INDEPENDENCE TESTING IN HIGH DIMENSIONS

BY PATRICK BASTIAN[1,a], HOLGER DETTE[1,b] AND JOHANNES HEINY[1,c]

[1]Ruhr-Universität Bochum, [a]patrick.bastian@rub.de

[1]Ruhr-Universität Bochum, [b]holger.dette@rub.de

[1]Ruhr-Universität Bochum, [c]johannes.heiny@rub.de

This paper takes a different look on the problem of testing the mutual independence of the components of a high-dimensional vector. Instead of testing if all pairwise associations (e.g. all pairwise Kendall's τ) between the components vanish, we are interested in the (null)-hypothesis that all pairwise associations do not exceed a certain threshold in absolute value. The consideration of these hypotheses is motivated by the observation that in the high-dimensional regime, it is rare, and perhaps impossible, to have a null hypothesis that can be exactly modeled by assuming that all pairwise associations are precisely equal to zero.

The formulation of the null hypothesis as a composite hypothesis makes the problem of constructing tests non-standard and in this paper we provide a solution for a broad class of dependence measures, which can be estimated by U-statistics. In particular we develop an asymptotic and a bootstrap level α-test for the new hypotheses in the high-dimensional regime. We also prove that the new tests are minimax-optimal and demonstrate good finite sample properties by means of a small simulation study.

1. Introduction. Measuring dependence and testing for independence are fundamental problems in statistics and since the early work of Pearson (1920), Kendall (1938), Hoeffding (1948b) and Blum et al. (1961) numerous authors have worked in this area (for some more recent references, see Gretton et al., 2008; Székely et al., 2007; Heller et al., 2012; Dette et al., 2012; Bergsma and Dassios, 2014; Albert et al., 2015; Geenens and Lafaye de Micheaux, 2020; Chatterjee, 2021, among many others). Similarly, testing for mutual independence of the components of a vector has found considerable attention in the literature and exemplary we refer to Narain (1950), Roy (1957), Lee (1971), Nagao (1973), and Chapter 9 in the book of Anderson (1984). However, it is well known that the last-named tests do not perform well if the dimension, say p, is comparable to or even larger than the sample size, say n, and in recent years many authors have worked on testing for mutual independence of the components in the high-dimensional regime, where the dimension p converges with the sample size n to infinity.

Independence testing of high-dimensional (mostly) Gaussian data has been considered by Bai et al. (2009), Jiang and Yang (2013), Jiang and Qi (2015), Chen and Kato (2017), Bodnar et al. (2019) and Dette and Dörnemann (2020), among others, who investigated the asymptotic properties of likelihood ratio tests. Other authors consider more general distributions, where the dependence between two components of the vectors is estimated by different covariance/correlation statistics such as Pearson's r, Spearman's ρ, and Kendall's τ, and different functions are used to aggregate these estimates of the pairwise dependencies. For example, Bao et al. (2015) and Li et al. (2021) use linear spectral statistics of the matrix of estimates, while Schott (2005); Qiu and Chen (2012); Yao et al. (2018) and Leung and Drton

MSC2020 subject classifications: 62G10, 62F40, 62C20.

Keywords and phrases: independence testing, relevant association, U-statistics, gaussian approximation, bootstrap, minimax optimality.

(2018) propose tests based on the Frobenius norm. Further very popular methods of aggregating estimates of the pairwise dependencies are maximum-type tests, which have good power properties against sparse alternatives and have been investigated for various covariance/correlation statistics in Jiang (2004); Zhou (2007); Liu et al. (2008); Li et al. (2010); Cai and Jiang (2012); Shao and Zhou (2014); Han et al. (2017); Drton et al. (2020); Heiny et al. (2021) and He et al. (2021) among others.

These tests differ in the distributional assumptions, the way of aggregation and in the considered measures to quantify the dependence between two components. However, a common feature of all cited references consists in the fact that statistical tests are proposed for the hypotheses

$$
\begin{aligned}
&H_0^{\text{exact}}: \;\; d_{ij} = 0 \quad \text{for all } 1 \le i < j \le p\,, \\
&H_1^{\text{exact}}: \;\; d_{ij} \neq 0 \quad \text{for at least one pair } (i,j) \text{ with } 1 \le i < j \le p\,,
\end{aligned}
\tag{1.1}
$$

where $d_{ij} = d(X_{1i}, X_{1j})$ is a (population) measure of dependence between the two components X_{1i} and X_{1j} of the p-dimensional random vector $X_1 = (X_{11}, \dots, X_{1p})^\top$, such as the covariance $\mathrm{Cov}(X_{1i}, X_{1j})$.

In the present paper we take a different point of view on the problem of testing the mutual independence of the components of a high-dimensional vector. Our work is motivated by the paper of Berger and Delampady (1987) who argue that it is rare, and perhaps impossible, to have a null hypothesis that can be exactly modeled by a parameter being exactly 0. In the context of independence testing this means, that in many applications, in particular in the high-dimensional regime, it is often unlikely that all $p(p-1)/2$ associations (measured by d_{ij}) satisfy $d_{ij} = 0$ ($1 \le i < j \le p$). As a consequence one uses a formulation of the null hypothesis in (1.1), which is believed to be not true, and for sufficiently large sample size any consistent test will detect an arbitrary small deviation from the null hypothesis, which might not be scientifically of interest. Problems of this type are particularly relevant in the big-data era, where the sample size and dimension are usually large.

As an alternative we propose to investigate if all associations (measured by the quantities d_{ij}) are in some sense “small”. For this purpose we consider the hypotheses

$$
\begin{aligned}
&H_0: \;\; |d_{ij}| \le \Delta \quad \text{for all } 1 \le i < j \le p\,, \\
&H_1: \;\; |d_{ij}| > \Delta \quad \text{for at least one pair } (i,j) \text{ with } 1 \le i < j \le p\,,
\end{aligned}
\tag{1.2}
$$

where $\Delta > 0$ is a given threshold. Note that (1.1) is obtained from (1.2) for $\Delta = 0$, but in the present paper we are not interested in this case, because we are aiming to detect only dependencies exceeding a given positive threshold. The rejection of H_0 in (1.2) allows to decide at a controlled type I error that at least one association is larger than the given threshold Δ. On the other hand, interchanging the null hypothesis in (1.2) and developing an appropriate test allows to decide at a controlled type I error that all measures $|d_{ij}|$ are smaller than Δ (note that interchanging the null-hypothesis and alternative does not make sense for the hypotheses in (1.1)). We also note that hypotheses of the form (1.2) are frequently used in biostatistics for inference on one-dimensional parameters (see, for example, the monographs of Chow and Liu, 1992; Wellek, 2010).

The purpose of the present paper is the development of statistical tests for hypotheses of the form (1.2) in the high-dimensional regime, where the dependence measures d_{ij} can be estimated by U-statistics. Typical examples include the classical covariance, Kendall's τ, Hoeffding's D, Blum-Kiefer-Rosenblatt's R, Bergsma–Dassios–Yanagimoto's τ^* and a dominating term of Spearman's rank correlation ρ. As Jiang (2004); Zhou (2007); Liu et al. (2008); Han et al. (2017) and Drton et al. (2020) we consider maximum-type tests, and allow the dimension p to grow exponentially with n. We develop a new asymptotic and a new bootstrap test for the hypotheses in (1.2) and investigate their statistical properties. Compared to the "classical" hypotheses in (1.1) the composite structure of the hypotheses in (1.2) makes both tasks non-standard. On the one hand, the asymptotic analysis of estimators of $\max_{1\le i<j\le p}|d_{ij}|$ by Poisson approximation techniques (see, for example, Arratia et al., 1989) is very demanding due to the additional dependencies under the null hypothesis in (1.2). On the other hand, further challenges arise in the development of bootstrap procedures, since "generating data under the null $H_0 : \max_{1\le i<j\le p}|d_{ij}| \le \Delta$" is not straightforward for the composite hypotheses in (1.2).

In Section 2 we consider testing problems of the form (1.2) in a more general context and propose an asymptotic level α test, which is (uniformly) consistent against local alternatives, where the maximum deviation is at least $\Delta + c\sqrt{\log d}/\sqrt{n}$ for some constant $c > 0$ (here $d = p(p-1)/2$ is the number of terms over which the maximum is taken). The proof of these properties is based on the weak convergence of an appropriately normalized maximum statistic to a Gumbel distribution under suitable assumptions on the dependence structure, sample size and dimension. As such assumptions are often hard to justify in statistical practice and the convergence rates in extreme value theory are usually very slow, we develop in Section 2.2 a non-standard bootstrap test for the hypotheses of the form (1.2) and prove its validity. In Section 3 we specialize these results to the problem of testing hypotheses of the form (1.2), where the associations d_{ij} are given by the covariances, Kendall's τ, a dominating term of Spearman's ρ, Hoeffding's D, Blum-Kiefer-Rosenblatt's R and Bergsma–Dassios–Yanagimoto's τ^*. In particular, we prove that for many dependence measures the tests proposed in this paper are minimax-optimal against local alternatives of the form $\max_{1\le i<j\le p}|d_{ij}| = \Delta + c\sqrt{\log p}/\sqrt{n}$. Note that these rates coincide with the minimax-optimal rates for testing the classical hypotheses (1.1), that is $\Delta = 0$, if dependencies are measured by Spearman's ρ and Kendall's τ correlations, see, for example, Han et al. (2017). In Section 4 we demonstrate by means of a simulation study that the developed methodology has good finite sample properties. Finally, all technical proofs and details are deferred to an online supplement (see Sections A and B).

2. Testing for relevant deviations. In this section we consider the testing problems in a slightly more general but notationally simpler form as described in the introduction. The case of testing for relevant deviations of the entries in a matrix of pairwise dependence measures is a special case of the following discussion (see Example 2.1) and will be addressed in Section 3 in more detail. To be precise, let $X_1, \dots, X_n$ denote independent identically distributed p-dimensional random vectors with distribution function F. Note that formally F depends on the dimension p, which varies with n, but we will not reflect this dependence in our notation throughout this paper. For some positive integer m let

$$h = (h_1, \dots, h_d)^\top : (\mathbb{R}^p)^m \to \mathbb{R}^d \tag{2.1}$$

denote a measurable symmetric function with finite expectation

$$\theta_F = (\theta_1, \dots, \theta_d)^\top = \mathbb{E}_F[h(X_1, \dots, X_m)] \in \mathbb{R}^d\,, \tag{2.2}$$

which defines our parameter of interest. In order to estimate the parameter θ_F we consider the U-statistic of order m

$$U = (U_1, \dots, U_d)^\top = \binom{n}{m}^{-1} \sum_{1\le l_1<\ldots<l_m\le n} h(X_{l_1}, \dots, X_{l_m})\,. \tag{2.3}$$

In the high-dimensional regime U-statistics have recently found considerable interest in the literature and we refer to Chen and Kato (2017); Chen (2018); Song et al. (2019); Kim (2020); Wang et al. (2021); Cheng et al. (2022) among others.

EXAMPLE 2.1. We briefly illustrate the notation for dependence measures between the components of high-dimensional vectors as introduced in Section 1. In particular, such U-statistics have been investigated by Han et al. (2017); Chen and Jiang (2018); Zhou et al. (2019); Drton et al. (2020) and He et al. (2021) in the context of independence testing by means of the classical hypotheses (1.1).
To be precise, for $1 \leq i < j \leq p$ let

$$d_{ij} = d(X_{1i}, X_{1j}) = \mathbb{E}_F[\tilde{h}(X_{1i}, X_{1j}, \dots, X_{mi}, X_{mj})] \tag{2.4}$$

denote a dependence measure between the ith and jth components of the random vector $X_1 = (X_{11}, \dots, X_{1p})^\top$, which can be expressed as the expectation of a kernel $\tilde{h} : \mathbb{R}^{2m} \to \mathbb{R}$ of order m evaluated at $(X_{1i}, X_{1j}, \dots, X_{mi}, X_{mj})$. In this case the function h in (2.1) is defined by

$$\begin{aligned} h(X_1, ..., X_m) &= \mathrm{vech}\big((h_{ij}(X_1, ..., X_m))_{i,j=1,...,p}\big) \\ &= \mathrm{vech}\big((\tilde{h}(X_{1i}, X_{1j}, ..., X_{mi}, X_{mj}))_{i,j=1,...,p}\big) , \end{aligned}$$

where the second equality defines the functions $h_{ij} : \mathbb{R}^{pm} \to \mathbb{R}$ in an obvious manner and vech$(\cdot)$ is the operator that stacks the columns above the diagonal of a symmetric $p \times p$ matrix as a vector with $d = p(p-1)/2$ components. Note that the index (i, j) in the definition of the function h_{ij} is only used to emphasize that each h_{ij} acts on different components of the vectors $X_1, \dots, X_m$. Similarly, the vector θ_F is defined by $\theta_F = \mathrm{vech}\big((d_{ij})_{i,j=1,...,p}\big)$, and the components of the vector $U = \mathrm{vech}\big((U_{ij})_{i,j=1,...,p}\big)$ in (2.3) are given by

$$\begin{aligned} U_{ij} &= \binom{n}{m}^{-1} \sum_{1 \leq l_1 < \ldots < l_m \leq n} h_{ij}(X_{l_1}, ..., X_{l_m}) \\ &= \binom{n}{m}^{-1} \sum_{1 \leq l_1 < \ldots < l_m \leq n} \tilde{h}(X_{l_1 i}, X_{l_1 j}, ..., X_{l_m i}, X_{l_m j}) . \end{aligned}$$

A more detailed discussion of specific dependence measures is postponed to Section 3.

Recall that, in this paper, we are not interested in testing the "classical" hypotheses $H_0 : \theta_F = 0$ versus $H_1 : \theta_F \neq 0$, but want to investigate if at least one of the components θ_i of the vector $\theta_F = (\theta_1, \dots, \theta_d)^\top$ exceeds a given threshold $\Delta > 0$, that is

$$H_0 : \max_{i=1}^{d} |\theta_i| \leq \Delta \ \text{ versus } \ H_1 : \max_{i=1}^{d} |\theta_i| > \Delta \ , \tag{2.5}$$

where Δ denotes the largest deviation that is still considered as negligible. Hypotheses of this form are often called *relevant* hypotheses. In the case $d = 1$ these hypotheses (more precisely the interchanged hypotheses $H_0 : \max_{i=1}^{d} |\theta_i| > \Delta$ versus $H_1 : \max_{i=1}^{d} |\theta_i| \leq \Delta$) have found considerable attention in the biostatistics literature (see, for example the monographs by Chow and Liu, 1992; Wellek, 2010), but - despite their importance - they have not been studied intensively in the high-dimensional regime. In what follows, we will construct tests for hypotheses of the form (2.5) based on asymptotic theory of a (standardized) estimator of $\max_{i=1}^{d} |\theta_i|$ and also develop (under substantially weaker assumptions) a non-standard bootstrap test in the high-dimensional regime, where we allow the dimension d to grow exponentially with n.

2.1. *An asymptotic level α test.* Recall the definition of the parameter $\theta_F = \mathbb{E}_F[U] = \mathbb{E}_F[h(X_1, \ldots, X_m)] \in \mathbb{R}^d$ in (2.2), where $X_1, \ldots, X_m \sim F$ are independent p-dimensional random vectors with distribution F (the dependence on p is omitted here for simplicity). In Example 2.1 and in most cases of practical interest, d is given as a function of p, but our theoretical results are more generally stated in a U-statistics framework that only depends on the dimension of the vector θ_F. We denote by $\mathcal{F}$ the class of all distribution functions on $\mathbb{R}^p$ for which $\mathbb{E}_F[U]$ exists, and we set $\theta_i = \theta_{F,i} = \mathbb{E}_F[h_i(X_1, \ldots, X_m)]$ to be the ith component of θ_F, where h_i, the ith component of the vector h in (2.1), is a symmetric kernel of order m. Define by

$$U_i = \binom{n}{m}^{-1} \sum_{1 \le l_1 < \ldots < l_m \le n} h_i(X_{l_1}, \ldots, X_{l_m}), \qquad i = 1, \ldots, d, \tag{2.6}$$

the corresponding estimate of θ_i. Under standard assumptions the statistics U_i are unbiased and consistent estimators of the parameters θ_i ($i = 1, \ldots, d$), and therefore it is reasonable to reject the null hypothesis in (2.5) for large values of $\max_{i=1}^d |U_i|$. For technical reasons we consider the quantities U_i^2 instead of $|U_i|$ and compare their maximum with Δ^2. Corresponding results for $|U_i|$, which estimates $|\theta_i|$, are briefly mentioned in Remark 2.8 (b).
We note that

$$U_i^2 - \Delta^2 = (U_i - \theta_i)^2 + 2\theta_i(U_i - \theta_i) - (\Delta^2 - \theta_i^2) \tag{2.7}$$

and introduce the notations

$$\zeta_{1,i} = \mathrm{Var}_F(h_{1,i}(X_1)) \quad \text{and} \quad h_{1,i}(x) = \mathbb{E}_F[h_i(X_1, \ldots, X_m) | X_1 = x]\,. \tag{2.8}$$

If $\zeta_{1,i} > 0$, the kernel h_i of the statistic U_i is called non-degenerate. Note that this property depends on the kernel h_i and on the distribution F. In particular, for composite null hypotheses of the form (2.5), there may exist different distributions, say $F_1, F_2 \in \mathcal{F}$, both corresponding to parameters θ_{F_1} and θ_{F_2} in the null hypothesis such that the kernel is degenerate under F_1 and non-degenerate under F_2, that is $0 = \mathrm{Var}_{F_1}(h_{1,i}(X_1)) < \mathrm{Var}_{F_2}(h_{1,i}(X_1))$. In the latter case the statistic U_i is asymptotically normal distributed with mean θ_i and variance $m^2\zeta_{1,i}/n$. Therefore, it is reasonable to standardize the differences $U_i^2 - \Delta^2$ appropriately before taking the maximum. We propose to use the test statistic

$$\mathcal{T}_{n,\Delta} := \max_{1 \le i \le d} \frac{U_i^2 - \Delta^2}{2\,\hat{\sigma}_i\,\Delta} \tag{2.9}$$

for testing the hypotheses in (2.5), where

$$\hat{\sigma}_i^2 := \frac{m^2(n-1)}{n(n-m)^2} \sum_{k=1}^n (q_{k,i} - U_i)^2 \tag{2.10}$$

is a Jackknife based estimator of the variance of U_i and $q_{k,i}$ is defined by

$$q_{k,i} := \binom{n-1}{m-1}^{-1} \sum_{1 \le l_1 < .. < l_{m-1} \le n, l_j \ne k} h_i(X_k, X_{l_1}, \ldots, X_{l_{m-1}})$$

(see Zhou et al., 2019, for details). The null hypothesis in (2.5) is rejected, whenever

$$\mathcal{T}_{n,\Delta} > \frac{q_{1-\alpha}}{a_d} + b_d\,, \tag{2.11}$$

where $q_{1-\alpha} = -\log(\log(\frac{1}{1-\alpha}))$ is the $(1-\alpha)$-quantile of the standard Gumbel distribution with distribution function $\exp(-\exp(-x)), x \in \mathbb{R}$, and

$$a_d = \sqrt{2 \log d} \quad \text{and} \quad b_d = a_d - \frac{\log(\log d) + \log(4\pi)}{2a_d}.$$

In the following discussion we will show that this test has asymptotic level α. An important step in these arguments is a proof of the weak convergence

$$\lim_{n\to\infty} \mathbb{P}\big(a_d(\mathcal{T}_{n,\Delta} - b_d) \le x\big) = \exp(-\exp(-x))\,, \qquad x \in \mathbb{R}\,, \tag{2.12}$$

in the case $|\theta_1| = |\theta_2| = \ldots = |\theta_d| = \Delta > 0$. Note that this choice corresponds to the most extreme case in the null hypothesis (1.2) which means that the probabilities of rejection by the test (2.11) for all other parameter constellations under the null hypothesis $H_0 : \max_{i=1}^d |\theta_i| \le \Delta$ are bounded by this scenario and in many cases substantially smaller.
Under additional assumptions on the kernels h_i we can also prove that the test (2.11) is minimax optimal, see Section 3.5 for a discussion of this property in the context of dependence measures. Interestingly, it turns out that for deriving these properties it is not necessary to assume that the kernels h_i are non-degenerate for all distributions F corresponding to the null hypothesis (see the discussion below, in particular Assumption (A2)).
In what follows, we will need the function $\psi_\beta(x) = \exp(x^\beta) - 1$ and the corresponding Orlicz norm

$$\|Z\|_{\psi_\beta} := \inf\{\nu > 0 : \mathbb{E}[\psi_\beta(|Z|/\nu)] \le 1\} \tag{2.13}$$

of a real-valued random variable Z. We continue by spelling out several regularity assumptions that are required for proving the weak convergence in (2.12).

(A1) For some constant $\beta \in (0,2]$ there exist a non-negative sequence $(B_n)_{n\in\mathbb{N}}$ and a constant $D > 0$ such that for all $d = d(n), n \in \mathbb{N}$,

$$\max_{1\le i\le d} \|h_i(X_1,\ldots,X_m) - \theta_i\|_{\psi_\beta} \le B_n\,,$$

$$\max_{1\le i\le d} \zeta_{1,i} \le D\,,$$

$$\max_{1\le i\le d} \mathbb{E}_F[(h_{1,i}(X_1) - \theta_i)^4] \le DB_n^2\,.$$

(A2) There exist constants $\underline{b} > 0$ and $c \in (0,\Delta)$ such that

$$\min_{1\le i\le d, |\theta_i|>c} \zeta_{1,i} > \underline{b}$$

for all $d = d(n), n \in \mathbb{N}$.

(A3) Let $\kappa_{i,j} = \mathrm{Corr}_F(h_{1,i}(X_1), h_{1,j}(X_1)) \in (-1,1)$ denote the correlation between $h_{1,i}(X_1)$ and $h_{1,j}(X_1)$. There exist a constant $\epsilon > 0$ and a sequence $\gamma_n = o(1)$ such that for all $d = d(n)$, $n \in \mathbb{N}$

$$\sum_{1\le i\ne j\le d} \frac{|\kappa_{i,j}|}{\sqrt{1-\kappa_{i,j}^2}} \exp\left(-\frac{(2-\epsilon)\log d}{1+|\kappa_{i,j}|}\right) \le \gamma_n\,.$$

Assumption $(A1)$ is a technical condition that captures a uniform tail probability decay from which we will deduce concentration inequalities for the the components of the U-statistic defined in (2.3). Note that this condition is always satisfied if the kernel h is bounded. Assumption $(A2)$ is a uniform non-degeneracy requirement which is a standard condition for deriving Gaussian approximation results, see for instance Chen (2018); Chernozhukov et al. (2019) among others. We emphasize that this assumption is only required here for the parameters θ_i which are (uniformly) bounded away from 0. This covers most cases of practical interest, where a degenerate kernel appears in the case $\theta_i = 0$, but the kernel is non-degenerate, whenever $\theta_i \ne 0$. Roughly speaking, for the problem of testing composite hypotheses of the form (2.5) the distinction between the degenerate and non-degenerate case is basically not

necessary if Assumption (A2) is satisfied (see Section 3.4 for a more detailed discussion in the context of dependence measures). Finally, Assumption $(A3)$ ensures that we can approximate the maximum of dependent normal distributed random variables by the maximum of independent ones. We already emphasize at this point that this assumption will not be required for the bootstrap test, which will be developed in Section 2.2 later on.
Our first result shows that the test defined in (2.11) has asymptotic level α (uniformly over a given class of distributions). For a precise statement consider the set of all distribution functions on $\mathbb{R}^d$ satisfying Assumptions (A1) - (A3), and define

$$V_0 := \left\{ z = (z_1, \ldots, z_d)^\top \in \mathbb{R}^d \;\middle|\; \max_{1 \le i \le d} |z_i| \le \Delta \right\} \tag{2.14}$$

as the parameter space corresponding to the null hypothesis in (2.5). Note that these sets depend on n (through the dimension $d = d(n)$). We define

$$\mathcal{H}_0(\Delta) := \big\{ F \in \mathcal{F} \mid \theta_F \in V_0 \;,\; F \text{ satisfies Assumptions (A1), (A2), (A3)} \big\} \tag{2.15}$$

as the set of distribution functions satisfying the null hypothesis (and the basic assumptions) with existing expectation $\mathbb{E}_F[U]$. Note that $\mathcal{H}_0(\Delta)$ depends on the constants $\underline{b}$, D and on n (through the dimension $d = d(n)$ and sequence $(B_n)_{n \in \mathbb{N}}$) which is not reflected in our notation.

THEOREM 2.2. *If Assumptions (A1), (A2), (A3) are satisfied,* $\log d = o(n^\gamma)$ *with* $0 \le \gamma \le \frac{1}{2/\beta + 1}$ *and*

$$\frac{B_n^2 \big(\log(nd)\big)^{4+2/\beta}}{n} = o(1) \;, \qquad n \to \infty \,, \tag{2.16}$$

then for any $\alpha \in (0,1)$

$$\limsup_{n \to \infty} \sup_{F \in \mathcal{H}_0(\Delta)} \mathbb{P}\Big(\mathcal{T}_{n,\Delta} > \frac{q_{1-\alpha}}{a_d} + b_d \Big) \le \alpha \,. \tag{2.17}$$

REMARK 2.3.

(1) For the proof of Theorem 2.2 we proceed in two steps: first we use Gaussian approximation techniques (see Chen, 2018; Chernozhukov et al., 2019, for example) and then compare the resulting Gaussian vector with a Gaussian vector with i.i.d components under the additional assumption (A3) on the dependence structure of the vector X_1. The maximum of the latter Gaussian vector then converges to a Gumbel distribution under suitable assumptions on the dependence structure, sample size and dimension.

(2) Note that the statement (2.17) addresses the worst case under the null hypotheses $H_0 : \max_{i=1}^d |\theta_i| \le \Delta$ (uniformly over the class of distributions defined by (2.15)), and that there exist also vectors θ_F such that equality holds in (2.17). For example, it follows from the proofs in Section A that equality holds in the case, where $\theta_i = \Delta$ for all $1 \le i \le d$. Moreover, an inspection of the arguments given in the proof of Theorem 2.2 further shows that

$$\limsup_{n \to \infty} \sup_{F \in \mathcal{H}_0(\Delta)} \mathbb{P}\Big(\mathcal{T}_{n,\Delta} > \frac{q_{1-\alpha}}{a_d} + b_d \Big) < \alpha,$$

whenever the proportion of indices $i \in \{1, \ldots, d\}$ with $\theta_i = \Delta$ is asymptotically strictly smaller than 1. In particular, we have

$$\lim_{n \to \infty} \sup_{F \in \mathcal{H}_0(\Delta)} \mathbb{P}\Big(\mathcal{T}_{n,\Delta} > \frac{q_{1-\alpha}}{a_d} + b_d \Big) = 0 \;,$$

whenever $\sup_{d\in\mathbb{N}} \max_{i=1}^{d} |\theta_i| < \Delta$. Thus for many parameter constellations in the null hypothesis the type I error of the test (2.11) will be much smaller than α, which is an appealing property of the test.

Next we turn to the consistency of the test (2.11) and define

$$V(c) = \Big\{ z \in \mathbb{R}^d \;\Big|\; \max_{1\le i\le d} |z_i| \ge \Delta + cB_n((\log d)/n)^{1/2} \Big\}$$

as a set of alternatives (note that for a bounded kernel h the sequence B_n can be chosen as a constant sequence). We will study the power of the test (2.11) against alternatives in the set

$$\mathcal{H}_1(c) = \Big\{ F \in \mathcal{F} \mid \theta_F \in V(c)\,;\; F \text{ satisfies Assumption (A1)} \Big\}\,. \tag{2.18}$$

THEOREM 2.4. *If $\log d = o(n^{\gamma})$ with $0 \le \gamma \le \frac{1}{2/\beta+1}$, then there exists a constant $c > 0$, only depending on γ and β, such that*

$$\lim_{n\to\infty} \inf_{F\in\mathcal{H}_1(c)} \mathbb{P}\Big(\mathcal{T}_{n,\Delta} > \frac{q_{1-\alpha}}{a_d} + b_d \Big) = 1\,.$$

The choice of the sequence $(B_n)_{n\in\mathbb{N}}$ depends on the tail behavior of the random variables $h_i(X_1,\dots,X_m)$ and $h_{1,i}(X_k)$ and the condition (2.16) puts a further restriction on the growth rate of the dimension. For example, if the sequence $(B_n)_{n\in\mathbb{N}}$ is bounded, Theorem 2.2 is applicable with an exponentially growing dimension d, i.e. $\log d = o(n^{1/(4+2/\beta)})$ which results in the rate $\log d = o(n^{1/5})$ if $h_1(X_1,\dots,X_m),\dots,h_d(X_1,\dots,X_m)$ are sub-Gaussian random variables. Note that this property implies that the random variables $h_{1,1}(X_1),\dots,h_{1,d}(X_1)$ are Sub-Gaussian as well. Under additional assumptions on the kernel h it can also be proved that the rate $\sqrt{\log(d)/n}$ in Theorem 2.4 is in fact minimax optimal and cannot be improved by other tests. We discuss this optimality property in the context of bivariate dependence measures in Section 3.5.

2.2. *Bootstrap.* The use of the asymptotic quantiles in the decision rule (2.11) is attractive from a computational point of view. On the other hand the basic statement of weak convergence (2.12) used to establish its validity requires additional assumptions regarding the dependence structure of the components of the random vectors X_i as formulated in Assumption (A3). Moreover, it is well–known that the rate of convergence in results of this type is typically rather slow and the nominal level of the test (2.11) will not be well approximated.
In this section we discuss a bootstrap approach to solve these problems. As usual in applications of the bootstrap in testing hypotheses this requires simulating the distribution of the statistic $\mathcal{T}_{n,\Delta}$ in (2.9) under an appropriate configuration of the null hypothesis $H_0 : \max_{1\le i\le d} |\theta_i| \le \Delta$. While this task is relatively easy in the case of the "classical" null hypothesis corresponding to the case $\Delta = 0$ it is significantly more difficult for the composite hypotheses corresponding to $\Delta > 0$ as considered in this paper. The approach proposed here is based on bootstrap data generated at the "boundary" of the hypotheses in (2.5), that is $\max_{1\le i\le d} |\theta_i| = \Delta$.
To be precise, let $X_1^*,\dots,X_n^*$ be drawn with replacement from $X_1,\dots,X_n$ and define for $i = 1,\dots,d$ by

$$U_i^* = \binom{n}{m}^{-1} \sum_{1\le l_1<\ldots<l_m\le n} h_i(X_{l_1}^*,\dots,X_{l_m}^*) \tag{2.19}$$

a bootstrap analogue of the statistic introduced in (2.6). Note that the conditional expectation of U_i^* given $X_1, \dots, X_n$ is given by the V-statistic

$$V_i = \mathbb{E}_F\big[U_i^* | X_1, \dots, X_n\big] = \frac{1}{n^m} \sum_{l_1,\dots,l_m=1}^{n} h_i(X_{l_1}, \dots, X_{l_m}) \tag{2.20}$$

(see, for example, Chen, 2018). Next we define a truncated version of V_i, that is

$$V_{i,\Delta} = \begin{cases} V_i\,, & \text{if } |V_i| \le \Delta \\ \Delta\,, & \text{otherwise} \end{cases} \qquad i = 1, \dots, d \tag{2.21}$$

and note that $\big|\mathbb{E}[U_i^* - V_i + V_{i,\Delta} | X_1, \dots, X_n]\big| \le \Delta$ a.s. We finally define

$$\mathcal{T}_n^* = \max_{1 \le i \le d} \frac{(U_i^* - V_i + V_{i,\Delta})^2 - V_{i,\Delta}^2}{2\,\hat{\sigma}_i\,\Delta} \tag{2.22}$$

as the bootstrap analogue of the statistic $\mathcal{T}_{n,\Delta}$ defined in (2.9) and denote by $q_{1-\alpha}^*$ the $(1-\alpha)$-quantile of the distribution of $\mathcal{T}_n^*$. We propose to reject the null hypothesis in (2.5), whenever

$$\mathcal{T}_{n,\Delta} > q_{1-\alpha}^*\,. \tag{2.23}$$

The next result shows that this procedure defines a (uniformly) consistent and asymptotic level α test for the hypotheses (2.5). We emphasize that we do not require Assumption (A3) for this statement and that in this sense the bootstrap test is valid under more general assumptions than the asymptotic test (2.11). This comes at the cost of a slight loss of sensitivity as the bootstrap data might have a larger conditional ψ_β-Orlicz norm than the original data. Additionally, we need some conditions on the entries $h_i(X_1, \dots, X_1, X_{m-k}, \dots, X_m)$ for all $1 \le k \le m$, which are known as von-Mises conditions in the literature (see Bickel and Freedman, 1981, for example). More precisely, we make the following assumption.

