Estimating minimum effect with outlier selection

Alexandra Carpentier, Sylvain Delattre, Etienne Roquain, Nicolas Verzelen

HAL Id: hal-03173293
https://hal.inrae.fr/hal-03173293

Submitted on 1 Jun 2022

Estimating minimum effect with outlier selection

Alexandra Carpentier, Sylvain Delattre, Etienne Roquain and Nicolas Verzelen

Abstract: We introduce one-sided versions of Huber's contamination model, in which corrupted samples tend to take larger values than uncorrupted ones. Two intertwined problems are addressed: estimation of the mean of uncorrupted samples (minimum effect) and selection of corrupted samples (outliers). Regarding the minimum effect estimation, we derive the minimax risks and introduce adaptive estimators to the unknown number of contaminations. Interestingly, the optimal convergence rate highly differs from that in classical Huber's contamination model. Also, our analysis uncovers the effect of particular structural assumptions on the distribution of the contaminated samples. As for the problem of selecting the outliers, we formulate the problem in a multiple testing framework for which the location/scaling of the null hypotheses are unknown. We rigorously prove how estimating the null hypothesis is possible while maintaining a theoretical guarantee on the amount of the falsely selected outliers, both through false discovery rate (FDR) or post hoc bounds. As a by-product, we address a long-standing open issue on FDR control under equi-correlation, which reinforces the interest of removing dependency when making multiple testing.

AMS 2000 subject classifications: Primary 62G10; secondary 62C20.
Keywords and phrases: minimax rate, contamination, Hermite polynomials, moment matching, sparsity, multiple testing, false discovery rate, post hoc, selective inference, equi-correlation.

1. Introduction

We are interested in a statistical framework where some data have been corrupted. Depending on how one defines and considers the corruption, such problems have been addressed by different fields in statistics such as robust estimation or sparse modeling. In the former, Huber's contamination model [46, 47] is the prototypical setting for handling this problem. It assumes that among n observations $Y_1, \ldots, Y_n$, most of them follow some normal distribution $\mathcal{N}(\theta, \sigma^2)$ whereas the corrupted ones are arbitrarily distributed. In sparse modeling, one typically assumes that the data $Y_1, \ldots, Y_n$ are normally distributed with mean γ_i where $\gamma_i = \theta$ for uncorrupted samples and arbitrary $\gamma_i \neq \theta$ for corrupted samples (see [10] for a related model).

However, in some practical problems, corrupted samples do not take arbitrary values and satisfy a structural assumption. Consider for instance the following situation where Y_i's are measurements of a pollutant, coming from n sensors spread out at n locations of a city. The background value for this pollutant in the city is θ, but, due to local pollution effects, some sensors may record larger values at some locations. Health authorities are then interested in evaluating the degree of background pollution and in finding where the most affected regions in the city are.

In this work, we introduce one-sided contamination models taking into account the structural assumption that corrupted samples tend to take larger values than uncorrupted ones. Then, we consider the twin problems of estimating the distribution of the uncorrupted samples and identifying the corrupted samples.

1.1. Models and objectives

1.1.1. One-sided Contamination Model (OSC)

We first introduce a one-sided counterpart of Huber contamination model for which some samples Y_i's follow a $\mathcal{N}(\theta, \sigma^2)$ distribution, whereas the remaining samples are positively contaminated, that is, have a distribution that stochastically dominates $\mathcal{N}(\theta, \sigma^2)$, but is otherwise arbitrary.

More formally, we assume that

$$Y_i = \theta + \sigma \varepsilon_i, \quad 1 \leq i \leq n , \tag{1}$$

where $\sigma > 0$ is some standard deviation parameter (either equal to 1 or unknown), $\theta \in \mathbb{R}$ is a fixed *minimum effect* and the ε_i are independent noise random variables. Denoting π_i the unknown distribution of the noise, we assume that, for some k, the distribution $\pi = \otimes_{i=1}^n \pi_i$ of ε belongs to the set

$$\overline{\mathcal{M}}_k = \left\{ \pi = \otimes_{i=1}^n \pi_i \; : \; \pi_i \succeq \mathcal{N}(0,1), \; \sum_{i=1}^n \mathbb{1}_{\{\pi_i \succ \mathcal{N}(0,1)\}} \leq k \right\} , \tag{2}$$

where $\succeq$ (resp. $\succ$) denotes the stochastic domination (resp. strict stochastic domination). In $\overline{\mathcal{M}}_k$, at most k distributions π_i's are allowed to strictly dominate the Gaussian measure. The model (1) satisfies the heuristic explanation described above. If $\pi \in \overline{\mathcal{M}}_k$, then at least $n - k$ samples are non-contaminated and are distributed as $\mathcal{N}(\theta, \sigma)$ whereas the remaining contaminated samples stochastically dominate this distribution.

In this model, henceforth referred as the One-Sided Contamination (OSC) model, the parameter θ corresponds to the expectation of the non-contaminated samples. If $k \leq n - 1$, it also satisfies

$$\theta = \min_{1 \leq i \leq n} \mathbb{E}(Y_i) , \tag{3}$$

and interprets therefore as a minimum theoretical effect. In particular, this model is identifiable for $k \in [n/2, n-1]$, whereas it is not in the classical Huber's model.

Throughout the paper, the probability (resp. expectation) in model (1) is denoted by $\mathbb{P}_{\theta,\pi,\sigma}$ (resp. $\mathbb{E}_{\theta,\pi,\sigma}$). The parameter σ is dropped in the notation whenever $\sigma = 1$.

1.1.2. One-sided Gaussian Contamination Model (gOSC)

In analogy with the sparse Gaussian vector model, we also consider a specific case of OSC model where the contaminated samples are still assumed to be normally distributed, that is, the π_i's are Gaussian distribution with unit variance and positive mean μ_i/σ where $\mu \in \mathbb{R}_+^n$ is a *contamination effect*. In that case, the model can be rewritten as

$$Y_i = \theta + \mu_i + \sigma \xi_i, \quad 1 \leq i \leq n , \tag{4}$$

where ξ_i's are i.i.d. $\mathcal{N}(0,1)$ distributed and $\mu \in \mathbb{R}_+^n$ is unknown. Defining the mean vector

$$\gamma = \theta + \mu , \tag{5}$$

we deduce that Y follows a normal distribution with unknown mean γ and variance $\sigma^2 I_n$ whereas θ corresponds to $\min_i \gamma_i$, that is, the minimum component of the mean vector.

To formalize the connection with the OSC model, we let $\varepsilon_i = \mu_i/\sigma + \xi_i$ and $\pi_i = \mathcal{N}(\mu_i/\sigma, 1)$ for all i. Then, (4) is a particular case of (1) since $\mathcal{N}(\mu_i/\sigma, 1) \succeq \mathcal{N}(0, 1)$. In analogy with the OSC model where we define a collection $\overline{\mathcal{M}}_k$ prescribing the number of contaminated samples to be less or equal to k, we introduce

$$\mathcal{M}_k = \left\{ \mu \in \mathbb{R}_+^n \, : \, \sum_{i=1}^n \mathbb{1}_{\{\mu_i \neq 0\}} \le k \right\} \, . \tag{6}$$

In what follows, we refer to the model (4) as One-Sided Gaussian Contamination (gOSC) model. The probability (resp. expectation) in that model (4) is denoted by $\mathbb{P}_{\theta,\mu,\sigma}$ (resp. $\mathbb{E}_{\theta,\mu,\sigma}$). Whenever we assume that the variance parameter σ is known and is equal to one, the subscript σ is dropped in the above notation.

1.1.3. Objectives

We are interested in the two following intertwined problems:

- *Objective 1: Optimal estimation of the minimum effect.* We aim at establishing the minimax estimation rates of θ both in OSC (1) and in gOSC (4) models. In particular, we explore the role of the one-sided assumption for the computation on such estimation rates. As explained below, this problem is at the crossroads of several lines of research such as robust estimation and non-smooth linear functional estimation.
- *Objective 2: controlled selection of the outliers.* Here, we are interested in finding the contaminated samples. In the Gaussian case (gOSC), this is equivalent to selecting the positive entries of μ in (5). Adopting a multiple testing framework, we aim at designing a selection procedure with suitable false discovery rate (FDR) control [5] and providing a uniformly valid post hoc bound [39, 42]. The difficulty stems from the fact the minimum effect θ is unknown. In contrast to objective 1 where the contaminated samples were considered as nuisance quantities, in this second objective the contaminated samples are now interpreted as the signal whereas θ is a nuisance parameter.

Furthermore, Objective 2 is intrinsically connected to the problem of removing the correlation when making (one-sided) multiple testing from Gaussian equi-correlated test statistics: when the equi-correlation is carried by the latent factor θ, we can remove this correlation by subtracting an estimator of θ to the test statistics. Although this simple strategy is quite common (see, e.g., [35] and references therein), assessing the theoretical performances of such a procedure is a longstanding question in the multiple testing literature. In this work, we establish a positive answer to this question, by showing that it is possible to (asymptotically) control the FDR while having (at least) the same power as if the test statistics had been independent.

In the remainder of the introduction, we first describe our contribution for minimum effect estimation and then turn to outlier selection.

1.2. Optimal estimation of the minimum effect

Given the sparsity $k \in \{1, \ldots, n-1\}$ and $\sigma^2 = 1$, we define the L_1 minimax estimation risk of θ for both gOSC (4) and OSC (1) models:

$$\mathcal{R}[k, n] = \inf_{\widehat{\theta}} \sup_{(\theta,\mu) \in \mathbb{R} \times \mathcal{M}_k} \mathbb{E}_{\theta,\mu}\big[|\widehat{\theta} - \theta|\big]; \quad \overline{\mathcal{R}}[k, n] = \inf_{\widehat{\theta}} \sup_{\theta \in \mathbb{R}, \pi \in \overline{\mathcal{M}}_k} \mathbb{E}_{\theta,\pi}\big[|\widehat{\theta} - \theta|\big] \, . \tag{7}$$

First, we characterize these minimax risks by deriving matching (up to numerical constants) lower and upper bounds, this uniformly over all numbers k of contaminated data, see Sections 2 and 3. The results are summarized in Table 1 below. It is mostly interesting to compare these orders of magnitude with those derived for the Huber contamination model with k contamination. From e.g. [14, Sec.2], we derive[1], that for $k < n/2$, the minimax risk is of order $\min(n^{-1/2}, \frac{k}{n})$. For $k \leq \sqrt{n}$, the rate is parametric in all three models. For $k \in (\sqrt{n}, n/2)$, one-sided contamination lead to some $\sqrt{\log(k^2/n)}$ gain over the Huber's model, whereas assuming that the contaminations are Gaussian lead to an additional logarithmic gain. For $k \in [n/2, n-1]$, recall that Huber's model is not identifiable whereas the one-sided contamination model is, and we identify various minimax rates. For a fixed proportion (k/n) of contaminated samples, the optimal rate still converges to 0 at a polylogarithmic rate. For slowly decaying (with n) proportion $\frac{n-k}{n}$ of non-contaminated samples, the estimation rate still goes to 0.

	General bound	$1 \leq k \leq 2\sqrt{n}$	$2\sqrt{n} \leq k \leq n/2$	$n/2 \leq k \leq n-1$
$\overline{\mathcal{R}}[k,n]$	$\dfrac{\log\left(\frac{n}{n-k}\right)}{\log^{1/2}(1+\frac{k^2}{n})}$	$n^{-1/2}$	$\dfrac{k/n}{\log^{1/2}(k^2/n)}$	$\dfrac{\log\left(\frac{n}{n-k}\right)}{\log^{1/2} n}$
$\mathcal{R}[k,n]$	$\dfrac{\log^2\left(1+\sqrt{\frac{k}{n-k}}\right)}{\log^{3/2}\left(1+(\frac{k}{\sqrt{n}})^{2/3}\right)}$	$n^{-1/2}$	$\dfrac{k/n}{\log^{3/2}(k^2/n)}$	$\dfrac{\log^2\left(\frac{n}{n-k}\right)}{\log^{3/2} n}$

TABLE 1
Minimax estimation risks of θ (up to numerical constants).

For both models (OSC and gOSC), we also devise estimation procedures that are adaptive to the unknown number k of contaminated samples. Finally, we consider the case where the noise level σ in (4) unknown, see Section 4. We prove that, in OSC model, adaptation to unknown σ is possible and characterize the optimal estimation risk for σ.

OSC: Technical aspects and connection to robust estimation. As explained earlier, OSC (1) model is a one-sided counterpart of Huber's contamination model [46, 47] - see also [63] for the historical reference on the concept of contamination and [58, 55] for more recent reviews. From a technical perspective, minimax bounds for OSC proceed from the same general ideas as for Huber's contamination model with a twist. For the latter, the empirical median turns out to be optimal [47]. In OSC model, there is a benefit of using other empirical quantiles. Since the contaminations are one-sided, the left tail is indeed less perturbed than the right tail. Correcting for the bias and choosing suitably a quantile, we prove that the resulting estimator achieves (up to constants) the optimal rate $\overline{\mathcal{R}}[k,n]$. Adaptation to unknown k is performed via Lepski's method whereas adaptation to unknown σ is based on a difference of empirical quantiles.

gOSC: Technical aspects and connection to non-smooth functional estimation. Pinpointing the minimax risk in the Gaussian contamination model (gOSC) is much more

---

[1]Actually, the results in [14] are proved for a model where the number of contaminated sample follows a Binomial distribution with parameters $(k, k/n)$, but the proofs straightforwardly extend to our setting

technical. Indeed, standard estimators, as those based on quantiles for instance, are not optimal in that setting. The key idea of our upper bound is to invert a collection of local tests of the form "$\theta \geq u$" vs "$\theta < u$" for $u \in \mathbb{R}$, by following an approach from [13] developed for sparsity testing. Recall that γ_i in (5) stands for the expectation of Y_i. If "$\theta \geq u$", then $\sum_i \mathbb{1}_{\gamma_i < u} = 0$ whereas under the alternative, one has $\sum_i \mathbb{1}_{\gamma_i < u} \geq n - k$. Thus, this boils down to estimating the non smooth-functional, $\sum_i \mathbb{1}_{\gamma_i < u}$.

Since the seminal work [48, 22] (for respectively the linear and quadratic functional), there is an extensive literature on estimating smooth functionals of the mean of a Gaussian vector. Under a sparsity assumption, the problem has been investigated in [11, 73, 16, 17], and has some deep connections with problem of signal detection [50, 3].

However, estimation of non-smooth functionals (such as $\sum_{i=1}^n |\gamma_i|^q$ for $q \in (0, 1]$)) is significantly more involved even without sparsity assumptions. For related papers, see e.g. [53, 10, 12, 52, 60, 45, 75, 51, 13, 54, 15]. For that class of problem, one powerful approach, coined as polynomial approximation [60, 45], amounts to build a best polynomial approximation of the non-smooth function and plug them with unbiased estimators of the moment $\sum_{i=1}^n \gamma_i^s$ for some integers $s = 1, \ldots, s_{\max}$. Unfortunately, we cannot rely on this strategy for estimating $\sum_i \mathbb{1}_{\gamma_i < u}$, mainly because the contaminated γ_i's may be arbitrarily large. In a related setting, where the contaminated means $\gamma_i \neq \theta$ are distributed according to some smooth prior distributions supported on $\mathbb{R}$, [10] have pinpointed the optimal rate by relying on empirical Fourier transform (see also [21]). However, this approach falls down in our framework because the contaminated γ_i's are arbitrary. In this work, we introduce a new strategy that combines polynomial approximation methods with the empirical Laplace transform.

As for the minimax lower bound, we rely on moment matching techniques following the approach of [60] and recently applied to other non-smooth functional models [12, 45, 75, 13].

1.3. Controlled selection of the outliers

This section presents the state of the art and our contributions for the second objective, that is, controlled selection of the outliers. Our approach relies on multiple testing paradigm and builds upon some of our results on the estimation of θ.

1.3.1. Multiple testing formulation

Our second objective is to identify the active set of outliers in the general model (1). Again, we emphasize that what we call outliers becomes in this part the quantities of interest (e.g., the city locations with abnormal pollutant concentration in our motivating example). In OSC model, we formulate this selection problem under the form of n simultaneous tests of

$$H_{0,i} : \text{``}\pi_i = \mathcal{N}(0, 1)\text{''} \text{ against } H_{1,i} : \text{``}\pi_i \succ \mathcal{N}(0, 1)\text{''}, \text{ for all } 1 \leq i \leq n.$$

(Remember that "$\succ$" stands for strict stochastic domination). In the specific case of gOSC model (4), this problem reduces to simultaneously test

$$H_{0,i} : \text{``}\mu_i = 0\text{''} \text{ against } H_{1,i} : \text{``}\mu_i > 0\text{''}. \tag{8}$$

We denote the set of non-outlier coordinates by $\mathcal{H}_0(\pi) = \{1 \leq i \leq n : \pi_i = \mathcal{N}(0, 1)\}$, and the set of outlier coordinates by $\mathcal{H}_1(\pi) = \{1 \leq i \leq n : \pi_i \succ \mathcal{N}(0, 1)\}$.

The cardinal of $\mathcal{H}_0(\pi)$ (resp. $\mathcal{H}_1(\pi)$) is denoted by $n_0(\pi)$ (resp. $n_1(\pi)$). Hence, $\pi \in \overline{\mathcal{M}}_k$, means that the number of outliers is $n_1(\pi) \leq k$. Thus, our selection problem amounts to

estimate $\mathcal{H}_1(\pi)$ (or equivalently $\mathcal{H}_0(\pi)$). The dependence in π of $\mathcal{H}_0(\pi)$, $\mathcal{H}_1(\pi)$, $n_0(\pi)$, $n_1(\pi)$ is sometimes removed for simplicity.

For any procedure declaring as outliers the elements of $R \subset \{1, \ldots, n\}$, we quantify the amount of false positives in R by a classical metric, introduced in [5], which is called the false discovery proportion of R:

$$\text{FDP}(\pi, R) = \frac{|R \cap \mathcal{H}_0(\pi)|}{|R| \vee 1} \ , \tag{9}$$

which corresponds to the proportion of errors among the set R of selected outliers. The expectation of this quantity is the false discovery rate $\text{FDR}(\pi, R) = \mathbb{E}_{\theta,\pi,\sigma}[\text{FDP}(\pi, R)]$, which can be considered as the standard generalization of the single testing type I error rate in large scale multiple testing. The true discovery proportion is then defined by

$$\text{TDP}(\pi, R) = \frac{|R \cap \mathcal{H}_1(\pi)|}{n_1(\pi) \vee 1} \ , \tag{10}$$

and corresponds to the proportion of (correctly) selected outliers among the set of false null hypotheses. The expectation of this quantity $\mathbb{E}_{\theta,\pi,\sigma}[\text{TDP}(\pi, R)]$ is a widely used analogue of the power in single testing, see, e.g., [68, 2, 64]. Our contribution falls into two frameworks:

- *Multiple testing*: find a procedure selecting a subset $R \subset \{1, \ldots, n\}$ as close as possible to $\mathcal{H}_1(\pi)$, i.e. that has a TDP as high as possible while maintaining a controlled FDR;
- *Post hoc bound*: provide a confidence bound on $\text{FDP}(\pi, S)$, uniformly valid over all possible $S \subset \{1, \ldots, n\}$.

While the first objective is a classical multiple testing aim, see, e.g., [5, 6, 33, 37], the second objective, relatively new, has been proposed in [38, 39, 42]. It is connected to the burgeoning research field of selective inference, see, e.g., [7] and references therein. The rationale behind developing such a bound is that, since the control is uniform, the probability coverage is guaranteed even if the user chooses an arbitrary S, possibly using the same data Y and possibly several times. In other words, the commonly used "data-snooping" is allowed with such bound. We denote the outlier selected set either by R or S depending on the considered issue: R is typically a procedure designed by the statistician, whereas S is chosen by the user.

