High-dimensional Gaussian and bootstrap approximations for robust means

Anders Bredahl Kock*
University of Oxford
Department of Economics
10 Manor Rd, Oxford OX1 3UQ
anders.kock@economics.ox.ac.uk

David Preinerstorfer
WU Vienna University of Economics and Business
Institute for Statistics and Mathematics
Welthandelsplatz 1, 1020 Vienna
david.preinerstorfer@wu.ac.at

First version: April 9, 2025
This version: November, 2025

Abstract

Recent years have witnessed much progress on Gaussian and bootstrap approximations to the distribution of sums of independent random vectors with dimension d large relative to the sample size n. However, for any number of moments $m > 2$ that the summands may possess, there exist distributions such that these approximations break down if d grows faster than the polynomial barrier $n^{\frac{m}{2}-1}$. In this paper, we establish Gaussian and bootstrap approximations to the distributions of winsorized and trimmed means that allow d to grow at an exponential rate in n as long as $m > 2$ moments exist. The approximations remain valid under some amount of adversarial contamination. Our implementations of the winsorized and trimmed means do not require knowledge of m. As a consequence, the performance of the approximation guarantees "adapts" to m.

*Kock's research was supported by the European Research Council (ERC) grant number 101124535 – HIDI (UKRI EP/Z002222/1). He is also a member of, and grateful for support from, i) the Aarhus Center for Econometrics (ACE), funded by the Danish National Research Foundation grant number DNRF186, and ii) the Center for Research in Energy: Economics and Markets (CoRe).

1 Introduction

Let $X_1, \ldots, X_n$ be a sample of i.i.d. random vectors in $\mathbb{R}^d$ with mean vector μ and covariance matrix Σ. Furthermore, let $S_n = n^{-1/2} \sum_{i=1}^{n} (X_i - \mu)$. Since the seminal paper of Chernozhukov et al. (2013) there has been substantial interest in Gaussian approximations to the distribution of S_n when d is large relative to n. Letting $Z \sim \mathsf{N}_d(0, \Sigma)$ and $\mathcal{H}$ be the class of (generalized) hyperrectangles in $\mathbb{R}^d$, that is the class of all sets of the form

$$H = \left\{ x \in \mathbb{R}^d : a_j \leq x_j \leq b_j \text{ for all } j = 1, \ldots, d \right\},$$

where $-\infty \leq a_j \leq b_j \leq \infty$ for all $j = 1, \ldots, d$, increasingly refined upper bounds have been established on

$$\rho_n := \sup_{H \in \mathcal{H}} \Big| \mathbb{P}\left(S_n \in H\right) - \mathbb{P}\left(Z \in H\right) \Big| \tag{1}$$

and related quantities, cf., e.g., Chernozhukov et al. (2017a); Deng and Zhang (2020); Lopes et al. (2020); Kuchibhotla and Rinaldo (2020); Das and Lahiri (2021); Koike (2021); Kuchibhotla et al. (2021); Lopes (2022); Chernozhuokov et al. (2022); Fang et al. (2023); Chernozhukov et al. (2023b); Koike (2024). We refer to the review in Chernozhukov et al. (2023a) for further references. For example, when the entries of $X_i = (X_{i,1}, \ldots, X_{i,d})'$ are (uniformly) sub-exponential, $\rho_n \to 0$ if $d = d(n) = o\left(\exp(n^{1/5})\right)$, cf. Chernozhuokov et al. (2022). Thus, d can grow exponentially fast with n and this rate can be further improved under additional assumptions on the distribution of the X_i such as, e.g., variance decay conditions on Σ as in Lopes et al. (2020) or eigenvalue conditions as in Fang and Koike (2021); Kuchibhotla and Rinaldo (2020); Chernozhukov et al. (2023b).

Despite the progress on such high-dimensional Gaussian approximations for S_n, it follows from Remark 2 in Zhang and Wu (2017) and Theorem 2.1 in Kock and Preinerstorfer (2024) that for every $m \in (2, \infty)$ there exist i.i.d. random vectors $X_1, \ldots, X_n$ with independent entries $X_{ij} \sim P_m$, and P_m depending neither on n nor d, having mean zero, variance one, and finite mth absolute moment, such that if for some $\xi \in (0, \infty)$ it holds that $\limsup_{n\to\infty} \frac{d}{n^{m/2-1+\xi}} > 0$, then $\limsup_{n\to\infty} \rho_n = 1$. In particular, for any given $m \in (2, \infty)$, the Gaussian approximation $\rho_n \to 0$ does not hold uniformly over all distributions with bounded mth moments when d grows exponentially in n.

Conversely, it is a simple consequence of Theorem 2 in Chernozhukov et al. (2023a), cf. Theorem 2.2 in Kock and Preinerstorfer (2024), that $\rho_n \to 0$ uniformly over a large class of distributions with bounded mth moments if there exists a $\xi \in (0, \infty)$ such that $\lim_{n\to\infty} \frac{d}{n^{m/2-1-\xi}} = 0$. Hence, a critical phase transition occurs for the asymptotic behaviour of ρ_n at $d = n^{m/2-1}$. As d passes this threshold from below, the limit of ρ_n jumps from zero to one. For example, for $m = 3$ one can construct $X_{i,j}$ with bounded third moments such that $\rho_n \to 1$ if $d = n^{1/2+\xi}$ for ξ arbitrarily close to zero. Thus, even in a regime where d grows (much) *slower* than n, the Gaussian approximation to the distribution of S_n can break down completely if the X_i only possess three moments.

Motivated by this phase transition, Resende (2024) recently studied the case where S_n is replaced by a suitably *trimmed* mean and $\mathcal{H}$ is replaced by the subfamily of "one-sided" intervals $\mathcal{R}$, say, i.e., the class of all sets of the form $R = \{x \in \mathbb{R}^d : x_j \leq t_j \text{ for all } j = 1, \ldots, d\}$, where $t_j \in \mathbb{R}$ for all $j = 1, \ldots, d$. He obtained Gaussian approximations that are informative even when the X_i only possess $m > 2$ moments and d grows exponentially fast in n. This is of fundamental importance, as it shows that one can break through the barrier $d = n^{m/2-1}$ faced by ρ_n, which is based on S_n. The exact permitted growth rate of d depends on m. As $m \to \infty$, his result allows d to grow almost as fast as $\exp(n^{1/6})$ for the Gaussian approximation as well as an empirical bootstrap and as fast as $\exp(n^{1/8})$ for a multiplier bootstrap. Furthermore, the trimming ensures that these approximations remain valid even when some of the X_i have been adversarially contaminated prior to being given to the statistician. This is in stark contrast to statistics based on the sample mean S_n, which have a breakdown point of $1/n$ (S_n can be changed to any value by manipulating only one of the vectors X_i). A potential drawback of the trimmed mean studied in Resende (2024) is that the amount of trimming needed depends on m (cf. Theorem 2 in Resende (2024)), which is typically unknown in practice. Thus, if one constructs an estimator based on an m higher than the actual number of moments that the X_i possess one does not have any approximations guarantees for the trimmed mean, whereas the guarantees one obtains may be suboptimal if the X_i possess more moments than used in the construction of the trimmed mean.

We also mention the work of Liu and Lopes (2024) who, motivated by the poor performance of S_n in the presence of heavy tails, even established a dimension-independent bootstrap approximation over $\mathcal{R}$ for certain *robust max statistics* related to the winsorized means we study under the conditions of L^4-L^2 moment equivalence, a variance decay condition on Σ, and restrictions on the Frobenius norm of certain submatrices of the correlation matrix of the X_i. Robustness to outliers or other sources of (adversarial) data contamination were not investigated.

In this paper we do not impose any structural assumptions on Σ (apart from positive variances) and obtain Gaussian approximations to the distributions of winsorized and trimmed means over $\mathcal{H}$, which contains the family $\mathcal{R}$ studied in Resende (2024). The winsorization and trimming points are suitably chosen order statistics and we only require $\log(d) = o\left(n^{\frac{m-2}{5m-2}}\right)$, which is exponential for all $m > 2$ albeit with a small exponent for m close to two. Apart from our Gaussian approximations being valid over a larger family of sets, an important advantage over the trimmed mean studied in Resende (2024) is that one does not need to know the number of moments m that the X_i possess in order to implement our winsorized and trimmed means — they "adapt" to m. Furthermore, as $m \to \infty$, we allow d to grow almost as fast as $\exp(n^{1/5})$, improving on the rate in Resende (2024), and thus "recover" the best known rate (cf. Remark 2 in Chernozhukov et al. (2023a)) for Gaussian approximations based on the sample mean of X_i with *sub-exponential* entries. This rate remains valid for the bootstrap procedures that we consider and all results are robust to some adversarial contamination. Our bootstrap approximations are based on a novel covariance matrix estimator, for which we establish performance guarantees in the presence of adversarial contamination and only $m > 2$ moments in Section 3.1. This estimator does not require knowledge of any unknown population quantities. In particular, its performance guarantees adapt to the unknown m, which may be of independent interest.

In contrast to the present paper, which focuses exclusively on the canonical problem of Gaussian approximations in $\mathbb{R}^d$, Resende (2024) also considers Gaussian approximations over VC-subgraph classes of functions and applies his results to vector mean estimation under general norms.

2 Gaussian approximations for winsorized means

We first present our approximations to the distributions of high-dimensional winsorized means. Section 5 outlines the corresponding results for the version of the trimmed means we study.

Recall that $X_1, \ldots, X_n$ is a sample of i.i.d. random vectors in $\mathbb{R}^d$ with $X_i = (X_{i,1}, \ldots, X_{i,d})'$ for $i = 1, \ldots, n$. Let $\mu = (\mu_1, \ldots, \mu_d)' = \mathbb{E}X_1$, Σ be the covariance matrix of X_1, and for $m \in [2, \infty)$ let $\sigma_{m,j}^m := \mathbb{E}|X_{1,j} - \mu_j|^m$, all of which are well-defined under Assumption 2.1 below. We suppress the dependence of $d = d(n)$ on n in our notation.

In this section, our main focus is to establish Gaussian approximations for winsorized means that are valid for d growing exponentially in n imposing only that the $X_{1,j}$ possess $m > 2$ moments, $j = 1, \dots, d$. An added benefit of the winsorization is that the Gaussian approximations are robust to some amount of *adversarial contamination*. Under such contamination an adversary inspects the sample and returns a corrupted sample $\tilde{X}_1, \dots, \tilde{X}_n$ to the statistician satisfying that

$$\left|\left\{i \in \{1, \dots, n\} : \tilde{X}_i \neq X_i\right\}\right| \leq \overline{\eta}_n n, \tag{2}$$

where $\overline{\eta}_n \in (0, 1/2)$ is a non-random and known upper bound on the fraction of contaminated observations. Which of $\tilde{X}_i$ differ from X_i as well as their values can depend on the uncontaminated sample $X_1, \dots, X_n$. Adversarial contamination has become a popular criterion to evaluate robustness of a statistic against as it allows for many forms of data manipulation, cf. Lai et al. (2016), Cheng et al. (2019), Diakonikolas et al. (2019), Hopkins et al. (2020), Lugosi and Mendelson (2021), Minsker and Ndaoud (2021), Bhatt et al. (2022), Depersin and Lecué (2022), Dalalyan and Minasyan (2022), Minasyan and Zhivotovskiy (2023), Minsker (2023), Oliveira et al. (2025). The recent book by Diakonikolas and Kane (2023) provides further references and discussion of various contamination settings. Since the sample mean has a breakdown point of $1/n$, Gaussian approximations based on S_n are not robust to adversarial contamination (or large outliers).

In all asymptotic statements $n \to \infty$. Throughout, we impose the following assumption (for various values of m).

Assumption 2.1. The $X_1, \dots, X_n$ are i.i.d. random vectors in $\mathbb{R}^d$ with, $\mathbb{E}|X_{1,j}|^m < \infty$ for some $m \in (2, \infty)$ and all $j = 1, \dots, d$. Suppose that there exist strictly positive constants b_1 and b_2 such that $\min_{j=1,\dots,d} \sigma_{2,j} \geq b_1$ and $\sigma_m := \max_{j=1,\dots,d} \sigma_{m,j} \leq b_2$. The actually observed adversarially contaminated random vectors (in $\mathbb{R}^d$) are denoted $\tilde{X}_1, \dots, \tilde{X}_n$ and satisfy (2).

Imposing lower and upper bounds on moments of the $X_{1,j}$ is commonplace when establishing upper bounds on ρ_n in (1), cf., e.g., the results in the overview Chernozhukov et al. (2023a). Let us emphasize that all of our results are valid (in particular) *absent* adversarial contamination, i.e., for $\overline{\eta}_n = 0$, which is the case studied in the literature on upper bounds on ρ_n summarized in the introduction.

2.1 The winsorized means

For real numbers $x_1, \ldots, x_n$, denote by $x_1^* \leq \ldots \leq x_n^*$ their non-decreasing rearrangement. Let $-\infty < \alpha \leq \beta < \infty$ and

$$\phi_{\alpha,\beta}(x) = \begin{cases} \alpha & \text{if } x < \alpha \\ x & \text{if } x \in [\alpha, \beta] \\ \beta & \text{if } x > \beta. \end{cases}$$

We establish Gaussian and bootstrap approximations to the distribution of centered winsorized means $S_{n,W} \in \mathbb{R}^d$ where

$$S_{n,W,j} = n^{-1/2} \sum_{i=1}^{n} \left(\phi_{\hat{\alpha}_j, \hat{\beta}_j}(\tilde{X}_{i,j}) - \mu_j \right), \qquad j = 1, \ldots, d, \tag{3}$$

with $\hat{\alpha}_j = \tilde{X}^*_{\lceil \varepsilon_n n \rceil, j}$ and $\hat{\beta}_j = \tilde{X}^*_{\lceil (1-\varepsilon_n) n \rceil, j}$ for $\varepsilon_n \in (0, 1/2)$. Thus, for each coordinate j, the winsorization points $\hat{\alpha}_j$ and $\hat{\beta}_j$ are order statistics of the contaminated data $\tilde{X}_{1,j}, \ldots, \tilde{X}_{n,j}$.[1] Under adversarial contamination it is clear that even $S_{n,W}$ can perform arbitrarily badly unless at least the smallest and largest $\overline{\eta}_n n$ observations are winsorized. Thus, one must choose $\varepsilon_n \geq \overline{\eta}_n$. In particular, we study ε_n of the form

$$\varepsilon_n = \lambda_1 \cdot \overline{\eta}_n + \lambda_2 \cdot \frac{\log(dn)}{n}, \qquad \lambda_1 \in (1, \infty) \text{ and } \lambda_2 \in (0, \infty).$$

To make efficient use of the data, it is desirable to establish Gaussian approximations with ε_n, and thus λ_1 and λ_2, as small as possible. We now discuss the choice of λ_1 and λ_2.

Consider first the case of $\overline{\eta}_n = 0$ (no contamination), which is the setting in which high-dimensional Gaussian approximations for the sample mean based S_n have been studied primarily (cf. the literature summarized in the introduction). In this case, one can choose $\lambda_2 = 6.05$ whereas the choice of λ_1 is irrelevant. In fact, as will be seen in Remark 2.1, even smaller choices of λ_2 are possible.

When one suspects that the data may have been contaminated, corresponding to $\overline{\eta}_n > 0$, there is a tradeoff between the sizes of λ_1 and λ_2 in our implementation of the winsorized

[1] For the purpose of construction estimators of $\mu \in \mathbb{R}$ with finite-sample sub-Gaussian concentration properties, related winsorized mean estimators were recently studied in Lugosi and Mendelson (2021) and Kock and Preinerstorfer (2025).

means, which we parameterize by $c \in (1, \sqrt{1.5})$. In particular, following Kock and Preinerstorfer (2025), one can choose

$$\lambda_1 = \lambda_{1,c} = \frac{c}{1 - \sqrt{2(c^2-1)}}$$

and

$$\lambda_2 = \lambda_{2,c} = \lambda_{2,c}(n,d) = \left[\frac{c}{3[1-\sqrt{2(c^2-1)}]} \vee c\left(\sqrt{2\frac{c+1}{c-1}} + \frac{1}{3}\right)\right] \wedge c\left(\sqrt{\frac{n}{2\log(dn)}} + \frac{1}{3}\right). \tag{4}$$

In the sequel ε_n refers to (we suppress its dependence on $c \in (1, \sqrt{1.5})$ and d)

$$\varepsilon_n = \lambda_{1,c} \cdot \overline{\eta}_n + \lambda_{2,c} \cdot \frac{\log(dn)}{n}. \tag{5}$$

Since $c \mapsto \lambda_{1,c}$ is a (strictly increasing) bijection from $(1, \sqrt{1.5})$ to $(1, \infty)$, any value of $\lambda_{1,c} \in (1, \infty)$ can be achieved by a suitable choice of c.[2]

Remark 2.1. In the important case of $\overline{\eta}_n = 0$ one can nearly minimize $\lambda_{2,c}$, and thus the number of winsorized observations, by checking whether i) $c = \tilde{c}$ with $\tilde{c}$ being the minimizer of the term $f(c)$ in square brackets in (4) or ii) c arbitrarily close to one yields the smallest value of $\lambda_{2,c}$.[3] In particular, $\lambda_{2,c}$ chosen in this way will never exceed 6.05 since $f(\tilde{c}) \leq 6.05$.

2.2 Gaussian approximation

We now present a high-dimensional Gaussian approximation result for $S_{n,W}$ over $\mathcal{H}$ (defined prior to (1)) in the form of an upper bound on

$$\rho_{n,W} := \sup_{H \in \mathcal{H}} \Big| \mathbb{P}\left(S_{n,W} \in H\right) - \mathbb{P}\left(Z \in H\right) \Big|, \qquad \text{where } Z \sim \mathsf{N}_d(0, \Sigma),$$

and where the dependence of $S_{n,W}$ on c, via ε_n, is suppressed notationally. We assume throughout that $d \geq 2$ and $n > 3$ (such that, e.g., $\sqrt{\log(d)} > 0$).

[2] To achieve $\lambda_{1,c} = A \in (1, \infty)$, set $c = \frac{\sqrt{2}\sqrt{3A^4 - A^2} - A}{2A^2 - 1}$.

[3] To see this, note that $f(c) = \left[\frac{c}{3[1-\sqrt{2(c^2-1)}]} \vee c\left(\sqrt{2\frac{c+1}{c-1}} + \frac{1}{3}\right)\right]$ is minimized by equating the two terms at $\tilde{c} = \frac{1}{17}(-4 + 3\sqrt{66}) \approx 1.198$ which results in $f(\tilde{c}) = 6.041346$ whereas $c \mapsto c\left(\sqrt{\frac{n}{2\log(dn)}} + \frac{1}{3}\right)$ is strictly increasing in c for given values of n and d.

Theorem 2.1. *Fix $c \in (1, \sqrt{1.5})$, and let Assumption 2.1 be satisfied with $m > 2$. If $\varepsilon_n \in (0, 1/2)$, with ε_n as in (5), then*

$$\rho_{n,W} \le C\left(\left[\frac{\log^{5-\frac{2}{m}}(dn)}{n^{1-\frac{2}{m}}}\right]^{\frac{1}{4}} + \left[\overline{\eta}_n^{1-\frac{1}{m}} + \left[\frac{\log(dn)}{n}\right]^{1-\frac{1}{m}}\right]\sqrt{n\log(d)}\right) + C\left(\log^2(d)\left[\overline{\eta}_n^{1-\frac{2}{m}} + \left[\frac{\log(dn)}{n}\right]^{1-\frac{2}{m}}\right]\right)^{1/2}, \tag{6}$$

where C is a constant depending only on b_1, b_2, c and m.

