of the HL estimator over the quantile regression method.

Based on the non-asymptotic linear expansion in (2.4), we further derive the Cramér-type moderate deviation to quantify the relative error of the normal approximation for $\widehat{\theta}$ in the following theorem, which has important applications to large-scale inference (Fan et al., 2007; Liu and Shao, 2014; Xia et al., 2018; Zhou et al., 2018).

Theorem 2.3. Let $\{\delta_n\}_{n\geq1}$ be a sequence of positive numbers satisfying $\sqrt{n}\delta_n\to\infty$. Then, under the conditions of Theorem 2.1, we have

$$\left|\frac{\mathbb{P}\{\sqrt{n}(\widehat{\theta}-\theta)/\sigma_\theta>z\}}{1-\Phi(z)}-1\right|\leq C\left\{\frac{1+z^3}{\sqrt{n}}+(1+z)\delta_n+2(z\vee1)\sqrt{2\pi}\exp\left(\frac{z^2}{2}-\sqrt{n}\delta_n\right)\right\}, \tag{2.7}$$

uniformly for $0<z\leq o(\delta_n^{-1}\wedge n^{1/4}\sqrt{\delta_n})$, where $C<\infty$ is a positive constant independent of z and n. In particular, when $\delta_n\asymp n^{-1/6}$, we have

$$\left|\frac{\mathbb{P}\{\sqrt{n}(\widehat{\theta}-\theta)/\sigma_\theta>z\}}{1-\Phi(z)}-1\right|\to0,$$

uniformly for $0<z\leq o(n^{1/6})$.

It is worth mentioning that taking $\delta_n \asymp n^{-1/6}$ yields the wideest possible range $0 < z \leq o(n^{1/6})$ for the relative error in (2.7) to vanish, which is also optimal for the Cramér-type moderate deviation results (Petrov, 1975; Fan et al., 2007; Liu and Shao, 2014; Zhou et al., 2018; Fan et al., 2019; Chen and Zhou, 2020; Fang et al., 2020). Next, we proceed to estimate the location shift parameter between two distributions via the HL estimator.

2.2 Two-sample Problem

A variety of applications use two-sample location shift estimation and inference, such as testing gene differences, quantifying treatment effects, and detecting change points. Accordingly, this section examines the two-sample estimation and inference of the population location shift parameter.

Let $Y_j = \theta^\circ + \varepsilon_j$, $j \in [m]$, be another sample of i.i.d. real-valued random variables independent of $\{X_1, \ldots, X_n\}$, and we aim at constructing confidence interval for $\Theta = \theta - \theta^\circ$. Throughout this section, it is assumed that

$$\Theta = \inf\{t \in \mathbb{R} : \mathcal{U}(t) \geq 1/2\}, \qquad \mathcal{U}(t) = \mathbb{P}(X_1 - Y_1 \leq t). \tag{2.8}$$

The existing literature on HL estimators mainly deals with the case where ξ_1 and ε_1 are identically distributed (Hodges and Lehmann, 1963; Lehmann, 1963; Bauer, 1972; Rosenkranz, 2010). In contrast, it should be noted that the assumption imposed in (2.8) is satisfied as long as median$(\varepsilon_1 - \xi_1) = 0$, which is much more general than the identical distribution. In the sequel, following Hodges and Lehmann (1963), the two sample HL estimator for Θ is given by

$$\widehat{\Theta} = \text{median}\{X_i - Y_j : i \in [n], j \in [m]\}. \tag{2.9}$$

Before proceeding, we present the following assumption on the relative sample sizes of the involved random samples.

Assumption 2.1. There exists a positive constant $\bar{\eta} < 1$ such that $\bar{\eta} \leq (n/m) \leq 1/\bar{\eta}$.

Assumption 2.1 is a natural condition which ensures the sample sizes to be comparable. Such a requirement is commonly imposed for two sample estimation and inference (Bai and Saranadasa, 1996; Chen and Qin, 2010; Li and Chen, 2012; Chang et al., 2017; Zhang et al., 2020). In what follows, we write $N = nm/(n+m)$ for simplicity. The sub-Gaussian-type deviation inequality and the non-asymptotic Bahadur representation of the two-sample HL estimator $\widehat{\Theta}$ are established in the subsequent theorem.

Theorem 2.4. Assume that there exist positive constants $\bar{c}_0$ and $\bar{\kappa}_0$ such that $\inf_{|\delta| \leq \bar{c}_0} \mathcal{U}'(\Theta + \delta) \geq \bar{\kappa}_0 > 0$. Then for any $z > 0$, we have

$$\mathbb{P}(|\widehat{\Theta} - \Theta| > z) \leq 4 \exp\{-2(n \wedge m)\bar{\kappa}_0^2(z \wedge \bar{c}_0)^2\}.$$

Furthermore, assume that $\sup_{t \in \mathbb{R}} |\mathcal{U}'(t)| < \infty$ and there exist positive constants $\bar{c}_1$ and $\bar{\kappa}_1$ such that $\sup_{|\delta| \leq \bar{c}_1} |\mathcal{U}''(\Theta + \delta)| \leq \bar{\kappa}_1$. Then, under Assumption 2.1, for any $0 < z = o(N)$, we have

$$\mathbb{P}\left\{\left|\widehat{\Theta} - \Theta - \frac{1}{\mathcal{U}'(\Theta)}\left\{\frac{1}{n}\sum_{i=1}^{n} G(\xi_i) - \frac{1}{m}\sum_{j=1}^{m} F(\varepsilon_j)\right\}\right| > \frac{C_1(z \vee 1)}{N}\right\} \leq C_2 \exp(-z),$$

8

where $G(t) = \mathbb{P}(\varepsilon_1 \leq t)$ stands for the cumulative distribution function of ε_1 and $C_1, C_2 < \infty$ are positive constants depending only on $\bar{c}_0, \bar{\kappa}_0, \bar{c}_1, \bar{\kappa}_1, \bar{\eta}$ and $\|\mathcal{U}'\|_\infty$.

Theorem 2.4 presents the non-asymptotic approximation of the HL estimator $\widehat{\Theta}$, where the approximator also enjoys sub-Gaussian tails without posing any constraints on the moments of ξ_1 and ε_1. Equipped with this, we establish the Berry-Esseen bound and Cramér type moderate deviation of $\widehat{\Theta}$, respectively, in the following theorem. Before proceeding, we define the asymptotic variance of $\sqrt{N}(\widehat{\Theta} - \Theta)$ to be

$$\widetilde{\sigma}_\Theta^2 = \frac{1}{\{\mathcal{U}'(\Theta)\}^2} \left(\frac{n}{n+m} \operatorname{Var}\{F(\varepsilon_1)\} + \frac{m}{n+m} \operatorname{Var}\{G(\xi_1)\} \right).$$

Theorem 2.5. Under the conditions of Theorem 2.4, we have

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P}\left\{ \frac{\sqrt{N}(\widehat{\Theta} - \Theta)}{\widetilde{\sigma}_\Theta} \leq z \right\} - \Phi(z) \right| \leq \frac{C_1 \log N}{\sqrt{N}},$$

where $C_1 < \infty$ is a positive constant independent of N. Moreover, let $\{\delta_N\}_{N \geq 1}$ be a sequence of positive constants satisfying $\sqrt{N}\delta_N \to \infty$. Then, we further achieve

$$\left| \frac{\mathbb{P}\{\sqrt{N}(\widehat{\Theta} - \Theta)/\widetilde{\sigma}_\Theta \geq z\}}{1 - \Phi(z)} - 1 \right| \leq C_2 \left\{ \frac{1 + z^3}{\sqrt{N}} + (1+z)\delta_N + 2(z \vee 1)\sqrt{2\pi} \exp\left(\frac{z^2}{2} - \sqrt{N}\delta_N \right) \right\},$$

uniformly for $0 < z \leq o(\delta_N^{-1} \wedge N^{1/4}\sqrt{\delta_N})$, where $C_2 < \infty$ is a positive constant independent of z and N. In particular, when $\delta_N \asymp N^{-1/6}$, we have

$$\left| \frac{\mathbb{P}\{\sqrt{N}(\widehat{\Theta} - \Theta)/\widetilde{\sigma}_\Theta \geq z\}}{1 - \Phi(z)} - 1 \right| \to 0,$$

uniformly for $0 < z \leq o(N^{1/6})$.

One observes that the asymptotic distributions of $\widehat{\theta}$ and $\widehat{\Theta}$ involve many unknown quantities such as density functions and population parameters θ and Θ. In the following section, we utilize the bootstrap method to construct confidence intervals for the parameters of interest.

3 Bootstrap Calibration

In this section, we propose a weighted bootstrap method to construct confidence intervals for θ and Θ, rather than directly estimating those involved unknown terms in asymptotic variances using the brute force methods. The reason is that the direct estimation approach always necessitates the additional selection of tuning parameters and imposes moment conditions. Additionally, the bootstrap calibration performs admirably with finite samples, particularly when the sample size is modest. Therefore, in the sections that follow, we outline the bootstrap procedures for both the one- and two-sample problems.

3.1 Boostrap for One-sample Problem

Recall that the one-sample HL estimator is given by $\widehat{\theta} = \arg\min_{\nu \in \mathbb{R}} \sum_{i \neq j \in [n]} |X_i + X_j - 2\nu|$. Throughout this paper, we focus on the weighted bootstrap procedure in which the bootstrap estimate of $\widehat{\theta}$ is defined by minimizing the randomly perturbed objective function. More specifically, let $\omega_1, \ldots, \omega_n \in \mathbb{R}$ be i.i.d. non-negative random variables with $\mathbb{E}(\omega_1) = 1$ and $\text{Var}(\omega_1) = 1$. Then the weighted bootstrap estimate of $\widehat{\theta}$ is given by

$$\widehat{\theta}^\star = \arg\min_{\nu \in \mathbb{R}} \sum_{i \neq j \in [n]} \omega_i \omega_j |X_i + X_j - 2\nu|.$$

A natural candidate of the bootstrap weight above would be ω_i sampled from a 2·Bernoulli(0.5) distribution (the multiplication of 2 is to guarantee the previous normalization condition). In this case, the bootstrap estimator $\widehat{\theta}^\star$ has the simple closed-form expression as follows,

$$\widehat{\theta}^\star = \text{median}\{(X_i + X_j)/2 : i \neq j \in \mathcal{S}\}, \tag{3.1}$$

which is the same as the sub-sampled HL estimator computed based on the dataset $\{X_i : i \in \mathcal{S}\}$ where $\mathcal{S} = \{i \in [n] : \omega_i \neq 0\}$, and we concentrate on this type of bootstrap calibration procedure in what follows.

Let $B_n = \sum_{i \neq j \in [n]} \omega_i \omega_j$ denote the total number of *Walsh averages* in (3.1) and denote $V_{ij} = [\mathbb{I}\{\xi_i + \xi_j \leq 0\} - U_n(\theta)]/U'(\theta)$ for each $i \neq j \in [n]$. In the subsequent theorem, we establish the non-asymptotic Bahadur representation of the bootstrap estimator $\widehat{\theta}^\star$ and approximated distribution of bootstrap samples.

Theorem 3.1. Under the conditions of Theorem 2.1, for any $\omega, z > 0$ such that $(\omega \vee z) = o(n)$, with probability at least $1 - C_1 \exp(-\omega)$, we have

$$\mathbb{P}^\star\left\{|\widehat{\theta}^\star - \widehat{\theta}| > C_2\left(\frac{\omega + \log n}{n} + \sqrt{z/n}\right)\right\} \leq C_3 \exp(-z),$$

$$\mathbb{P}^\star\left\{\left|\widehat{\theta}^\star - \widehat{\theta} - \frac{1}{B_n}\sum_{i \neq j \in [n]}(\omega_i \omega_j - 1)V_{ij}\right| > \frac{C_4(z + \omega + \log n)}{n}\right\} \leq C_5 \exp(-z), \tag{3.2}$$

where $\mathbb{P}^\star(\cdot) = \mathbb{P}(\cdot | X_1, \ldots, X_n)$ stands for the conditional probability and C_1–C_7 are positive constants depending only on $c_0, \kappa_0, c_1, \kappa_1$ and $\|f\|_\infty$.

The non-asymptotic linear expansion in (3.2) enables us to derive the asymptotic normality of the bootstrap estimator $\widehat{\theta}^\star$. Combined with the Berry-Esseen bound in Theorem 2.2, we further establish a non-asymptotic upper bound on the Kolmogorov distance between the distribution functions of $\widehat{\theta} - \theta$ and $\widehat{\theta}^\star - \widehat{\theta}$. More specifically, with probability at least $1 - C \exp(-\omega)$, we have

$$\sup_{z \in \mathbb{R}}\left|\mathbb{P}^\star\left(|\widehat{\theta}^\star - \widehat{\theta}| \leq z\right) - \mathbb{P}\left(|\widehat{\theta} - \theta| \leq z\right)\right| \leq \frac{C'(\omega + \log n)}{\sqrt{n}}, \tag{3.3}$$

where C and C' are positive constants independent of n. Consequently, we are equipped to construct confidence interval for θ in a data-driven way. For any significance level $\alpha \in (0, 1)$, let

$$q^\star_{1-\alpha} = \inf\left\{z \in \mathbb{R} : \mathbb{P}^\star\left(|\widehat{\theta}^\star - \widehat{\theta}| \leq z\right) \geq 1 - \alpha\right\}.$$

Then the $(1 - \alpha) \times 100\%$ confidence interval for θ is given by $\mathbb{CI}(\theta, 1 - \alpha) = \{\widehat{\theta} - q^\star_{1-\alpha}, \widehat{\theta} + q^\star_{1-\alpha}\}$.

3.2 Bootstrap in Two-sample Problem

This section is devoted to constructing confidence intervals for Θ for the two-sample problem. Let $\omega_{n+1}, \ldots, \omega_{n+m} \in \mathbb{R}$ be i.i.d. 2·Bernoulli(0.5) random variables independent of $\{\omega_1, \ldots, \omega_n\}$. Following (3.1), the bootstrap estimator for $\widehat{\Theta}$ is defined by

$$\widehat{\Theta}^\star = \text{median}\{X_i - Y_j : i \in \mathcal{S}^X, j \in \mathcal{S}^Y\}, \tag{3.4}$$

where $\mathcal{S}^X = \{i \in [n] : \omega_i \neq 0\}$ and $\mathcal{S}^Y = \{j \in [m] : \omega_{j+n} \neq 0\}$. It is worth noting that $\widehat{\Theta}^\star$ is equivalent to the sub-sampled HL estimator based on the two datasets $\{X_i : i \in \mathcal{S}^X\}$ and $\{Y_j : j \in \mathcal{S}^Y\}$. With these necessary tools at hands, the non-asymptotic Bahadur representation of the bootstrap estimator $\widehat{\Theta}^\star$ and approximated distribution of bootstrap samples are developed in the following theorem.