1.3.2. Relation to the first objective and to previous literature

In OSC model (1), solving the above multiple testing issues is challenging primarily because of the unknown parameters θ and σ. Indeed, this entails that the scaling of the null distribution (i.e. the distribution under the null hypothesis) is unknown. A natural idea is to design a two-stage procedure: first, we estimate θ and σ by some estimators $\widehat{\theta}$ and $\widehat{\sigma}$ (actually this is precisely what we do in the first part of this paper). Then, in a testing stage, we apply a standard multiple testing procedure to the rescaled observation $Y_i' = (Y_i - \widehat{\theta})/\widehat{\sigma}$.

Estimating the null distribution in a multiple testing context has been popularized in a series of work of Efron, see [24, 26, 27]. Through careful data analyses, Efron noticed that the theoretical null distribution often turns out to be wrong in practical situations, which can lead to an uncontrolled increase of false positives. To address this issue, Efron recommends to estimate the scaling parameters of the null distribution (θ, σ here) by "central matching", that is, by fitting a parametric curve to the trimmed data. In his work, Efron provides compelling empirical evidence on his approach. However, up to our knowledge, the FDP and TDP of such

two-stage testing procedures has never been theoretically controlled. Note that estimating the null in a multiple testing context was also the motivation of the minimax results of [53, 10], although the corresponding multiple testing procedure was not studied. We recall that these previous studies are all developed in the two-sided context, whereas our focus is on the one-sided shape constraint.

1.3.3. Summary of our results

In Section 5, we show that some minor modification of the quantile-based estimators $\widehat{\theta}$, $\widehat{\sigma}$ introduced for OSC model, can be used to estimate the null distribution to rescale the p-value process, and can then be suitably combined with classical multiples testing procedures:

1. A new $(\widehat{\theta}, \widehat{\sigma})$-rescaled Benjamini-Hochberg procedure R is defined and proved to enjoy the following FDR controlling property: in general model (1), for any $\pi \in \overline{\mathcal{M}}_k$, with $k = \lfloor 0.9n \rfloor$ (not anti-sparse signal),

$$\left(\mathbb{E}_{\theta, \pi, \sigma} \left(\text{FDP}(\pi, R) \right) - \frac{n_0}{n} \alpha \right)_+ \lesssim \log(n)/n^{1/16} \ .$$

   In addition, we derive a power result showing that the power (TDP) of this procedure is close to the one of the (θ, σ)-rescaled Benjamini-Hochberg procedure (under mild conditions). The latter is an oracle benchmark that would require the exact knowledge of θ and σ.
2. A new $(\widehat{\theta}, \widehat{\sigma})$-rescaled post hoc bound $\overline{\text{FDP}}(\cdot)$ is proposed, satisfying, for any $\pi \in \overline{\mathcal{M}}_k$, with $k = 0.9n$,

$$\left(1 - \alpha - \mathbb{P}_{\theta, \pi, \sigma} \left(\forall S \subset \{1, \ldots, n\}, \ \text{FDP}(\pi, S) \leq \overline{\text{FDP}}(S) \right) \right)_+ \lesssim \log(n)/n^{1/16} \ .$$

To our knowledge, these are the first theoretical results that fully validate Efron's principle of empirical null correction in a specific multiple testing problem.

For bounding the type I error rates, the technical argument used in our proof is close in spirit to recent studies [62, 49] (among others): the idea is to divide the data into two "orthogonal" parts (small or large Y_i's), the first part being used for the rescaling and the second one for testing. For the power result, our formal argument is entirely new to our knowledge, as this kind of results is rarely met in the literature.

1.3.4. Application to decorrelation in multiple testing

It is well known that Efron's methodology on empirical null correction can be applied to reduce the effect of correlations between the tests, as noted by Efron himself [25, 28] where he mentioned that "there is a lot at stake here". Several following work supported this assertion, especially by decomposing the covariance matrix of the data into factors, see [35, 59, 30, 29]. However, strong theoretical results on the corrected multiple testing procedure are still not available.

Meanwhile, another branch of the literature aimed at incorporating known and unknown dependence into multiple testing procedures, for instance, by resampling-based approach [74, 66, 67, 65, 23, 4] or by directly incorporating the known dependence structure [44, 19, 9]. However, as noted for instance in the discussion of [70], even for very simple correlation

structures, no multiple testing procedure has yet been proved to control the FDR while having an optimal TDP.

In Section 5.5, we apply our two-step procedure to address the multiple testing problem in the one-sided Gaussian equi-correlation case (with nonnegative equi-correlation ρ). This model (or its block diagonal variant) is often used as a concrete test bed in multiple testing literature, see, e.g., [56, 18, 20] among others. It turns out that this model can be written under the form of gOSC model (4) with a random value of θ (the variable carrying the equi-correlation) and an unknown variance $\sigma = (1 - \rho)^{1/2}$. Hence, we can directly apply our $(\widehat{\theta}, \widehat{\sigma})$-rescaled Benjamini-Hochberg procedure introduced above to solve the problem: we show that the new procedure has performances close to the BH procedure under independence (and even with a slight increase of the signal to noise ratio). Even if the model is somewhat specific, this shows that correcting the dependence can be fully theoretically justified. To illustrate numerically the benefit of such an approach, Figure 1 displays a ROC-type curve for four different versions of corrected BH procedure. A full description of the simulation setting and additional experiments are provided in Section 5.5.
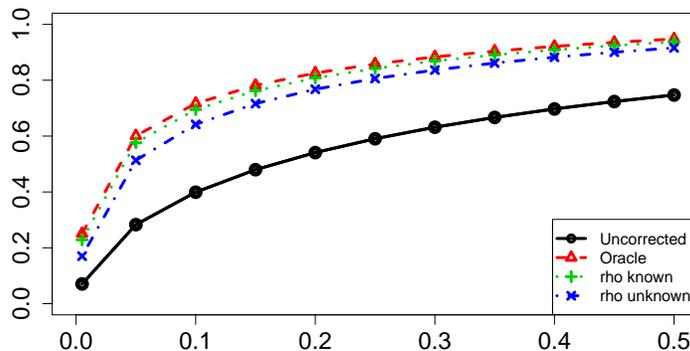

FIG 1. *X-axis: targeted FDR level $\alpha \in \{0.005, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5\}$, Y-axis: TDP (power) averaged over 100 replications for four different procedures (see text in Section 5.5). The model is the one-sided Gaussian with equi-correlation $\rho = 0.3$. The parameters used are $n = 10^6$, $\Delta = 2.5$, $k/n = 0.1$.*

1.4. Notation

For $x > 0$, we write $\lfloor x \rfloor^{(\log_2)}$ (resp. $\lceil x \rceil^{(\log_2)}$) for $2^{\lfloor \log_2(x) \rfloor}$ (resp. $2^{\lceil \log_2(x) \rceil}$) the largest (resp. smallest) dyadic number smaller (resp. higher) than x. Similarly, $\lfloor x \rfloor_{\text{even}}$ is the largest even integer which is not higher than x.

For $x \in \mathbb{R}^n$, $x_{(k)}$ is the k-th smallest element of $\{x_i, 1 \leq i \leq n\}$. We also write $x_{(\ell:m)}$ for the ℓ-smallest element among $\{x_i, 1 \leq i \leq m\}$, for some integer $1 \leq m \leq n$.

In the sequel, c, c' denote numerical positive constants whose values may vary from line to line. For two sequences $(u_t)_{t \in \mathcal{T}}$ and $(v_t)_{t \in \mathcal{T}}$, we write that for all $t \in \mathcal{T}$, $u_t \lesssim v_t$ (resp. for all $t \in \mathcal{T}$, $u_t \gtrsim v_t$), if there exists a universal constant $c > 0$ such that for all $t \in \mathcal{T}$, $u_t \leq c\, v_t$ (resp. for all $t \in \mathcal{T}$, $u_t \geq c\, v_t$). We write $u_t \asymp v_t$ if $u_t \lesssim v_t$ and $v_t \lesssim u_t$.

For X, Y two real random variables with respective cumulative distribution functions F_X, F_Y, we write $X \succeq Y$ if for all $x \in \mathbb{R}$, we have $F_X(x) \leq F_Y(x)$. We write $X \succ Y$ if

$X \succeq Y$ and if there exists $x \in \mathbb{R}$, such that $F_X(x) < F_Y(x)$. We also denote $P \succeq Q$ (resp. $P \succ Q$) whenever $X \succeq Y$ (resp. $X \succ Y$) for $X \sim P$ and $Y \sim Q$.

For the standard normal distribution, we write Φ for its cumulative distribution function, $\bar{\Phi} = 1 - \Phi$ and ϕ for its usual density.

2. Estimation of θ in the gOSC model (4)

In this section, we consider the problem of estimating θ in the Gaussian contamination model (4) and investigate the L_1 minimax risk defined in (7). We assume throughout this section that $\sigma^2 = 1$.

2.1. Lower bound

Theorem 2.1. *There exists a universal constant $c > 0$ such that for any positive integer n and for any integer $k \in [1, n-1]$,*

$$\mathcal{R}[k,n] \geq c \frac{\log^2 \left(1 + \left(\frac{k}{n-k}\right)^{1/2}\right)}{\log^{3/2} \left(1 + \left(\frac{k}{\sqrt{n}}\right)^{2/3}\right)} \ . \tag{11}$$

The proof of this theorem is given in Section A.1. The main tool for proving this lower bound is moment matching: we build two priors on the parameter γ that are related to two different values of θ (as far as possible) while having about $\log n$ first moments that coincide. This is done in an implicit way by using the Hahn-Banach theorem together with properties of Chebychev polynomials, by using techniques close to [54, 13].

Let us distinguish between the three following regimes (see also Table 1):

- for $k \leq \sqrt{n}$, the lower bound (11) is of order $n^{-1/2}$, which is the parametric rate that would hold in the case of no contamination (*i.e.*, $k = 0$);
- for $k \in (\sqrt{n}, \zeta n)$ with $\zeta \in (0, 1)$, the lower bound is of the order $(k/n) \log^{-3/2}(k/\sqrt{n})$. In particular, in the non-sparse case $k = \lceil n/2 \rceil$, we obtain $\log^{-3/2} n$;
- for larger k, e.g., $n/2 \leq k \leq n-1$, the lower bound on the minimax risk is of order $\log^2(\frac{n}{n-k}) \log^{-3/2}(n)$. In particular, for $k = n - 1$, the lower bound is of order $\log^{1/2} n$.

In the remainder of this section, we match these lower bounds by considering three different estimators of θ, corresponding to the three regimes discussed above. They are then combined to derive an adaptive estimator.

2.2. Upper bound for small and large k

For small and for large values k the optimal risk is achieved by simple quantile estimators. For $k \leq n^{1/2}$, we consider the empirical median defined by

$$\widehat{\theta}_{\mathrm{med}} = Y_{(\lceil n/2 \rceil)} \ . \tag{12}$$

The following result holds for $\widehat{\theta}_{\mathrm{med}}$ (note that it is stated in the more general OSC model (1)).

Proposition 2.2. *Consider OSC model* (1) *with* $\sigma = 1$. *Then there exist universal positive constants* c_1, c_2 *and a universal positive integer* n_0 *such that the following holds. For any* $n \geq n_0$, *any* $k \leq n/10$, *any* $\pi \in \overline{\mathcal{M}}_k$ *and any* $\theta \in \mathbb{R}$, *we have*

$$\mathbb{P}_{\theta,\pi}\left[|\widehat{\theta}_{\mathrm{med}} - \theta| \geq \frac{3(k+1)}{2(n-k)} + 3\frac{\sqrt{(n+1)x}}{n-k}\right] \leq e^{-x}, \qquad \text{for all } x \leq c_1 n,$$

$$\mathbb{E}_{\theta,\pi}\left[|\widehat{\theta}_{\mathrm{med}} - \theta|\right] \leq \frac{3(k+1)}{2(n-k)} + \frac{c_2}{\sqrt{n}}.$$

A proof is provided in Section B.1. A consequence if that, for $k \leq \sqrt{n}$, the empirical median $\widehat{\theta}_{\mathrm{med}}$ achieves the parametric rate $n^{-1/2}$, which turns out to be optimal in this regime, see Theorem 2.1. Note that in that regime $k \geq \sqrt{n}$, the empirical medial was already known to achieve this parametric rate in the more general Huber's contamination model, that allows for two sided contaminations.

When k is really close n, there are very few non contaminated data. Since $\theta = \min_i \gamma_i$ in this model (4), we consider debiased empirical minimum estimator

$$\widehat{\theta}_{\mathrm{min}} = Y_{(1)} + \overline{\Phi}^{-1}(1/n), \tag{13}$$

where we recall that $\overline{\Phi}^{-1}(1/n) = \sqrt{2\log(n)} + O(1)$, see Section D.2. The following result holds for $\widehat{\theta}_{\mathrm{min}}$ (note that it is also stated in the more general OSC model (1)).

Proposition 2.3. *Consider OSC model* (1) *with* $\sigma = 1$. *Then there exists some universal positive integer* n_0 *such that for any* $n \geq n_0$, *any* $\pi \in \overline{\mathcal{M}}_{n-1}$ *and* $\theta \in \mathbb{R}$, *the estimator* $\widehat{\theta}_{\mathrm{min}}$ *satisfies*

$$\mathbb{P}_{\theta,\pi}\left[|\widehat{\theta}_{\mathrm{min}} - \theta| \geq 2\sqrt{2\log n}\right] \leq \frac{2}{n}; \quad \mathbb{E}_{\theta,\pi}\left[|\widehat{\theta}_{\mathrm{min}} - \theta|\right] \leq 2\sqrt{2\log n} + 1.$$

A proof is provided in Section B.1. From Theorem 2.1, the estimator $\widehat{\theta}_{\mathrm{min}}$ turns out to be optimal when k is very close to n, e.g. when k larger than $n - n^{\epsilon}$ for a fixed $\epsilon \in (0, 1)$, that is when very few samples are non-contaminated.

2.3. Upper bound in the intermediate regime

In the previous section, we have introduced estimators that are optimal in the regimes where $k \leq \sqrt{n}$ and where k is very close to n, respectively. The intermediate case turns out to be much more involved.

Let $q \geq 2$ be an even integer whose value will be fixed below. Let $a = 3[1 + \log(3 + 2\sqrt{2})] \approx 8.29$ and $q_{\max} = \lfloor \frac{1}{2a}\log n \rfloor_{\mathrm{even}} - 2$, where $\lfloor \cdot \rfloor_{\mathrm{even}}$ is defined in Section 1.4. Let us also introduce two rough estimators $\widehat{\theta}_{\mathrm{up}}$ and $\widehat{\theta}_{\mathrm{low},q}$ such that θ is proved to belong to $[\widehat{\theta}_{\mathrm{low},q}, \widehat{\theta}_{\mathrm{up}}]$ with high probability. Let $\widehat{\theta}_{\mathrm{up}} = Y_{(1)} + 2\sqrt{\log n}$. For any positive and even integer q, define $\widehat{\theta}_{\mathrm{low},q} = \widehat{\theta}_{\mathrm{med}} - \overline{v}$ with $\overline{v} = \pi^2/(144\, q_{\max}^{3/2})$ if $q \leq \frac{3}{10a}\log n$ and $\widehat{\theta}_{\mathrm{low},q} = -\infty$ for larger q.

To explain the intuition behind our procedure, assume for the purpose of the discussion, that we have access to the mean $\gamma_i = \theta + \mu_i$ and that instead of estimating θ, we simply want to test whether θ is greater than u or not. Thus, our aim is to define a suitable function

of γ_i which is close to zero for $\gamma_i \geq u$ and the largest possible when $\gamma_i < u$. Since at least $n - k$'s of the γ_i's are equal to θ, a large value of this function would entail that $\theta < u$. This can be achieved by building $g_q : \mathbb{R} \mapsto \mathbb{R}$ such that $|g_q(x)| \leq 1$ and for $x \in (-\infty, 0]$ and $g_q(x)$ large for $x > 0$ (assuming $u = 0$, without loss of generality). If the interval $(-\infty, 0]$ had been replaced by $[-1, 1]$ and the function g_q was restricted to be a polynomial, this would look like a polynomial extremum problem, which is achieved by a Chebychev polynomial (see Section D.1 for some definitions and properties). To handle the non-bounded interval $(-\infty, 0]$, we map $(-\infty, 0]$ to $(-1, 1]$ using the function $x \mapsto 2e^x - 1$ before using Chebychev polynomials of order q. Denoting by T_q the Chebychev polynomial of degree q, this leads us to considering the function

$$g_q(x) = T_q(2e^x - 1) = \sum_{j=0}^{q} a_{j,q} e^{xj}, \quad x \in \mathbb{R}, \tag{14}$$

where the coefficients $a_{j,q}$ are defined in (115). It follows from the definition of Chebychev polynomials that $g_q(x)$ belongs to $[-1, 1]$ for $x \leq 0$ and $g_q(x) = \cosh[q \arg \cosh(2e^x - 1)]$ for $x > 0$.

Now consider, for $\lambda > 0$ and $u \in \mathbb{R}$, the function $\psi_{q,\lambda}(u)$ defined by

$$\psi_{q,\lambda}(u) = \frac{1}{n} \sum_{i=1}^{n} g_q(\lambda(u - \gamma_i)) = \frac{1}{n} \sum_{i=1}^{n} g_q(\lambda(u - \theta - \mu_i)). \tag{15}$$

This functions depends on the γ_i's. Since all μ_i's are non negative, it follows from the above observation, that $|\psi_{q,\lambda}(u)| \leq 1$ for all $u \leq \theta$. Conversely, for $u \geq \theta$, $\psi_{q,\lambda}(u)$ is lower bounded as follows

$$\psi_{q,\lambda}(u) \geq -\frac{k}{n} + \frac{n - k}{n} g_q(\lambda(u - \theta)), \tag{16}$$

which is bounded away from 1 as long as $u - \theta$ is large enough. As a consequence, the smallest number u_* that satisfies $\psi_{q,\lambda}(u_*) > 1$ should be close (in some sense) to θ.

Obviously, we do not have access to the function $\psi_{q,\lambda}$ as it requires the knowledge of the γ_i's or more precisely of quantities of the form $e^{-j\lambda\gamma_i}$. Nevertheless, we can still build an unbiased estimator of such quantities relying on the empirical Laplace transform of Y. Given $\lambda > 0$ and $u \in \mathbb{R}$ define

$$\widehat{\eta}_\lambda(u) = n^{-1} \sum_{i=1}^{n} e^{\lambda(u - Y_i) - \lambda^2/2}, \qquad \eta_\lambda(u) = n^{-1} \sum_{i=1}^{n} e^{\lambda(u - \theta - \mu_i)}. \tag{17}$$

Since all Y_i's are independent with normal distribution of unit variance, we have $\mathbb{E}[\widehat{\eta}_\lambda(u)] = \eta_\lambda(u)$. This leads us to considering the statistic

$$\widehat{\psi}_{q,\lambda}(u) = \sum_{j=0}^{q} a_{j,q} \widehat{\eta}_{j \cdot \lambda}(u), \tag{18}$$

which is an unbiased estimator of $\psi_{q,\lambda}(u)$ for any fixed $\lambda > 0$ and $u \in \mathbb{R}$. Since $\widehat{\psi}_{q,\lambda}(u)$ approximates $\psi_{q,\lambda}(u)$, it is tempting to take $\widehat{\theta}_q$ as the smallest value such that $\widehat{\psi}_{q,\lambda}(u)$ is bounded away from 1.

Intuitively, $\widehat{\psi}_{q,\lambda}(u)$ is large compared to 1 when $u > \theta$. This is why we define $\widehat{\theta}_q$ by inverting the function $\widehat{\psi}_{q,\lambda}(.)$. More precisely, for an even integer $q \le q_{\max}$, we define $\lambda_q = \sqrt{2/q}$ and the estimator $\widehat{\theta}_q$ by

$$\widehat{\theta}_q = \inf \left\{ u \in \left[\widehat{\theta}_{\mathrm{low},q}, \widehat{\theta}_{\mathrm{up}} \right] : \widehat{\psi}_{q,\lambda_q}(u) > 1 + \frac{e^{aq}}{\sqrt{n}} \right\}, \tag{19}$$

with the convention $\inf\{\emptyset\} = \widehat{\theta}_{\mathrm{up}}$.