In particular, $\rho_{n,W} \to 0$ if $\sqrt{n\log(d)}\overline{\eta}_n^{1-\frac{1}{m}} \to 0$ and $\log(d)/n^{\frac{m-2}{5m-2}} \to 0$.

Consider the case of $\overline{\eta}_n = 0$. Theorem 2.1 then shows that winsorized means can break through the *polynomial* growth rate barrier $d = n^{m/2-1}$ that d must obey for the Gaussian approximation error to the distribution of S_n, i.e., ρ_n in (1), to converge to zero. In particular, $\rho_{n,W} \to 0$ if only $\log(d) = o\left(n^{\frac{m-2}{5m-2}}\right)$, which allows for d growing *exponentially* in n for any $m > 2$ albeit with a small exponent for m close to two.

Thus, the winsorized mean obeys a Gaussian approximation result over $\mathcal{H}$ under heavy tails and adversarial contamination, which is in analogy to the Gaussian approximation result over $\mathcal{R} \subseteq \mathcal{H}$ in Resende (2024) for the trimmed mean analyzed there. However, as discussed in the introduction, the dependence of the trimmed mean estimator in Resende (2024) on m implies that if the X_i have fewer moments than the chosen m then there are no approximation guarantees. If, on the other hand, the X_i have more moments than the specified m then the guarantees may be suboptimal. In contrast, the implementation of our winsorized mean does not depend on m — it "adapts" to it. Furthermore, we recover the rate $d = o\left(\exp(n^{1/5})\right)$ as $m \to \infty$, which is currently the best available for Gaussian approximations based on S_n with sub-exponential X_i, instead of $d = o\left(\exp(n^{1/6})\right)$ for the trimmed mean analyzed in Resende (2024). In case the entries of the X_i are heavy-tailed in the sense of possessing exactly, e.g., $m = 4$ moments, $\rho_{n,W} \to 0$ if $\log(d) = o(n^{1/9})$ whereas the trimmed mean analyzed in Resende (2024) allows $\log(d) = o(n^{3/35})$ if implemented with $m = 3$, $\log(d) = o(n^{4/35})$ if implemented with $m = 4$, and provides no guarantees if implemented with $m > 4$. Thus, unless one has reliable information on the number of moments the data possesses (knows $m \approx 4$), our method adapting to m allows for larger d.

To prove Theorem 2.1, we first show that the order statistics $\hat{\alpha}_j$ and $\hat{\beta}_j$ can (essentially) be replaced by closely related population quantiles $Q_{\varepsilon_n,j}$ and $Q_{1-\varepsilon_n,j}$ of the $X_{1,j}$ such that one can analyze the non-random winsorization functions $\phi_n(x) := \phi_{Q_{\varepsilon_{n,j}}, Q_{1-\varepsilon_{n,j}}}(x)$ instead of $\phi_{\hat{\alpha}_j, \hat{\beta}_j}(x)$ for all $j = 1, \ldots, d$:[4]

$$
\begin{aligned}
S_{n,W,j} \approx \underbrace{\frac{1}{n^{1/2}} \sum_{i=1}^{n} \left[\phi_n(\tilde{X}_{i,j}) - \mu_j\right]}_{T_{n,j}} &= \underbrace{\frac{1}{n^{1/2}} \sum_{i=1}^{n} \left[\phi_n(\tilde{X}_{i,j}) - \phi_n(X_{i,j})\right]}_{I_{n,j,1}} \\
&+ \underbrace{\frac{1}{n^{1/2}} \sum_{i=1}^{n} \left[\phi_n(X_{i,j}) - \mathbb{E}\phi_n(X_{i,j})\right]}_{I_{n,j,2}} \\
&+ \underbrace{\frac{1}{n^{1/2}} \sum_{i=1}^{n} \left[\mathbb{E}\phi_n(X_{i,j}) - \mu_j\right]}_{I_{n,j,3}} .
\end{aligned}
$$

The term $\max_{j=1,\ldots,d} |I_{n,j,1}|$ isolates the effect of the adversarial contamination of the data and $\max_{j=1,\ldots,d} |I_{n,j,3}|$ quantifies how far the winsorized means $\mathbb{E}\phi_n(X_{1,j})$ are from the population means μ_j of interest. Next, note that $I_{n,j,2}$ is a sample average of *bounded* i.i.d. random variables. Thus, letting $Z_n \sim \mathsf{N}_d(0, \Sigma_{\phi_n})$ with Σ_{ϕ_n} being the covariance matrix of $I_{n,2} = (I_{n,1,2}, \ldots, I_{n,d,2})'$, one can apply, e.g., the Gaussian approximation for sums of sub-exponential random vectors from Chernozhuokov et al. (2022) to this term to show that

$$
\sup_{H \in \mathcal{H}} \left| \mathbb{P}\left(I_{n,2} \in H\right) - \mathbb{P}\left(Z_n \in H\right) \right| \qquad \text{is small.}
$$

Furthermore, we show that $\max_{1 \leq j,k \leq d} |\Sigma_{\phi_n,j,k} - \Sigma_{j,k}|$ is sufficiently small for the Gaussian-to-Gaussian comparison inequality as stated in Proposition 2.1 in the previous reference (cf. also Proposition 2 in Chernozhukov et al. (2023a)) to imply that

$$
\mathbb{P}\left(Z_n \in H\right) \qquad \text{in the previous display can be replaced by} \qquad \mathbb{P}\left(Z \in H\right).
$$

Finally, we show that for $l \in \{1, 3\}$ one has that $\max_{j=1,\ldots,d} |I_{n,j,l}|$ are sufficiently small for a Gaussian anti-concentration inequality to imply that these can be "ignored". Therefore, $\mathbb{P}\left(I_{n,2} \in H\right)$ can be replaced by $\mathbb{P}\left(T_n \in H\right)$ in the penultimate display, where $T_n =$

[4]A similar control of the order statistics was also used in Lugosi and Mendelson (2021). Resende (2024) studied the trimmed mean by relating it to the winsorized mean.

$(T_{n,1},\ldots,T_{n,d})'$.

3 Bootstrap approximations for winsorized means

Similarly to Gaussian approximations for S_n, cf. the references in the introduction, the one for $S_{n,W}$ in Theorem 2.1 is not directly useful for statistical inference since the covariance matrix Σ of the approximating distribution $\mathsf{N}_d(0,\Sigma)$ is typically unknown. Because the Gaussian-to-Gaussian comparison inequality as stated in Proposition 2.1 in Chernozhuokov et al. (2022) (cf. also Proposition 2 in Chernozhukov et al. (2023a)) shows that for $Z_1 \sim \mathsf{N}_d(0,\Sigma^{(1)})$ and $Z_2 \sim \mathsf{N}_d(0,\Sigma^{(2)})$ with $\min_{j=1,\ldots,d}\Sigma^{(2)}_{j,j} > b$ for some $b > 0$, it holds that

$$\sup_{H\in\mathcal{H}}\Big|\mathbb{P}\left(Z_1 \in H\right) - \mathbb{P}\left(Z_2 \in H\right)\Big| \leq C\left(\max_{1\leq j,k\leq d}\big|\Sigma^{(1)}_{j,k} - \Sigma^{(2)}_{j,k}\big|\log^2(d)\right)^{1/2}, \tag{7}$$

for some $C = C(b)$, one can approximate the unknown $\mathbb{P}(Z \in H)$ from Theorem 2.1 if an estimator $\hat{\Sigma}_n$ satisfying an upper bound on $\log^2(d)\max_{1\leq j,k\leq d}\big|\hat{\Sigma}_{n,j,k} - \Sigma_{j,k}\big|$ can be exhibited. The sample covariance matrix can be used for d growing exponentially in n when i) $\overline{\eta}_n = 0$ and ii) the X_i have sub-exponential entries.[5] However, since we allow for $\overline{\eta}_n > 0$ and only impose the existence of $m > 2$ moments, the sample covariance matrix cannot be used in our context.

There has been a recent interest in constructing estimators of Σ which perform well under heavy tails (and frequently also adversarial contamination). For example, estimators with precision guarantees in the entrywise maximal distance $\max_{1\leq j,k\leq d}\big|\hat{\Sigma}_{n,j,k} - \Sigma_{j,k}\big|$ needed in (7), have been proposed in Ke et al. (2019) in the setting of heavy-tailed X_i. These estimators are based on, e.g., entrywise truncation or the median-of-means principle and the practical choice of the needed tuning parameters (which depend on unknown population quantities) is also discussed there.

In the next section we construct the estimator $\tilde{\Sigma}_n$ based on suitably winsorized observations $\tilde{X}_i$, for which we establish performance guarantees even for $\bar{\eta}_n > 0$ and when the X_i possess only $m > 2$ moments. Our estimator does not depend on unknown population quantities and its performance guarantees "adapt" to the unknown m. We stress that Theorem 3.2 below is modular in the sense that it remains valid for any estimator $\hat{\Sigma}_n$ satisfying a bound as in Theorem 3.1.

[5] See Section 4.1 of Kuchibhotla and Chakrabortty (2022) for properties of the sample covariance matrix when the X_i have sub-Weibull entries (generalizing sub-exponential distributions).

3.1 Estimating Σ

Imposing only $m > 2$ moments to exist we now construct an estimator of Σ with precision guarantees in the maximal entrywise norm. To avoid making assumptions on μ, let

$$Y_i = \frac{1}{\sqrt{2}}\left(X_{2i} - X_{2i-1}\right) \qquad \text{and} \qquad \tilde{Y}_i = \frac{1}{\sqrt{2}}\left(\tilde{X}_{2i} - \tilde{X}_{2i-1}\right), \qquad i = 1, \dots, \lfloor n/2 \rfloor. \tag{8}$$

Clearly, $Y_1, \dots, Y_{\lfloor n/2 \rfloor}$ are i.i.d. mean zero with covariance matrix Σ. *In the sequel we assume for convenience that n is even.* Let $N := n/2$ and set

$$\varepsilon_n' = c2\overline{\eta}_n + c\sqrt{\frac{\log(d^2N)}{2N}}, \qquad c \in (1, \infty). \tag{9}$$

Writing $\hat{a}_j = \tilde{Y}^*_{\lceil \varepsilon_n' N \rceil, j}$ and $\hat{b}_j = \tilde{Y}^*_{\lceil (1-\varepsilon_n')N \rceil, j}$ for $j = 1, \dots, d$, define $\tilde{\Sigma}_n$ as the matrix with entries

$$\tilde{\Sigma}_{n,j,k} = \frac{1}{N}\sum_{i=1}^{N} \phi_{\hat{a}_j, \hat{b}_j}(\tilde{Y}_{i,j})\phi_{\hat{a}_k, \hat{b}_k}(\tilde{Y}_{i,k}), \qquad 1 \leq j, k \leq d. \tag{10}$$

$\tilde{\Sigma}_n$ is positive semi-definite and symmetric by virtue of being a Gram matrix and obeys the following precision guarantee.

Theorem 3.1. *Fix $c \in (1, \infty)$, and let Assumption 2.1 be satisfied with $m > 2$. If $\varepsilon_n' \in (0, 1/2)$, with ε_n' as in (9), then for a constant $C = C(b_2, c, m)$,*

$$\mathbb{P}\left(\max_{1 \leq j,k \leq d}\left|\tilde{\Sigma}_{n,j,k} - \Sigma_{j,k}\right| > C\left[\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n}\right)^{\frac{1}{2}-\frac{1}{m}}\right]\right) \leq \frac{24}{n}. \tag{11}$$

Theorem 3.1 shows that for any $m > 2$ it is possible for $\max_{1 \leq j,k \leq d}|\tilde{\Sigma}_{n,j,k} - \Sigma_{j,k}|$ to converge to zero in probability even when d grows exponentially in n. Finally, we reiterate that the approximation in Theorem 3.2 below remains valid for *any* estimator from the burgeoning literature on large covariance matrix estimation satisfying (11): the result is not specific to $\tilde{\Sigma}_n$.

3.2 Bootstrap consistency

Equipped with the estimator $\tilde{\Sigma}_n$, the following theorem justifies approximating $\mathbb{P}\left(S_{n,W} \in H\right)$ for $H \in \mathcal{H}$ by sampling repeatedly from $\mathsf{N}_d(0, \tilde{\Sigma}_n)$.

Theorem 3.2. *Fix $c \in (1, \sqrt{1.5})$, and let Assumption 2.1 be satisfied with $m > 2$. Let $n > 24$. If $\varepsilon_n, \varepsilon_n' \in (0, 1/2)$, with ε_n as in (5) and ε_n' as in (9), and $\tilde{Z} \sim \mathsf{N}_d(0, \tilde{\Sigma}_n)$ conditionally on $\tilde{X}_1, \ldots, \tilde{X}_n$, it holds with probability at least $1 - \frac{24}{n}$ that*

$$\begin{aligned}\tilde{\rho}_{n,W} &:= \sup_{H \in \mathcal{H}} \left| \mathbb{P}\left(S_{n,W} \in H\right) - \mathbb{P}\left(\tilde{Z} \in H \mid \tilde{X}_1, \ldots, \tilde{X}_n\right) \right| \\ &\leq \mathfrak{A}_n + C\left(\log^2(d)\left[\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n}\right)^{\frac{1}{2}-\frac{1}{m}}\right]\right)^{1/2},\end{aligned}$$

where $\mathfrak{A}_n$ is the upper bound on $\rho_{n,W}$ in (6) of Theorem 2.1 and C is a constant depending only on b_1, b_2, c and m.

In particular, $\tilde{\rho}_{n,W} \to 0$ in probability if $\sqrt{n \log(d)} \overline{\eta}_n^{1-\frac{1}{m}} \to 0$ and $\log(d)/n^{\frac{m-2}{5m-2}} \to 0$.

Draws from $\mathsf{N}_d(0, \tilde{\Sigma}_n)$ can be obtained efficiently by, e.g., the following multiplier bootstrap: Let $\xi_1, \ldots, \xi_N$ be i.i.d. $N_1(0,1)$, independent of $\tilde{X}_1, \ldots, \tilde{X}_n$, and set

$$S_{n,\mathrm{MB}} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_i V_i \qquad \text{where} \qquad V_i = \left(\phi_{\hat{a}_1, \hat{b}_1}(\tilde{Y}_{i,1}), \ldots, \phi_{\hat{a}_d, \hat{b}_d}(\tilde{Y}_{i,d})\right)' \in \mathbb{R}^d, \; i = 1, \ldots, N.$$

Observe that conditionally on $\tilde{X}_1, \ldots, \tilde{X}_n$ the distribution of $S_{n,\mathrm{MB}}$ is $\mathsf{N}_d(0, \tilde{\Sigma}_n)$. Generating a draw from $\mathsf{N}_d(0, \tilde{\Sigma}_n)$ via this multiplier bootstrap may be numerically preferable to first drawing from $\mathsf{N}_d(0, \mathsf{I}_d)$ and then premultiplying this by $\tilde{\Sigma}_n^{1/2}$ since the calculation of the matrix square root may be costly for d large.

Remark 3.1 ($m \geq 4$). The results so far in this section have imposed $m > 2$ only. In case $m \geq 4$, implying that the $X_{1,j} X_{1,k}$ have second moments for all $1 \leq j, k \leq d$, one can choose

$$\varepsilon_n' = 2\lambda_{1,c}\overline{\eta}_n + \lambda_{2,c}(N, d^2)\frac{\log(d^2 N)}{N}, \qquad \text{for } c \in (1, \sqrt{1.5}) \tag{12}$$

where the second summand tends to zero faster in $N = n/2$ than the second summand in (9), cf. Remark D.2 in the appendix. The resulting estimator, Σ_n', say, satisfies

$$\mathbb{P}\left(\max_{1 \leq j,k \leq d} \left|\check{\Sigma}_{n,j,k} - \Sigma_{j,k}\right| > C\left[\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n}\right)^{\frac{1}{2}} + \left(\frac{\log(dn)}{n}\right)^{1-\frac{2}{m}}\right]\right) \leq \frac{24}{n},$$

where C is a constant depending only on b_2, c, and m. The bootstrap consistency of

Theorem 3.2 also carries over with this improved rate when $\check{\Sigma}_n$ replaces $\tilde{\Sigma}_n$.

4 Normalized winsorized means

In practice one often normalizes the data by an estimate of $\sigma_{2,j}$ to bring the variables on the same scale. We now describe how the Gaussian and bootstrap approximations established so far remain valid upon normalization by the diagonal elements $\tilde{\sigma}_{n,j} = \tilde{\Sigma}_{n,j,j}^{1/2}$ of the robust estimator $\tilde{\Sigma}_n$ of Σ, cf. (10). Writing $D = \operatorname{diag}(\sigma_{2,1}, \ldots, \sigma_{2,d})$ and $\Sigma_0 = D^{-1}\Sigma D^{-1}$ for the correlation matrix of the (centered) uncontaminated $X_1, \ldots, X_n$, one has the following Gaussian approximation for the vector $S_{n,W,S}$ of normalized winsorized means with elements (we leave the quotients undefined if one of the variance estimators equals 0)

$$S_{n,W,S,j} = \frac{1}{\sqrt{n}\tilde{\sigma}_{n,j}} \sum_{i=1}^{n} \left[\phi_{\hat{\alpha}_j,\hat{\beta}_j}(\tilde{X}_{i,j}) - \mu_j\right], \qquad j = 1, \ldots, d. \tag{13}$$

Theorem 4.1. *Fix $c \in (1, \sqrt{1.5})$, and let Assumption 2.1 be satisfied with $m > 2$. If $\varepsilon_n, \varepsilon_n' \in (0, 1/2)$, with ε_n as in (5) and ε_n' as in (9), then for $Z' \sim \mathsf{N}_d(0, \Sigma_0)$,*

$$\begin{aligned}\rho_{n,W,S} &:= \sup_{H\in\mathcal{H}} \Big|\mathbb{P}\left(S_{n,W,S} \in H\right) - \mathbb{P}\left(Z' \in H\right)\Big| \\ &\le C\left(\mathfrak{A}_n + \sqrt{\log(d)\log(dn)}\left[\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n}\right)^{\frac{1}{2}-\frac{1}{m}}\right]\right),\end{aligned} \tag{14}$$

where $\mathfrak{A}_n$ is the upper bound on $\rho_{n,W}$ in (6) of Theorem 2.1 and C is a constant depending only on b_1, b_2, c and m.

In particular, $\rho_{n,W,S} \to 0$ if $\sqrt{n\log(d)}\overline{\eta}_n^{1-\frac{1}{m}} \to 0$ and $\log(d)/n^{\frac{m-2}{5m-2}} \to 0$.

Theorem 4.1 allows for the same (exponential) growth rate of d (for all $m > 2$) and contamination rate $\overline{\eta}_n$ as Theorem 2.1 for non-normalized data. In analogy to Σ in Theorem 2.1, Σ_0 is unknown in Theorem 4.1. Letting $\tilde{D}_n = \operatorname{diag}(\tilde{\sigma}_{n,1}, \ldots, \tilde{\sigma}_{n,d})$ and $\tilde{\Sigma}_{n,0} = \tilde{D}_n^{-1}\tilde{\Sigma}_n\tilde{D}_n^{-1}$, we have the following analogue to the bootstrap approximation in Theorem 3.2.