(A1') Let Assumption (A1) hold and assume that the constant $\beta \in (0, 2]$ and the sequence $(B_n)_{n\in\mathbb{N}}$ satisfy additionally

$$\max_{1\le i\le d} \|h_i(X_{j_1}, \dots, X_{j_m})\|_{\psi_\beta} \le B_n\,,$$

for all $j_1, \dots, j_m \in \{1, \dots, n\}$.

THEOREM 2.5. *Let Assumptions (A1') and (A2) be satisfied, assume that* $\log d = o(n^\gamma)$ *with* $0 \le \gamma \le \frac{1}{2/\beta+1}$ *and that*

$$\frac{B_n^2(\log(nd))^{5+2/\beta}}{n} + \frac{B_n^3(\log(nd))^{1+2/\beta}}{\sqrt{n}} = o(1)\,, \qquad n \to \infty\,. \tag{2.24}$$

(1) *For any* $\alpha \in (0, 1)$ *it follows that*

$$\limsup_{n\to\infty} \sup_{F \in \mathcal{H}_{0,boot}(\Delta)} \mathbb{P}\Big(\mathcal{T}_{n,\Delta} > q_{1-\alpha}^*\Big) \le \alpha\,, \tag{2.25}$$

where

$$\mathcal{H}_{0,boot}(\Delta) := \big\{F \in \mathcal{F} \mid \theta_F \in V_0;\ F \text{ satisfies Assumptions (A1'), (A2)}\big\} \tag{2.26}$$

and V_0 *is defined in* (2.14).

(2) *For a sufficiently large constant* c*, which only depends on* γ *and* β*, it follows that*

$$\lim_{n\to\infty} \inf_{F\in\mathcal{H}_1(c(\log(nd))^{1/\beta})} \mathbb{P}\Big(\mathcal{T}_{n,\Delta} > q^*_{1-\alpha}\Big) = 1 \,, \tag{2.27}$$

where the set $\mathcal{H}_1(c)$ *is defined in* (2.18)*. Moreover, if the kernel* h *in* (2.3) *is bounded, then the set* $\mathcal{H}_1(c(\log(nd))^{1/\beta})$ *in* (2.27) *can be replaced by* $\mathcal{H}_1(c)$*.*

REMARK 2.6.

(1) Note that the sets $\mathcal{H}_0(\Delta)$ and $\mathcal{H}_{0,boot}(\Delta)$ defined in (2.15) and (2.26), respectively, satisfy $\mathcal{H}_0(\Delta) \subset \mathcal{H}_{0,boot}(\Delta)$. This means part (1) of Theorem 2.5 holds under weaker assumptions than Theorem 2.2.
(2) Comparing the statement (2.27) for the power of the bootstrap test (2.23) with Theorem 2.4 about the power of the asymptotic test (2.11), we observe that for unbounded kernels there is an additional factor $(\log(nd))^{1/\beta}$ in the definition of the set of alternatives $\mathcal{H}_1$. This factor is a consequence of an inflation in the tails of the conditional distribution of the bootstrap data for unbounded kernels. As a consequence the bootstrap test can detect local alternatives converging to the null at the rate $(\log(nd))^{1/\beta}\sqrt{(\log d)/n}$ and this rate improves to $\sqrt{(\log d)/n}$ in the case of bounded kernels.
(3) We emphasize that the test (2.23) has similar properties as described in Remark 2.3 for the test (2.11), which uses the quantiles of the Gumbel distribution. In particular, under the null hypothesis (1.2) the rejection probability is asymptotically α if $|\theta_i| = \Delta$ for all $1 \le i \le d$, and, by Theorem 2.5, this is an upper bound for the rejection probability under the null. Consequently, the type I error can be much smaller than α if $|\theta_i|$ is substantially smaller than Δ for many indices $1 \le i \le d$, where the extreme case appears if $\theta_i = 0$ for all $1 \le i \le d$.
(4) Under additional assumptions on the kernel h it can also be proved that the test (2.23) is optimal in the sense that no other test can detect alternatives converging with a faster rate than $B_n\sqrt{\log(d)/n}$ to the null hypotheses. We give more details and illustrate this property in Section 3.5 for the bivariate dependence measures considered in Example 2.1.
(5) Naive algorithms for calculating higher order U statistics result in prohibitive run times of order n^m already when considering the case $d = 2$. Fortunately there are software packages providing optimized algorithms that calculate rank based U-Statistics in time $n\log(n)$, see for instance the R package "independence" from Even-Zohar (2020). Similar techniques can be used to shorten the computation times of the quantities V_i and $\hat{\sigma}_i$ for rank based statistics.

REMARK 2.7. As mentioned in the introduction the theory can be extended for testing the interchanged hypotheses

$$H_0^{\text{int}} : \max_{i=1}^{d} |\theta_i| > \Delta \quad \text{versus} \quad H_1^{\text{int}} : \max_{i=1}^{d} |\theta_i| \le \Delta \tag{2.28}$$

For the sake of brevity we restrict ourselves to a bootstrap test, which rejects the null hypothesis in (2.28), whenever

$$\mathcal{T}_{n,\Delta} \le q^{**}_{\alpha} \,, \tag{2.29}$$

where the statistic $\mathcal{T}_{n,\Delta}$ is defined in (2.9) and the bootstrap quantile q^{**}_{α} is obtained as follows. Let $X_1^*, ..., X_n^*$ be drawn with replacement from $X_1, .., X_n$, recall the definitions (2.19) and (2.20) and replace the definition of $V_{i,\Delta}$ in (2.21), by

$$V_{i,\Delta} = \begin{cases} V_i \,, & \text{if } |V_i| > \Delta \\ \Delta \,, & \text{otherwise} \end{cases} \qquad i = 1, \dots, d \,.$$

Then we define the statistic

$$\mathcal{T}_n^{*,\text{int}} := \max_{1\le i\le d} \frac{(U_i^* - V_i + V_{i,\Delta})^2 - V_{i,\Delta}^2}{2\hat{\sigma}_i\Delta}$$

and denote by q_α^{**} its corresponding α-quantile. Using similar arguments as given in the proof of Theorem 2.5 we can show that the decision rule (2.29) defines a (uniformly) consistent and asymptotic level α test for the hypotheses (2.28).

REMARK 2.8 (Alternative tests).
(a) A careful inspection of the proofs in the online supplement shows that it is possible to construct a bootstrap procedure without normalizing the variance of each component. To be precise we consider the test statistic

$$\mathcal{T}_{n,\Delta}^{\text{nv}} := \sqrt{n} \max_{1\le i\le d} U_i^2 - \Delta^2 \ , \tag{2.30}$$

which is obtained from (2.9) by omitting the normalizing factors $\hat{\sigma}_i$. This statistic does not converge weakly to a Gumbel distribution. However, a Gaussian approximation and corresponding construction of a bootstrap procedure is still possible.
For this purpose let $X_1^*, \dots, X_n^*$ be drawn with replacement from $X_1, \dots, X_n$ and define U_i^*, V_i and $V_{i,\Delta}$ by (2.19), (2.20) and (2.21), respectively. We then obtain a bootstrap analogue of the statistic (2.30) by

$$\mathcal{T}_n^{*,\text{nv}} := \sqrt{n} \max_{1\le i\le d} \{(U_i^* - V_i + V_{i,\Delta})^2 - V_{i,\Delta}^2\}$$

and denote by $q_{1-\alpha}^{*,\text{nv}}$ the corresponding $(1-\alpha)$-quantile. The null hypothesis in (2.5) is rejected, whenever

$$\mathcal{T}_n^{*,\text{nv}} > q_{1-\alpha}^{*,\text{nv}} . \tag{2.31}$$

For this test an analogue of Theorem 2.5 can be proved which even allows us to relax condition (2.24) slightly as we do not need to take into account errors that are incurred by approximating the variances anymore. However condition (A2) is still required as the result still relies crucially on Gaussian approximations. The details are omitted for the sake of brevity.

(b) The consideration of the squared U-statistics in (2.9) and (2.30) was made for technical reasons. In fact, similar results can be shown, if U_i^2 is replaced by $|U_i|$ $(i = 1, \dots, d)$. To be precise note that an essential step in the proof of Theorem 2.2 - 2.5 is the decomposition (2.7). Under the null hypothesis and the alternative the properties of the tests are determined by the parameter vectors, for which all components do not vanish, that is $\theta_i \neq 0$ for all $i = 1, \dots, d$. Whenever $\theta_i \neq 0$, the quadratic term in (2.7) is negligible and the linear term is dominating, which is analyzed using Gaussian approximation techniques (see Appendix A for details). Now, if we consider the non-normalized case and define the test statistic by

$$\mathcal{T}_{n,\Delta}^{\text{abs}} := \sqrt{n} \max_{1\le i\le d} |U_i| - \Delta \ ,$$

one observes a property similar to (2.7) that facilitates the application of a Gaussian approximation. More precisely, whenever $\theta_i \neq 0$, we have

$$|U_i| - \Delta = \text{sign}(U_i)U_i - \Delta = \text{sign}(\theta_i)(U_i - \theta_i) + (\text{sign}(\theta_i)\theta_i - \Delta) \tag{2.32}$$

with high probability. Therefore a valid bootstrap procedure is obtained as follows. Let $X_1^*, \dots, X_n^*$ be drawn with replacement from $X_1, \dots, X_n$ and recall definitions (2.19), (2.20) and (2.21). We then define the bootstrap statistic as

$$\mathcal{T}_{n,\Delta}^{*,\text{abs}} := \sqrt{n} \max_{1\le i\le d} \{|U_i^* - V_i + V_{i,\Delta}| - |V_{i,\Delta}|\}$$

and denote by $q_{1-\alpha}^{*,\mathrm{abs}}$ its $(1-\alpha)$ quantile. The null hypothesis (2.5) is rejected, whenever

$$\mathcal{T}_n^{*,\mathrm{abs}} > q_{1-\alpha}^{*,\mathrm{abs}} . \tag{2.33}$$

For this test one can obtain an analog of Theorem 2.5 using similar arguments as in the quadratic case, where (2.7) is replaced by (2.32) to linearize the test statistic. The details are omitted for the sake of brevity. As for the test (2.31) the condition (2.24) can be slightly relaxed, because no approximation of the variances is required.

REMARK 2.9 (Classical hypotheses). With the choice $\Delta = 0$ the non-normalized bootstrap tests (2.31) and (2.33) can also be used for testing the classical hypotheses in (1.1), provided that the representation (2.2) for the parameter of interest holds with a U-statistic which is non-degenerate under the null hypothesis. This follows by a careful inspection of the arguments given in the proofs of Theorem 2.5 in the online supplement. In such cases these tests provide an interesting alternative to the tests constructed by asymptotic arguments, see for instance Han et al. (2017); Zhou et al. (2019) and Drton et al. (2020).

3. Relevant dependencies in high-dimension. In this section we apply the methodology in the context of bivariate dependence measures between the components of high-dimensional vectors as considered in the introduction. The relation between this problem and the general formulation in Section 2 is described in Example 2.1. Recall the definition of the dependence measure in (2.4) for the kernel $\tilde{h}$, the notation $X_k = (X_{k1}, \ldots, X_{kp})^\top$ and write

$$\begin{aligned} U_{ij} &= \binom{n}{m}^{-1} \sum_{1 \le l_1 < \ldots < l_m \le n} h_{ij}(X_{l_1}, \ldots, X_{l_m}) \\ &= \binom{n}{m}^{-1} \sum_{1 \le l_1 < \ldots < l_m \le n} \tilde{h}(X_{l_1 i}, X_{l_1 j}, \ldots, X_{l_m i}, X_{l_m j}) \end{aligned} \tag{3.1}$$

for the corresponding U-statistic, where the second equality defines the functions $h_{ij} : \mathbb{R}^{pm} \to \mathbb{R}$ in an obvious manner. We now discuss several dependence measures separately. For the sake of brevity we restrict ourselves to the bootstrap test introduced in Section 2.2, which is defined by

$$\mathcal{T}_{n,\Delta} > q_{1-\alpha}^* , \tag{3.2}$$

where $\mathcal{T}_{n,\Delta} = \max_{1 \le i < j \le p} (U_{ij}^2 - \Delta^2)/(2\hat{\sigma}_{ij}\Delta)$ and $q_{1-\alpha}^*$ denotes the $(1-\alpha)$-quantile of the corresponding bootstrap distribution.

3.1. *Covariance.* The sample covariance matrix

$$\big(\hat{\Sigma}_{ij}\big)_{i,j=1,\ldots,p} = \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \bar{X}_n)(X_k - \bar{X}_n)^\top ,$$

where $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$ denotes the sample mean of $X_1, \ldots, X_n$, is the commonly used unbiased estimate for the covariance matrix $\Sigma = \mathrm{Cov}_F(X_1) = \mathbb{E}_F\big[(X_1 - \mathbb{E}_F[X_1])(X_1 - \mathbb{E}_F[X_1])^\top\big]$.

The covariance is a special case of (3.1) choosing $h(x_1, x_2) = (x_1 - x_2)(x_1 - x_2)^\top/2$, and we refer to Bai et al. (2009); Chen and Kato (2017), among others, who considered independence testing of the classical hypotheses in (1.1) for covariances. We now consider the problem of testing the relevant hypotheses (1.2), where $d_{ij} = \mathrm{Cov}(X_{1i}, X_{1j})$, $1 \le i < j \le p$. For the problem of testing relevant hypotheses of the form (1.2) an application of the results of Section 2.2 yields the following result.

COROLLARY 3.1. *If $\log p = o(n^\gamma)$ with $0 \le \gamma \le (5+4/\beta)^{-1} \wedge (2+8/\beta))^{-1}$, then the bootstrap test (3.2) with $U_{ij} = \hat{\Sigma}_{ij}$ is (uniformly) consistent and has (uniform) asymptotic level α over the classes of distributions $\mathcal{H}_1(c(\log(nd))^{2/\beta})$ and $\mathcal{H}_{0,boot}(\Delta)$ defined in (2.18) and (2.26) respectively, where the conditions (A1') and (A2) have to be replaced (and are implied) by*

(C1) *There exist constants $\beta \in (0,2]$ and $C > 0$ such that for all $p = p(n), n \in \mathbb{N}$*

$$\max_{1\le i\le p} \|X_i - \mathbb{E}[X_i]\|_{\psi_\beta} \le C\,.$$

(C2) *For some constant $\underline{b} > 0$ and $c \in (0,\Delta)$ we have*

$$\min_{1\le i<j\le p, |\hat{\Sigma}_{ij}|>c} \mathit{Var}_F[(X_{1i} - \mathbb{E}_F[X_{1i}])(X_{1j} - \mathbb{E}_F[X_{1j}])] \ge \underline{b}$$

for all $p = p(n), n \in \mathbb{N}$.

Note that for a normal distribution Assumption (C2) holds whenever there exists a uniform positive lower bound for the diagonal elements of Σ.

3.2. *Kendall's τ.* A very popular measure of (monotonic) dependence between the ith and jth component of the vector $X_1 = (X_{11},\dots,X_{1p})^\top$ is Kendall's τ coefficient given by $\tau_{ij} = \mathbb{E}_F[\mathrm{sign}(X_{1i} - X_{2i})\mathrm{sign}(X_{1j} - X_{2j})]$ with empirical version

$$\hat{\tau}_{ij} = \frac{2}{n(n-1)} \sum_{1\le k<l\le n} \mathrm{sign}(X_{ki} - X_{li})\,\mathrm{sign}(X_{kj} - X_{lj})\,.$$

Here the kernel is given by

$$h_{ij}(x_1,x_2) = \tilde{h}(x_{1i},x_{1j},x_{2i},x_{2j}) = \mathrm{sign}(x_{1i} - x_{2i})\mathrm{sign}(x_{1j} - x_{2j})$$

and the vector U is defined by $U = \mathrm{vech}\big((\hat{\tau}_{ij})_{i,j=1,\dots,p}\big)$. The classical testing problem (1.1) with $d_{ij} = \mathbb{E}_F[\mathrm{sign}(X_{1i} - X_{2i})\mathrm{sign}(X_{1j} - X_{2j})]$ was considered by Han et al. (2017); Leung and Drton (2018); Zhou et al. (2019) and Li et al. (2021) in the high dimensional regime. For the problem of testing relevant hypotheses of the form (1.2) an application of the results of Section 2.2 yields the following result.

COROLLARY 3.2. *If $\log p = o(n^\gamma)$ holds with $0 \le \gamma \le \frac{1}{6}$, then the bootstrap test (3.2) with $U_{ij} = \hat{\tau}_{ij}$ is (uniformly) consistent and has (uniform) asymptotic level α over the classes of distributions $\mathcal{H}_1(c)$ and $\mathcal{H}_{0,boot}(\Delta)$ defined in (2.18) and (2.26) respectively, where condition (A1') can be omitted (because the kernel is bounded) and condition (A2) is replaced by*

(T1) *There exist constants $\underline{b} > 0$ and $c \in (0,\Delta)$ such that*

$$\min_{1\le i<j\le p, |\tau_{ij}|>c} \mathit{Var}_F\left[\mathbb{E}_F[\mathit{sign}(X_{1i} - X_{2i})\mathit{sign}(X_{1j} - X_{2j})|X_1]\right] \ge \underline{b}\,.$$

for all $p = p(n), n \in \mathbb{N}$.

3.3. *The dominating term of Spearman's ρ.* Let Q^i_{nk} be the rank of X_{ki} among $X_{1i},\dots,X_{ni}$ and consider Spearman's rank correlation coefficient

$$\rho_{ij} = \frac{\sum_{k=1}^n (Q^i_{nk} - (n+1)/2)(Q^j_{nk} - (n+1)/2)}{\sqrt{\sum_{k=1}^n (Q^i_{nk} - (n+1)/2)^2 \sum_{k=1}^n (Q^j_{nk} - (n+1)/2)^2}}\,,$$

which defines another popular measure of dependence between the ith and jth component of the vector $X_1 = (X_{11}, \ldots, X_{1p})^\top$. While ρ_{ij} ist not a U-statistic, it was shown by Hoeffding (1948a) that it can be decomposed as follows

$$\rho_{ij} = \frac{n-2}{n+1}\hat{\rho}_{ij} + \frac{3}{n+1}\hat{\tau}_{ij}\,,$$

where the dominating term

$$\hat{\rho}_{ij} = \frac{6}{n(n-1)(n-2)} \sum_{1 \le k_1 < k_2 < k_3 \le n} \mathrm{sign}(X_{k_1 i} - X_{k_2 i})\mathrm{sign}(X_{k_1 j} - X_{k_3 j})$$

is a U-statistic of degree 3 with bounded kernel

$$h_{ij}(x_1, x_2, x_3) = \tilde{h}(x_{1i}, x_{1j}, x_{2i}, x_{2j}, x_{3i}, x_{3j}) = \mathrm{sign}(x_{1i} - x_{2i})\mathrm{sign}(x_{1j} - x_{3j}).$$

The classical testing problem for this statistic and continuous data was considered by Han et al. (2017) and Leung and Drton (2018). For the problem of testing relevant hypotheses of the form (1.2) an application of the results of Section 2.2 yields the following result.

COROLLARY 3.3. *If* $\log p = o(n^\gamma)$ *holds with* $0 \le \gamma \le \frac{1}{6}$*, then the bootstrap test* (3.2) *with* $U_{ij} = \hat{\rho}_{ij}$ *is (uniformly) consistent and has (uniform) asymptotic level* α *over the classes of distributions* $\mathcal{H}_1(c)$ *and* $\mathcal{H}_{0,boot}(\Delta)$ *defined in* (2.18) *and* (2.26) *respectively, where condition (A1') can be omitted (because the kernel is bounded) and condition (A2) is replaced by*

(S1) *There exist constants* $\underline{b} > 0$ *and* $c \in (0, \Delta)$ *such that*

$$\min_{1 \le i < j \le p, |\hat{\rho}_{ij}| > c} \mathit{Var}_F\left[\mathbb{E}_F[\mathit{sign}(X_{1i} - X_{2i})\mathit{sign}(X_{1j} - X_{3j})|X_1]\right] \ge \underline{b}\,.$$

for all $p = p(n), n \in \mathbb{N}$.

3.4. *Dependence measures with degenerate kernel* . While Kendall's τ and Spearman's ρ only capture monotonic dependencies between two random variables there are a number of higher order U-statistics that are able to capture any form of dependency between two random vectors. Exemplary, we mention here Hoeffding's D (Hoeffding, 1948b), Blum-Kiefer-Rosenblatt's R (Blum et al., 1961) and Bergsma–Dassios–Yanagimoto's τ^* (Bergsma and Dassios, 2014). Note that in the case of independence (reflecting the classical null hypothesis in (1.1)) the kernels corresponding to these U-statistics are degenerate. On the other hand, if the components are dependent (which corresponds to the classical alternative), all three statistics are non-degenerate for a large class of distributions. In such cases the general theory developed in Section 3 is applicable as well. Before going into details we emphasize that similar results as presented below can be derived for other types of dependence measure which can be estimated by U-statistics with a degenerate kernel under independence such as the distance correlation introduced by Székely et al. (2007), see Theorem 4.1 in Edelmann et al. (2021).

To be precise we recall the definition of the U-statistics considered in Hoeffding (1948b); Blum et al. (1961); Bergsma and Dassios (2014). Let $z_1, \ldots, z_6$ be p-dimensional vectors of the form $z_i = (z_{i1}, \ldots, z_{ip})^\top$, define

$$\mathbb{1}^{k}_{j_1,j_2,j_3} := \mathbb{1}\{z_{j_1 k} \le z_{j_3 k}\} - \mathbb{1}\{z_{j_2 k} \le z_{j_3 k}\}\,,$$

$$\mathbb{1}^{j_3,j_4,k}_{j_1,j_2} := \mathbb{1}\{z_{j_1 k} < z_{j_3 k}\}\mathbb{1}\{z_{j_1 k} < z_{j_4 k}\}\mathbb{1}\{z_{j_2 k} < z_{j_3 k}\}\mathbb{1}\{z_{j_2 k} < z_{j_4 k}\}\,,$$

and consider the kernels

$$h_{ij}^D(z_1,...,z_5) := \frac{1}{16} \sum_{1\le j_1\ne ...\ne j_5\le 5} \mathbb{1}^i_{j_1,j_2,j_5}\mathbb{1}^i_{j_3,j_4,j_5}\mathbb{1}^j_{j_1,j_2,j_5}\mathbb{1}^j_{j_3,j_4,j_5}\,,$$

$$h_{ij}^R(z_1,...,z_6) := \frac{1}{32} \sum_{1\le j_1\ne ...\ne j_6\le 6} \mathbb{1}^i_{j_1,j_2,j_5}\mathbb{1}^i_{j_3,j_4,j_5}\mathbb{1}^j_{j_1,j_2,j_6}\mathbb{1}^j_{j_3,j_4,j_6}\,,$$

$$h_{ij}^{\tau^*}(z_1,...,z_4) := \frac{1}{16} \sum_{1\le j_1\ne ...\ne j_4\le 4} (\mathbb{1}^{j_2,j_4,i}_{j_1,j_3}+\mathbb{1}^{j_1,j_3,i}_{j_2,j_4}-\mathbb{1}^{j_2,j_3,i}_{j_1,j_4}-\mathbb{1}^{j_1,j_4,i}_{j_2,j_3})\,,$$

$$\times\,(\mathbb{1}^{j_2,j_4,j}_{j_1,j_3}+\mathbb{1}^{j_1,j_3,j}_{j_2,j_4}-\mathbb{1}^{j_2,j_3,j2}_{j_1,j_4}-\mathbb{1}^{j_1,j_4,j}_{j_2,j_3})\,.$$

Note that h_{ij}^D, h_{ij}^R and $h_{ij}^{\tau^*}$ define symmetric kernels of orders $5, 6$ and 4 respectively. The corresponding matrices of empirical dependence measures calculated from the sample $X_1,...,X_n \in \mathbb{R}^p$ are then given by

$$\hat{D} = (\hat{D}_{ij})_{1\le i<j\le p} = \Big(\binom{n}{5}^{-1} \sum_{1\le j_1<...<j_5\le n} h_{ij}^D(X_{j_1},\dots,X_{j_5})\Big)_{1\le i<j\le p},$$

$$\hat{R} = (\hat{R}_{ij})_{1\le i<j\le p} = \Big(\binom{n}{6}^{-1} \sum_{1\le j_1<...<j_6\le n} h_{ij}^R(X_{j_1},\dots,X_{j_6})\Big)_{1\le i<j\le p}.$$

$$\hat{\tau}^* = (\hat{\tau}^*_{ij})_{1\le i<j\le p} = \Big(\binom{n}{4}^{-1} \sum_{1\le j_1<...<j_4\le n} h_{ij}^{\tau^*}(X_{j_1},\dots,X_{j_4})\Big)_{1\le i<j\le p}.$$

The classical testing problem (1.1), where the dependence measure d_{ij} is either given by $D_{ij} = \mathbb{E}_F[h_{ij}^D(X_1,\dots,X_5)]$, $R_{ij} = \mathbb{E}_F[h_{ij}^R(X_1,\dots,X_6)]$ or $\tau^*_{ij} = \mathbb{E}_F[h_{ij}^{\tau^*}(X_1,\dots,X_4)]$ was considered by Drton et al. (2020) in the high dimensional regime. For the problem of testing relevant hypotheses of the form (1.2) an application of the results of Section 2.2 yields the following result.

COROLLARY 3.4. *If $\log p = o(n^\gamma)$ holds with $0 \le \gamma \le \frac{1}{6}$, then the bootstrap test (3.2) with U_{ij} given by either $\hat{D}_{ij}$, $\hat{R}_{ij}$ or $\hat{\tau}^*_{ij}$ is (uniformly) consistent and has (uniform) asymptotic level α over the classes of distributions $\mathcal{H}_1(c)$ and $\mathcal{H}_{0,boot}(\Delta)$ defined in (2.18) and (2.26) respectively, where condition (A1') can be omitted (because the kernels are bounded) and condition (A2) is replaced by*

(D1) *There exist constants $\underline{b} > 0$ and $c \in (0,\Delta)$ such that*

$$\min_{1\le i<j\le p, |D_{ij}|>c} Var_F\left[\mathbb{E}_F[h_{ij}^D(X_1,...,X_5)|X_1]\right] \ge \underline{b}\,.$$

for all $p = p(n), n \in \mathbb{N}$.

in the case of Hoeffding's D, by

(R1) *There exist constants $\underline{b} > 0$ and $c \in (0,\Delta)$ such that*

$$\min_{1\le i<j\le p, |R_{ij}|>c} Var_F\left[\mathbb{E}_F[h_{ij}^R(X_1,...,X_6)|X_1]\right] \ge \underline{b}\,.$$

for all $p = p(n), n \in \mathbb{N}$.

in the case of Blum-Kiefer-Rosenblatt's R, and by

(TA1) *There exist constants* $\underline{b} > 0$ *and* $c \in (0, \Delta)$ *such that*

$$\min_{1 \le i < j \le p, |\tau^*_{ij}| > c} Var_F \left[\mathbb{E}_F [h^{\tau^*}_{ij}(X_1, ..., X_4) | X_1] \right] \ge \underline{b} \, .$$

for all $p = p(n), n \in \mathbb{N}$.

for Bergsma–Dassios–Yanagimoto's τ^*.

3.5. *Minimax optimality.* Recall that, by Theorems 2.4 and 2.5, both the asymptotic test and the bootstrap test (under the additional assumption of a bounded kernel) correctly reject the null hypothesis in (2.5) if at least one entry of the vector θ is larger than $\Delta + CB_n\sqrt{\log(d)/n}$. In this section we will show that in many situations, where the sequence $(B_n)_{n\in\mathbb{N}}$ is bounded this rate cannot be improved. These cases include all dependence measures discussed in Sections 3.1 – 3.4. To be precise, we define

$$\mathcal{T}_\alpha := \left\{ T_\alpha \mid \sup_{F \in \mathcal{H}_0(\Delta)} \mathbb{P}(T_\alpha \text{ does not reject } H_0) \le \alpha \right\}$$

as the set of all tests with (uniform) level α.

We begin with a result for the covariances, that is $d_{ij} = \mathrm{Cov}_F(X_{1i}, X_{1j})$ $(1 \le i < j \le p$. For the sake of simplicity, we assume without loss of generality that $d_{ii} = \mathrm{Var}(X_{1i}) = 1$ $(i = 1, \dots, p)$, the general case is obtained by a scaling argument. Note that in this case only values $\Delta \in (0, 1)$ are useful thresholds for the hypotheses (2.5). We then obtain the following result.

THEOREM 3.5. *Assume that the dependence measure* d_{ij} *in* (2.4) *is given by* $d_{ij} = \mathrm{Cov}_F(X_{1i}, X_{1j})$ *and* $d_{ii} = 1$ $(i, j = 1, \dots, p)$*; so we have* $d = p(p-1)/2$*. Further let* c_0, α, β *denote positive constants such that* $c_0 < 1 - \Delta$ *and* $\alpha + \beta < 1$*. If* $\log(p)/n \to 0$ *and* $\log(p)n/p^2 \to 0$*, as* $n \to \infty$*, then we have for sufficiently large* n *and* p

$$\inf_{T_\alpha \in \mathcal{T}_\alpha} \sup_{F \in \mathcal{H}_1(c_0)} \mathbb{P}(T_\alpha \textit{ does not reject } H_0) \ge 1 - \alpha - \beta \, . \tag{3.3}$$

The proof of (3.3) uses the fact that the supremum of the probabilities with respect to the distributions $F \in \mathcal{H}_1(c_0)$ can be bounded from below by the supremum taken over all centered multivariate normal distributions in $\mathcal{H}_1(c_0)$, where the covariance matrices have the following form. All diagonal elements are 1, except of two off-diagonal elements all off-diagonal elements are equal to Δ and the two remaining off-diagonal elements are given by $\Delta + \rho$. Because this argument does not depend on the specific dependence measure under consideration, a careful inspection of the proof of Theorem 3.5 shows that statements of the form (3.3) are also available for dependence measures, which, under the assumption of a normal distribution, can be represented as a function of the correlation. More precisely, let $d_{ij}(F) = d(X_{1i}, X_{1j})$ denote a bivariate dependence measure, such that

$$d_{ij}(N_1, N_2) = g(\rho) \tag{3.4}$$

for a normal distributed vector $(N_1, N_2)^\top \sim \mathcal{N}_2\big(0, \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)\big)$, where $g : (-1, 1) \to \mathbb{R}$ is a differentiable function with non-vanishing derivative at some $\rho \in g^{-1}(\{\Delta\})$.