Theorem 3.2. Under the conditions of Theorem 2.4, for any $\omega, z > 0$ such that $(\omega \vee z) = o(N)$, with probability at least $1 - C_1 \exp(-\omega)$, we have

$$\mathbb{P}^\star \left\{ |\widehat{\Theta}^\star - \widehat{\Theta}| > C_2 \left(\frac{\omega + \log N}{N} + \sqrt{z/N} \right) \right\} \leq C_3 \exp(-z),$$

$$\mathbb{P}^\star \left\{ \left| \widehat{\Theta}^\star - \widehat{\Theta} - \frac{1}{\mathcal{B}_n} \sum_{i=1}^n \sum_{j=1}^m (\omega_i \omega_{j+n} - 1) \mathcal{V}_{ij} \right| > \frac{C_4(z + \omega + \log N)}{N} \right\} \leq C_5 \exp(-z),$$

where $\mathbb{P}^\star(\cdot) = \mathbb{P}(\cdot | X_1, \ldots, X_n, Y_1, \ldots, Y_{n+m})$ stands for the conditional probability, $\mathcal{B}_n = \sum_{i=1}^n \sum_{j=1}^m \omega_i \omega_{j+n}$ is the total number of pairwise differences in (3.4) and $\mathcal{V}_{ij} = [\mathbb{I}\{\xi_i \leq \varepsilon_j\} - (nm)^{-1} \sum_{i=1}^n \sum_{j=1}^m \mathbb{I}\{\xi_i \leq \varepsilon_j\}]/\mathcal{U}'(\Theta)$ for each $i \in [n]$ and $j \in [m]$. In addition, C_1–C_5 are positive constants depending only on $c_0, \kappa_0, c_1, \kappa_1, \bar{\eta}$ and $\|\mathcal{U}'\|_\infty$.

We then obtain the Berry-Esseen bound and build confidence intervals for Θ based on the non-asymptotic expansion of the two-sample bootstrap estimator $\widehat{\Theta}^\star$ using similar arguments in §3.1.

4 Large-scale Multiple Testing

With the advancement of technology, large-scale, high-dimensional data have been extensively collected over the past two decades in a variety of fields such as medicine, biology, genetics, earth science, and finance. In large-scale regimes, there are inevitably heavy-tailed noises and it is crucial to develop robust statistical inference procedures. Yet, existing research that infers location shifts via Huber-type estimates calls for variable-dependent tuning parameters and moment limitations (Fan et al., 2019; Sun et al., 2020). This type of technique is hard to apply efficiently and faithfully to large-scale inferences due to the choices of tuning values. Moment constraints also exclude a large number of heavy-tailed distributions. To remedy the issues, this section focuses on extending the HL estimation to high dimensions and developing *tuning-free* and *moment-free* high-dimensional multiple testing procedures.

4.1 Large-Scale Testing for One-sample Problem

In this section, we investigate high-dimensional multiple testing using the HL estimator for one-sample data. Let $\boldsymbol{X}_i = \boldsymbol{\theta} + \boldsymbol{\xi}_i$, $i \in [n]$, be i.i.d. p-dimensional random vectors, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top$ is a p-dimensional vector of unknown parameters and $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n \in \mathbb{R}^p$ are i.i.d. random vectors. With building blocks presented in the previous section, we first proceed to constructing simultaneous confidence intervals for $\boldsymbol{\theta}$ using Gaussian approximation and bootstrap calibrations.

4.1.1 Gaussian Approximation

The primary goal of this section is to construct simultaneous confidence intervals for $\boldsymbol{\theta}$. To this end, we develop a Gaussian approximation for the maximum deviation $\max_{\ell \in [p]} |\widehat{\theta}_\ell - \theta_\ell|$ following the intuition of recently developed high dimensional distributional theory (Chernozhukov et al., 2017; Chernozhuokov et al., 2022). More specifically, let $\boldsymbol{Z} = (Z_1, \ldots, Z_p)^\top$ be a p-dimensional centered Gaussian random vector with

$$\mathrm{Cov}(Z_k, Z_\ell) = \frac{4 \, \mathrm{Cov}\{F_k(-\xi_{1k}), F_\ell(-\xi_{1\ell})\}}{U_k'(\theta_k) U_\ell'(\theta_\ell)}, \quad k, \ell \in [p], \tag{4.1}$$

where $U_\ell(t) = \mathbb{P}(X_{1\ell} + X_{2\ell} \le 2t)$ and $U_\ell'(t)$ stands for its derivative, and $F_\ell(t) = \mathbb{P}(\xi_{1\ell} \le t)$. We have the high dimensional Gaussian approximation in the following theorem.

Theorem 4.1. Assume that there exist positive constants c_0, κ_0, c_1 and κ_1 such that

$$\min_{\ell \in [p]} \inf_{|\delta| \le c_0} U_\ell'(\theta_\ell + \delta) \ge \kappa_0 \ \text{ and } \ \max_{\ell \in [p]} \sup_{|\delta| \le c_1} |U_\ell''(\theta_\ell + \delta)| \le \kappa_1.$$

Then, we have

$$\sup_{z > 0} \left| \mathbb{P}\left(\max_{\ell \in [p]} \sqrt{n} |\widehat{\theta}_\ell - \theta_\ell| \le z \right) - \mathbb{P}\left(\max_{\ell \in [p]} |Z_\ell| \le z \right) \right| \le \frac{C \log^{5/4}(pn)}{n^{1/4}}. \tag{4.2}$$

We consider testing the global null hypotheses

$$H_0 : \theta_\ell = 0 \text{ for all } \ell \in [p] \ \text{ versus } \ H_1 : \theta_\ell \ne 0 \text{ for some } \ell \in [p]. \tag{4.3}$$

Based on the marginal HL estimators $\{\widehat{\theta}_\ell\}_{\ell \in [p]}$, one shall reject the null hypothesis H_0 in (4.3) when $\max_{\ell \in [p]} |\widehat{\theta}_\ell|$ exceeds certain threshold that depends on the distribution of $\max_{\ell \in [p]} |Z_\ell|$. However, in light of (4.1), the distribution depends on the unknown distribution functions F_ℓ. Therefore, to approximate the distribution of $\max_{\ell \in [p]} |Z_\ell|$, we also propose to use bootstrap procedure. In specific, in one-sample regime, recall that $\mathcal{S} = \{i \in [n] : \omega_i \ne 0\}$. For each $\ell \in [p]$, define the bootstrap estimate of θ_ℓ as $\widehat{\theta}_\ell^\star = \mathrm{median}\{(X_{i\ell} + X_{j\ell})/2 : i \ne j \in \mathcal{S}\}$. It is worth mentioning that these bootstrap estimators $\{\widehat{\theta}_\ell^\star\}_{\ell \in [p]}$ can be efficiently computed in practice. For $\alpha \in (0,1)$, let $Q_{1-\alpha}^\star = \inf\{z \in \mathbb{R} : \mathbb{P}^\star(\max_{\ell \in [p]} |\widehat{\theta}_\ell^\star - \widehat{\theta}_\ell| \le z) \ge 1 - \alpha\}$ denote the $(1-\alpha)$th quantile of the bootstrap statistic $\max_{\ell \in [p]} |\widehat{\theta}_\ell^\star - \widehat{\theta}_\ell|$. With the help of bootstrap, we manage to estimate the quantiles of the approximated distribution efficiently and the corresponding results are presented in the following theorem.

Theorem 4.2. Under the conditions of Theorem 4.1, we have

$$\left| \mathbb{P}\left(\max_{\ell \in [p]} |\widehat{\theta}_\ell - \theta_\ell| > Q^\star_{1-\alpha} \right) - \alpha \right| \le \frac{C \log^{5/4}(pn)}{n^{1/4}}.$$

Theorem 4.2 reveals that the proposed bootstrap procedure can efficiently estimate the quantiles of the approximated distribution. This allows for the direct construction of simultaneous data-driven confidence intervals for $\boldsymbol{\theta}$. In addition, when the null-hypothesis of (4.3) is rejected, it is essential to conduct multiple testing to identify significant individuals and control false discovery proportion (FDP). We next address this problem in the following section.

4.1.2 Multiple Testing

The goal of this section is to conduct multiple testing to identify statistically significant individuals with controlled false discovery proportions. Specifically, we consider simultaneously testing the hypotheses

$$H_{0,\ell} : \theta_\ell = 0 \ \text{ versus } \ H_{1,\ell} : \theta_\ell \ne 0, \ \text{ for } \ \ell \in [p].$$

Let $\mathcal{H}_0 = \{\ell \in [p] : \theta_\ell = 0\}$ denote the set of true null hypotheses with cardinality $|\mathcal{H}_0|$. For each $\ell \in [p]$, let P_ℓ denote the p-value for testing the individual hypothesis $H_{0,\ell}$. For any prescribed threshold $t \in (0,1)$, we shall reject the null hypothesis $H_{0,\ell}$ whenever $P_\ell < t$. Then the false discovery proportion is defined by

$$\mathrm{FDP}(t) = V(t)/\max\{R(t), 1\}, \tag{4.4}$$

where $V(t) = \sum_{\ell \in \mathcal{H}_0} \mathbb{I}\{P_\ell \le t\}$ denotes the number of false discoveries and $R(t) = \sum_{\ell=1}^p \mathbb{I}\{P_\ell \le t\}$ is the number of total discoveries. Note that the denominator is observable but $V(t)$ is not. When $|\mathcal{H}_0|$ tends to infinity, $V(t) \approx t|\mathcal{H}_0| \le tp$. This can be used to give an upper bound of $\mathrm{FDP}(t)$. In many applications, $|\mathcal{H}_0| \approx p$, and Storey (2002) gives an estimator for $|\mathcal{H}_0|$ and incorporates it into $\mathrm{FDP}(t)$ estimator.

It is worth noting that the p-values $\{P_\ell\}_{\ell \in [p]}$ are computed by constructing inferential test statistics, with pivotal limiting distributions, based on the normal distribution calibration (Fan et al., 2007). As illustrated above, to construct a test statistic for each hypothesis with pivotal asymptotic distribution, the asymptotic variance of the HL estimator always depends on the unknown components and the traditional quantile-based approach is not scalable in the ultra-high dimensional scenario. To remedy this issue, we leverage bootstrap to proceed the analysis. Specifically, let $\{\omega_{i\ell} : i \in [n], \ell \in [p]\}$ be i.i.d. non-negative random variables generated in the same way with those in §3.1 and denote $\mathcal{S}_\ell = \{i \in [n] : \omega_{i\ell} \ne 0\}$ for each $\ell \in [p]$. Similar to (3.1), the bootstrap estimate of $\widehat{\theta}_\ell$ is defined by

$$\widehat{\theta}^\star_\ell = \mathrm{median}\{(X_{i\ell} + X_{j\ell})/2 : i \ne j \in \mathcal{S}_\ell\}.$$

Consequently, our p-values are derived as $P_\ell = \mathbb{P}(|\widehat{\theta}^\star_\ell - \widehat{\theta}_\ell| > |\widehat{\theta}_\ell| \, | \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ and let $P_{(1)} \le P_{(2)} \le \ldots \le P_{(p)}$ denote the ordered p-values. In order to choose t properly to control the FDP, we adopt

the distribution-free procedure proposed by Benjamini and Hochberg (1995). Specifically, for any significance level $\alpha \in (0, 1)$, the data-dependent threshold is $t_{BH} = P_{(\ell_{BH})}$, where ℓ_{BH} is given by

$$\ell_{BH} = \max\{\ell \in [p] : P_{(\ell)} \leq \alpha\ell/p\}.$$

Recall that an estimate of FDP(t) is $\widehat{\text{FDP}}(t) = pt/R(t)$, following the discussion after (4.4). Then a natural choice of threshold is

$$\widehat{t} = \sup\{t : \widehat{\text{FDP}}(t) \leq \alpha\} = \sup\{t : t \leq \alpha R(t)/p\} = t_{BH}.$$

This provides a simple explanation on the choice of t_{BH}.

In what follows, we assume that $|\mathcal{H}_0|/p \to \pi_0 \in (0, 1]$. For each $k, \ell \in [p]$, we define the correlation measure as $\rho_{k\ell} = \text{Corr}\{F_k(-\xi_{1k}), F_\ell(-\xi_{1\ell})\}$, and impose the assumption quantifying dependence between measurements below.

Assumption 4.1. There exists a positive constant $0 < \rho < 1$ such that $\max_{k \neq \ell} |\rho_{k\ell}| \leq \rho$ and

$$\max_{\ell \in [p]} \sum_{k=1}^{p} \mathbb{I}\left\{|\rho_{k\ell}| > \frac{1}{(\log p)^{2+\kappa}}\right\} = \mathcal{O}(p^\phi),$$

for some $\kappa > 0$ and $\phi \in (0, (1 - \rho)/(1 + \rho))$.

Assumption 4.1 requires that for every variable, the number of other variables, whose correlations with the given variable exceed certain threshold, does not grow too fast. It is worth noting that this is a commonly imposed condition in large-scale multiple testing problems (Liu and Shao, 2014). Based on this assumption, we next summarize the theoretical results in the following Theorem 4.3.

Theorem 4.3. Assume that $\log p = o(n^{1/5})$ and

$$\varpi_p := \left|\left\{\ell \in [p] : |\theta_\ell/\sigma_\ell| \geq \lambda_0 \sqrt{(2\log p)/n}\right\}\right| \to \infty, \tag{4.5}$$

for some $\lambda_0 > 2$, where $\sigma_\ell^2 = 4\text{Var}\{F_\ell(-\xi_{1\ell})\}/\{U_\ell'(\theta_\ell)\}^2$ for each $\ell \in [p]$. Then, under Assumption 4.1 and the conditions of Theorem 4.1, we have

$$\left|\frac{\text{FDP}(t_{BH})}{|\mathcal{H}_0|/p} - \alpha\right| \xrightarrow{\mathbb{P}} 0. \tag{4.6}$$

Theorem 4.3 develops theoretical guarantees for the consistency of FDP control procedure. To further support the derived results, we make the following several remarks.