Theorem 4.2. *Fix $c \in (1, \sqrt{1.5})$, and let Assumption 2.1 be satisfied with $m > 2$. Let $n > 24$. If $\varepsilon_n, \varepsilon_n' \in (0, 1/2)$, with ε_n as in (5) and ε_n' as in (9), and $\tilde{Z}' \sim \mathsf{N}_d(0, \tilde{\Sigma}_{n,0})$*

conditionally on $\tilde{X}_1, \ldots, \tilde{X}_n$, it holds with probability at least $1 - \frac{24}{n}$ that

$$\begin{aligned}\tilde{\rho}_{n,W,S} &:= \sup_{H \in \mathcal{H}} \left| \mathbb{P}\left(S_{n,W,S} \in H\right) - \mathbb{P}\left(\tilde{Z}' \in H \mid \tilde{X}_1, \ldots, \tilde{X}_n\right) \right| \\ &\leq \mathfrak{B}_n + C \left(\log^2(d) \left[\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n} \right)^{\frac{1}{2} - \frac{1}{m}} \right] \right)^{1/2}, \end{aligned} \tag{15}$$

where $\mathfrak{B}_n$ is the upper bound on $\rho_{n,W,S}$ in (14) of Theorem 4.1 and C is a constant depending only on b_1, b_2, c and m.

In particular, $\tilde{\rho}_{n,W,S} \to 0$ in probability if $\sqrt{n \log(d)} \overline{\eta}_n^{1-\frac{1}{m}} \to 0$ and $\log(d)/n^{\frac{m-2}{5m-2}} \to 0$.

5 Trimmed means

In this section we show that the Gaussian and bootstrap approximations established so far for *winsorized* means carry over to (suitably) *trimmed* means allowing for exactly the same growth rates of $\overline{\eta}_n$ and d. For ε_n as in (5), let $I_n := \left\{ \lceil \varepsilon_n n \rceil, \ldots, \lceil (1 - \varepsilon_n) n \rceil \right\}$ with cardinality

$$|I_n| = \lceil (1 - \varepsilon_n) n \rceil - \lceil \varepsilon_n n \rceil + 1 = n - \lfloor \varepsilon_n n \rfloor - \lceil \varepsilon_n n \rceil + 1,$$

and consider the vector of trimmed means $S_{n,T} \in \mathbb{R}^d$ with entries ($j = 1, \ldots, d$)

$$S_{n,T,j} = \frac{\sqrt{n}}{|I_n|} \sum_{i \in I_n} \left[\tilde{X}^*_{i,j} - \mu_j \right] = \frac{\sqrt{n}}{n - \lfloor \varepsilon_n n \rfloor - \lceil \varepsilon_n n \rceil + 1} \sum_{i = \lceil \varepsilon_n n \rceil}^{\lceil (1-\varepsilon_n) n \rceil} \left[\tilde{X}^*_{i,j} - \mu_j \right]. \tag{16}$$

Analogously to the winsorized means in (3), the trimmed mean in (16) can be implemented in a fully data-driven way: Its amount of trimming is governed by ε_n in (5), which does not depend on any unknown population quantities.[6] This sets this trimmed mean apart from the one studied in Theorem 2 in Resende (2024) where the amount of trimming depends on m. Furthermore, the bound in Theorem 5.1 below is valid over the larger family of sets $\mathcal{H}$. In analogy to $\rho_{n,W}$ define

$$\rho_{n,T} = \sup_{H \in \mathcal{H}} \left| \mathbb{P}\left(S_{n,T} \in H\right) - \mathbb{P}\left(Z \in H\right) \right|.$$

[6] As is typical in the literature on inference under data contamination, an upper bound $\overline{\eta}_n$ on the contamination rate must be supplied, however.

The following theorem is the trimmed mean counterpart to Theorem 2.1.

Theorem 5.1. *Fix $c \in (1, \sqrt{1.5})$, and let Assumption 2.1 be satisfied with $m > 2$. If $\varepsilon_n \in (0, 1/2)$, with ε_n as in (5), then*

$$\rho_{n,T} \leq \mathfrak{A}_n + C\sqrt{n \log(d)} \left(\overline{\eta}_n^{1-\frac{1}{m}} + \left[\frac{\log(dn)}{n}\right]^{1-\frac{1}{m}} \right), \tag{17}$$

where $\mathfrak{A}_n$ is the upper bound on $\rho_{n,W}$ in (6) of Theorem 2.1 and C is a constant depending only on b_1, b_2, c and m.

In particular, $\rho_{n,T} \to 0$ if $\sqrt{n \log(d)} \overline{\eta}_n^{1-\frac{1}{m}} \to 0$ and $\log(d)/n^{\frac{m-2}{5m-2}} \to 0$.

Theorem 5.1 is proven by establishing that $S_{n,T}$ is sufficiently close to $S_{n,W}$ in the supremum-norm order to use a Gaussian anti-concentration inequality to deduce (17) from (6) in Theorem 2.1. This explains the presence of the additional summand on the right-hand side in (17). The same discussion as that following Theorem 2.1 also applies to the trimmed mean in (16).

The following theorem is the trimmed mean analogue to Theorem 3.2 on bootstrap approximation to the distribution of winsorized means.

Theorem 5.2. *Fix $c \in (1, \sqrt{1.5})$, and let Assumption 2.1 be satisfied with $m > 2$. Let $n > 24$. If $\varepsilon_n, \varepsilon_n' \in (0, 1/2)$, with ε_n as in (5) and ε_n' as in (9), and $\tilde{Z} \sim \mathsf{N}_d(0, \tilde{\Sigma}_n)$ conditionally on $\tilde{X}_1, \ldots, \tilde{X}_n$, it holds with probability at least $1 - \frac{24}{n}$ that*

$$\begin{aligned}\tilde{\rho}_{n,T} &:= \sup_{H \in \mathcal{H}} \left| \mathbb{P}\left(S_{n,T} \in H\right) - \mathbb{P}\left(\tilde{Z} \in H \mid \tilde{X}_1, \ldots, \tilde{X}_n\right) \right| \\ &\leq \mathfrak{C}_n + C\left(\log^2(d) \left[\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n}\right)^{\frac{1}{2}-\frac{1}{m}} \right] \right)^{1/2},\end{aligned}$$

where $\mathfrak{C}_n$ is the upper bound on $\rho_{n,T}$ in (17) of Theorem 5.2 and C is a constant depending only on b_1, b_2, c and m.

In particular, $\tilde{\rho}_{n,W} \to 0$ in probability if $\sqrt{n \log(d)} \overline{\eta}_n^{1-\frac{1}{m}} \to 0$ and $\log(d)/n^{\frac{m-2}{5m-2}} \to 0$.

Theorem 5.2 allows for exactly the same growth rates of $\overline{\eta}_n$ and d as Theorem 3.2 for winsorized means. Note that we have chosen to keep the estimator $\tilde{\Sigma}_n$, which is based on winsorized means. Of course, one could also use a trimmed mean based estimator obeying the same performance guarantees (in fact, *any* estimator obeying the same performance guarantees as $\tilde{\Sigma}_n$ in (11) suffices, cf. also the discussion following (7)).

References

BHATT, S., G. FANG, P. LI, AND G. SAMORODNITSKY (2022): "Minimax m-estimation under adversarial contamination," in *International Conference on Machine Learning*, PMLR, 1906–1924.

CHENG, Y., I. DIAKONIKOLAS, AND R. GE (2019): "High-dimensional robust mean estimation in nearly-linear time," in *Proceedings of the thirtieth annual ACM-SIAM symposium on discrete algorithms*, SIAM, 2755–2771.

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

——— (2017a): "Central limit theorems and bootstrap in high dimensions," *Annals of Probability*, 45, 2309–2352.

——— (2017b): "Detailed proof of Nazarov's inequality," *arXiv preprint arXiv:1711.10696*.

CHERNOZHUKOV, V., D. CHETVERIKOV, K. KATO, AND Y. KOIKE (2023a): "High-dimensional data bootstrap," *Annual Review of Statistics and Its Application*, 10, 427–449.

CHERNOZHUKOV, V., D. CHETVERIKOV, AND Y. KOIKE (2023b): "Nearly optimal central limit theorem and bootstrap approximations in high dimensions," *Annals of Applied Probability*, 33, 2374–2425.

CHERNOZHUOKOV, V., D. CHETVERIKOV, K. KATO, AND Y. KOIKE (2022): "Improved central limit theorem and bootstrap approximations in high dimensions," *Annals of Statistics*, 50, 2562–2586.

CHOW, Y. S. AND W. J. STUDDEN (1969): "Monotonicity of the variance under truncation and variations of Jensen's inequality," *The Annals of Mathematical Statistics*, 40, 1106–1108.

DALALYAN, A. S. AND A. MINASYAN (2022): "All-in-one robust estimator of the gaussian mean," *Annals of Statistics*, 50, 1193–1219.

Das, D. and S. Lahiri (2021): "Central Limit Theorem in high dimensions: The optimal bound on dimension growth rate," *Transactions of the American Mathematical Society*, 374, 6991–7009.

Deng, H. and C.-H. Zhang (2020): "Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors," *Annals of Statistics*, 48, 3643–3671.

Depersin, J. and G. Lecué (2022): "Robust sub-Gaussian estimation of a mean vector in nearly linear time," *Annals of Statistics*, 50, 511–536.

Diakonikolas, I., G. Kamath, D. Kane, J. Li, A. Moitra, and A. Stewart (2019): "Robust estimators in high-dimensions without the computational intractability," *SIAM Journal on Computing*, 48, 742–864.

Diakonikolas, I. and D. Kane (2023): *Algorithmic high-dimensional robust statistics*, Cambridge University Press.

Fang, X. and Y. Koike (2021): "High-dimensional central limit theorems by Stein's method," *Annals of Applied Probability*, 31, 1660–1686.

Fang, X., Y. Koike, S.-H. Liu, and Y.-K. Zhao (2023): "High-dimensional Central Limit Theorems by Stein's method in the degenerate case," *arXiv preprint arXiv:2305.17365*.

Hopkins, S., J. Li, and F. Zhang (2020): "Robust and heavy-tailed mean estimation made simple, via regret minimization," *Advances in Neural Information Processing Systems*, 33, 11902–11912.

Ke, Y., S. Minsker, Z. Ren, Q. Sun, and W.-X. Zhou (2019): "User-friendly covariance estimation for heavy-tailed distributions," *Statistical Science*, 34, 454–471.

Kock, A. B. and D. Preinerstorfer (2024): "A remark on moment-dependent phase transitions in high-dimensional Gaussian approximations," *Statistics & Probability Letters*, 211, 110149.

——— (2025): "Winsorized mean estimation with heavy tails and adversarial contamination," *arXiv preprint 2504.08482*.

Koike, Y. (2021): "Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles," *Japanese Journal of Statistics and Data Science*, 4, 257–297.

——— (2024): "High-dimensional bootstrap and asymptotic expansion," *arXiv preprint arXiv:2404.05006*.

Kuchibhotla, A. K. and A. Chakrabortty (2022): "Moving beyond sub-Gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression," *Information and Inference: A Journal of the IMA*, 11, 1389–1456.

Kuchibhotla, A. K., S. Mukherjee, and D. Banerjee (2021): "High-dimensional CLT: Improvements, non-uniform extensions and large deviations," *Bernoulli*, 27, 192–217.

Kuchibhotla, A. K. and A. Rinaldo (2020): "High-dimensional CLT for sums of non-degenerate random vectors: $n^{-1/2}$-rate," *arXiv preprint arXiv:2009.13673*.

Lai, K. A., A. B. Rao, and S. Vempala (2016): "Agnostic estimation of mean and covariance," in *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, IEEE, 665–674.

Liu, M. and M. E. Lopes (2024): "Robust max statistics for high-dimensional inference," *arXiv preprint arXiv:2409.16683*.

Lopes, M. (2022): "Central limit theorem and bootstrap approximation in high dimensions: Near 1/n rates via implicit smoothing," *Annals of Statistics*, 50, 2492–2513.

Lopes, M., Z. Lin, and H.-G. Müller (2020): "Bootstrapping max statistics in high dimensions: Near-parametric rates under weak variance decay and application to functional and multinomial data," *Annals of Statistics*, 48, 1214–1229.

Lugosi, G. and S. Mendelson (2021): "Robust multivariate mean estimation: The optimality of trimmed mean," *Annals of Statistics*, 49, 393–410.

Minasyan, A. and N. Zhivotovskiy (2023): "Statistically optimal robust mean and covariance estimation for anisotropic gaussians," *arXiv preprint arXiv:2301.09024*.

MINSKER, S. (2023): "Efficient median of means estimator," in *The Thirty Sixth Annual Conference on Learning Theory*, PMLR, 5925–5933.

MINSKER, S. AND M. NDAOUD (2021): "Robust and efficient mean estimation: an approach based on the properties of self-normalized sums," *Electronic Journal of Statistics*, 15, 6036–6070.

OLIVEIRA, R., P. ORENSTEIN, AND Z. RICO (2025): "Finite-sample properties of the trimmed mean," *arXiv preprint arXiv:2501.03694*.

RESENDE, L. (2024): "Robust high-dimensional Gaussian and bootstrap approximations for trimmed sample means," *arXiv preprint arXiv:2410.22085*.

ROCKAFELLAR, R. T. (1997): *Convex Analysis*, Princeton University Press.

ZHANG, D. AND W. B. WU (2017): "Gaussian approximation for high dimensional time series," *Annals of Statistics*, 45, 1895–1919.

A A useful decomposition

We begin by outlining a decomposition of $S_{n,W}$ used in the proof of Theorem 2.1. For $p \in (0,1)$ and a random variable X, denote by $Q_p(X)$ the p-quantile of the distribution of X, that is

$$Q_p(Z) = \inf\left\{z \in \mathbb{R} : \mathbb{P}(X \le z) \ge p\right\}.$$

Suppressing the dependence on $c \in (1, \sqrt{1.5})$, which is fixed throughout, for $\varepsilon_n \in (0, 0.5)$ let

$$\underline{\alpha}_j := \underline{\alpha}_{c,j} := Q_{\varepsilon_n - c^{-1}\varepsilon_n}(X_{1,j}), \qquad \text{and} \qquad \overline{\alpha}_j := \overline{\alpha}_{c,j} := Q_{\varepsilon_n + c^{-1}\varepsilon_n}(X_{1,j}) \tag{A.1}$$

as well as

$$\underline{\beta}_j := \underline{\beta}_{c,j} := Q_{1-\varepsilon_n - c^{-1}\varepsilon_n}(X_{1,j}), \qquad \text{and} \qquad \overline{\beta}_j := \overline{\beta}_{c,j} := Q_{1-\varepsilon_n + c^{-1}\varepsilon_n}(X_{1,j}). \tag{A.2}$$

By definition $\hat{\alpha}_j = \tilde{X}^*_{\lceil \varepsilon_n n\rceil, j}$ and $\hat{\beta}_j = \tilde{X}^*_{\lceil (1-\varepsilon_n) n\rceil, j}$. Lemma G.2 and the union bound, together with obvious monotonicity properties of $(a,b) \mapsto \phi_{a,b}$, show that with probability at least $1 - \frac{4}{n}$ we have simultaneously for $j = 1, \dots, d$ (the inequalities $\phi_{\underline{\alpha}_j, \underline{\beta}_j} \le \phi_{\hat{\alpha}_j, \hat{\beta}_j} \le \phi_{\overline{\alpha}_j, \overline{\beta}_j}$ and hence)

$$\frac{1}{n}\sum_{i=1}^{n}\left[\phi_{\underline{\alpha}_j, \underline{\beta}_j}(\tilde{X}_{i,j}) - \mu_j\right] \le \frac{1}{n}\sum_{i=1}^{n}\left[\phi_{\hat{\alpha}_j, \hat{\beta}_j}(\tilde{X}_{i,j}) - \mu_j\right] \le \frac{1}{n}\sum_{i=1}^{n}\left[\phi_{\overline{\alpha}_j, \overline{\beta}_j}(\tilde{X}_{i,j}) - \mu_j\right]. \tag{A.3}$$

The far right-hand of side of (A.3) can be decomposed as

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left[\phi_{\overline{\alpha}_j, \overline{\beta}_j}(\tilde{X}_{i,j}) - \mu_j\right] &= \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left[\phi_{\overline{\alpha}_j, \overline{\beta}_j}(\tilde{X}_{i,j}) - \phi_{\overline{\alpha}_j, \overline{\beta}_j}(X_{i,j})\right]}_{\overline{I}_{n,j,1}} \\
&\quad + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left[\phi_{\overline{\alpha}_j, \overline{\beta}_j}(X_{i,j}) - \mathbb{E}\phi_{\overline{\alpha}_j, \overline{\beta}_j}(X_{i,j})\right]}_{\overline{I}_{n,j,2}} \\
&\quad + \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left[\mathbb{E}\phi_{\overline{\alpha}_j, \overline{\beta}_j}(X_{i,j}) - \mu_j\right]}_{\overline{I}_{n,j,3}}.
\end{aligned} \tag{A.4}$$

Similarly, the left-hand side of (A.3) can be decomposed as

$$\frac{1}{n}\sum_{i=1}^{n}\left[\phi_{\underline{\alpha}_j,\underline{\beta}_j}(\tilde{X}_{i,j})-\mu_j\right]=\underline{I}_{n,j,1}+\underline{I}_{n,j,2}+\underline{I}_{n,j,3}, \tag{A.5}$$

with $\underline{I}_{n,j,k}$, $j=1,\ldots,d,\ k=1,2,3$, defined analogously to the $\overline{I}_{n,j,k}$. Define

$$\overline{I}_{n,k}=\max_{j=1,\ldots,d}\left|\overline{I}_{n,j,k}\right| \qquad \text{and} \qquad \underline{I}_{n,k}=\max_{j=1,\ldots,d}\left|\underline{I}_{n,j,k}\right|, \qquad k=1,3, \tag{A.6}$$

as well as

$$\overline{I}_{n,2}=\left(\overline{I}_{n,1,2},\ldots,\overline{I}_{n,d,2}\right)' \qquad \text{and} \qquad \underline{I}_{n,2}=\left(\underline{I}_{n,1,2},\ldots,\underline{I}_{n,d,2}\right)'.$$

Throughout, $\overline{\mathbb{R}}=\mathbb{R}\cup\{-\infty,\infty\}$. Thus, writing $Y_{n,j}=n^{-1/2}\sum_{i=1}^{n}\left[\phi_{\hat{\alpha}_j,\hat{\beta}_j}(\tilde{X}_{i,j})-\mu_j\right]$ and $Y_n=\left(Y_{n,1},\ldots,Y_{n,d}\right)'$, one has with probability at least $1-\frac{4}{n}$ that

$$\sqrt{n}(\underline{I}_{n,2}-\underline{I}_{n,1}-\underline{I}_{n,3})\leq Y_n\leq\sum_{k=1}^{3}\sqrt{n}\overline{I}_{n,k}, \tag{A.7}$$

where *here and in the sequel* (i) all inequalities between vectors are understood elementwise, and (ii) with some abuse of notation we define, for a vector $x\in\overline{\mathbb{R}}^m$, say, and a real number a, the sum $x+a$ coordinatewise as $(x_1+a,\ldots,x_m+a)'$, with the usual convention that $\infty+a=\infty$ and $-\infty+a=-\infty$. In the next section we study the left- and right-hand sides of the previous display.