COROLLARY 3.6. *The conclusion of Theorem 3.5 remains valid for any bivariate dependence measure* d_{ij}*, which satisfies* (3.4) *and for which there exists a constant* $a \in (-1, 1)$ *such that* $|g(a)| = \Delta$ *and* $\mathrm{sign}(g'(a)) = \mathrm{sign}(g(a))$.

REMARK 3.7. We conclude this section with some examples of dependence measures, where Corollary 3.6 is applicable. Note that Theorem 3.5 gives a lower bound for all tests. Thus it also applicable for dependence measures, which can be estimated by U-statistics.

(1) A prominent dependence measure that fulfills this assumption is Kendall's τ for which it holds that $\tau_{ij} = (2/\pi)\arcsin(\rho)$. A similar result holds for Spearman's ρ, here we have $\rho_{ij} = (6/\pi)\arcsin(\rho/2)$. Another obvious choice is the Pearson correlation for which $g(\rho) = \rho$ is the identity function.
(2) For a centered normal distribution Hoeffding's D, Blum-Kiefer-Rosenblatt's R and Bergsma–Dassios– Yanagimoto's τ^*s, which are considered in Section 3.4, can be expressed in terms of ρ, such that (3.4) holds. We expect that the assumptions of Corollary 3.6 are satisfied as well, but we do not work out the details here for the sake of brevity.

4. Finite sample properties. In this section we report the results of a small simulation study conducted in order to investigate the finite sample properties of the proposed tests for the relevant hypotheses (1.2). We focus on Kendall's τ and the bootstrap test (2.23), its non-normalized version defined by (2.31) and the test (2.33), which uses the statistics $|U_{ij}|$ instead of their squares U_{ij}^2.
As distributions we consider the centered p-dimensional normal distribution with covariance matrix Σ, that is

$$X_1, \dots, X_n \sim \mathcal{N}_p(0, \Sigma) \tag{4.1}$$

and the centered p-dimensional t-distribution with $f = 3$ degrees of freedom and scale matrix Σ, that is

$$X_1, \dots, X_n \sim t_f(0, \Sigma) \ , \tag{4.2}$$

with density

$$g_{f,\Sigma}(x) = \frac{\Gamma((f+p)/2)}{\Gamma(f/2) f^{p/2} \pi^{p/2} |\Sigma|^{1/2}} \Big(1 + \frac{1}{f} x^\top \Sigma x\Big)^{-(f+p)/2} .$$

We generate data from the models (4.1) and (4.2) for sample sizes $n \in \{50, 100\}$ and dimension $p \in \{100, 200, 400\}$, where we investigate 3 choices for the covariance matrices Σ and $(f/(f-2))\Sigma$ in (4.1) and (4.2) respectively, that is

$$\mathrm{Diag}_p(1-\rho, \dots, 1-\rho) + \rho J_p \ , \tag{$M1$}$$

$$\mathrm{Diag}_p(1, \dots, 1) + \rho \textstyle\sum_{1 \le i < j \le \lfloor p/\sqrt{2} \rfloor} (e_i e_j^\top + e_j e_i^\top) \ , \tag{$M2$}$$

$$\mathrm{Diag}_p(1, \dots, 1) + \rho (e_i e_j^\top + e_j e_i^\top) \ . \tag{$M3$}$$

Here $\mathrm{Diag}_p(a_1, \dots, a_p)$ denotes a diagonal $p \times p$ matrix with diagonal entries $a_1, \dots, a_p$, J_p denotes the $p \times p$ matrix with all entries equal to 1, e_j is the jth standard basis vector and ρ is a constant that varies depending on whether or not on one wants generate data whose Kendall's τ exceeds the threshold or not. In model ($M1$) we have equal correlation between all components of X_1, whereas in model ($M3$) only the ith and jth components of X_1 are correlated. Model ($M2$) defines an intermediate case with a block-diagonal correlation matrix, where the first $\lfloor p/\sqrt{2} \rfloor$ components have the same correlation and the remaining components are uncorrelated. All numerical results presented in the following discussion are based on 1000 simulation runs and 100 bootstrap replications.
We investigate the test for the hypothesis of a relevant deviation from independence between the components of a high-dimensional vector, if the dependencies are measured by Kendall's τ, as discussed in Section 3.2. Thus, the hypotheses are given by

$$H_0: \max_{1 \le i < j \le p} |\tau_{ij}| \le \Delta \quad \text{versus} \quad H_1: \max_{1 \le i < j \le p} |\tau_{ij}| > \Delta \ , \tag{4.3}$$

where we choose the threshold $\Delta = 0.1$. Note that the distributions in (4.1) and (4.2) are elliptical, which implies the relation

$$\tau_{ij} = \frac{2}{\pi} \arcsin\big(\mathrm{Corr}(X_{1i,}, X_{1j})\big)$$

between Kendall's τ and the off-diagonal elements of the matrices Σ and $(f/(f-2))\Sigma$ in (4.1) and (4.2) respectively (see Lindskog et al., 2003).

4.1. *Test statistics involving* U_{ij}^2. We begin studying the type I error of the bootstrap test (2.23), which is based on a maximum of normalized statistics involving squares of the U-statistics U_{ij}. As pointed out in Sections 2 and 3, the (asymptotic) level of the bootstrap test is substantially smaller than the nominal level α if $\max_{1\le i<j\le p} |\tau_{ij}| < 0.1$ (we emphasize again that this is a very desirable property). Therefore, we concentrate on the case where at least one of the bivariate dependence measures satisfies $|\tau_{ij}| = 0.1$, which corresponds to the choice $\rho = \sin(\pi/20)$ in model ($M1$) - ($M3$). Note that the matrix in ($M1$) represents the situation, where $|\tau_{ij}| = \Delta = 0.1$ for all $1 \le i < j \le p$, which corresponds to the "full boundary" of the hypotheses (4.3). The matrix in ($M3$) represents a case which is closer to the "interior" of the null hypothesis (only two off-diagonal elements have a Kendall's τ equal to 0.1, but for all other entries Kendall's τ is equal to 0). For the matrix ($M2$) about 50% of the off-diagonal elements have a Kendall's τ equal to 0.1. Therefore, from the discussion in Sections 2 and 3, we expect that for model ($M1$) the simulated level should be close to 0.1, while it should be substantially smaller than 0.1 in the two other cases. Moreover, this effect should be more visible for model ($M3$) than for ($M2$).

(n,p)	$(50,100)$	$(50,200)$	$(50,400)$	$(100,100)$	$(100,200)$	$(100,400)$
($M1$)	0.081	0.054	0.043	0.164	0.103	0.112
($M2$)	0.026	0.011	0.009	0.100	0.078	0.067
($M3$)	0.000	0.000	0.000	0.000	0.000	0.000
($M1$)	0.115	0.192	0.152	0.160	0.275	0.337
($M2$)	0.030	0.037	0.035	0.127	0.124	0.123
($M3$)	0.000	0.000	0.000	0.000	0.000	0.000

TABLE 1

Simulated rejection probabilities of the test (2.23) under the null hypothesis in (4.3) (nominal level $\alpha = 0.1$). Upper part: multivariate normal distribution; lower part: multivariate t-distribution with 3 degrees of freedom.

The corresponding rejection probabilities under the null hypothesis of the test (2.23) are shown in Table 1 and confirm the asymptotic theory. For normal data we observe that the test keeps its nominal level $\alpha = 0.1$ in almost all cases under consideration. More precisely, for the matrix Σ in ($M1$) the simulated level is close to the nominal level $\alpha = 0.1$ for all pairs (n,p) except in the case $(n,p) = (100,100)$. For the matrix in ($M3$) (only two off-diagonal elements have a Kendall's τ equal to 0.1, but for all other entries Kendall's τ is 0) the type I error is approximately 0. On the other hand, for the matrix ($M2$) (about 50% of the off-diagonal elements have a Kendall's τ equal to 0.1, but for all other entries Kendall's τ is 0) the type I error is larger than for the matrix ($M3$) but still smaller than the nominal level $\alpha = 0.1$. We note that these properties are desirable for composite hypotheses of the form (4.3). For some parameter constellations under the null hypothesis the rejection probabilities of the test have to approximate the nominal level $\alpha = 0.1$. This reflects the "worst case" under the null hypothesis (corresponding to the situation $\tau_{ij} = \Delta = 0.1$ for all $1 \le i < j \le p$). On the other hand most scenarios under the null-hypothesis yield a much smaller type I error. We also emphasise that the observed numerical results confirm the theoretical properties

(n, p)	$(50, 100)$	$(50, 200)$	$(50, 400)$	$(100, 100)$	$(100, 200)$	$(100, 400)$
$(M1)$	0.062	0.042	0.050	0.130	0.113	0.098
$(M2)$	0.016	0.007	0.009	0.074	0.045	0.034
$(M3)$	0.000	0.000	0.000	0.000	0.000	0.000
$(M1)$	0.102	0.084	0.075	0.154	0.180	0.149
$(M2)$	0.026	0.018	0.008	0.061	0.059	0.060
$(M3)$	0.000	0.000	0.000	0.000	0.000	0.000

TABLE 2

Simulated rejection probabilities of the test (2.31) under the null hypothesis in (4.3) (nominal level $\alpha = 0.1$). Upper part: multivariate normal distribution; lower part: multivariate t-distribution with 3 degrees of freedom.

mentioned in Section 2 and 3. For data generated from the multivariate t distribution the picture is slightly different. In model $(M2)$ and $(M3)$ the test keeps its nominal level in most cases. This is also the case in scenario $(M1)$ for dimension $p = 100$. On the other hand, for dimension $p = 200, 400$ the test significantly exceeds the desired significance level in model $(M1)$. The deviations become smaller if one is testing the hypotheses (1.2) with a smaller threshold than $\Delta = 0.1$ such as $\Delta = 0.05$ (these results are not displayed for the sake of brevity). A potential explanation of the observed exceedance in these cases is that the normalization by the variance estimators (2.10) may yield to some instabilities for more heavy tailed data.

Therefore, we next investigate the approximation of the nominal level by the non-normalized version of the test (2.5), which is defined in equation (2.31) in Remark 2.8. The corresponding empirical type I error rates are displayed in Table 2. Compared to the test (2.5) we observe an improvement of the approximation of the nominal level in scenario $(M1)$. While this is satisfactory in the case of a normal distribution, the rejection probabilities are still a little to large for t-distributed data if the sample sizes is $n = 100$ (again the deviations become smaller if the threshold $\Delta = 0.05$ is used in the hypotheses (1.2)). However, the test (2.31) keeps the nominal level well for the two other models $(M2)$ and $(M3)$ and all combination of n and p.

The power curves of the tests (2.23) and (2.31) are displayed in Figures 1 and 2, respectively, where we show the rejection probabilities of the test (2.23) as a function of Kendall's $\tau = \frac{2}{\pi} \arcsin \rho$ for sample size and dimension given by $(n, p) = (50, 100)$ and $(n, p) = (100, 100)$. The results reflect our theoretical findings. The rejection rates increase with the distance to the null-hypothesis and the sample size for all three covariance structures. Moreover, the largest power is obtained for the covariance matrix $(M1)$ followed by $(M2)$ and $(M3)$. A comparison of the upper and lower parts in the figures shows that the tests have lower power for t-distributed data. Comparing Figures 1 and 2 we observe that the results of the tests (2.23) and (2.31) under the alternative are comparable in most cases (with slight advantages of the test (2.23)). Only for model $(M3)$ with $(p, n) = (100, 100)$ we observe that the test (2.23) has a substantially larger power.

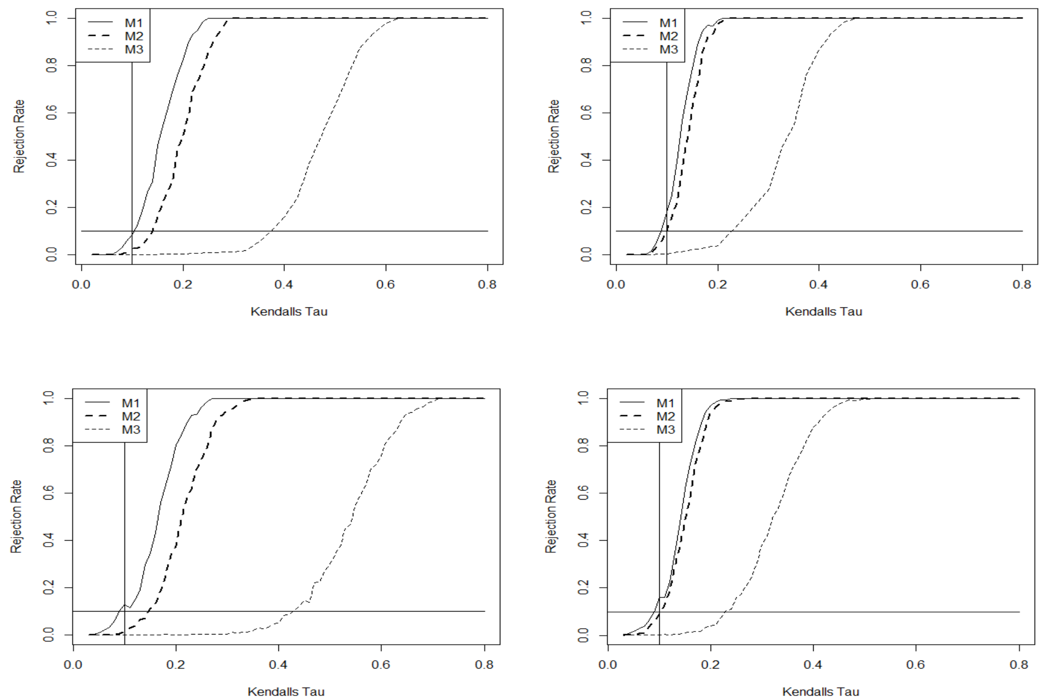

FIG 1. *Simulated rejection probabilities of the test* (2.23) *for the hypotheses* (1.2) *with* $\Delta = 0.1$*. The dimension is* $p = 100$*, and the sample sizes are* $n = 50$ *(left panels) and* $n = 100$ *(right panels). Upper part: normal distributed data; Lower part:* t_3*-distributed data.*

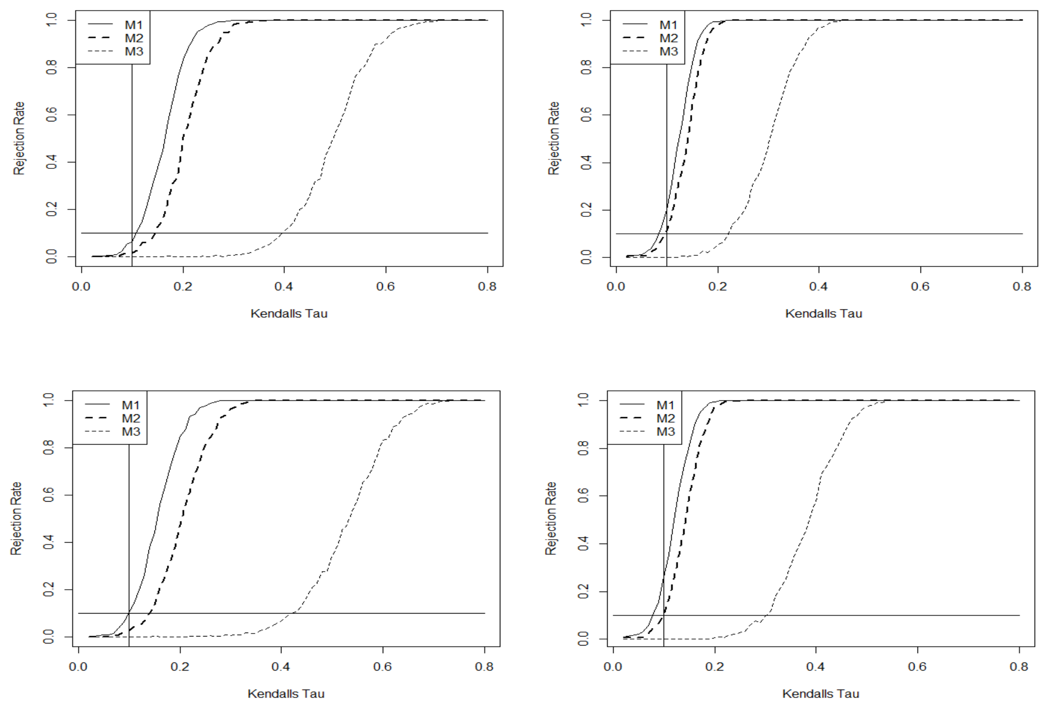

FIG 2. *Simulated rejection probabilities of the test* (2.31) *for the hypotheses* (1.2) *with* $\Delta = 0.1$. *The dimension is* $p = 100$, *and the sample sizes are* $n = 50$ *(left panels) and* $n = 100$ *(right panels). Upper part: normal distributed data; Lower part:* t_3*-distributed data.*

4.2. *Test statistics involving* $|U_{ij}|$. In this section we investigate the bootstrap test (2.33) that uses the absolute value $|U_{ij}|$ instead of U_{ij}^2 in the defintion of the test statistic (see Remark 2.8(b)). For the sake of comparison we consider the same scenarios as in Section 4.1 and study the properties of the test for the hypotheses (1.2) with $\Delta = 0.1$. The empirical type I error rates are shown in Table 3 and we observe that the test (2.33) keeps the nominal level in all cases under consideration (in particular also in the "worst case" scenario ($M1$), where all (pairwise) Kendall's taus satisfy $\tau_{ij} = 0.1$, and the data is heavy tailed). Again we observe in the two other scenarios ($M2$) and ($M3$) a smaller type I error rate than for the scenario ($M1$), which agrees with our theoretical findings in Section 2 and 3.

(n,p)	$(50,100)$	$(50,200)$	$(50,400)$	$(100,100)$	$(100,200)$	$(100,400)$
($M1$)	0.026	0.013	0.015	0.048	0.047	0.029
($M2$)	0.005	0.008	0.003	0.017	0.017	0.009
($M3$)	0.000	0.000	0.000	0.000	0.000	0.000
($M1$)	0.052	0.025	0.014	0.081	0.066	0.044
($M2$)	0.014	0.010	0.009	0.027	0.019	0.023
($M3$)	0.000	0.000	0.000	0.000	0.000	0.000

TABLE 3
Simulated rejection probabilities of the test (2.33) *under the null hypothesis in* (4.3) *(nominal level* $\alpha = 0.1$*). Upper part: multivariate normal distribution; lower part: multivariate* t*-distribution with* 3 *degrees of freedom.*

In Figure 3 we display the empirical rejection probabilities as a function of Kendall's $\tau = \frac{2}{\pi} \arcsin \rho$ where the sample size and dimension are given by $(n,p) = (50,100)$ and $(n,p) =$

$(100, 100)$. We consider again the covariance structures $(M1)$ - $(M3)$ and a multivariate normal and t_3-distribution.
Once again, the results are in line with our theoretical findings. The rejection rates increase with the distance to the null-hypothesis and the sample size for all three covariance structures. Moreover, the largest power is obtained for the covariance matrix $(M1)$ followed by $(M2)$ and $(M3)$. Comparing the upper and the lower parts we observe a loss in power for t-distributed data. It is also of interest to compare these results with the non-normalized test (2.31) in Figure 2. While the differences are small in the case $(p, n) = (50, 100)$, they are more visible for $(p, n) = (50, 100)$. In other words: the test (2.33) keeps the nominal level in all cases under consideration, but compared to the test (2.31) this advantage comes with the price of a slight loss in power in the case $(p, n) = (50, 100)$.

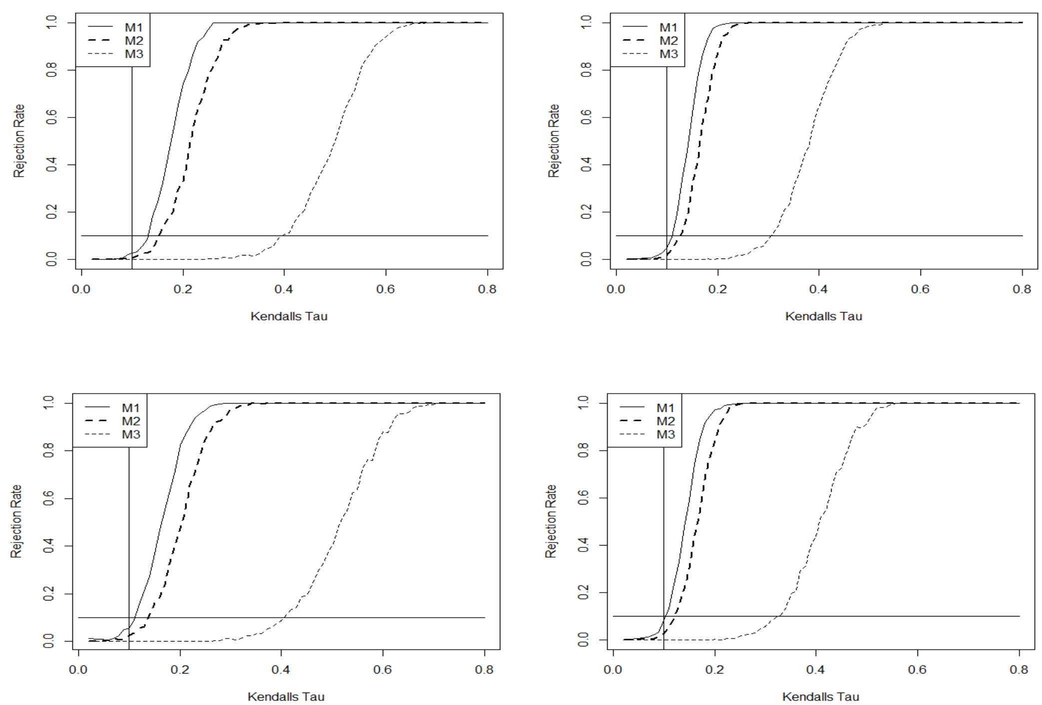

FIG 3. *Simulated rejection probabilities of the test* (2.33) *for the hypotheses* (1.2) *with* $\Delta = 0.1$. *The dimension is* $p = 100$, *and the sample sizes are* $n = 50$ *(left panels) and* $n = 100$ *(right panels). Upper part: normal distributed data; Lower part:* t_3*-distributed data.*

Acknowledgements. This work was partially supported by the DFG Research unit 5381 *Mathematical Statistics in the Information Age*, project number 460867398.

APPENDIX A: ONLINE SUPPLEMENT: PROOFS

In this section we provide proofs of our theoretical results. These are rather involved and we proceed in several steps. In Section A.1, we begin with the analysis of the variance estimators $\hat{\sigma}_i^2$ defined in (2.10). These results are used in the proofs of Theorems 2.2 and 2.4, which are provided in Section A.2. The proof of the consistency of the bootstrap test can be found in Section A.3. Several arguments given in this section rely on sophisticated technical results, which will be provided in Section B.

Notation: Throughout this section we use the symbol $a_n \lesssim b_n$ to denote $a_n \leq C\, b_n$ for some generic positive constant C not depending on n whose concrete value may change from line to line. We also introduce $1 - o_K(1)$ as a shorthand for any term of the form

$$1 - C_1/(nd) - C_2(\log(nd))^{1/2+1/\beta}/\sqrt{n} - C_3(\log(nd))^{\beta} n^{-\gamma/\beta},$$

where the non-negative constants C_1, C_2 and C_3 may only depend on γ and β. We remark that in many cases some of the factors in the summands will be 0.
Moreover $\|x\|_\infty = \max_{i=1}^d |x_i|$ denotes the maximum norm of a d-dimensional vector, where the dimension of x will always be clear from the context. We also note that many bounds could be stated with $\log(d)$ in place of $\log(nd)$ at the cost of slight changes to terms involving $1 - o_K(1)$. The only places where we pay close attention to the difference between the two is when we inspect the consistency properties of the two tests. Also note that we write $\mathbb{E}$ instead of $\mathbb{E}_F$ for the sake of notational convenience.

A.1. Variance Estimation.

From (2.10), recall the definition of the variance estimator $\hat{\sigma}_i^2$. The following theorem characterizes the uniform convergence rate of the differences $\{n\hat{\sigma}_i^2 - m^2\zeta_{1,i} \mid i = 1, \ldots, d\}$ with $\zeta_{1,i}$ defined in (2.8).

THEOREM A.1. *If Assumption (A1) is satisfied and $\log d = o(n^\gamma)$ for $\gamma \leq \frac{1}{4/\beta+1}$, we have*

$$\max_{1\leq i\leq d} |n\hat{\sigma}_i^2 - m^2\zeta_{1,i}| \lesssim B_n^2 \sqrt{\frac{\log(nd)}{n}}$$

with probability at least $1 - o_K(1)$, where the hidden constant in the inequality depends only on β.

PROOF. We will use similar arguments as given in the proof of Lemma A.1 in Zhou et al. (2019). Some difficulties arise as in contrast to this work we consider U-statistics with unbounded kernels. First, we define a centralized version of the U-statistics in (2.6) and the leave one out estimator below (2.10),

$$\bar{U}_i := U_i - \theta_i \quad \text{and} \quad \bar{q}_{k,i} := \binom{n-1}{m-1}^{-1} \sum_{1\leq l_1<\ldots<l_{m-1}\leq n, l_j\neq k} g_i(X_k, X_{l_1}, \ldots, X_{l_{m-1}}),$$

respectively, where $g_i(X_{l_1}, \ldots, X_{l_m}) = h_i(X_{l_1}, \ldots, X_{l_m}) - \theta_i$ and $1 \leq i \leq d$. A simple calculation shows that

$$\hat{\sigma}_i^2 = \frac{m^2(n-1)}{n(n-m)^2} \sum_{k=1}^n (q_{k,i} - U_i)^2 = \frac{m^2(n-1)}{n(n-m)^2} \sum_{k=1}^n (\bar{q}_{k,i} - \bar{U}_i)^2.$$

Setting $g_{1,i}(x) = \mathbb{E}[g_i(X_1,\ldots,X_m)|X_1 = x]$, $\bar{g}_{1,i} = \frac{1}{n}\sum_{j=1}^n g_{1,i}(X_j)$ and using the triangle inequality then yields

$$|n\hat{\sigma}_i^2 - m^2\zeta_{1,i}| \leq \left|\frac{m^2(n-1)}{(n-m)^2}\sum_{k=1}^n \left[(\bar{q}_{k,i} - \bar{U}_i)^2 - (g_{1,i}(X_k) - \bar{g}_{1,i})^2\right]\right| + \left|\frac{m^2(n-1)}{(n-m)^2}\sum_{k=1}^n (g_{1,i}(X_k) - \bar{g}_{1,i})^2 - m^2\zeta_{1,i}\right| =: M_i^{(1)} + M_i^{(2)}$$

for $1 \leq i \leq d$. Therefore, the claim of Theorem A.1 is a consequence of the following two Lemmas A.2 and A.3. □

LEMMA A.2. *Under the conditions of Theorem A.1 we have with probability at least* $1 - o_K(1)$ *that*

$$\max_{1\leq i\leq d} M_i^{(1)} \lesssim B_n^2\sqrt{\frac{\log(nd)}{n}}\,.$$

PROOF. Recalling that $\sum_{k=1}^n (g_{1,i}(X_k) - \bar{g}_{1,i})^2 = \sum_{k=1}^n g_{1,i}^2(X_k) - n\bar{g}_{1,i}^2$ and using the triangle inequality yields

$$M_i^{(1)} \lesssim \left|\bar{U}_i^2 - \bar{g}_{1,i}^2\right| + \left|\frac{m^2(n-1)}{(n-2)^2}\sum_{k=1}^n \left(\bar{q}_{k,i}^2 - g_{1,i}^2(X_k)\right)\right|. \tag{A.1}$$

For the first term we use Lemma B.18 from the online supplement, as we will use it repeatedly throughout the remaining proofs we will explain its application in detail one time. We apply it separately to the U-Statistics $\bar{U}_i$ and $\bar{g}_{1,i}$ which fulfill the required conditions by the first equation in assumption (A1) and the assumption that $\gamma \leq \frac{1}{4/\beta+1}$, note that we will always use the version of the bound containing $\log(nd)$ except when considering consistency properties.

$$\max_{1\leq i\leq d}|\bar{U}_i^2 - \bar{g}_{1,i}^2| = \max_{1\leq i\leq d}|(\bar{U}_i - \bar{g}_{1,i})(\bar{U}_i + \bar{g}_{1,i})| \lesssim B_n^2\frac{\log(nd)}{n}$$

with probability at least $1 - o_K(1)$. For the second term in (A.1) a more sophisticated analysis is necessary which we facilitate by decomposing

$$\sum_{1\leq l_1<\ldots<l_{m-1}\leq n, l_j\neq k} g_i(X_k, X_{l_1},\ldots,X_{l_{m-1}}) = A_{n,m}g_{1,i}(X_k) + B_{n,m}S_i + \Gamma_{k,i}\,,$$

where $A_{n,m} = \binom{n-1}{m-1} - \binom{n-2}{m-2}$, $B_{n,m} = \binom{n-2}{m-2}$, $S_i = \sum_{l=1}^n g_{1,i}(X_l)$ and

$$\Gamma_{k,i} = \sum_{1\leq l_1<\ldots<l_{m-1}\leq n, l_j\neq k}\left(g_i(X_k, X_{l_1}\ldots,X_{l_{m-1}}) - g_{1,i}(X_k) - \sum_{j=1}^{m-1} g_{1,i}(X_{l_j})\right)\,.$$

By the definition of $\bar{q}_{k,i}$, we then have

$$\bar{q}_{k,i} = \frac{A_{n,m}g_{1,i}(X_k) + B_{n,m}S_i + \Gamma_{k,i}}{\binom{n-1}{m-1}},$$

which leaves us with the task to bound

$$J := \max_{1\leq i\leq d}\left|\frac{m^2(n-1)}{(n-2)^2}\sum_{k=1}^n \left[\bar{q}_{k,i}^2 - g_{1,i}^2(X_k)\right]\right|\,.$$

Setting $V_i^2 = \sum_{k=1}^n g_{1,i}^2(X_k)$, $D_{n,m} = \binom{n-1}{m-1}$ and $\Lambda_i^2 = \sum_{k=1}^n \Gamma_{k,i}^2$, we have

$$\sum_{k=1}^n \bar{q}_{k,i}^2 = \frac{1}{D_{n,m}^2}\Big\{A_{n,m}^2 V_i^2 + \Lambda_i^2 + (nB_{n,m}^2 + 2A_{n,m}B_{n,m})S_i^2 + 2A_{n,m}\sum_{k=1}^n g_{1,i}(X_k)\Gamma_{k,i} + 2B_{n,m}S_i\sum_{k=1}^n \Gamma_{k,i}\Big\}$$

which together with the Cauchy-Schwarz inequality (for J_4 and J_5) yields

$$J \lesssim J_1 + J_2 + J_3 + J_4 + J_5 \,,$$

where

$$J_1 = \max_{1\le i\le d}\left|\frac{(A_{n,m}^2 - D_{n,m}^2)V_i^2}{nD_{n,m}^2}\right| \lesssim \max_{1\le i\le d}\frac{V_i^2}{n^2}\,,$$

$$J_2 = \max_{1\le i\le d}\frac{\Lambda_i^2}{nD_{n,m}^2}\,,$$

$$J_3 = \max_{1\le i\le d}\frac{(nB_{n,m}^2 + 2A_{n,m}B_{n,m})S_i^2}{nD_{n,m}^2} \lesssim \max_{1\le i\le d}\frac{S_i^2}{n^2}\,,$$

$$J_4 = \max_{1\le i\le d}\frac{2A_{n,m}V_i\Lambda_i}{nD_{n,m}^2} \lesssim \max_{1\le i\le d}\frac{V_i\Lambda_i}{n^2}\,,$$

$$J_5 = \max_{1\le i\le d}\frac{2B_{n,m}|S_i|\sqrt{n}\Lambda_i}{nD_{n,m}^2} \lesssim \max_{1\le i\le d}\frac{V_i\Lambda_i}{n^2}\,.$$