Remark 4.1. Condition (4.5) is nearly optimal for controlling the false discovery proportion. More specifically, as shown in Proposition 2.1 in Liu and Shao (2014), if the number of alternative hypotheses is fixed instead of tending to infinity, the B-H approach fails to control the FDP at any level $0 < \beta < 1$ even if the true p-values for $\mathcal{H}_{0,\ell}$ are known. □

4.2 Large-Scale Two-Sample Tests

In this section, we study the large-scale two-sample testings. In specific, let $\boldsymbol{Y}_j = \boldsymbol{\theta}^\circ + \boldsymbol{\varepsilon}_j, j \in [m]$, be another sample of i.i.d. p-dimensional random vectors independent of $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}$, where $\boldsymbol{\theta}^\circ = (\theta_1^\circ, \ldots, \theta_p^\circ)^\top \in \mathbb{R}^p$. Let $\boldsymbol{\Theta} = \boldsymbol{\theta} - \boldsymbol{\theta}^\circ = (\Theta_1, \ldots, \Theta_p)^\top \in \mathbb{R}^p$ be the location shift parameter. Following (2.9), the HL estimator for Θ_ℓ is given by $\widehat{\Theta}_\ell = \text{median}\{X_{i\ell} - Y_{j\ell} : i \in [n], j \in [m]\}$. A detailed global test for whether $\theta_\ell = \theta_\ell^\circ$ for all $\ell \in [p]$ is given in §E.4.

We next conduct a simultaneously test on the hypotheses

$$H_{0,\ell} : \theta_\ell = \theta_\ell^\circ \text{ versus } H_{1,\ell} : \theta_\ell \neq \theta_\ell^\circ, \text{ for } \ell \in [p],$$

where $\mathcal{H}_0^\diamond = \{\ell \in [p] : \theta_\ell = \theta_\ell^\circ\}$ denote the set of true null hypotheses and $|\mathcal{H}_0^\diamond| = \sum_{\ell=1}^p \mathbb{I}\{\theta_\ell = \theta_\ell^\circ\}$ is its cardinality. Throughout this section, we assume that $|\mathcal{H}_0^\diamond|/p \to \pi_0 \in (0, 1]$. For each $\ell \in [p]$, let $P_\ell^\diamond$ denote the p-value for testing whether $\theta_\ell = \theta_\ell^\circ$. For any prescribed threshold $t \in (0, 1)$, we shall reject the null hypothesis $\theta_\ell = \theta_\ell^\circ$ whenever $P_\ell^\diamond < t$. The primary goal is to control the following false discovery proposition $\text{FDP}^\diamond(t)$,

$$\text{FDP}^\diamond(t) = V^\diamond(t)/\max\{R^\diamond(t), 1\},$$

by selecting a proper t, where $V^\diamond(t) = \sum_{\ell \in \mathcal{H}_0^\diamond} \mathbb{I}\{P_\ell^\diamond \leq t\}$ and $R^\diamond(t) = \sum_{\ell=1}^p \mathbb{I}\{P_\ell^\diamond \leq t\}$.

To approximate the unknown involved asymptotic distributions, we let $\{\omega_{i\ell} : i \in [n+m], \ell \in [p]\}$ be i.i.d. non-negative random variables generated in the same way with those in §3.2 and denote $\mathcal{S}_\ell^X = \{i \in [n] : \omega_{i\ell} \neq 0\}$ and $\mathcal{S}_\ell^Y = \{j \in [(n+1) : (m+n)] : \omega_{j\ell} \neq 0\}$ for each $\ell \in [p]$. Following (3.4), the bootstrap estimate for $\widehat{\Theta}_\ell$ is defined by

$$\widehat{\Theta}_\ell^\star = \text{median}\{X_{i\ell} - Y_{j\ell} : i \in \mathcal{S}_\ell^X, j \in \mathcal{S}_\ell^Y\}.$$

Then, for each $\ell \in [p]$, the p-value is given by $P_\ell^\diamond = \mathbb{P}(|\widehat{\Theta}_\ell^\star - \widehat{\Theta}_\ell| > |\widehat{\Theta}_\ell||\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$, and $P_{(1)}^\diamond \leq P_{(2)}^\diamond \leq \ldots \leq P_{(p)}^\diamond$ are the ordered p-values. Following Benjamini and Hochberg (1995), the data-dependent threshold $t_{BH}^\diamond = P_{(\ell_{BH}^\diamond)}^\diamond$, where $\ell_{BH}^\diamond$ is chosen as $\ell_{BH}^\diamond = \max\{\ell \in [p] : P_{(\ell)}^\diamond \leq \alpha\ell/p\}$.

With these necessary tools at hand, in the paragraph that follows, we present the required assumptions and the main theorem, respectively. In specific, Assumption 4.2 provides the formal condition that quantifies the level of dependence, while Theorem 4.4 presents the theoretical assurances for two-sample multiple testing.

Assumption 4.2. There exist positive constants $0 < \rho^\diamond < 1$ such that $\max_{k \neq \ell} |\rho_{k\ell}^\diamond| \leq \rho^\diamond$, where $\rho_{k\ell}^\diamond = \text{Corr}(\bar{G}_{nk} - \bar{F}_{mk}, \bar{G}_{n\ell} - \bar{F}_{m\ell})$, with $\bar{G}_{n\ell} = n^{-1} \sum_{i=1}^n G_\ell(\xi_{i\ell})$ and $\bar{F}_{m\ell} = m^{-1} \sum_{j=1}^m F_\ell(\varepsilon_{j\ell})$ for each $\ell \in [p]$. Moreover, for some $\kappa^\diamond > 0$ and $0 < \phi_\diamond < (1 - \rho^\diamond)/(1 + \rho^\diamond)$, we have

$$\max_{\ell \in [p]} \sum_{k=1}^p \mathbb{I}\left\{|\text{Corr}\{G_k(\xi_{1k}), G_\ell(\xi_{1\ell})\}| \vee |\text{Corr}\{F_k(\varepsilon_{1k}), F_\ell(\varepsilon_{1\ell})\}| > \frac{1}{(\log p)^{2+\kappa^\diamond}}\right\} = \mathcal{O}(p^{\phi_\diamond}).$$

Theorem 4.4. Assume that $\log p = o(N^{1/5})$ and

$$\varpi_p^\diamond := \left|\left\{\ell \in [p] : |(\theta_\ell - \theta_\ell^\circ)/\tilde{\sigma}_\ell| \geq \lambda_0 \sqrt{(2\log p)/N}\right\}\right| \to \infty, \tag{4.7}$$

for some $\lambda_0 > 2$, where $\widetilde{\sigma}_\ell = N[\mathrm{Var}\{F_\ell(\varepsilon_{1\ell})\}/m + \mathrm{Var}\{G_\ell(\xi_{1\ell})\}/n]/\{\mathcal{U}'_\ell(\Theta_\ell)\}^2$ for each $\ell \in [p]$. Then, under Assumption 4.2 and the conditions of Theorem E.1, we have

$$\left| \frac{\mathrm{FDP}^\diamond(t^\diamond_{BH})}{|\mathcal{H}^\diamond_0|/p} - \alpha \right| \xrightarrow{\mathbb{P}} 0. \tag{4.8}$$

5   Numerical Studies

In this section, we use simulation experiments to verify the theoretical findings in the paper. In specific, we validate the results on Gaussian approximation and FDP control in via experiments in §5.1 and §5.2, respectively.

5.1   Numerical Studies for Global Tests

We let $n = m = 300, p = 400$ and generate the variables $\{\mathbf{X}_i\}_{i=1}^n$ and $\{\mathbf{Y}_j\}_{j=1}^m$ following the setting in §4.1 and §4.2, respectively, where the involved random variables $\xi_i \in \mathbb{R}^p, \varepsilon_j \in \mathbb{R}^p, (i \in [n], j \in [m])$ follow several distributions described below. Cases 1-3 are devoted primarily to one-sample tests, whereas Cases 4-6 are for two-sample tests. Moreover, the notation $\mathrm{diff}(F)$ denotes the distribution that is generated by the differences between two independent random variables drawn from distribution F.

- Case 1: Scaled t_3 distribution with $\xi_{i,k} \sim 0.3 \cdot t_3, (i,k) \in [n] \times [p]$.

- Case 2: Mixture of Pareto distribution with shape parameter 2 and standard Gaussian distribution, namely, $\xi_{i,k} \sim \mathrm{diff}(0.2 \cdot \mathrm{Pareto}(2) + 0.8 \cdot N(0,1)), (i,k) \in [n] \times [p]$.

- Case 3: Mixture of Gaussian distributions, namely, $\xi_i \in \mathbb{R}^p \sim 0.2 \cdot N(0, 10\Sigma) + 0.8 \cdot N(0, \Sigma), i \in [n]$, where $\Sigma_{c,d} = 0.7^{|c-d|}, (c,d) \in [p] \times [p]$.

- Case 4: Gaussian distributions but with different covariance, namely, $\xi_i \sim N(0, 1.5\Sigma)$ and $\varepsilon_j \sim N(0, \Sigma), (i,j) \in [n] \times [m]$, where the covariance matrix is the same with that in case 3.

- Case 5: Mixture of Gaussian distributions with $\xi_i \in \mathbb{R}^p \sim 0.2 \cdot N(0, 10\Sigma) + 0.8 \cdot N(0, \Sigma), i \in [n]$, where the covariance matrix is the same with that in case 3, and mixture of Pareto distribution with shape parameter 2 and standard Gaussian distribution, namely, $\varepsilon_{j,k} \sim \mathrm{diff}(0.2 \cdot \mathrm{Pareto}(2) + 0.8 \cdot N(0,1)), (j,k) \in [m] \times [p]$.

- Case 6: Scaled t_3 distribution, where $\xi_{i,k} \sim 0.3 \cdot t_3, (i,k) \in [n] \times [p]$ and $\varepsilon_{j,k} \sim 0.3 \cdot t_3, (j,k) \in [n] \times [p]$.

As a first step, we validate the results of Gaussian approximation of HL estimator by conducting the tests in (4.3) and (E.3) with threshold $\alpha = 0.05$, respectively. We set the first 50 entries of $\boldsymbol{\theta} \in \mathbb{R}^p$ (or $\boldsymbol{\Theta} = \boldsymbol{\theta} - \boldsymbol{\theta}_0 \in \mathbb{R}^p$) given in §4.1 and §4.2 as μ, and the other entries as 0, where μ increases from 0 to 0.25. When $\mu = 0$, the null-hypothesis test holds; otherwise, the alternative holds. The size or power of the tests in (4.3) and (E.3) are computed via averaged outcomes from

500 replications of the methods in §4.1.1 and §E.4, respectively. In addition, for every replication, we conduct the weighted bootstrap method 300 times to compute the critical value of the test.

In addition, under the same experimental settings given above, we also compare the performance of HL estimator with the sample mean estimator ($\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{X}_i$ in one-sample case and $\sqrt{nm/(n+m)}(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i - \frac{1}{m} \sum_{j=1}^{m} \boldsymbol{Y}_j)$ in two-sample test). In this scenario, the critical values for tests (4.3) and (E.3) via sample mean estimators are computed via 300 bootstrap samples from joint Gaussian distribution $N(\boldsymbol{0}, \widehat{\Sigma})$, where $\widehat{\Sigma}$ is the sample covariance matrix ($\widehat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^{\top}$ in one-sample test and $\widehat{\Sigma} = \frac{n-1}{n+m-2} \sum_{i=1}^{n} (\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^{\top} + \frac{m-1}{m+n-2} \sum_{j=1}^{m} (\boldsymbol{Y}_j - \bar{\boldsymbol{Y}})(\boldsymbol{Y}_j - \bar{\boldsymbol{Y}})^{\top}$ in two-sample test). The results are then summarized in Table 2.

Estimator	μ	One-sample			Two-sample		
		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
HL	$\mu = 0$	0.054	0.048	0.052	0.040	0.046	0.045
	$\mu = 0.05$	0.880	0.062	0.060	0.080	0.084	0.282
	$\mu = 0.10$	1.000	0.480	0.464	0.164	0.140	1.000
	$\mu = 0.15$	1.000	0.982	0.942	0.322	0.320	1.000
	$\mu = 0.20$	1.000	1.000	1.000	0.728	0.688	1.000
	$\mu = 0.25$	1.000	1.000	1.000	0.944	1.000	1.000
Mean	$\mu = 0$	0.004	0.000	0.044	0.052	0.000	0.000
	$\mu = 0.05$	0.042	0.000	0.046	0.122	0.000	0.000
	$\mu = 0.10$	0.540	0.000	0.478	0.200	0.000	0.166
	$\mu = 0.15$	0.818	0.002	0.940	0.384	0.002	0.718
	$\mu = 0.20$	1.000	0.006	1.000	0.742	0.008	0.858
	$\mu = 0.25$	1.000	0.010	1.000	0.968	0.020	0.900

Table 2: Sizes and powers for testing the global null using Gaussian approximation via HL estimator and sample mean estimator, respectively.

We conclude from Table 2 that, in terms of the HL estimator, for all scenarios, the sizes of the test are roughly 0.05 when the null hypothesis is true ($\mu = 0$). Therefore, this validates the results of the Gaussian approximation. On the other hand, when the alternative holds, the power of the tests via HL estimator rises quickly to 1 as μ increases. This demonstrates the effectiveness of the HL test statistics. However, in terms of sample mean estimator, one observes that when the null holds, the size of the test is approximately 0 for most scenarios, whereas, when alternative holds, the power is much less than that of HL estimator. Thus, this further confirms the efficiency and robustness of HL estimator.

5.2 Numerical Studies for FDP Control

We validate the theoretical findings for FDP control under multiple regimes and also compare the performance of HL estimator with student's t-statistics given in Liu and Shao (2014). In specific, we maintain most of the settings mentioned in §5.1. The only difference is that we let the first 50

entries of $\boldsymbol{\theta}$ and $\boldsymbol{\Theta}$ be μ, where $\mu \in \{0.5, 0.3\}$, and the rest ones being 0. The target false discover proportion α is varied uniformly from 0.05 to 0.25 and the empirical FDP are computed via the procedures in §4.1.2 and §4.2, respectively. Meanwhile, for both of these two statistics, besides the FDP, the true positive proportion (TPP) of the test $\text{TPP}(t_{BH}) = \sum_{\ell \in \mathcal{H}_1} \mathbb{I}\{P_\ell \leq t_{BH}\}/|\mathcal{H}_1|$ is also computed. The numerical outcomes are summarized in Table 3 and Table 4 (corresponds to cases with $\mu = 0.5$ and $\mu = 0.3$, respectively).