B Preparatory lemmas

Let $\overline{\Sigma}_{\varepsilon_n}$ and $\underline{\Sigma}_{\varepsilon_n}$ be the matrices with (j,k)th entry

$$\overline{\Sigma}_{\varepsilon_n,j,k}=\mathbb{E}\left[\left(\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{1,j})-\mathbb{E}\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{1,j})\right)\left(\phi_{\overline{\alpha}_k,\overline{\beta}_k}(X_{1,k})-\mathbb{E}\phi_{\overline{\alpha}_k,\overline{\beta}_k}(X_{1,k})\right)\right]$$

and

$$\underline{\Sigma}_{\varepsilon_n,j,k}=\mathbb{E}\left[\left(\phi_{\underline{\alpha}_j,\underline{\beta}_j}(X_{1,j})-\mathbb{E}\phi_{\underline{\alpha}_j,\underline{\beta}_j}(X_{1,j})\right)\left(\phi_{\underline{\alpha}_k,\underline{\beta}_k}(X_{1,k})-\mathbb{E}\phi_{\underline{\alpha}_k,\underline{\beta}_k}(X_{1,k})\right)\right],$$

respectively. The following lemma bounds the distance of these covariance matrices to the covariance matrix Σ of the vector X_1.

Lemma B.1. *Fix $c \in (1, \sqrt{1.5})$, and let Assumption 2.1 be satisfied with $m > 2$. If $\varepsilon_n \in (0, 0.5)$, with ε_n as in (5), it holds that*

$$\max_{j,k \in \{1,\dots,d\}} \left(\left| \overline{\Sigma}_{\varepsilon_n,j,k} - \Sigma_{j,k} \right| \vee \left| \underline{\Sigma}_{\varepsilon_n,j,k} - \Sigma_{j,k} \right| \right) \leq 32\sigma_m^2 \left(\frac{3c-1}{c-1} \right) \left(\frac{c+1}{c} \right) \varepsilon_n^{1-\frac{2}{m}}. \tag{B.1}$$

Proof. We only establish (B.1) for $\max_{j,k \in \{1,\dots,d\}} \left| \overline{\Sigma}_{\varepsilon_n,j,k} - \Sigma_{j,k} \right|$ as the proof is identical for $\max_{j,k \in \{1,\dots,d\}} \left| \underline{\Sigma}_{\varepsilon_n,j,k} - \Sigma_{j,k} \right|$.

Fix $1 \leq j \leq d$ and note that

$$\begin{aligned}
&\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{1,j}) - \mathbb{E}\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{1,j}) \\
&= X_{1,j} + (\overline{\alpha}_j - X_{1,j})\mathbb{1}\left(X_{1,j} < \overline{\alpha}_j\right) + (\overline{\beta}_j - X_{1,j})\mathbb{1}\left(X_{1,j} > \overline{\beta}_j\right) \\
&- \mathbb{E}\left(X_{1,j} + (\overline{\alpha}_j - X_{1,j})\mathbb{1}\left(X_{1,j} < \overline{\alpha}_j\right) + (\overline{\beta}_j - X_{1,j})\mathbb{1}\left(X_{1,j} > \overline{\beta}_j\right)\right) \\
&= \left[X_{1,j} - \mathbb{E}X_{1,j}\right] + \left[(\overline{\alpha}_j - X_{1,j})\mathbb{1}\left(X_{1,j} < \overline{\alpha}_j\right) - \mathbb{E}(\overline{\alpha}_j - X_{1,j})\mathbb{1}\left(X_{1,j} < \overline{\alpha}_j\right)\right] \\
&+ \left[(\overline{\beta}_j - X_{1,j})\mathbb{1}\left(X_{1,j} > \overline{\beta}_j\right) - \mathbb{E}(\overline{\beta}_j - X_{1,j})\mathbb{1}\left(X_{1,j} > \overline{\beta}_j\right)\right] \\
&= T_{1,j} + T_{2,j} + T_{3,j}.
\end{aligned}$$

Thus, for $j, k \in \{1, \dots, d\}$, it follows by the Cauchy-Schwarz inequality that

$$\begin{aligned}
\left| \overline{\Sigma}_{\varepsilon_n,j,k} - \Sigma_{j,k} \right| &= \left| \mathbb{E}(T_{1,j} + T_{2,j} + T_{3,j})(T_{1,k} + T_{2,k} + T_{3,k}) - \mathbb{E}(T_{1,j}T_{1,k}) \right| \\
&\leq |\mathbb{E}T_{1,j}T_{2,k}| + |\mathbb{E}T_{1,j}T_{3,k}| + |\mathbb{E}T_{2,j}T_{1,k}| + |\mathbb{E}T_{2,j}T_{2,k}| \\
&+ |\mathbb{E}T_{2,j}T_{3,k}| + |\mathbb{E}T_{3,j}T_{1,k}| + |\mathbb{E}T_{3,j}T_{2,k}| + |\mathbb{E}T_{3,j}T_{3,k}| \\
&\leq |\mathbb{E}T_{1,j}T_{2,k}| + |\mathbb{E}T_{1,j}T_{3,k}| + |\mathbb{E}T_{2,j}T_{1,k}| + |\mathbb{E}T_{3,j}T_{1,k}| \\
&+ \left(\mathbb{E}T_{2,j}^2 \mathbb{E}T_{2,k}^2\right)^{1/2} + \left(\mathbb{E}T_{2,j}^2 \mathbb{E}T_{3,k}^2\right)^{1/2} + \left(\mathbb{E}T_{3,j}^2 \mathbb{E}T_{2,k}^2|\right)^{1/2} + \left(\mathbb{E}T_{3,j}^2 \mathbb{E}T_{3,k}^2\right)^{1/2}.
\end{aligned} \tag{B.2}$$

We proceed by bounding the far right-hand side of (B.2). To this end, note first that for all $j = 1, \dots, d$ it holds that

$$\begin{aligned}
\mathbb{E}T_{2,j}^2 &= \mathbb{E}\left[(\overline{\alpha}_j - X_{1,j})\mathbb{1}\left(X_{1,j} < \overline{\alpha}_j\right) - \mathbb{E}(\overline{\alpha}_j - X_{1,j})\mathbb{1}\left(X_{1,j} < \overline{\alpha}_j\right)\right]^2 \\
&\leq \mathbb{E}(\overline{\alpha}_j - X_{1,j})^2 \mathbb{1}\left(X_{1,j} < \overline{\alpha}_j\right).
\end{aligned}$$

Recall that $\overline{\alpha}_j = Q_{\varepsilon_n + c^{-1}\varepsilon_n}(X_{1,j})$ such that $\mathbb{P}(X_{1,j} < \overline{\alpha}_j) \le \varepsilon_n + c^{-1}\varepsilon_n$. Thus, it follows by Lemma G.1, followed by Hölder's inequality, that

$$\begin{aligned}
\mathbb{E}T_{2,j}^2 &\le \mathbb{E}(\mu_j + \sigma_m/(1-\varepsilon_n - c^{-1}\varepsilon_n)^{1/m} - X_{1,j})^2 \mathbb{1}\left(X_{1,j} < \overline{\alpha}_j\right) \\
&\le \left(\mathbb{E}\big|\mu_j - X_{1,j} + \sigma_m/(1-\varepsilon_n - c^{-1}\varepsilon_n)^{1/m}\big|^m\right)^{2/m} \mathbb{P}(X_{1,j} < \overline{\alpha}_j)^{1-\frac{2}{m}} \\
&\le \left(2^m\sigma_m^m + 2^m\sigma_m^m/(1-\varepsilon_n - c^{-1}\varepsilon_n)\right)^{2/m} (\varepsilon_n + c^{-1}\varepsilon_n)^{1-\frac{2}{m}} \\
&\le 4\sigma_m^2\left(1 + 1/(1-\varepsilon_n - c^{-1}\varepsilon_n)^{2/m}\right)\left(\frac{c+1}{c}\right)^{1-\frac{2}{m}} \varepsilon_n^{1-\frac{2}{m}}. \qquad \text{(B.3)}
\end{aligned}$$

By the same arguments, using $\mathbb{P}(X_{1,j} > \overline{\beta}_j) = 1 - \mathbb{P}(X_{1,j} \le \overline{\beta}_j) \le \varepsilon_n - c^{-1}\varepsilon_n$,

$$\mathbb{E}T_{3,j}^2 \le 4\sigma_m^2\left(1 + 1/(1-\varepsilon_n + c^{-1}\varepsilon_n)^{2/m}\right)\left(\frac{c-1}{c}\right)^{1-\frac{2}{m}} \varepsilon_n^{1-\frac{2}{m}}. \qquad \text{(B.4)}$$

Furthermore,

$$\begin{aligned}
\mathbb{E}T_{1,j}T_{2,k} &= \mathbb{E}\left([X_{1,j} - \mu_j]\left[(\overline{\alpha}_k - X_{1,k})\mathbb{1}\left(X_{1,k} < \overline{\alpha}_k\right) - \mathbb{E}(\overline{\alpha}_k - X_{1,k})\mathbb{1}\left(X_{1,k} < \overline{\alpha}_k\right)\right]\right) \\
&= \mathbb{E}\left([X_{1,j} - \mu_j](\overline{\alpha}_k - X_{1,k})\mathbb{1}\left(X_{1,k} < \overline{\alpha}_k\right)\right).
\end{aligned}$$

Recalling that $\overline{\alpha}_j = Q_{\varepsilon_n + c^{-1}\varepsilon_n}(X_{1,j})$, it follows by Lemma G.1, followed by Hölder's inequality, that

$$\begin{aligned}
&\left|\mathbb{E}T_{1,j}T_{2,k}\right| \\
&\le \mathbb{E}\left(|X_{1,j} - \mu_j||\mu_k + \sigma_m/(1-\varepsilon_n - c^{-1}\varepsilon_n)^{1/m} - X_{1,k}|\mathbb{1}\left(X_{1,k} < \overline{\alpha}_k\right)\right) \\
&\le \mathbb{E}\left(|X_{1,j} - \mu_j||X_{1,k} - \mu_k| + |X_{1,j} - \mu_j|\sigma_m/(1-\varepsilon_n - c^{-1}\varepsilon_n)^{1/m}\right)\mathbb{1}\left(X_{1,k} < \overline{\alpha}_k\right) \\
&\le \left(\mathbb{E}\left(|X_{1,j} - \mu_j||X_{1,k} - \mu_k| + |X_{1,j} - \mu_j|\sigma_m/(1-\varepsilon_n - c^{-1}\varepsilon_n)^{1/m}\right)^{\frac{m}{2}}\right)^{\frac{2}{m}} (\varepsilon_n + c^{-1}\varepsilon_n)^{1-\frac{2}{m}} \\
&\le \left(2^{m/2}(\sigma_m^m + \sigma_{m/2}^{m/2}\sigma_m^{m/2}/(1-\varepsilon_n - c^{-1}\varepsilon_n)^{1/2}\right)^{2/m}\left(\frac{c+1}{c}\right)^{1-\frac{2}{m}} \varepsilon_n^{1-\frac{2}{m}} \\
&\le 2\sigma_m^2\left(1 + 1/(1-\varepsilon_n - c^{-1}\varepsilon_n)^{1/m}\right)\left(\frac{c+1}{c}\right)^{1-\frac{2}{m}} \varepsilon_n^{1-\frac{2}{m}}. \qquad \text{(B.5)}
\end{aligned}$$

By the same arguments,

$$\left|\mathbb{E}T_{1,j}T_{3,k}\right| \le 2\sigma_m^2\left(1 + 1/(1-\varepsilon_n + c^{-1}\varepsilon_n)^{1/m}\right)\left(\frac{c-1}{c}\right)^{1-\frac{2}{m}} \varepsilon_n^{1-\frac{2}{m}}. \qquad \text{(B.6)}$$

Therefore, observing that none of the upper bounds in (B.3)–(B.6) depend on j and k and that the one in (B.3) is the largest one, we bound each of the eight terms in (B.2) by this. Hence,

$$\max_{j,k\in\{1,\ldots,d\}}\left|\overline{\Sigma}_{\varepsilon_n,j,k} - \Sigma_{j,k}\right| \leq 32\sigma_m^2\left(1 + 1/(1-\varepsilon_n - c^{-1}\varepsilon_n)^{2/m}\right)\left(\frac{c+1}{c}\right)^{1-\frac{2}{m}}\varepsilon_n^{1-\frac{2}{m}}$$
$$\leq 32\sigma_m^2\left(\frac{3c-1}{c-1}\right)\left(\frac{c+1}{c}\right)\varepsilon_n^{1-\frac{2}{m}},$$

the last estimate following from $\varepsilon_n \in (0, 0.5)$, $c > 1$, and $m > 2$. □

The following lemma applies a high-dimensional Gaussian approximation to $\underline{I}_{n,2}$ and $\overline{I}_{n,2}$, which are both sums of winsorized, and hence bounded, random variables.

Lemma B.2. *Fix $c \in (1, \sqrt{1.5})$, and let Assumption 2.1 be satisfied with $m > 2$. If $\varepsilon_n \in (0, 0.5)$, with ε_n as in (5), and there exists a strictly positive constants $\mathfrak{b}_1$ such that for all $j = 1, \ldots, d$*

$$\mathbb{E}\left[\phi_{\underline{\alpha}_j,\underline{\beta}_j}(X_{1,j}) - \mathbb{E}\phi_{\underline{\alpha}_j,\underline{\beta}_j}(X_{1,j})\right]^2 > \mathfrak{b}_1^2 \qquad \text{and} \qquad \mathbb{E}\left[\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{1,j}) - \mathbb{E}\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{1,j})\right]^2 > \mathfrak{b}_1^2,$$

then, for $\overline{Z} \sim \mathsf{N}_d(0, \overline{\Sigma}_{\varepsilon_n})$ and $\underline{Z} \sim \mathsf{N}_d(0, \underline{\Sigma}_{\varepsilon_n})$,

$$\sup_{H\in\mathcal{H}}\left|\mathbb{P}\left(\frac{1}{n^{1/2}}\sum_{i=1}^n\left[\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{i,j}) - \mathbb{E}\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{i,j})\right] \in H\right) - \mathbb{P}\left(\overline{Z} \in H\right)\right|$$

and

$$\sup_{H\in\mathcal{H}}\left|\mathbb{P}\left(\frac{1}{n^{1/2}}\sum_{i=1}^n\left[\phi_{\underline{\alpha}_j,\underline{\beta}_j}(X_{i,j}) - \mathbb{E}\phi_{\underline{\alpha}_j,\underline{\beta}_j}(X_{i,j})\right] \in H\right) - \mathbb{P}\left(\underline{Z} \in H\right)\right|$$

are bounded from above by

$$C\left(\frac{12}{[(c-1)/c]^{1/m}}\right)^{\frac{1}{2}}\sigma_m^{1/2}\left(\frac{\log^{5-\frac{2}{m}}(dn)}{n^{1-\frac{2}{m}}}\right)^{\frac{1}{4}},$$

where C is a constant depending only on $\mathfrak{b}_1$ and b_2.

Proof. We establish the first bound by verifying the conditions of Theorem 2.1 in Chernozhuokov et al. (2022), cf. in particular the display following that theorem. The second bound is proven analogously. First, note that for $j = 1, \dots, d$

$$\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{i,j}) - \mathbb{E}\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{i,j}) \leq \overline{\beta}_j - \overline{\alpha}_j = Q_{1-\varepsilon_n+c^{-1}\varepsilon_n}(X_{1,j}) - Q_{\varepsilon_n+c^{-1}\varepsilon_n}(X_{1,j}),$$

Furthermore, by Lemma G.1 and $\varepsilon_n \geq \lambda_{2,c}(\delta, n)\frac{\log(dn)}{n} \geq \frac{\log(dn)}{3n}$, the far right-hand side of the previous display is upper bounded by

$$\frac{\sigma_m}{(\varepsilon_n - c^{-1}\varepsilon_n)^{1/m}} + \frac{\sigma_m}{(\varepsilon_n + c^{-1}\varepsilon_n)^{1/m}} \leq \frac{2\sigma_m}{[(c-1)/c]^{1/m}\varepsilon_n^{1/m}} \leq \frac{6\sigma_m}{[(c-1)/c]^{1/m}}\left(\frac{n}{\log(dn)}\right)^{1/m}.$$

Similarly,

$$\begin{aligned}\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{i,j}) - \mathbb{E}\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{i,j}) \geq \overline{\alpha}_j - \overline{\beta}_j &= Q_{\varepsilon_n+c^{-1}\varepsilon_n}(X_{1,j}) - Q_{1-\varepsilon_n+c^{-1}\varepsilon_n}(X_{1,j}) \\ &\geq -\frac{6\sigma_m}{[(c-1)/c]^{1/m}}\left(\frac{n}{\log(dn)}\right)^{1/m},\end{aligned}$$

and we define

$$B_n = \frac{12\sigma_m}{[(c-1)/c]^{1/m}}\left(\frac{n}{\log(dn)}\right)^{1/m}.$$

By construction, $\left|\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{i,j}) - \mathbb{E}\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{i,j})\right| \leq \frac{1}{2}B_n$ for all $i = 1, \dots, n$ and $j = 1, \dots, d$. Furthermore, by assumption, $\mathbb{E}\left[\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{1,j}) - \mathbb{E}\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{1,j})\right]^2 > \mathfrak{b}_1^2$, as well as

$$\begin{aligned}\mathbb{E}\left[\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{1,j}) - \mathbb{E}\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{1,j})\right]^4 &\leq B_n^2\mathbb{E}\left[\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{1,j}) - \mathbb{E}\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{1,j})\right]^2 \\ &\leq B_n^2\mathbb{E}\left[X_{1,j} - \mathbb{E}(X_{1,j})\right]^2 \leq B_n^2\sigma_2^2 \leq B_n^2 b_2^2;\end{aligned}$$

a proof of the second inequality can be found in, e.g., Corollary 3 in Chow and Studden (1969). Therefore, by Theorem 2.1 and the display following it in Chernozhuokov et al. (2022), there exists a constant C depending only on $\mathfrak{b}_1$ and b_2 such that

$$\begin{aligned}&\sup_{H\in\mathcal{H}}\left|\mathbb{P}\left(\frac{1}{n^{1/2}}\sum_{i=1}^{n}\left[\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{i,j}) - \mathbb{E}\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{i,j})\right] \in H\right) - \mathbb{P}\left(\overline{Z} \in H\right)\right| \\ \leq &C\left(\frac{B_n^2\log^5(dn)}{n}\right)^{\frac{1}{4}} = C\left(\frac{12}{[(c-1)/c]^{1/m}}\right)^{\frac{1}{2}}\sigma_m^{1/2}\left(\frac{\log^{5-\frac{2}{m}}(dn)}{n^{1-\frac{2}{m}}}\right)^{\frac{1}{4}}.\end{aligned}$$

□

Lemma B.3. *Fix $c \in (1, \sqrt{1.5})$, and let Assumption 2.1 be satisfied with $m > 2$. If $\varepsilon_n \in (0, 0.5)$, with ε_n as in (5), then for $\overline{I}_{n,3}$ and $\underline{I}_{n,3}$ as defined in (A.6), it holds that*

$$\underline{I}_{n,3} \vee \overline{I}_{n,3} \leq \sigma_m f(c,m) \left(\overline{\eta}_n^{1-\frac{1}{m}} + \left[\frac{\log(dn)}{n} \right]^{1-\frac{1}{m}} \right),$$

for a function $f : (1, \sqrt{1.5}) \times [1, \infty) \to [0, \infty)$ of c and m only.