In the remainder of this proof, we will bound the terms $J_1,\dots,J_5$ separately. For J_1 we have by Lemma B.11 that $\left\|g_{1,i}^2(X_k)\right\|_{\psi_{\beta/2}} \le B_n^2$ so that Lemma B.18 in the online supplement yields that

$$J_1 \lesssim \max_{1\le i\le d}\frac{V_i^2}{n^2} = \max_{1\le i\le d}\frac{\sum_{k=1}^n g_{1,i}^2(X_k) - \zeta_{1,i}}{n^2} + \frac{\zeta_{1,i}}{n^2} \lesssim B_n^2\sqrt{\frac{\log(nd)}{n^3}} \tag{A.2}$$

with probability at least $1 - 3/(nd) - C(\log(nd))^{1/2+1/\beta}/\sqrt{n}$.
Regarding the term J_2, we define the set

$$A_n = \Big\{\max_{\substack{1\le i\le d\\ 1\le k,l_1,\dots,l_{m-1}\le n}}\Big(g_i(X_k, X_{l_1}\dots,X_{l_{m-1}}) - g_{1,i}(X_k) - \sum_{j=1}^{m-1} g_{1,i}(X_{l_j})\Big) \le C_\beta B_n(\log(nd))^{1/\beta}\Big\}\,,$$

where the constant C_β is chosen such that $\mathbb{P}(A_n) \ge 1 - \frac{1}{nd}$. Indeed, using the union bound and Lemma B.8 in the online supplement it is easy to see that by choosing C_β appropriately (this can be done universally with only dependence on β) we obtain $\mathbb{P}(A_n) \ge 1 - \frac{1}{nd}$.
Conditional on X_k we now apply Lemma B.18 in the online supplement to $\Gamma_{k,i}$ on the set A_n with $K = B_n(\log(nd))^{1/\beta}$ to obtain

$$\mathbb{P}\Big(\Big\{\max_{1\le i\le d, 1\le k\le n}\Gamma_{k,i}/D_{n,m} \lesssim B_n(\log(nd))^{1/\beta}\sqrt{\tfrac{\log(nd)}{n}}\Big\} \cap A_n\Big)$$

$$
\begin{aligned}
&= \sum_{k=1}^{n} \mathbb{P}\Big(\Big\{ \max_{1\le i\le d} \Gamma_{k,i}/D_{n,m} \lesssim B_n (\log(nd))^{1/\beta} \sqrt{\tfrac{\log(nd)}{n}} \Big\} \cap A_n \,\Big|\, \tilde{B}_k \Big) \frac{1}{n} \\
&= \sum_{k=1}^{n} \mathbb{E}\Big[\mathbb{P}\Big(\Big\{ \max_{1\le i\le d} \Gamma_{k,i}/D_{n,m} \lesssim B_n (\log(nd))^{1/\beta} \sqrt{\tfrac{\log(nd)}{n}} \Big\} \cap A_n \,\Big|\, X_k, \tilde{B}_k \Big) \Big] \frac{1}{n} \\
&\ge 1 - o_K(1)
\end{aligned}
$$

where we used that $\Gamma_{k,i}$ and $\Gamma_{k,j}$ have the same distribution to obtain the second line and define $\tilde{B}_l$ as the event that $\max_{1\le i\le d, 1\le k\le n} \Gamma_{k,i}/D_{n,m} = \max_{1\le i\le d} \Gamma_{l,i}/D_{n,m}$. Using the definition of Λ_i and recalling that $\gamma \le \frac{1}{4/\beta+1}$, this yields

$$
J_2 = \max_{1\le i\le d} \frac{\Lambda_i^2}{nD^2} \lesssim B_n^2 \frac{\log(nd)}{n} (\log(nd))^{2/\beta} \lesssim B_n^2 \sqrt{\frac{\log(nd)}{n}} \tag{A.3}
$$

with probability at least $1 - \frac{4}{nd} - C(\log(d))^{1/2+1/\beta}/\sqrt{n}$.
For J_3 we have by Lemma B.13 in the online supplement that

$$
J_3 \lesssim \max_{1\le i\le d} \frac{S_i^2}{n^2} \lesssim B_n^2 \frac{\log(nd)}{n} \tag{A.4}
$$

with probability at least $1 - \frac{3}{nd}$. Finally, regarding J_4 and J_5 we have by the calculations for J_2

$$
J_l \lesssim \max_{1\le i\le d} \frac{V_i \Lambda_i}{n^2} = \max_{1\le i\le d} \sqrt{\frac{V_i^2 \Lambda_i^2}{n^4}} \lesssim B_n^2 \sqrt{\frac{\log(nd)}{n}}, \qquad l = 4, 5, \tag{A.5}
$$

with probability at least $1 - \frac{4}{nd} - C(\log(d))^{1/2+1/\beta}/\sqrt{n}$ provided that $\gamma \le \frac{1}{4/\beta+1}$. Combining (A.2), (A.3), (A.4) and (A.5) shows that $J \lesssim B_n^2 \sqrt{(\log(nd))/n}$ with probability at least $1 - o_K(1)$. □

LEMMA A.3. *Under the conditions of Theorem A.1, we have with probability at least* $1 - o_K(1)$ *that*

$$
\max_{1\le i\le d} M_i^{(2)} \lesssim B_n^2 \sqrt{\frac{\log(nd)}{n}}.
$$

PROOF. Recalling that $\sum_{k=1}^{n} (g_{1,i}(X_k) - \bar{g}_{1,i})^2 = \sum_{k=1}^{n} g_{1,i}(X_k)^2 - n\bar{g}_{1,i}^2$ as well as $\zeta_{1,i} = \mathbb{E}[g_{1,i}^2(X_1)]$ yields that

$$
\left| \frac{m^2(n-1)}{(n-m)^2} \sum_{k=1}^{n} (g_{1,i}(X_k) - \bar{g}_{1,i})^2 - m^2 \zeta_{1,i} \right| \lesssim \left| \frac{1}{n} \sum_{k=1}^{n} g_{1,i}(X_k)^2 - \zeta_{1,i} \right| + \bar{g}_{1,i}^2 .
$$

Note that $\gamma \le \frac{1}{4/\beta-1}$. We then apply Lemma B.13, Lemma B.11 and Lemma B.9 in the online supplement to obtain that with probability at least $1 - C/(nd)$,

$$
\max_{1\le i\le d} \left| \frac{1}{n} \sum_{k=1}^{n} \left(g_{1,i}^2(X_k) - \zeta_{1,i} \right) \right| \lesssim B_n^2 \sqrt{\frac{\log(nd)}{n}}
$$

and

$$
\max_{1\le i\le d} \bar{g}_{1,i}^2 \lesssim B_n^2 \frac{\log(nd)}{n} .
$$

□

A.2. Proof of the results in Section 2.1.

A.2.1. *Preliminaries.* The main step in the proofs of Theorem 2.2 and 2.4 is a weak convergence result for the statistic

$$\mathcal{T}_n := \max_{1\le i\le d} \frac{U_i^2-\theta_i^2}{2\hat{\sigma}_i|\theta_i|}$$

in the case where $|\theta|_{\min} := \min_{1\le i\le d}|\theta_i| > c$ for some constant $c>0$. To prepare its proof we first replace the variance estimates $\hat{\sigma}_i^2$ by the population variances using Lemma A.4 and then apply the Gaussian approximation in Lemma A.5 to the linearized statistic $\mathcal{T}_n$ assuming that $\log d$ and the constants B_n in Assumption (A1) do not grow too fast. To this end, we recall the notation (2.8) and define

$$S_n = \sqrt{n}\max_{1\le i\le d} \frac{\frac{1}{n}\sum_{k=1}^n h_{1,i}(X_k)-\theta_i}{\sqrt{\zeta_{1,i}}}\operatorname{sign}(\theta_i). \tag{A.6}$$

LEMMA A.4. *If Assumptions (A1) and (A2) are satisfied and $|\theta|_{min} > c$ for some positive constant c, then it holds*

$$\left|\max_{1\le i\le d}\frac{U_i^2-\theta_i^2}{2\hat{\sigma}_i|\theta_i|} - S_n\right| \lesssim B_n^3\frac{\log(nd)}{\sqrt{n}} + B_n\frac{\log(nd)}{n^{1/2-\gamma/\beta}}$$

with probability at least $1-o_K(1)$. Here the constants hidden in $\lesssim$ only depend on the quantities $c,\gamma,\beta,\underline{b}$, and therefore the estimate is uniform for the subsets of the classes $\mathcal{H}_0(\Delta)$ and $\mathcal{H}_1$ defined in (2.15) and (2.18), respectively, for which $|\theta|_{min} > c$.

PROOF. By Theorem A.1, we have

$$\max_{1\le 1\le d}\left|\sqrt{n}\hat{\sigma}_i - m\sqrt{\zeta_{1,i}}\right| = \max_{1\le 1\le d}\frac{\left|n\hat{\sigma}_i^2-m^2\zeta_{1,i}\right|}{\sqrt{n}\hat{\sigma}_i+m\sqrt{\zeta_{1,i}}} \lesssim B_n^2\sqrt{\frac{\log(nd)}{n}}$$

up to a constant depending only on β and $\underline{b}$, and therefore,

$$\begin{aligned}
&\left|\sqrt{n}\max_{1\le i\le d}\frac{U_i^2-\theta_i^2}{2\sqrt{n}\hat{\sigma}_i|\theta_i|} - \sqrt{n}\max_{1\le i\le d}\frac{U_i^2-\theta_i^2}{2m\sqrt{\zeta_{1,i}}|\theta_i|}\right| \\
&\le \max_{1\le i\le d}\left|\sqrt{n}\hat{\sigma}_i - m\sqrt{\zeta_{1,i}}\right| \sqrt{n}\max_{1\le i\le d}\left|\frac{U_i^2-\theta_i^2}{2m\sqrt{\zeta_{1,i}}\sqrt{n}\hat{\sigma}_i|\theta_i|}\right| \\
&\lesssim B_n^2\sqrt{\frac{\log(nd)}{n}}\ \max_{1\le i\le d}\left|\frac{U_i^2-\theta_i^2}{2m\sqrt{\zeta_{1,i}}\hat{\sigma}_i|\theta_i|}\right|.
\end{aligned}$$

By the same arguments, the triangle inequality and writing

$$\hat{T}_{n,1} = \max_{1\le i\le d}\left|\theta_i\frac{U_i-\theta_i}{m\sqrt{\zeta_{1,i}}\hat{\sigma}_i|\theta_i|}\right| \quad \text{and} \quad \hat{T}_{n,2} = \max_{1\le i\le d}\left|\frac{(U_i-\theta_i)^2}{2m\sqrt{\zeta_{1,i}}\hat{\sigma}_i|\theta_i|}\right|,$$

we obtain that

$$\max_{1\le i\le d}\left|\frac{U_i^2-\theta_i^2}{2m\sqrt{\zeta_{1,i}}\hat{\sigma}_i|\theta_i|}\right| \le \hat{T}_{n,1}+\hat{T}_{n,2} \lesssim B_n\sqrt{\log(nd)}, \tag{A.7}$$

with probability at least $1 - o_K(1)$, where the last inequality in (A.7) follows from Lemma B.18 in the online supplement. Therefore, the constant in this inequality only depends on the constants γ, c and β. Combining the two estimates we conclude

$$\left| \sqrt{n} \max_{1\le i\le d} \frac{U_i^2 - \theta_i^2}{2\sqrt{n}\hat{\sigma}_i|\theta_i|} - \sqrt{n} \max_{1\le i\le d} \frac{U_i^2 - \theta_i^2}{2m\sqrt{\zeta_{1,i}}|\theta_i|} \right| \lesssim B_n^3 \frac{\log(nd)}{\sqrt{n}} \,. \tag{A.8}$$

We then observe that

$$\begin{aligned} \sqrt{n} \max_{1\le i\le d} \theta_i \frac{U_i - \theta_i}{m\sqrt{\zeta_{1,i}}|\theta_i|} &\le \sqrt{n} \max_{1\le i\le d} \frac{U_i^2 - \theta_i^2}{2m\sqrt{\zeta_{1,i}}|\theta_i|} \\ &\le \sqrt{n} \max_{1\le i\le d} \theta_i \frac{U_i - \theta_i}{m\sqrt{\zeta_{1,i}}|\theta_i|} + \sqrt{n} \max_{1\le i\le d} \frac{(U_i - \theta_i)^2}{2m\sqrt{\zeta_{1,i}}|\theta_i|} \\ &\lesssim \sqrt{n} \max_{1\le i\le d} \theta_i \frac{U_i - \theta_i}{m\sqrt{\zeta_{1,i}}|\theta_i|} + B_n^2 \frac{\log(nd)}{\sqrt{n}} \,, \end{aligned} \tag{A.9}$$

where the last inequality follows by Lemma B.18 with probability at least $1 - o_K(1)$ and the hidden constant depends only on γ and β. Using the estimate (B.17) (with $t = B_n(\log(nd))/(n^{1-\gamma/\beta})$) in the proof of Lemma B.18, we get

$$\begin{aligned} \left| \sqrt{n} \max_{1\le i\le d} \theta_i \frac{U_i - \theta_i}{m\sqrt{\zeta_{1,i}}|\theta_i|} - S_n \right| &\le \sqrt{n} \max_{1\le i\le d} \left| \frac{\frac{1}{n}\sum_{k=1}^n (h_{1,i}(X_k) - \theta_i) - (U_i - \theta_i)}{\sqrt{\zeta_{1,i}}} \right| \\ &\lesssim B_n \frac{\log(nd)}{n^{1/2-\gamma/\beta}} \end{aligned} \tag{A.10}$$

with probability at least $1 - C(\log(nd))^\beta n^{-\gamma/\beta}$, where the constants in both inequalities depend only on β and γ. Combining (A.8), (A.9) and (A.10) yields the desired result. □

We will now provide a Gaussian approximation for S_n, which is a consequence of Lemma B.3 in the online supplement. Note that the conditions of Lemma B.3 are satisfied because of Assumption (A1), (A2) and Jensen's inequality.

LEMMA A.5. *Under the assumptions of Theorem 2.2 and $|\theta|_{\min} > c$ we have, up to some constant C depending only on $\beta, D, \underline{b}, c$, that*

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}(S_n \le x) - \mathbb{P}(S_n^G \le x) \right| \le C \left(\frac{B_n^2 (\log(nd))^{4+2/\beta}}{n} \right)^{1/4} ,$$

where S_n^G is defined as in (A.6) with the difference that the vectors $X_1, \ldots, X_k$ are replaced by independent centered Gaussian vectors with covariance matrix $\Gamma = (\Gamma_{ij})_{1\le i,j\le d}$ defined by

$$\Gamma_{ij} = \mathrm{Corr}(h_{1,i}(X_1), h_{1,j}(X_1)) sign(\theta_i\theta_j) \,. \tag{A.11}$$

Note that the sole dependence on $\beta, D, \underline{b}, c$ of the bound implies that it is valid uniformly for the subsets of the classes $\mathcal{H}_0, \mathcal{H}_1$ in (2.15) and (2.18) for which $|\theta|_{\min} > c$.

By the Schur product theorem, Γ is a positive semidefinite matrix as it is the Hadamard product of a correlation matrix and the rank one matrix $(\mathrm{sign}(\theta_i\theta_j))_{1\le i,j\le d}$ which are both positive semidefinite.

A.2.2. *Proof of Theorem 2.2.*. We recall $|\theta_i| \leq \Delta$ for all $1 \leq i \leq d$. Fix some $0 < c_0 < \Delta$ and consider the following decomposition of $\{1, \ldots, d\}$:

$$I_1 = \{1 \leq i \leq d : \quad |\theta_i| > c_0\},$$
$$I_2 = \{1 \leq i \leq d : \quad 0 < |\theta_i| \leq c_0\},$$
$$I_3 = \{1 \leq i \leq d : \quad |\theta_i| = 0\}.$$

Using the definition of $\mathcal{T}_{n,\Delta}$ in (2.9), we note that

$$\mathcal{T}_{n,\Delta} = \max\left\{\max_{i \in I_1} \frac{U_i^2 - \Delta^2}{2\hat{\sigma}_i \Delta}, \max_{i \in I_2} \frac{U_i^2 - \Delta^2}{2\hat{\sigma}_i \Delta}, \max_{i \in I_3} \frac{U_i^2 - \Delta^2}{2\hat{\sigma}_i \Delta}\right\}.$$

First, we show that the second and third terms in the right-hand maximum are negligible for our purposes. For the third term we use Lemma B.18 and $\theta_i = 0$ for all $i \in I_3$ to obtain

$$\max_{i \in I_3} \frac{U_i^2 - \Delta^2}{2\hat{\sigma}_i \Delta} \lesssim \max_{i \in I_3} \frac{B_n^2 \log(nd)/\sqrt{n} - \sqrt{n}\Delta^2}{2\hat{\sigma}_i \sqrt{n}} \tag{A.12}$$

with probability at least $1 - o_K(1)$. Due to the assumption $\zeta_{1,i} \leq D$ and the fact that $B_n^2 \log(nd)/\sqrt{n} \ll \sqrt{n}\Delta^2$ this diverges to $-\infty$ with rate at least $\sqrt{n}$. Note that all constants in these inequalities depend only on γ, β and D.

For the second term we use that

$$\begin{aligned} \max_{i \in I_2} \frac{U_i^2 - \Delta^2}{2\hat{\sigma}_i \Delta} &= \max_{i \in I_2} \frac{U_i^2 - \theta_i^2 + (\theta_i^2 - \Delta^2)}{2\hat{\sigma}_i \Delta} \\ &\lesssim \max_{i \in I_2} \frac{B_n \log(d)/\sqrt{n} + \sqrt{n}(\theta_i^2 - \Delta^2)}{2\hat{\sigma}_i \sqrt{n}}, \end{aligned} \tag{A.13}$$

where the last inequality holds with probability at least $1 - o_K(1)$ by the same calculation as in (A.12) and the decomposition of $U_i^2 - \theta_i^2$ into a linear and quadratic part as in (A.9). For the same reasons as in (A.12) we conclude that the right-hand side of (A.13) converges to $-\infty$ with rate at least $\sqrt{n}$ (note that $|\theta_i| \leq c_0 < \Delta$ for all $i \in I_2$). We again stress the fact that all constants in these inequalities depend only on γ, β, D and c_0.

We hence obtain uniformly for all distributions in $\mathcal{H}_0(\Delta)$

$$\mathbb{P}\left(a_d\left(\mathcal{T}_{n,\Delta} - b_d\right) > q_{1-\alpha}\right) = \mathbb{P}\left(a_d\left(\sqrt{n}\max_{i \in I_1} \frac{U_i^2 - \Delta^2}{2\hat{\sigma}_i \Delta} - b_d\right) > q_{1-\alpha}\right) + o(1),$$

and it remains to show that the probability on the right hand side is asymptotically bounded by α uniformly in $\mathcal{H}_0(\Delta)$. As $|\theta_i| \leq \Delta$ we obtain the bound

$$\mathcal{T}_n(I_1) := \max_{i \in I_1} \frac{U_i^2 - \Delta^2}{2\hat{\sigma}_i \Delta} \leq \max_{i \in I_1} \frac{U_i^2 - \theta_i^2}{2\hat{\sigma}_i \Delta}.$$

Let C_1 denote the constants hidden in $\lesssim$ in Lemma A.4 and let C_2 denote the hidden constants in Lemma A.5 (these constants depend only on $\gamma, \beta, c_0, \underline{b}, D$). Defining

$$c_{\gamma,\beta}^{(1)} := C_1 \frac{B_n^3(\log(nd))}{\sqrt{n}} + \frac{B_n(\log(nd))}{n^{1/2 - \gamma/\beta}},$$

$$c_{\gamma,\beta}^{(2)} := C_2 \left(\frac{B_n^2(\log(nd))^{4+2/\beta}}{n}\right)^{1/4},$$

$$c_{\gamma,\beta}^{(3)} := c_{\gamma,\beta}^{(1)} \sqrt{\log(nd)} + c_{\gamma,\beta}^{(2)}$$

and using Lemma A.4, Lemma A.5 and Nazarovs Inequality (see **?**) we obtain that

$$\begin{aligned}
\mathbb{P}\big(a_d\big(\mathcal{T}_n(I_1)-b_d\big)>q_{1-\alpha}\big) &\le \mathbb{P}\big(\max_{i\in I_1}(S_n)_i > q_{1-\alpha}/a_d - c^{(1)}_{\gamma,\beta}+b_d\big)\\
&\le \mathbb{P}\big(\max_{i\in I_1}(S_n^G)_i > q_{1-\alpha}/a_d - c^{(1)}_{\gamma,\beta}+b_d\big)+c^{(2)}_{\gamma,\beta}\\
&\le \mathbb{P}\big(\max_{i\in I_1}(S_n^G)_i > q_{1-\alpha}/a_d +b_d\big)+c^{(3)}_{\gamma,\beta}\\
&\le \mathbb{P}\big(\max_{i\in I_1}Z_i > q_{1-\alpha}/a_d +b_d\big)+c^{(3)}_{\gamma,\beta}+\gamma_n\\
&\le \mathbb{P}\big(\max_{1\le i\le d}Z_i > q_{1-\alpha}/a_d +b_d\big)+c^{(3)}_{\gamma,\beta}+\gamma_n
\end{aligned}$$

where Z is a d-dimensional random vector with independent standard normal components, S_n^G is defined in Lemma A.5, $\gamma_n \to 0$ is the sequence in Assumption (A3) and the second to last line is obtained by the normal comparison Lemma from Leadbetter et al. (1983) (Theorem 4.2.1) . Note that these estimates are uniform with respect the distribution in $\mathcal{H}_0(\Delta)$ (as the constants C_1, C_2 depend only on $\gamma,\beta,c,\underline{b},D,\gamma_n$) and that last probability does not depend on $\mathcal{H}_0(\Delta)$. Therefore, taking the $\limsup$ and the supremum with respect to $\mathcal{H}_0(\Delta)$ yields

$$\begin{aligned}
\limsup_{n\to\infty}\sup_{F\in\mathcal{H}_0(\Delta)}\mathbb{P}\left(a_d\left(\mathcal{T}_{n,\Delta}-b_d\right)>q_{1-\alpha}\right) &= \limsup_{n\to\infty}\sup_{F\in\mathcal{H}_0(\Delta)}\mathbb{P}\big(a_d\big(\mathcal{T}_n(I_1)-b_d\big)>q_{1-\alpha}\big)\\
&\le \lim_{n\to\infty}\mathbb{P}\big(\|Z\|_\infty > q_{1-\alpha}/a_d+b_d\big)=\alpha\ ,
\end{aligned}$$

which proves the assertion of Theorem 2.2.

A.2.3. *Proof of Theorem 2.4.* Let i_0 be an index such that $|\theta_{i_0}| = \max_{1\le i\le d}|\theta_i| > \Delta$; note that i_0 can depend on n, which is not reflected by our notation. Then we have

$$\mathcal{T}_{n,\Delta} \ge \frac{U_{i_0}^2-\theta_{i_0}^2}{2\hat\sigma_{i_0}\Delta}+\frac{\theta_{i_0}^2-\Delta^2}{2\hat\sigma_{i_0}\Delta}\ .$$

By the same arguments as for (A.7) we obtain that with probability $1-o_K(1)$

$$\left|\frac{U_{i_0}^2-\theta_{i_0}^2}{2\hat\sigma_{i_0}\Delta}\right| \lesssim B_n\sqrt{\log(d)}\frac{1}{\sqrt{n}\hat\sigma_{i_0}}$$

while the second term converges to ∞ at rate $\frac{\sqrt{n}\xi_n}{\sqrt{n}\hat\sigma_{i_0}}$ with $\xi_n = \theta_{i_0}^2-\Delta^2$ by the same arguments as in (A.12). These bounds depend only on the constants γ,β and B_n in the Assumption (A1) and therefore hold uniformly over the class $\mathcal{H}_1$ defined in (2.18). This yields the desired conclusion whenever $\xi_n \ge CB_n\sqrt{\frac{\log(d)}{n}}$ for some large enough constant C as $(\frac{q_{1-\alpha}}{a_d}+b_d)\Delta \simeq \sqrt{\log d}$ and $\sqrt{n}\hat\sigma_{i_0}\lesssim B_n$ by Lemma A.1. Here $a\simeq b$ denotes $c_1a\le b\le c_2a$ for some constants c_1,c_2 that do not depend on n.

A.3. Proof of the results in Section 2.2. Let $\xi_k = (\xi_{k1},\dots,\xi_{kn})^\top, 1\le k\le n$, be independent identically multinomial $\mathcal{M}(1;\frac{1}{n},\dots,\frac{1}{n})$ distributed random vectors independent of $X_1,\dots,X_n$, that is $\mathbb{P}(\xi_{k1}=y_1,\dots,\xi_{kn}=y_n)=1/n$ for $(y_1,\dots,y_n)\in\{0,1\}^n$ such that $y_1+\dots+y_n=1$. Then a sample $X_1^*,\dots,X_n^*$ drawn with replacement from $X_1,\dots,X_n$ can be represented as

$$X_k^* = X\xi_k = \sum_{j=1}^n \xi_{kj}X_j\ ,$$

where $X = (X_1, \ldots, X_n) \in \mathbb{R}^{p \times n}$. We denote by $\mathbb{P}^*$ and $\mathbb{E}^*$ the probabilities and expectations conditional on $X_1, \ldots, X_n$. We also recall the definition of the statistic U_i^* in (2.19) and note that $\mathbb{E}^*[U_i^*] = V_i$ (see (2.20)), so that conditional on $X_1, \ldots, X_n$ the quantity $U^* - V = (U_1^* - V_1, \ldots, U_d^* - V_d)^\top$ is a U-Statistic of the random variables $\xi_1, \ldots, \xi_n$. We start with several auxiliary results, which are required for the proof of Theorem 2.5 in Section A.3.1.

A.3.1. *Some preparations.* We first observe that the conditional mean of the Bootstrap statistic is close to the mean of the original statistic, this will be used multiple times in some of the following approximations when terms involving $\|V - V_\Delta\|_\infty$ appear, where we used the definition $V_\Delta = (V_{1,\Delta}, ..., V_{d,\Delta})$.

LEMMA A.6. *Under the assumptions of Theorem 2.5 we have that*

$$\|V - \theta\|_\infty \lesssim B_n \sqrt{\frac{\log(nd)}{n}},$$

with $\mathbb{P}$-probability at least $1 - o_K(1)$ where all constants involved depend only on γ and β which implies that the bound holds uniformly for the classes $\mathcal{H}_0(\Delta)$ and $\mathcal{H}_{0,boot}(\Delta)$ defined in (2.15) and (2.26) , respectively.

PROOF. We first decompose V (see (2.20) for its definition) into its diagonal and non-diagonal parts

$$\begin{aligned} V = {} & \frac{1}{n^m} \sum_{k=1}^{n} h(X_k, \ldots, X_k) + \frac{1}{n^m} \sum_{1 \le l_1 \ne l_2 = \ldots = l_m \le n} h(X_{l_1}, \ldots, X_{l_m}) \\ & + \cdots + \frac{1}{n^m} \sum_{1 \le l_1 \ne \ldots \ne l_m \le n} h(X_{l_1}, \ldots, X_{l_m}) \end{aligned} \tag{A.14}$$

Applying Lemma B.13 to the diagonal part yields, up to some constant depending only on β and γ,

$$\left\| \frac{1}{n^m} \sum_{k=1}^{n} h(X_k, \ldots, X_k) \right\|_\infty \lesssim B_n \sqrt{\frac{\log(nd)}{n^{2m-1}}}$$

with $\mathbb{P}$-probability at least $1 - o_K(1)$, where we also used Lemma B.8 to uniformly bound the mean of $h(X_k, \ldots, X_k)$ by a multiple of B_n that depends only on β. Next, we will exemplary inspect the term

$$\frac{1}{n^m} \sum_{1 \le l_1 \ne l_2 \le n} h(X_{l_1}, X_{l_2}, \ldots, X_{l_2})$$

in detail, all other terms (except the very last) in the decomposition of V_i can be treated analogously. Note that $H(x_1, x_2) = h(x_1, x_2, \ldots, x_2)$ defines a non-symmetric kernel of order two whose associated U-statistic is given by the preceding equation, which can be symmetrized without changing the value of the associated U-statistic. Applying Lemma B.18 then yields

$$\left\| \frac{1}{n^m} \sum_{1 \le l_1 \ne l_2 \le n} h(X_{l_1}, \ldots, X_{l_2}) \right\|_\infty \lesssim B_n \sqrt{\frac{\log(nd)}{n^{2m-3}}} + n^{-(m-2)}$$

with probability at least $1 - o_K(1)$ for some constant C that depends only on β (note that the mean is negligible by the same arguments as for the first term). The same arguments show that all terms in (A.14) (except the last one) are of smaller order than $B_n\sqrt{(\log(nd))/n}$. Finally, for the remaining term in (A.14), we have by Lemma B.18 that

$$\Big\| \frac{1}{n^m} \sum_{1\le l_1\neq \ldots\neq l_m\le n} h(X_{l_1},\ldots,X_{l_m}) - \theta \Big\|_\infty \lesssim B_n\sqrt{\frac{\log(nd)}{n}},$$

with $\mathbb{P}$-probability at least $1-o_K(1)$, which proves the assertion of the lemma. □

Next, we set

$$S_n^* = \sqrt{n}\max_{1\le i\le d}\theta_i \frac{U_i^*-V_i}{m\sqrt{\zeta_{1,i}}\Delta}\,, \tag{A.15}$$

which is a linearized version of $\mathcal{T}_n^*$ (see (2.22) for its definition). We will show that $\mathcal{T}_n^*$ is well approximated by S_n^*. This will allow us to apply Gaussian approximation results to approximate the distribution of $\mathcal{T}_n^*$.

LEMMA A.7. *If the assumptions of Theorem 2.5 are satisfied,* $\min_{1\le i\le d}\zeta_{1,i}\ge \underline{b}>0$ *and* $\max_{i=1}^d|\theta_i|\le\Delta$, *we have that*

$$|\mathcal{T}_n^* - S_n^*| \lesssim B_n^2\frac{(\log(nd))^{1+2/\beta}}{\sqrt{n}} + B_n^3\frac{(\log(nd))^{1+1/\beta}}{\sqrt{n}} \tag{A.16}$$

holds with $\mathbb{P}^*$ *probability at least* $1-o_K(1)$ *on a set of* $\mathbb{P}$*-probability at least* $1-o_K(1)$. *Here the constant in inequality* (A.16) *depends only on* $\beta,\gamma,\underline{b}$. *This implies that* (A.16) *holds uniformly for the subset of the class* $\mathcal{H}_{0,boot}(\Delta)$ *in* (2.26) *for which* $\min_{1\le i\le d}\zeta_{1,i}\ge\underline{b}>0$.