Estimator	α	0.05	0.10	0.15	0.20	0.25
HL	Case 1	0.068 (1.000)	0.110 (1.000)	0.152 (1.000)	0.212 (1.000)	0.256 (1.000)
	Case 2	0.062 (1.000)	0.116 (1.000)	0.154 (1.000)	0.198 (1.000)	0.254 (1.000)
	Case 3	0.052 (1.000)	0.102 (1.000)	0.150 (1.000)	0.188 (1.000)	0.248 (1.000)
	Case 4	0.058 (1.000)	0.122 (1.000)	0.170 (1.000)	0.238 (1.000)	0.273 (1.000)
	Case 5	0.064 (1.000)	0.103 (1.000)	0.142 (1.000)	0.187 (1.000)	0.234 (1.000)
	Case 6	0.062 (1.000)	0.092 (1.000)	0.151 (1.000)	0.193 (1.000)	0.232 (1.000)
Student's t	Case 1	0.071 (1.000)	0.109 (1.000)	0.168 (1.000)	0.228 (1.000)	0.276 (1.000)
	Case 2	0.033 (0.976)	0.075 (0.976)	0.118 (0.990)	0.161 (0.996)	0.217 (1.000)
	Case 3	0.042 (1.000)	0.087 (1.000)	0.131 (1.000)	0.206 (1.000)	0.255 (1.000)
	Case 4	0.063 (1.000)	0.127 (1.000)	0.181 (1.000)	0.246 (1.000)	0.299 (1.000)
	Case 5	0.026 (0.962)	0.064 (0.962)	0.110 (0.988)	0.159 (0.994)	0.195 (1.000)
	Case 6	0.048 (1.000)	0.106 (1.000)	0.168 (1.000)	0.196 (1.000)	0.272 (1.000)

Table 3: Empirical FDP and TPP versus the nominal level α of the test via HL estimator and student's t-statistics when $\mu = 0.5$. The numbers outside and inside the brackets are averaged empirical false discovery proportions (FDP) and averaged true positive proportions (TPP), respectively, from 50 replications of the experiments in §4.1.2 and §4.2. For every replication, we conduct bootstrap 300 times for every dimension to compute the corresponding empirical p-values for HL estimator using the method in §4.1.2 and §4.2. In addition, the p-values for student's t-statistics are computed via quantiles of the standard Gaussian distribution.

Compared with the student's t-statistics, the empirical FDPs of HL estimator are closer to the theoretical thresholds in most cases and outcomes based on HL estimator have larger true positive proportion (TPP). This validates the theory of FDP control and confirms the benefit and effectiveness of using HL estimator when heavy-tailed error exists. In addition, we also further compare the performance of HL estimator with the student's t-statistics in both one-sample and two-sample tests when the noises follow t_1 distribution, where the HL estimator also performs much better. Interested readers are referred to §A.1 for more details.

6   Conclusion

In large-scale data analysis, conventional methods are ineffective since outliers and variables with heavy tails can easily corrupt the data. To resolve heavy-tailed contamination, existing tech-

Estimator	α	0.05	0.10	0.15	0.20	0.25
HL	Case 1	0.048 (1.000)	0.112 (1.000)	0.154 (1.000)	0.202 (1.000)	0.256 (1.000)
	Case 2	0.074 (1.000)	0.133 (1.000)	0.178 (1.000)	0.222 (1.000)	0.260 (1.000)
	Case 3	0.054 (1.000)	0.104 (1.000)	0.153 (1.000)	0.210 (1.000)	0.254 (1.000)
	Case 4	0.006 (0.984)	0.048 (0.998)	0.108 (1.000)	0.176 (1.000)	0.232 (1.000)
	Case 5	0.000 (0.992)	0.024 (0.992)	0.092 (0.992)	0.166 (0.992)	0.200 (0.992)
	Case 6	0.055 (1.000)	0.095 (1.000)	0.138 (1.000)	0.184 (1.000)	0.254 (1.000)
Student's t	Case 1	0.044 (1.000)	0.098 (1.000)	0.166 (1.000)	0.204 (1.000)	0.263 (1.000)
	Case 2	0.011 (0.910)	0.056 (0.910)	0.106 (0.930)	0.161 (0.968)	0.205 (0.984)
	Case 3	0.052 (1.000)	0.106 (1.000)	0.150 (1.000)	0.198 (1.000)	0.259 (1.000)
	Case 4	0.004 (1.000)	0.046 (1.000)	0.110 (1.000)	0.158 (1.000)	0.210 (1.000)
	Case 5	0.000 (0.870)	0.000 (0.952)	0.056 (0.984)	0.100 (0.990)	0.170 (1.000)
	Case 6	0.046 (1.000)	0.106 (1.000)	0.140 (1.000)	0.180 (1.000)	0.262 (1.000)

Table 4: Empirical FDP and TPP versus the nominal level α of the test via HL estimator and student's t-statistics when $\mu = 0.3$. The remain captions are the same with those in Table 3.

niques, such as Huberized mean, truncation, and median of means, always require additional tuning parameters and moment restrictions. Consequently, they cannot effectively be scaled to the high-dimensional applications with fidelity. Using the well-known Hodge-Lehmann estimator, the constraint on moment and tuning parameters can be removed. However, its non-asymptotic and large-scale properties have never been investigated. This paper fills this important gap by contributing a finite-sample analysis of the HL estimator, generalizing it to large-scale studies, and proposing tuning-free and moment-free high-dimensional testing methods.

There are various potential future directions that merit further studies. First, we permit mild measurement dependence while controlling the FDP in both the one- and two-sample regimes. However, in reality, high-dimensional data can exhibit strong dependency such as those collected in the field of economics, finance, genomics, and meteorology. To solve such strong dependence problems, factor-adjusted multiple testing via Huber-type estimation has been proposed (Fan et al., 2019). However, using Huber-type loss will result in extra tuning parameters and moment constraints. As a result, factor adjustments can be incorporated into the large-scale HL estimation and testing procedures in order to develop tuning-free procedures that can be adapted to a strong dependence scenario. Second, in terms of computation, we need to compute the median of $\mathcal{O}(n^2)$ pairs. However, in the one sample estimation regime, if we use the sub-sampling idea, one may compute the median of $n/2$ non-overlapping pair averages:

$$\theta^\tau = \text{median}\left(\frac{X_{\tau(2i-1)} + X_{\tau(2i)}}{2}, i \in [n/2]\right)$$

with a random permutation τ on $[n]$. By taking 5 or more permutations and the averages of these estimates, numerically, the averaged one approximates well the Hodge-Lehmann estimator and only

requires $\mathcal{O}(n)$ samples (Fan et al., 2020b). Therefore, it is also interesting to derive non-asymptotic analysis for this estimator based on sub-sampling.

References

Arcones, M. A. (1995). The asymptotic accuracy of the bootstrap of u-quantiles. *Ann. Statist.* 1802–1822.

Arcones, M. A. (1996). The bahadur-kiefer representation for u-quantiles. *Ann. Statist.*, **24** 1400–1422.

Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statist. Sinica*, **6** 311–329.

Bauer, D. F. (1972). Constructing confidence sets using rank statistics. *J. Amer. Statist. Assoc.*, **67** 687–690.

Belloni, A. and Chernozhukov, V. (2011). ℓ_1-penalized quantile regression in high-dimensional sparse models. *Ann. Statist.*, **39** 82–130.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, **57** 289–300.

Bentkus, V. K. and Dzindzalieta, D. (2015). A tight gaussian bound for weighted sums of rademacher random variables. *Bernoulli*, **21** 1231–1237.

Blanchard, G. and Roquain, É. (2009). Adaptive false discovery rate control under independence and dependence. *Journal of Machine Learning Research*, **10**.

Brownlees, C., Joly, E. and Lugosi, G. (2015). Empirical risk minimization for heavy-tailed losses. *The Annals of Statistics*, **43** 2507–2536.

Bühlmann, P. and Van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications.* Springer Science & Business Media.

Catoni, O. (2012). Challenging the empirical mean and empirical variance: a deviation study. In *Annales de l'IHP Probabilités et statistiques*, vol. 48.

Chang, J., Zheng, C., Zhou, W.-X. and Zhou, W. (2017). Simulation-based hypothesis testing of high dimensional means under covariance heterogeneity. *Biometrics*, **73** 1300–1310.

Chen, S. X. and Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.*, **38** 808–835.

Chen, X. and Zhou, W.-X. (2020). Robust inference via multiplier bootstrap. *Ann. Statist.*, **48** 1665–1691.

Chen, Y., Chi, Y., Fan, J., Ma, C. et al. (2021). Spectral methods for data science: A statistical perspective. *Foundations and Trends® in Machine Learning*, **14** 566–806.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.*, **45** 2309–2352.

Chernozhuokov, V., Chetverikov, D., Kato, K. and Koike, Y. (2022). Improved central limit theorem and bootstrap approximations in high dimensions. *Ann. Statist.*, **50** 2562–2586.

Chi, Z. (2007). On the performance of fdr control: constraints and a partial solution. *The Annals of Statistics*, **35** 1409–1431.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative finance*, **1** 223.

Eklund, A., Nichols, T. E. and Knutsson, H. (2016). Cluster failure: Why fmri inferences for spatial extent have inflated false-positive rates. *Proceedings of the national academy of sciences*, **113** 7900–7905.

Fan, J., Fan, Y. and Barut, E. (2014). Adaptive robust variable selection. *Ann. Statist.*, **42** 324–351.

Fan, J., Gu, Y. and Zhou, W.-X. (2022a). How do noise tails impact on deep relu networks? *arXiv preprint arXiv:2203.10418*.

Fan, J., Hall, P. and Yao, Q. (2007). To how many simultaneous hypothesis tests can normal, Student's *t* or bootstrap calibration be applied? *J. Amer. Statist. Assoc.*, **102** 1282–1288.

Fan, J., Ke, Y., Sun, Q. and Zhou, W.-X. (2019). FarmTest: factor-adjusted robust multiple testing with approximate false discovery control. *J. Amer. Statist. Assoc.*, **114** 1880–1893.

Fan, J., Li, Q. and Wang, Y. (2017). Estimation of high dimensional mean regression in the absence of symmetry and light tail assumptions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, **79** 247–265.

Fan, J., Li, R., Zhang, C.-H. and Zou, H. (2020a). *Statistical foundations of data science.* Chapman and Hall/CRC.

Fan, J., Lou, Z. and Yu, M. (2022b). Are latent factor regression and sparse regression adequate? *arXiv preprint arXiv:2203.01219*.

Fan, J., Ma, C. and Wang, K. (2020b). Comment on "a tuning-free robust and efficient approach to high-dimensional regression". *Journal of the American Statistical Association*, **115** 1720–1725.

Fan, J., Wang, W. and Zhu, Z. (2021). A shrinkage principle for heavy-tailed data: high-dimensional robust low-rank matrix recovery. *Ann. Statist.*, **49** 1239–1266.

Fan, J., Yang, Z. and Yu, M. (2022c). Understanding implicit regularization in over-parameterized single index model. *Journal of the American Statistical Association* 1–14.

Fan, J. and Yao, Q. (2017). *The elements of financial econometrics.* Cambridge University Press.

Fang, X., Luo, L. and Shao, Q.-M. (2020). A refined Cramér-type moderate deviation for sums of local statistics. *Bernoulli*, **26** 2319–2352.

Ferreira, J. A. and Zwinderman, A. H. (2006). On the Benjamini-Hochberg method. *Ann. Statist.*, **34** 1827–1849.

Finotello, F. and Di Camillo, B. (2015). Measuring differential gene expression with rna-seq: challenges and strategies for data analysis. *Briefings in functional genomics*, **14** 130–142.

Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control. *The annals of statistics*, **32** 1035–1061.

Giné, E., Latała, R. and Zinn, J. (2000). Exponential and moment inequalities for u-statistics. In *High Dimensional Probability II*. Springer, 13–38.

Goldstein, L., Minsker, S. and Wei, X. (2018). Structured signal recovery from non-linear and heavy-tailed measurements. *IEEE Transactions on Information Theory*, **64** 5513–5530.

Gupta, S., Ellis, S. E., Ashar, F. N., Moes, A., Bader, J. S., Zhan, J., West, A. B. and Arking, D. E. (2014). Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nature communications*, **5** 1–8.

Hastie, T., Tibshirani, R., Friedman, J. H. and Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*, vol. 2. Springer.

Hastie, T., Tibshirani, R. and Wainwright, M. (2015). Statistical learning with sparsity. *Monographs on statistics and applied probability*, **143** 143.

He, X. and Shao, Q.-M. (1996). A general bahadur representation of m-estimators and its application to linear regression with nonstochastic designs. *The Annals of Statistics*, **24** 2608–2630.

He, X. and Shao, Q.-M. (2000). On parameters of increasing dimensions. *Journal of multivariate analysis*, **73** 120–135.

Hodges, J. L., Jr. and Lehmann, E. L. (1963). Estimates of location based on rank tests. *Ann. Math. Statist.*, **34** 598–611.

Høyland, A. (1965). Robustness of the Hodges-Lehmann estimates for shift. *Ann. Math. Statist.*, **36** 174–197.

Hsu, D. and Sabato, S. (2016). Loss minimization and parameter estimation with heavy tails. *The Journal of Machine Learning Research*, **17** 543–582.

Huber, P. J. (1973). Robust regression: asymptotics, conjectures and monte carlo. *The annals of statistics* 799–821.

Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of economic perspectives*, **15** 143–156.

Lehmann, E. L. (1963). Nonparametric confidence intervals for a shift parameter. *Ann. Math. Statist.*, **34** 1507–1512.

Li, J. and Chen, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist.*, **40** 908–940.

Li, J. and Tibshirani, R. (2013). Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. *Stat. Methods Med. Res.*, **22** 519–536.

Li, J., Witten, D. M., Johnstone, I. M. and Tibshirani, R. (2012). Normalization, testing, and false discovery rate estimation for rna-sequencing data. *Biostatistics*, **13** 523–538.

Liu, W. and Shao, Q.-M. (2014). Phase transition and regularized bootstrap in large-scale t-tests with false discovery rate control. *Ann. Statist.*, **42** 2003–2025.

Loh, P.-L. (2017). Statistical consistency and asymptotic normality for high-dimensional robust m-estimators. *The Annals of Statistics*, **45** 866–896.

Mammen, E. (1989). Asymptotics with increasing dimension for robust regression with applications to the bootstrap. *The Annals of Statistics* 382–400.