Proof. Lemma G.4 and subadditivity of $z \mapsto z^{1-\frac{1}{m}}$ along with

$$\varepsilon_n \leq \frac{c}{1 - \sqrt{2(c^2-1)}} \cdot \overline{\eta}_n + \left[\frac{c}{3[1 - \sqrt{2(c^2-1)}]} \vee c \left(\sqrt{2 \frac{c+1}{c-1}} + \frac{1}{3} \right) \right] \frac{\log(dn)}{n}$$

implies that for a non-negative function f of c and m only

$$\underline{I}_{n,3} \vee \overline{I}_{n,3} \leq \sigma_m f(c,m) \left(\overline{\eta}_n^{1-\frac{1}{m}} + \left[\frac{\log(dn)}{n} \right]^{1-\frac{1}{m}} \right).$$

□

The following lemma collects some simple limits calculations for later reference.

Lemma B.4. *Let $m > 2$ and assume that $\sqrt{n \log(d)} \overline{\eta}_n^{1-\frac{1}{m}} \to 0$ as well as $\log(d)/n^{\frac{m-2}{5m-2}} \to 0$. Then*

$$\log^2(d) \left(\overline{\eta}_n^{1-\frac{2}{m}} + \left[\frac{\log(dn)}{n} \right]^{1-\frac{2}{m}} \right) \to 0, \tag{B.7}$$

$$\left[\frac{\log^{5-\frac{2}{m}}(dn)}{n^{1-\frac{2}{m}}} \right]^{\frac{1}{4}} + \left(\overline{\eta}_n^{1-\frac{1}{m}} + \left[\frac{\log(dn)}{n} \right]^{1-\frac{1}{m}} \right) \sqrt{n \log(d)} \to 0, \tag{B.8}$$

$$\log^2(d) \left(\overline{\eta}_n^{1-\frac{2}{m}} + \left[\frac{\log(dn)}{n} \right]^{\frac{1}{2}-\frac{1}{m}} \right) \to 0, \tag{B.9}$$

and

$$\sqrt{\log(d)\log(dn)}\left(\overline{\eta}_n^{1-\frac{2}{m}} + \left[\frac{\log(dn)}{n}\right]^{\frac{1}{2}-\frac{1}{m}}\right) \to 0. \tag{B.10}$$

Proof. We begin by proving (B.7), to which end we first establish that $\log^2(d)\overline{\eta}_n^{1-\frac{2}{m}} \to 0$. Since $\log(d)/n^{\frac{m-2}{5m-2}} \to 0$, one has that $\log^2(d)\overline{\eta}_n^{1-\frac{2}{m}} = o\left(n^{\frac{2m-4}{5m-2}}\overline{\eta}_n^{1-\frac{2}{m}}\right)$ and

$$n^{\frac{2m-4}{5m-2}}\overline{\eta}_n^{1-\frac{2}{m}} \to 0 \iff n^{\frac{2m^2-4m}{(5m-2)(m-2)}}\overline{\eta}_n \to 0.$$

The latter convergence is satisfied since $\sqrt{n}\overline{\eta}^{1-\frac{1}{m}} \to 0$ by assumption, which is equivalent to $n^{\frac{m}{2(m-1)}}\overline{\eta}_n \to 0$, and $n^{\frac{2m^2-4m}{(5m-2)(m-2)}}\overline{\eta}_n \leq n^{\frac{m}{2(m-1)}}\overline{\eta}_n \to 0$ for $m > 2$.

Next,

$$\log^2(d)\left[\frac{\log(d)}{n}\right]^{1-\frac{2}{m}} = \frac{\log(d)^{3-\frac{2}{m}}}{n^{1-\frac{2}{m}}} \to 0 \iff \frac{\log(d)}{n^{\frac{m-2}{3m-2}}} \to 0,$$

the latter convergence following from $\log(d)/n^{\frac{m-2}{5m-2}} \to 0$ by assumption.

Finally,

$$\log^2(d)\left[\frac{\log(n)}{n}\right]^{1-\frac{2}{m}} \to 0$$

by a standard subsequence argument, considering separately the cases of subsequences along which i) $n \leq d$ for which the convergence follows from the penultimate display and ii) $d < n$ for which the convergence follows from $m > 2$. This establishes (B.7).

To prove (B.8), note that

$$\frac{\log^{5-\frac{2}{m}}(dn)}{n^{1-\frac{2}{m}}} \to 0 \iff \frac{\log(d)}{n^{\frac{m-2}{5m-2}}} \to 0,$$

the latter convergence being true by assumption. Furthermore, $\sqrt{n\log(d)}\overline{\eta}_n^{1-\frac{1}{m}} \to 0$ by assumption and it remains to prove that

$$\sqrt{n\log(d)}\left(\frac{\log(dn)}{n}\right)^{1-\frac{1}{m}} \to 0.$$

To this end, note that

$$\sqrt{n\log(d)}\left(\frac{\log(d)}{n}\right)^{1-\frac{1}{m}} = \frac{\left[\log(d)\right]^{\frac{3}{2}-\frac{1}{m}}}{n^{\frac{1}{2}-\frac{1}{m}}} \to 0 \iff \frac{\log(d)}{n^{\frac{m-2}{3m-2}}} \to 0,$$

where the latter convergence was already verified above. The convergence in the penultimate display now follows by considering separately subsequences along which $n \leq d$ and $d < n$, respectively (as in the concluding argument of the proof of (B.7)).

Next, (B.9) is true since we have already established $\log^2(d)\overline{\eta}_n^{1-\frac{2}{m}} \to 0$ and

$$\log^2(d)\left(\frac{\log(d)}{n}\right)^{\frac{1}{2}-\frac{1}{m}} = \frac{\left[\log(d)\right]^{\frac{5}{2}-\frac{1}{m}}}{n^{\frac{1}{2}-\frac{1}{m}}} \to 0 \iff \frac{\log(d)}{n^{\frac{m-2}{5m-2}}} \to 0,$$

the latter convergence being true by assumption. Conclude, once more, by a subsequence argument.

Finally, we establish (B.10) by showing that any subsequence possesses a further subsequence along which (B.10) is true. Thus, fix a subsequence. In case there exists a subsequence thereof along which $n \leq d$, then it suffices to show that along this subsequence

$$\log(d)\left(\overline{\eta}_n^{1-\frac{2}{m}} + \left[\frac{\log(d)}{n}\right]^{\frac{1}{2}-\frac{1}{m}}\right) \to 0,$$

which is implied by (B.9).

In the remaining case (where there does not exists a further subsequence along which $n \leq d$), we have $d < n$ for n sufficiently large, so that it suffices to show that

$$\log(n)\left(\overline{\eta}_n^{1-\frac{2}{m}} + \left[\frac{\log(n)}{n}\right]^{\frac{1}{2}-\frac{1}{m}}\right) \to 0.$$

Since $m > 2$ this, in turn, is true if

$$\log(n)\overline{\eta}_n^{1-\frac{2}{m}} = \sqrt{n}\overline{\eta}_n^{1-\frac{1}{m}} \cdot \frac{\log(n)}{\sqrt{n}}\overline{\eta}_n^{-\frac{1}{m}} \to 0,$$

which, since $\sqrt{n}\overline{\eta}_n^{1-\frac{1}{m}} \to 0$ by assumption, is true in case $\frac{\log(n)}{\sqrt{n}}\overline{\eta}_n^{-\frac{1}{m}}$ is bounded. In case $\frac{\log(n)}{\sqrt{n}}\overline{\eta}_n^{-\frac{1}{m}}$ is unbounded, $\frac{\log(n)}{\sqrt{n}}\overline{\eta}_n^{-\frac{1}{m}} \geq 1$ along a further subsequence. The latter implies $\overline{\eta}_n \leq \frac{\log^m(n)}{n^{\frac{m}{2}}}$, such that

$$\log(n)\overline{\eta}_n^{1-\frac{2}{m}} \leq \frac{[\log(n)]^{m-1}}{n^{\frac{m-2}{2}}} \to 0,$$

because $m > 2$. □

The following lemma shows that to establish Theorem 2.1, it suffices to prove it for the class of one-sided rectangles $H = \times_{j=1}^{d}[-\infty, t_j]$ with $t_j \in \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ for $j = 1, \ldots, d$.

Lemma B.5. *Suppose Theorem 2.1 holds for $\rho_{n,W}$ throughout replaced by*

$$\rho_{n,W}^{os} := \sup_{t \in \overline{\mathbb{R}}^d} \left| \mathbb{P}\left(S_{n,W} \leq t\right) - \mathbb{P}\left(Z \leq t\right) \right|. \tag{B.11}$$

Then, Theorem 2.1 holds (with a different constant C).

Proof. Fix a and b in $\overline{\mathbb{R}}^d$ such that $-\infty \leq a_j \leq b_j \leq \infty$ and let $H = \times_{j=1}^{d}[a_j, b_j] \in \mathcal{H}$. Clearly,

$$\mathbb{P}\left(S_{n,W} \in H\right) = \mathbb{P}\left((S_{n,W}', -S_{n,W}')' \leq (b', -a')'\right). \tag{B.12}$$

Furthermore, by (3)

$$-S_{n,W,j} = n^{-1/2} \sum_{i=1}^{n} \left(\phi_{-\hat{\beta}_j, -\hat{\alpha}_j}(-\tilde{X}_{i,j}) - (-\mu_j)\right), \qquad j = 1, \ldots, d.$$

Thus, the vectors $(X_i', -X_i')'$ and $(\tilde{X}_i', -\tilde{X}_i')'$ in $\mathbb{R}^{2d}$ satisfy Assumption 2.1 with d there replaced by $2d$ (but with the same m, b_1, b_2 and $\overline{\eta}_n$). Note that the covariance matrix of $(X_1', -X_1')'$ is

$$\Xi = \begin{pmatrix} \Sigma & -\Sigma \\ -\Sigma & \Sigma \end{pmatrix},$$

and let $Z_2 \sim \mathsf{N}_{2d}(0, \Xi) =: \nu'$. Hence, the version of Theorem 2.1 that replaces $\rho_{n,W}$ by $\rho_{n,W}^{os}$ (and which is assumed to hold true) applies to $(S_{n,W}', -S_{n,W}')'$ and yields, for C^*

a constant depending only on b_1, b_2, c and m, the upper bound

$$C^* \left(\left[\frac{\log^{5-\frac{2}{m}}(dn)}{n^{1-\frac{2}{m}}} \right]^{\frac{1}{4}} + \left[\overline{\eta}_n^{1-\frac{1}{m}} + \left[\frac{\log(dn)}{n} \right]^{1-\frac{1}{m}} \right] \sqrt{n \log(d)} \right)$$
$$+ C^* \left(\log^2(d) \left[\overline{\eta}_n^{1-\frac{2}{m}} + \left[\frac{\log(dn)}{n} \right]^{1-\frac{2}{m}} \right] \right)^{1/2}$$

on

$$\left| \mathbb{P}\left((S_{n,W}, -S_{n,W}) \leq (b', -a')' \right) - \mathbb{P}\left(Z_2 \leq (b', -a')' \right) \right|, \tag{B.13}$$

Furthermore, for $\nu = \mathsf{N}_d(0, \Sigma)$, ν' is the image measure of ν under the mapping $\mathbb{R}^d \ni z \mapsto (z, -z) \in \mathbb{R}^{2d}$, that is $\nu' = \nu \circ \left(z \mapsto (z, -z) \right)^{-1}$. Thus,

$$\begin{aligned} \mathbb{P}\left(Z_2 \leq (b', -a')' \right) &= \nu' \left([-\infty, b] \times [-\infty, -a] \right) \\ &= \nu \left(z \in \mathbb{R}^d : (z, -z) \in [-\infty, b] \times [-\infty, -a] \right) \\ &= \mathbb{P}\left((Z, -Z) \leq (b, -a) \right) \\ &= \mathbb{P}\left(Z \in H \right). \end{aligned} \tag{B.14}$$

Combining (B.12) and (B.14) with the upper bound in (B.13) obtained above (which does not depend on a or b) delivers the claim. $\square$

C Proof of Theorem 2.1

By Lemma B.5 it suffices to prove (6) for hyperractangles H on the form $H = \times_{j=1}^d [-\infty, t_j]$ with $t_j \in \overline{\mathbb{R}}$. To this end, by Lemma G.3,

$$\underline{I}_{n,1} \vee \overline{I}_{n,1} \leq 2 \left[\frac{1 - \sqrt{2(c^2-1)}}{(c-1)} \right]^{\frac{1}{m}} \sigma_m \overline{\eta}_n^{1-\frac{1}{m}} := I_{n,1} \tag{C.1}$$

with $I_{n,1}$ being non-random. Thus, from (A.7) it holds with probability at least $1 - \frac{4}{n}$ that

$$\underline{Y}_n := \sqrt{n} \underline{I}_{n,2} - \sqrt{n}(I_{n,1} + \underline{I}_{n,3}) \leq Y_n \leq \sqrt{n} \overline{I}_{n,2} + \sqrt{n}(I_{n,1} + \overline{I}_{n,3}) =: \overline{Y}_n.$$

Next, it holds for all $t = (t_1, \ldots, t_d)$ with $t_j \in \overline{\mathbb{R}}$ that

$$\mathbb{P}\left(Y_n \le t\right) \ge \mathbb{P}\left(\overline{Y}_n \le t, Y_n \le \overline{Y}_n\right) \ge \mathbb{P}\left(\overline{Y}_n \le t\right) + \mathbb{P}\left(Y_n \le \overline{Y}_n\right) - 1 \ge \mathbb{P}\left(\overline{Y}_n \le t\right) - \frac{4}{n}.$$

Similarly,

$$\mathbb{P}\left(Y_n \le t\right) \le \mathbb{P}\left(\underline{Y}_n \le t\right) + \mathbb{P}\left(\underline{Y}_n > Y_n\right) \le \mathbb{P}\left(\underline{Y}_n \le t\right) + \frac{4}{n}.$$

Thus,

$$\begin{aligned}\sup_{t\in\overline{\mathbb{R}}^d}\left|\mathbb{P}\left(Y_n \le t\right) - \mathbb{P}\left(Z \le t\right)\right| &\le \sup_{t\in\overline{\mathbb{R}}^d}\left|\mathbb{P}\left(\underline{Y}_n \le t\right) - \mathbb{P}\left(Z \le t\right)\right| \\ &\quad + \sup_{t\in\overline{\mathbb{R}}^d}\left|\mathbb{P}\left(\overline{Y}_n \le t\right) - \mathbb{P}\left(Z \le t\right)\right| + \frac{4}{n}. \qquad \text{(C.2)}\end{aligned}$$

We proceed by bounding the second summand on the right-hand side of the previous display (the argument for the first summand is analogous and hence skipped). To this end, consider first the case of

$$32b_2^2\left(\frac{3c-1}{c-1}\right)\left(\frac{c+1}{c}\right)\varepsilon_n^{1-\frac{2}{m}} \le \frac{b_1^2}{2}, \qquad \text{(C.3)}$$

which by Lemma B.1 implies that $\min_{j=1,\ldots,d}\overline{\Sigma}_{\varepsilon_n,j,j} \ge b_1^2/2$. Next, by the definition of $\overline{Y}_n$

$$\mathbb{P}\left(\overline{Y}_n \le t\right) = \mathbb{P}\left(\sqrt{n}\overline{I}_{n,2} \le t - \sqrt{n}\left(I_{n,1} + \overline{I}_{n,3}\right)\right), \qquad t \in \overline{\mathbb{R}}^d,$$

whence, noting also that $I_{n,1}$ and $\overline{I}_{n,3}$ are non-random, Lemma B.2 applied with $\mathfrak{b}_1^2 = b_1^2/2$ implies the existence of a constant C_1 depending only on b_1, b_2, c and m such that

$$\sup_{t\in\overline{\mathbb{R}}^d}\left|\mathbb{P}\left(\overline{Y}_n \le t\right) - \mathbb{P}\left(\overline{Z} \le t - \sqrt{n}\left(I_{n,1} + \overline{I}_{n,3}\right)\right)\right| \le C_1\left(\frac{\log^{5-\frac{2}{m}}(dn)}{n^{1-\frac{2}{m}}}\right)^{\frac{1}{4}},$$

where $\overline{Z} \sim \mathsf{N}_d(0, \overline{\Sigma}_{\varepsilon_n})$. Furthermore, by the Gaussian anti-concentration inequality as stated in Theorem 1 of Chernozhukov et al. (2017b), (cf. also Lemma A.1 in Chernozhukov et al. (2017a)),

$$
\begin{aligned}
0 &\geq \mathbb{P}\left(\overline{Z} \leq t - \sqrt{n}\,(I_{n,1} + \overline{I}_{n,3})\right) - \mathbb{P}\left(\overline{Z} \leq t\right) \\
&\geq -\frac{\sqrt{n}\,(I_{n,1} + \overline{I}_{n,3})}{\sqrt{\min_{j=1,\ldots,d} \overline{\Sigma}_{\varepsilon_n,j,j}}} \left(\sqrt{2\log(d)} + 2\right) \\
&\geq -C_2 \left(\overline{\eta}_n^{1-\frac{1}{m}} + \left[\frac{\log(dn)}{n}\right]^{1-\frac{1}{m}}\right) \sqrt{n\log(d)},
\end{aligned}
$$

the final inequality following from Lemma B.3, (C.1), $\min_{j=1,\ldots,d} \overline{\Sigma}_{\varepsilon_n,j,j} \geq b_1^2/2$, and C_2 being a constant depending on b_1, b_2, c and m only. Thus, combining the previous two displays, there exists a constant C_3 depending on b_1, b_2, c and m only such that when (C.3) is satisfied one has

$$
\sup_{t\in\overline{\mathbb{R}}^d} \left|\mathbb{P}\left(\overline{Y}_n \leq t\right) - \mathbb{P}\left(\overline{Z} \leq t\right)\right| \leq C_3 \left(\left[\frac{\log^{5-\frac{2}{m}}(dn)}{n^{1-\frac{2}{m}}}\right]^{\frac{1}{4}} + \left[\overline{\eta}_n^{1-\frac{1}{m}} + \left[\frac{\log(dn)}{n}\right]^{1-\frac{1}{m}}\right] \sqrt{n\log(d)}\right).
$$

Finally, by the Gaussian-to-Gaussian comparison inequality as stated in Proposition 2.1 of Chernozhuokov et al. (2022) and Lemma B.1, there exists a constant C_4, depending on b_1, b_2 and c only (C_4 changes value in the second inequality below), such that

$$
\begin{aligned}
\sup_{t\in\overline{\mathbb{R}}^d} \left|\mathbb{P}\left(\overline{Z} \leq t\right) - \mathbb{P}\left(Z \leq t\right)\right| &\leq C_4 \left(\log^2(d)\varepsilon_n^{1-\frac{2}{m}}\right)^{1/2} \\
&\leq C_4 \left(\log^2(d) \left[\overline{\eta}_n^{1-\frac{2}{m}} + \left[\frac{\log(dn)}{n}\right]^{1-\frac{2}{m}}\right]\right)^{1/2}.
\end{aligned}
$$

The previous two displays yield that there exists a constant C depending only on b_1, b_2, c and m such that when (C.3) is satisfied it holds that

$$
\begin{aligned}
\sup_{t\in\overline{\mathbb{R}}^d} \left|\mathbb{P}\left(\overline{Y}_n \leq t\right) - \mathbb{P}\left(Z \leq t\right)\right| &\leq C \left(\log^2(d) \left[\overline{\eta}_n^{1-\frac{2}{m}} + \left[\frac{\log(dn)}{n}\right]^{1-\frac{2}{m}}\right]\right)^{1/2} \\
&+ C \left(\left[\frac{\log^{5-\frac{2}{m}}(dn)}{n^{1-\frac{2}{m}}}\right]^{\frac{1}{4}} + \left[\overline{\eta}_n^{1-\frac{1}{m}} + \left[\frac{\log(dn)}{n}\right]^{1-\frac{1}{m}}\right] \sqrt{n\log(d)}\right).
\end{aligned}
$$

This implies the claimed bound on $\rho_{n,W}$ in (6) by using that $4/n$ in (C.2) is dominated by, e.g., $\left[\frac{\log(dn)}{n}\right]^{1-\frac{1}{m}}\sqrt{n\log(d)}$ (if necessary adjust C).