PROOF. We start by noting that an analogue of Lemma B.12 in the online supplement (which considers the maximum with respect to two indices) and Assumption (A1') show that up to some universal constant

$$\max_{1\le i\le d, 1\le j_1<\ldots<j_m\le n}\|h_i(X\xi_{j_1},\ldots,X\xi_{j_m})\|_\infty \lesssim B_n(\log(dn))^{1/\beta}$$

with $\mathbb{P}$-probability at least $1-o_K(1)$. Part $iii)$ of Lemma B.8 then yields

$$\max_{1\le i\le d, 1\le j_1<\ldots<j_m\le n}\|h_i(X\xi_{j_1},\ldots,X\xi_{j_m})-V_i\|_{\psi_2}^* \lesssim B_n(\log(dn))^{1/\beta} \tag{A.17}$$

up to some universal constant, where $\|Z\|_{\psi_2}^* := \inf\{\nu>0 : \mathbb{E}^*[\psi_\beta(|Z|/\nu)]\le 1\}$ denotes the Orlicz-Norm (of a real-valued random variable Z) with respect to the conditional expectation $\mathbb{E}^*$.

Next we observe by the triangle inequality that

$$\begin{aligned}|\mathcal{T}_n^* - S_n^*| \lesssim &\Big|\max_{1\le i\le d}V_{i,\Delta}\frac{U_i^*-V_i}{\hat\sigma_i\Delta} - \sqrt{n}\max_{1\le i\le d}V_i\frac{U_i^*-V_i}{m\sqrt{\zeta_{1,i}}\Delta}\Big| \\ &+\Big|\mathcal{T}_n^* - \max_{1\le i\le d}V_{i,\Delta}\frac{U_i^*-V_i}{\hat\sigma_i\Delta}\Big| + \Big|\sqrt{n}\max_{1\le i\le d}V_i\frac{U_i^*-V_i}{m\sqrt{\zeta_{1,i}}\Delta} - S_n^*\Big|.\end{aligned} \tag{A.18}$$

For the second summand we have

$$0\le \mathcal{T}_n^* - \max_{1\le i\le d}V_{i,\Delta}\frac{U_i^*-V_i}{\hat\sigma_i\Delta} \le \max_{1\le i\le d}\frac{(U_i^*-V_i)^2}{2\hat\sigma_i\Delta} \tag{A.19}$$

and, provided that $|\theta_i| \leq \Delta$ for all i, we claim

$$\left|\max_{1\leq i\leq d} V_{i,\Delta}\frac{U_i^*-V_i}{\hat\sigma_i\Delta} - \sqrt{n}\max_{1\leq i\leq d} V_i\frac{U_i^*-V_i}{m\sqrt{\zeta_{1,i}}\Delta}\right| \lesssim B_n^3\frac{(\log(nd))^{1+1/\beta}}{\sqrt{n}}\,, \tag{A.20}$$

$$\left|\max_{1\leq i\leq d}\frac{(U_i^*-V_i)^2}{2\hat\sigma_i\Delta} - \sqrt{n}\max_{1\leq i\leq d}\frac{(U_i^*-V_i)^2}{2m\sqrt{\zeta_{1,i}}\Delta}\right| \lesssim B_n^4\frac{(\log(nd))^{3/2+2/\beta}}{n} \tag{A.21}$$

with $\mathbb{P}$-probability at least $1-o_K(1)$, where the constants in the inequalities depend only on β,γ and $\underline{b}$. This will help bounding the right hand term in (A.19) while simultaneously taking care of the first summand in (A.18). In view of (A.17), an application of Lemma B.18 yields for the vector $U^*=(U_1^*,\dots,U_d^*)^\top$, that

$$\|U^*-V\|_\infty \lesssim B_n\frac{(\log(nd))^{1/2+1/\beta}}{\sqrt{n}} \tag{A.22}$$

with $\mathbb{P}^*$ probability at least $1-o_K(1)$ with the bound depending only on β and γ. Since by Assumption $\min\limits_{1\leq i\leq d, |\theta_i|>c}\zeta_{1,i}\geq \underline{b}$ this gives

$$\sqrt{n}\max_{1\leq i\leq d}\frac{(U_i^*-V_i)^2}{2m\sqrt{\zeta_{1,i}}\Delta} \lesssim B_n^2\frac{(\log(nd))^{1+2/\beta}}{\sqrt{n}} \tag{A.23}$$

with $\mathbb{P}^*$ probability at least $1-o_K(1)$. Combining (A.21) and (A.23) yields

$$\begin{aligned}\left|\mathcal{T}_n^* - \max_{1\leq i\leq d} V_{i,\Delta}\frac{U_i^*-V_i}{\hat\sigma_i\Delta}\right| &\lesssim B_n^2\frac{(\log(nd))^{1+2/\beta}}{\sqrt{n}} + B_n^4\frac{(\log(nd))^{3/2+2/\beta}}{n}\\ &\lesssim B_n^2\frac{(\log(nd))^{1+2/\beta}}{\sqrt{n}}\,.\end{aligned} \tag{A.24}$$

The estimate (A.20) is obtained as follows. First, we use the inequality

$$\begin{aligned}&\sqrt{n}\left|\max_{1\leq i\leq d} V_{i,\Delta}\frac{U_i^*-V_i}{\sqrt{n}\hat\sigma_i\Delta} - \max_{1\leq i\leq d} V_i\frac{U_i^*-V_i}{m\sqrt{\zeta_{1,i}}\Delta}\right|\\ &\qquad\lesssim \sqrt{n}\,\|V-V_\Delta\|_\infty\,\|U^*-V\|_\infty \max_{1\leq i\leq d}\frac{\left|\sqrt{n}\hat\sigma_i - m\zeta_{1,i}\right|}{\sqrt{n}\hat\sigma_i}\,.\end{aligned}$$

Secondly, we use (A.22) and Theorem A.1 to bound the terms involving U^* and $\hat\sigma_i$. Recalling (2.21), we get

$$|V_i - V_{i,\Delta}| = \begin{cases}0 & \text{if } |V_i|\leq\Delta\\ |V_i-\Delta| & \text{otherwise}\,.\end{cases}$$

As long as $|\theta_i|\leq\Delta$ and when (A.22) holds we can bound the latter quantity uniformly by $\|V-\theta\|_\infty$ so that Lemma A.6 is applicable to derive (A.20) with the hidden constants depending only on β,γ and $\underline{b}$. The bound (A.21) is obtained similarly.

For the last term on the right-hand side of (A.18) we observe that

$$\begin{aligned}\left|\sqrt{n}\max_{1\leq i\leq d} V_i\frac{U_i^*-V_i}{m\sqrt{\zeta_{1,i}}\Delta} - S_n^*\right| &\lesssim \sqrt{n}\,\|V-\theta\|_\infty\,\|U^*-V\|_\infty\\ &\lesssim B_n^2\frac{(\log(nd))^{1+1/\beta}}{\sqrt{n}}\end{aligned} \tag{A.25}$$

with $\mathbb{P}$-probability at least $1-o_K(1)$ by virtue of Lemma A.6.
Combining (A.20), (A.24) and (A.25) yields, up to some constant depending only on $\gamma, \beta, \underline{b}$, that

$$\left|\mathcal{T}_n^* - \sqrt{n}\max_{1\le i\le d} V_i \frac{U_i^* - V_i}{m\sqrt{\zeta_{1,i}}\Delta}\right| \lesssim B_n^2 \frac{(\log(nd))^{1+2/\beta}}{\sqrt{n}} + B_n^3 \frac{(\log(nd))^{1+1/\beta}}{\sqrt{n}}$$

with $\mathbb{P}^*$ probability at least $1-o_K(1)$ on a set of $\mathbb{P}$-probability at least $1-o_K(1)$.

□

In the next step we decompose the statistic U^* into a linear and a non-linear part. The linear part of the Hoeffding decomposition (for more details see Hoeffding (1948b)) of U^* conditional on $X_1,\dots,X_n$ is given by

$$h_1^X(\xi_1) = \mathbb{E}^*[h(X\xi_1,\dots,X\xi_m)|\xi_1] = \frac{1}{n^{m-1}} \sum_{l_1,\dots,l_{m-1}=1}^{n} h(X\xi_1, X_{l_1},\dots,X_{l_{m-1}})\,.$$

To proceed we need the notation

$$S_{n,1}^* = \sqrt{n}\max_{1\le i\le d} \theta_i \frac{\frac{1}{n}\sum_{j=1}^n h_{1,i}^X(\xi_j) - V_i}{\sqrt{\zeta_{1,i}}\Delta}\,.$$

LEMMA A.8. *Under the assumptions of Theorem 2.5 we have for the statistic S_n^* in (A.15) that*

$$|S_n^* - S_{n,1}^*| \lesssim B_n \frac{(\log(nd))^{1+1/\beta}}{n^{1/2-\gamma/\beta}}$$

with $\mathbb{P}^*$-probability at least $1-n^{-\gamma/\beta}$ whenever (A.17) holds. Here the constant in the inequality depends only on β, and therefore the inequality holds uniformly over the classes $\mathcal{H}_{0,boot}(\Delta)$ and $\mathcal{H}_1$ defined in (2.26) and (2.18).

PROOF. By Theorem 5.1 in Song et al. (2019) and Markov's inequality the non-linear part of the Hoeffding decomposition is bounded by some multiple of $B_n\frac{(\log(nd))^{1+1/\beta}}{n^{1-\gamma/\beta}}$ that depends only on β with $\mathbb{P}^*$-probability at least $1-n^{-\gamma/\beta}$ whenever (A.17) holds. □

The final result of this section provides a Gaussian approximation for the statistic $S_{n,1}^*$. Note that $h_1^X(\xi_i)$ is not the bootstrap version of $h_1(X_i)$ and therefore Lemma B.3 is not applicable. Instead we will utilize a Gaussian approximation together with a bound on the distance of two Gaussian random vectors by the difference of their covariance matrices and their dimension. Recalling the definition of Γ from (A.11), we define the $d\times d$ diagonal matrix $B = \mathrm{Diag}(\zeta_{1,1}^{-1/2},\dots,\zeta_{1,d}^{-1/2})$ and put

$$\begin{aligned}\hat\Gamma &:= B\mathrm{Cov}^*(h_1^X(\xi_1))B \\ &= B\Big(\frac{1}{n^{2m-1}}\sum_{l,l_1,\dots,l_{2m-2}}^{n} h(X_l, X_{l_1},\dots,X_{l_{m-1}})h(X_l,X_{l_m},\dots,X_{l_{2m-2}})^\top - VV^\top\Big)B\,,\end{aligned}$$

where Cov^* is the covariance operator with respect to the conditional expectation $\mathbb{E}^*$. $\hat\Gamma$ is a rescaled version of the (conditional) covariance matrix of the vector $h_1^X(\xi_1)$. Further, we introduce the matrices $\hat\Lambda$ and Λ with entries

$$\hat\Lambda_{ij} = \hat\Gamma_{ij}\theta_i\theta_j \quad \text{and} \quad \Lambda_{ij} = \Gamma_{ij}\theta_i\theta_j\,, \qquad i,j=1,\dots,d\,.$$

In the following discussion the symbol $a \le b$ for vectors $a, b \in \mathbb{R}^d$ means coordinate-wise inequality.

LEMMA A.9. *Let $Z \sim N(0,\Lambda)$ and $Z^X \sim N(0,\hat{\Lambda})$ conditional on $X_1,\dots,X_n$. Suppose that the assumptions of Theorem 2.5 hold and that $|\theta|_{\min} > c > 0$ for some constant c. Then we have*

$$\sup_{\mathbf{x}\in\mathbb{R}^d} \left|\mathbb{P}(Z \le \mathbf{x}) - \mathbb{P}^*(Z^X \le \mathbf{x})\right| \lesssim \left(\frac{B_n^2 (\log(nd))^5}{n}\right)^{1/6}$$

with $\mathbb{P}$-probability at least $1-o_K(1)$. Additionally, whenever (A.17) holds, we have

$$\sup_{x\in\mathbb{R}} \left|\mathbb{P}^*(Z^X \le (x,...,x)^\top) - \mathbb{P}^*(\Delta\, S^*_{n,1} \le x)\right| \lesssim \left(\frac{B_n^2 (\log(nd))^{5+\frac{2}{\beta}}}{n}\right)^{1/4}. \tag{A.26}$$

The constants in both inequalities depend only on β and γ. Therefore, both inequalities hold uniformly in the subsets of the classes $\mathcal{H}_{0,boot}(\Delta)$ and $\mathcal{H}_1$, defined in (2.26) and (2.18), for which $|\theta|_{\min} > c > 0$.

PROOF. We employ a decomposition into U-statistics of orders up to $2m-1$

$$\begin{aligned}&\frac{1}{n^{2m-1}} \sum_{l,l_1,\dots,l_{2m-2}}^{n} h(X_l, X_{l_1},\dots,X_{l_{m-1}}) h(X_l, X_{l_m},\dots,X_{l_{2m-2}})^\top \\ &= \frac{1}{n^{2m-1}} \sum_{l\neq l_1\neq\dots\neq l_{2m-2}}^{n} h(X_l, X_{l_1},\dots,X_{l_{m-1}}) h(X_l, X_{l_m},\dots,X_{l_{2m-2}})^\top + R_n\,,\end{aligned}$$

where the term R_n contains all sums, where at least two of the indices l_i and l_j $(i\neq j)$ coincide (compare with the the proof of Lemma A.6). We then apply Lemma B.18 to each U-statistic appearing in the above decomposition to obtain, up to some constant depending only on γ and β, that

$$\max_{1\le i,j\le d} \left|\hat{\Gamma}_{ij} - \Gamma_{ij}\right| \lesssim B_n \sqrt{\frac{\log(nd)}{n}}$$

with probability at least $1-o_K(1)$. Finally, we use the Gaussian to Gaussian comparison from Lemma C.1 from Chen (2018) to establish the desired result.
The second bound (A.26) is an immediate consequence of Lemma B.3 in the online supplement. Note that conditions (A) and (W), which are required for Lemma B.3, are satisfied with $B_n(\log(nd))^{1/\beta}$ instead of B_n with $\mathbb{P}$-probability at least $1-o_K(1)$, which follows from similar arguments as for the first bound and the fact that $\left\|h^X_{1,i} - V_i\right\|_\infty$ is bounded by $B_n(\log(nd))^{1/\beta}$ with $\mathbb{P}$-probability at least $1-1/(nd)$.

□

A.3.2. *Proof of Theorem 2.5.* We start with the proof of (2.25). First assume that $|\theta|_{\min} > c > 0$. A combination of Lemmas A.7 and A.8 yields that under the null hypothesis

$$|S^*_{n,1} - \mathcal{T}^*_n| \lesssim c_{n,\epsilon} := B_n \frac{(\log(nd))^{1+1/\beta}}{n^{1/2-\gamma/\beta}} + \frac{B_n^3(\log(nd))^{1+1/\beta} + B_n^2(\log(nd))^{1+2/\beta}}{\sqrt{n}}$$

with $\mathbb{P}^*$-probability at least $1-o_K(1)$ on a set of $\mathbb{P}$-probability at least $1-o_K(1)$, where all involved constants depend only on β,γ,c and $\underline{b}$. We hence obtain

$$\mathbb{P}^*(S^*_{n,1} > t + c_{n,\epsilon}) - o_K(1) \le \mathbb{P}^*(\mathcal{T}^*_n > t) \le \mathbb{P}^*(S^*_{n,1} > t - c_{n,\epsilon}) + o_K(1)\,. \tag{A.27}$$

Nazarov's inequality (see for example **?**) combined with the second part of Lemma A.9 then yields, up to some constant depending only on $\beta, \gamma, \underline{b}, c$, that

$$\sup_{t\in\mathbb{R}} \left|\mathbb{P}^*(S^*_{n,1} > t \pm c_{n,\epsilon}) - \mathbb{P}^*(S^*_{n,1} > t)\right| \lesssim \left(\frac{B_n^2(\log(nd))^{5+\frac{2}{\beta}}}{n}\right)^{1/4} + c_{n,\epsilon}\sqrt{\log d}\,.$$

In conjunction with (A.27) and the first part of Lemma A.9, we obtain

$$\sup_{t\in\mathbb{R}} \left|\mathbb{P}^*(\mathcal{T}^*_n \le t) - \mathbb{P}(Z/\Delta \le (t,\dots,t)^\top)\right| \lesssim d^{(1)}_{n,\epsilon}$$

with $\mathbb{P}$-probability at least $1 - o_K(1)$, where $Z \sim N(0,\Lambda)$ and

$$d^{(1)}_{n,\epsilon} := \left(\frac{B_n^2(\log(nd))^{5+\frac{2}{\beta}}}{n}\right)^{1/4} + \left(\frac{(\log(nd))^5 B_n^2}{n}\right)^{1/6} + c_{n,\epsilon}\sqrt{\log d}.$$

We now derive a similar Gaussian approximation for the quantity

$$T_{n,\Delta} = \max_{1\le i\le d} \frac{U_i^2 - \theta_i^2}{2\hat{\sigma}_i\Delta}\,.$$

Using Lemma A.4 as well as the same arguments as above (with Lemma B.4 replacing Nazarov's inequality), we get

$$\sup_{t\in\mathbb{R}} \left|\mathbb{P}(T_{n,\Delta} \le t) - \mathbb{P}(Z/\Delta \le (t,...,t)^\top\right| \lesssim d^{(2)}_{n,\epsilon}\,,$$

where

$$d^{(2)}_{n,\epsilon} = \left(\frac{B_n^2(\log(nd))^{4+\frac{2}{\beta}}}{n}\right)^{1/4} + \frac{B_n^3(\log(nd))^{3/2}}{\sqrt{n}} + \frac{B_n(\log(nd))^{3/2}}{n^{1/2-\gamma/\beta}}.$$

where all constants involved depend only on $\beta, \gamma, \underline{b}$ and c. Since $d^{(2)}_{n,\epsilon} \lesssim d^{(1)}_{n,\epsilon}$, we deduce

$$\sup_{t\in\mathbb{R}} |\mathbb{P}(T_{n,\Delta} \le t) - \mathbb{P}^*(\mathcal{T}^*_n \le t)| \lesssim d^{(1)}_{n,\epsilon}$$

with $\mathbb{P}$-probability at least $1 - o_K(1)$. Because $\mathcal{T}_{n,\Delta} \le T_{n,\Delta}$ this yields (2.25) in the case $|\theta|_{\min} > c$.

We conclude the proof considering the case where $|\theta|_{\min}$ is not bounded away from zero. In this case we define for some sufficiently small $c > 0$ the set $I := \{1 \le i \le d \colon |\theta_i| > c\}$. By the arguments in the proof of Theorem 2.2 we observe that

$$\mathcal{T}_{n,\Delta} = \mathcal{T}^I_{n,\Delta} := \max_{i\in I} \frac{U_i^2 - \Delta^2}{2\hat{\sigma}_i\Delta}$$

with high probability uniformly with respect to the class $\mathcal{H}_0$. Let $\mathcal{T}^{**}_n$ denote the analogue of the statistic $\mathcal{T}^*_n$ defined in (2.22), where the maximum is only taken over the set I, and denote by $\hat{q}^{**}_{1-\alpha}$ the corresponding $(1-\alpha)$-quantile. Observing that $q^*_{1-\alpha} \ge q^{**}_{1-\alpha}$, we have by the arguments given in the above paragraph and the first part of this proof that

$$\begin{aligned}\limsup_{n\to\infty} \sup_{F\in\mathcal{H}_0(\Delta)} \mathbb{P}(\mathcal{T}_{n,\Delta} \ge \hat{q}^*_{1-\alpha}) &= \limsup_{n\to\infty} \sup_{F\in\mathcal{H}_0(\Delta)} \mathbb{P}(\mathcal{T}^I_{n,\Delta} \ge \hat{q}^*_{1-\alpha}) \\ &\le \limsup_{n\to\infty} \sup_{F\in\mathcal{H}_0(\Delta)} \mathbb{P}(\mathcal{T}^I_{n,\Delta} \ge \hat{q}^{**}_{1-\alpha}) \le \alpha\,,\end{aligned}$$

which yields (2.25) and completes the proof under the null hypothesis.

Finally, we turn to the consistency part of Theorem 2.5. We have already seen in the proof of Theorem 2.4 that there exists some constant $C > 0$ such that for $\xi = \max_{1\le i\le d} \theta_i^2 - \Delta^2 = \theta_{i_0}^2 - \Delta^2$

$$\mathcal{T}_{n,\Delta} \geq O_{\mathbb{P}}\left(B_n\sqrt{\log(d)}\frac{1}{\sqrt{n}\hat{\sigma}_{i_0}}\right) + \frac{C\xi\sqrt{n}}{\sqrt{n}\hat{\sigma}_{i_0}}\,.$$

uniformly over $\mathcal{H}_1$. Note that for $\xi \downarrow 0$ we have $\xi \simeq \max_{1\le i\le d}|\theta_i| - \Delta$. On the other hand, the arguments used in the proof of Lemma A.7 show that

$$\mathcal{T}_n^* \lesssim B_n(\log d)^{1/2}(\log(nd))^{1/\beta}$$

with $\mathbb{P}^*$-probability at least $1 - o_K(1)$ on a set of $\mathbb{P}$-probability at least $1 - o_K(1)$ which implies that any fixed quantile of T_n^* is eventually bounded (up to some constant that does not change with n) by $B_n(\log d)^{1/2}(\log(nd))^{1/\beta}$ with $\mathbb{P}$-probability at least $1 - o_K(1)$. Moreover, if the kernel h in (2.3) is bounded we can obtain (A.16) without the additional factor $(\log(nd))^{1/\beta}$, which yields

$$\mathcal{T}_n^* \lesssim B_n(\log d)^{1/2}$$

and hence establishes the improved rate in Theorem 2.5 for bounded kernels.

A.3.3. *Proof of Theorem 3.5.* Let I_p be the p-dimensional identity matrix and $J_{a,b}$ the $a \times b$ matrix filled with ones and $J_p := J_{p,p}$. Let U_k be the $p \times 2$ matrix with entries $U_{k,11} = U_{k,k2} = 1$ and $U_{k,ij} = 0$ otherwise and write $e_1, \ldots, e_p$ for the canonical basis vectors of $\mathbb{R}^p$. We then define $\Sigma_{p,a} = (1-a)I_p + aJ_p$ and $C = J_2 - I_2$. Set $M_0 = \Sigma_{p,\Delta}$ and

$$M_k = (1-\Delta)I_p + \Delta J_p + \rho e_1 e_k^\top + \rho e_k e_1^\top = \Sigma_{p,\Delta} + \rho U_k C U_k^\top\,, \quad 2 \le k \le p\,,$$

where $\rho = c_0(\log(p)/n)^{1/2}$ for some small constant $c_0 = c_0(\Delta)$, which will be specified later. Note that for sufficiently small ρ, the matrices M_k are correlation matrices.
Let μ_p be the uniform measure on the set $\mathcal{F}(p) = \{M_2, \ldots, M_p\}$. We denote by $\mathbb{P}_\Sigma = \mathcal{N}_p(0,\Sigma) \otimes \ldots \otimes \mathcal{N}_p(0,\Sigma)$ the product probability measure induced by n i.i.d. p-dimensional random vectors $Z_1, \ldots, Z_n \sim \mathcal{N}_p(0,\Sigma)$ and define $\mathbb{P}_{\mu_p} = \int \mathbb{P}_\Sigma d\mu_p(\Sigma)$. Let $\mathbb{P}_0$ denote the n-fold product probability measure of $\mathcal{N}_p(0, M_0)$. By the same arguments as in the proof of Theorem 5 in Han et al. (2017), we obtain

$$\text{(A.28)}\quad \inf_{T_\alpha \in \mathcal{T}_\alpha} \sup_{\Sigma \in \mathcal{F}(p)} \mathbb{P}_\Sigma(T_\alpha \text{ does not reject } H_0) \geq 1 - \alpha - \frac{1}{2}\big(\mathbb{E}_{\mathbb{P}_0}[\mathcal{L}_{\mu_p}^2(Y)] - 1\big)^{1/2}\,,$$

where

$$\mathcal{L}_{\mu_p}(y) = \frac{\mathrm{d}\mathbb{P}_{\mu_p}}{\mathrm{d}\mathbb{P}_0}(y) = \frac{1}{p-1}\sum_{k=2}^{p}\Big[\prod_{i=1}^{n}\frac{|M_0|^{1/2}}{|M_k|^{1/2}}\exp\big(-\frac{1}{2}y_i^\top(M_k^{-1} - M_0^{-1})y_i\big)\Big]$$

with $|M_k|$ being the determinant of M_k. Squaring and taking expectations yields

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_0}\big[\mathcal{L}_{\mu_p}^2(Y)\big] = \frac{1}{(p-1)^2}\sum_{k,l=2}^{p}\mathbb{E}_{\mathbb{P}_0}\Big[\prod_{i=1}^{n}&\frac{|M_0|^{1/2}}{|M_k|^{1/2}}\frac{|M_0|^{1/2}}{|M_l|^{1/2}} \\ &\times \exp\big(-\frac{1}{2}Y_i^\top(M_k^{-1} + M_l^{-1} - 2M_0^{-1})Y_i\big)\Big]\,,\end{aligned}$$

where $Y = (Y_1, ..., Y_n)$ and the random vectors $Y_1, \ldots, Y_n$ are independent with distribution $\mathbb{P}_0 = \mathcal{N}_p(0, M_0)$. By definition, the matrix M_k is a rank two perturbation of M_0 and thus we can obtain its inverse by the Woodbury matrix identity. Lengthy but straightforward calculations then yield

$$M_0^{-1} - M_k^{-1} = \frac{1}{a(a+2b)} \left[(u-v)U_k + vJ_{p,2}\right] \left((a+2b)I_2 - bJ_2\right) \left[(u-v)U_k^\top + vJ_{2,p}\right],$$

where

$$u := \frac{1}{1-\Delta}\Big(1 - \frac{\Delta}{1+(p-1)\Delta}\Big), \qquad v := \frac{-\Delta}{(1-\Delta)(1+(p-1)\Delta)},$$
$$a := \frac{1}{1-\Delta} - \frac{1}{\rho}, \quad b := \frac{1}{\rho} - \frac{\Delta}{(1-\Delta)(1+(p-1)\Delta)}.$$

Denoting $T^{kl} = M_k^{-1} + M_l^{-1} - 2M_0^{-1}$ we have by standard results on the moment generating function of a Gaussian quadratic form that

$$\mathbb{E}\left[\exp\left(-\frac{1}{2} Y_i^\top T^{kl} Y_i\right)\right] = \left|I_d + T^{kl}\Sigma_{p,\Delta}\right|^{-1/2}.$$

We will show below that these determinants attain only two values depending on whether $k = l$ or $k \neq l$. Hence, observing that $|M_k| = |M_2|$ for $k = 2, \ldots, d$ we obtain

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_0}\left[\mathcal{L}_{\mu_p}^2(Y)\right] &= \frac{1}{p-1}\prod_{i=1}^n \frac{|M_0|}{|M_2|}\left|I_p + T^{22}\Sigma_{p,\Delta}\right|^{-1/2} \\ &+ \frac{p-2}{p-1}\prod_{i=1}^n \frac{|M_0|}{|M_2|}\left|I_p + T^{23}\Sigma_{p,\Delta}\right|^{-1/2} =: A_{11} + A_{22}.\end{aligned} \tag{A.29}$$

We now investigate the different terms separately. First we consider the ratio $|M_0|/|M_2|$ which appears in both terms in (A.29). Using the fact that the eigenvalues of an equicorrelation matrix $\Sigma_{p,a}$ are $1-a$ with multiplicity $p-1$ and $1+(p-1)a$ with multiplicity 1 we have

$$|M_0| = (1-\Delta)^{p-1}(1+(p-1)\Delta).$$

For M_2 we have the block decomposition

$$M_2 = \begin{pmatrix} \Sigma_{2,\Delta+\rho} & \Delta J_{2,p-2} \\ \Delta J_{p-2,2} & \Sigma_{p-2,\Delta} \end{pmatrix}$$

from which we deduce that

$$\begin{aligned}|M_2| &= \left|\Sigma_{2,\Delta+\rho}\right| \left|\Sigma_{p-2,\Delta} - \Delta^2 J_{2,p-2}\Sigma_{2,\Delta+\rho}^{-1} J_{p-2,2}\right| \\ &= \left|\Sigma_{2,\Delta+\rho}\right| \left|\frac{\Delta(1-\Delta+\rho)}{1+\Delta+\rho} J_{p-2} + (1-\Delta)I_{p-2}\right| \\ &= (1-\Delta-\rho)(1+\Delta+\rho)(1-\Delta)^{p-3}\left[(p-2)\frac{\Delta(1-\Delta+\rho)}{1+\Delta+\rho} + 1 - \Delta\right].\end{aligned}$$

Hence, we get

$$\frac{|M_0|}{|M_k|} = \frac{(1-\Delta)^2(1+(p-1)\Delta}{(1-\Delta-\rho)(1+\Delta+\rho)\left[(p-2)\frac{\Delta(1-\Delta+\rho)}{1+\Delta+\rho} + 1 - \Delta\right]}. \tag{A.30}$$

Next we consider the determinant involving the matrix T^{kk}. We start by observing that

$$T^{kk}\Sigma_{p,\Delta} = 2[(u-v)U_k + vJ_{p,2}]MU_k^\top,$$

where $M := \frac{-1}{a(a+2b)}((a+2b)I_2 - bJ_2)$. An application of the Weinstein–Aronszajn identity Akritas et al. (1996) then yields

$$\begin{aligned}|I_p + T^{kk}\Sigma_{p,\Delta}| &= |I_2 + 2MU_k^\top[(u-v)U_k + vJ_{p,2}]| \\ &= |I_2 + 2M[(u-v)I_2 + vJ_2]| \\ &= \left|I_2 - \frac{2}{a(a+2b)}[(a+2b)(u-v)I_2 + ((a+2b)v - b(u+v))J_2]\right| \\ &= |q_1 I_2 + q_2 J_2| = (2q_2 + q_1)q_1\,, \end{aligned} \tag{A.31}$$

where

$$\begin{aligned} q_1 &= 1 - \frac{2}{a}(u-v) = \frac{1+\rho-\Delta}{1-\rho-\Delta}\,, \\ q_2 &= \frac{-2v}{a} + \frac{2b(u+v)}{a(a+2b)} = \frac{\rho(1-\Delta)2((p-1)\Delta+1)}{(\rho-1+\Delta)((-p+1)\Delta^2 + ((p-3)\rho + d - 2)\Delta + \rho + 1)}. \end{aligned}$$

Combining (A.30) and (A.31) then yields

$$\begin{aligned}\log A_{11} &= \log\Big[\frac{1}{p-1}\prod_{i=1}^n \frac{|M_0|}{|M_2|}\,\big|I_p + T^{22}\Sigma_{p,\Delta}\big|^{-1/2}\Big] \\ &= -\log(p-1) + \frac{n}{2}\Big[4\log(1-\Delta) + 2\log(1+(p-1)\Delta) \\ &\quad - \log((1-\Delta)^2 - \rho^2) - \log((1+(p-2)\Delta(1-\Delta) - \Delta^2)^2 - ((p-3)\Delta+1)^2\rho^2)\Big] \\ &= -\log(p-1) + \frac{n}{2}\Big[\frac{-C}{p} + o(p^{-1}) + \frac{2\rho^2}{(1-\Delta)^2}\Big]\,, \qquad n\to\infty\,, \end{aligned}$$

where we used a Taylor expansion for $\log(1+x)$ in the last step (assuming that $\rho\to 0$) and C is some positive constant. Therefore we obtain

$$A_{11} = o(1) \tag{A.32}$$

if we choose $\rho^2 = c_0^2\log(p)/n$, where the constant c_0 satisfies $c_0 < 1-\Delta$.
For the determinant involving T^{kl} in the A_{22} term in (A.29) we obtain by straightforward calculations that

$$\begin{aligned}|I_p + T^{kl}\Sigma_{d,\Delta}| &= |I_p + T^{23}\Sigma_{p,\Delta}| \\ &= \left|I_3 - \frac{1}{a(a+2b)}\begin{pmatrix} 2[u(a+b)-vb] & v(a+b)-ub & v(a+b)-ub \\ v(a+b)-ub & u(a+b)-vb & av \\ v(a+b)-ub & av & u(a+b)-vb\end{pmatrix}\right|. \end{aligned}$$

Tedious but straightforward calculations yield

$$\Big(\frac{|M_0|}{|M_2|}\Big)^2 |I_p + T^{kl}\Sigma_{p,\Delta}|^{-1} = \frac{f}{g}\,,$$

where

$$\begin{aligned} f &= ((-p+1)\Delta^2 + ((\rho+1)p - 3\rho - 2)\Delta + \rho + 1)(1+(p-1)\Delta)^2(-1+\Delta)^4\,, \\ g &= \Big(1 + (p-1)^2\Delta^4 + (-2p^2+6p-4)\Delta^3 + (\rho^2 p + p^2 - 3\rho^2 - 6p + 6)\Delta^2 + (\rho^2 + 2p - 4)\Delta\Big) \\ &\quad \times (1-\Delta+\rho)\Big(1 + (p-1)\Delta^3 + (-2p+2\rho+3)\Delta^2 + (\rho^2 + p - 2\rho - 3)\Delta\Big)\,. \end{aligned}$$

Once again assuming $\rho \to 0$, taking the logarithm of f/g and using the Taylor expansion of $\log(1+x)$ yields that $\log(f/g) = O(p^{-2} + \rho p^{-1})$ so that

$$A_{22} = \frac{p-2}{p-1}\left[\prod_{i=1}^{n} \frac{|M_0|}{|M_2|}\left|I_p + T^{23}\Sigma_{p,\Delta}\right|^{-1/2}\right] = \exp(o(1))(1+o(1)) = 1 + o(1)\,,$$

where we used $\log(p)n/p^2 = o(1)$. Observing (A.29) and (A.32) we obtain

$$\mathbb{E}_{\mathbb{P}_0}\left[\mathcal{L}^2_{\mu_p}(Y)\right] = A_{11} + A_{22} = 1 + o(1)$$

and the assertion of the theorem follows from (A.28), completing the proof.