Minsker, S. (2015). Geometric median and robust estimation in banach spaces. *Bernoulli*, **21** 2308–2335.

Minsker, S. (2018). Sub-gaussian estimators of the mean of a random matrix with heavy-tailed entries. *The Annals of Statistics*, **46** 2871–2903.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M. and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by rna sequencing. *Science*, **320** 1344–1349.

Nemirovsky, A. S. and Yudin, D. B. a. (1983). *Problem complexity and method efficiency in optimization.* Wiley-Interscience Series in Discrete Mathematics, John Wiley & Sons, Inc., New York.

Petrov, V. V. (1975). *Sums of independent random variables.* Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82, Springer-Verlag, New York-Heidelberg. Translated from the Russian by A. A. Brown.

Rosenkranz, G. K. (2010). A note on the hodges–lehmann estimator. *Pharmaceutical statistics*, **9** 162–167.

Shendure, J. and Ji, H. (2008). Next-generation dna sequencing. *Nature biotechnology*, **26** 1135–1145.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, **20** 147–162.

Storey, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **64** 479–498.

Storey, J. D. (2003). The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann. Statist.*, **31** 2013–2035.

Storey, J. D., Taylor, J. E. and Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **66** 187–205.

Sun, Q., Zhou, W.-X. and Fan, J. (2020). Adaptive huber regression. *Journal of the American Statistical Association*, **115** 254–265.

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, vol. 48. Cambridge University Press.

Wang, B. and Fan, J. (2022). Robust matrix completion with heavy-tailed noise. *arXiv preprint arXiv:2206.04276*.

Wang, L., Peng, B., Bradic, J., Li, R. and Wu, Y. (2020). A tuning-free robust and efficient approach to high-dimensional regression. *Journal of the American Statistical Association*, **115** 1700–1714.

Wang, L., Peng, B. and Li, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *Journal of the American Statistical Association*, **110** 1658–1669.

Wang, Z., Gerstein, M. and Snyder, M. (2009). Rna-seq: a revolutionary tool for transcriptomics. *Nature reviews genetics*, **10** 57–63.

Xia, Y., Cai, T. T. and Li, H. (2018). Joint testing and false discovery rate control in high-dimensional multivariate regression. *Biometrika*, **105** 249–269.

Yang, Z., Balasubramanian, K. and Liu, H. (2017). High-dimensional non-gaussian single index models via thresholded score function estimation. In *International conference on machine learning*. PMLR.

Yohai, V. J. and Maronna, R. A. (1979). Asymptotic behavior of m-estimators for the linear model. *The Annals of Statistics* 258–268.

Zhang, J.-T., Guo, J., Zhou, B. and Cheng, M.-Y. (2020). A simple two-sample test in high dimensions based on L^2-norm. *J. Amer. Statist. Assoc.*, **115** 1011–1027.

Zheng, Q., Peng, L. and He, X. (2015). Globally adaptive quantile regression with ultra-high dimensional data. *Annals of statistics*, **43** 2225.

Zhou, W.-X., Bose, K., Fan, J. and Liu, H. (2018). A new perspective on robust M-estimation: finite sample theory and applications to dependence-adjusted multiple testing. *Ann. Statist.*, **46** 1904–1931.

A   Appendix

In the following sections, we present additional simulation results as well as the proofs of all the theoretical results in the main paper.

A.1   Additional Simulation

In this section, we conduct additional simulations for FDP control using HL estimator and student's t-statistics, respectively. We keep all settings as in §5.2, except that the noises are generated from the following case 7 and case 8, where we change the distribution in case 1 and case 6 from t_3 to t_1. The numerical outcomes are summarized in Table 5 and Table 6.

- Case 7: Scaled t_1 distribution with $\xi_{i,k} \sim 0.3 \cdot t_1, (i,k) \in [n] \times [p]$.

- Case 8: Scaled t_1 distribution, where $\xi_{i,k} \sim 0.3 \cdot t_1, (i,k) \in [n] \times [p]$ and $\varepsilon_{j,k} \sim 0.3 \cdot t_1, (j,k) \in [n] \times [p]$.

Estimator	α	0.05	0.10	0.15	0.20	0.25
HL	Case 7	0.036 (1.000)	0.082 (1.000)	0.132 (1.000)	0.180 (1.000)	0.222 (1.000)
	Case 8	0.069 (1.000)	0.105 (1.000)	0.168 (1.000)	0.225 (1.000)	0.272 (1.000)
Student's t	Case 7	0.000 (0.282)	0.000 (0.284)	0.000 (0.320)	0.000 (0.320)	0.000 (0.320)
	Case 8	0.000 (0.274)	0.000 (0.274)	0.000 (0.274)	0.000 (0.322)	0.000 (0.322)

Table 5: Empirical FDP and TPP versus the nominal level α of the test via HL estimator and student's t-statistics when $\mu = 0.5$. The remain captions are the same with those in Table 3.

Estimator	α	0.05	0.10	0.15	0.20	0.25
HL	Case 7	0.036 (1.000)	0.82 (1.000)	0.122 (1.000)	0.179 (1.000)	0.224 (1.000)
	Case 8	0.058 (1.000)	0.105 (1.000)	0.157 (1.000)	0.214 (1.000)	0.252 (1.000)
Student's t	Case 7	0.000 (0.282)	0.000 (0.282)	0.000 (0.282)	0.000 (0.284)	0.000 (0.284)
	Case 8	0.000 (0.134)	0.000 (0.134)	0.000 (0.134)	0.000 (0.134)	0.000 (0.134)

Table 6: Empirical FDP and TPP versus the nominal level α of the test via HL estimator and student's t-statistics when $\mu = 0.3$. The other captions are the same with those in Table 3.

We conclude from Table 5 and Table 6, the HL estimator outperforms the student's t-statistics when the noise follows t_1 distribution in terms of the FDP and TPP. Therefore, this further confirms the robustness of HL estimator.

B Lemmas

Lemma B.1. Let $S_n = \sum_{i=1}^n Y_i$, where $Y_1, \ldots, Y_n \in \mathbb{R}$ are independent random variables such that $a_i \leq Y_i \leq b_i$ for each $i \in [n]$. Then, for any $z > 0$, we have

$$\mathbb{P}(|S_n - \mathbb{E}S_n| > z) \leq 2 \exp\left\{ -\frac{2z^2}{\sum_{i=1}^n (b_i - a_i)^2} \right\}.$$

Lemma B.2. Let $Y_1, \ldots, Y_n \in \mathbb{R}$ be i.i.d. random variables and $H_n = \{n(n-1)\}^{-1} \sum_{i \neq j \in [n]} h(X_i, X_j)$, where $h(x, y)$ is a symmetric function with $a \leq h(x, y) \leq b$ for some $a \leq b \in \mathbb{R}$. Then, for any $z > 0$, we have

$$\mathbb{P}(H_n - \mathbb{E}H_n > z) \leq \exp\left\{ -\frac{nz^2}{(b-a)^2} \right\}.$$

Let $V_1, \ldots, V_m \in \mathbb{R}$ be another sample of i.i.d. random variables independent of $\{Y_1, \ldots, Y_n\}$ and let $\mathcal{H}_{n,m} = (nm)^{-1} \sum_{i=1}^n \sum_{j=1}^m \bar{h}(Y_i, V_j)$, where $\bar{h}(x, y)$ is bounded such that $a \leq \bar{h}(x, y) \leq b$ for some $a \leq b \in \mathbb{R}$. Then, for any $z > 0$, we have

$$\mathbb{P}(\mathcal{H}_{n,m} - \mathbb{E}\mathcal{H}_{n,m} > z) \leq \exp\left\{ -\frac{2(n \wedge m)z^2}{(b-a)^2} \right\}.$$

C Proof of Theoretical Results in §2

C.1 Proof of Theorem 2.1

For simplicity of notation, we write $W_n(t) = U_n(\theta + t) - U(\theta + t)$ for $t \in \mathbb{R}$. Define $R(X_i, X_j, t) = \mathbb{I}\{\theta < (X_i + X_j)/2 \leq \theta + t\}$ and

$$\bar{R}(X_i, X_j, t) = R(X_i, X_j, t) - \mathbb{E}\{R(X_i, X_j, t)|X_i\} - \mathbb{E}\{R(X_i, X_j, t)|X_j\} + \mathbb{E}\{R(X_i, X_j, t)\}.$$

With this notation, we have $W_n(t) - W_n(0) = \bar{R}_n^\diamond(t) + \bar{R}_n^\circ(t)$, where

$$\bar{R}_n^\diamond(t) = \frac{2}{n} \sum_{i=1}^n [\mathbb{E}\{R(X_i, X_j, t)|X_i\} - \mathbb{E}\{R(X_i, X_j, t)\}] \text{ and } \bar{R}_n^\circ(t) = \frac{1}{n(n-1)} \sum_{i \neq j \in [n]} \bar{R}(X_i, X_j, t).$$

Lemma C.1. For any $0 < t \leq 1/(2\|f\|_\infty)$, there exists a universal positive constant C such that

$$\mathbb{P}(|\bar{R}_n^\circ(t)| > z) \leq C \exp\left[-\frac{1}{C} \min\left\{ \frac{n^2 z^2}{t\|f\|_\infty}, \left(\frac{n^3 z^2}{t\|f\|_\infty}\right)^{1/3}, \frac{nz}{\sqrt{t\|f\|_\infty}}, nz^{1/2} \right\} \right]. \tag{C.1}$$

Proof of Lemma C.1. Observe that $\bar{R}(X_i, X_j, t)$ is a bounded canonical kernel of X_i and X_j for any $t \in \mathbb{R}$. Hence it is straightforward to derive (C.1) by Theorem 3.3 in Giné et al. (2000). □

Lemma C.2. For any $0 < \Lambda \leq 1/(2\|f\|_\infty)$, with probability at least $1 - C\exp(-z)$, we have

$$\sup_{|t| \leq \Lambda} |W_n(t) - W_n(0)| \leq C_1 \left\{ \left(\frac{K+z}{n}\right)^2 + \Lambda\|f\|_\infty \left(\frac{z + C_2}{n}\right)^{1/2} + \frac{K+z}{n}(\Lambda\|f\|_\infty)^{1/2} \right\}, \tag{C.2}$$

where $K = \min\{k \in \mathbb{N} : 2^k \geq n\}$.

27

Proof of Lemma C.2. For any $h \in [2^K]$ and $t \in (\Lambda(h-1)2^{-K}, \Lambda h 2^{-K}]$, we have

$$W_n(t) - W_n(0) \leq U_n\left(\theta + \frac{\Lambda h}{2^K}\right) - U\left(\theta + \frac{\Lambda(h-1)}{2^K}\right) - W_n(0)$$

$$\leq W_n\left(\frac{\Lambda h}{2^K}\right) - W_n(0) + U\left(\theta + \frac{\Lambda h}{2^K}\right) - U\left(\theta + \frac{\Lambda(h-1)}{2^K}\right)$$

$$\leq W_n\left(\frac{\Lambda h}{2^K}\right) - W_n(0) + \frac{\Lambda\|U'\|_\infty}{2^K}.$$

Similarly, we have

$$W_n(t) - W_n(0) \geq W_n\left(\frac{\Lambda(h-1)}{2^K}\right) - W_n(0) - \frac{\Lambda\|U'\|_\infty}{2^K}.$$

Consequently, we obtain

$$\sup_{0 < t \leq \Lambda} |W_n(t) - W_n(0)| \leq \max_{h \in [2^K]} \left| W_n\left(\frac{\Lambda h}{2^K}\right) - W_n(0) \right| + \frac{\Lambda\|U'\|_\infty}{2^K}$$

$$\leq \underbrace{\max_{h \in [2^K]} \left| \bar{R}_n^\circ\left(\frac{\Lambda h}{2^K}\right) \right|}_{\Gamma_K^\circ} + \underbrace{\max_{h \in [2^K]} \left| \bar{R}_n^\diamond\left(\frac{\Lambda h}{2^K}\right) \right|}_{\Gamma_K^\diamond} + \frac{\Lambda\|U'\|_\infty}{2^K}.$$

For each $h \in [2^K]$, by (C.1), with probability at least $1 - C\exp(-z)$,

$$\Gamma_K^\circ \leq C\left\{ \left(\frac{K+z}{n}\right)^2 + \frac{K+z}{n}\sqrt{\Lambda\|U'\|_\infty} \right\},$$

where $C > 0$ is a universal constant.

$$\Gamma_K^\diamond \leq \sum_{k=1}^K \max_{h \in [2^k]} \left| \bar{R}_n^\diamond\left(\frac{\Lambda h}{2^k}\right) - \bar{R}_n^\diamond\left(\frac{\Lambda(h-1)}{2^k}\right) \right| + \max_{h \in [2]} \left| \bar{R}_n^\diamond\left(\frac{\Lambda h}{2}\right) \right| =: \Gamma_{K,1}^\diamond + \Gamma_{K,2}^\diamond.$$

For each $k \in [K]$, by Lemma B.1, with probability at least $1 - \exp(-z)$, we have

$$\max_{h \in [2^k]} \left| \bar{R}_n^\diamond\left(\frac{\Lambda h}{2^k}\right) - \bar{R}_n^\diamond\left(\frac{\Lambda(h-1)}{2^k}\right) \right| \leq C\Lambda\|f\|_\infty 2^{-k}\sqrt{\frac{k+z}{n}}.$$

Hence with the same probability, we have

$$\Gamma_{K,1}^\diamond \leq \sum_{k=1}^K \left(C\Lambda\|f\|_\infty 2^{-k}\sqrt{\frac{2k+z}{n}} \right) \leq C_1\Lambda\|f\|_\infty\sqrt{\frac{z+C_2}{n}}.$$

Similarly, it follows that

$$\mathbb{P}\left(\Gamma_{K,2}^\diamond > C_3\Lambda\|f\|_\infty\sqrt{(2z)/n} \right) \leq C_4\exp(-z).$$

Putting all these pieces together, we obtain (C.2). $\qquad\square$

Proof of Theorem 2.1. For any $z \geq 0$, since $U(t)$ is a cumulative distribution function, we have

$$\frac{1}{2} > U_n(\theta + z) = U(\theta + z) - \{U(\theta + z) - U_n(\theta + z)\}$$

$$\geq U(\theta) + \int_0^{z \wedge c_0} U'(\theta + \nu)d\nu - \{U(\theta + z) - U_n(\theta + z)\}$$

$$\geq \frac{1}{2} + \kappa_0(z \wedge c_0) - \{U(\theta + z) - U_n(\theta + z)\}.$$

By the definition of $\widehat{\theta}$ in (C.4), it follows that $\mathbb{P}(\widehat{\theta} > \theta + z) \leq \mathbb{P}\{1/2 > U_n(\theta + z)\}$. Therefore

$$\mathbb{P}(\widehat{\theta} > \theta + z) \leq \mathbb{P}\{U(\theta + z) - U_n(\theta + z) > \kappa_0(z \wedge c_0)\} \leq \exp\{-n\kappa_0^2(z \wedge c_0)^2\},$$

where the last inequality follows from the Hoeffding inequality in Lemma B.2.