Theorem 2.4. *Consider gOSC model* (4) *with known variance* $\sigma = 1$*. There exist universal positive constants* c_1*,* c_2*,* c_3*, and* n_0 *such that the following holds for any* $n \ge n_0$*, any integer* $k \in \left[e^{2a}\sqrt{n}, n - 64n^{1-1/(4a)} \right)$*, any* $\mu \in \mathcal{M}_k$ *and any* $\theta \in \mathbb{R}$*. The estimator* $\widehat{\theta}_{q_k}$ *defined by* (19) *with* $q_k = \lfloor \frac{1}{a} \log \left(\frac{k}{\sqrt{n}} \right) \rfloor_{\mathrm{even}} \wedge q_{\max}$ *satisfies*

$$\mathbb{P}_{\theta,\mu} \left(\widehat{\theta}_{q_k} \notin \left[\theta, \; \theta + c_1 \frac{\log^2 \left(1 + \sqrt{\frac{k}{n-k}} \right)}{\log^{3/2} \left(\frac{k}{\sqrt{n}} \right)} \right] \right) \le c_3 \left(\frac{\sqrt{n}}{k} \right)^{4/3} \log^3 \left(\frac{k}{\sqrt{n}} \right), \tag{20}$$

and

$$\mathbb{E}_{\theta,\mu} \left[|\widehat{\theta}_{q_k} - \theta| \right] \le c_2 \frac{\log^2 \left(1 + \sqrt{\frac{k}{n-k}} \right)}{\log^{3/2} \left(\frac{k}{\sqrt{n}} \right)}. \tag{21}$$

A proof is provided in Section B.3. This result shows that $\widehat{\theta}_{q_k}$ has a maximum risk of order $\frac{k}{n} \log^{-3/2}(k/\sqrt{n})$ in the regime $k \in [e^{2a}\sqrt{n}, n - 64n^{1-1/(4a)})$. Combined with the lower bound of Theorem 2.1, we have shown that $\widehat{\theta}_{q_k}$ is minimax in the intermediate regime.

Remark 2.5. *Let us emphasize that in the regime* $e^{2a}\sqrt{n} \le k \le \lfloor n/2 \rfloor$*, the minimax risk is of order* $(k/n)\log^{-3/2}(n)$*, which is faster than the minimax rate* $(k/n)\log^{-1/2}(n)$ *that we would obtained in a two-sided deconvolution problem, as in [10] where* $k/n \propto n^{-\beta}$ *(and by considering the extreme case where there is no regularity assumption, that is,* $\alpha = 0$ *with their notation).*

Remark 2.6. *If we are only interested in a probability bound* (20) *and not in the moment bound* (21)*, the preliminary estimators* $\widehat{\theta}_{\mathrm{low},q}$ *and* $\widehat{\theta}_{\mathrm{up}}$ *are not needed: the estimator could be computed by taking the minimum over* $\mathbb{R}$ *in* (19)*.*

2.4. Adaptative estimation

In this section, we combine the three estimators studied in the above section to obtain an estimator that is adaptive with respect to the parameter k. The method relies is a Goldenshluger-Lepski approach, see, e.g., [61, 43, 57].

To unify notation, we write henceforth $\widehat{\theta}_0$ for the median estimator $\widehat{\theta}_{\mathrm{med}}$ and $\widehat{\theta}_{q_{\max}+2}$ for the minimum estimator $\widehat{\theta}_{\mathrm{min}}$. In order to obtain an adaptive procedure, we select one of the estimators $\{\widehat{\theta}_q, q \in \{0, 2, \ldots, q_{\max}, q_{\max} + 2\}\}$ as follows:

$$\widehat{q} = \min \left\{ q \in \{0, \ldots, q_{\max} + 2\} \; \text{s.t.} \; |\widehat{\theta}_q - \widehat{\theta}_{q'}| \le \delta_{q'} \; \text{for all} \; q' > q \right\}, \tag{22}$$

where the thresholds are chosen such that $\delta_q = 10 \frac{e^{a(q+2)}}{\sqrt{n} q^{3/2}}$ for $q \in \{2, \ldots, q_{\max} - 2\}$, $\delta_{q_{\max}} = \frac{25}{q_{\max}^{3/2}}$ and $\delta_{q_{\max}+2} = 4\sqrt{2 \log n}$ (the value of a being the same as in Section 2.3).

Theorem 2.7. *Consider gOSC model* (4) *with known variance* $\sigma = 1$. *There exist universal positive constants* c_1, c_2, c_3, *and* n_0 *such that the following holds. For any* $n \geq n_0$, *for any integer* $k \in [1, n-1]$, *any* $\theta \in \mathbb{R}$, *and any* $\mu \in \mathcal{M}_k$, *the adaptive estimator* $\widehat{\theta}_{\text{ad}} = \widehat{\theta}_{\hat{q}}$ *satisfies*

$$\mathbb{P}_{\theta,\mu} \left[|\widehat{\theta}_{\text{ad}} - \theta| > c_1 \frac{\log^2 \left(1 + \sqrt{\frac{k}{n-k}}\right)}{\log^{3/2} \left(1 + (\frac{k}{\sqrt{n}})^{2/3}\right)} \right] \leq c_2 \left(\frac{\sqrt{n}}{k \vee \sqrt{n}} \right)^{4/3} \log^3 \left(\frac{k \vee (2\sqrt{n})}{\sqrt{n}} \right), \quad (23)$$

and

$$\mathbb{E}_{\theta,\mu} \left[|\widehat{\theta}_{\text{ad}} - \theta| \right] \leq c_3 \frac{\log^2 \left(1 + \sqrt{\frac{k}{n-k}}\right)}{\log^{3/2} \left(1 + (\frac{k}{\sqrt{n}})^{2/3}\right)} . \quad (24)$$

A proof is given in Section B.4. The risk bound in (24) matches the minimax lower bound of Theorem 2.1 for all $k = 1, \ldots, n-1$. The estimator $\widehat{\theta}_{\text{ad}}$ is therefore minimax adaptive with respect to k.

Remark 2.8. *Theorem 2.7 shows that the rate of estimation is not affected by the adaptation step. This is specific to our problem, for which the deviations of our estimators are very small when compared to its bias when* $k \geq \sqrt{n}$, *whereas a single estimator, the empirical median, has already good performances over the range* $k \leq \sqrt{n}$.

3. Estimation of θ in the general OSC model

In this section, we study the estimation problem in the general OSC model (1). Hence, the contaminations are not assumed anymore to be Gaussian. Throughout this section, σ is assumed to be known and equal to 1. Recall that the corresponding L_1 minimax risk is given by (7).

3.1. Lower bound

We first show that estimating θ becomes more difficult under this model than for the gOSC case.

Theorem 3.1. *There exists a universal positive constant* c *such that for any positive integer* n *and for any integer* $k \in [1, n-1]$,

$$\overline{\mathcal{R}}[k, n] \geq c \frac{\log \left(\frac{n}{n-k} \right)}{\log^{1/2}(1 + \frac{k^2}{n})} . \quad (25)$$

A proof is provided in Section A.2. Let us comment briefly the order of this lower bound, by going back to the three aforementioned regimes (see also Table 1):

- for $k \leq \sqrt{n}$, the lower bound (25) is of order $n^{-1/2}$, which is the parametric rate, hence is the same as for the Gaussian case;
- for $k \in (\sqrt{n}, \zeta n)$ with $\zeta \in (0, 1)$, the lower bound is of the order $(k/n) \log^{-1/2}(k/\sqrt{n})$, so is strictly slower than with the Gaussian assumption (additional factor of order $\log(k/\sqrt{n})$). In particular, in the non-sparse case $k = \lceil n/2 \rceil$, this gives a lower bound of order $\log^{-1/2}(n)$ (in contrast to $\log^{-3/2}(n)$ in the Gaussian model)

- for larger k, e.g., $n/2 \leq k \leq n-1$, the lower bound is of order $\log(n/(n-k)) \log^{-1/2}(n)$. In comparison to gOSC, there is an additional factor of order $\log(n)/\log(n/(n-k))$. Nevertheless, in the extreme case $k = n-1$, the two lower bounds are of order $\log^{1/2}(n)$.

In the next subsection, these lower bounds are proved to be sharp.

3.2. Upper bound

In this subsection, we introduce a bias-corrected quantile estimator that matches the minimax lower bound of Theorem 3.1. Consider some $\pi \in \overline{\mathcal{M}}_k$. Let $\xi = (\xi_1, \ldots, \xi_n)$ denote a standard Gaussian vector. The starting point is the following: on the one hand, all random variables $Y_i - \theta$ stochastically dominate ξ_i so that $Y_{(q)} - \theta \succeq \xi_{(q)}$. On the other hand, $Y_{(q)}$ is stochastically dominated by the q-th smallest observation among the *non-contaminated* data Y_j. As a consequence, we have

$$\xi_{(q)} \preceq Y_{(q)} - \theta \preceq \xi_{(q:(n-k))} \ , \tag{26}$$

where we recall that $\xi_{(q:(n-k))}$ is the q-th largest observation among the $n - k$ first observations of ξ. Since $\xi_{(q)}$ is concentrated around $\overline{\Phi}^{-1}(q/n)$, this leads to introducing the debiased estimator

$$\widetilde{\theta}_q = Y_{(q)} + \overline{\Phi}^{-1}(q/n) \quad 1 \leq q \leq \lceil n/2 \rceil \ . \tag{27}$$

In view of (12), we have that $\widetilde{\theta}_1 = \widehat{\theta}_{\min}$ while $\widetilde{\theta}_{\lceil n/2 \rceil}$ is almost equal to the empirical median $\widehat{\theta}_{\mathrm{med}}$ (up to the additive $\overline{\Phi}^{-1}(\lceil n/2 \rceil/n)$ term which is of order $1/n$ so is negligible). The following theorem bounds the error of $\widetilde{\theta}_q$ for a wide range of q.

Theorem 3.2. *Consider OSC model* (1) *with known variance* $\sigma = 1$. *There exist universal positive constants* c_1, c_2, c_2', c_3, c_4 *such that the following holds. For all positive integers* $k \leq n-1$, *any* q *such that* $c_4 \log n \leq q \leq (0.7(n-k)) \wedge \lceil n/2 \rceil$, *any* $\theta \in \mathbb{R}$ *and any* $\pi \in \overline{\mathcal{M}}_k$, *the estimator* $\widetilde{\theta}_q$ *satisfies*

$$\mathbb{P}_{\theta,\pi}\left[-c_2\sqrt{\frac{x}{q[\log(\frac{n-k}{q}) \vee 1]}} \leq \widetilde{\theta}_q - \theta \leq c_1\frac{\log\left(\frac{n}{n-k}\right)}{\sqrt{\log\left(\frac{n-k}{q}\right) \vee 1}} + c_2\sqrt{\frac{x}{q[\log(\frac{n-k}{q}) \vee 1]}}\right] \geq 1 - 2e^{-x} \ , \tag{28}$$

for all $0 < x < c_3 q$ *and*

$$\mathbb{E}_{\theta,\pi}\left[|\widetilde{\theta}_q - \theta|\right] \leq c_1\frac{\log\left(\frac{n}{n-k}\right)}{\sqrt{\log\left(\frac{n-k}{q}\right) \vee 1}} + c_2'\frac{1}{\sqrt{q[\log(\frac{n-k}{q}) \vee 1]}} \ . \tag{29}$$

A proof is given in Section B.5. The risk bound in (29) exhibits a bias/variance trade-off as a function of q via the quantities

$$b(q) = \frac{\log\left(\frac{n}{n-k}\right)}{\sqrt{\log\left(\frac{n-k}{q}\right) \vee 1}} \ ; \quad s(q) = \frac{1}{\sqrt{q[\log(\frac{n-k}{q}) \vee 1]}} \ .$$

The quantity $s(q)$ is a deviation term that decreases with q and whose minimum is of the order of $n^{-1/2}$. This minimum is achieved for $q = \lceil n/2 \rceil$ and the corresponding estimator is

close to the empirical median. The quantity $b(q)$ is a bias term which increases slowly with q. Its minimum is of the order of $\log\left(\frac{n}{n-k}\right)\log^{-1/2}(n-k)$ and is achieved for q constant (or of the order of $\log n$). The corresponding estimators are extreme quantiles such as $\widetilde{\theta}_1 = \widehat{\theta}_{\min}$.

Note also that the condition $c_4 \log(n) \leq q \leq 0.7(n-k)$ cannot be met when k is too close to n (i.e. $n - k < (c_4/0.7)\log(n)$). Hence, Theorem 3.2 is silent in that regime. Nevertheless, this case is addressed by the minimum estimator $\widetilde{\theta}_1 = \widehat{\theta}_{\min}$ already studied in Proposition 2.3.

To achieve the minimax risk, it remains to suitably choose q as a function of k. In view of $b(.)$ and $s(.)$, when k is large, one should consider a smaller q and therefore more extreme quantile in order to decrease the bias. More precisely, we define

$$q_k = \begin{cases} \lceil n/2 \rceil & \text{if } k \in [1, 4\sqrt{n}) \ ; \\ \lceil \frac{n^{5/4}}{k^{1/2}} \rceil^{(\log_2)} & \text{if } k \in [4\sqrt{n}, n - n^{4/5}] \ ; \\ 1 & \text{if } k \in (n - n^{4/5}, n - 1] \ . \end{cases} \tag{30}$$

In the very sparse situation ($k \leq 4\sqrt{n}$), $\widetilde{\theta}_{q_k}$ corresponds to the empirical median. For k increasing to n, q_k goes smoothly to $n^{1/4}$. Finally, when k is very close to n, we consider the minimum estimator $\widetilde{\theta}_1$. Other choices of q_k may also lead to optimal risk bounds and the choice (30) is made to simplify the proofs.

Corollary 3.3. *Consider OSC model* (1) *with known variance* $\sigma = 1$. *There exist universal positive constants* c *and* n_0 *such that the following holds. For any integer* $n \geq n_0$, *any integer* $k \in \{1, \ldots, n-1\}$, *any* $\theta \in \mathbb{R}$ *and any* $\pi \in \overline{\mathcal{M}}_k$, *the estimator* $\widetilde{\theta}_{q_k}$ *satisfies*

$$\mathbb{E}_{\theta,\pi}\left[|\widetilde{\theta}_{q_k} - \theta|\right] \leq c \frac{\log\left(\frac{n}{n-k}\right)}{\log^{1/2}(1 + \frac{k^2}{n})} \ . \tag{31}$$

A proof is given in Section B.5. The estimation rate of the estimator $\widetilde{\theta}_{q_k}$ matches the minimax lower bound given in Theorem 3.1. However, it is not adaptive because it uses the value of k.

3.3. Adaptive estimation

We now provide a procedure that adapts to k, by following a Goldenshluger-Lepski approach. Let $\mathcal{Q}$ denote the collections of values of q_k when k goes from 1 to $n - 1$. This collection contains 1, $\lceil n/2 \rceil$ and a dyadic sequence from $n^{1/4}$ to $n/2$ (roughly). To build an adaptive procedure, we select among the estimators $\{\widetilde{\theta}_q, q \in \mathcal{Q}\}$ in the following way:

$$\widehat{q} = \max\left\{q \in \mathcal{Q} \text{ s.t. } |\widehat{\theta}_q - \widehat{\theta}_{q'}| \leq \delta_{q'} \text{ for all } q' < q\right\} , \tag{32}$$

where

$$\delta_q = c_0 \begin{cases} \sqrt{\log(n)} & \text{if } q < \sqrt{2}n^{1/4} \ ; \\ \frac{n^{1/6}}{q^{2/3}\sqrt{\log(\frac{n}{q})\vee 1}} & \text{otherwise} , \end{cases} \tag{33}$$

where the constant c_0 is large enough (and depends on c_1 and c_2' in Theorem 3.2). Note that at most three elements in $\mathcal{Q}$ are less than $\sqrt{2}n^{1/4}$.

Proposition 3.4. *Consider OSC model* (1) *with known variance* $\sigma = 1$*. There exist universal positive constants* c *and* n_0 *such that the following holds. For any integer* $n \geq n_0$*, any integer* $k \in \{1, \ldots, n-1\}$*, any* $\theta \in \mathbb{R}$*, and any* $\pi \in \overline{\mathcal{M}}_k$*, the estimator* $\widetilde{\theta}_{\mathrm{ad}} = \widetilde{\theta}_{\hat{q}}$ *(see* (30) *and* (32)*) satisfies*

$$\mathbb{E}_{\theta,\pi}\left[|\widetilde{\theta}_{\mathrm{ad}} - \theta|\right] \leq c\frac{\log\left(\frac{n}{n-k}\right)}{\log^{1/2}\left(1 + \frac{k^2}{n}\right)} \ .$$

A proof is given in Section B.5. The above result shows that, as in the Gaussian case, adaptation with respect to k can be achieved without any loss.

4. Unknown variance

In this section, we consider OSC model (1) for which the noise variance σ^2 is unknown. We derive the minimax risks and estimators for θ and σ in that setting.

4.1. Lower bound

First note that, obviously, the lower bound (25) for estimating θ ais also a valid lower bound for the minimax risk

$$\inf_{\widetilde{\theta}} \sup_{\theta \in \mathbb{R}, \sigma > 0, \pi \in \overline{\mathcal{M}}_k} \mathbb{E}_{\theta,\pi,\sigma}\left[\frac{|\widetilde{\theta} - \theta|}{\sigma}\right] \ ,$$

corresponding to the OSC model (1) where σ is unknown.

Now, let us provide a lower bound for the estimation risk of σ. As above, it is enough to consider the case where θ is known and equal to zero. This corresponds to the minimax risk

$$\overline{\mathcal{R}}_v[k, n] = \inf_{\widetilde{\sigma}} \sup_{\sigma > 0, \pi \in \overline{\mathcal{M}}_k} \mathbb{E}_{0,\pi,\sigma}\left[\frac{|\widetilde{\sigma} - \sigma|}{\sigma}\right] \ . \tag{34}$$

The following theorem provides a lower bound for $\overline{\mathcal{R}}_v[k, n]$ (and therefore also a lower bound on the minimax risk with arbitrary unknown θ):

Theorem 4.1. *There exists a universal positive constant* c *such that for any integer* $n \geq 2$ *and any* $k = 1, \ldots, n-2$*, we have*

$$\overline{\mathcal{R}}_v[k, n] \geq c\frac{\log\left(\frac{n}{n-k}\right)}{\log\left(1 + \frac{k}{n^{1/2}}\right)} \ . \tag{35}$$

A proof is given in Section A.3. For $k \leq \sqrt{n}$, the lower bound (35) is of order $n^{-1/2}$. For $k \in [\sqrt{n}, \zeta n]$ (with $\zeta \in (0, 1)$ fixed), the risk is of order $k/[n \log(k^2/n)]$ which is faster by a $\log^{1/2}(k^2/n)$ term than for mean estimation. When $n - k = n^{\gamma}$ with $\gamma \in (0, 1)$ (almost no uncontaminated data), the relative rate of convergence is at least constant.

In the next section, we prove that these lower bounds on θ and σ are all sharp (up to numerical constants).

4.2. Upper bound

Since the model is translation invariant, estimating the variance can be done without knowing θ. This is done by considering rescaled differences of empirical quantiles. More precisely, for two positive integers $1 \leq q' \leq q \leq n$, let

$$\widetilde{\sigma}_{q,q'} = \frac{Y_{(q)} - Y_{(q')}}{\overline{\Phi}^{-1}(q'/n) - \overline{\Phi}^{-1}(q/n)} \ , \tag{36}$$

with the convention $0/0 = 0$. When $k = 0$ (no contamination), $Y_{(q)}$ (resp. $Y_{(q')}$) should be close to $\theta - \sigma\overline{\Phi}^{-1}(q/n)$ (resp. $\theta - \sigma\overline{\Phi}^{-1}(q'/n)$) so that, intuitively, $\widetilde{\sigma}_{q,q'}$ should be close to σ. Then, to estimate θ, we simply plug $\widetilde{\sigma}_{q,q'}$ into the quantile estimators considered in Section 3.2. More precisely, we consider

$$\widetilde{\theta}_{q,q'} = Y_{(q)} + \widetilde{\sigma}_{q,q'}\overline{\Phi}^{-1}\left(\frac{q}{n}\right) \ . \tag{37}$$

Given $k \in \{1, \ldots, n-1\}$, q_k is taken as in (30) and

$$q'_k = \begin{cases} \lceil n/3 \rceil & \text{if } k \in [1, 4\sqrt{n}) \ ; \\ \lfloor \frac{n^{7/4}}{k^{3/2}} \rfloor^{(\log_2)} & \text{if } k \in [4\sqrt{n}, n - n^{4/5}] \ ; \\ 1 & \text{if } k \in (n - n^{4/5}, n - 2] \ . \end{cases} \tag{38}$$

For sparse contaminations ($k \leq 4\sqrt{n}$), $\widetilde{\sigma}_{q_k,q'_k}$ is a rescaled difference of the empirical median and the empirical quantile of order $1/3$. For a larger number of contaminations, more extreme quantiles are considered. For $k \geq n - n^{4/5}$, we simply take $\widetilde{\sigma}_{q_k,q'_k} = 0$.