APPENDIX B: FURTHER TECHNICAL DETAILS

B.1. Randomized Lindeberg Method. In this section we state two important auxiliary results (Lemmas B.3 and B.4), which will be used in the proofs of our main results in Section A. They are a consequence of a general Gaussian approximation result (Theorem B.1), which is proved in Section B.1.2 via the iterative randomized Lindeberg method.

B.1.1. *A Gaussian approximation and its consequences.* Let $V_1, \ldots, V_n, Z_1, \ldots, Z_n$ denote independent random vectors in $\mathbb{R}^d$, where $V_i = (V_{i1}, \ldots, V_{id})^\top$ and $Z_i = (Z_{i1}, \ldots, Z_{id})^\top$ for $i = 1, \ldots, n$. We also assume that the following conditions hold for the vectors $V_1, \ldots, V_n, Z_1, \ldots, Z_n$. There exists a sequence of constants B_n such that:

Condition V: There exists a constant $C_v > 0$ such that for all j

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[V_{ij}^2 + Z_{ij}^2] \leq C_v, \quad \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[V_{ij}^4 + Z_{ij}^4] \leq B_n^2 C_v\,.$$

Condition P: There exists a constant $C_p \geq 1$ such that for all i

$$\mathbb{P}\left(\|V_i\|_\infty \vee \|Z_i\|_\infty > C_p B_n (\log(dn))^{1/\beta}\right) \leq \frac{1}{n^4}\,.$$

Condition B: There exists a constant $C_b > 0$ such that for all i

$$\mathbb{E}[\|V_i\|_\infty^8 + \mathbb{E}[\|Z_i\|_\infty^8] \leq C_b B_n^8 (\log(dn))^{8/\beta}\,.$$

Condition A: There exists a constant $C_a > 0$ such that for all $(y,t) \in \mathbb{R}^d \times \mathbb{R}_+$, we have

$$\mathbb{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i \leq y + t\right) - \mathbb{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i \leq y\right) \leq C_a t \sqrt{\log d}\,.$$

Here $y + t$ means addition of t to every component of y. The following result, which will be proved in Section B.1.2, will be crucial for Lemmas B.3 - B.4. Its proof uses distributional approximations via the Iterative Randomized Lindeberg Method and is structurally the same as in Chernozhukov et al. (2019). However, we require a weaker decay in the tails at the cost of a weaker bound.

THEOREM B.1 (Iterative Randomized Lindeberg Method). *Suppose that conditions V,P,B and A are satisfied. In addition, suppose that for some positive constant C_m*

$$\max_{1 \leq j,k \leq d}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbb{E}[V_{ij}V_{ik}] - \mathbb{E}[Z_{ij}Z_{ik}])\right| \leq C_m B_n (\log(dn))^{1/\beta}\,,$$

$$\max_{1 \leq j,k,l \leq d}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbb{E}[V_{ij}V_{ik}V_{il}] - \mathbb{E}[Z_{ij}Z_{ik}Z_{il}])\right| \leq C_m B_n^2 (\log(dn))^{2/\beta}\,.$$

Then it holds

$$\sup_{y\in\mathbb{R}^d}\left|\mathbb{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n V_i\le y\right)-\mathbb{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i\le y\right)\right|\le C\left(\frac{B_n^2(\log(dn))^{4+2/\beta}}{n}\right)^{1/4},$$

where $C>0$ is a constant depending only on C_v, C_p, C_b, C_a, C_m.

Theorem B.1 has several important consequences, which are now stated in Lemma B.3 and Lemma B.4 and used in the proofs in Section A. For a precise formulation we require the following assumptions.
Let $X_1,\dots,X_n\in\mathbb{R}^d$ denote i.i.d. centred random vectors, $X_i=(X_{i1},\dots,X_{id})^\top$, satisfying the following Assumptions:

(A): There exists a sequence of constants $(B_n)_{n\in\mathbb{N}}$ such that for $1\le j\le d$ we have $\|X_{1j}\|_{\psi_\beta}\lesssim B_n$ for some $0<\beta\le 2$.
(W): There exist constants $\sigma_{min}>0$ and $D>0$ such that for all j

$$\sigma_{min}\le\frac{1}{n}\sum_{i=1}^n\mathbb{E}[X_{ij}^2]\le D\quad\text{and}\quad\frac{1}{n}\sum_{i=1}^n\mathbb{E}[X_{ij}^4]\le B_n^2D\,.$$

We begin with a result describing the deviation between the empirical moments of the centered vectors

$$\tilde{X}_k=(\tilde{X}_{k1},\dots,\tilde{X}_{kd})^\top:=X_k-\bar{X}=(X_{k1}-\bar{X}_1,\dots,X_{kd}-\bar{X}_d)^\top,$$

where $\bar{X}_j=\frac{1}{n}\sum_{i=1}^n X_{ij}$, and the covariance matrix $\mathbb{E}[X_kX_k^\top]$.

LEMMA B.2. *Suppose that assumptions (A) and (W) hold. Then there exists a universal constant $c>0$, constants $C,D>0$ and $n_0\in\mathbb{N}$ depending only on β, σ_{min} and B_n such that for all $n\ge n_0$ the inequality*

$$B_n^2(\log(dn))^{4+2/\beta}\le cn$$

implies that the inequalities

$$\frac{\sigma_{min}}{2}\le\frac{1}{n}\sum_{i=1}^n\tilde{X}_{ij}^2\le D,\tag{B.1}$$

$$\frac{1}{n}\sum_{i=1}^n\tilde{X}_{ij}^4\le B_n^2D,\tag{B.2}$$

$$\max_{1\le k,j\le p}\left|\frac{1}{\sqrt{n}}\sum_{k=1}^n(\tilde{X}_{ik}\tilde{X}_{ij}-E[X_{ik}X_{ij}])\right|\le CB_n(\log(dn))^{1/\beta},\tag{B.3}$$

$$\max_{1\le k,j,l\le p}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(\tilde{X}_{ik}\tilde{X}_{ij}\tilde{X}_{il}-E[X_{ik}X_{ij}X_{il}])\right|\le CB_n^2(\log(dn))^{2/\beta}\tag{B.4}$$

hold jointly with probability at least $1-1/n$.

PROOF. Let $A=5L(C_1+C_2)$ for some C_1,C_2 to be specified later and denote by $\mathcal{A}$ the event that the inequalities

$$\max_{1\le k\le d}\left|\frac{1}{\sqrt{n}}\sum_{k=1}^n X_{ik}\right|\le A\sqrt{\log(dn)},$$

$$\max_{1\le k,j\le d}\left|\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_{ik}X_{ij}-E[X_{ii}X_{ij}])\right|\le AB_n(\log(dn))^{1/\beta},$$

$$\max_{1\le k,j,l\le d}\left|\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_{ik}X_{ij}X_{kl}-E[X_{ii}X_{ij}X_{il}])\right|\le AB_n^2(\log(dn))^{2/\beta},$$

$$\max_{1\le k,j,l,r\le d}\left|\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_{ik}X_{ij}X_{kl}X_{ir}-E[X_{ii}X_{ij}X_{il}X_{ir}])\right|\le AB_n^3(\log(dn))^{3/\beta}$$

hold jointly. Noting that

$$\begin{aligned}&\max_{1\le k,j\le d}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\tilde{X}_{ik}\tilde{X}_{ij}-E[X_{ii}X_{ij}])\right|\\&\le\max_{1\le k,j\le d}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_{ik}X_{ij}-E[X_{ii}X_{ij}])\right|+\sqrt{n}\max_{1\le k\le d}|\bar{X}_k|^2\end{aligned}$$

and

$$\begin{aligned}&\max_{1\le k,j,l\le d}\left|\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(\tilde{X}_{ik}\tilde{X}_{ij}\tilde{X}_{il}-E[X_{ik}X_{ij}X_{il}])\right|\\&\le\max_{1\le k,j,l\le d}\left|\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_{ik}X_{ij}X_{il}-E[X_{ik}X_{ij}X_{il}])\right|\\&+2\sqrt{n}\max_{1\le k\le d}|\bar{X}_k|^3+\max_{1\le k,j,l\le d}|\bar{X}_l|\left|\frac{3}{\sqrt{n}}\sum_{i=1}^{n}X_{ik}X_{ij}\right|\end{aligned}$$

yields the bounds (B.3) and (B.4) on $\mathcal{A}$. Considering

$$\begin{aligned}\max_{1\le k\le d}\left|\frac{1}{n}\sum_{i=1}^{n}(\tilde{X}_{ik}^2-\mathbb{E}[X_{ik}^2])\right|&\le\max_{1\le k,j\le d}\left|\frac{1}{n}\sum_{i=1}^{n}(\tilde{X}_{ik}\tilde{X}_{ij}-\mathbb{E}[X_{ik}X_{ij}])\right|\\&\le\frac{CB_n(\log(dn))^{1/\beta}}{\sqrt{n}}\le\frac{\sigma_{min}}{2}\end{aligned}$$

yields (B.1) on $\mathcal{A}$, and (B.2) follows by similar considerations.

We now show that we can find C_1, C_2 such that $\mathcal{A}$ has probability at least $1-1/n$. Fix $m\in\{1,2,3,4\}$ and let $P=\{1,\dots,d\}^m$. We denote $y^h=y_{h_1}\dots y_{h_m}$ for any $y=(y_1,\dots,y_d)^\top\in\mathbb{R}^d$ and $h=(h_1,\dots,h_m)^\top\in P$. As X_{ij} have ψ_β-norms uniformly bounded by B_n we obtain by standard calculations and (W) that

$$\max_{h\in P}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(X_i^h-\mathbb{E}[X_i^h])^2]\le\max_{h\in P}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(X_i^h)^2]\le C_1^2B_n^{2(m-1)}\log(dn)^{2(m-2)/\beta\vee 0},$$

where the constant C_1 depends only on β and D. By Lemma B.11 and Lemma B.12 in Section B.2 we obtain that

$$\mathbb{E}\left[\max_{1\le i\le n,h\in P}(X_i^h-\mathbb{E}[X_i^h])^2\right]\le C_2^2B_n^{2m}(\log(nd))^{2m/\beta},$$

where the constant C_2 depends only on β, m and K. Therefore, it follows from Lemma B.16 that

$$\begin{aligned}
&\mathbb{E}\left[\max_{h\in P}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i^h-\mathbb{E}[X_i^h])\right|\right] \\
&\leq L\left(C_1 B_n^{m-1}\log(nd)^{((m-2)/\beta+1/2)\vee 1/2}+\frac{C_2(\log(nd))^{m/\beta+1}}{\sqrt{n}}\right) \\
&\leq L(C_1+C_2)B_n^{m-1}(\log(nd))^{(m-1)/\beta\vee 1/2}\,.
\end{aligned}$$

Now applying Lemma B.16 with $t=3L(C_1+C_2)B_n^{m-1}(\log(nd))^{(m-1)/\beta\vee 1/2}$, $\nu=1$ and $\beta=\frac{2m}{\beta}$ we obtain for $n\geq n_0$ (where n_0 depends only on β, m, K and L) that

$$\max_{h\in P}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i^h-\mathbb{E}[X_i^h])\right| > 5L(C_1+C_2)B_n^{m-1}(\log(nd))^{(m-1)/\beta\vee 1/2}$$

with probability at most

$$\frac{1}{(dn)^3}+3\exp\left(-C\left(\frac{\sqrt{n}B_n^{m-1}(\log(nd))^{(m-1)/\beta\vee 1/2}}{B_n^m(\log(nd))^{2m/\beta}}\right)^{\beta/(2m)}\right)\leq\frac{1}{(nd)^3}+\frac{3}{n^3}\leq\frac{1}{n}\,.$$

Here C is a constant, which depends only on β, m, K and L and we have used that $B_n^2(\log(nd))^{4+2/\beta}\leq n$ for the first inequality.

□

We note that, using Lemma B.12, Assumptions (A) and (W) imply Conditions A,B, P and V and therefore Theorem B.1 is applicable in the following discussion. We start with a preliminary result regarding the quantities

$$T_n=\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i \quad\text{and}\quad T_n^*=\frac{1}{\sqrt{n}}\sum_{i=1}^n e_k(X_i-\bar{X})\,. \tag{B.5}$$

Here $e_1,\dots,e_n$, which we will sometimes call multipliers, are independent random variables (independent of $X=(X_1,\dots,X_n)$) such that $e_i=e_{i,1}+e_{i,2}$, where $e_{i,1}$ and $e_{i,2}$ are independent, $e_{i,1}\sim N(0,\sigma^2)$ and $e_{i,2}$ has a two point distribution with $\mathbb{E}[e_i]=0$, $\mathbb{E}[e_i^2]=\mathbb{E}[e_i^3]=1$ (see Lemma 7.3 in Chernozhukov et al. (2019) for more details and note that σ can be chosen universally).

LEMMA B.3. *Suppose that Conditions (A) and (W) hold, then, with probability at least $1-2/n$, we have*

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}(T_n\leq x)-\mathbb{P}(T_n^*\leq x|X)\right|\leq C\left(\frac{B_n^2(\log(nd))^{4+2/\beta}}{n}\right)^{1/4}, \tag{B.6}$$

where the constant C only depends on σ_{min}, β.
Further, let T_n^G denote the analogue of the statistic T_n in (B.5), where the random variables X_k have been replaced by independent zero mean Gaussian vectors with the same covariance structure. Then

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}(T_n\leq x)-\mathbb{P}(T_n^G\leq x)\right|\leq C\left(\frac{B_n^2(\log(nd))^{4+2/\beta}}{n}\right)^{1/4}. \tag{B.7}$$

PROOF. First assume that $Y_1, \dots, Y_n$ are vectors in $\mathbb{R}^d$ such that

$$\max_{1\le k\le d} \|Y_k\|_\infty \le KB_n \left(5\log(dn)\right)^{1/\beta} ,$$

$$\frac{\sigma_{min}}{2} \le \frac{1}{n}\sum_{i=1}^n Y_{ik}^2 \le D \,,$$

$$\frac{1}{n}\sum_{i=1}^n Y_{ik}^4 \le B_n^2 D \,,$$

$$\max_{1\le k,j\le d} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_{ik}Y_{ij} - E[X_{ik}X_{ij}]) \right| \le C_m B_n (\log(dn))^{1/\beta} \,,$$

$$\max_{1\le k,j,l\le d} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_{ik}Y_{ij}Y_{il} - E[X_{ik}X_{ij}X_{il}]) \right| \le C_m B_n^2 (\log(dn))^{2/\beta} \,.$$

Recall the definition of the multipliers in the paragraph following equation (B.5). We will apply Theorem B.1 with $Z_i = e_i Y_i$ and $V_i = X_i$. The Conditions V, P and B follow immediately from the properties of Y with C_v, C_p, C_b only depending on σ_{min} and B_n. Condition A follows from the Gaussianity of $e_{i,1}$ with C_a depending only on σ_{min} and σ (first condition on $e_{i,2}$ and then use Lemma 8.3 from Chernozhukov et al. (2019)). The remaining conditions in Theorem B.1 follow easily from the properties of e_i and Y_i. We hence obtain

$$\sup_{y\in\mathbb{R}} \left| \mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \le y \right) - \mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n e_i Y_i \le y \right) \right| \le K_2 \left(\frac{B_n^2 (\log(nd))^{4+2/\beta}}{n} \right)^{1/4} ,$$

where K_2 depends only on σ_{min}. By Lemma B.12 and Lemma B.2 the random vectors $X_i - \bar{X}$ satisfy the assumptions stated for the vectors Y_i (with probability close to 1), and we obtain

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(T_n \le x) - \mathbb{P}(T_n^* \le x | X)| \le K_2 \left(\frac{B_n^2 (\log(nd))^{4+2/\beta}}{n} \right)^{1/4}$$

with probability at least $1 - 2/n$, establishing (B.6). For the second inequality (B.7) we define

$$R_n^* := \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{e}_i (X_i - \bar{X}) \,,$$

where the multipliers $\tilde{e}_1, \dots, \tilde{e}_n$ are now chosen as in Corollary 5.2 of Chernozhukov et al. (2019), with $v = 0, \alpha = 1/2$ and $\beta = 3/2$. More precisely, we sample $\tilde{e}_i$ independently from the distribution that is given by $4\nu - 1$ where $\nu \sim \text{Beta}(1/2, 3/2)$. Note that $\mathbb{E}[\tilde{e}_i] = 0$ and $\mathbb{E}[\tilde{e}_i^2] = 1$. We then obtain by similar arguments as above that

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(T_n \le x) - \mathbb{P}(R_n^* \le x | X)| \le C \left(\frac{B_n^2 (\log(nd))^{4+2/\beta}}{n} \right)^{1/4}$$

with probability at least $1 - 2/n$. We let A_n be the event that the first three inequalities in Lemma B.2 hold. Then $\mathbb{P}(A_n) \ge 1 - 1/n$. On A_n we may apply first Corollary 5.2 of Chernozhukov et al. (2019) which gives

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(T_n^{\tilde{G}} \le x | X) - \mathbb{P}(R_n^* \le x | X)| \le C \left(\frac{B_n^2 (\log(nd))^5}{n} \right)^{1/4} ,$$

where $T_n^{\tilde{G}}$ is defined analogously to T_n^G for a certain Gaussian process $\tilde{G}$. The Gaussian to Gaussian comparison from Corollary 5.1 of Chernozhukov et al. (2019) then yields that on A_n we further have

$$\sup_{x\in\mathbb{R}}|\mathbb{P}(T_n^{\tilde{G}} \leq x|X) - \mathbb{P}(T_n^G \leq x)| \leq C\left(\frac{B_n^2(\log(nd))^{4+2/\beta}}{n}\right)^{1/4}$$

as we can bound $\|\mathrm{vech}(\Sigma_G) - \mathrm{vech}(\Sigma_{\tilde{G}})\|_\infty$ by $B_n \log(dn)^{1/\beta}$ due to (B.3). □

LEMMA B.4. *Suppose that Conditions (A) and (W) hold. Then for any $x \in \mathbb{R}$ and $t > 0$ we have*

$$\mathbb{P}(T_n \leq x+t) - \mathbb{P}(T_n \leq x) \leq C\left(t\sqrt{\log d} + \left(\frac{B_n^2(\log(nd))^{4+2/\beta}}{n}\right)^{1/4}\right).$$

PROOF. For some constant C only depending on σ_{min} and B_n we get

$$\begin{aligned}\mathbb{P}(T_n \leq x+t) - \mathbb{P}(T_n \leq x) &\leq \left|\mathbb{P}(T_n \leq x+t) - \mathbb{P}(T_n^G \leq x+t)\right| \\ &+ \left|\mathbb{P}(T_n \leq x) - \mathbb{P}(T_n^G \leq x)\right| + \left|\mathbb{P}(T_n^G \leq x+t) - \mathbb{P}(T_n^G \leq x)\right| \\ &\leq 2C\left(\frac{B_n^2(\log(nd))^{4+2/\beta}}{n}\right)^{1/4} + Ct\sqrt{\log d},\end{aligned}$$

where for the last line we used (B.7) and the Gaussian anti-concentration property from (Chernozhukov et al., 2019, Lemma 8.3). □

B.1.2. *Proof of Theorem B.1.* We will establish Theorem B.1 via the Iterative Randomized Lindeberg Method. The proof is structurally the same as in Chernozhukov et al. (2019) but asks for weaker decay in the tails at the cost of a weaker bound. We begin by introducing some notation which we will be used in this section.
For $\epsilon \in \{0,1\}^n$ we set

$$S_{n,\epsilon}^V = \frac{1}{\sqrt{n}}\sum_{i=1}^n(\epsilon_i V_i + (1-\epsilon_i)Z_i) \quad \text{and} \quad S_n^Z = \frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i\,.$$

Let $\epsilon^0 = (1,\dots,1)$, $D = [\log(n)] + 1$ and define random vectors $\epsilon^1,\dots,\epsilon^D \in \{0,1\}^n$ such that i) $\epsilon_i^s = 0$ if $\epsilon_i^{s-1} = 0$ and ii) for $I_{s-1} = \{i = 1,\dots,n : \epsilon_i^{s-1} = 1\}$, the random variables $\{\epsilon_i^s\}_{i\in I_{s-1}}$ are exchangeable conditional on ϵ^{s-1} and satisfy

$$\mathbb{P}\left(\sum_{i\in I_{s-1}}\epsilon_i^s = k\,|\,\epsilon^{s-1}\right) = \frac{1}{1+|I_{s-1}|}\,, \qquad k = 0,\dots,|I_{s-1}|\,.$$

As remarked in Chernozhukov et al. (2019), these properties uniquely determine the joint distribution of $\epsilon^1,\dots,\epsilon^D$ which we also assume independent of $V_1,\dots,V_n,Z_1,\dots,Z_n$. For positive constants $B_{n,1,s}, B_{n,2,s}$ and

$$\begin{aligned}\mathcal{E}_{i,jk}^V &= \mathbb{E}[V_{ij}V_{ik}]\,, \quad \mathcal{E}_{i,jkl}^V = \mathbb{E}[V_{ij}V_{ik}V_{il}]\,, \\ \mathcal{E}_{i,jk}^Z &= \mathbb{E}[Z_{ij}Z_{ik}]\,, \quad \mathcal{E}_{i,jkl}^Z = \mathbb{E}[Z_{ij}Z_{ik}Z_{il}]\,,\end{aligned}$$

we denote by $\mathcal{A}_s$ the event

$$\left\{\max_{1\leq j,k\leq d}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\epsilon_i^s(\mathcal{E}_{i,jk}^V - \mathcal{E}_{i,jk}^Z)\right| \leq B_{n,1,s}\,, \max_{1\leq j,k,l\leq d}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\epsilon_i^s(\mathcal{E}_{i,jkl}^V - \mathcal{E}_{i,jkl}^Z)\right| \leq B_{n,2,s}\right\}.$$

We also fix a five times continuously differentiable and decreasing function $g_0 : \mathbb{R} \to \mathbb{R}$ such that

$$\text{i) } g_0(t) \geq 0\,, \quad \text{ii) } g_0(t) = 0 \text{ when } t \geq 1\,, \quad \text{and iii) } g_0(t) = 1 \text{ when } t \leq 0\,.$$

Clearly we can bound the first five derivatives of this function uniformly by some constant C_g. The bounds in the following proofs and results depend on the particular choice of g_0, but as we may choose some g_0 that works universally we suppress that dependence.

Next we let $\phi > 0, \beta = \phi \log d$, set $g(t) = g_0(\phi t)$ and define for $\omega \in \mathbb{R}^d$ the softmax function

$$F(w) = \beta^{-1} \log \left(\sum_{j=1}^{d} \exp(\beta w_j) \right) .$$

It is easy to check that

$$g(t) = \begin{cases} 1 & \text{if } t \leq 0 \\ 0 & \text{if } t \geq \phi^{-1} \end{cases}$$

and $\max_{1 \leq j \leq d} w_j \leq F(w) \leq \max_{1 \leq j \leq d} w_j + \phi^{-1}$. For $y \in \mathbb{R}^d$ we now define the function

$$m^y(w) = g(F(w - y))\,, \qquad w \in \mathbb{R}^d\,,$$

and its partial derivatives up to fifth order, for instance we write

$$m^y_{jklrh}(w) = \frac{\partial^5 m^y(w)}{\partial w_j \partial w_k \partial w_l \partial w_r \partial w_h}\,, \qquad j,k,l,r,h = 1,\ldots,d\,.$$

From Chernozhukov et al. (2019) we know that there exist functions $U^y_{jk}, U^y_{jkl}, U^y_{jklr}, U^y_{jklrh} : \mathbb{R}^d \to \mathbb{R}$ with the following 3 properties.

i) $|m^y_I(w)| \leq U^y_I(w)$ where I is any of the index sets $jk, jkl, jklr$ or $jklrh$.
ii) For any $w_1, w_2 \in \mathbb{R}^d$ such that $\beta \left\| w_2 \right\|_\infty \leq 1$ we have

(B.8) $$U^y_{jklr}(w_1 + w_2) \lesssim U^y_{jklr}(w_1), \quad U^y_{jklrh}(w_1 + w_2) \lesssim U^y_{jklrh}(w_1)\,.$$

iii) For the same I as in i) we have uniformly in w

(B.9) $$\sum_I U^y_I(w) \lesssim \phi^{|I|} (\log d)^{|I|-1}\,.$$

Lastly we define

$$\mathcal{I}^y := m^y(S^V_{n,\epsilon^s}) - m^y(S^Z_n)\,,$$
$$h^y(Y;x) := \mathbb{1}\{-x < \max_{1\leq j\leq d}(Y_j - y_j) \leq x\}\,, \qquad x > 0\,,$$
$$\varrho_\epsilon := \sup_{y \in \mathbb{R}^d} |\mathbb{P}(S^V_{n,\epsilon} \leq y) - \mathbb{P}(S^Z_n \leq y)|\,.$$

We now state and prove three auxiliary results, which be essential for the proof of Theorem B.1.

LEMMA B.5. *Suppose that conditions V,P,B and A are satisfied. Then for any* $d = 0, \ldots, D-1$ *and any* $\phi > 0$ *such that*

(B.10) $$C_p B_n \phi (\log(dn))^{1+1/\beta} \leq \sqrt{n}$$

on the event $\mathcal{A}_s$, we have

$$\varrho_{\epsilon^s} \lesssim \frac{\sqrt{\log d}}{\phi} + \frac{B_n^2\phi^4(\log(dn))^{3+2/\beta}}{n^2} + \left(\frac{\sqrt{\log d}}{\phi} + \mathbb{E}[\varrho_{\epsilon^{s+1}}|\epsilon^s]\right)\left(\frac{B_{n,1,s}\phi^2\log d}{\sqrt{n}} + \frac{B_{n,2,s}\phi^3(\log d)^2}{n} + \frac{B_n^2\phi^4(\log d)^3}{n}\right)$$

up to a constant depending only on C_v, C_p, C_b, C_a.

PROOF. Fix $s = 0, \dots, D-1$ and $e^s \in \{0,1\}^n$ such that if $\epsilon^s = e^s$, then $\mathcal{A}_s$ holds. All following arguments will be conditional on $\epsilon^s = e^s$, for the sake of brevity we will make this conditioning implicit and write $\mathbb{P}(\cdot)$ and $\mathbb{E}[\cdot]$ instead of $\mathbb{P}(\cdot|\epsilon^s = e^s)$ and $\mathbb{E}[\cdot|\epsilon^s = e^s]$. We denote

$$W = (W_1, \dots, W_d)^\top = \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\epsilon_i^{s+1}V_i + (1-\epsilon_i^{s+1})Z_i\right) .$$

We will split the proof into two steps and three auxiliary calculations where we prove bounds that are used in the first two steps. In the first step we establish the bound

$$\sup_{y\in\mathbb{R}^d}|\mathbb{E}[\mathcal{I}^y]| \lesssim \frac{B_n^2\phi^4(\log(dn))^{3+2/\beta}}{n^2} + \left(\frac{\sqrt{\log d}}{\phi} + \mathbb{E}[\varrho_{\epsilon^{s+1}}|\epsilon^s]\right)\left(\frac{B_{n,1,s}\phi^2\log d}{\sqrt{n}} + \frac{B_{2,n,s}\phi^3(\log d)^2}{n} + \frac{B_n^2\phi^4(\log d)^3}{n}\right)$$

and in the second step we show that

$$\varrho_{\epsilon^s} \lesssim \frac{\sqrt{\log d}}{\phi} + \sup_{y\in\mathbb{R}^d}|\mathbb{E}[\mathcal{I}^y]| \tag{B.11}$$

which then yields the desired claim.