Recall that $W_n(t) = U_n(\theta + t) - U(\theta + t)$ for $t \in \mathbb{R}$. Taking $\Lambda = \sqrt{z/(n\kappa_0^2)}$ yields $\mathbb{P}(|\widehat{\theta} - \theta| > \Lambda) \leq 2\exp(-z)$. By Lemma C.2, with probability at least $1 - C\exp(-z)$, we have

$$\sup_{|\delta| \leq \Lambda} |W_n(\delta) - W_n(0)| \leq \frac{C_1 \|f\|_\infty (z + c)}{n\kappa_0}.$$

We then obtain

$$\left| \widehat{\theta} - \theta - \frac{U_n(\widehat{\theta}) - U_n(\theta)}{U'(\theta)} \right| \leq \frac{|W_n(\widehat{\theta} - \theta) - W_n(0)|}{\kappa_0} + \frac{\kappa_1|\widehat{\theta} - \theta|^2}{\kappa_0} \leq \frac{C_1\|f\|_\infty(z+c)}{n\kappa_0^2} + \frac{\kappa_1 z}{n\kappa_0^3}.$$

Finally, denote

$$\Delta_\theta := \frac{1/2 - U_n(\theta)}{U'(\theta)} - \frac{2}{nU'(\theta)} \sum_{i=1}^n \left\{ \frac{1}{2} - F(-\xi_i) \right\}$$

$$= \frac{1}{n(n-1)U'(\theta)} \sum_{i \neq j \in [n]} \left(F(-\xi_i) + F(-\xi_j) - \mathbb{I}\{\xi_i + \xi_j \leq 0\} - \frac{1}{2} \right).$$

Observe that $|F(-\xi_i) + F(-\xi_j) - \mathbb{I}\{\xi_i + \xi_j \leq 0\} - 1/2| \leq 2$ uniformly for $i \neq j \in [n]$. Then (2.4) follows from Theorem 3.3 in Giné et al. (2000). □

C.2   Proof of Theorem 2.2

Proof of Theorem 2.2. For simplicity of notation, denote

$$T_n = \frac{\sqrt{n}(\widehat{\theta} - \theta)}{\sigma_\theta} \quad \text{and} \quad T_n^\sharp = \frac{1}{\sqrt{n \, \text{Var}\{F(-\xi_1)\}}} \sum_{i=1}^n \left\{ \frac{1}{2} - F(-\xi_i) \right\}. \tag{C.3}$$

Note that $\max_{i \in [n]} |1/2 - F(-\xi_i)| \leq 1/2$. Hence $\sup_{z \in \mathbb{R}} |\mathbb{P}(T_n^\sharp \leq z) - \Phi(z)| \leq Cn^{-1/2}$. Let $\mathcal{C} < \infty$ be a sufficiently large positive constant. Then it follows from (2.4) that

$$\mathbb{P}\left(|T_n - T_n^\sharp| > \frac{\mathcal{C}\log n}{\sqrt{n}} \right) \leq \frac{C_1}{\sqrt{n}}.$$

Consequently, we obtain

$$\sup_{z \in \mathbb{R}} |\mathbb{P}(T_n \leq z) - \Phi(z)| \leq \sup_{z \in \mathbb{R}} |\mathbb{P}(T_n^\sharp \leq z) - \Phi(z)| + \sup_{z \in \mathbb{R}} |\mathbb{P}(T_n \leq z) - \mathbb{P}(T_n^\sharp \leq z)| \leq \mathcal{C}(\log n)/\sqrt{n}.$$

□

C.3 Proof of Theorem 2.3

Proof of Theorem 2.3. Recall the definitions of T_n and $T_n^\sharp$ in (C.3). Since $\max_{i \in [n]} |1/2 - F(-\xi_i)| \le 1/2$, it follows that

$$\left| \frac{\mathbb{P}(T_n^\sharp > z)}{1 - \Phi(z)} - 1 \right| \le \frac{C(1 + z^3)}{\sqrt{n}} \ \text{ for } \ 0 \le z \le C_0 n^{1/6},$$

where $C_0 > 0$ is any fixed constant and C is a positive constant depending only on C_0. Denote $\bar\Phi(\cdot) = 1 - \Phi(\cdot)$. By (2.4), we have

$$\begin{aligned}
\mathbb{P}(T_n > z) &\le \mathbb{P}(T_n^\sharp > z - \delta_n) + C_1 \exp(-\sqrt{n}\delta_n) \\
&\le \left\{ 1 + \frac{C(1 + z^3)}{\sqrt{n}} \right\} \bar\Phi(z - \delta_n) + C_1 \exp(-\sqrt{n}\delta_n) \\
&\le \left\{ 1 + \frac{C(1 + z^3)}{\sqrt{n}} \right\} \{1 + (1 + z)\delta_n \exp(z\delta_n)\} \bar\Phi(z) + C_1 \exp(-\sqrt{n}\delta_n).
\end{aligned}$$

By Mill's inequality, for any $z > 0$,

$$z\sqrt{2\pi} \exp\left(\frac{z^2}{2} \right) \le \frac{1}{\bar\Phi(z)} \le 2(z \vee 1)\sqrt{2\pi} \exp\left(\frac{z^2}{2} \right).$$

Consequently, it follows that

$$\frac{\mathbb{P}(T_n > z)}{1 - \Phi(z)} - 1 \le \frac{C(1 + z^3)}{\sqrt{n}} + C_2(1 + z)\delta_n + 2C_1(z \vee 1)\sqrt{2\pi} \exp\left(\frac{z^2}{2} - \sqrt{n}\delta_n \right).$$

Similarly, we have

$$1 - \frac{\mathbb{P}(T_n > z)}{1 - \Phi(z)} \le \frac{C(1 + z^3)}{\sqrt{n}} + C_2(1 + z)\delta_n + C_1 z\sqrt{2\pi} \exp\left(\frac{z^2}{2} - \sqrt{n}\delta_n \right).$$

Putting all these pieces together, we obtain (2.7). $\qquad\square$

C.4 Proof of Theorem 2.4

Define the two-sample U-process $\mathcal{U}_{n,m}(t) = (nm)^{-1} \sum_{i=1}^n \sum_{j=1}^m \mathbb{I}\{X_i - Y_j \le t\}$. Then the HL estimator $\widehat\Theta$ in (2.9) can be equivalently expressed as the sample median of the U-process $\mathcal{U}_{n,m}(t)$, namely,

$$\widehat\Theta = \inf\{t \in \mathbb{R} : \mathcal{U}_{n,m}(t) \ge 1/2\}. \tag{C.4}$$

The remaining steps for proving Theorem 2.4 can be derived by following similar steps as in the proof of Lemma C.2 and proof of Theorem 2.1.

C.5 Proof of Theorem 2.5

For simplicity of notation, we write $\mathcal{T}_N = \sqrt{N}(\widehat{\Theta} - \Theta)/\widetilde{\Theta}$ and

$$\mathcal{T}_N^\sharp = \frac{\bar{G}_n - \bar{F}_m}{\sqrt{\mathrm{Var}(\bar{G}_n - \bar{F}_m)}},$$

where $\bar{G}_n = n^{-1}\sum_{i=1}^n G(\xi_i)$ and $\bar{F}_m = m^{-1}\sum_{j=1}^m F(\varepsilon_j)$. Then, similar to proof of Theorem 2.2, it follows that

$$\sup_{z\in\mathbb{R}}|\mathbb{P}(\mathcal{T}_N \le z) - \Phi(z)| \le \sup_{z\in\mathbb{R}}|\mathbb{P}(\mathcal{T}_N \le z) - \mathbb{P}(\mathcal{T}_N^\sharp \le z)| + \sup_{z\in\mathbb{R}}|\mathbb{P}(\mathcal{T}_N^\sharp \le z) - \Phi(z)| \lesssim \frac{\log N}{\sqrt{N}}.$$

The proof of the Cramér-type moderate deviation for $\widehat{\Theta}$ can be derived by following similar proof procedures of Theorem 2.3. Therefore, we decide to omit the details.

D   Proof of Results in §3

Define

$$\widehat{\sigma}_\theta^2 = \frac{4}{n\{U'(\theta)\}^2}\sum_{i=1}^n\left(\frac{1}{n-1}\sum_{j\neq i}\mathbb{I}\{\xi_i + \xi_j \le 0\} - U_n(\theta)\right)^2.$$

Lemma D.1. Under the conditions of Theorem 2.1, we have

$$\mathbb{P}\left(\left|\frac{\widehat{\sigma}_\theta}{\sigma_\theta} - 1\right| > z\right) \lesssim \exp(-C_U n) + \exp(-C_U nz) + \exp(-C_U nz^2), \tag{D.1}$$

where C_U is a positive constant depending only on U.

Proof of Lemma D.1. For each $i \in [n]$, denote

$$W_i = \frac{1}{n-1}\sum_{j\neq i}\mathbb{I}\{\xi_i + \xi_j \le 0\} - F(-\xi_i).$$

Recall that $U_n(\theta) = \{n(n-1)\}^{-1}\sum_{i\neq j\in[n]}\mathbb{I}\{\xi_i + \xi_j \le 0\}$. Hence

$$\widehat{\sigma}_\theta^2 - \underbrace{\frac{4}{n\{U'(\theta)\}^2}\sum_{i=1}^n\left\{F(-\xi_i) - \frac{1}{2}\right\}^2}_{\widetilde{\sigma}_\theta^2} = \frac{4}{n\{U'(\theta)\}^2}\sum_{i=1}^n W_i^2 - \frac{4|U_n(\theta) - 1/2|^2}{\{U'(\theta)\}^2}$$

$$+ \frac{8}{n\{U'(\theta)\}^2}\sum_{i=1}^n W_i\left\{F(-\xi_i) - \frac{1}{2}\right\} =: \Delta_{\sigma,1} - \Delta_{\sigma,2} + \Delta_{\sigma,3}.$$

Since $\mathbb{E}(\widetilde{\sigma}_\theta^2) = \sigma_\theta^2$ and $\max_{i\in[n]}|F(-\xi_i) - 1/2| \le 1/2$, it follows from Lemma B.1 that

$$\mathbb{P}(|\widetilde{\sigma}_\theta^2 - \sigma_\theta^2| > z) \le 2\exp(-C_U nz^2),$$

31

where C_U is a positive constant depending only on the density function U'. Note that $\{\mathbb{I}\{\xi_i + \xi_j \leq 0\} - F(-\xi_i)\}_{j \neq i}$ is a sequence of independent centered random variables conditional on ξ_i for each $i \in [n]$. Hence, by Hoeffding's inequality, we have $\mathbb{P}(\max_{i \in [n]} W_i^2 > z) \leq 2n \exp(-Cnz)$. Therefore

$$\mathbb{P}(\Delta_{\sigma,1} > z) \leq \mathbb{P}\left(\frac{4}{\{U'(\theta)\}^2} \max_{i \in [n]} W_i^2 > z\right) \leq 2n \exp(-C_U nz).$$

Consequently, by the Cauchy-Schwarz inequality, we have $|\Delta_{\sigma,3}|^2 \leq 4\widetilde{\sigma}_\theta^2 \Delta_{\sigma,1}$ and

$$\mathbb{P}(|\Delta_{\sigma,4}| > z) \leq 2\exp(-C_U n) + 2n \exp(-C_U nz^2).$$

By Lemma B.2, we have $\mathbb{P}(|U_n(\theta) - 1/2| > z) \leq 2\exp(-nz^2)$ and $\mathbb{P}(\Delta_{\sigma,3} > z) \leq 2\exp(-C_U nz)$. Putting all these pieces together, we obtain (D.1). $\qquad\square$

D.1 Proof of Theorem 3.1

Define the bootstrap U-process

$$U_n^\star(t) = \frac{1}{B_n} \sum_{i \neq j \in [n]} \omega_i \omega_j \mathbb{I}\left\{\frac{X_i + X_j}{2} \leq t\right\}, \quad \text{where } B_n = \sum_{i \neq j \in [n]} \omega_i \omega_j.$$

Define $W_n^\star(t) = U_n^\star(\theta + t) - U(\theta + t)$ and $W_n^\diamond(t) = U_n^\star(\theta + t) - U_n(\theta + t)$ for $t \in \mathbb{R}$. We first introduce some notation. Let $K = \min\{k \in \mathbb{N} : 2^k \geq n\}$. For each $i \neq j \in [n]$ and $h \in [2^K]$, denote

$$\mathcal{R}\left(X_i, X_j, \frac{\Lambda h}{2^K}\right) = R\left(X_i, X_j, \frac{\Lambda h}{2^K}\right) - U_n\left(\theta + \frac{\Lambda h}{2^K}\right) + U_n(\theta)$$

$$= R\left(X_i, X_j, \frac{\Lambda h}{2^K}\right) - \frac{1}{n(n-1)} \sum_{i \neq j \in [n]} R\left(X_i, X_j, \frac{\Lambda h}{2^K}\right).$$

Lemma D.2. Let $0 < \Lambda \leq 1/(2\|f\|_\infty)$. Under $\mathcal{E}_K = \{\Gamma_n \leq \Gamma_\diamond\}$, for any $z \leq Cn$, we have

$$\mathbb{P}^\star\left\{\sup_{|t| \leq \Lambda} |W_n^\star(t) - W_n^\star(0)| > \sup_{|t| \leq \Lambda} |W_n(t) - W_n(0)| + \frac{\Lambda\|U'\|_\infty}{2^K} + \frac{C_1 n(z + K) + C_2\sqrt{\Gamma_\diamond(z + K)}}{2n^2}\right\}$$
$$\leq C\exp(-z) + \exp\left(-\frac{n}{16}\right).$$

(D.2)