Proposition 4.2. *Consider OSC model* (1) *with unknown variance* σ^2 *and the quantity* q_k *and* q'_k *defined in* (30) *and* (38). *There exist universal positive constants* c, c', *and* n_0 *such that the following holds. For any* $n \geq n_0$, *for any positive integer* $k \leq n - 2$, *any* $\theta \in \mathbb{R}$, *any* $\sigma > 0$ *and any* $\pi \in \overline{\mathcal{M}}_k$, *we have*

$$\mathbb{E}_{\theta,\pi,\sigma}\left[|\widetilde{\sigma}_{q_k,q'_k} - \sigma|/\sigma\right] \leq c\frac{\log\left(\frac{n}{n-k}\right)}{\log(1 + \frac{k}{n^{1/2}})} \ ; \tag{39}$$

$$\mathbb{E}_{\theta,\pi,\sigma}\left[|\widetilde{\theta}_{q_k,q'_k} - \theta|/\sigma\right] \leq c'\frac{\log\left(\frac{n}{n-k}\right)}{\log^{1/2}(1 + \frac{k^2}{n})} \ . \tag{40}$$

A proof is given in Section B.5.

The above proposition together with the lower bounds of Section 4.2 implies that $\widetilde{\sigma}_{q_k,q'_k}$ and $\widetilde{\theta}_{q_k,q'_k}$ are minimax estimator of σ and θ, respectively. In particular, not knowing the variance does not increase the minimax rate when estimating θ.

5. Controlled selection of outliers

In this section, we focus on the general OSC model (1) (with unknown σ) and now turn to the identification of the outliers. As described in Section 1.3, this can be reformulated as a multiple testing problem (see also notation therein).

5.1. Rescaled p-values

As already discussed in Section 1.3, ensuring good multiple testing properties in OSC is challenging because the scaling parameters θ and σ are unknown. A natural approach is then to use the rescaled observations $Y_i' = (Y_i - \widehat{\theta})/\widehat{\sigma}$, $1 \leq i \leq n$, where $\widehat{\theta}$, $\widehat{\sigma}$ are some suitable estimators of θ and σ. To formalize further this idea, let us consider the corrected p-values

$$p_i(u, s) = \overline{\Phi}\left(\frac{Y_i - u}{s}\right), \ \ u \in \mathbb{R}, \ \ s > 0, \ \ 1 \leq i \leq n . \tag{41}$$

The perfectly corrected p-values thus correspond to

$$p_i^\star = p_i(\theta, \sigma), \ \ 1 \leq i \leq n . \tag{42}$$

These oracle p-values cannot be used in practice, because they depend on the unknown parameters θ and σ. Our general aim is to build estimators $\widehat{\theta}$, $\widehat{\sigma}$ such that the theoretical performance of the corrected p-values $p_i(\hat{\theta}, \hat{\sigma})$ mimic those of the oracle p-values $p_i^\star$, when plugged into standard multiple testing or post hoc procedures. If the use of modified p-values and plug-in estimators has often been advocated since the seminal work of Efron [24], proving the convergence of the behavior of the corrected p-values towards the oracle one is, up to our knowledge, new. The challenge is to precisely quantify how the estimation error affects the FDP/TDP metrics. For this, a key point is the following relation between $p_i(u, s)$ and $p_i^\star$:

$$\{p_i(u, s) \leq t\} = \{p_i^\star \leq U_{u,s}(t)\}, \ \ i \in \{1, \ldots, n\}, \ t \in [0, 1] , \tag{43}$$

where

$$U_{u,s}(t) = \overline{\Phi}\left(\frac{s}{\sigma}\overline{\Phi}^{-1}(t) + \frac{u - \theta}{\sigma}\right) \ ; \ \ U_{u,s}^{-1}(v) = \overline{\Phi}\left(\frac{\sigma}{s}\overline{\Phi}^{-1}(v) + \frac{\theta - u}{s}\right) . \tag{44}$$

Furthermore, a useful property is that the order of the p-values does not change after rescaling. We will denote

$$0 = p_{(0)}(u, s) \leq p_{(1)}(u, s) \leq \cdots \leq p_{(n)}(u, s) , \tag{45}$$

the ordered elements of $\{p_i(u, s), 1 \leq i \leq n\}$. We also denote $0 = p_{(0:\mathcal{H}_0)}(u, s) \leq p_{(1:\mathcal{H}_0)}(u, s) \leq \cdots \leq p_{(n_0:\mathcal{H}_0)}(u, s)$ the ordered elements of the subset $\{p_i(u, s), i \in \mathcal{H}_0\}$, that is, of the p-value set corresponding to false outliers (or, equivalently, true null hypotheses).

5.2. Upper-biased estimators

This section provides estimators $\widetilde{\theta}_+$, $\widetilde{\sigma}_+$ that will be suitable to make the p-value rescaling. They are similar to the estimators introduced in Sections 3.2 and 4.2. However, since minimax estimation and false outliers control do not use the same risk metrics, we need to slightly modify these estimators, especially by making them upper-biased (which roughly means that the null hypotheses are favored).

For $q_n = \lfloor n^{3/4} \rfloor$ and $q_n' = \lfloor n^{1/4} \rfloor$, let us consider

$$\begin{cases} \widetilde{\theta}_+ = Y_{(q_n)} + \widetilde{\sigma}_+ \overline{\Phi}^{-1}\left(\frac{q_n}{n}\right) \ ; \\ \widetilde{\sigma}_+ = \dfrac{Y_{(q_n)} - Y_{(q_n')}}{\overline{\Phi}^{-1}(q_n'/(n - k_0)) - \overline{\Phi}^{-1}(q_n/n)} \ , \end{cases} \tag{46}$$

for some parameter $k_0 \leq \lfloor 0.9n \rfloor$. The key difference with the estimators $\widetilde{\theta}_{q,q'}, \widetilde{\sigma}_{q,q'}$ of Section 4 is the quantity k_0 in the denominator of $\widetilde{\sigma}_+$. The following result holds.

Proposition 5.1. *Consider OSC model* (1) *with unknown variance* σ^2. *Then there exists two universal positive constants* c, c' *such that the following holds for any positive integer* n, *for any* $\theta \in \mathbb{R}$, $\sigma > 0$, *for any* $\pi \in \overline{\mathcal{M}}_k$ *with* $k = \lfloor 0.9n \rfloor$. *Choosing* k_0 *such that* $n_1(\pi) \leq k_0 \leq \lfloor 0.9n \rfloor$ *within the estimators* $\widetilde{\theta}_+$, $\widetilde{\sigma}_+$ (46), *we have*

$$\mathbb{P}_{\theta,\pi,\sigma}(\widetilde{\theta}_+ - \theta \leq -\sigma n^{-1/16}) \leq c/n \; ; \tag{47}$$

$$\mathbb{P}_{\theta,\pi,\sigma}(\widetilde{\sigma}_+ - \sigma \leq -\sigma n^{-1/16}) \leq c/n \; ; \tag{48}$$

$$\mathbb{P}_{\theta,\pi,\sigma}\left(|\widetilde{\theta}_+ - \theta| \geq \sigma\big(c'(k_0/n)\log^{-1/2}(n) + n^{-1/16}\big)\right) \leq c/n \; ; \tag{49}$$

$$\mathbb{P}_{\theta,\pi,\sigma}\left(|\widetilde{\sigma}_+ - \sigma| \geq \sigma\big(c'(k_0/n)\log^{-1}(n) + n^{-1/16}\big)\right) \leq c/n \; . \tag{50}$$

Proposition 5.1 is proved in Section B.5. It is strongly related to Theorem 3.2 above, although the statement is slightly different because of the introduced bias in the estimators. Inequalities (47–48) entail that the estimators are (with high probability) above the targeted quantity minus a polynomial term, which will be particularly suitable for obtaining a control on the false positives (FDR control and post hoc bounds). Inequalities (49–50) are two-sided, which is useful for studying the power of the rescaled procedures: there is an additional error term of order $(k_0/n)\log^{-a}(n)$, $a \in \{1/2, 1\}$, where k_0 corresponds to a known upper bound of the number of contaminated coordinates in π.

The assumption $n_1(\pi) \leq k_0 \leq \lfloor 0.9n \rfloor$ in Proposition 5.1 is not very restrictive: it means that the number of outliers is bounded by above by some quantities k_0, which is used in the definition of the estimators (46). For instance, taking $k_0 = \lfloor 0.7n \rfloor$ means that we assume that there is no more than 70% of outliers in the data, which is fair.

Finally note that our multiple testing analysis will only rely on the deviation bounds (47–50). As a consequence, other estimators satisfying these properties can be used for scaling.

5.3. FDR control for selected outliers

The Benjamini-Hochberg (BH) procedure is probably the most famous and widely used multiple testing procedure since its introduction in [5]. Here, the rescaled BH procedure (of nominal level α), denoted $\mathrm{BH}_\alpha(u, s)$ is defined from the *p*-value family $p_i(u, s), 1 \leq i \leq n$, as follows:

- Order the *p*-values as in (45);
- Consider $\widehat{\ell}_\alpha(u, s) = \max\{\ell \in \{0, 1, \ldots, n\} \; : \; p_{(\ell)}(u, s) \leq \alpha\ell/n\}$;
- Reject $H_{0,i}$ for any i such that $p_i(u, s) \leq \widehat{t}_\alpha(u, s)\}$, for $\widehat{t}_\alpha(u, s) = \alpha\widehat{\ell}_\alpha(u, s)/n$.

The procedure, identified to set of the selected outliers, is then given by

$$\mathrm{BH}_\alpha(u, s) = \{1 \leq i \leq n \; : \; p_i(u, s) \leq \widehat{t}_\alpha(u, s)\} \; . \tag{51}$$

The famous FDR-controlling result of [5, 6] can be re-interpreted as follows: the BH procedure using the perfectly corrected *p*-values (42), that is, $\mathrm{BH}_\alpha^\star = \mathrm{BH}_\alpha(\theta, \sigma)$, satisfies

$$\mathbb{E}_{\theta,\pi,\sigma}\left(\mathrm{FDP}(\pi, \mathrm{BH}_\alpha^\star)\right) = \frac{n_0}{n}\alpha \leq \alpha, \quad \text{for all } \theta, \pi, \sigma.$$

This comes from the fact that the perfectly corrected *p*-values (42) are independent, with uniform marginal distributions under the null hypothesis.

Recall the estimators $\widetilde{\theta}_+$ and $\widetilde{\sigma}_+$ defined in (46) with the tuning parameter k_0. The next result gives the behavior of the rescaled procedure $\mathrm{BH}_\alpha(\widetilde{\theta}_+, \widetilde{\sigma}_+)$ both in terms of FDP and TDP.

Theorem 5.2. *Consider OSC model* (1) *with unknown variance* σ^2. *Then, there exists a universal positive constant* c *such that the following holds. For any* $\alpha \in (0, 0.4)$, $\theta \in \mathbb{R}$, $\sigma > 0$, *and any* $\pi \in \overline{\mathcal{M}}_{\lfloor 0.9n \rfloor}$ *such that* $n_1(\pi) \leq k_0 \leq \lfloor 0.9n \rfloor$, *we have*

$$\left(\mathbb{E}_{\theta,\pi,\sigma} \left(FDP(\pi, BH_\alpha(\widetilde{\theta}_+, \widetilde{\sigma}_+)) \right) - \frac{n_0}{n}\alpha \right)_+ \leq c \, \log(n)/n^{1/16} \ . \tag{52}$$

Additionally, for any sequences $\epsilon_n \in (0, 1)$ *tending to zero with* $\epsilon_n \gg \log^{-1/2}(n)$, *for any sequence* $\pi = \pi_n$, $k_0 = k_{0,n}$ *with* $n_1(\pi_n) \leq k_{0,n} \leq \lfloor 0.9n \rfloor$ *and* $n_1(\pi_n)/n \asymp k_{0,n}/n$, *we have for all* θ, σ,

$$\limsup_n \left\{ \mathbb{E}_{\theta,\pi,\sigma} \left(TDP(\pi_n, BH_\alpha^\star) \right) - \mathbb{E}_{\theta,\pi,\sigma} \left(TDP(\pi_n, BH_{\alpha(1+\epsilon_n)}(\widetilde{\theta}_+, \widetilde{\sigma}_+)) \right) \right\} \leq 0 \ . \tag{53}$$

In a nutshell, inequalities (52) and (53) show that the procedure $BH_\alpha(\widetilde{\theta}_+, \widetilde{\sigma}_+)$ behaves similarly to the oracle procedure $BH_\alpha^\star$, both in terms of false discovery rate control and power. These can be seen as a first validation of Efron's theory on empirical null distribution estimation for FDR control.

The proof of this theorem is given in Section C.1. Compared to the usual FDR proofs of the existing literature, there are two additional difficulties: first the independence assumption between the corrected p-values is not satisfied anymore, because the correction terms are random; second, the quantity $FDP(\pi, BH_\alpha(\widetilde{\theta}_+, \widetilde{\sigma}_+))$ is not monotone in the estimators $\widetilde{\theta}_+$, $\widetilde{\sigma}_+$, because of the denominator in the FDP. However, the specific properties of $\widetilde{\theta}_+$, $\widetilde{\sigma}_+$ given in Proposition 5.1 will be enough to get our result: first, these estimators are biased upwards with an error term vanishing at a polynomial rate $n^{-1/16}$, which is enough for false positive control. As for the power, we should consider the bias downwards, which is of order $(k_0/n) \log^{-a}(n)$, $a \in \{1/2, 1\}$. It turns out that the error term induced in the power is of order $(k_0/n) \log^{-a}(n) \log(n/n_1)$, which tends to 0 when $k_0/n \asymp n_1/n$ both in the sparse and non-sparse cases.

5.4. Post hoc bound for selected outliers

We now turn to the problem of finding a post hoc bound, that is, a confidence bound for $FDP(\pi, S)$ which is valid uniformly over all $S \subset \{1, \ldots, n\}$. In [39, 41], the authors showed that post hoc bound can be derived from a simple inequality, called the Simes inequality. This inequality has a long history since the original work of Simes [71] and is still a very active research area, see, e.g., [8, 32].

Specifically, the following property holds for the perfectly corrected p-values $p_i^\star$:

$$\mathbb{P}_{\theta,\pi,\sigma} \left[\exists \ell \in \{1, \ldots, n_0\} \, : \, p_{(\ell:\mathcal{H}_0)}^\star \leq \alpha\ell/n \right] \leq \alpha, \ \ \alpha \in (0, 1) \ , \tag{54}$$

where $p_{(1:\mathcal{H}_0)}^\star \leq \cdots \leq p_{(n_0:\mathcal{H}_0)}^\star$ are the ordered elements of $\{p_i^\star, i \in \mathcal{H}_0\}$.

When replacing the perfect p-values by the estimated ones, the next result shows that Simes inequality is approximately valid.

Theorem 5.3. *Consider OSC model* (1) *with unknown variance* σ^2. *Then, there exists a universal positive constant* c *such that the following holds. For any* $\alpha \in (0, 0.4)$, $\theta \in \mathbb{R}$, $\sigma > 0$, *and any* $\pi \in \overline{\mathcal{M}}_{\lfloor 0.9n \rfloor}$ *such that* $n_1(\pi) \leq k_0 \leq \lfloor 0.9n \rfloor$, *we have*

$$\left(\mathbb{P}_{\theta,\pi,\sigma} \left[\exists \ell \in \{1, \ldots, n_0\} \, : \, p_{(\ell:\mathcal{H}_0)}(\widetilde{\theta}_+, \widetilde{\sigma}_+) \leq \alpha\ell/n \right] - \alpha \right)_+ \leq c \, \log(n)/n^{1/16} \ . \tag{55}$$

The proof is given in Section C.2. It uses that $\widetilde{\theta}_+$, $\widetilde{\sigma}_+$ are bias upwards thanks to Proposition 5.1, together with the monotonicity of the criterion. Let us now define the following data-driven quantity

$$\overline{\text{FDP}}(S; u, s) = 1 \wedge \left\{ \min_{\ell \in \{1,\ldots,n\}} \left(\sum_{i \in S} \mathbb{1}\{p_i(u,s) > \alpha\ell/n\} + \ell - 1 \right) / (|S| \vee 1) \right\}, \ S \subset \{1, \ldots, n\} . \tag{56}$$

The next result shows that (56) is an upper-bound for the FDP, uniformly valid over all the possible selection sets S.

Corollary 5.4. *There exists a numerical constant c such that the following holds. Under the conditions of Theorem 5.3, the bound $\overline{FDP}(\cdot; \widetilde{\theta}_+, \widetilde{\sigma}_+)$ defined by (56) satisfies the following property:*

$$\left(1 - \alpha - \mathbb{P}_{\theta,\pi,\sigma} \left(\forall S \subset \{1, \ldots, n\}, \ FDP(\pi, S) \le \overline{FDP}(S; \widetilde{\theta}_+, \widetilde{\sigma}_+) \right) \right)_+ \le c \, \log(n)/n^{1/16} . \tag{57}$$

The proof is standard from [39, 41] and is given in Section C.3 for completeness. From Corollary 5.4, we deduce, for any selection procedure $\widehat{S}$ possibly depending on the data in an arbitrary way, that the quantity $\overline{\text{FDP}}(\widehat{S}; \widetilde{\theta}_+, \widetilde{\sigma}_+)$ is a valid confidence bound for $\text{FDP}(\pi, \widehat{S})$. Hence, this provides a statistical guarantee on the proportion of false outliers in any data-driven set.

5.5. Application to decorrelation

In this section, we consider a multiple testing issue in the Gaussian equi-correlated model, that corresponds to observe

$$Y \sim \mathcal{N}(m, \Gamma), \quad m \in \mathbb{R}^n, \quad \Gamma = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}, \tag{58}$$

for some unknown $\rho \in [0, 1)$ and some unknown m_i, $1 \le i \le n$. The probability (resp. expectation) in this model is denoted $\mathbb{P}_{m,\rho}$ (resp. $\mathbb{E}_{m,\rho}$). We consider the problem of testing simultaneously:

$$H_{0,i} : \text{``}m_i = 0\text{''} \text{ against } H_{1,i} : \text{``}m_i > 0\text{''}, \text{ for all } 1 \le i \le n. \tag{59}$$

Classically, this model can be rewritten as follows

$$Y_i = m_i + \rho^{1/2} W + (1 - \rho)^{1/2} \zeta_i, \quad 1 \le i \le n, \tag{60}$$

for $W, \zeta_i, 1 \le i \le n$, i.i.d. $\mathcal{N}(0, 1)$. This model shares strong similarities with the gOSC model (and therefore also OSC model), because, conditionally on W, the Y_i's follows model (4)-(5) with $\gamma_i = m_i + \rho^{1/2} W$, $\theta = \rho^{1/2} W$, $\mu_i = m_i$, $\xi_i = \zeta_i$ and $\sigma^2 = 1 - \rho > 0$. In particular, the multiple testing problem (59) is the same as (8) and we can define $\mathcal{H}_0(m), n_0(m), \mathcal{H}_1(m), n_1(m), \text{FDP}(m, \cdot)$ and $\text{TDP}(m, \cdot)$ accordingly, see Section 1.3.1. Whereas the classical p-values $p_i = \overline{\Phi}(Y_i), 1 \le i \le n$ lead to some pessimistic behaviour, we can use the empirically re-scaled p-values to get the following result.