Step 1. Let $\mathcal{S}_n$ be the set of permutations on $\{1, \dots, |I_s|\}$ and let σ be a random variable that is distributed uniformly on $\mathcal{S}_n$ and also independent of $V_1, \dots, V_n, Z_1, \dots, Z_n$ and ϵ^{s+1}. Writing

$$W_i^\sigma = \frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}V_{\sigma(j)} + \frac{1}{\sqrt{n}}\sum_{j=i+1}^{|I_s|}Z_{\sigma(j)} + \frac{1}{\sqrt{n}}\sum_{j\notin I_s}Z_j, \quad \text{for all } i = 1, \dots, |I_s|,$$

it follows by Lemma B.14 that for any function $m : \mathbb{R}^d \to \mathbb{R}$ and any $i \in I_s$,

$$\mathbb{E}[m(W)] = \mathbb{E}\left[\frac{\sigma^{-1}(i)}{|I_s|+1}m\left(W_{\sigma^{-1}(i)}^\sigma + \frac{V_i}{\sqrt{n}}\right) + (1 - \frac{\sigma^{-1}(i)}{|I_s|+1})m\left(W_{\sigma^{-1}(i)}^\sigma + \frac{Z_i}{\sqrt{n}}\right)\right]$$

Fixing some $y \in \mathbb{R}^d$ we observe that

$$\mathcal{I}^y = \sum_{i=1}^{|I_s|}\left(m\left(W_i^\sigma + \frac{V_{\sigma(i)}}{\sqrt{n}}\right) - m\left(W_i^\sigma + \frac{Z_{\sigma(i)}}{\sqrt{n}}\right)\right)$$

and let

$$f(t) = \sum_{i=1}^{|I_s|}\mathbb{E}\left(m\left(W_i^\sigma + \frac{tV_{\sigma(i)}}{\sqrt{n}}\right) - m\left(W_i^\sigma + \frac{tZ_{\sigma(i)}}{\sqrt{n}}\right)\right), \quad \text{for } t \in [0,1].$$

Clearly $\mathbb{E}[\mathcal{I}^y]=f(1)$ and by Taylor's expansion

$$f(1)=f(0)+f^{(1)}(0)+\frac{f^{(2)}(0)}{2}+\frac{f^{(3)}(0)}{6}+\frac{f^{(4)}(\bar t)}{24}$$

for some $\bar t\in(0,1)$. Clearly $f(0)=0$ and because of $\mathbb{E}[V_{ij}]=\mathbb{E}[Z_{ij}]=0$ we also obtain $f^{(1)}(0)=0$.
We defer the bounds of $|f^{(2)}(0)|, |f^{(3)}(0)|$ and $|f^{(4)}(\bar t)|$ to the three auxiliary calculations.

Step 2. We observe that

$$\begin{aligned}\mathbb{P}(S^V_{n,\epsilon^s}\le y)&\le \mathbb{P}(F(S^V_{n,\epsilon^s}-y-\phi^{-1})\le 0)\le \mathbb{E}[m^{y+\phi^{-1}}(S^V_{n,\epsilon^s})]\\&\le \mathbb{E}[m^{y+\phi^{-1}}(S^Z_n)]+|E[\mathcal{I}^{y+\phi^{-1}}]|\le \mathbb{P}(S^Z_n\le y+2\phi^{-1})+|E[\mathcal{I}^{y+\phi^{-1}}]|\\&\le \mathbb{P}(S^Z_n\le y)+2C_a\phi^{-1}\sqrt{\log d}+|E[\mathcal{I}^{y+\phi^{-1}}]|\,.\end{aligned}$$

Similarly, we obtain

$$\mathbb{P}(S^V_{n,\epsilon^s}\le y)\ge \mathbb{P}(S^Z_n\le y)-2C_a\phi^{-1}\sqrt{\log d}-|E[\mathcal{I}^{y+\phi^{-1}}]|\,.$$

Combining these bounds yields (B.11).
Auxiliary Calculation 1. We calculate a bound for $|f^{(2)}(0)|$ by utilizing the representation

$$\begin{aligned}f^{(2)}(0)&=\frac{1}{n}\sum_{i=1}^{|I_s|}\sum_{j,k=1}^{d}\mathbb{E}[m^y_{jk}(W^\sigma_i)(V_{\sigma(i)j}V_{\sigma(i)k}-Z_{\sigma(i)j}Z_{\sigma(i)k})]\\&=\frac{1}{n}\sum_{i\in I_s}\sum_{j,k=1}^{d}\mathbb{E}[m^y_{jk}(W^\sigma_{\sigma^{-1}(i)})(V_{ij}V_{ik}-Z_{ij}Z_{ik})]\\&=\frac{1}{n}\sum_{i\in I_s}\sum_{j,k=1}^{d}\mathbb{E}[m^y_{jk}(W^\sigma_{\sigma^{-1}(i)})](\mathcal{E}^V_{i,jk}-\mathcal{E}^Z_{i,jk})\,,\end{aligned}$$

where we used the independence of $W^\sigma_{\sigma^{-1}(i)}$ and $V_{ij}V_{ik}-Z_{ij}Z_{ik}$ when conditioning on σ in the third line. Denoting

$$R^\sigma_{i,jk}=m^y_{jk}(W^\sigma_{\sigma^{-1}(i)})-\frac{\sigma^{-1}(i)}{|I_s|+1}m^y_{jk}\left(W^\sigma_{\sigma^{-1}(i)}+\frac{V_i}{\sqrt n}\right)-(1-\frac{\sigma^{-1}(i)}{|I_s|+1})m^y_{jk}\left(W^\sigma_{\sigma^{-1}(i)}+\frac{Z_i}{\sqrt n}\right)$$

we obtain the decomposition $f^{(2)}(0)=\mathcal{I}_{2,1}+\mathcal{I}_{2,2}$, where

$$\begin{aligned}\mathcal{I}_{2,1}&=\frac{1}{n}\sum_{i\in I_s}\sum_{j,k=1}^{d}\mathbb{E}[m^y_{jk}(W)](\mathcal{E}^V_{i,jk}-\mathcal{E}^Z_{i,jk})\,,\\ \mathcal{I}_{2,2}&=\frac{1}{n}\sum_{i\in I_s}\sum_{j,k=1}^{d}\mathbb{E}[R^\sigma_{i,jk}](\mathcal{E}^V_{i,jk}-\mathcal{E}^Z_{i,jk})\,.\end{aligned}$$

We first bound $\mathcal{I}_{2,1}$ by

$$|\mathcal{I}_{2,1}|\le\sum_{j,k=1}^{d}\mathbb{E}[|m^y_{jk}(W)|]\max_{1\le j,k\le p}\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon^s_i(\mathcal{E}^V_{i,jk}-\mathcal{E}^Z_{i,jk})\right|\le\frac{B_{n,1,s}}{\sqrt n}\sum_{j,k=1}^{d}\mathbb{E}[|m^y_{jk}(W)|]\,.$$

Recalling the definition of m^y and h^y, we see that $m^y_{jk}(W) = h^y(W;\phi^{-1})m^y_{jk}(W)$. Thus, since

$$\mathcal{P} := \mathbb{P}\left(-\phi^{-1} \le \max_{1\le j\le p}\frac{1}{\sqrt{n}}\sum_{i=1}^n (Z_{ij}-y_j) \le \phi^{-1}\right) \le \frac{2C_a\sqrt{\log d}}{\phi} \tag{B.12}$$

by Condition A, the basic properties of U_{jk} and the definitions of the quantities involved imply (note that $m^y_{jk} \le U_{jk}$ and (B.9), (B.12))

$$\begin{aligned}
\sum_{j,k=1}^d \mathbb{E}[|m^y_{jk}(W)|] &= \sum_{j,k=1}^d \mathbb{E}[|h^y(W;\phi^{-1})m^y_{jk}(W)|] \\
&\le \sum_{j,k=1}^d \mathbb{E}[|h^y(W;\phi^{-1})U_{jk}(W)|] \\
&\lesssim \phi^2 \log d\, \mathbb{P}\left(-\phi^{-1} < \max_{1\le j\le d}(W_j - y_j) \le \phi^{-1}\right) \\
&\le \phi^2 \log d\,(2\mathbb{E}[\varrho_{\epsilon^{s+1}}] + \mathcal{P}) \\
&\lesssim \phi^2 \log d\left(\mathbb{E}[\varrho_{\epsilon^{s+1}}] + \frac{\sqrt{\log d}}{\phi}\right), \qquad \text{(B.12)}
\end{aligned}$$

and therefore,

$$|\mathcal{I}_{2,1}| \lesssim \frac{B_{n,1,d}\phi^2\log d}{\sqrt{n}}\left(\mathbb{E}[\varrho_{\epsilon^{s+1}}] + \frac{\sqrt{\log d}}{\phi}\right).$$

To bound $\mathcal{I}_{2,2}$ we use the same Taylor expansion as above and get

$$\begin{aligned}
|\mathbb{E}[R^\sigma_{i,jk}]| \le &\sum_{l,r=1}^d \mathbb{E}\left[m^y_{jklr}\left(W^\sigma_{\sigma^{-1}(i)} + \frac{\bar{t}V_i}{\sqrt{n}}\right)\frac{V_{il}V_{ir}}{n}\right] \\
&+ \sum_{l,r=1}^d \mathbb{E}\left[m^y_{jklr}\left(W^\sigma_{\sigma^{-1}(i)} + \frac{\bar{t}Z_i}{\sqrt{n}}\right)\frac{Z_{il}Z_{ir}}{n}\right],
\end{aligned}$$

which yields $|\mathcal{I}_{2,2}| \le \mathcal{I}_{2,2,1} + \mathcal{I}_{2,2,2}$, where

$$\mathcal{I}_{2,2,1} = \frac{1}{n^2}\sum_{i\in I_s}\sum_{j,k,l,r=1}^n \mathbb{E}\left[\left|m^y_{jklr}\left(W^\sigma_{\sigma^{-1}(i)} + \frac{\bar{t}V_i}{\sqrt{n}}\right)V_{il}V_{ir}\right|\right]|\mathcal{E}^V_{i,jk} - \mathcal{E}^Z_{i,jk}|,$$

$$\mathcal{I}_{2,2,2} = \frac{1}{n^2}\sum_{i\in I_s}\sum_{j,k,l,r=1}^n \mathbb{E}\left[\left|m^y_{jklr}\left(W^\sigma_{\sigma^{-1}(i)} + \frac{\bar{t}Z_i}{\sqrt{n}}\right)Z_{il}Z_{ir}\right|\right]|\mathcal{E}^V_{i,jk} - \mathcal{E}^Z_{i,jk}|.$$

Next, we will bound $\mathcal{I}_{2,2,1}$. Setting $x = C_pB_n(\log(dn))^{1/\beta}/\sqrt{n} + \phi^{-1}$ and $\bar{V}_i = \mathbb{1}\{\|V_i\|_\infty \le C_pB_n(\log(dn))^{1/\beta}\}$, we have

$$\begin{aligned}
&\sum_{l,r=1}^d \mathbb{E}\left[\bar{V}_i\left|m^y_{jklr}\left(W^\sigma_{\sigma^{-1}(i)} + \frac{\bar{t}V_i}{\sqrt{n}}\right)V_{il}V_{ir}\right|\right] \\
&= \sum_{l,r=1}^d \mathbb{E}\left[\bar{V}_i h^y\left(W^\sigma_{\sigma^{-1}(i)};x\right)\left|m^y_{jklr}\left(W^\sigma_{\sigma^{-1}(i)} + \frac{\bar{t}V_i}{\sqrt{n}}\right)V_{il}V_{ir}\right|\right]
\end{aligned}$$

$$\leq \sum_{l,r=1}^{d} \mathbb{E}\left[\bar V_i h^y\left(W^{\sigma}_{\sigma^{-1}(i)};x\right) U^y_{jklr}\left(W^{\sigma}_{\sigma^{-1}(i)} + \frac{\bar t V_i}{\sqrt{n}}\right)|V_{il}V_{ir}|\right]$$

$$\lesssim \sum_{l,r=1}^{d} \mathbb{E}\left[\bar V_i h^y\left(W^{\sigma}_{\sigma^{-1}(i)};x\right) U^y_{jklr}\left(W^{\sigma}_{\sigma^{-1}(i)}\right)|V_{il}V_{ir}|\right], \tag{B.13}$$

where the first equality follows by the definitions of the involved quantities and the second inequality follows by (B.8). Setting $\bar Z_i = \mathbb{1}\{\|Z_i\|_\infty \leq C_p B_n(\log(dn))^{1/\beta}\}$ we bound the above expectation by

$$\begin{aligned}
&\mathbb{E}\left[\bar V_i h^y\left(W^{\sigma}_{\sigma^{-1}(i)};x\right) U^y_{jklr}\left(W^{\sigma}_{\sigma^{-1}(i)}\right)|V_{il}V_{ir}|\right]\\
&\lesssim \mathbb{E}\left[\bar V_i \bar Z_i h^y\left(W^{\sigma}_{\sigma^{-1}(i)};x\right) U^y_{jklr}\left(W^{\sigma}_{\sigma^{-1}(i)}\right)\right]\mathbb{E}[|V_{il}V_{ir}|]\\
&\lesssim \mathbb{E}\left[h^y\left(W;2x\right) U^y_{jklr}\left(W\right)\right]\mathbb{E}[|V_{il}V_{ir}|],
\end{aligned} \tag{B.14}$$

where the inequalities follow by Condition P, the definitions of h^y, W and $W^{\sigma}_{\sigma^{-1}(i)}$ and (B.8) as well as (B.10).

Hence we obtain, by the same arguments as for (B.12),

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i\in I_s}\sum_{j,k,l,r=1}^{d} \mathbb{E}\left[\bar V_i\left|m^y_{jklr}\left(W^{\sigma}_{\sigma^{-1}(i)} + \frac{\bar t V_i}{\sqrt{n}}\right)V_{il}V_{ir}\right|\right]|\mathcal{E}^V_{i,jk} - \mathcal{E}^Z_{i,jk}|\\
&\lesssim \frac{1}{n^2}\sum_{i\in I_s}\sum_{j,k,l,r=1}^{d} \mathbb{E}\left[h^y\left(W;2x\right) U^y_{jklr}\left(W\right)\right]\mathbb{E}[|V_{il}V_{ir}|]|\mathcal{E}^V_{i,jk} - \mathcal{E}^Z_{i,jk}|\\
&\lesssim \frac{B_n^2}{n}\sum_{j,k,l,r=1}^{d} \mathbb{E}\left[h^y\left(W;2x\right) U^y_{jklr}\left(W\right)\right] \lesssim \frac{B_n^2\phi^4(\log d)^3}{n}\left(\mathbb{E}[\varrho_{\epsilon^{s+1}}] + \frac{\sqrt{\log d}}{\phi}\right),
\end{aligned}$$

where we used that by Condition V

$$\begin{aligned}
\max_{1\leq j,j,l,r\leq d}\sum_{i=1}^{n}\mathbb{E}[|V_{il}V_{ir}|]|\mathcal{E}^V_{i,jk} - \mathcal{E}^Z_{i,jk}| &\lesssim \max_{1\leq j,j,l,r\leq d}\sum_{i=1}^{n}\mathbb{E}[|V_{il}V_{ir}|^2]| + |\mathcal{E}^V_{i,jk} - \mathcal{E}^Z_{i,jk}|^2\\
&\lesssim B_n^2 n\,.
\end{aligned}$$

Additionally we have

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i\in I_s}\sum_{j,k,l,r=1}^{d} \mathbb{E}\left[(1-\bar V_i)\left|m^y_{jklr}\left(W^{\sigma}_{\sigma^{-1}(i)} + \frac{\bar t V_i}{\sqrt{n}}\right)V_{il}V_{ir}\right|\right]|\mathcal{E}^V_{i,jk} - \mathcal{E}^Z_{i,jk}|\\
&\lesssim \frac{\phi^4(\log d)^3}{n}\sum_{i=1}^{n}\mathbb{E}[(1-\bar V_i)\,\|V_i\|_\infty^2]\\
&\lesssim \frac{B_n^2\phi^4(\log(nd))^{3+2/\beta}}{n^2},
\end{aligned}$$

where the first inequality follows by $m^y_I \leq U^y_I$ for appropriate index sets I as well as Condition V, and the second inequality follows from Hölder's inequality and Condition P as well as B.

Combining these two inequalities we obtain

$$\mathcal{I}_{2,2,1} \lesssim \frac{B_n^2 \phi^4 (\log d)^3}{n} \left(\mathbb{E}[\varrho_{\epsilon^{s+1}}] + \frac{\sqrt{\log d}}{\phi} \right) + \frac{B_n^2 \phi^4 (\log(nd))^{3+2/\beta}}{n^2} \, .$$

Similar arguments as for $\mathcal{I}_{2,2,1}$ also establish the same bound for $\mathcal{I}_{2,2,2}$ and therefore we conclude

$$|f^{(2)}(0)| \lesssim \left(\frac{B_{n,1,s}\phi^2 \log d}{\sqrt{n}} + \frac{B_n^2 \phi^4 (\log d)^3}{n} \right) \left(\mathbb{E}[\varrho_{\epsilon^{s+1}}] + \frac{\sqrt{\log d}}{\phi} \right) + \frac{B_n^2 \phi^4 (\log(nd))^{3+2/\beta}}{n^2} \, .$$

Auxiliary Calculation 2.
Just as in the beginning of the previous calculations we obtain

$$f^{(3)}(0) = \frac{1}{n^{3/2}} \sum_{i \in I_s} \sum_{j,k,l=1}^{d} \mathbb{E}[m_{jkl}^y (W_{\sigma^{-1}(i)}^{\sigma})](\mathcal{E}_{i,jkl}^V - \mathcal{E}_{i,jkl}^Z) \, .$$

Writing

$$R_{i,jkl}^{\sigma} = m_{jkl}^y (W_{\sigma^{-1}(i)}^{\sigma}) - \frac{\sigma^{-1}(i)}{|I_s|+1} m_{jkl}^y \left(W_{\sigma^{-1}(i)}^{\sigma} + \frac{V_i}{\sqrt{n}} \right)$$
$$- \left(1 - \frac{\sigma^{-1}(i)}{|I_s|+1} \right) m_{jkl}^y \left(W_{\sigma^{-1}(i)}^{\sigma} + \frac{Z_i}{\sqrt{n}} \right)$$

we have (just as above for $f^{(2)}(0)$) that $f^{(3)}(0) = \mathcal{I}_{3,1} + \mathcal{I}_{3,2}$, where

$$\mathcal{I}_{3,1} = \frac{1}{n^{3/2}} \sum_{i \in I_s} \sum_{j,k,l=1}^{d} \mathbb{E}[m_{jkl}^y (W)](\mathcal{E}_{i,jkl}^V - \mathcal{E}_{i,jkl}^Z) \, ,$$

$$\mathcal{I}_{3,2} = \frac{1}{n^{3/2}} \sum_{i \in I_s} \sum_{j,k,l=1}^{d} \mathbb{E}[R_{i,jkl}^{\sigma} (W)](\mathcal{E}_{i,jkl}^V - \mathcal{E}_{i,jkl}^Z) \, .$$

By the same arguments that we used to bound $\mathcal{I}_{2,1}$, we obtain

$$\mathcal{I}_{3,1} \lesssim \frac{B_{n,2,s}\phi^3 (\log d)^2}{n} \left(\mathbb{E}[\varrho_{\epsilon^{s+1}}] + \frac{\sqrt{\log d}}{\phi} \right) \, .$$

We also get by the same arguments as before that $|\mathcal{I}_{3,2}| \le \mathcal{I}_{3,2,1} + \mathcal{I}_{3,2,2}$ where

$$\mathcal{I}_{3,2,1} = \frac{1}{n^{5/2}} \sum_{i \in I_s} \sum_{j,k,l,r,h=1}^{d} \mathbb{E}\left[\left| m_{jklrh}^y \left(W_{\sigma^{-1}(i)}^{\sigma} + \frac{\bar{t} V_i}{\sqrt{n}} \right) V_{ir} V_{ih} \right| \right] |\mathcal{E}_{i,jkl}^V - \mathcal{E}_{i,jkl}^Z| \, ,$$

$$\mathcal{I}_{3,2,2} = \frac{1}{n^{5/2}} \sum_{i \in I_s} \sum_{j,k,l,r,h=1}^{d} \mathbb{E}\left[\left| m_{jklrh}^y \left(W_{\sigma^{-1}(i)}^{\sigma} + \frac{\bar{t} Z_i}{\sqrt{n}} \right) Z_{ir} Z_{ih} \right| \right] |\mathcal{E}_{i,jkl}^V - \mathcal{E}_{i,jkl}^Z| \, .$$

Moreover, we have

$$|\mathcal{E}_{i,jkl}^V| \le \mathbb{E}[|V_{ij} V_{ik} V_{il}|] = \mathbb{E}[\bar{V}_i |V_{ij} V_{ik} V_{il}|] + \mathbb{E}[(1 - \bar{V}_i)|V_{ij} V_{ik} V_{il}|]$$
$$\lesssim B_n (\log(dn))^{1/\beta} \mathbb{E}[|V_{ij} V_{ik}|] + B_n^3 (\log(dn))^{2/\beta} / n^2 \tag{B.15}$$

and similarly

$$|\mathcal{E}_{i,jkl}^Z| \lesssim B_n (\log(dn))^{1/\beta} \mathbb{E}[|V_{ij} V_{ik}|] + B_n^3 (\log(dn))^{2/\beta} / n^2 \, . \tag{B.16}$$

Just as in the previous auxiliary calculation we get

$$
\begin{aligned}
&\frac{1}{n^{5/2}} \sum_{i \in I_s} \sum_{j,k,l,r,h=1}^{d} \mathbb{E}\left[\bar{V}_i \left| m^y_{jklr}\left(W^\sigma_{\sigma^{-1}(i)} + \frac{\bar{t}V_i}{\sqrt{n}}\right) V_{il} V_{ir}\right|\right] |\mathcal{E}^V_{i,jkl} - \mathcal{E}^Z_{i,jkl}| \\
&\lesssim \frac{1}{n^{5/2}} \sum_{i \in I_s} \sum_{j,k,l,r,h=1}^{d} \mathbb{E}\left[h^y(W;2x)\, U^y_{jklrh}(W)\right] \mathbb{E}[|V_{ir}V_{ih}|] |\mathcal{E}^V_{i,jkl} - \mathcal{E}^Z_{i,jkl}| \\
&\lesssim \left(\frac{B_n^3 \phi^5 (\log(dn))^{4+1/\beta}}{n^{3/2}} + \frac{B_n^3 \phi^5 (\log(dn))^{4+2/\beta}}{n^{7/2}}\right)\left(\mathbb{E}[\varrho_{\epsilon^{s+1}}] + \frac{\sqrt{\log d}}{\phi}\right) \\
&\lesssim \frac{B_n^3 \phi^5 (\log(dn))^{4+1/\beta}}{n^{3/2}}\left(\mathbb{E}[\varrho_{\epsilon^{s+1}}] + \frac{\sqrt{\log d}}{\phi}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{1}{n^{5/2}} \sum_{i \in I_s} \sum_{j,k,l,r,h=1}^{d} \mathbb{E}\left[(1-\bar{V}_i) \left| m^y_{jklrh}\left(W^\sigma_{\sigma^{-1}(i)} + \frac{\bar{t}V_i}{\sqrt{n}}\right) V_{ir} V_{ih}\right|\right] |\mathcal{E}^V_{i,jkl} - \mathcal{E}^Z_{i,jkl}| \\
&\lesssim \frac{B_n \phi^5 (\log(dn))^4}{n^{3/2}} \sum_{i=1}^{n} \mathbb{E}\left[(1-\bar{V}_i)\, \|V_i\|_\infty^2\right] \lesssim \frac{B_n^3 \phi^5 (\log(dn))^{4+2/\beta}}{n^{5/2}},
\end{aligned}
$$

where we used that by Condition V and Hölder's inequality

$$
|\mathcal{E}^V_{i,jkl}| \lesssim B_n n \quad \text{and} \quad |\mathcal{E}^Z_{i,jkl}| \lesssim B_n n\,.
$$

Thus,

$$
\mathcal{I}_{3,2,1} \lesssim \frac{B_n^3 \phi^5 (\log(dn))^{4+1/\beta}}{n^{3/2}}\left(\mathbb{E}[\varrho_{\epsilon^{s+1}}] + \frac{\sqrt{\log d}}{\phi}\right) + \frac{B_n^3 \phi^5 (\log(dn))^{4+2/\beta}}{n^{5/2}}
$$

and since the same bound holds for $\mathcal{I}_{3,2,2}$ we have that

$$
\mathcal{I}_{3,2} \lesssim \frac{B_n^3 \phi^5 (\log(dn))^{4+1/\beta}}{n^{3/2}}\left(\mathbb{E}[\varrho_{\epsilon^{s+1}}] + \frac{\sqrt{\log d}}{\phi}\right) + \frac{B_n^3 \phi^5 (\log(dn))^{4+2/\beta}}{n^{5/2}},
$$

which finally yields

$$
\begin{aligned}
|f^{(3)}(0)| \lesssim &\left(\frac{B_n^3 \phi^5 (\log(dn))^{4+1/\beta}}{n^{3/2}} + \frac{B_{n,2,s}\phi^3 (\log d)^2}{n}\right)\left(\mathbb{E}[\varrho_{\epsilon^{s+1}}] + \frac{\sqrt{\log d}}{\phi}\right) \\
&+ \frac{B_n^3 \phi^5 (\log(dn))^{4+2/\beta}}{n^{5/2}}.
\end{aligned}
$$

Auxiliary Calculation 3. We decompose $f^{(4)}(\bar{t}) = \mathcal{I}_{4,1} - \mathcal{I}_{4,2}$, where

$$
\begin{aligned}
\mathcal{I}_{4,1} &= \frac{1}{n^2} \sum_{i \in I_s} \sum_{j,k,l,r=1} \mathbb{E}\left[m^y_{jklr}\left(W^\sigma_{\sigma^{-1}(i)} + \frac{\bar{t}V_i}{\sqrt{n}}\right) V_{ij} V_{ik} V_{il} V_{ir}\right], \\
\mathcal{I}_{4,2} &= \frac{1}{n^2} \sum_{i \in I_s} \sum_{j,k,l,r=1} \mathbb{E}\left[m^y_{jklr}\left(W^\sigma_{\sigma^{-1}(i)} + \frac{\bar{t}Z_i}{\sqrt{n}}\right) Z_{ij} Z_{ik} Z_{il} Z_{ir}\right].
\end{aligned}
$$

Again denoting $x = C_p B_n (\log(dn))^{1/\beta}/\sqrt{n} + \phi^{-1}$ we have, by the same arguments leading to (B.13),

$$
\frac{1}{n^2} \sum_{i \in I_s} \sum_{j,k,l,r=1} \mathbb{E}\left[\bar{V}_i \left| m^y_{jklr}\left(W^\sigma_{\sigma^{-1}(i)} + \frac{\bar{t}V_i}{\sqrt{n}}\right) V_{ij} V_{ik} V_{il} V_{ir}\right|\right]
$$

$$\lesssim \frac{1}{n^2}\sum_{i\in I_s}\sum_{j,k,l,r=1} \mathbb{E}\left[h^y(W^\sigma_{\sigma^{-1}(i)};x)U^y_{jklr}\left(W^\sigma_{\sigma^{-1}(i)}\right)\right]\mathbb{E}[|V_{ij}V_{ik}V_{il}V_{ir}|].$$

We also obtain

$$\mathbb{E}[h^y(W^\sigma_{\sigma^{-1}(i)};x)U^y_{jklr}(W^\sigma_{\sigma^{-1}(i)})]\lesssim \mathbb{E}[h^y(W;2x)U^y_{jklr}(W)]$$

by the same arguments as those leading to (B.14). Hence,

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i\in I_s}\sum_{j,k,l,r=1}\mathbb{E}\left[\bar V_i\left|m^y_{jklr}\left(W^\sigma_{\sigma^{-1}(i)}+\frac{\bar t V_i}{\sqrt{n}}\right)V_{ij}V_{ik}V_{il}V_{ir}\right|\right]\\
&\lesssim \frac{1}{n^2}\sum_{j,k,l,r=1}\mathbb{E}\left[h^y(W;2x)U^y_{jklr}(W)\right]\sum_{i=1}^n\max_{1\le j,k,l,r\le p}\mathbb{E}[|V_{ij}V_{ik}V_{il}V_{ir}|]\\
&\lesssim \frac{B_n^2\phi^4(\log d)^3}{n}\left(\mathbb{E}[\varrho_{\epsilon^{s+1}}]+\frac{\sqrt{\log d}}{\phi}\right),
\end{aligned}$$

where the second inequality follows from the properties of U^y and the arguments leading up to (B.12). Moreover, we have

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i\in I_s}\sum_{j,k,l,r=1}\mathbb{E}\left[(1-\bar V_i)\left|m^y_{jklr}\left(W^\sigma_{\sigma^{-1}(i)}+\frac{\bar t V_i}{\sqrt{n}}\right)V_{ij}V_{ik}V_{il}V_{ir}\right|\right]\\
&\lesssim \frac{\phi^4(\log d)^3}{n^2}\sum_{i=1}^n\mathbb{E}[(1-\bar V_i)\,\|V_i\|_\infty^4]\lesssim \frac{B_n^4\phi^4(\log d)^3(\log(nd))^{4/\beta}}{n^3}\\
&\le \frac{B_n^2\phi^4(\log(nd))^{1+2/\beta}}{n^2}
\end{aligned}$$

by Condition B and $m^y_I\lesssim U^y_I$. Clearly the same bounds also hold for $\mathcal{I}_{4,2}$ which finally establishes

$$|f^{(4)}(\bar t)|\lesssim \frac{B_n^2\phi^4(\log(nd))^{1+2/\beta}}{n^2}+\frac{B_n^2\phi^4(\log d)^3}{n}\left(\mathbb{E}[\varrho_{\epsilon^{s+1}}]+\frac{\sqrt{\log d}}{\phi}\right).$$

□

LEMMA B.6. *Suppose that the conditions of Lemma B.5 are satisfied. Then there exists a constant $K>0$ depending only on C_v, C_p, C_b such that for all $s=0,\dots,D$, if $B_{n,1,s+1}\ge B_{n,1,s}+KB_n(\log(nd))^{1/2}$ and $B_{n,2,s+1}\ge B_{n,2,s}+KB_n^2(\log(dn))^{1/2+2/\beta}$, then for any constant $\phi>0$ satisfying (B.10) we have*

$$\begin{aligned}
\mathbb{E}[\varrho_{\epsilon^s}\mathbb{1}\{\mathcal{A}_s\}]\lesssim&\frac{\sqrt{\log d}}{\phi}+\frac{B_n^2\phi^4(\log(dn))^{3+2/\beta}}{n^2}+\left(\frac{\sqrt{\log d}}{\phi}+\mathbb{E}[\varrho_{\epsilon^{s+1}}\mathbb{1}\{\mathcal{A}_{s+1}\}]\right)\\
&\times\left(\frac{B_{n,1,s}\phi^2\log d}{\sqrt{n}}+\frac{B_{2,n,s}\phi^3(\log d)^2}{n}+\frac{B_n^2\phi^4(\log d)^3}{n}\right)
\end{aligned}$$

up to a constant only depending on C_v, C_p, C_b, C_a.