Proof of Lemma D.2. For any $h \in [2^K]$ and $t \in (\Lambda(h-1)2^{-K}, \Lambda h2^{-K}]$, we have

$$W_n^\star(t) - W_n^\star(0) = U_n^\star(\theta + t) - U(\theta + t) - \{U_n^\star(\theta) - U(\theta)\}$$
$$\leq U_n^\star\left(\theta + \frac{\Lambda h}{2^K}\right) - U\left(\theta + \frac{\Lambda(h-1)}{2^K}\right) - \{U_n^\star(\theta) - U(\theta)\}$$
$$\leq W_n^\star\left(\frac{\Lambda h}{2^K}\right) - W_n^\star(0) + \frac{\Lambda\|U'\|_\infty}{2^K}$$
$$\leq W_n^\diamond\left(\frac{\Lambda h}{2^K}\right) - W_n^\diamond(0) + \sup_{|t| \leq \Lambda} |W_n(t) - W_n(0)| + \frac{\Lambda\|U'\|_\infty}{2^K}$$

32

Similarly, we have

$$W_n^\star(t) - W_n^\star(0) \geq W_n^\diamond\left(\frac{\Lambda(h-1)}{2^K}\right) - W_n^\diamond(0) - \sup_{|t| \leq \Lambda} |W_n(t) - W_n(0)| - \frac{\Lambda\|U'\|_\infty}{2^K}.$$

Consequently, we have

$$\sup_{0 < t \leq \Lambda} |W_n^\star(t) - W_n(0)| \leq \max_{h \in [2^K]} \left| W_n^\diamond\left(\frac{\Lambda h}{2^K}\right) - W_n^\diamond(0) \right| + \sup_{|t| \leq \Lambda} |W_n(t) - W_n(0)| + \frac{\Lambda\|U'\|_\infty}{2^K}$$

Hence it suffices to upper bound $\max_{h \in [2^K]} |W_n^\diamond(\Lambda h/2^K) - W_n^\diamond(0)|$. For each $h \in [2^K]$, we have

$$
\begin{aligned}
W_n^\diamond\left(\frac{\Lambda h}{2^K}\right) - W_n^\diamond(0) &= \frac{1}{N} \sum_{i \neq j \in [n]} \mathcal{R}\left(X_i, X_j, \frac{\Lambda h}{2^K}\right)(\omega_i\omega_j - 1) \\
&= \frac{1}{N} \sum_{i \neq j \in [n]} \mathcal{R}\left(X_i, X_j, \frac{\Lambda h}{2^K}\right)(\omega_i - 1)(\omega_j - 1) \\
&\quad + \frac{2}{N} \sum_{j=1}^n \left\{ \sum_{i \neq j} \mathcal{R}\left(X_i, X_j, \frac{\Lambda h}{2^K}\right) \right\}(\omega_j - 1) \\
&=: \Delta_{W,h}^\flat + \Delta_{W,h}^\sharp.
\end{aligned}
$$

Recall that $B_n = \sum_{i \neq j \in [n]} \omega_i\omega_j$, where $\omega_1, \ldots, \omega_n$ are i.i.d. random variables such that $0 \leq \omega_i \leq 2$ and $\mathbb{E}(\omega_i) = 1$ for each $i \in [n]$. Hence, it follows from Lemma B.2 that $\mathbb{P}(B_n > 2n(n-1)) \leq \exp(-n/16)$. We first upper bound $\max_{h \in [2^K]} |\Delta_{W,h}^\flat|$. Observe that $\Delta_{W,h}^\flat$ is a degenerate second order U-statistic with

$$\max_{i \neq j \in [n]} \max_{h \in [2^K]} \left| \mathcal{R}\left(X_i, X_j, \frac{\Lambda h}{2^K}\right) \right| \leq 1.$$

Hence, by Theorem 3.3 in Giné et al. (2000),

$$\mathbb{P}^\star\left(\max_{h \in [2^K]} |\Delta_{W,h}^\flat| > \frac{z}{2n^2}\right) \leq C2^K \exp\left[-\frac{1}{C} \min\left\{ \frac{z^2}{n^2}, \frac{z}{n}, \left(\frac{z}{\sqrt{n}}\right)^{2/3}, \sqrt{z} \right\} \right] + \exp\left(-\frac{n}{16}\right).$$

We now upper bound $\max_{h \in [2^K]} |\Delta_{W,h}^\sharp|$. For simplicity of notation, denote $R_{ij,h} = R(X_i, X_j, \Lambda h/2^K)$ and $\bar{R}_{n,h} = \{n(n-1)\}^{-1} \sum_{i \neq j \in [n]} R_{ij,h}$. Define

$$\Gamma_n = \max_{h \in [2^K]} \sum_{j=1}^n \left| \sum_{i \neq j} \mathcal{R}\left(X_i, X_j, \frac{\Lambda h}{2^K}\right) \right|^2. \tag{D.3}$$

Hence, under $\mathcal{E}_K$, by Theorem 1.1 in Bentkus and Dzindzalieta (2015), it follows that

$$\mathbb{P}^\star\left\{ \max_{h \in [2^K]} |\Delta_{W,h}^\sharp| > \frac{C_1\sqrt{\Gamma_\diamond(z+K)}}{2n^2} \right\} \leq C_2 \exp(-z) + \exp\left(-\frac{n}{16}\right).$$

Putting all these pieces together, we obtain (D.2). $\qquad\square$

Lemma D.3. Let Γ_n be defined in (D.3). Assume that $\Lambda \le 1/(2\|f\|_\infty)$ and $\log n \le n\Lambda\|f\|_\infty$. Then we have

$$\mathbb{P}(\Gamma_n \le Cn^3\Lambda^2\|f\|_\infty^2) \ge 1 - C_1\exp(-n\Lambda\|f\|_\infty). \tag{D.4}$$

Proof of Lemma D.3. By the triangle inequality,

$$\Gamma_n \lesssim \max_{h\in[2^K]} \sum_{j=1}^n \left|\sum_{i\ne j}\{R_{ij,h} - \mathbb{E}(R_{ij,h}|X_j)\}\right|^2 + \max_{h\in[2^K]} n^2\sum_{j=1}^n |\mathbb{E}(R_{ij,h}|X_j) - \mathbb{E}(R_{ij,h})|^2$$
$$+ \max_{h\in[2^K]} n^3|\bar{R}_{n,h} - \mathbb{E}(\bar{R}_{n,h})|^2 =: \Gamma_{n,1} + \Gamma_{n,2} + \Gamma_{n,3}.$$

For each $j \in [n]$, conditional on X_j, $\{R_{ij,h} - \mathbb{E}(R_{ij,h}|X_j)\}_{i\ne j}$ are independent centered random variables with

$$\max_{h\in[2^K]}\max_{j\in[n]}|R_{ij,h} - \mathbb{E}(R_{ij,h}|X_j)| \le 1 \quad \text{and} \quad \max_{h\in[2^K]}\max_{j\in[n]}\mathbb{E}\{|R_{ij,h} - \mathbb{E}(R_{ij,h}|X_j)|^2|X_j\} \le \Lambda\|f\|_\infty.$$

Then, by the Bernstein inequality, with probability at least $1 - C\exp(-z)$, we have

$$\Gamma_{n,1} \le n\max_{h\in[2^K]}\max_{j\in[n]}\left|\sum_{i\ne j}\{R_{ij,h} - \mathbb{E}(R_{ij,h}|X_j)\}\right|^2 \le C_1\{n(z + \log n)^2 + n^2\Lambda\|f\|_\infty(z + \log n)\}.$$

We now bound $\Gamma_{n,2}$. Observe that $\{|\mathbb{E}(R_{ij,h}|X_j) - \mathbb{E}(R_{ij,h})|^2\}_{j\in[n]}$ are independent random variables with

$$\max_{h\in[2^K]}\max_{j\in[n]}|\mathbb{E}(R_{ij,h}|X_j) - \mathbb{E}(R_{ij,h})|^2 \le \Lambda^2\|f\|_\infty^2.$$

Hence, it follows from Lemma B.1 that

$$\mathbb{P}\left[\Gamma_{n,2} \le C\left\{n^3\Lambda^2\|f\|_\infty^2 + n^{5/2}\Lambda^2\|f\|_\infty^2\sqrt{(z + \log n)}\right\}\right] \ge 1 - \exp(-z).$$

Recall that $W_n(t) = \{n(n-1)\}^{-1}\sum_{i\ne j\in[n]} R(X_i, X_j, t)$. Hence $\bar{R}_{n,h} - \mathbb{E}(\bar{R}_{n,h}) = W_n(\Lambda h/2^K) - W_n(0)$ for each $h \in [2^K]$. Therefore, by Lemma C.2, with probability at least $1 - C\exp(-z)$, we have

$$\Gamma_{n,3} \le Cn^3\left\{\left(\frac{K+z}{n}\right)^2 + \Lambda\|f\|_\infty\sqrt{\frac{z+c}{n}} + \frac{K+z}{n}\sqrt{\Lambda\|f\|_\infty}\right\}^2.$$

Putting all these pieces together, we obtain (D.4). $\qquad\square$

Proof of Theorem 3.1. By Theorem 2.1, for any $\omega > 0$ such that $\omega \le n\kappa_0^2 c_0^2$, we have

$$\mathbb{P}\left(|\hat{\theta} - \theta| > \sqrt{\omega/(n\kappa_0^2)}\right) \le 2\exp(-\omega).$$

Combined with Lemma C.2, for any $z > 0$ such that $z \leq 1/(4\|f\|_\infty)$, with probability at least $1 - C_1 \exp(-\omega)$, we have

$$
\begin{aligned}
&|U_n(\widehat{\theta} + z) - U_n(\widehat{\theta}) - \{U(\widehat{\theta} + z) - U(\widehat{\theta})\}| \\
&\qquad \leq |U_n(\widehat{\theta} + z) - U_n(\theta) - \{U(\widehat{\theta} + z) - U(\theta)\}| \\
&\qquad\quad + |U_n(\widehat{\theta}) - U_n(\theta) - \{U(\widehat{\theta}) - U(\theta)\}| \\
&\qquad \leq C_2 \left(\|f\|_\infty z \sqrt{\frac{\omega + C}{n}} + \frac{K + \omega}{n} \sqrt{\|f\|_\infty z} + \frac{K + \omega}{n\kappa_0} \|f\|_\infty \right).
\end{aligned}
$$

Similar to the proof of Theorem 2.1, for any $z \geq 0$ such that $\widehat{\theta}^\star > \widehat{\theta} + z$, we have

$$
\begin{aligned}
\frac{1}{2} &> U_n^\star(\widehat{\theta} + z) \geq U_n(\widehat{\theta} + z) - |U_n^\star(\widehat{\theta} + z) - U_n(\widehat{\theta} + z)| \\
&\geq U_n(\widehat{\theta}) + \left\{ U(\widehat{\theta} + z) - U(\widehat{\theta}) \right\} - |U_n^\star(\widehat{\theta} + z) - U_n(\widehat{\theta} + z)| \\
&\quad - |U_n(\widehat{\theta} + z) - U_n(\widehat{\theta}) - \{U(\widehat{\theta} + z) - U(\widehat{\theta})\}| \\
&\geq \frac{1}{2} + \frac{\kappa_0 z}{2} - |U_n^\star(\widehat{\theta} + z) - U_n(\widehat{\theta} + z)| - C_2 \left(\frac{K + \omega}{n} \sqrt{\|f\|_\infty z} + \frac{K + \omega}{n\kappa_0} \|f\|_\infty \right) \\
&\geq \frac{1}{2} + \frac{\kappa_0 z}{2} - |U_n^\star(\widehat{\theta} + z) - U_n(\widehat{\theta} + z)| - \frac{2C_2(K + \omega)}{n\kappa_0} \|f\|_\infty.
\end{aligned}
$$

Then it follows from Lemma B.1 that

$$
\mathbb{P}^\star \left\{ |\widehat{\theta}^\star - \widehat{\theta}| > \frac{2}{\kappa_0} \sqrt{\frac{z}{n}} + \frac{4C_2 \|f\|_\infty (K + \omega)}{n\kappa_0^2} \right\} \leq 2 \exp(-C_3 z) + \exp\left(-\frac{n}{16}\right).
$$

By Lemma D.2 and Lemma D.3, with probability at least $1 - C \exp(-\omega)$, we have

$$
\mathbb{P}^\star \left\{ |U_n^\star(\widehat{\theta}^\star) - U(\widehat{\theta}^\star) - \{U_n^\star(\theta) - U(\theta)\}| > \frac{\mathcal{C}(K + z + \omega)}{n} \right\} \leq C \exp(-z) + \exp\left(-\frac{n}{16}\right),
$$

where $\mathcal{C}$ is a positive constant depending only on $\kappa_0, \|f\|_\infty$ and c_0. Therefore, combined with the fact that $\sup_{|\delta| \leq c_1} |U''(\theta + \delta)| \leq \kappa_1$, we obtain

$$
\left| \widehat{\theta}^\star - \theta - \frac{1}{B_n U'(\theta)} \sum_{i \neq j \in [n]} \omega_i \omega_j \left(\mathbb{I}\{\xi_i + \xi_j \leq 0\} - \frac{1}{2} \right) \right| \leq \frac{\mathcal{C}(K + z + \omega)}{n}.
$$

Combining these with Theorem 2.1, we obtain (3.2). $\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof of (3.3). For each $i \in [n]$, denote

$$
\mathcal{D}_i = \frac{2}{U'(\theta)} \left(\frac{1}{n - 1} \sum_{j \neq i} \mathbb{I}\{\xi_i + \xi_j \leq 0\} - U_n(\theta) \right).
$$

Then if follows from (3.2) that with probability at least $1 - C \exp(-\omega)$, we have

$$
\mathbb{P}^\star \left\{ \left| \widehat{\theta}^\star - \widehat{\theta} - \frac{1}{n} \sum_{i=1}^n \mathcal{D}_i (\omega_i - 1) \right| > \frac{C(\omega + \log n)}{n} \right\} \leq C n^{-c}. \tag{D.5}
$$

By Lemma D.1, we have $\mathbb{P}(\widehat{\sigma}_\theta \geq \sigma_\theta/2) \geq 1 - C\exp(-n)$. Consequently, by Berry-Esseen theorem, with probability at least $1 - C\exp(-n)$, we have

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}^\star\left(\frac{1}{\sqrt{n}\widehat{\sigma}_\theta}\sum_{i=1}^n \mathcal{D}_i(\omega_i - 1) \leq z\right) - \Phi(z)\right| \leq \frac{C}{\sqrt{n}}.$$

Combining this with (D.5) yields (3.3). $\qquad\square$

D.2   Proof of Theorem 3.2

Based on Theorem 2.4, the proof of Theorem 3.2 can be derived by following the proof of Theorem 3.1. Thus we omit the details.