Corollary 5.5. *Consider the model* (58). *There exists a universal positive constant c such that the following holds. For any* $\alpha \in (0, 0.4)$, *any* $\rho \in (0, 1]$, *any mean m satisfying* $n_1(m) \leq k_0 \leq 0.9n$ *(that is, m has at most* $0.9n$ *non-zero coordinates), we have*

(i) *the procedure* $BH_\alpha(\widetilde{\theta}_+, \widetilde{\sigma}_+)$ *defined in Section* 5.3 *satisfies*

$$\left(\mathbb{E}_{m,\rho} \left(FDP(m, BH_\alpha(\widetilde{\theta}_+, \widetilde{\sigma}_+)) \right) - \frac{n_0}{n} \alpha \right)_+ \leq c \log(n)/n^{1/16} .$$

Moreover, for any sequence $\epsilon_n \in (0, 1)$ *tending to zero with* $\epsilon_n \gg \log^{-1/2}(n)$, *for any sequences* m_n *and* $k_0 = k_{0,n}$ *with* $n_1(m_n) \leq k_{0,n}$ *and* $n_1(m_n)/n \asymp k_{0,n}/n$, *we have*

$$\limsup_n \left\{ \mathbb{E}' \left(TDP(m_n, BH_\alpha) \right) - \mathbb{E}_{m_n, \rho} \left(TDP(m_n, BH_{\alpha(1+\epsilon_n)}(\widetilde{\theta}_+, \widetilde{\sigma}_+)) \right) \right\} \leq 0 ,$$

where $\mathbb{E}'$ *denotes the expectation in the model* (58) *in which* $\rho' = 0$ *(independence) and* $m_n' = m_n(1-\rho)^{-1/2}$.

(ii) *the bound* $\overline{FDP}(\cdot; \widetilde{\theta}_+, \widetilde{\sigma}_+)$ *defined by* (56) *satisfies*

$$\left(1 - \alpha - \mathbb{P}_{m,\rho} \left(\forall S \subset \{1, \ldots, n\}, \ FDP(m, S) \leq \overline{FDP}(S; (\widetilde{\theta}_+, \widetilde{\sigma}_+)) \right) \right)_+ \leq c \log(n)/n^{1/16} .$$
$$\tag{61}$$

This result is a direct consequence of Theorems 5.2 and 5.3 by integrating w.r.t. W, so the proof is omitted.

In a nutshell, these results indicate that the analysis of the FDR under independence can be extendeds in the one-sided equi-correlated model, with an additional improvement due to variance reduction by a factor $(1-\rho)$, which can be significant under strong dependence. The condition $n_1(\pi) \leq k_0 \leq \lfloor 0.9n \rfloor$ is not very restrictive as we can always choose $k_0 = \lfloor 0.9n \rfloor$. However, this choice leads to a conservative estimator $\widetilde{\sigma}_+$ and it is obviously better to choose k_0 as close as possible to n_1, the number of non zero coordinates of m. The power result indicates that choosing $k_0/n \asymp n_1/n$ is enough from an asymptotic point of view.

Compared to the state of the art, and especially the work aiming at correcting the dependencies by estimating factors [35, 59, 30, 29], this result is to our knowledge the first one that shows that the corrected procedures is rigorously controlling the desired multiple testing criteria, with some power optimality. In particular, our result shows that the sparsity is not required to make a rigorous dependency correction: the correction is also theoretically valid when $n_1 = \lfloor 3n/4 \rfloor$ for instance. This is due to the one-sided structure of our model and would certainly be not true in the two-sided case. Overall, while our setting is admittedly somewhat narrowed, we argue that this is an important "proof of concept" that supports previous studies and may pave the way for future developments in the area of dependence correction via principal factor approximation.

5.6. *Numerical experiments*

In this section, we illustrate Corollary 5.5 with numerical experiments. We consider the equi-correlation model (58), with a mean m taken of the form $m_i = \Delta$, $1 \leq i \leq n_1$ and $m_i = 0$ otherwise. The results of our experiments are qualitatively the same for other kinds of alternative means, see Section E.

In our simulations, we consider four types of rescaled p-values $p_i(u, s)$, $1 \leq i \leq n$, see (41):

- Uncorrected: $u = 0$, $s = 1$;
- Oracle: $u = \theta = \rho^{1/2} W$ and $s = \sigma = (1 - \rho)^{1/2}$;
- Correlation ρ known:

$$u = \widetilde{\theta}_+(\sigma) = Y_{(q_n)} + \sigma \, \overline{\Phi}^{-1}\left(\frac{q_n}{n}\right), \ \ q_n = \lfloor n^{3/4} \rfloor \, , \tag{62}$$

and $s = \sigma = (1 - \rho)^{1/2}$;
- Correlation ρ unknown: $u = \widetilde{\theta}_+$ and $s = \widetilde{\sigma}_+$ for the estimators defined by (46) with $k_0 = n_1(m)$ (a value of k_0 avoiding a too biased estimation of σ).

Each of these rescaled p-values are used either for FDR control via the Benjamini-Hochberg procedure $\mathrm{BH}_\alpha(u, s)$ (51) (Figure 2), or to get post hoc bound via the Simes bound $\overline{\mathrm{FDP}}(\cdot; u, s)$ (56) (Figure 3).

New corrected BH procedure Figure 2 displays the performances of the rescaled BH procedure. As we can see, the result of this experiment corroborates Proposition 5.5: while the FDR control is maintained, the power of the new procedure is greatly improved. More importantly, estimating the variable W substantially stabilizes the picture and make the FDP/TDP much less variant. This figure also allows to feel the price to pay for estimating ρ, as the procedure using the true value of ρ is closer to the oracle and less variant.
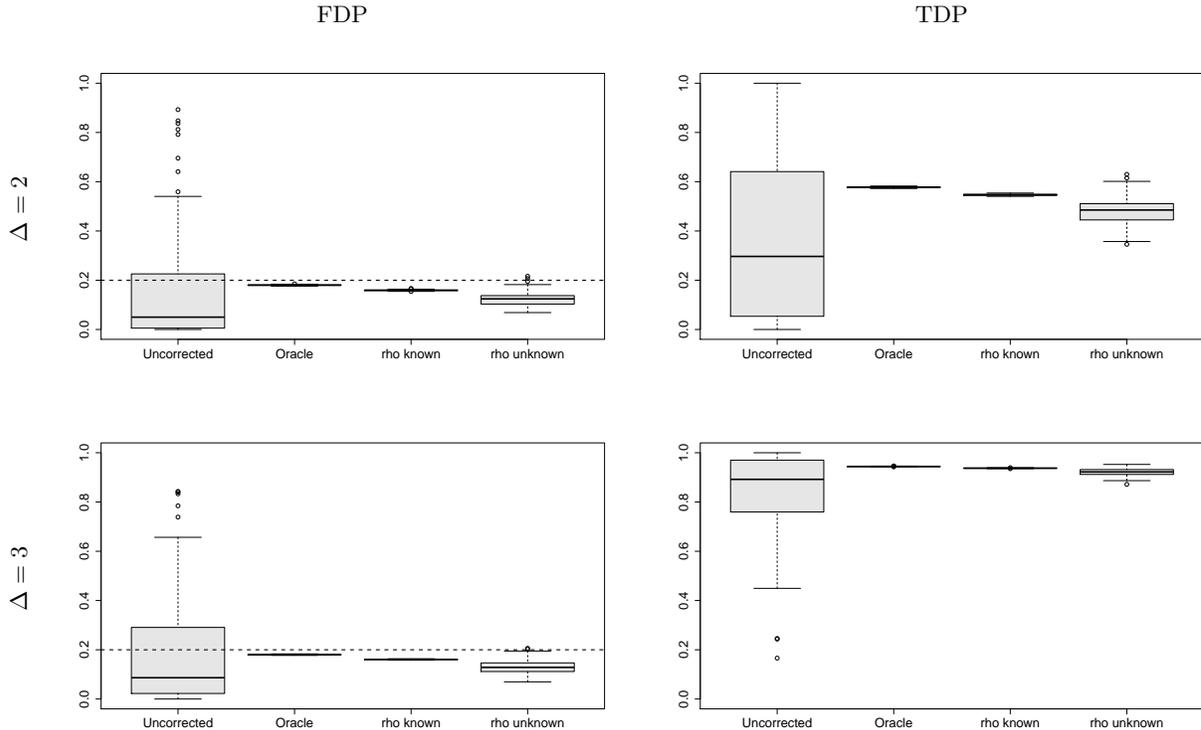

FIG 2. *Boxplot of the FDP and TDP of the BH procedure, with different type of p-value rescaling, see text. The values of the parameters are $n = 10^6$, $\rho = 0.3$, $k/n = 0.1$, $\alpha = 0.2$, 100 replications are done to evaluate the FDP/TDP.*

Post selection bound   Here, we evaluate the quality of the rescaled Simes post hoc bounds. Since these bounds are meant to be uniform over all the possible selection sets S, there is no obvious choice for the set S on which these bounds should be computed. A possible choice, in line with the recent work [40], is to choose a "typical" family of sets $(S_t)_t$ and to look at the quality of the so-derived confidence envelope $t \mapsto \overline{\text{FDP}}(S_t)$ for $t \mapsto \text{FDP}(S_t)$. Here, the subset family $\{S_t, 1 \leq t \leq 200\}$ is built as follows: each S_t is composed of the indices of the t largest values of $\{m_i, 1 \leq i \leq n\}$. Hence, we have simply here $S_t = \{1, \ldots, t\}$ and $\text{FDP}(S_t) = (t - n_1)_+ / t$.

Figure 3 reports the values of the obtained confidence envelopes, for the four types of p-value rescaling described above. Each time, the confidence envelope should be above the true value of the FDP (bold red line), with high probability, and the closer the bound to this quantity, the sharper the bound. The conclusion is similar to the FDP/TDP case. We see less variability for the corrected bounds, especially around the "neck" of the curves ($s \approx n_1 = 50$), which is a point of interest.

Appendix A: Proofs of minimax lower bounds

A.1. Proof of Theorem 2.1 (gOSC)

A.1.1. Extreme cases

First, for $k = 0$, estimating θ amounts to estimating the mean of a Gaussian random variable based on n observation. For this problem, the minimax risk is widely known to be of order $1/\sqrt{n}$ (see standard statistical textbooks such as [72]). Since $\mathcal{R}(k, n)$ is nondecreasing with respect to k if follows that, for all integers k, $\mathcal{R}(k, n) \geq cn^{-1/2}$ for some universal constant $c > 0$. For $k' = n - \lfloor n^{1/4} \rfloor$, we shall prove prove that $\mathcal{R}[k', n] \geq c'\sqrt{\log(n)}$, which in turn implies $\mathcal{R}[k, n] \geq c'\sqrt{\log(n)}$ for all $k \geq k'$.

Lemma A.1. *There exists a constant $c' > 0$ such that for $k' = n - \lfloor n^{1/4} \rfloor$, we have $\mathcal{R}[k', n] \geq c'\sqrt{\log(n)}$.*

This lower bound straightforwardly follows from a reduction of the estimation problem to a detection problem for which minimax separation distance have already been derived. The proof will also serve as a warm-up for the more challenging case $k \in [\sqrt{n}, n - n^{1/4}]$.

Proof of Lemma A.1. To prove this lemma we reduce the problem of estimating θ to a signal detection problem. Write $\underline{k} = n - k' = \lfloor n^{1/4} \rfloor$ for short. Let $a \in (0, 1)$ be a constant that will be fixed later. Given any k, denote $\mathcal{P}[k, n]$, the collection of subset of $\{1, \ldots, n\}$ of size k. Define $\theta_0 = -a\sqrt{\log(n)}$ and for any $S \in \mathcal{P}[\underline{k}, n]$ let μ_S denote the vector such that $(\mu_S)_i = |\theta_0|$ if $i \notin S$ and 0 otherwise. Note that μ_S is k'-sparse, that is, $\mu_S \in \mathcal{M}_{k'}$. It follows from the
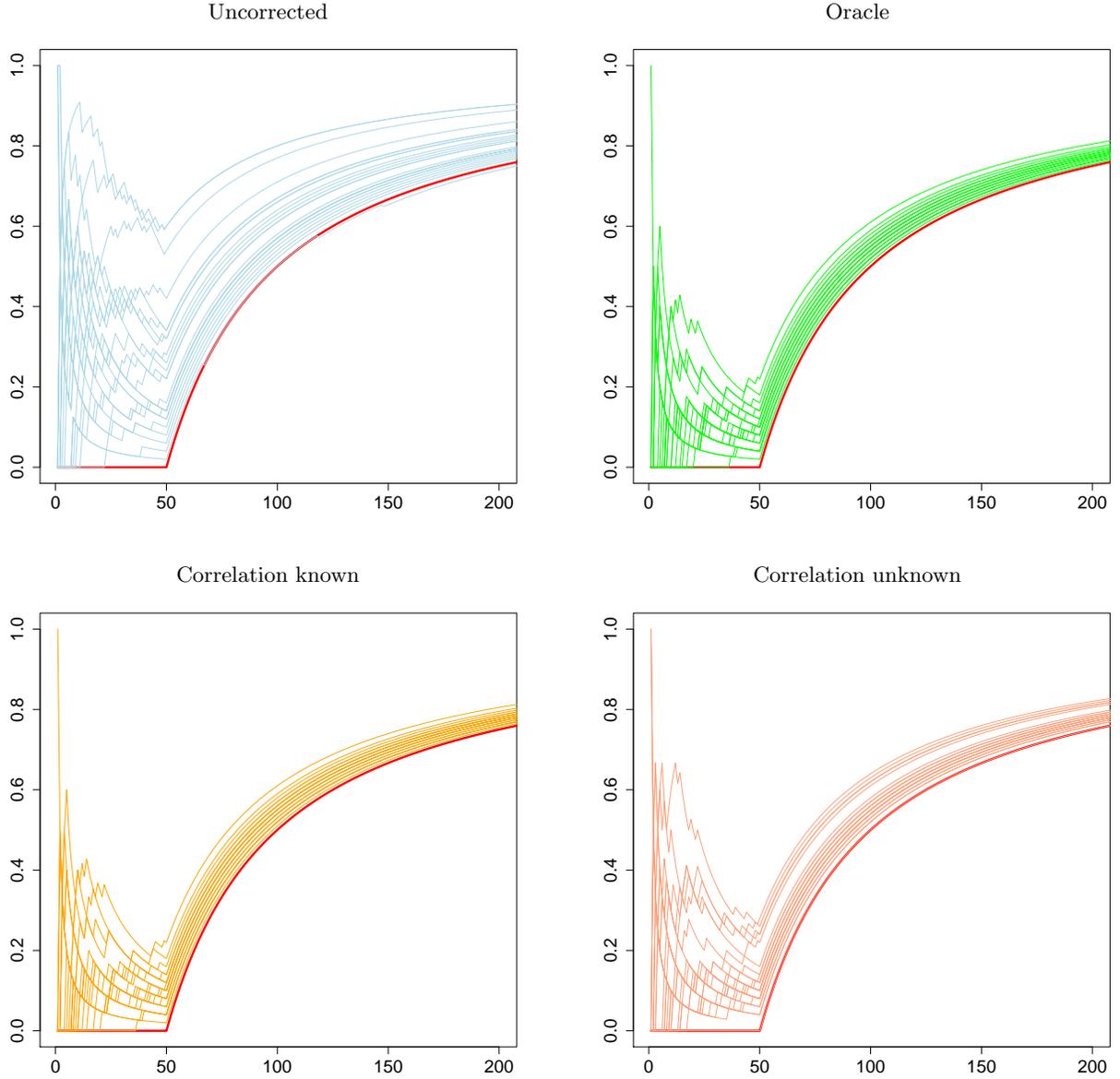
FIG 3. *Plot of 25 trajectories of different post hoc bounds $t \mapsto \overline{FDP}(S_t; \cdot, \cdot)$ for S_t corresponding to the top-t largest means (see (56) for the definition of $\overline{FDP}$ and the text for the exact definition of S_t). From top-left to bottom-right: uncorrected bound $t \mapsto \overline{FDP}(S_t; 0, 1)$; oracle bound $t \mapsto \overline{FDP}(S_t; \theta, \sigma)$; bound with $\sigma = (1 - \rho)^{1/2}$ known $t \mapsto \overline{FDP}(S_t; \widetilde{\theta}_+(\sigma), \sigma)$; bound with σ unknown $t \mapsto \overline{FDP}(S_t; \widetilde{\theta}_+, \widetilde{\sigma}_+)$. The estimators $\widetilde{\theta}_+, \widetilde{\sigma}_+$ (resp. the estimator $\widetilde{\theta}_+(\sigma)$) are given by (46) (resp. 62). In each of these pictures, the unknown value of $t \mapsto FDP(S_t)$ is displayed with the red bold line. The values of the parameters are $n = 10^3$, $\rho = 0.3$, $k/n = 0.05$, $\alpha = 0.2$ and $\Delta = 4$.*

definition of the minimax risk

$$
\begin{aligned}
\mathcal{R}[k', n] \;\; &\geq \;\; \frac{1}{2} \inf_{\widehat{\theta}} \left[\mathbb{E}_{0,0}[|\widehat{\theta}|] + \max_{S \in \mathcal{P}[\underline{k}, n]} \mathbb{E}_{\theta_0, \mu_S}[|\widehat{\theta} - \theta_0|] \right] \\
&\geq \;\; \frac{a}{4} \sqrt{\log(n)} \left[\mathbb{P}_{0,0}[\widehat{T} = 0] + \max_{S \in \mathcal{P}[\underline{k}, n]} \mathbb{P}_{\theta_0, \mu_S}[\widehat{T} = 1] \right].
\end{aligned}
$$

by letting $\widehat{T} = \mathbb{1}_{\widehat{\theta} \geq \theta_0/2}$ and by using $(|\theta_0|/2)\mathbb{1}_{\widehat{T}=0} \leq |\widehat{\theta}|$ and $(|\theta_0|/2)\mathbb{1}_{\widehat{T}=1} \leq |\widehat{\theta} - \theta_0|$. Thus

$$\mathcal{R}[k', n] \geq \frac{a}{4}\sqrt{\log(n)} \inf_{\widehat{T}} \left[\mathbb{P}_{0,0}[\widehat{T} = 0] + \max_{S \in \mathcal{P}[\underline{k},n]} \mathbb{P}_{\theta_0,\mu_S}[\widehat{T} = 1] \right].$$

As a consequence, if a is chosen small enough such that no test is able to decipher reliably between $\mathbb{P}_{0,0}$ and $\{\mathbb{P}_{\theta_0,\mu_S}, \ S \in \mathcal{P}[\underline{k},n]\}$, then the minimax risk is of the order of $a\sqrt{\log(n)}$. Note that this problem amounts to testing in a simple Gaussian white noise model $\mathcal{N}(\gamma, I_n)$ whether the mean vector γ is zero or if $\gamma = \theta_0 + \mu_S$ is $\underline{k}$-sparse with negative non-zero values that are all equal to $-a\sqrt{\log(n)}$. Quantifying the difficulty of this problem is classical in the statistical literature and has be done for instance in [3] (for non-necessarily positive values). For the sake of completeness, we shall provide exhaustive arguments. Denote $\overline{\mathbb{P}}_{\underline{k}} = |\mathcal{P}[\underline{k},n]|^{-1} \sum_S \mathbb{P}_{\theta_0,\mu_S}$ the mixture measure when S is sampled uniformly in $\mathcal{P}[\underline{k},n]$. Since the supremum is larger than the mean, we obtain

$$\mathcal{R}[k', n] \geq \frac{a}{4}\sqrt{\log(n)} \inf_{\widehat{T}} \left[\mathbb{P}_{0,0}[\widehat{T} = 0] + \overline{\mathbb{P}}_{\underline{k}}[\widehat{T} = 1] \right]$$

$$\geq \frac{a}{4}\sqrt{\log(n)}(1 - \|\mathbb{P}_{0,0} - \overline{\mathbb{P}}_{\underline{k}}\|_{TV}) \geq \frac{a}{4}\sqrt{\log(n)} \left(1 - \sqrt{\chi^2(\overline{\mathbb{P}}_{\underline{k}}; \mathbb{P}_{0,0})} \right) , \quad (63)$$

since the χ^2 discrepancy between distributions dominates the square of the total variation distance $\|.\|_{TV}$, see Section 2.4 in [72]. Writing L the likelihood ratio of $\overline{\mathbb{P}}_{\underline{k}}$ over $\mathbb{P}_{0,0}$ and, for $S \in \mathcal{P}[\underline{k},n]$, L_S the likelihood ratio of $\mathbb{P}_{\theta_0,\mu_S}$ over $\mathbb{P}_{0,0}$, and π the uniform measure over $\mathcal{P}[\underline{k},n]$, we have

$$\begin{aligned}
1 + \chi^2(\overline{\mathbb{P}}_{\underline{k}}; \mathbb{P}_{0,0}) &= \mathbb{E}_{0,0}[L^2] = \mathbb{E}_{0,0}[\pi^{\otimes 2}(L_S L_{S'})] \\
&= \pi^{\otimes 2}\left[\mathbb{E}_{0,0}(L_S L_{S'}) \right] \\
&= \pi^{\otimes 2}\left[e^{\theta_0^2 |S \cap S'|} \right] ,
\end{aligned}$$

where the last equation follows from simple computations for normal distribution. Note that $|S \cap S'|$ is distributed as an hypergeometric random variable Z with parameter $(\underline{k}, \underline{k}, n)$. It has been observed in [1] (see the proof of Proposition 20.6 therein) that there exists σ-field $\mathcal{B}$ and a random variable W following a Binomial distribution with parameter $(\underline{k}, \underline{k}/n)$ such that $Z = \mathbb{E}[W|\mathcal{B}]$. Then, it follows from Jensen inequality, that

$$\begin{aligned}
1 + \chi^2(\overline{\mathbb{P}}_{\underline{k}}; \mathbb{P}_{0,0}) &\leq \mathbb{E}[\exp[\theta_0^2 W]] = \left[1 + \frac{\underline{k}}{n}\left(e^{\theta_0^2} - 1 \right) \right]^{\underline{k}} \\
&\leq \exp\left[\frac{\underline{k}^2}{n} e^{\theta_0^2} \right] \leq \exp\left[n^{-1/2+a^2} \right] ,
\end{aligned}$$

by definition of θ_0 and $\underline{k}$. Fixing $a = 1/2$, and coming back to (63), we obtain

$$\mathcal{R}[k', n] \geq c'\sqrt{\log(n)} \left(1 - \sqrt{\exp(n^{-1/4}) - 1} \right) ,$$

which is larger than some $c\sqrt{\log(n)}$ for $n \geq 5$. The result is also valid for $n < 5$ by considering a constant c small enough.