If, on the other hand, (C.3) is not satisfied then

$$32b_2^2\left(\frac{3c-1}{c-1}\right)\left(\frac{c+1}{c}\right)\varepsilon_n^{1-\frac{2}{m}} > \frac{b_1^2}{2} \qquad \Longleftrightarrow \qquad \varepsilon_n > \left[\frac{b_1^2}{64b_2^2\left(\frac{3c-1}{c-1}\right)\left(\frac{c+1}{c}\right)}\right]^{\frac{m}{m-2}} =: K(b_1,b_2,c,m),$$

where $K = K(b_1,b_2,c,m)$ does not depend on n. Thus, by the definition of ε_n (cf. also Footnote 3)

$$\overline{\eta}_n > \frac{K}{2\lambda_{1,c}} \qquad \text{or} \qquad \frac{\log(dn)}{n} > \frac{K}{2f(c)},$$

where $f(c) = \frac{c}{3[1-\sqrt{2(c^2-1)}]} \vee c\left(\sqrt{2\frac{c+1}{c-1}} + \frac{1}{3}\right)$. In either case, since $\rho_{n,W} \leq 1$ the bound in (6) remains valid by adjusting the constant C, if necessary.

Finally, $\rho_{n,W} \to 0$ by (B.7) and (B.8) of Lemma B.4.

D Proofs for Section 3

In order to prove Theorem 3.1, we introduce an intermediate estimator $\hat{\Sigma}_n$ of Σ that may be of independent interest. Its properties are established by an application of the union bound (over the entries of the covariance matrix) to the one-dimensional winsorized mean estimator in Theorem 5.1 of Kock and Preinerstorfer (2025). Subsequently, the properties of $\tilde{\Sigma}_n$ are established by showing it is sufficiently close to $\hat{\Sigma}_n$.

To define $\hat{\Sigma}_n$, recall that $Y_i = \frac{1}{\sqrt{2}}(X_{2i} - X_{2i-1})$ and $\tilde{Y}_i = \frac{1}{\sqrt{2}}(\tilde{X}_{2i} - \tilde{X}_{2i-1})$ for $i = 1,\dots,N = n/2$, cf. (8). Next, write $U_{i,j,k} = Y_{i,j}Y_{i,k}$ as well as $\tilde{U}_{i,j,k} = \tilde{Y}_{i,j}\tilde{Y}_{i,k}$ for $i = 1,\dots,n/2$ and $1 \leq j,k \leq d$. For ε_n' as in (9), define $\hat{\alpha}_{j,k} = \tilde{U}^*_{\lceil \varepsilon_n' N\rceil,j,k}$ as well as $\hat{\beta}_{j,k} = \tilde{U}^*_{\lceil (1-\varepsilon_n')N\rceil,j,k}$. Finally, let $\hat{\Sigma}_n$ be the estimator with elements

$$\hat{\Sigma}_{n,j,k} = \frac{1}{N}\sum_{i=1}^{N}\phi_{\hat{\alpha}_{j,k},\hat{\beta}_{j,k}}\left(\tilde{U}_{i,j,k}\right), \qquad 1 \leq j,k \leq d. \tag{D.1}$$

Theorem D.1. *Fix $c \in (1,\infty)$, and let Assumption 2.1 be satisfied with $m > 2$. If $\varepsilon_n' \in$*

$(0, 1/2)$, *for* ε'_n *as in* (9), *then for a constant* $C = C(b_2, c, m)$

$$\mathbb{P}\left(\max_{1\le j,k\le d}\left|\hat{\Sigma}_{n,j,k} - \Sigma_{j,k}\right| > C\left[\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n}\right)^{\frac{1}{2}-\frac{1}{m}}\right]\right) \le \frac{12}{n}. \tag{D.2}$$

Remark D.1. The estimator $\hat{\Sigma}_n$, although symmetric, need not be positive semi-definite (PSD). *If* $\hat{\Sigma}_n$ is not PSD (which can easily be checked), PSD can be enforced by projecting $\hat{\Sigma}_n$ onto the convex cone $\mathbf{S}_+^d$ of $d \times d$ symmetric PSD matrices in the maximum entry-wise distance. Thus, one replaces $\hat{\Sigma}_n$ by[7]

$$\hat{\Sigma}_{n,\mathrm{PSD}} \in \arg\min_{S\in\mathbf{S}_+^d} \max_{1\le j,k\le d}\left|S_{j,k} - \hat{\Sigma}_{n,j,k}\right|.$$

By the minimizing property of $\hat{\Sigma}_{n,\mathrm{PSD}}$ and $\Sigma \in \mathbf{S}_+^d$, the triangle inequality implies that

$$\begin{aligned}\max_{1\le j,k\le d}\left|\hat{\Sigma}_{n,j,k,\mathrm{PSD}} - \Sigma_{j,k}\right| &\le \max_{1\le j,k\le d}\left|\hat{\Sigma}_{n,j,k,\mathrm{PSD}} - \hat{\Sigma}_{n,j,k}\right| + \max_{1\le j,k\le d}\left|\hat{\Sigma}_{n,j,k} - \Sigma_{j,k}\right| \\ &\le 2\max_{1\le j,k\le d}\left|\hat{\Sigma}_{n,j,k} - \Sigma_{j,k}\right|,\end{aligned} \tag{D.3}$$

so that $\hat{\Sigma}_{n,\mathrm{PSD}}$ also satisfies (D.2) (with C replaced by $2C$).

Proof of Theorem D.1. We set up for an elementwise application of Theorem 5.1 in Kock and Preinerstorfer (2025). To this end, fix $1 \le j, k \le d$. Note that $U_{i,j,k},\ i = 1, \ldots, n/2$ is i.i.d. with $\mathbb{E}U_{1,j,k} = \Sigma_{j,k}$. Furthermore, at most $\overline{\eta}_n n = 2\overline{\eta}_n N$ of the $\tilde{U}_{i,j,k}$ differ from $U_{i,j,k}$. Next, by the Cauchy-Schwarz inequality,

$$\begin{aligned}\mathbb{E}\left|U_{1,j,k} - \mathbb{E}U_{1,j,k}\right|^{m/2} \le 2^{m/2}\mathbb{E}\left|U_{1,j,k}\right|^{m/2} &= \mathbb{E}\left|\left(X_{2,j} - X_{1,j}\right)\left(X_{2,k} - X_{1,k}\right)\right|^{m/2} \\ &\le \left(\mathbb{E}\left|X_{2,j} - X_{1,j}\right|^m \mathbb{E}\left|X_{2,k} - X_{1,k}\right|^m\right)^{1/2}.\end{aligned}$$

Thus, since for all $j = 1, \ldots, d$

$$\mathbb{E}\left|X_{2,j} - X_{1,j}\right|^m = \mathbb{E}\left|(X_{2,j} - \mu_j) - (X_{1,j} - \mu_j)\right|^m \le 2^m\sigma_m^m,$$

[7]The convex set of minimizers is non-empty (and closed) by, e.g., Theorem 27.1 in Rockafellar (1997).

the two previous displays imply that

$$\left(\mathbb{E}\left|U_{1,j,k} - \mathbb{E}U_{1,j,k}\right|^{m/2}\right)^{2/m} \leq 4\sigma_m^2. \tag{D.4}$$

Applying Theorem 5.1 in Kock and Preinerstorfer (2025) with n *there* equal to N, m *there* equal to $m/2 > 1$, σ_m *there* equal to $4\sigma_m^2 \leq 4b_2^2$, X_i *there* equal to $U_{i,j,k}$, $\tilde{X}_i$ *there* equal to $\tilde{U}_{i,j,k}$, η *there* equal to $2\overline{\eta}_n$, and $\delta = \frac{6}{d^2N} = \frac{12}{d^2n}$ [inserting these choices there imply that their corresponding ε_c' coincides with ε_n' in (9)], where we also note that $\delta \in (0,1)$ (because, by assumption, throughout $d \geq 2$ and $n > 3$), yields for the pair $1 \leq j,k \leq d$

$$\mathbb{P}\left(\left|\hat{\Sigma}_{n,j,k} - \Sigma_{j,k}\right| > C\left[\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n}\right)^{\frac{1}{2}-\frac{1}{m}}\right]\right) \leq \frac{12}{d^2n},$$

where C depends on b_2, c, and m only and we have inserted $N = n/2$. (D.2) now follows by the union bound. □

Remark D.2. If $m \geq 4$, one can use ε_n' as in (12) and instead appeal to Theorem 3.1 in Kock and Preinerstorfer (2025) [this value of ε_n' being implied by the same choices as right after (D.4)] to show that the resulting estimator $\check{\Sigma}_n$ satisfies

$$\mathbb{P}\left(\max_{1\leq j,k\leq d}\left|\check{\Sigma}_{n,j,k} - \Sigma_{j,k}\right| > C\left[\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n}\right)^{\frac{1}{2}} + \left(\frac{\log(dn)}{n}\right)^{1-\frac{2}{m}}\right]\right) \leq \frac{12}{n},$$

and thus has a better dependence on n and d than $\hat{\Sigma}_n$.

Theorem D.2. *Fix $c \in (1, \sqrt{1.5})$, let Assumption 2.1 be satisfied with $m > 2$. Let $n > 12$. If $\varepsilon_n, \varepsilon_n' \in (0, 1/2)$, with ε_n as in (5) and ε_n' as in (9), then for $\hat{Z} \sim \mathsf{N}_d(0, \hat{\Sigma}_{n,\mathrm{PSD}})$ conditionally on $\hat{\Sigma}_{n,\mathrm{PSD}}$ as in Remark D.1, cf. also (D.1), positive semi-definite, it holds with probability at least $1 - \frac{12}{n}$ that*

$$\begin{aligned}\hat{\rho}_{n,W} &:= \sup_{H\in\mathcal{H}}\left|\mathbb{P}\left(S_{n,W} \in H\right) - \mathbb{P}\left(\hat{Z} \in H \mid \tilde{X}_1, \ldots, \tilde{X}_n\right)\right| \\ &\leq \mathfrak{A}_n + C\left(\log^2(d)\left[\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n}\right)^{\frac{1}{2}-\frac{1}{m}}\right]\right)^{1/2}\end{aligned}$$

where $\mathfrak{A}_n$ is the upper bound on $\rho_{n,W}$ in (6) of Theorem 2.1 and C is a constant depending only on b_1, b_2, c and m.

In particular, $\hat{\rho}_{n,W} \to 0$ *in probability if* $\sqrt{n \log(d)}\overline{\eta}_n^{1-\frac{1}{m}} \to 0$ *and* $\log(d)/n^{\frac{m-2}{5m-2}} \to 0$.

Proof of Theorem D.2. By the triangle inequality $\hat{\rho}_{n,W}$ is bounded from above by the sum of

$$\rho_{n,W} = \sup_{H \in \mathcal{H}} \left| \mathbb{P}\left(S_{n,W} \in H\right) - \mathbb{P}\left(Z \in H\right) \right|,$$

where we recall that $Z \sim \mathsf{N}_d(0, \Sigma)$, and

$$B_n := \sup_{H \in \mathcal{H}} \left| \mathbb{P}\left(Z \in H\right) - \mathbb{P}\left(\hat{Z} \in H \mid \tilde{X}_1, \ldots, \tilde{X}_n\right) \right|.$$

First, $\rho_{n,W} \leq \mathfrak{A}_n$, where $\mathfrak{A}_n$ is the upper bound on $\rho_{n,W}$ in Theorem 2.1. Next, by the Gaussian-to-Gaussian comparison inequality as stated in Proposition 2.1 of Chernozhuokov et al. (2022) (cf. also Proposition 2 in Chernozhukov et al. (2023a)),

$$B_n \leq C\left(\log^2(d) \max_{1 \leq j,k \leq d} \left|\hat{\Sigma}_{n,j,k,\text{PSD}} - \Sigma_{j,k}\right|\right)^{1/2} \leq C\left(\log^2(d)\left[\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n}\right)^{\frac{1}{2}-\frac{1}{m}}\right]\right)^{1/2},$$

the last inequality holding with probability at least $1 - \frac{12}{n}$ by Remark D.1 and C being a constant depending on b_1, b_2, c and m only. That $\hat{\rho}_{n,W} \to 0$ in probability follows from from (B.7)–(B.9) of Lemma B.4. $\square$

We now prove Theorem 3.1 by showing that $\tilde{\Sigma}_n$ is suitably close to $\hat{\Sigma}_n$.

Proof of Theorem 3.1. The theorem is proved by showing that for a constant C depending only on b_2, c, and m,

$$\mathbb{P}\left(\max_{1 \leq j,k \leq d} \left|\tilde{\Sigma}_{n,j,k} - \hat{\Sigma}_{n,j,k}\right| > C\left(\overline{\eta}_n^{1-\frac{2}{m}} + \left[\frac{\log(dn)}{n}\right]^{\frac{1}{2}-\frac{1}{m}}\right)\right) \leq \frac{12}{n}, \tag{D.5}$$

which together with the triangle inequality and Theorem D.1 yields the desired conclusion. To prove (D.5), recall the notation prior to (D.1), fix $1 \leq j, k \leq d$, and note that with

$$A_{n,j,k} := \left\{i \in \{1, \ldots, N\} : \hat{\alpha}_{j,k} \leq \tilde{U}_{i,j,k} \leq \hat{\beta}_{j,k} \text{ and } \hat{a}_j \leq \tilde{Y}_{i,j} \leq \hat{b}_j \text{ and } \hat{a}_k \leq \tilde{Y}_{i,k} \leq \hat{b}_k\right\},$$

one has that

$$\phi_{\hat{\alpha}_{j,k},\hat{\beta}_{j,k}}(\tilde{U}_{i,j,k}) - \phi_{\hat{a}_j,\hat{b}_j}(\tilde{Y}_{i,j})\phi_{\hat{a}_k,\hat{b}_k}(\tilde{Y}_{i,k}) = \tilde{U}_{i,j,k} - \tilde{Y}_{i,j}\tilde{Y}_{i,k} = 0 \qquad \text{for } i \in A_{n,j,k}.$$

Hence,

$$\tilde{\Sigma}_{n,j,k} - \hat{\Sigma}_{n,j,k} = \frac{1}{N}\sum_{i\in A^c_{n,j,k}} \left[\phi_{\hat{\alpha}_{j,k},\hat{\beta}_{j,k}}(\tilde{U}_{i,j,k}) - \phi_{\hat{a}_j,\hat{b}_j}(\tilde{Y}_{i,j})\phi_{\hat{a}_k,\hat{a}_k}(\tilde{Y}_{i,k})\right],$$

and for every $i \in A^c_{n,j,k}$

$$\left|\phi_{\hat{\alpha}_{j,k},\hat{\beta}_{j,k}}(\tilde{U}_{i,j,k}) - \phi_{\hat{a}_j,\hat{b}_j}(\tilde{Y}_{i,j})\phi_{\hat{a}_k,\hat{b}_k}(\tilde{Y}_{i,k})\right| \le \left(|\hat{\alpha}_{j,k}| \vee |\hat{\beta}_{j,k}|\right) + \left(|\hat{a}_j| \vee |\hat{b}_j|\right)\left(|\hat{a}_k| \vee |\hat{b}_k|\right). \tag{D.6}$$

To bound the right-hand side of the previous display, observe that since $n = 2N$,

$$\left|\left\{i \in \{1,\dots,N\} : \tilde{U}_{i,j,k} \ne U_{i,j,k}\right\}\right| \le 2\overline{\eta}_n N \quad \text{and} \quad \left|\left\{i \in \{1,\dots,N\} : \tilde{Y}_{i,j} \ne Y_{i,j}\right\}\right| \le 2\overline{\eta}_n N$$

for $1 \le j,k \le d$. Thus, by Lemma G.5 with n *there* being N, $\overline{\eta}_n$ *there* being $2\overline{\eta}_n$, $\delta = 6/(d^2N)$, [which implies the choice of ε'_n in (9)], noting that $d^2N > 6$ (recall that $d \ge 2$ and $n > 3$ is assumed throughout),

$$Q_{\varepsilon'_n - c^{-1}\varepsilon'_n}(U_{1,j,k}) \le \hat{\alpha}_{j,k} \le \hat{\beta}_{j,k} \le Q_{1-\varepsilon'_n + c^{-1}\varepsilon'_n}(U_{1,j,k})$$

with probability at least $1 - 2\cdot\frac{\delta}{6} = 1 - \frac{2}{d^2N}$, and where the second inequality used that $\varepsilon'_n \in (0,1/2)$. Therefore, with at least this probability,

$$|\hat{\alpha}_{j,k}| \vee |\hat{\beta}_{j,k}| \le |Q_{\varepsilon'_n - c^{-1}\varepsilon'_n}(U_{1,j,k})| \vee |Q_{1-\varepsilon'_n + c^{-1}\varepsilon'_n}(U_{1,j,k})|.$$

By the same argument, it holds with probability at least $1 - \frac{4}{d^2N}$ that

$$\begin{aligned}\left(|\hat{a}_j| \vee |\hat{b}_j|\right)\left(|\hat{a}_k| \vee |\hat{b}_k|\right) \le &\left(|Q_{\varepsilon'_n - c^{-1}\varepsilon'_n}(Y_{1,j})| \vee |Q_{1-\varepsilon'_n + c^{-1}\varepsilon'_n}(Y_{1,j})|\right) \\ &\cdot \left(|Q_{\varepsilon'_n - c^{-1}\varepsilon'_n}(Y_{1,k})| \vee |Q_{1-\varepsilon'_n + c^{-1}\varepsilon'_n}(Y_{1,k})|\right).\end{aligned}$$

Thus, with probability at least $1 - \frac{6}{d^2N}$ the right-hand side of (D.6) is bounded from above by

$$\left(|Q_{\varepsilon_n'-c^{-1}\varepsilon_n'}(U_{1,j,k})| \vee |Q_{1-\varepsilon_n'+c^{-1}\varepsilon_n'}(U_{1,j,k})|\right) + \left(|Q_{\varepsilon_n'-c^{-1}\varepsilon_n'}(Y_{1,j})| \vee |Q_{1-\varepsilon_n'+c^{-1}\varepsilon_n'}(Y_{1,j})|\right) \cdot \left(|Q_{\varepsilon_n'-c^{-1}\varepsilon_n'}(Y_{1,k})| \vee |Q_{1-\varepsilon_n'+c^{-1}\varepsilon_n'}(Y_{1,k})|\right).$$