PROOF. Fix $s=0,\dots,D-1$ and $\phi>0$ such that (B.10) holds. By Lemma B.5 we have

$$\begin{aligned}
\mathbb{E}[\varrho_{\epsilon^s}\mathbb{1}\{\mathcal{A}_s\}]\lesssim&\frac{\sqrt{\log d}}{\phi}+\frac{B_n^2\phi^4(\log(dn))^{3+2/\beta}}{n^2}+\left(\frac{\sqrt{\log d}}{\phi}+\mathbb{E}[\varrho_{\epsilon^{s+1}}\mathbb{1}\{\mathcal{A}_d\}]\right)\\
&\times\left(\frac{B_{n,1,s}\phi^2\log d}{\sqrt{n}}+\frac{B_{2,n,s}\phi^3(\log d)^2}{n}+\frac{B_n^2\phi^4(\log d)^3}{n}\right)
\end{aligned}$$

up to a constant only depending on C_v, C_p, C_B, C_a. Hence the claim of the lemma follows if we can show that

$$\mathbb{E}[\varrho_{\epsilon^{s+1}}\mathbb{1}\{\mathcal{A}_s\}] \le \mathbb{E}[\varrho_{\epsilon^{s+1}}\mathbb{1}\{\mathcal{A}_{s+1}\}] + \frac{4}{n}\,.$$

We have

$$\begin{aligned}\mathbb{E}[\varrho_{\epsilon^{s+1}}\mathbb{1}\{\mathcal{A}_s\}] &= \mathbb{E}[\varrho_{\epsilon^{s+1}}\mathbb{1}\{\mathcal{A}_s\}\mathbb{1}\{\mathcal{A}_{s+1}\}] + \mathbb{E}[\varrho_{\epsilon^{s+1}}\mathbb{1}\{\mathcal{A}_s\}(1-\mathbb{1}\{\mathcal{A}_{s+1}\})]\\ &\le \mathbb{E}[\varrho_{\epsilon^{s+1}}\mathbb{1}\{\mathcal{A}_{s+1}\}] + \mathbb{E}[\mathbb{1}\{\mathcal{A}_s\}(1-\mathbb{1}\{\mathcal{A}_{s+1}\})]\\ &\le \mathbb{E}[\varrho_{\epsilon^{s+1}}\mathbb{1}\{\mathcal{A}_{s+1}\}] + 1 - \mathbb{P}(\mathcal{A}_{s+1}|\mathcal{A}_s)\end{aligned}$$

where we used that $0 \le \varrho_{\epsilon^{s+1}} \le 1$ for the first inequality. Now Lemma B.15 yields

$$\begin{aligned}&\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i^{s+1}(\mathcal{E}_{i,jk}^V - \mathcal{E}_{i,jk}^Z)\right| > \left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i^{s}(\mathcal{E}_{i,jk}^V - \mathcal{E}_{i,jk}^Z)\right| + t\middle|\epsilon^s\right)\\ &\qquad\le 2\exp\left(-\frac{nt^2}{32\sum_{i=1}^n(\mathcal{E}_{i,jk}^V - \mathcal{E}_{i,jk}^Z)^2}\right) \le 2\exp\left(-\frac{t^2}{128B_n^2C_v}\right),\end{aligned}$$

where the last inequality is due to Condition V. Setting $t = 8B_n\sqrt{6C_v\log(dn)}$ and recalling that

$$\max_{1\le j,k\le d}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i^{s}(\mathcal{E}_{i,jk}^V - \mathcal{E}_{i,jk}^Z)\right| \le B_{n,1,d}$$

on $\mathcal{A}_s$ we obtain by the tower property of conditional probabilities that for any $B_{n,1,s+1} \ge B_{n,1,s} + t$

$$\mathbb{P}\left(\max_{1\le j,k\le d}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i^{s}(\mathcal{E}_{i,jk}^V - \mathcal{E}_{i,jk}^Z)\right| > B_{n,1,d+1}\middle|\mathcal{A}_s\right) \le \frac{2p^2}{(nd)^3} \le \frac{2}{n}\,.$$

We recall (B.15) and (B.16) which follow by Conditions P and B. Hence we find that

$$\frac{32}{n}\sum_{i=1}^n(\mathcal{E}_{i,jkl}^V - \mathcal{E}_{i,jkl}^Z)^2 \le CB_n^4(\log(dn))^{4/\beta}$$

for some constant C only depending on C_v, C_p and C_b. We hence obtain by the same arguments as above

$$\begin{aligned}&\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i^{s+1}(\mathcal{E}_{i,jkl}^V - \mathcal{E}_{i,jkl}^Z)\right| > \left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i^{s}(\mathcal{E}_{i,jkl}^V - \mathcal{E}_{i,jkl}^Z)\right| + t\middle|\epsilon^s\right)\\ &\le 2\exp\left(-\frac{nt^2}{32\sum_{i=1}^n(\mathcal{E}_{i,jkl}^V - \mathcal{E}_{i,jkl}^Z)^2}\right) \le 2\exp\left(-\frac{t^2}{128C(\log(dn))^{4/\beta}}\right).\end{aligned}$$

Applying this inequality with $t = \sqrt{3C}B_n^2(\log(dn))^{1/2+2/\beta}$ yields that for any $B_{n,2,s+1} \ge B_{n,2,s} + t$, we have

$$\mathbb{P}\left(\max_{1\le j,k,l\le d}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i^{s+1}(\mathcal{E}_{i,jkl}^V - \mathcal{E}_{i,jkl}^Z)\right| > B_{n,2,s+1}\middle|\mathcal{A}_s\right) \le \frac{2p^3}{(nd)^3} \le \frac{2}{n}\,.$$

Thus $1 - \mathbb{P}(\mathcal{A}_{s+1}|\mathcal{A}_s) \le 4/n$ which completes the proof. $\square$

LEMMA B.7. *For any constant $\phi > 0$ such that (B.10) holds we have*

$$\mathbb{E}[\varrho_{\epsilon^s}\mathbb{1}\{\mathcal{A}_D\}] \le \frac{1}{n}.$$

PROOF. Recall that $D = [\log(n)] + 1$ and note that $\varrho_{\epsilon^D} = 0$ if $\epsilon^D = (0, \ldots, 0)$. Moreover, by Markov's Inequality,

$$\begin{aligned}
\mathbb{P}(\epsilon^D \neq (0, \ldots, 0)) &= \mathbb{P}\left(\sum_{i=1}^n \epsilon_i^D \ge 1\right) \le \mathbb{E}[\sum_{i=1}^n \epsilon_i^D] \\
&= \mathbb{E}[\mathbb{E}[\sum_{i=1}^n \epsilon_i^D | \sum_{i=1}^n \epsilon_i^{D-1}]] = \mathbb{E}[\frac{1}{2}\sum_{i=1}^n \epsilon_i^{D-1}] \\
&= \ldots = \mathbb{E}[\frac{1}{2^D}\sum_{i=1}^n \epsilon_i^0] = \frac{n}{2^D} \le \frac{n}{2^{4\log(n)}} \le \frac{1}{n}.
\end{aligned}$$

It follows that

$$\mathbb{E}[\varrho_{\epsilon^D}\mathbb{1}\{\mathcal{A}_D\}] \le \mathbb{E}[\varrho_{\epsilon^D}] \le \mathbb{P}(\epsilon^D \neq (0, \ldots, 0)) \le \frac{1}{n}.$$

$\square$

PROOF OF THEOREM B.1. Throughout the proof we will assume that

$$C_p^4 B_n^2 (\log(dn))^{4+2/\beta} \le n$$

since otherwise the claim follows immediately.
Let K be the constant from Lemma B.6 and for all $s = 0, \ldots, D$ define $B_{n,1,s} = C_1 B_n (s + 1)(\log(nd))^{1/\beta}$ and $B_{n,2,s} = C_1 B_n^2 (s+1)(\log(nd))^{1/2+2/\beta}$ where $C_1 = K + C_M$ so that both $\mathcal{A}_0$ and the requirements for Lemma B.6 hold. Now we define for $s = 0, \ldots, D$

$$f_s = \inf\left\{x \ge 1 : \mathbb{E}[\varrho_{\epsilon^s}\mathbb{1}\{\mathcal{A}_s\}] \le x\left(\frac{B_n^2(\log(dn))^{4+2/\beta}}{n}\right)^{1/4}\right\}$$

and for all $s = 0, \ldots, D$ we apply Lemma B.6 with

$$\phi = \phi_s = \frac{n^{1/4}}{B_n^{1/4}(\log(dn))^{1/2+1/(2\beta)}((d+1)f_{s+1})^{1/3}}.$$

Noting that

$$\begin{aligned}
\frac{B_n^2\phi^4(\log(dn))^{3+2/\beta}}{n^2} \le \frac{\log(dn)}{n} \le \frac{B_n^2 C_p(\log(dn))^{1/4}}{n^{1/4}} &\le \frac{C_p\sqrt{\log d}}{\phi} \\
&\le C_p((s+1)f_{s+1})^{1/3}\left(\frac{B_n^2(\log(nd))^{4+2/\beta}}{n}\right)^{1/4}
\end{aligned}$$

and

$$\begin{aligned}
\frac{B_{n,1,s}\phi^2\log d}{\sqrt{n}} &\le \frac{C_1(s+1)}{((s+1)f_{s+1})^{2/3}} \\
\frac{B_{n,2,s}\phi^3(\log d)^2}{n} + \frac{B_n^2\phi^4(\log(dn))^3}{n} &\le \frac{C_1+1}{f_{s+1}}
\end{aligned}$$

we get for $s = 0, \dots D$

$$\mathbb{E}[\varrho_{\epsilon^s}\mathbb{1}\{\mathcal{A}_s\}] \le C_2(f_{s+1}^{2/3} + (s+1)^{2/3} + 1)\left(\frac{B_n^2(\log(dn))^{4+2/\beta}}{n}\right)^{1/4}$$

for some constant C_2 depending only on C_v, C_p, C_b, C_a, C_m. Hence we obtain

$$f_s \le C_2(f_{s+1}^{2/3} + (s+1)^{2/3} + 1)).$$

Clearly $f_D = 1$ due to the previous lemma. A simple induction then shows that

$$f_s \le C(s+1)$$

for some constant $C \ge 1$ depending only on C_2. We then finally obtain

$$\varrho_{\epsilon^0}\mathbb{1}\{\mathcal{A}_0\} = \mathbb{E}[\varrho_{\epsilon^0}\mathbb{1}\{\mathcal{A}_0\}] \le C\left(\frac{B_n^2(\log(dn))^{4+2/\beta}}{n}\right)^{1/4}.$$

□

B.2. Sub-Weibull Random Variables. In this section we collect some results on sub-Weibull random variables, which are mainly taken from Kuchibhotla and Chakrabortty (2020). Recalling the definition of the Orlicz norm in (2.13), a random variable X is called sub-Weibull of order β, denoted sub-Weibull(β), if

$$\|X\|_{\psi_\beta} < \infty\,,$$

where $\psi_\beta(x) = \exp(x^\beta) - 1$. We also occasionally call $\|X\|_{\psi_\beta}$ its β-parameter. This definition includes the important sub-exponential ($\beta = 1$) and sub-Gaussian ($\beta = 2$) cases. Clearly sub-Weibull(β) random variables possess exponential tail decay rates, more precisely $\mathbb{P}(|X| \ge t) \le 2\exp(-t^\beta/\|X\|_{\psi_\beta}^\beta)$. The following result is a slight refinement of this statement, which for instance can be found in Kuchibhotla and Chakrabortty (2020).

LEMMA B.8. *For any random variable X and constant $\beta > 0$ the following are equivalent:*

i) $\|X\|_{\psi_\beta} = K_1$,
ii) $\mathbb{P}(|X| \ge t) \le 2\exp\left(-\frac{t^\beta}{K_2^\beta}\right)$,
iii) $\sup_{p\ge1}\frac{\|X\|_p}{p^{1/\beta}} = K_3$,

where we have $K_1 \lesssim K_2 \lesssim K_3 \lesssim K_1$ up to constants only depending on β. Note that the third formulation yields a quasi-triangle inequality for the β-parameter of sums of finitely many random variables.

PROOF. If i) holds, ii) follows from Markov's inequality.

If ii) holds, it follows that

$$\begin{aligned}\mathbb{E}[|X|^p] &= \int_0^\infty \mathbb{P}(|X|^p > t)dt = \int_0^\infty \mathbb{P}(|X| > t^{1/p})dt \\ &\le 2\int_0^\infty \exp(-t^{\beta/p}/K_2^\beta)dt = 2K_2^p\frac{p}{\beta}\Gamma(p/\beta)\,.\end{aligned}$$

Taking the $p-$th root and recalling $\Gamma(x) \le x^{1/x}$ then yields $\frac{\|X\|_p}{p^{1/\beta}} \le C_\beta K_2$, which implies iii).

If iii) holds, we have for some $K > 0$

$$\mathbb{E}[\exp(|X|^\beta/K^\beta) - 1] = \sum_{n=1}^{\infty} \frac{\mathbb{E}[|X|^{\beta n}]}{K^{\beta n} n!} \leq \sum_{n=1}^{\infty} (\beta n)^n \frac{K_3^{\beta n}}{K^{\beta n} n!} \leq \sum_{n=1}^{\infty} (\beta e)^n \left(\frac{K_3}{K}\right)^{\beta n}.$$

Now there exists a constant C_β only depending on β such that $K = C_\beta K_3$ yields that the last term is bounded by 1. This implies i).

□

LEMMA B.9. *Let X be a random variable with $\|X\|_{\psi_\beta} < \infty$. Then for any sigma algebra $\mathcal{B}$ we have that $\|\mathbb{E}[X|\mathcal{B}]\|_{\psi_\beta} \leq C_\beta \|X\|_{\psi_\beta}$.*

PROOF. This follows immediately from Lemma B.8 and the fact that conditional expectations are $\mathcal{L}_p$ contractions. □

LEMMA B.10. *Let $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$ be the average of sub-Weibull(2) random variables with 2-parameter σ. Then $\bar{X}_n$ is sub-Weibull(2) with 2-parameter at least $\frac{\sigma C}{\sqrt{n}}$ and at most $\frac{\sigma \bar{C}}{\sqrt{n}}$ for some universal constants $C, \bar{C} > 0$.*

LEMMA B.11. *Let $X_1, \dots, X_n$ be random variables with $\|X_k\|_{\psi_{\beta_k}} < \infty$ $(k = 1, \dots, n)$. Then for $\frac{1}{\beta} = \sum \frac{1}{\beta_k}$ we have*

$$\left\|\prod_{k=1}^n X_k\right\|_{\psi_\beta} \leq \prod_{k=1}^n \|X_k\|_{\psi_{\beta_k}}.$$

LEMMA B.12. *Assume that $X_i = (X_{i1}, \dots, X_{id})^\top$, $1 \leq i \leq n$, are random vectors whose components X_{ij}, $1 \leq j \leq d$, are sub-Weibull(β) random variables with $\|X_{ij}\|_{\psi_\beta} \leq K$. Then for $d \geq 2$ we have*

$$\max_{1\leq i\leq d} \|X_i\|_\infty \leq K\left(5\log(dn)\right)^{1/\beta}$$

with probability at least $1 - 1/(2n^4)$.

PROOF. Using the union bound and Lemma B.8, we obtain for any $x > 0$,

$$\mathbb{P}(\max_{1\leq i\leq n, 1\leq j\leq d} |X_{ij}| > x) \leq dn \max_{1\leq i\leq n, 1\leq j\leq d} \mathbb{P}(|X_{ij}| > x) \leq 2dn \exp\left(-\frac{x^\beta}{K^\beta}\right).$$

Taking $x = K\left(5\log(dn)\right)^{1/\beta}$ yields the desired claim. □

LEMMA B.13 (Kuchibhotla and Chakrabortty (2020), Theorem 3.4). *Let $X_1, \dots, X_n$ be independent d-dimensional random vectors with mean zero and components satisfying $\|X_{ij}\|_{\psi_\beta} \leq K_n$ for some $\beta \leq 2$. Setting*

$$\Gamma_n := \max_{1\leq j\leq d} \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_{ij}^2]$$

we have for $t > 0$, with probability at least $1 - 3e^{-t}$,

$$\left\|\frac{1}{n}\sum_{i=1}^n X_i\right\|_\infty \lesssim \sqrt{\frac{\Gamma_n(t + \log d)}{n}} + K_n \frac{(\log(2n))^{1/\beta}(t + \log d)^{1/\beta^*}}{n}$$

up to some constant depending only on β and where $\beta^* = \min(1,\beta)$. In particular, noting that $\Gamma_n \lesssim K_n^2$ up to a constant depending only on β, we have for $t = \log d$ and $\frac{1}{n} \lesssim (\log d)^{1-2/\beta^*}$ that

$$\left\| \frac{1}{n}\sum_{i=1}^n X_i \right\|_\infty \lesssim K_n \sqrt{\frac{\log d}{n}}$$

holds with probability at least $1 - 3/d$. When $\log d \lesssim n^\gamma$ this holds as long as $\gamma \le \frac{1}{2/\beta^* - 1}$.

B.3. Further technical details. All results in this section are taken from Chernozhukov et al. (2019), but we will list them here for sake of completeness.

LEMMA B.14. *(Chernozhukov et al., 2019, Lemma 7.2) Let $\mathcal{S}_n$ be the set of all permutations of $\{1,\dots,n\}$. Let $X_1,\dots,X_n$, $Y_1,\dots,Y_n$ be sequences of vectors in $\mathbb{R}^d$. Let U be a random variable with uniform distribution on $[0,1]$ and σ be uniformly distributed on $\mathcal{S}_n$ and also independent from U. For $k = 1,\dots,n$ denote*

$$W_k^\sigma = \sum_{j=1}^{k-1} X_{\sigma(j)} + \sum_{j=k+1}^{n} Y_{\sigma(j)}$$

and

$$W_k = \begin{cases} W^\sigma_{\sigma^{-1}(k)} + X_k\,, & \text{if } U \le \frac{\sigma^{-1}(k)}{n+1}\,, \\ W^\sigma_{\sigma^{-1}(k)} + Y_k\,, & \text{if } U > \frac{\sigma^{-1}(k)}{n+1}\,. \end{cases}$$

Then the distribution of W_k does not depend on k and there exists a random vector $\epsilon = (\epsilon_1,\dots,\epsilon_n)$ with values in $\{0,1\}^n$ such that the distribution of W_k is equal to that of

$$\sum_{i=1}^n (\epsilon_i X_i + (1-\epsilon_i) Y_i)$$

In particular, the random variables ϵ_i are exchangeable and their sum is uniformly distributed on $\{0,\dots,n\}$.

LEMMA B.15. *(Chernozhukov et al., 2019, Lemma7.1) Let $a_1,\dots,a_n$ be some constants in $\mathbb{R}$ and let $X_1,\dots,X_n$ be exchangeable random variables such that $|X_i| \le 1$ almost surely. Then*

$$\mathbb{P}\left(\left| \sum_{i=1}^n a_i X_i \right| \ge \left| \sum_{i=1}^n a_i \right| + t \right) \le 2\exp\left(-\frac{t^2}{32\sum_{i=1}^n a_i^2} \right)$$

for all $t > 0$.

LEMMA B.16. *(Chernozhukov et al., 2017, Lemma7.1) Let $X_1,\dots,X_n$ be independent centered random vectors in $\mathbb{R}^d$ with $d \ge 2$. Define $Z = \max_{1\le j\le d} |\sum_{i=1}^n X_{ij}|$, $M = \max_{1\le i\le n, 1\le j\le d} |X_{ij}|$ and $\sigma^2 = \max_{1\le j\le d} \sum_{i=1}^n \mathbb{E}[X_{ij}^2]$. Then*

$$E[Z] \le L(\sigma\sqrt{\log d} + \sqrt{E[M^2]}\,\log d)$$

for some universal constant L. Moreover, for every $\nu > 0, \beta \in (0,1]$ and $t > 0$ we have

$$\mathbb{P}(Z \ge (1+\nu)\mathbb{E}[Z] + t) \le \exp(-t^2/(3\sigma^2)) + 3\exp\left(-\frac{t^\beta}{K\,\|M\|_{\psi_\beta}^\beta} \right)$$

for some universal constant K that depends only on ν and β.

B.4. Concentration Inequalities for U-Statistics.

DEFINITION B.17. *Consider a symmetric and measurable function* $h = (h_1, \ldots, h_d)^\top : (\mathbb{R}^p)^m \to \mathbb{R}^d$ *together with a collection of iid random variables* $X_1, \ldots, X_n \in \mathbb{R}^p$. *We define the associated U-statistic* U_n *of order* m *by*

$$U_n = \binom{n}{m}^{-1} \sum_{1 \le l_1 < \ldots < l_m \le n} h(X_{l_1}, \ldots, X_{l_m}) \,.$$

For $x \in \mathbb{R}^d$ *we write*

$$h_{1,i}(x) = \mathbb{E}[h(X_1, \ldots, X_m) | X_1 = x] \,, \qquad 1 \le i \le d \,,$$

and set $h_{(1)}(x) = (h_{1,1}(x), \ldots, h_{1,d}(x))^\top$.

LEMMA B.18. *Consider a mean zero U-Statistic* U_n *of order* m *as defined above. Provided that* $\max_{1 \le i \le d} \|h_i(X_1, \ldots, X_m)\|_{\psi_\beta} \le K$ *for some* $2 \ge \beta > 0$ *and that* $\log d = o(n^\gamma)$ *for* $\gamma \le \frac{1}{2/\beta + 1}$ *it holds*

$$\|U_n\|_\infty \lesssim K \sqrt{\frac{\log d}{n}}$$

with probability at least $1 - 3/d - C(\log d)^{1/2 + 1/\beta}/\sqrt{n}$ *for some universal constant* $C > 0$. *Note that the same bound with* $\log(nd)$ *instead of* $\log d$ *holds with probability at least* $1 - 3/(nd) - C(\log d)^{1/2 + 1/\beta}/\sqrt{n}$

PROOF. By Theorem 5.1 from Song et al. (2019) we obtain that

$$\mathbb{E} \left\| U_n - \frac{m}{n} \sum_{k=1}^n h_{(1)}(X_k) \right\|_\infty \lesssim K \left(\frac{(\log d)^{1+1/\beta}}{n} \right)$$

up to some universal constant that depends only on m and β. Using Markov's inequality we deduce that

$$\mathbb{P} \left(\left\| U_n - \frac{m}{n} \sum_{k=1}^n h_{(1)}(X_k) \right\|_\infty > t \right) \lesssim \frac{K}{t} \left(\frac{(\log d)^{1+1/\beta}}{n} \right) . \tag{B.17}$$

For the linear part of U_n we obtain by Lemmas B.13 and B.9 that

$$\left\| \frac{m}{n} \sum_{k=1}^n h_{(1)}(X_k) \right\|_\infty \lesssim K \sqrt{\frac{\log d}{n}}$$

with probability at least $1 - 3/d$ as long as $\log d = o(n^\gamma)$ where $\gamma \le \frac{1}{2/\beta^* - 1}$. Setting $t = K\sqrt{\frac{\log d}{n}}$ in (B.17) then yields

$$\|U_n\|_\infty \lesssim K \sqrt{\frac{\log d}{n}}$$

with probability at least $1 - 3/d - C(\log d)^{1/2 + 1/\beta}/\sqrt{n}$ up to some universal constant C that depends only on m and β. The second bound is obtained by the same arguments but with a different choice of t. $\square$

REFERENCES

Akritas, A. G., Akritas, E. K., and Malaschonok, G. I. (1996). Various proofs of sylvester's (determinant) identity. *Mathematics and Computers in Simulation*, 42(4):585–593. Sybolic Computation, New Trends and Developments.

Albert, M., Bouret, Y., Fromont, M., and Reynaud-Bouret, P. (2015). Bootstrap and permutation tests of independence for point processes. *The Annals of Statistics*, 43(6):2537–2564.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, second edition.

Arratia, R., Goldstein, L., and Gordon, L. (1989). Two Moments Suffice for Poisson Approximations: The Chen-Stein Method. *The Annals of Probability*, 17(1):9 – 25.

Bai, Z., Jiang, D., Yao, J.-F., and Zheng, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Annals of Statistics*, 37:3822–3840.

Bao, Z., Lin, L.-C., Pan, G., and Zhou, W. (2015). Spectral statistics of large dimensional spearman's rank correlation matrix and its application. *The Annals of Statistics*, 43(6):2588–2623.

Berger, J. O. and Delampady, M. (1987). Testing Precise Hypotheses. *Statistical Science*, 2(3):317 – 335.

Bergsma, W. and Dassios, A. (2014). A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli*, 20(2):1006 – 1028.

Bickel, P. J. and Freedman, D. A. (1981). Some Asymptotic Theory for the Bootstrap. *The Annals of Statistics*, 9(6):1196 – 1217.

Blum, J. R., Kiefer, J., and Rosenblatt, M. (1961). Distribution Free Tests of Independence Based on the Sample Distribution Function. *Ann. Math. Statist.*, 32(2):485 – 498.

Bodnar, T., Dette, H., and Parolya, N. (2019). Testing for independence of large dimensional vectors. *The Annals of Statistics*, 47(5):2977 – 3008.

Cai, T. T. and Jiang, T. (2012). Phase transition in limiting distributions of coherence of high-dimensional random matrices. *J. Multivar. Anal.*, 107:24–39.

Chatterjee, S. (2021). A new coefficient of correlation. *J. Amer. Statist. Assoc.*, 116(536):2009–2022.

Chen, H. and Jiang, T. (2018). A study of two high-dimensional likelihood ratio tests under alternative hypotheses. *Random Matrices: Theory and Applications*, 07(01):1750016.

Chen, X. (2018). Gaussian and bootstrap approximations for high-dimensional u-statistics and their applications. *Annals of Statistics*, 46:642–678.

Chen, X. and Kato, K. (2017). Randomized incomplete u-statistics in high dimensions. *The Annals of Statistics*, 47(6):3127–3156.

Cheng, G., Liu, Z., and Peng, L. (2022). Gaussian approximations for high-dimensional non-degenerate u-statistics via exchangeable pairs. *Statistics & Probability Letters*, 182:109295.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability*, 45(4):2309–2352.

Chernozhukov, V., Chetverikov, D., Kato, K., and Koike, Y. (2019). Improved Central Limit Theorem and bootstrap approximations in high dimensions. Papers 1912.10529, arXiv.org.

Chow, S.-C. and Liu, P.-J. (1992). *Design and Analysis of Bioavailability and Bioequivalence Studies*. Marcel Dekker, New York.

Dette, H. and Dörnemann, N. (2020). Likelihood ratio tests for many groups in high dimensions. *Journal of Multivariate Analysis*, 178:104605.

Dette, H., Siburg, K. F., and Stoimenov, P. A. (2012). A copula-based non-parametric measure of regression dependence. *Scand. J. Stat.*, 40(1):21–41.

Drton, M., Han, F., and Shi, H. (2020). High-dimensional consistent independence testing with maxima of rank correlations. *Annals of Statistics*, 48:3206–3227.

Edelmann, D., Terzer, T., and Richards, D. (2021). A basic treatment of the distance covariance. *Sankhya B*, 83(S1):12–25.

Even-Zohar, C. (2020). Independence: Fast rank tests. *Arxiv Preprint arXiv:2010.09712*.

Geenens, G. and Lafaye de Micheaux, P. (2020). The Hellinger correlation. *J. Amer. Statist. Assoc.*, pages 1–15.

Gretton, A., Fukumizu, K.and Teo, C., Song, L., Schölkopf, B., and Smola, A. (2008). A kernel statistical test of independence. In Platt, J., Koller, D., Singer, Y., and Roweis, S., editors, *Advances in Neural Information Processing Systems*, volume 20, pages 585–592. Curran Associates, Inc.

Han, F., Chen, S., and Liu, H. (2017). Distribution-free tests of independence in high dimensions. *Biometrika*, 104(4):813–828.

He, Y., Xu, G., Wu, C., and Pan, W. (2021). Asymptotically independent U-statistics in high-dimensional testing. *The Annals of Statistics*, 49(1):154 – 181.

Heiny, J., Mikosch, T., and Yslas, J. (2021). Point process convergence for the off-diagonal entries of sample covariance matrices. *Ann. Appl. Probab.*, 31(2):538–560.

Heller, R., Heller, Y., and Gorfine, M. (2012). A consistent multivariate test of association based on ranks of distances. *Biometrika*, 100(2):503–510.

Hoeffding, W. (1948a). A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical Statistics*, 19(3):293 – 325.

Hoeffding, W. (1948b). A Non-Parametric Test of Independence. *The Annals of Mathematical Statistics*, 19(4):546 – 557.

Jiang, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *The Annals of Applied Probability*, 14(2):865 – 880.

Jiang, T. and Qi, Y. (2015). Likelihood ratio tests for high-dimensional normal distributions. *Scandinavian Journal of Statistics*, 42(4):988–1009.

Jiang, T. and Yang, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *The Annals of Statistics*, 41(4):2029 – 2074.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2):81–93.

Kim, I. (2020). Multinomial goodness-of-fit based on u-statistics: High-dimensional asymptotic and minimax optimality. *Journal of Statistical Planning and Inference*, 205:74–91.

Kuchibhotla, A. K. and Chakrabortty, A. (2020). Moving beyond sub-gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression. *arXiv Preprint, arXiv:1804.02605*.

Leadbetter, M., Lindgren, G., and Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Springer Series in Statistics. Springer, Germany.

Lee, Y.-S. (1971). Distribution of the Canonical Correlations and Asymptotic Expansions for Distributions of Certain Independence Test Statistics. *The Annals of Mathematical Statistics*, 42(2):526 – 537.

Leung, D. and Drton, M. (2018). Testing independence in high dimensions with sums of rank correlations. *The Annals of Statistics*, 46(1):280 – 307.

Li, D., Liu, W.-D., and Rosalsky, A. (2010). Necessary and sufficient conditions for the asymptotic distribution of the largest entry of a sample correlation matrix. *Probability Theory and Related Fields*, 148:5–35.

Li, Z., Wang, Q., and Li, R. (2021). Central limit theorem for linear spectral statistics of large dimensional kendall's rank correlation matrices and its applications. *Annals of Statistics*, 49(3):1569–1593.

Lindskog, F., McNeil, A., and Schmock, U. (2003). Kendall's tau for elliptical distributions. *In: Bol G., Nakhaeizadeh G., Rachev S.T., Ridder T., Vollmer KH. (eds) Credit Risk. Contributions to Economics. Physica-Verlag*, pages 149–156.

Liu, W.-D., Lin, Z., and Shao, Q.-M. (2008). The asymptotic distribution and berry: Esseen bound of a new test for independence in high dimension with an application to stochastic optimization. *The Annals of Applied Probability*, 18(6):2337–2366.

Nagao, H. (1973). On Some Test Criteria for Covariance Matrix. *The Annals of Statistics*, 1(4):700 – 709.

Narain, R. D. (1950). On the Completely Unbiassed Character of Tests of Independence in Multivariate Normal Systems. *The Annals of Mathematical Statistics*, 21(2):293 – 298.

Pearson, K. (1920). Notes on the history of correlation. *Biometrika*, 13(1):25–45.

Qiu, Y. and Chen, S. X. (2012). Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. *The Annals of Statistics*, 40(3).

Roy, S. (1957). *Some aspects of multivariate analysis.* Wiley. John Wiley & Sons, Inc., New York.

Schott, J. R. (2005). Testing for complete independence in high dimensions. *Biometrika*, 92(4):951–956.

Shao, Q.-M. and Zhou, W.-X. (2014). Necessary and sufficient conditions for the asymptotic distributions of coherence of ultra-high dimensional random matrices. *The Annals of Probability*, 42(2):623 – 648.

Song, Y., Chen, X., and Kato, K. (2019). Approximating high-dimensional infinite-order U-statistics: Statistical and computational guarantees. *Electronic Journal of Statistics*, 13(2):4794 – 4848.

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769 – 2794.

Wang, R., Zhu, C., Volgushev, S., and Shao, X. (2021). Inference for change points in high dimensional data via self-normalization. *arXiv preprint, arXiv:1905.08446*.

Wellek, S. (2010). *Testing statistical hypotheses of equivalence and noninferiority*. CRC Press.

Yao, S., Zhang, X., and Shao, X. (2018). Testing mutual independence in high dimension via distance covariance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(3):455–480.

Zhou, C., Han, F., Zhang, X.-S., and Liu, H. (2019). An extreme-value approach for testing the equality of large U-statistic based correlation matrices. *Bernoulli*, 25(2):1472 – 1503.

Zhou, W. (2007). Asymptotic distribution of the largest off-diagonal entry of correlation matrices. *Transactions of the American Mathematical Society*, 359(11):5345–5363.