$\square$

We now turn to the case $k \in [\sqrt{n}, n - n^{1/4}]$. As in the previous proof, we shall first reduce the problem to a two point testing problem and then compute an upper bound of the total variation distance. However, the reasoning here is much more involved.

A.1.2. Step 1*: Two point reduction*

Given any two distributions ν_0 and ν_1 on $\mathbb{R}_+$ and any $\theta_0 < \theta_1$ in $\mathbb{R}$, we denote, for $i = 0, 1$ $\mathbf{P}_i = \int \mathbb{P}_{\theta_i,\mu} \nu_i^{\otimes n}(d\mu)$ the mixture distribution where the components of μ are i.i.d. sampled according to the distribution ν_i. We start with the following general reduction lemma.

Lemma A.2 (Reduction). *For any θ_0, θ_1, ν_0 and ν_1, we have*

$$\mathcal{R}[k, n] \geq \frac{|\theta_0 - \theta_1|}{4} \left[1 - \|\mathbf{P}_0 - \mathbf{P}_1\|_{TV} - \sum_{i=0,1} \nu_i^{\otimes n}(\mu \notin \mathcal{M}_k) \right] . \tag{64}$$

Proof of Lemma A.2. For short, we write $\mathcal{B}_k = \{\|\mu\|_0 > k\}$. Starting from the definition of the minimax risk, we have

$$
\begin{aligned}
\mathcal{R}[k, n] &\geq \frac{1}{2} \inf_{\widehat{\theta}} \left(\sup_{\mu \in \mathcal{M}_k} \mathbb{E}_{\theta_0,\mu}[|\widehat{\theta} - \theta_0|] + \sup_{\mu \in \mathcal{M}_k} \mathbb{E}_{\theta_1,\mu}[|\widehat{\theta} - \theta_1|] \right) \\
&\geq \frac{|\theta_0 - \theta_1|}{4} \inf_{\widehat{\theta}} \left(\sup_{\mu \in \mathcal{M}_k} \mathbb{P}_{\theta_0,\mu} \left[\widehat{\theta} > \frac{\theta_0 + \theta_1}{2} \right] + \sup_{\mu \in \mathcal{M}_k} \mathbb{P}_{\theta_1,\mu} \left[\widehat{\theta} \leq \frac{\theta_0 + \theta_1}{2} \right] \right) \\
&\overset{(i)}{\geq} \frac{|\theta_0 - \theta_1|}{4} \inf_{\widehat{\theta}} \left[\mathbf{P}_0 \left(\widehat{\theta} > \frac{\theta_0 + \theta_1}{2} \right) - \nu_0^{\otimes n}(\mathcal{B}_k) + \mathbf{P}_1 \left(\widehat{\theta} \leq \frac{\theta_0 + \theta_1}{2} \right) - \nu_1^{\otimes n}(\mathcal{B}_k) \right] \\
&\geq \frac{|\theta_0 - \theta_1|}{4} \left[1 - \sup_{\mathcal{A}} |\mathbf{P}_0(\mathcal{A}) - \mathbf{P}_1(\mathcal{A})| - \nu_0^{\otimes n}(\mathcal{B}_k) - \nu_1^{\otimes n}(\mathcal{B}_k) \right] \\
&= \frac{|\theta_0 - \theta_1|}{4} \left[1 - \|\mathbf{P}_0 - \mathbf{P}_1\|_{TV} - \nu_0^{\otimes n}(\mathcal{B}_k) - \nu_1^{\otimes n}(\mathcal{B}_k) \right] ,
\end{aligned}
$$

where we use in (i) that both $\mathbf{P}_0 \left[\widehat{\theta} > \frac{\theta_0 + \theta_1}{2} \right] \leq \nu_0^{\otimes n}(\mathcal{B}_k) + \sup_{\mu \in \mathcal{M}_k} \mathbb{P}_{\theta_0,\mu} \left[\widehat{\theta} > \frac{\theta_0 + \theta_1}{2} \right]$ and $\mathbf{P}_1 \left[\widehat{\theta} \leq \frac{\theta_0 + \theta_1}{2} \right] \leq \nu_0^{\otimes n}(\mathcal{B}_k) + \sup_{\mu \in \mathcal{M}_k} \mathbb{P}_{\theta_1,\mu} \left[\widehat{\theta} \leq \frac{\theta_0 + \theta_1}{2} \right]$. $\square$

As a consequence, we will carefully choose the θ_i's and ν_i's in such a way that $|\theta_0 - \theta_1|$ is the largest possible while the total variation distance $\|\mathbf{P}_0 - \mathbf{P}_1\|_{TV}$ is bounded away from one and the measures $\nu_i^{\otimes n}$ are concentrated on $\mathcal{M}_k$. To this end, we consider in the sequel some measures with an extra mass on 0, but in a light fashion so that the convergence rate is preserved.

In the sequel, we define $k_0 = \max(k/2, n - 3(n - k)/2) < k$ and we assume

$$\nu_i(\{0\}) \geq 1 - \frac{k_0}{n} , \text{ for } i = 0, 1. \tag{65}$$

As a consequence, under $\nu_i^{\otimes n}$, $\|\mu\|_0$ is stochastically dominated by a Binomial distribution with parameters $(n, k_0/n)$. By Chebychev inequality, we have

$$\nu_i^{\otimes n}(\mu \notin \mathcal{M}_k) \leq \frac{k_0(n - k_0)}{n \min^2[k/2, (n-k)/2]} \leq \frac{2 \min(k, n-k)}{\min^2[k/2, (n-k)/2]} \leq \frac{8}{k \wedge (n-k)} ,$$

which is smaller than $8n^{-1/4}$ since $k \in [\sqrt{n}, n - n^{1/4}]$. For n large enough, this is smaller than 0.45 and together with Lemma A.2, we obtain

$$\mathcal{R}[k, n] \geq \frac{|\theta_0 - \theta_1|}{4} [0.55 - \|\mathbf{P}_0 - \mathbf{P}_1\|_{TV}] . \tag{66}$$

In view of (66), the challenging part is to control the total variation distance between the mixture distributions $\mathbf{P}_0$ and $\mathbf{P}_1$. Contrary to the situation we dealt with in Lemma A.1, one cannot easily derive a closed form formula for the χ^2 distance between two mixtures. Instead, we shall rely on a general upper bound for mixture of normal distributions. Remember that ϕ denotes the standard normal measure and that, given a real probability measure π, we write $\pi * \phi$ for the corresponding convolution measure.

Lemma A.3. *For two real probability measures π_0 and π_1, assume that $\pi_1(\{0\}) > 0$ and that the supports of π_0 and π_1 are bounded. Then we have*

$$\|(\pi_0 * \phi)^{\otimes n} - (\pi_1 * \phi)^{\otimes n}\|_{TV}^2 \leq \frac{n}{\pi_1(\{0\})} \sum_{\ell \geq 1} \left(\int x^\ell (d\pi_0(x) - d\pi_1(x)) \right)^2 / \ell! .$$

Proof of Lemma A.3. Let f_0 (resp. f_1) the density associated to $\pi_0 * \phi$ (resp. $\pi_1 * \phi$) and let Z_0 (resp. Z_1) be a random variable distributed according to f_0 (resp. f_1). It follows from Le Cam's inequalities and tensorization identities for Hellinger distances [72, Section 2.4] that

$$\|(\pi_0 * \phi)^{\otimes n} - (\pi_1 * \phi)^{\otimes n}\|_{TV}^2 \leq \int \cdots \int \left(\prod_{i=1}^n f_0^{1/2}(y_i) - \prod_{i=1}^n f_1^{1/2}(y_i) \right)^2 dy_1 \ldots dy_n$$

$$\leq n \int \left(f_0^{1/2}(y) - f_1^{1/2}(y) \right)^2 dy . \tag{67}$$

Obviously, we have $f_1(y) \geq \pi_1(\{0\})\phi(y)$, which enforces

$$\int \left(f_0^{1/2}(y) - f_1^{1/2}(y) \right)^2 dy = \int \frac{(f_0(y) - f_1(y))^2}{\left(f_0^{1/2}(y) + f_1^{1/2}(y) \right)^2} dy$$

$$\leq \pi_1^{-1}(\{0\}) \int (f_0(y) - f_1(y))^2 / \phi(y) dy . \tag{68}$$

Next, we use Hermite's polynomials $(H_k(\cdot)/(k!)^{1/2})_{k \geq 0}$ as an Hilbert basis of $L^2(\mathbb{R}, \phi)$ (the space of square integrable function with respect to the normal measure) and the relation $\frac{\phi(y-x)}{\phi(y)} = 1 + \sum_{\ell \geq 1} H_\ell(y) x^\ell / \ell!$ (see, e.g., (1.1) in [34]) to obtain that the rhs of (68) is upper-bounded by

$$\pi_1^{-1}(\{0\}) \int \left(\int \frac{\phi(y-x)}{\phi(y)} d\pi_0(x) - \int \frac{\phi(y-x)}{\phi(y)} d\pi_1(x) \right)^2 \phi(y) dy$$

$$= \pi_1^{-1}(\{0\}) \int \left(\sum_{\ell \geq 1} \frac{H_\ell(y)}{\ell!} \left(\int x^\ell d\pi_0(x) - \int x^\ell d\pi_1(x) \right) \right)^2 \phi(y) dy$$

$$= \pi_1^{-1}(\{0\}) \sum_{\ell \geq 1} \frac{1}{\ell!} \left(\int x^\ell (d\pi_0(x) - d\pi_1(x)) \right)^2 , \tag{69}$$

where we used in the last line the orthonormality of the Hermite polynomials. This concludes the proof. $\square$

Now, if we take $\theta_1 = 0$ and we define $\pi_i = \delta_{\theta_i} * \nu_i$ (where δ_x is the Dirac measure), we have $\mathbf{P}_i = (\pi_i * \phi)^{\otimes n}$ and we are in position to apply Lemma A.3. If we further assume that, for

some integer $m > 2$ and some $M > 0$ the support of π_0 and π_1 is included in $[-M, M]$ and that their m first moments are matching

$$\int x^\ell d\pi_0(x) = \int x^\ell d\pi_1(x), \quad \forall \ell = 1, \ldots, m , \tag{70}$$

we can derive from Lemma A.3 and (65) that

$$
\begin{aligned}
\|\mathbf{P}_0 - \mathbf{P}_1\|_{TV}^2 &\leq \frac{n}{1 - k_0/n} \sum_{\ell > m} \left(\int x^\ell (d\pi_0(x) - d\pi_1(x)) \right)^2 / \ell! \\
&\leq \frac{n^2}{n - k_0} \sum_{\ell > m} \frac{1}{\ell!} \left[\frac{2k_0}{n} M^\ell + |\theta_0|^\ell \right]^2 \\
&\leq \frac{2n^2}{n - k_0} \sum_{\ell > m} \left[\frac{4k_0^2}{n^2} \left(\frac{eM^2}{\ell} \right)^\ell + \left(\frac{e|\theta_0|^2}{\ell} \right)^\ell \right] .
\end{aligned}
$$

Then, if $|\theta_0|$, m, and M are such that $m \geq 2e(M^2 \vee \theta_0^2)$, we obtain

$$\|\mathbf{P}_0 - \mathbf{P}_1\|_{TV}^2 \leq \frac{8k_0^2}{n - k_0} 2^{-m} + 2\frac{n^2}{n - k_0} \left(\frac{e\theta_0^2}{m} \right)^m .$$

If m is large enough this will imply that $\|\mathbf{P}_0 - \mathbf{P}_1\|_{TV} \leq 0.5$. Putting everything together and coming back to (66), we conclude that

$$\mathcal{R}[k, n] \geq \frac{|\theta_0|}{80} , \tag{71}$$

if there exists θ_0, π_0 and π_1 such that for $m_0 = \lceil \log(64k_0^2/(n-k_0))/\log(2) \rceil$ and $M_0 = \sqrt{m_0/2e}$

$$
\begin{cases}
\min(\pi_0(\{\theta_0\}), \pi_1(\{0\})) \geq 1 - \frac{k_0}{n} ; \\
\pi_0 \text{ (resp. } \pi_1\text{) is supported on } [\theta_0, M_0] \text{ (resp. } [0, M_0]) ; \\
\int x^\ell d\pi_0(x) = \int x^\ell d\pi_1(x), \quad \forall \ell = 1, \ldots, m_0 ; \\
\theta_0^2 \leq \frac{m_0}{e} \left[\left(\frac{n-k_0}{16n^2} \right)^{1/m_0} \wedge \frac{1}{2} \right] .
\end{cases}
\tag{72}
$$

The remainder of the proof is devoted to demonstrate the existence of such π_0 and π_1, for $|\theta_0|$ taken as large as possible.

A.1.3. Step 3: Existence of π_0 and π_1

Lemma A.4. *For any positive numbers $M > 0$, $\eta > 0$ and any positive integer m, there exist two probability measures π_0 and π_1 respectively supported on $[-\eta M, M]$ and $[0, M]$, whose m first moments are matching and such that*

$$\min(\pi_0(\{-\eta M\}), \pi_1(\{0\})) \geq \left[1 + \eta m^2 e^{2\sqrt{\eta}m} \right]^{-1} .$$

Proof of Lemma A.4. Consider the space $\mathcal{C}_0$ of continuous functions from $[0, 1]$ to $\mathbb{R}$, endowed with the supremum norm. Let $\mathcal{P}_k$ be the subset $\mathcal{C}_0$ made of polynomials of degree at most m. Consider the linear map $\Lambda : \mathcal{P}_N \to \mathbb{R}$ defined by $\Lambda(P) = P(-\eta) - P(0)$. Then, by the

Hahn-Banach theorem, Λ can be extended into a linear map $\widetilde{\Lambda}$ on the whole subspace $\mathcal{C}_0$ without increasing its operator norm, that is

$$\sup_{\substack{f\in\mathcal{C}_0 \\ \|f\|_\infty\le 1}} |\widetilde{\Lambda}(f)| = \sup_{\substack{f\in\mathcal{P}_m \\ \|f\|_\infty\le 1}} |\Lambda(f)| \ .$$

Since, for $f\in\mathcal{P}_m$, $|f(-\eta)-f(0)| \le \eta\sup_{x\in[-\eta,0]}|f'(x)|$, we derive from Markov's theorem (Lemma D.2 (ii) and (i)) that

$$\sup_{\substack{f\in\mathcal{C}_0 \\ \|f\|_\infty\le 1}} |\widetilde{\Lambda}(f)| \le \frac{\eta m^2}{1+\eta}\cosh[m\,\mathrm{arccosh}(1+2\eta)] \le \eta m^2 e^{2\eta\sqrt{m}} =: a^\star(m,\eta) \ , \tag{73}$$

where we used that $\cosh(x)\le e^x$ and $\mathrm{arccosh}(1+x)\le\sqrt{2x}$. Next, by Riesz representation theorem, there exists a signed measure $\underline{\pi}$ such that $\widetilde{\Lambda}(f) = \int f\,d\pi$ for all $f\in\mathcal{C}_0$. Decomposing $\underline{\pi} = \underline{\pi}^+ - \underline{\pi}^-$ as a difference of positive measure supported on $[0,1]$, if follows from the values of $\Lambda(x^\ell)$ for $\ell=0,\dots,m$, that $|\underline{\pi}^+|=|\underline{\pi}^-|$ and $\int x^\ell(d\underline{\pi}^+(x)-d\underline{\pi}^-(x)) = (-\eta)^\ell$. Besides, the total variation norm $\|\underline{\pi}\|_{TV} = 2|\underline{\pi}^+|$ is upper bounded by $a^\star(m,\eta)$. Let now define the two probability measures

$$\pi_1 = \frac{1}{1+\|\pi\|_{TV}/2}\left(\delta_0 + \underline{\pi}^+(\cdot M)\right) \ ; \qquad \pi_0 = \frac{1}{1+\|\pi\|_{TV}/2}\left(\delta_{-\eta} + \underline{\pi}^-(\cdot M)\right) \ .$$

Obviously, the m first moments of π_0 and π_1 are matching and $\min(\pi_0(\{-\eta M\}),\pi_1(\{0\})) \ge (1+a^\star(m,\eta))^{-1}$. $\qquad\square$

Note that, for any $x,t>0$, if $x\le 0.5\log(1+t)$, we have $xe^x\le t$. Applying Lemma A.4 with m_0, M_0 and any η such that

$$\eta\le\eta_0 = \frac{\log^2\left(1+\sqrt{\frac{k_0}{n-k_0}}\right)}{4m_0^2} \ ,$$

we conclude that $\min(\pi_0(\{-\eta M\}),\pi_1(\{0\})) \ge 1 - k_0/n$. As a consequence, if we choose θ_0 negative with

$$|\theta_0| \le \frac{M_0\log^2\left(1+\sqrt{\frac{k_0}{n-k_0}}\right)}{4m_0^2} \bigwedge \left(\sqrt{\frac{m_0}{2e}}\left(\frac{n-k_0}{16n^2}\right)^{1/(2m_0)}\right) \ ,$$

then, there exist π_0 and π_1 satisfying Conditions (72). From (71), we conclude that

$$\mathcal{R}[k,n] \ge c\left[\frac{\log^2\left(1+\sqrt{\frac{k_0}{n-k_0}}\right)}{m_0^{3/2}} \bigwedge \left(m_0^{1/2}\left(\frac{n-k_0}{16n^2}\right)^{1/(2m_0)}\right)\right] \ .$$

In fact, the second expression in the rhs is always larger (up to a numerical constant) than the first one. Indeed, for $m_0=1$ (which corresponds to $k_0\lesssim\sqrt{n}$), this expression is of order $1/\sqrt{n}$. When $k_0\le n^{1/3}$ and $m_0\ge 2$, this expression is higher than $n^{-1/4}$ which is again higher than the first term. For $k_0\in(n^{1/3},n-1]$, the expression in the rhs is higher than $\sqrt{\log(n)}$

which is again no less than the left expression. In view of the definition of k_0, we have proved that

$$\mathcal{R}[k,n] \geq c \frac{\log^2\left(1 + \sqrt{\frac{k}{n-k}}\right)}{\log^{3/2}\left(1 + \frac{k}{\sqrt{n}}\right)} ,$$

which concludes the proof of Theorem 2.1

A.2. Proof of Theorem 3.1 (OSC)

Note that, for $k \leq 4\sqrt{n}$ and for $n - k \leq 2\sqrt{n}$, the minimax lower bound (25) is a consequence of the lower bound of Theorem 2.1 for the specific gOSC model. As a consequence, we only have to show the result for $k \in (2\sqrt{n}, n - 2\sqrt{n})$. Consider one such k. As in the proof of Theorem 2.1, we define $k_0 = \max(k/2, n - 3(n-k)/2) < k$.