Next, note that $\mathbb{E}U_{1,j,k} = \Sigma_{j,k}$, $\mathbb{E}\left|U_{1,j,k} - \mathbb{E}U_{1,j,k}\right|^{m/2} \leq 2^m \sigma_m^m$ (cf. (D.4)), $\mathbb{E}Y_{1,j} = 0$, and $\mathbb{E}|Y_{1,j}|^m \leq 2^{m/2}\sigma_m^m$ for all $1 \leq j, k \leq d$. Hence, Lemma G.1 implies that the previous display is bounded from above by

$$\left(|\Sigma_{j,k}| + \frac{4\sigma_m^2}{(\varepsilon_n' - c^{-1}\varepsilon_n')^{2/m}}\right) + \frac{2\sigma_m^2}{(\varepsilon_n' - c^{-1}\varepsilon_n')^{2/m}} \leq \frac{7\sigma_m^2}{(\varepsilon_n' - c^{-1}\varepsilon_n')^{2/m}},$$

the inequality following from $|\Sigma_{j,k}| \leq \sigma_2^2 \leq \sigma_m^2$ and $\varepsilon_n' - c^{-1}\varepsilon_n' \in (0,1)$. Thus, since $|A_{n,j,k}^c| \leq 6\varepsilon_n' N$, with probability at least $1 - \frac{6}{d^2 N}$ it holds that

$$\begin{aligned}\left|\tilde{\Sigma}_{n,j,k} - \hat{\Sigma}_{n,j,k}\right| &\leq 6\varepsilon_n' \cdot \frac{7\sigma_m^2}{(\varepsilon_n' - c^{-1}\varepsilon_n')^{2/m}} \leq 42\sigma_m^2 [c/(c-1)]^{2/m} {\varepsilon_n'}^{1-\frac{2}{m}} \\ &\leq C\left(\overline{\eta}_n^{1-\frac{2}{m}} + \left[\frac{\log(d^2N)}{N}\right]^{\frac{1}{2}-\frac{1}{m}}\right),\end{aligned}$$

where we inserted ε_n' from (9), used subadditivity of $z \mapsto z^{1-\frac{2}{m}}$, and C is a constant depending only on b_2, c, and m. Hence, (D.5) follows by the union bound over the d^2 entries of the covariance matrices and $N = n/2$ upon adjusting multiplicative constants. □

Proof of Theorem 3.2. The proof is almost identical to that of Theorem D.2, but is included for completeness. By the triangle inequality $\tilde{\rho}_{n,W}$ is bounded from above by the sum of

$$\rho_{n,W} = \sup_{H \in \mathcal{H}} \left|\mathbb{P}\left(S_{n,W} \in H\right) - \mathbb{P}\left(Z \in H\right)\right|,$$

where we recall that $Z \sim \mathsf{N}_d(0, \Sigma)$, and

$$B_n := \sup_{H \in \mathcal{H}} \left|\mathbb{P}\left(Z \in H\right) - \mathbb{P}\left(\tilde{Z} \in H \mid \tilde{X}_1, \ldots, \tilde{X}_n\right)\right|.$$

First, $\rho_{n,W} \leq \mathfrak{A}_n$, where $\mathfrak{A}_n$ is the upper bound on $\rho_{n,W}$ in Theorem 2.1. Next, by the Gaussian-to-Gaussian comparison inequality as stated in Proposition 2.1 of Chernozhuokov et al. (2022) (cf. also Proposition 2 in Chernozhukov et al. (2023a)),

$$B_n \leq C\left(\log^2(d) \max_{1\leq j,k\leq d}\left|\tilde{\Sigma}_{n,j,k} - \Sigma_{j,k}\right|\right)^{1/2} \leq C\left(\log^2(d)\left[\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n}\right)^{\frac{1}{2}-\frac{1}{m}}\right]\right)^{1/2},$$

the last inequality holding with probability at least $1 - \frac{24}{n}$ by Theorem 3.1 and C being a constant depending on b_1, b_2, c and m only. That $\tilde{\rho}_{n,W} \to 0$ in probability follows from (B.7)–(B.9) of Lemma B.4. □

E Proofs for Section 4

The following "two-sided" Gaussian anti-concentration inequality is a simple consequence of the "one-sided" one stated in Theorem 1 of Chernozhukov et al. (2017b), (cf. also Lemma A.1 in Chernozhukov et al. (2017a)), but as we could not pinpoint it in the literature we state it here for completeness. For all u, v in $\overline{\mathbb{R}}^d$, we define the set $[u, v] = \left\{x \in \mathbb{R}^d : u_j \leq x_j \leq v_j \text{ for all } j = 1, \ldots, d\right\}$, which may be empty. Recall the notational conventions introduced after (A.7).

Lemma E.1. *Let Z in $\mathbb{R}^d$ with $d \geq 1$ be such that $Z \sim \mathsf{N}_d(0, \Sigma)$ with $\Sigma_{j,j} \geq \underline{\sigma}^2$ for all $j = 1, \ldots, d$ and some $\underline{\sigma}^2 > 0$. Then, for all real numbers δ_1 and δ_2 and all a and b in $\overline{\mathbb{R}}^d$, it holds that*

$$\mathbb{P}\left(Z \in [a + \delta_1, b + \delta_2]\right) \leq \mathbb{P}\left(Z \in [a, b]\right) + \frac{\overline{\delta}}{\underline{\sigma}}\left(\sqrt{2\log(d)} + 4\right), \tag{E.1}$$

where $\overline{\delta} = |\delta_1| \vee |\delta_2|$, and

$$\mathbb{P}\left(Z \in [a + \delta_1, b + \delta_2]\right) \geq \mathbb{P}\left(Z \in [a, b]\right) - \frac{\overline{\delta}}{\underline{\sigma}}\left(\sqrt{2\log(d)} + 4\right). \tag{E.2}$$

Proof. Consider first (E.1). Clearly,

$$\mathbb{P}\left(Z \in [a + \delta_1, b + \delta_2]\right) \leq \mathbb{P}\left(Z \in [a - \overline{\delta}, b + \overline{\delta}]\right) = \mathbb{P}\left((Z', -Z')' \leq (b' + \overline{\delta}, -a' + \overline{\delta})'\right).$$

By Theorem 1 of Chernozhukov et al. (2017b), which trivially remains valid for y there taking values in $\overline{\mathbb{R}}^d$, the far right-hand side of the previous display is bounded from above by the sum of $\mathbb{P}(Z \in [a,b])$ and

$$\frac{\overline{\delta}}{\underline{\sigma}}\left(\sqrt{2\log(2d)}+2\right) = \frac{\overline{\delta}}{\underline{\sigma}}\left(\sqrt{2(\log(2)+\log(d))}+2\right) \leq \frac{\overline{\delta}}{\underline{\sigma}}\left(\sqrt{\log(d)}+2+\sqrt{2(\log(2))}\right),$$

which yields the desired result. To prove (E.2), note that

$$\mathbb{P}\left(Z \in [a+\delta_1, b+\delta_2]\right) \geq \mathbb{P}\left(Z \in [a+\overline{\delta}, b-\overline{\delta}]\right) = \mathbb{P}\left((Z', -Z') \leq (b'-\overline{\delta}, -a'-\overline{\delta})'\right),$$

which, by Theorem 1 of Chernozhukov et al. (2017b), is bounded from below by $\mathbb{P}(Z \in [a,b])$ minus the left-hand side of the penultimate display, implying (E.2). □

We will use the following notation in the proof of Theorem 4.1. For $x \in \mathbb{R}^d$, let $\|x\|_\infty = \max_{j=1,\dots,d} |x_j|$. For any $A \subseteq \mathbb{R}^d$ and $\zeta > 0$, let

$$A^{\zeta,\infty} = \left\{x \in \mathbb{R}^d : \inf_{y \in A} \|x-y\|_\infty \leq \zeta\right\}.$$

Furthermore,

$$A^{-\zeta,\infty} = \left\{x \in \mathbb{R}^d : \mathcal{B}_\infty(x,\zeta) \subseteq A\right\} \quad \text{where} \quad \mathcal{B}_\infty(x,\zeta) = \left\{y \in \mathbb{R}^d : \|y-x\|_\infty \leq \zeta\right\}.$$

Proof of Theorem 4.1. We first establish (14), and assume that

$$C_1\left[\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n}\right)^{\frac{1}{2}-\frac{1}{m}}\right] \leq \frac{b_1^2}{2}, \tag{E.3}$$

where C_1 is the constant from Theorem 3.1 depending only on b_2, c, and m. This is without loss of generality, because if (E.3) does not hold, we can conclude (14) immediately as in the end of the proof of Theorem 2.1.

Recall the definition of $S_{n,W,S}$ in (13), let $T_n \in \mathbb{R}^d$ have entries

$$T_{n,j} = \frac{1}{\sqrt{n}\sigma_{2,j}} \sum_{i=1}^n \left[\phi_{\hat{\alpha}_j,\hat{\beta}_j}(\tilde{X}_{i,j}) - \mu_j\right], \qquad j = 1,\dots,n,$$

and, observe that (grant the quotients are well-defined)

$$A_n := \left\| S_{n,W,S} - T_n \right\|_\infty \leq \max_{j=1,\ldots,d} \left| \frac{1}{\tilde{\sigma}_{n,j}} - \frac{1}{\sigma_{2,j}} \right| \cdot \max_{j=1,\ldots,d} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\phi_{\hat{\alpha}_j, \hat{\beta}_j}(\tilde{X}_{i,j}) - \mu_j \right] \right|. \quad \text{(E.4)}$$

By Theorem 3.1, $\min_{j=1,\ldots,d} \sigma_{2,j} \geq b_1$, and the mean-value theorem, there exists a constant C_2 depending only on b_1, b_2, c, and m such that on a set of probability at least $1-24/n$ it holds that $\tilde{\sigma}_{n,j} \geq b_1/\sqrt{2} > 0$ for every $j = 1, \ldots, d$ (we used (E.3)) and

$$\max_{j=1,\ldots,d} \left| \frac{1}{\tilde{\sigma}_{n,j}} - \frac{1}{\sigma_{2,j}} \right| = \max_{j=1,\ldots,d} \frac{|\tilde{\sigma}_{n,j} - \sigma_{2,j}|}{\tilde{\sigma}_{n,j}\sigma_{2,j}} \leq C_2 \left(\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n} \right)^{\frac{1}{2}-\frac{1}{m}} \right). \quad \text{(E.5)}$$

Furthermore, as $\sigma_2 \leq \sigma_m \leq b_2$, the union bound and $\mathbb{P}(|\boldsymbol{z}| > t) \leq 2\exp(-t^2/2)$ for $t \geq 0$ and $\boldsymbol{z} \sim \mathsf{N}_1(0,1)$ yields

$$\mathbb{P}\left(\max_{j=1,\ldots,d} |Z_j| > b_2 \sqrt{2\log(2dn)} \right) \leq \frac{1}{n}.$$

Thus, by Theorem 2.1, and writing

$$\begin{aligned} r_n := & \left[\frac{\log^{5-\frac{2}{m}}(dn)}{n^{1-\frac{2}{m}}} \right]^{\frac{1}{4}} + \left[\overline{\eta}_n^{1-\frac{1}{m}} + \left[\frac{\log(dn)}{n} \right]^{1-\frac{1}{m}} \right] \sqrt{n \log(d)} \\ & + \left(\log^2(d) \left[\overline{\eta}_n^{1-\frac{2}{m}} + \left[\frac{\log(dn)}{n} \right]^{1-\frac{2}{m}} \right] \right)^{1/2}, \end{aligned}$$

it follows that there exists a constant K_1 depending only on b_1, b_2, c, and m such that

$$\mathbb{P}\left(\max_{j=1,\ldots,d} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\phi_{\hat{\alpha}_j, \hat{\beta}_j}(\tilde{X}_{i,j}) - \mu_j \right] \right| > b_2 \sqrt{2\log(2dn)} \right) \leq K_1 r_n + \frac{1}{n} \leq K_1 r_n, \quad \text{(E.6)}$$

where the value of K_1 is suitably adjusted to justify the second inequality. Hence, by (E.4)–(E.6) there exists a constant C_3 depending only on b_1, b_2, c, and m such that

$$\mathbb{P}\left(A_n \leq C_3 \sqrt{\log(dn)} \left[\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n} \right)^{\frac{1}{2}-\frac{1}{m}} \right] \right) \geq 1 - \frac{24}{n} - K_1 r_n \geq 1 - K r_n, \quad \text{(E.7)}$$

where K depends only on b_1, b_2, c, and m. Thus, writing

$$\overline{A}_n := C_3 \sqrt{\log(dn)} \left[\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n} \right)^{\frac{1}{2}-\frac{1}{m}} \right],$$

it holds for all $H \in \mathcal{H}$ that

$$\left\{ S_{n,W,S} \in H \right\} \subseteq \left\{ T_n \in H^{\overline{A}_n,\infty} \right\} \bigcup \left\{ A_n > \overline{A}_n \right\}. \tag{E.8}$$

Writing $Y_i = D^{-1}X_i$, $\tilde{Y}_i = D^{-1}\tilde{X}_i$ as well as $\check{\alpha}_j = \tilde{Y}^*_{\lceil \varepsilon_n n \rceil, j}$ and $\check{\beta}_j = \tilde{Y}^*_{\lceil (1-\varepsilon_n) n \rceil, j}$, note that

$$T_{n,j} = \frac{1}{\sqrt{n}\sigma_{2,j}} \sum_{i=1}^{n} \left[\phi_{\hat{\alpha}_j, \hat{\beta}_j}(\tilde{X}_{i,j}) - \mu_j \right] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\phi_{\check{\alpha}_j, \check{\beta}_j}(\tilde{Y}_{i,j}) - \mathbb{E}Y_{1,j} \right], \qquad \text{for } j = 1, \ldots, d,$$

such that by Theorem 2.1 and the covariance matrix of Y_1 being $\Sigma_0 = D^{-1}\Sigma D^{-1}$, there exists a constant C depending only on b_1, b_2, c, and m such that (using that $H^{\overline{A}_n,\infty} \in \mathcal{H}$)

$$\left| \mathbb{P}\left(T_n \in H^{\overline{A}_n,\infty} \right) - \mathbb{P}\left(Z' \in H^{\overline{A}_n,\infty} \right) \right| \leq C r_n.$$

Furthermore, by Lemma E.1 (applied with $\delta_1 = -\overline{A}_n$ and $\delta_2 = \overline{A}_n$) and $Z' \sim \mathsf{N}_d(0, \Sigma_0)$ where $\Sigma_{0,j,j} = 1$ for all $j = 1, \ldots, d$,

$$0 \leq \mathbb{P}\left(Z' \in H^{\overline{A}_n,\infty} \right) - \mathbb{P}\left(Z' \in H \right) \leq \overline{A}_n \left(\sqrt{2\log(d)} + 4 \right) \leq 7\overline{A}_n \sqrt{\log(d)},$$

the last inequality following from $d \geq 2$. Using this in (E.8) along with (E.7) yields that

$$\mathbb{P}\left(S_{n,W,S} \in H \right) - \mathbb{P}\left(Z'_j \in H \right) \leq (C+K) r_n + 7\overline{A}_n \sqrt{\log(d)}.$$

Finally, since

$$\left\{ T_n \in H^{-\overline{A}_n,\infty} \right\} \subseteq \left\{ S_{n,W,S} \in H \right\} \bigcup \left\{ A_n > \overline{A}_n \right\},$$

the same arguments as those following (E.8) lead to

$$-(C+K) r_n - 7\overline{A}_n \sqrt{\log(d)} \leq \mathbb{P}\left(S_{n,W,S} \in H \right) - \mathbb{P}\left(Z' \in H \right),$$

which yields (14) under the assumption of (E.3), upon adjusting constants as the upper and lower bounds just obtained do not depend on H.

Finally, since $r_n \to 0$ by (B.7) and (B.8) as well as $\overline{A}_n\sqrt{\log(d)} \to 0$ by (B.10) of Lemma B.4 it follows that $\rho_{n,W,S} \to 0$. □

Proof of Theorem 4.2. The proof is similar to that of Theorem 3.2, but we include it here for completeness. By the triangle inequality $\tilde{\rho}_{n,W,S}$ is bounded from above by the sum of

$$\rho_{n,W,S} = \sup_{H\in\mathcal{H}}\left|\mathbb{P}\left(S_{n,W,S} \in H\right) - \mathbb{P}\left(Z' \in H\right)\right|,$$

where we recall that $Z' \sim \mathsf{N}_d(0, \Sigma_0)$, and

$$B_n := \sup_{H\in\mathcal{H}}\left|\mathbb{P}\left(Z' \in H\right) - \mathbb{P}\left(\hat{Z}' \in H \mid \tilde{X}_1, \ldots, \tilde{X}_n\right)\right|.$$

First, $\rho_{n,W,S} \le \mathfrak{B}_n$, where $\mathfrak{B}_n$ is the upper bound on $\rho_{n,W,S}$ in Theorem 4.1. Next, by the Gaussian-to-Gaussian comparison inequality as stated in Proposition 2.1 of Chernozhuokov et al. (2022) (cf. also Proposition 2 in Chernozhukov et al. (2023a)),

$$B_n \le C\left(\log^2(d)\max_{1\le j,k\le d}\left|\tilde{\Sigma}_{n,0,j,k} - \Sigma_{0,j,k}\right|\right)^{1/2},$$

for an absolute constant C (since $\Sigma_{0,j,j} = 1$ for all $j = 1, \ldots, d$).

Consider the first the case of

$$C_1\left[\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n}\right)^{\frac{1}{2}-\frac{1}{m}}\right] \le \frac{b_1^2}{2}, \tag{E.9}$$

where C_1 is the constant from Theorem 3.1 depending only on b_2, c, and m. Thus, by that theorem there exists a set E_n of probability at least $1 - \frac{24}{n}$ on which

$$\max_{1\le j,k\le d}\left|\tilde{\Sigma}_{n,j,k} - \Sigma_{j,k}\right| \le C_1\left[\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n}\right)^{\frac{1}{2}-\frac{1}{m}}\right] \le \frac{b_1^2}{2}.$$

Therefore, on E_n, $\min_{j=1,\ldots,d}\tilde{\sigma}_{n,j}^2 \ge b_1^2/2$ because by assumption $\min_{j=1,\ldots,d}\sigma_{2,j}^2 \ge b_1^2$. Furthermore, note that for all $1 \le j,k \le d$

$$\tilde{\Sigma}_{n,0,j,k} - \Sigma_{0,j,k} = \frac{\tilde{\Sigma}_{n,j,k}}{\tilde{\sigma}_{n,j}\tilde{\sigma}_{n,j}} - \frac{\Sigma_{j,k}}{\sigma_{2,j}\sigma_{2,k}} = \frac{(\tilde{\Sigma}_{n,j,k} - \Sigma_{j,k})\sigma_{2,j}\sigma_{2,k} + \Sigma_{j,k}(\sigma_{2,j}\sigma_{2,k} - \tilde{\sigma}_{n,j}\tilde{\sigma}_{n,k})}{\tilde{\sigma}_{n,j}\tilde{\sigma}_{n,k}\sigma_{2,j}\sigma_{2,k}},$$

which is well-defined on E_n. Thus, by the mean-value theorem, there exists a constant C_2 depending only on b_1, b_2, c, and m such that on E_n.

$$\max_{1\leq j,k\leq d}\left|\tilde{\Sigma}_{n,0,j,k} - \Sigma_{0,j,k}\right| \leq C_2 \left(\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n}\right)^{\frac{1}{2}-\frac{1}{m}}\right).$$

Hence, with probability at least $1 - \frac{24}{n}$

$$B_n \leq C\left(\log^2(d)\left[\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n}\right)^{\frac{1}{2}-\frac{1}{m}}\right]\right)^{1/2}$$

for a constant C depending only on b_1, b_2, c, and m which implies the bound on $\tilde{\rho}_{n,W,S}$ in (15) in case of (E.9).