E   Proof of Results in §4

E.1   Proof of Theorem 4.1

Proof of Theorem 4.1. For each $i \in [n]$ and $\ell \in [p]$, denote $D_{i\ell} = \{1 - 2F_\ell(-\xi_{i\ell})\}/U'_\ell(\theta_\ell)$. Define $\bar{\boldsymbol{D}}_n = (\bar{D}_{n1},\ldots,\bar{D}_{np})^\top \in \mathbb{R}^p$, where $\bar{D}_{n\ell} = n^{-1}\sum_{i=1}^n D_{i\ell}$ for each $\ell \in [p]$. Note that $|D_{i\ell}| \leq 1/\kappa_0$ uniformly for $i \in [n]$ and $\ell \in [p]$. Hence, it follows from Theorem 2.1 in Chernozhuokov et al. (2022) that

$$\sup_{z\in\mathbb{R}}|\mathbb{P}(\sqrt{n}\|\bar{\boldsymbol{D}}_n\|_\infty \leq z) - \mathbb{P}(\|\boldsymbol{Z}\|_\infty \leq z)| \leq \frac{C\log^{5/4}(pn)}{n^{1/4}}.$$

Let $\delta^\sharp = C\log(pn)/\sqrt{n}$. By Theorem 2.1, it follows that

$$\sup_{z\in\mathbb{R}}|\mathbb{P}(\sqrt{n}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_\infty \leq z) - \mathbb{P}(\sqrt{n}\|\bar{\boldsymbol{D}}_n\|_\infty \leq z)|$$
$$\leq \mathbb{P}(\sqrt{n}\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} - \bar{\boldsymbol{D}}_n\|_\infty > \delta^\sharp) + \sup_{z\in\mathbb{R}}\mathbb{P}(z < \sqrt{n}\|\bar{\boldsymbol{D}}_n\|_\infty \leq z + \delta^\sharp)$$
$$\lesssim n^{-c} + \delta^\sharp\sqrt{\log p} + \frac{\log^{5/4}(pn)}{n^{1/4}} \leq \frac{C\log^{5/4}(pn)}{n^{1/4}}.$$

Putting all these pieces together, we obtain (4.2). $\qquad\square$

E.2   Proof of Theorem 4.2

Proof of Theorem 4.2. For each $\ell \in [p]$ and $i \in [n]$, denote

$$\mathcal{D}_{i\ell} = \frac{2}{\sigma_\ell U'_\ell(\theta_\ell)}\left(\frac{1}{n-1}\sum_{j\neq i}\mathbb{I}\{\xi_{i\ell} + \xi_{j\ell} \leq 0\} - U_{n\ell}(\theta_\ell)\right).$$

Let $\bar{\boldsymbol{\mathcal{D}}}_n = (\bar{\mathcal{D}}_{n1}, \ldots, \bar{\mathcal{D}}_{np})^\top$, where $\bar{\mathcal{D}}_{n\ell} = n^{-1} \sum_{i=1}^n \mathcal{D}_{i\ell}$ for each $\ell \in [p]$. By the triangle inequality,

$$\max_{\ell \in [p]} \sum_{i=1}^n (D_{i\ell} - \mathcal{D}_{i\ell})^2 \lesssim \max_{\ell \in [p]} \sum_{i=1}^n \left(\frac{1}{n-1} \sum_{j \neq i} \mathbb{I}\{\xi_{i\ell} + \xi_{j\ell} \leq 0\} - F_\ell(-\xi_{i\ell}) \right)^2 + n \max_{\ell \in [p]} \left| U_{n\ell}(\theta_\ell) - \frac{1}{2} \right|^2$$

$$\leq n \max_{\ell \in [p]} \max_{i \in [n]} \left| \frac{1}{n-1} \sum_{j \neq i} \mathbb{I}\{\xi_{i\ell} + \xi_{j\ell} \leq 0\} - F_\ell(-\xi_{i\ell}) \right|^2 + n \max_{\ell \in [p]} \left| U_{n\ell}(\theta_\ell) - \frac{1}{2} \right|^2$$

$$=: \Delta_D^\diamond + \Delta_D^\circ.$$

By Lemma B.1, it follows that $\mathbb{P}\{\Delta_D^\diamond > C \log(np)\} \lesssim (np)^{-c}$. Recall that

$$U_\ell(\theta_\ell) = \frac{1}{n(n-1)} \sum_{i \neq j \in [n]} \mathbb{I}\{\xi_{i\ell} + \xi_{j\ell} \leq 0\}, \quad \ell \in [p].$$

Hence, it follows from Lemma B.2 that $\mathbb{P}\{\Delta_D^\circ > C \log(np)\} \lesssim (np)^{-c}$. Consequently, for any $\delta > 0$, by Theorem 1.1 in Bentkus and Dzindzalieta (2015), with probability at least $1 - C(np)^{-c}$, we have

$$\mathbb{P}^\star \left(\max_{\ell \in [p]} \left| \sum_{i=1}^n (D_{i\ell} - \mathcal{D}_{i\ell})(\omega_i - 1) \right| > \delta \right) \lesssim p \exp \left(-\frac{C\delta^2}{\log(np)} \right).$$

Combined with Lemma, with probability at least $1 - C(np)^{-c}$, we have

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P}^\star \left(\max_{\ell \in [p]} \left| \sum_{i=1}^n D_{i\ell}(\omega_i - 1) \right| \leq z \right) - \mathbb{P}^\star \left(\max_{\ell \in [p]} \left| \sum_{i=1}^n \mathcal{D}_{i\ell}(\omega_i - 1) \right| \leq z \right) \right|$$

$$\lesssim p \exp\{-C \log(np)\} + \frac{\log(np)\sqrt{\log p}}{\sqrt{n}} \leq \frac{\log(np)\sqrt{\log p}}{\sqrt{n}}.$$

$\square$

E.3 Proof of Theorem 4.3

Proof of Theorem 4.3. Following the proof of Theorem 2.3, it follows from Theorem 3.1 that with probability at least $1 - Cn^{-c}$, we have

$$\frac{\mathbb{P}^\star\{\sqrt{n}(\widehat{\theta}_\ell^\star - \widehat{\theta}_\ell) > z\}}{2\{1 - \Phi(z/\sigma_\ell)\}} = 1 + o(1), \tag{E.1}$$

uniformly for $0 < z \leq o(n^{1/6})$ and $\ell \in [p]$. By Lemma 1 in Storey et al. (2004), it follows that

$$t_S = \left\{ t \in [0, 1] : t \leq \frac{\alpha \max(\sum_{\ell=1}^p \mathbb{I}\{P_\ell \leq t\}, 1)}{p} \right\}.$$

By the definition of t_S, we have

$$t_S = \frac{\alpha \max(\sum_{\ell=1}^p \mathbb{I}\{P_\ell \leq t_S\}, 1)}{p}. \tag{E.2}$$

Observe that $t_S \geq \alpha/p$. Hence, it follows from (E.1) that $\mathbb{P}\{t_S > \max_{\ell \in [p]} \mathcal{G}_\ell^\star(\sigma_\ell(2\log p)^{1/2})\} \to 1$. Combining this with (E.2) yields

$$
\begin{aligned}
t_S &\geq \frac{\alpha}{p} \sum_{\ell=1}^p \mathbb{I}\{P_\ell \leq t_S\} \geq \frac{\alpha}{p} \sum_{\ell=1}^p \mathbb{I}\left\{ \mathcal{G}_\ell^\star(\sqrt{n}|\widehat{\theta}_\ell|) \leq \mathcal{G}_\ell^\star\left(\sigma_\ell\sqrt{2\log p} \right) \right\} \\
&\geq \frac{\alpha}{p} \sum_{\ell=1}^p \mathbb{I}\left\{ \sqrt{n}|\widehat{\theta}_\ell| \geq \sigma_\ell\sqrt{2\log p} \right\} \\
&\geq \frac{\alpha}{p} \sum_{\ell=1}^p \mathbb{I}\left\{ \frac{|\theta_\ell|}{\sigma_\ell} \geq \sqrt{2\log p} + \max_{\ell \in [p]} \frac{\sqrt{n}(\widehat{\theta}_\ell - \theta_\ell)}{\sigma_\ell} \right\}.
\end{aligned}
$$

For some $\lambda > 0$, define

$$
\mathcal{E}_\lambda = \left\{ \sqrt{n} \max_{\ell \in [p]} \frac{|\widehat{\theta}_\ell - \theta_\ell|}{\sigma_\ell} \leq (1+\lambda)\sqrt{2\log p} \right\}.
$$

Under $\mathcal{E}_\lambda$, it is straightforward to verify that

$$
t_S \geq \frac{\alpha}{p} \sum_{\ell=1}^p \mathbb{I}\left\{ \frac{|\theta_\ell|}{\sigma_\ell} \geq (2+\lambda)\sqrt{2\log p} \right\}
$$

By theorem 2.3, it follows that $\mathbb{P}(\mathcal{E}_\lambda^c) \leq Cp\exp\{-(1+\lambda)^2\log p\} \leq Cp^{-\lambda^2-2\lambda}$. Then, following the proof of Theorem 3.3 in Zhou et al. (2018), for any sequence $1 \leq b_p \to \infty$, it is straightforward to derive that

$$
\sup_{b_p/p \leq t \leq 1} \left| \frac{\sum_{\ell \in \mathcal{H}_0} \mathbb{I}\{P_\ell \leq t\}}{|\mathcal{H}_0|t} - 1 \right| \xrightarrow{\mathbb{P}} 0.
$$

Putting all these pieces together, we obtain (4.6). $\qquad\square$

E.4 Large-scale Two-sample Simultaneous Testing

We consider the following high dimensional two-sample global test,

$$
H_0 : \theta_\ell = \theta_\ell^\circ \ \text{for all} \ \ell \in [p] \ \text{versus} \ H_1 : \theta_\ell \neq \theta_\ell^\circ \ \text{for some} \ \ell \in [p]. \tag{E.3}
$$

Given the HL estimators $\widehat{\Theta}_\ell = \text{median}\{X_{i\ell} - Y_{j\ell} : i \in [n], j \in [m]\}$, $\ell \in [p]$, for $\Theta_\ell = \theta_\ell - \theta_\ell^\circ$, we shall the null hypothesis H_0 in (E.3) whenever $\max_{\ell \in [p]} |\widehat{\Theta}_\ell|$ exceeds some threshold. To this end, we develop a Gaussian approximation for $\max_{\ell \in [p]} |\widehat{\Theta}_\ell - \Theta_\ell|$. More specifically, let $\boldsymbol{\mathcal{Z}} = (\mathcal{Z}_1, \ldots, \mathcal{Z}_p)^\top$ be a p-dimensional centered Gaussian random vector with

$$
\text{Cov}(\mathcal{Z}_k, \mathcal{Z}_\ell) = \frac{N\,\text{Cov}(\bar{G}_{nk} - \bar{F}_{mk}, \bar{G}_{n\ell} - \bar{F}_{m\ell})}{\mathcal{U}_k'(\Theta_k)\mathcal{U}_\ell'(\Theta_\ell)}, \ \ k, \ell \in [p].
$$

In the following theorem, we establish a non-asymptotic upper bound for the Kolmogorov distance between the distribution functions of $\max_{\ell \in [p]} |\widehat{\Theta}_\ell - \Theta_\ell|$ and its Gaussian analogue $\max_{\ell \in [p]} |\mathcal{Z}_\ell|$.

Theorem E.1. Let $\mathcal{U}_\ell(t) = \mathbb{P}(X_{1\ell} - Y_{1\ell} \leq t)$ for each $\ell \in [p]$. Assume that there exist positive constants $\bar{c}_0, \bar{\kappa}_0, \bar{c}_1$ and $\bar{\kappa}_1$ such that

$$\min_{\ell \in [p]} \inf_{|\Lambda| \leq \bar{c}_0} \mathcal{U}'_\ell(\Theta_\ell + \Lambda) \geq \bar{\kappa}_0 > 0 \text{ and } \max_{\ell \in [p]} \sup_{|\Lambda| \leq \bar{c}_1} |\mathcal{U}''_\ell(\Theta_\ell + \Lambda)| \leq \bar{\kappa}_1.$$

Then, under Assumption 2.1, we have

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P}\left(\max_{\ell \in [p]} \sqrt{N}|\widehat{\Theta}_\ell - \Theta_\ell| \leq z \right) - \mathbb{P}\left(\max_{\ell \in [p]} |\mathcal{Z}_\ell| \leq z \right) \right| \leq \frac{C \log^{5/4}(Np)}{N^{1/4}}.$$

Proof of Theorem E.1. We are able to prove Theorem E.1 by using similar arguments as in Theorem 4.1. The major differences only lies in changing the one-sample to two-sample estimators. Thus, we decided to omit the corresponding details. □

Motivated by Theorem E.1, an asymptotic α-level test for (E.3) is given by $\mathbb{I}\{\max_{\ell \in [p]} |\widehat{\Theta}_\ell| > Q^\diamond_{1-\alpha}\}$, where $Q^\diamond_{1-\alpha}$ stands for the $(1-\alpha)$th quantile of the bootstrap statistic $\max_{\ell \in [p]} |\widehat{\Theta}^\star_\ell - \widehat{\Theta}_\ell|$, namely,

$$Q^\diamond_{1-\alpha} = \inf \left\{ z \in \mathbb{R} : \mathbb{P}^\star \left(\max_{\ell \in [p]} |\widehat{\Theta}^\star_\ell - \widehat{\Theta}_\ell| \leq z \right) \geq 1 - \alpha \right\}, \quad \alpha \in (0, 1).$$

The validity of the proposed test is justified via the following theorem.

Theorem E.2. Under the conditions of Theorem E.1, we have

$$\left| \mathbb{P}\left(\max_{\ell \in [p]} |\widehat{\Theta}_\ell - \Theta_\ell| > Q^\diamond_{1-\alpha} \right) - \alpha \right| \leq \frac{C \log^{5/4}(pN)}{N^{1/4}}.$$

Proof of Theorem E.2. The proof of Theorem E.2 can be derived similarly by following the proof procedure of Theorem 4.2 by replacing the one-sample estimator with the two-sample version. Therefore, we omit the details. □

E.5   Proof of Theorem 4.4

The proof of Theorem 4.4 can be derived similarly by following the proof procedure of Theorem 4.3 by changing the one-sample estimator to the two-sample version. Therefore, we omit the details.