As for the proof of Theorem 2.1, we shall rely on a two point Le Cam's approach but the construction of the distribution is quite different. Let us denote $\mathbb{P}_x$ for the distribution $\mathcal{N}(x,1)$ and ϕ_x for its usual density. Let $\theta > 0$ be a positive number whose value will be fixed later and define $\epsilon = k_0/n$. We shall introduce below two probability measures μ_0 and μ_1 that stochastically dominate $\mathbb{P}_{-\theta}$ and $\mathbb{P}_\theta$. Consider the mixture distribution $\vartheta_0 = (1-\epsilon)\mathbb{P}_{-\theta} + \epsilon\mu_0$ and $\vartheta_1 = (1-\epsilon)\mathbb{P}_\theta + \epsilon\mu_1$ and $\mathbf{P}_0 = \vartheta_0^{\otimes n}$ and $\mathbf{P}_1 = \vartheta_1^{\otimes n}$. Under $\mathbf{P}_0$, all variables Y_i are sampled independently and with probability $(1-\epsilon)$ follow the normal distribution $\mathbb{P}_{-\theta}$ and with probability ϵ follows the stochastically larger distribution μ_0. Let Z be a binomial variable with parameters (n, ϵ). Under $\mathbf{P}_0$, Z of the observations have been sampled according to μ_0 and the $n - Z$ remaining observations have been sampled according to $\mathbb{P}_{-\theta}$. Thus, up to an event of probability $\mathbb{P}[\mathcal{B}(n,\epsilon) \leq k]$, $\mathbf{P}_0$ is a mixture of distributions in $\overline{\mathcal{M}}_k$ whose corresponding functional is $-\theta$. The measure $\mathbf{P}_1$ satisfies the same property with $-\theta$ replaced by θ.

Arguing as in Lemma A.2, we therefore obtain

$$\overline{\mathcal{R}}[k,n] \geq \frac{\theta}{2}\left[1 - \|\mathbf{P}_0 - \mathbf{P}_1\|_{TV} - 2\mathbb{P}[Z > k]\right] .$$

The probability $\epsilon = k_0/n$ has been chosen small enough that $\mathbb{P}[Z > k]$ is vanishing for n large enough (see the proof of Theorem 2.1, Step 1) so that for such n, we obtain $\overline{\mathcal{R}}[k,n] \geq \frac{\theta}{2}[0.55 - \|\mathbf{P}_0 - \mathbf{P}_1\|_{TV}]$ (see (66)) and thus

$$\overline{\mathcal{R}}[k,n] \geq \frac{\theta}{40} \quad \text{if} \quad \|\mathbf{P}_0 - \mathbf{P}_1\|_{TV} \leq 0.5 . \tag{74}$$

In the sequel, we fix

$$\theta = \frac{\log\left(\frac{1}{1-\epsilon}\right)}{8\sqrt{\log(n\epsilon^2)}} \tag{75}$$

and we will shall build two measure μ_0 and μ_1 that enforce $\|\mathbf{P}_0 - \mathbf{P}_1\|_{TV} \leq 0.5$. In view of (25) and (74), this will conclude the proof.

Define

$$t_\theta = \frac{1}{2\theta}\log\left(\frac{1}{1-\epsilon}\right) . \tag{76}$$

We shall pick μ_0 and μ_1 in such a way that the densities of ϑ_0 and ϑ_1 are matching on the widest interval possible. Define μ_0 and μ_1 by their respective densities f_0 and f_1

$$f_0(x) = g_0(x) + h(x) , \qquad f_1(x) = g_1(x) + h(x) ,$$

where

$$g_0(x) = \frac{1-\epsilon}{\epsilon}[\phi_\theta(x) - \phi_{-\theta}(x)] \text{ if } x \in (0, t_\theta] \quad \text{and } g_0(x) = 0 \text{ else}$$

$$g_1(x) = \frac{1-\epsilon}{\epsilon}[\phi_{-\theta}(x) - \phi_\theta(x)] \text{ if } x \in [-t_\theta; 0) \quad \text{and } g_1(x) = 0 \text{ else} ,$$

and $h(x) = \mathbf{1}_{x>u} a\phi_{2\theta}(x)$ where $u > \max(n+\theta, t_\theta)$ and $a \geq 1$ are taken such that $\int (g_0(x) + h(x))dx = 1$. To ensure the existence of h (that is, of such a and u), we need to prove that $\int g_0(x)dx = \int g_1(x)dx < 1$. By definition (76) of t_θ, we have $\phi_\theta(x)/\phi_{-\theta}(x) < (1-\epsilon)^{-1}$ for all $x \in (0, t_\theta)$. This implies that $g_0(x) < \phi_{-\theta}(x)$ for all $x \in (0, t_\theta)$ and $g_1(x) < \phi_\theta(x)$ for all $x \in (-t_\theta, 0)$, which entails $\int g_0(x)dx = \int g_1(x)dx < 1$.

Also, the two measures μ_0 and μ_1 respectively satisfy $\mu_0 \succeq \mathbb{P}_{-\theta}$ and $\mu_1 \succeq \mathbb{P}_\theta$. Let us only prove the second inequality, the first one being simpler. Consider any $t \in \mathbb{R}$. Then, it follows from the definition of f_1 that

$$\int_{-\infty}^t f_1(x)dx = \begin{cases} 0 & \text{if } t < -t_\theta ; \\ \int_{-t_\theta}^t g_1(x)dx & \text{if } t \in [-t_\theta, 0] ; \\ \int_{-t_\theta}^0 g_1(x)dx & \text{if } t \in (0, u] ; \\ 1 - a\int_t^\infty \phi_{2\theta}(x)dx & \text{if } t > u . \end{cases}$$

Since $g_1(x) < \phi_\theta(x)$ for all $x \in (-t_\theta, 0)$, we readily obtain $\int_{-\infty}^t f_1(x)dx < \int_{-\infty}^t \phi_\theta(x)dx$ for all $t \leq u$. For $t > u$, we have $\int_t^\infty f_1(x)dx = a\int_t^\infty \phi_{2\theta}(x)dx \geq \int_t^\infty \phi_\theta(x)dx$, which implies $\mu_1 \succeq \mathbb{P}_\theta$.

It remains to prove that $\|\mathbf{P}_0 - \mathbf{P}_1\|_{TV} \leq 0.5$. Denote $H^2(\mathbf{P}_0; \mathbf{P}_1)$ the square Hellinger distance. As in the previous proof, it follows from Le Cam's inequalities and tensorization identities for Hellinger distances [72, Section 2.4] that

$$\|\mathbf{P}_0 - \mathbf{P}_1\|_{TV}^2 \leq H^2(\mathbf{P}_0; \mathbf{P}_1) = 2\left[1 - \left(1 - \frac{H^2(\vartheta_0; \vartheta_1)}{2}\right)^n\right]$$
$$\leq nH^2(\vartheta_0; \vartheta_1) .$$

As a consequence, for n large enough, one has $\|\mathbf{P}_0 - \mathbf{P}_1\|_{TV} \leq 0.5$ as long as $H^2(\vartheta_0; \vartheta_1) \leq (2n)^{-1}$. It remains to prove this last inequality. Write $v_0 = (1-\epsilon)\phi_{-\theta} + \epsilon(g_0 + h)$ the density of ϑ_0 and v_1 the density of ϑ_1. g_0 and g_1 have been chosen in such a way that v_0 and v_1 are matching on the interval $[-t_\theta; t_\theta]$.

$$\frac{1}{2}H^2(\vartheta_0; \vartheta_1) = 1 - \int \sqrt{v_0(x)v_1(x)}dx = \int (v_0(x) - \sqrt{v_0 v_1}(x))dx$$
$$= (1-\epsilon)\left[\int_{(-\infty;-t_\theta]\cup[t_\theta,\infty)} \left[\phi_{-\theta}(x) - \sqrt{\phi_{-\theta}(x)\phi_\theta(x)}\right]dx\right]$$
$$+ \int_{(u,\infty)} \left[v_0(x) - \sqrt{v_1 v_0}(x) - (1-\epsilon)\left[\phi_{-\theta}(x) - \sqrt{\phi_{-\theta}(x)\phi_\theta(x)}\right]\right]dx$$
$$\leq (1-\epsilon)\left[\int_{(-\infty;-t_\theta]\cup[t_\theta,\infty)} \left[\phi_{-\theta}(x) - \sqrt{\phi_{-\theta}(x)\phi_\theta(x)}\right]dx\right] + e^{-n^2/2} ,$$

where we used $v_0(x) \le v_1(x)$ for $x > u$. This leads us to

$$H^2(\vartheta_0; \vartheta_1) \le 2e^{-n^2/2} + 2(1 - \epsilon)\left[\overline{\Phi}(t_\theta + \theta) + \overline{\Phi}(t_\theta - \theta) - e^{-\theta^2/2}2\overline{\Phi}(t_\theta)\right] . \tag{77}$$

Since $k \ge 4\sqrt{n}$, we have $16\log(n\epsilon^2) \ge \log(1/(1 - \epsilon))$ which entails $\theta \le t_\theta/2$. This leads us to

$$
\begin{aligned}
\overline{\Phi}(t_\theta + \theta) + \overline{\Phi}(t_\theta - \theta) - 2\overline{\Phi}(t_\theta) &\le& \theta^2 \sup_{x \in [t_\theta - \theta; t_\theta + \theta]} |\phi'(x)| \\
&\le& \frac{\theta^2}{\sqrt{2\pi}}\frac{3t_\theta}{2}e^{-t_\theta^2/8} \\
&\le& \frac{3}{32\sqrt{2\pi}\log(n\epsilon^2)(n\epsilon^2)^2}\log^2\left(\frac{1}{1 - \epsilon}\right) \\
&\le& \frac{1}{8\sqrt{2\pi}n^2\epsilon^2(1 - \epsilon)^2} ,
\end{aligned}
$$

where we used the definitions (75) and (76) of θ and t_θ and $\log(n\epsilon^2) \ge 1$. Similarly, one has

$$\overline{\Phi}(t_\theta)(1 - e^{-\theta^2/2}) \le e^{-t_\theta^2/2}\frac{\theta^2}{2} \le \frac{1}{128\log(n\epsilon^2)(n\epsilon^2)^2}\log^2\left(\frac{1}{1 - \epsilon}\right) \le \frac{1}{128n^2\epsilon^2(1 - \epsilon)^2} .$$

Coming back to (77), we conclude that

$$H^2(\vartheta_0; \vartheta_1) \le 2e^{-n^2/2} + \frac{1}{8n^2\epsilon^2(1 - \epsilon)^2} \le 2e^{-n^2/2} + \frac{1}{8n(1 - cn^{-1/2})} ,$$

since $\epsilon \ge 2n^{-1/2}$ and $1 - \epsilon \ge 3n^{-1/2}$. For n large enough, we obtain $H^2(\vartheta_0; \vartheta_1) \le 1/(2n)$, which concludes the proof.

A.3. Proof of Theorem 4.1 (OSC)

This proof proceeds from the same approach as that of Theorem 3.1 but the construction of the prior distributions are quite different. For any fixed numerical constant $c_0 > 0$ and any $k \le c_0\sqrt{n}$, the lower bound in the theorem is parametric and is easily proved in a model without contamination. We assume henceforth that $k > c_0\sqrt{n}$ and we will fix the value of c_0 at the end of the proof. Also for $n - k \le 2\sqrt{n}$, the lower bound in Theorem 4.1 is of the order of a constant, so that we only have to prove the result for $n - k > 2\sqrt{n}$, so we also assume henceforth that $n - k \ge 2\sqrt{n}$.

As in the previous proof, we define $k_0 = \max(k/2, n - 3(n - k)/2) < k$. Let us denote $\mathbb{P}_{0,y}$ for the distribution $\mathcal{N}(0, y^2)$ and $\phi_{0,y}$ for its usual density. Let $\sigma > 1$ be a positive quantity that will be fixed later and let $\epsilon = k_0/n$. We shall introduce below two probability measures μ_0 and μ_1 that stochastically dominate $\mathbb{P}_{0,1}$ and $\mathbb{P}_{0,\sigma}$. Consider the mixture distribution $\vartheta_0 = (1 - \epsilon)\mathbb{P}_{0,1} + \epsilon\mu_0$ and $\vartheta_1 = (1 - \epsilon)\mathbb{P}_{0,\sigma} + \epsilon\mu_1$ and $\mathbf{P}_0 = \vartheta_0^{\otimes n}$ and $\mathbf{P}_1 = \vartheta_1^{\otimes n}$. Arguing as in the proof of Theorem 3.1 (and Lemma A.2), we obtain that

$$\overline{\mathcal{R}}_v[k, n] \ge \frac{\sigma - 1}{80} \quad \text{if} \quad \|\mathbf{P}_0 - \mathbf{P}_1\|_{TV} \le 0.5 . \tag{78}$$

In the sequel, we fix

$$\sigma = 1 + \frac{\log\left(\frac{1}{1 - \epsilon}\right)}{6\log(n\epsilon^2)} \wedge 1 . \tag{79}$$

and we will shall build two measure μ_0 and μ_1 that enforce $\|\mathbf{P}_0 - \mathbf{P}_1\|_{TV} \le 0.5$. In view of the two previous inequalities this will conclude the proof.

We shall pick μ_0 and μ_1 in such a way that the densities of ϑ_0 and ϑ_1 are matching on the widest interval possible. Denote μ_0 and μ_1 by their respective densities f_0 and f_1. In principle, we would like to take $f_0 = (1-\epsilon)/\epsilon[\phi_{0,\sigma} - \phi_{0,1}]_+$ and $f_1 = (1-\epsilon)/\epsilon[\phi_{0,1} - \phi_{0,\sigma}]_+$ as this would enforce $\|\mathbf{P}_0 - \mathbf{P}_1\|_{TV} = 0$. Unfortunately, such a choice is not possible as the corresponding measure μ_0 would not be a probability measure (and would not either dominate $\mathbb{P}_{0,1}$). The actual construction is a bit more involved. First, define

$$v_\sigma = \sqrt{\frac{2\sigma^2}{\sigma^2 - 1}\log(\sigma)}, \quad w_\sigma = \sqrt{\frac{2\sigma^2}{\sigma^2 - 1}\log\left(\frac{\sigma}{1-\epsilon}\right)} . \tag{80}$$

We have $\phi_{0,\sigma}(t) \ge \phi_{0,1}(t)$ if and only if $|t| \ge v_\sigma$ and $(1-\epsilon)\phi_{0,\sigma}(t) \ge \phi_{0,1}(t)$ for all $|t| \ge w_\sigma$. This implies

$$\int_0^{v_\sigma} \phi_{0,1}(x) - \phi_{0,\sigma}(x)dx > \int_{v_\sigma}^{w_\sigma} \phi_{0,\sigma}(x) - \phi_{0,1}(x)dx .$$

Thus, we can define $u_\sigma \in (0, v_\sigma)$ in such a way that

$$\int_{u_\sigma}^{v_\sigma} \phi_{0,1}(x) - \phi_{0,\sigma}(x)dx = \int_{v_\sigma}^{w_\sigma} \phi_{0,\sigma}(x) - \phi_{0,1}(x)dx . \tag{81}$$

Then, we take

$$f_0(x) = g_0(x) + h(x) , \qquad f_1(x) = g_1(x) + h(x) ,$$

where

$$g_0(x) = \frac{1-\epsilon}{\epsilon}[\phi_{0,\sigma}(x) - \phi_{0,1}(x)] \text{ if } |x| \in [v_\sigma, w_\sigma] \quad \text{and } g_0(x) = 0 \quad \text{else.}$$

$$g_1(x) = \frac{1-\epsilon}{\epsilon}[\phi_{0,1}(x) - \phi_{0,\sigma}(x)] \text{ if } |x| \in [u_\sigma, v_\sigma] \quad \text{and } g_1(x) = 0 \quad \text{else.}$$

By definition of v_σ and w_σ, g_0 is nonnegative and is smaller or equal to $\phi_{0,1}$. As a consequence, $\int g_0(x) < \int \phi_{0,1}(x)dx \le 1$. Besides, u_σ has been chosen in such a way that $\int g_0(x) = \int g_1(x)dx$. Finally, we define $h(x) = \mathbf{1}_{x>s}a\phi_{0,\sigma}(x)$ where $s > n\sigma + w_\sigma$ and $a \ge 1$ are taken such that $\int(g_0(x) + h(x))dx = 1$.

Since we assume that $\sigma(1-\epsilon) < 1$, observe that $(1-\epsilon)\phi_{0,1}(t) \le \phi_{0,\sigma}(t)$ for all $t \in \mathbb{R}$, which in turn implies that $g_1 \le \phi_{0,\sigma}$. Since $g_0 \le \phi_{0,1}$, it follows that the two measures μ_0 and μ_1 respectively satisfy $\mu_0 \succeq \mathbb{P}_{0,1}$ and $\mu_1 \succeq \mathbb{P}_{0,\sigma}$.

It remains to prove that $\|\mathbf{P}_0 - \mathbf{P}_1\|_{TV} \le 0.5$. As in the previous proof, we have $\|\mathbf{P}_0 - \mathbf{P}_1\|_{TV}^2 \le nH^2(\vartheta_0; \vartheta_1)$ and we only have to prove that $H^2(\vartheta_0; \vartheta_1) \le (2n)^{-1}$ for n large enough.