In case (E.9) is not satisfied, we conclude as in the end of the proof of Theorem 2.1. Finally, $\tilde{\rho}_{n,W,S} \to 0$ by (B.7)–(B.10) of Lemma B.4. □

F Proofs for Section 5

Proof of Theorem 5.1. Fix $j \in \{1, \ldots, d\}$.

Since $\phi_{\hat{\alpha}_j,\hat{\beta}_j}(\tilde{X}_{i,j}) = \tilde{X}_{i,j}$ for $i \in I_n = \left\{\lceil \varepsilon_n n\rceil, \ldots, \lceil (1-\varepsilon_n)n\rceil\right\}$ one has

$$\begin{aligned}
A_{n,j} :=& \left|S_{n,T,j} - S_{n,W,j}\right| \\
=& \sqrt{n}\left|\frac{1}{|I_n|}\sum_{i=\lceil \varepsilon_n n\rceil}^{\lceil (1-\varepsilon_n)n\rceil}\left[\tilde{X}^*_{i,j} - \mu_j\right] - \frac{1}{n}\sum_{i=1}^{n}\left[\phi_{\hat{\alpha}_j,\hat{\beta}_j}(\tilde{X}^*_{i,j}) - \mu_j\right]\right| \\
=& \sqrt{n}\left|\frac{\lfloor \varepsilon_n n\rfloor + \lceil \varepsilon_n n\rceil - 1}{n|I_n|}\sum_{i=\lceil \varepsilon_n n\rceil}^{\lceil (1-\varepsilon_n)n\rceil}\left[\tilde{X}^*_{i,j} - \mu_j\right] - \frac{\lceil \varepsilon_n n\rceil - 1}{n}\left[\hat{\alpha}_j - \mu_j\right] - \frac{\lfloor \varepsilon_n n\rfloor}{n}\left[\hat{\beta}_j - \mu_j\right]\right| \\
\leq& \sqrt{n}\left(\frac{2\lceil \varepsilon_n n\rceil}{n|I_n|}\sum_{i=\lceil \varepsilon_n n\rceil}^{\lceil (1-\varepsilon_n)n\rceil}\left|\tilde{X}^*_{i,j} - \mu_j\right| + \frac{\lceil \varepsilon_n n\rceil}{n}\left|\hat{\alpha}_j - \mu_j\right| + \frac{\lceil \varepsilon_n n\rceil}{n}\left|\hat{\beta}_j - \mu_j\right|\right).
\end{aligned}$$

Next, since

$$\frac{2\lceil \varepsilon_n n\rceil}{n|I_n|}\sum_{i=\lceil \varepsilon_n n\rceil}^{\lceil (1-\varepsilon_n)n\rceil}\left|\tilde{X}^*_{i,j}-\mu_j\right| \le \frac{2\lceil \varepsilon_n n\rceil}{n}\max_{i=\lceil \varepsilon_n n\rceil,\dots,\lceil (1-\varepsilon_n)n\rceil}\left|\tilde{X}^*_{i,j}-\mu_j\right|$$
$$\le \frac{2\lceil \varepsilon_n n\rceil}{n}\left[|\hat{\alpha}_j-\mu_j|\vee|\hat{\beta}_j-\mu_j|\right],$$

it suffices to bound $|\hat{\alpha}_j-\mu_j|\vee|\hat{\beta}_j-\mu_j|$ from above. To this end, by Lemma G.2 and $\hat{\alpha}_j-\mu_j\le\hat{\beta}_j-\mu_j$ it holds with probability at least $1-\frac{2}{dn}$ that

$$|\hat{\alpha}_j-\mu_j|\vee|\hat{\beta}_j-\mu_j| \le |Q_{\varepsilon_n-c^{-1}\varepsilon_n}(X_{1,j})-\mu_j|\vee|Q_{1-\varepsilon_n+c^{-1}\varepsilon_n}(X_{1,j})-\mu_j| \le \frac{\sigma_m}{(\varepsilon_n-c^{-1}\varepsilon_n)^{1/m}},$$

where the last estimate is by Lemma G.1. Therefore, there exists a constant C depending only on b_2, c, and m such that with probability at least $1-\frac{2}{dn}$,

$$A_{n,j}\le\frac{4\sqrt{n}\lceil \varepsilon_n n\rceil\sigma_m}{n(\varepsilon_n-c^{-1}\varepsilon_n)^{1/m}}\le C\sqrt{n}\varepsilon_n^{1-\frac{1}{m}}\le C\sqrt{n}\left(\overline{\eta}_n^{1-\frac{1}{m}}+\left[\frac{\log(dn)}{n}\right]^{1-\frac{1}{m}}\right)=:\overline{A}_n,$$

with C potentially changing values in the last inequality and subadditivity of $z\mapsto z^{1-\frac{1}{m}}$ was used along with the definition of ε_n in (5). Therefore, since the right-hand side does not depend on j, it follows by the union bound over $j=1,\dots,d$ that with probability at least $1-\frac{2}{n}$

$$A_n := \max_{j=1,\dots,d}A_{n,j}\le\overline{A}_n.$$

Next, observe that for all $H\in\mathcal{H}$ (recalling also the notation introduced prior to the proof of Theorem 4.1)

$$\left\{S_{n,T}\in H\right\}\subseteq\left\{S_{n,W}\in H^{\overline{A}_n,\infty}\right\}\bigcup\left\{A_n>\overline{A}_n\right\}.$$

By Theorem 2.1, $H^{\overline{A}_n,\infty}\in\mathcal{H}$, and Lemma E.1,

$$\mathbb{P}\left(S_{n,W}\in H^{\overline{A}_n,\infty}\right)-\mathbb{P}\left(Z\in H\right)\le\mathfrak{A}_n+\frac{\overline{A}_n}{b_1}\left(\sqrt{2\log(d)}+4\right)\le\mathfrak{A}_n+C\overline{A}_n\sqrt{\log(d)},$$

for a constant C depending only on b_1. Thus, since $\mathbb{P}\left(A_n > \overline{A}_n\right) \leq \frac{2}{n}$, we conclude that

$$\mathbb{P}\left(S_{n,T} \in H\right) - \mathbb{P}\left(Z \in H\right) \leq \mathfrak{A}_n + C\overline{A}_n\sqrt{\log(d)} + \frac{2}{n}. \tag{F.1}$$

Since also

$$\left\{S_{n,W} \in H^{-\overline{A}_n,\infty}\right\} \subseteq \left\{S_{n,T} \in H\right\} \bigcup \left\{A_n > \overline{A}_n\right\},$$

an identical argument shows that

$$-\mathfrak{A}_n - C\overline{A}_n\sqrt{\log(d)} - \frac{2}{n} \leq \mathbb{P}\left(S_{n,T} \in H\right) - \mathbb{P}\left(Z \in H\right), \tag{F.2}$$

which, together with the penultimate display, (F.1)–(F.2) not depending on $H \in \mathcal{H}$, and dominating $2/n$ by $\overline{A}_n\sqrt{\log(d)}$ implies the bound in (17). That $\rho_{n,T} \to 0$ follows from (B.7) and (B.8) of Lemma B.4. □

Proof of Theorem 5.2. The proof is almost identical to that of Theorem 3.2, but is included for completeness. By the triangle inequality $\tilde{\rho}_{n,T}$ is bounded from above by the sum of

$$\rho_{n,T} = \sup_{H\in\mathcal{H}}\left|\mathbb{P}\left(S_{n,T} \in H\right) - \mathbb{P}\left(Z \in H\right)\right|,$$

where we recall that $Z \sim \mathsf{N}_d(0, \Sigma)$, and

$$B_n := \sup_{H\in\mathcal{H}}\left|\mathbb{P}\left(Z \in H\right) - \mathbb{P}\left(\tilde{Z} \in H \mid \tilde{X}_1, \ldots, \tilde{X}_n\right)\right|.$$

First, $\rho_{n,T} \leq \mathfrak{C}_n$, where $\mathfrak{C}_n$ is the upper bound on $\rho_{n,T}$ in Theorem 5.1. Next, by the Gaussian-to-Gaussian comparison inequality as stated in Proposition 2.1 of Chernozhuokov et al. (2022) (cf. also Proposition 2 in Chernozhukov et al. (2023a)),

$$B_n \leq C\left(\log^2(d) \max_{1\leq j,k\leq d}\left|\tilde{\Sigma}_{n,j,k} - \Sigma_{j,k}\right|\right)^{1/2} \leq C\left(\log^2(d)\left[\overline{\eta}_n^{1-\frac{2}{m}} + \left(\frac{\log(dn)}{n}\right)^{\frac{1}{2}-\frac{1}{m}}\right]\right)^{1/2},$$

the last inequality holding with probability at least $1 - \frac{24}{n}$ by Theorem 3.1 and C being a constant depending on b_1, b_2, c and m only. That $\tilde{\rho}_{n,T} \to 0$ in probability follows from (B.7)–(B.9) of Lemma B.4. □

G Auxiliary lemmas

This section gathers some auxiliary lemmas, the proofs of which largely follow from related results in Kock and Preinerstorfer (2025).

The following standard lemma is Lemma B.2 from Kock and Preinerstorfer (2025). It bounds the difference between the mean and quantile of a distribution of a random variable Z (which is not necessarily continuous).

Lemma G.1. *Let Z satisfy $\sigma_m^m := \mathbb{E}|Z - \mathbb{E}Z|^m \in [0, \infty)$ for some $m \in [1, \infty)$. Then, for all $p \in (0, 1)$,*

$$\mathbb{E}Z - \frac{\sigma_m}{p^{1/m}} \leq Q_p(Z) \leq \mathbb{E}Z + \frac{\sigma_m}{(1-p)^{1/m}}. \tag{G.1}$$

The following lemma shows that for ε_n as defined in (5), the lower and upper ε_n order statistics of the contaminated data are close to related population quantiles of the uncontaminated data.

Lemma G.2. *Fix $j \in \{1, \ldots, d\}$, $c \in (1, \sqrt{1.5})$, $n \in \mathbb{N}$, and let Assumption 2.1 be satisfied. If $\varepsilon_n \in (0, 1/2)$ with ε_n as in (5), each of (G.2)–(G.5) below holds with probability at least $1 - \frac{1}{dn}$:*

$$\tilde{X}^*_{\lceil \varepsilon_n n \rceil, j} \geq Q_{\varepsilon_n - c^{-1}\varepsilon_n}(X_{1,j}); \tag{G.2}$$
$$\tilde{X}^*_{\lceil (1-\varepsilon_n) n \rceil, j} \geq Q_{1-\varepsilon_n - c^{-1}\varepsilon_n}(X_{1,j}); \tag{G.3}$$
$$\tilde{X}^*_{\lfloor \varepsilon_n n \rfloor + 1, j} \leq Q_{\varepsilon_n + c^{-1}\varepsilon_n}(X_{1,j}); \tag{G.4}$$
$$\tilde{X}^*_{\lfloor (1-\varepsilon_n) n \rfloor + 1, j} \leq Q_{1-\varepsilon_n + c^{-1}\varepsilon_n}(X_{1,j}). \tag{G.5}$$

Proof. Apply Lemma B.4 in Kock and Preinerstorfer (2025) with $\delta = \frac{6}{dn}$ (recall that $dn > 6$ is assumed throughout; cf. the sentence right before Theorem 2.1), $\eta = \overline{\eta}_n$ to each coordinate $j = 1, \ldots, d$ separately noting that ε_c *there* equals ε_n for $\delta = \frac{6}{dn}$ and that our Assumption 2.1 implies the assumptions *there*. □

Lemma G.3. *If $\varepsilon_n \in (0, 1/2)$ with ε_n as in (5), $c \in (1, \sqrt{1.5})$, and Assumption 2.1 is satisfied then, for all $j = 1, \ldots, d$,*

$$\left| \frac{1}{n} \sum_{i=1}^n \left[\phi_{\underline{\alpha}_j, \underline{\beta}_j}(\tilde{X}_{i,j}) - \phi_{\underline{\alpha}_j, \underline{\beta}_j}(X_{i,j}) \right] \right| \leq 2 \left[\frac{1 - \sqrt{2(c^2-1)}}{(c-1)} \right]^{1/m} \sigma_m \overline{\eta}_n^{1-\frac{1}{m}} \tag{G.6}$$

and

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left[\phi_{\overline{\alpha}_j,\overline{\beta}_j}(\tilde{X}_{i,j})-\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{i,j})\right]\right|\leq 2\left[\frac{1-\sqrt{2(c^2-1)}}{(c-1)}\right]^{1/m}\sigma_m\overline{\eta}_n^{1-\frac{1}{m}}. \tag{G.7}$$

Proof. Fix $j\in\{1,\ldots,d\}$ and recall the definitions of $\underline{\alpha}_j,\overline{\alpha}_j,\underline{\beta}_j$ and $\overline{\beta}_j$ from (A.1) and (A.2). The lemma now follows from applying Lemma B.5 in Kock and Preinerstorfer (2025) to $X_{1,j}$ with $\varepsilon=\varepsilon_n$, $a=c^{-1}\varepsilon_n$, recalling that $\sigma_{m,j}\leq\sigma_m$ and using that $\varepsilon_n\geq\lambda_{1,c}\overline{\eta}_n=\frac{c}{1-\sqrt{2(c^2-1)}}\cdot\overline{\eta}_n$, noting that our Assumption 2.1 implies the assumptions *there* □

Lemma G.4. *If $\varepsilon_n\in(0,1/2)$ with ε_n as in (5), $c\in(1,\sqrt{1.5})$, and Assumption 2.1 is satisfied, then for all $j=1,\ldots,d$,*

$$\left|\mathbb{E}\phi_{\underline{\alpha}_j,\underline{\beta}_j}(X_{1,j})-\mu_j\right|\leq\sigma_m\left(2\left(\frac{c-1}{c}\right)^{1-\frac{1}{m}}+\left(1+\left[\frac{c+1}{c-1}\right]^{\frac{1}{m}}\right)\left(\frac{c+1}{c}\right)^{1-\frac{1}{m}}\right)\varepsilon_n^{1-\frac{1}{m}} \tag{G.8}$$

and

$$\left|\mathbb{E}\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_1)-\mu_j\right|\leq\sigma_m\left(2\left(\frac{c-1}{c}\right)^{1-\frac{1}{m}}+\left(1+\left[\frac{c+1}{c-1}\right]^{\frac{1}{m}}\right)\left(\frac{c+1}{c}\right)^{1-\frac{1}{m}}\right)\varepsilon_n^{1-\frac{1}{m}}. \tag{G.9}$$

Proof. Fix $j\in\{1,\ldots,d\}$ and recall the definitions of $\underline{\alpha}_j,\overline{\alpha}_j,\underline{\beta}_j$ and $\overline{\beta}_j$ from (A.1) and (A.2). Since $\mathbb{E}\phi_{\underline{\alpha}_j,\underline{\beta}_j}(X_{1,j})-\mu_j\leq\mathbb{E}\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{1,j})-\mu_j$, (G.8) and (G.9) follow from Lemma B.7 in Kock and Preinerstorfer (2025) applied to $X_{1,j}$ with $\varepsilon=\varepsilon_n$, $a=c^{-1}\varepsilon_n$ and recalling that $\sigma_{m,j}\leq\sigma_m$ (noting that our Assumption 2.1 implies the assumptions *there*) such that

$$\begin{aligned}
&\left|\mathbb{E}\phi_{\underline{\alpha}_j,\underline{\beta}_j}(X_{1,j})-\mu_j\right|\vee\left|\mathbb{E}\phi_{\overline{\alpha}_j,\overline{\beta}_j}(X_{1,j})-\mu_j\right|\\
\leq&2\sigma_m(\varepsilon_n-c^{-1}\varepsilon_n)^{1-\frac{1}{m}}+\sigma_m\left(1+\left[\frac{\varepsilon_n+c^{-1}\varepsilon_n}{1-\varepsilon_n-c^{-1}\varepsilon_n}\right]^{\frac{1}{m}}\right)(\varepsilon_n+c^{-1}\varepsilon_n)^{1-\frac{1}{m}}\\
\leq&2\sigma_m\left(\frac{c-1}{c}\right)^{1-\frac{1}{m}}\varepsilon_n^{1-\frac{1}{m}}+\sigma_m\left(1+\left[\frac{c+1}{c-1}\right]^{\frac{1}{m}}\right)\left(\frac{c+1}{c}\right)^{1-\frac{1}{m}}\varepsilon_n^{1-\frac{1}{m}}\\
=&\sigma_m\left[2\left(\frac{c-1}{c}\right)^{1-\frac{1}{m}}+\left(1+\left[\frac{c+1}{c-1}\right]^{\frac{1}{m}}\right)\left(\frac{c+1}{c}\right)^{1-\frac{1}{m}}\right]\varepsilon_n^{1-\frac{1}{m}},
\end{aligned}$$

where the second inequality used that $(0,1)\ni x\mapsto x/(1-x)$ is strictly increasing such that $\varepsilon_n + c^{-1}\varepsilon_n \leq 0.5 + c^{-1}0.5$ implies

$$\frac{\varepsilon_n + c^{-1}\varepsilon_c}{1 - \varepsilon_n - c^{-1}\varepsilon_n} \leq \frac{0.5 + c^{-1}0.5}{1 - 0.5 - c^{-1}0.5} = \frac{c+1}{c-1}.$$

□

The following lemma, which is Lemma E.1 in Kock and Preinerstorfer (2025), is an analogue to Lemma G.2, replacing ε_n in (5) with ε_n' in (9).

Lemma G.5. *Fix $c \in (1, \infty)$, $n \in \mathbb{N}$, and $\delta \in (0, 1)$. Let $Z_1, \ldots, Z_n$ be i.i.d. real random variables and suppose that $\tilde{Z}_1, \ldots, \tilde{Z}_n$ satisfy that*

$$\left|\left\{i \in \{1, \ldots, n\} : \tilde{Z}_i \neq Z_i\right\}\right| \leq \overline{\eta}_n n.$$

If $\varepsilon_n' \in (0, 1/2)$ where

$$\varepsilon_n' = c\overline{\eta}_n + c\sqrt{\frac{\log(6/\delta)}{2n}},$$

each of (G.10)–(G.13) below holds with probability at least $1 - \frac{\delta}{6}$:

$$\tilde{Z}^*_{\lceil \varepsilon_n' n \rceil} \geq Q_{\varepsilon_n' - c^{-1}\varepsilon_n'}(Z_1); \tag{G.10}$$

$$\tilde{Z}^*_{\lceil (1-\varepsilon_n') n \rceil} \geq Q_{1-\varepsilon_n' - c^{-1}\varepsilon_n'}(Z_1); \tag{G.11}$$

$$\tilde{Z}^*_{\lfloor \varepsilon_n' n \rfloor + 1} \leq Q_{\varepsilon_n' + c^{-1}\varepsilon_n'}(Z_1); \tag{G.12}$$

$$\tilde{Z}^*_{\lfloor (1-\varepsilon_n') n \rfloor + 1} \leq Q_{1-\varepsilon_n' + c^{-1}\varepsilon_n'}(Z_1). \tag{G.13}$$