To compute this Hellinger distance, we first observe that the densities v_0 and v_1 associated to ϑ_0 and ϑ_1 are matching in $[-w_\sigma, u_\sigma] \cup [u_\sigma, w_\sigma]$. Together with the definition of f_0 and f_1 this leads us to

$$\begin{aligned}
\frac{1}{2}H^2(\vartheta_0; \vartheta_1) &= 1 - \int \sqrt{v_0(x)v_1(x)}dx = \int (v_0(x) - \sqrt{v_0 v_1}(x))dx \\
&= 2(1-\epsilon)\int_{[0,u_\sigma]\cup[w_\sigma,\infty)}\left[\phi_{0,1}(x) - \sqrt{\phi_{0,1}(x)\phi_{0,\sigma}(x)}\right]dx \\
&\quad + \int_{(s,\infty)}\left[v_0(x) - \sqrt{v_1 v_0}(x) + (1-\epsilon)\left(\sqrt{\phi_{0,1}(x)\phi_{0,\sigma}(x)} - \phi_{0,1}(x)\right)\right]dx .
\end{aligned}$$

Since $v_0(x) \leq v_1(x)$ for $x > s$, the last term is less or equal to $\int_s^\infty \phi_{0,\sigma}(x)dx \leq e^{n^2/2}$. It follows from (81) that

$$\int_0^{u_\sigma} (\phi_{0,1}(x) - \phi_{0,\sigma})(x)dx = \int_{w_\sigma}^\infty (\phi_{0,\sigma} - \phi_{0,1})(x)dx .$$

This leads us to

$$\frac{1}{2}H^2(\vartheta_0; \vartheta_1) \leq e^{-n^2/2} + (1-\epsilon) \int_{[0,u_\sigma] \cup [w_\sigma,\infty)} \phi_{0,1}(x) + \phi_{0,\sigma}(x) - 2\sqrt{\phi_{0,1}(x)\phi_{0,\sigma}(x)}dx . \quad (82)$$

For $\delta \in [-1/2, 1]$, Taylor's formula leads to $(2 + \delta - 2\sqrt{1+\delta}) \leq \delta^2/\sqrt{2}$. As a consequence, for any x such that $\phi_{0,\sigma}(x)/\phi_{0,1}(x) \in [1/2, 2]$, we have

$$\phi_{0,1}(x) + \phi_{0,\sigma}(x) - 2\sqrt{\phi_{0,1}(x)\phi_{0,\sigma}(x)} \leq \frac{(\phi_{0,1}(x) - \phi_{0,\sigma}(x))^2}{\sqrt{2}\phi_{0,1}(x)} = \frac{\phi_{0,1}(x)}{\sqrt{2}} \left[\frac{1}{\sigma} \exp\left(\frac{x^2(\sigma^2 - 1)}{2\sigma^2} \right) - 1 \right]^2 . \quad (83)$$

Define $z_\sigma = \sqrt{\frac{2\sigma^2}{\sigma^2-1}[\log(2\sigma) \wedge 1]}$. For any $|x| \leq z_\sigma$, we have $\phi_{0,\sigma}(x) \leq 2\phi_{0,1}(x)$. From the previous inequality, we derive that, for $|x| \in (w_\sigma; w_\sigma \vee z_\sigma)$,

$$
\begin{aligned}
\phi_{0,1}(x) + \phi_{0,\sigma}(x) - 2\sqrt{\phi_{0,1}(x)\phi_{0,\sigma}(x)} &\leq \frac{\phi_{0,1}(x)}{\sqrt{2}} \left[\frac{1}{\sigma} \exp\left(\frac{x^2(\sigma^2-1)}{2\sigma^2} \right) - 1 \right]^2 \\
&\leq \frac{\phi_{0,1}(x)}{\sqrt{2}} \left[2\frac{x^2(\sigma^2-1)}{2\sigma^2} + \frac{1}{\sigma} - 1 \right]_+^2 \\
&\leq 3\phi_{0,1}(x)x^4(\sigma-1)^2 .
\end{aligned}
$$

Since $\sigma \leq 2$, we have $\phi_{0,\sigma}(x) \geq \phi_{0,1}(x)/2$ for all x. As a consequence of (83), we obtain that, for $|x| \leq u_\sigma$,

$$\phi_{0,1}(x) + \phi_{0,\sigma}(x) - 2\sqrt{\phi_{0,1}(x)\phi_{0,\sigma}(x)} \leq \frac{(\phi_{0,1}(x) - \phi_{0,\sigma}(x))^2}{\sqrt{2}\phi_{0,1}(x)} \leq [\phi_{0,1}(x) - \phi_{0,\sigma}(x)]\frac{\sigma-1}{\sqrt{2}\sigma} .$$

Coming back to (82), we obtain

$$
\begin{aligned}
&\frac{\frac{1}{2}H^2(\vartheta_0; \vartheta_1) - e^{-n^2/2}}{1-\epsilon} \\
&\leq \frac{\sigma-1}{\sqrt{2}\sigma} \int_0^{u_\sigma} (\phi_{0,1}(x) - \phi_{0,\sigma}(x))dx + 3(\sigma-1)^2 \int_{w_\sigma}^{w_\sigma \vee z_\sigma} x^4\phi_{0,1}(x)dx + \phi_{0,1}\left(\frac{w_\sigma \vee z_\sigma}{\sigma}\right) \\
&\leq \frac{\sigma-1}{\sqrt{2}\sigma} \int_{w_\sigma}^\infty (\phi_{0,\sigma}(x) - \phi_{0,1}(x))dx + 3(\sigma-1)^2 \int_{w_\sigma}^\infty x^4\phi_{0,1}(x)dx + \phi_{0,1}\left(\frac{w_\sigma \vee z_\sigma}{\sigma}\right) \\
&\leq \frac{\sigma-1}{\sqrt{2}\sigma} \int_{w_\sigma/\sigma}^{w_\sigma} \phi_{0,1}(x)dx + 3(\sigma-1)^2[w_\sigma^3 + 6w_\sigma]\phi_{0,1}(w_\sigma) + \phi_{0,1}\left(\frac{w_\sigma \vee z_\sigma}{\sigma}\right) \\
&\leq w_\sigma \frac{(\sigma-1)^2}{\sqrt{2}\sigma^2}\phi_{0,1}\left(\frac{w_\sigma}{\sigma}\right) + 3(\sigma-1)^2[w_\sigma^3 + 6w_\sigma]\phi_{0,1}(w_\sigma) + \phi_{0,1}\left(\frac{w_\sigma \vee z_\sigma}{\sigma}\right) \\
&\leq (\sigma-1)^2 \left[3w_\sigma^3 + 7w_\sigma\right] + \phi_{0,1}\left(\frac{w_\sigma \vee z_\sigma}{\sigma}\right) ,
\end{aligned}
\quad (84)
$$

where we used the definition (81) of u_σ in the third line. To conclude, we come back to the definitions of w_σ, z_σ and σ

$$
\begin{aligned}
\frac{w_\sigma^2}{2\sigma^2} &= \frac{1}{\sigma^2 - 1} \log\left(\frac{\sigma}{1-\epsilon}\right) \geq 2\log(n\epsilon^2) ; \\
\frac{z_\sigma^2}{2\sigma^2} &= \frac{\log(2\sigma) \wedge 1}{\sigma^2 - 1} \geq \frac{\log(2)}{3(\sigma - 1)} = \frac{2\log(2)\log(n\epsilon^2)}{\log\left(\frac{1}{1-\epsilon}\right)} ; \\
\sigma - 1 &\leq \frac{\log(\frac{1}{1-\epsilon})}{\log(n\epsilon^2)} ; \\
w_\sigma^2 &\leq \frac{2}{\sigma - 1}\log\left(\frac{\sigma}{1-\epsilon}\right) \leq 2 + \frac{2}{\sigma - 1}\log\left(\frac{1}{1-\epsilon}\right) \leq 2 + [12\log(n\epsilon^2)] \vee [\log(\frac{1}{1-\epsilon})] .
\end{aligned}
$$

This implies that

$$
\phi_{0,1}\left(\frac{z_\sigma \vee w_\sigma}{\sigma}\right) \leq \exp\left[-2\log(n\epsilon^2)\left(1 \vee \frac{\log(2)}{\log\left(\frac{1}{1-\epsilon}\right)}\right)\right] ,
$$

which is less than $16n^{-2}$ since the maximum over $\epsilon \in [e^2/\sqrt{n}, 1 - 1/\sqrt{n}]$ is achieved at $\epsilon = 1/2$.

Coming back to (84), we conclude that

$$
\begin{aligned}
\frac{1}{2}H^2(\vartheta_0; \vartheta_1) &\leq e^{-n^2/2} + \frac{c'}{n^2} + c(1-\epsilon)\frac{\log^2\left(\frac{1}{1-\epsilon}\right)\log(n\epsilon^2) + \log^5\left(\frac{1}{1-\epsilon}\right)}{n^2\epsilon^4} \\
&\leq e^{-n^2/2} + \frac{c'}{n^2} + \begin{cases} \frac{c}{n}\frac{\log(n\epsilon^2)}{n\epsilon^2} & \text{if } \epsilon \leq 1/2 ; \\ \frac{c}{n^2} & \text{if } \epsilon > 1/2 . \end{cases}
\end{aligned}
$$

This last expression is less than $1/(2n)$ as soon as long as n is large enough and $n\epsilon^2$ is large enough, which is ensured if the constant c_0 introduced at the beginning of the proof is large enough. This concludes the proof.

Appendix B: Proofs of upper bounds

B.1. Proofs for the preliminary estimators : Propositions 2.2 and 2.3 (OSC)

Proof of Proposition 2.2 . We prove this result in the OSC model. Consider any $\mu \in \overline{\mathcal{M}_k}$. As argued in Section 3, we have the stochastic bounds

$$
\xi_{(\lceil n/2 \rceil)} \preceq \widehat{\theta}_{\text{med}} - \theta \preceq \xi_{(\lceil n/2 \rceil : n-k)} , \tag{85}
$$

where $\xi = (\xi_1, \ldots, \xi_n)$ is a standard Gaussian vector. Hence, we only have to control the deviations of $\xi_{(\lceil n/2 \rceil)}$ and of $\xi_{(\lceil n/2 \rceil : n-k)}$. Then, we apply Lemma D.7 with $q = \lceil n/2 \rceil$ to obtain

$$
\mathbb{P}_{\theta,\pi}\left[\widehat{\theta}_{\text{med}} - \theta + \overline{\Phi}^{-1}(\tfrac{\lceil n/2 \rceil}{n}) \leq -3\frac{\sqrt{(n+1)x}}{n}\right] \leq e^{-x} ,
$$

for all $x \leq cn$ (where c is some universal constant). As for the right deviations of $\widehat{\theta}_{med}$, we apply the deviation inequality (125) to $\xi_{(\lceil n/2 \rceil : n-k)}$ as $k \leq n/10$. This leads us to

$$
\mathbb{P}_{\theta,\pi}\left[\widehat{\theta}_{\text{med}} - \theta + \overline{\Phi}^{-1}(\tfrac{\lceil n/2 \rceil}{n-k}) \geq 3\frac{\sqrt{(n+1)x}}{n-k}\right] \leq e^{-x} ,
$$

for all $x \leq cn$. Then, Lemma D.5 ensures that

$$|\overline{\Phi}^{-1}(\tfrac{\lceil n/2 \rceil}{n-k})| = |\overline{\Phi}^{-1}(\tfrac{\lceil n/2 \rceil}{n-k}) - \overline{\Phi}^{-1}(1/2)| \leq \frac{3(k+1)}{2(n-k)} \ .$$

Similarly, we have $|\overline{\Phi}^{-1}(\tfrac{\lceil n/2 \rceil}{n})| \leq 3/(2n)$. We have proved the first result.

Let us now turn to the moment bound. Starting from (85), we get

$$
\begin{aligned}
\mathbb{E}_{\theta,\pi}\big[|\widehat{\theta}_{\mathrm{med}} - \theta|\big] &\leq \mathbb{E}_{\theta,\pi}\big[(\widehat{\theta}_{\mathrm{med}} - \theta)_+\big] + \mathbb{E}_{\theta,\pi}\big[(\theta - \widehat{\theta}_{\mathrm{med}})_+\big] \\
&\leq \mathbb{E}\big[(\xi_{(\lceil n/2 \rceil:n-k)})_+\big] + \mathbb{E}\big[(\xi_{(\lceil n/2 \rceil)})_-\big] \ .
\end{aligned}
$$

We have proved above deviation inequalities for these two random variables for probabilities larger than $e^{-c'n}$ (where c' is some universal constant). Write $Z_1 = \xi_{(\lceil n/2 \rceil:n-k)} + \overline{\Phi}^{-1}(\tfrac{\lceil n/2 \rceil}{n-k})$ and $Z_2 = \xi_{(\lceil n/2 \rceil)} + \overline{\Phi}^{-1}(\tfrac{\lceil n/2 \rceil}{n})$. We deduce from the previous deviation inequalities that

$$\mathbb{E}\big[(Z_1)_+ \mathbb{1}_{Z_1 \leq c'}\big] \leq \frac{3\sqrt{\pi(n+1)}}{\sqrt{2}(n-k)} \ , \quad \mathbb{E}\big[(Z_2)_- \mathbb{1}_{Z_2 \geq -c'}\big] \leq \frac{3\sqrt{\pi(n+1)}}{\sqrt{2}n} \ .$$

It remains to control $\mathbb{E}\big[(Z_1)_+ \mathbb{1}_{Z_1 > c'}\big]$ and $\mathbb{E}\big[(Z_1)_+ \mathbb{1}_{Z_1 \leq -c'}\big]$. Since these two random variables are Lipschitz functions of ξ, they follow the Gaussian concentration theorem. In particular, their variance is less than 1. Also, Z_1 and Z_2 concentrate well around their medians and around 0 (previous deviation inequality). Thus, their first moments is smaller than a constant. Cauchy-Schwarz inequality then yields

$$\mathbb{E}\big[(Z_1)_+ \mathbb{1}_{Z_1 > c'}\big] \leq \mathbb{P}^{1/2}[Z_1 \geq c'] \mathbb{E}^{1/2}[(Z_1)_+^2] \leq c e^{-c''n} \ .$$

Similarly, we have $\mathbb{E}\big[(Z_2)_- \mathbb{1}_{Z_2 \leq -c'}\big] \leq c e^{-c''n}$. This concludes the proof. $\qquad\square$

Proof of Proposition 2.3. For any θ and any $\pi \in \overline{\mathcal{M}}_{n-1}$, we have

$$\xi_{(1)} \preceq Y_{(1)} - \theta \preceq \xi_1 \ ,$$

where $\xi = (\xi_1, \dots, \xi_n)$ is a standard normal vector. Using the Gaussian tail bound, we derive that

$$\mathbb{P}_{\theta,\pi}\left[(\widehat{\theta}_{\min} - \theta) \in [\overline{\Phi}^{-1}(n^{-1}) - \overline{\Phi}^{-1}(n^{-2}), 2\overline{\Phi}^{-1}(n^{-1})]\right] \geq 1 - \frac{2}{n} \ .$$

By Lemma D.4, $2\overline{\Phi}^{-1}(n^{-1}) \leq 2\sqrt{2\log(n)}$, whereas Lemma D.5 ensures that $\overline{\Phi}^{-1}(n^{-2}) - \overline{\Phi}^{-1}(n^{-1}) \leq \sqrt{\log(n)/2} + O(1)$. Thus, the desired deviation bound holds for n large enough. Turning to the moment bound, we have the following decomposition

$$
\begin{aligned}
\mathbb{E}_{\theta,\pi}\left[|\widehat{\theta}_{\min} - \theta|\right] &\leq \mathbb{E}\left[(\widehat{\theta}_{\min} - \theta)_+ + (\theta - \widehat{\theta}_{\min})_+\right] \\
&\leq 2\sqrt{2\log(n)} + \mathbb{E}\left[(\xi_1 + \overline{\Phi}^{-1}(1/n))_+ \mathbb{1}_{\xi_1 + \overline{\Phi}^{-1}(1/n) \geq 2\sqrt{2\log(n)}}\right] \\
&\quad + \mathbb{E}\left[(\xi_{(1)} + \overline{\Phi}^{-1}(1/n))_- \mathbb{1}_{\xi_{(1)} + \overline{\Phi}^{-1}(1/n) \leq -2\sqrt{2\log(n)}}\right] \ .
\end{aligned}
$$

Let us focus on the first expectation, the second expectation being handled similarly. As a consequence of Cauchy-Schwarz inequality, we have

$$\mathbb{E}\left[(\xi_1 + \overline{\Phi}^{-1}(1/n))_+ \mathbb{1}_{\xi_1 + \overline{\Phi}^{-1}(1/n) \geq 2\sqrt{2\log(n)}}\right] \leq c\sqrt{\log(n)}\mathbb{P}^{1/2}[\xi_1 + \overline{\Phi}^{-1}(1/n) \geq 2\sqrt{2\log(n)}]$$

$$\leq c\sqrt{\frac{\log(n)}{n}},$$

where we used the above deviation inequality. We obtain

$$\mathbb{E}_{\theta,\pi}\left[|\widehat{\theta}_{\min} - \theta|\right] \leq 2\sqrt{2\log(n)} + c\sqrt{\frac{\log(n)}{n}},$$

which concludes the proof.

$\square$

B.2. Range for θ: analysis of $\widehat{\theta}_{\mathrm{up}}$ and $\widehat{\theta}_{\mathrm{low},q_k}$ (gOSC)

As a preliminary step for the proof of Theorem 2.4, we control the deviations of the rough estimators $\widehat{\theta}_{\mathrm{up}}$ and $\widehat{\theta}_{\mathrm{low},q_k}$. Recall that the tuning parameter q_k is defined in Theorem 2.4.

Lemma B.1 (Control of $\widehat{\theta}_{\mathrm{low},q_k}$). *There exist an universal constants $n_0 \geq 1$ and $c > 0$ such that the following holds for all $n \geq n_0$. For all $k \in [1, n]$ such that $q_k \leq \frac{3}{10a}\log n$, $\mu \in \mathcal{M}_k$ and all $\theta \in \mathbb{R}$, the estimator $\widehat{\theta}_{\mathrm{low},q_k}$ satisfies*

$$\mathbb{P}_{\theta,\mu}\left[\widehat{\theta}_{\mathrm{low},q_k} \geq \theta\right] \leq \frac{1}{n} ; \qquad \mathbb{E}_{\theta,\mu}\left[(\theta - \widehat{\theta}_{\mathrm{low},q_k})\mathbb{1}_{\widehat{\theta}_{\mathrm{low},q_k} \leq \theta - 2\overline{v}}\right] \leq \frac{c}{\sqrt{n}} .$$

Proof of Lemma B.1. Recall $q_k = \lfloor \frac{1}{a}\log\left(\frac{k}{\sqrt{n}}\right)\rfloor_{\mathrm{even}} \wedge q_{\max}$ so that $k \leq e^{2a}n^{4/5}$. By definition, we have $\widehat{\theta}_{\mathrm{low},q_k} = \widehat{\theta}_{\mathrm{med}} - \overline{v}$, which, thanks to Proposition 2.2, implies that

$$\mathbb{P}_{\theta,\mu}\left(\widehat{\theta}_{\mathrm{low},q_k} \geq \theta\right) \leq \mathbb{P}_{\theta,\mu}(\widehat{\theta}_{\mathrm{med}} - \theta \geq \overline{v})$$

$$\leq e^{-\left(\frac{n-k}{3}\overline{v} - \frac{k+1}{2}\right)^2/(n+1)} \leq e^{-cn\overline{v}^2} \leq e^{-c'n/\log^3(n)},$$

for some constant $c' > 0$ and n large enough (above, we have used that $k/(n-k) = O(n^{-1/5})$). The first bound follows. Let us turn to proving the second bound. From (85), one has $\widehat{\theta}_{\mathrm{low},q_k} - \theta \succeq \varepsilon_{(\lceil n/2 \rceil)} - \overline{v}$. As a consequence,

$$\mathbb{E}_{\theta,\mu}\left[(\theta - \widehat{\theta}_{\mathrm{low},q})\mathbb{1}_{\widehat{\theta}_{\mathrm{low},q} \leq \theta - 2\overline{v}}\right] \leq \mathbb{E}_{\theta,\mu}\left[(-\varepsilon_{(\lceil n/2 \rceil)} + \overline{v})\mathbb{1}_{\varepsilon_{(\lceil n/2 \rceil)} \leq -\overline{v}}\right]$$

$$\leq 2\,\mathbb{E}_{\theta,\mu}\left[(-\varepsilon_{(\lceil n/2 \rceil)})_+\right] \leq 2c_1/\sqrt{n},$$

where the last bound is for instance a consequence of Proposition 2.2 for $k = 0$. $\square$

Lemma B.2 (Control of $\widehat{\theta}_{\mathrm{up}}$). *There exists an universal integer $n_0 \geq 1$ such that for any $n \geq n_0$, $\mu \in \mathcal{M}_{n-1}$ and any $\theta \in \mathbb{R}$, the estimatorv $\widehat{\theta}_{\mathrm{up}}$ satisfies*

$$\mathbb{P}_{\theta,\mu}\left[\widehat{\theta}_{\mathrm{up}} < \theta\right] \leq \frac{1}{n} ; \quad \mathbb{E}_{\theta,\mu}\left[(\theta - \widehat{\theta}_{\mathrm{up}})_+\right] \leq \frac{1}{n} ; \quad \mathbb{E}_{\theta,\mu}\left[(\widehat{\theta}_{\mathrm{up}} - \theta)_+ \mathbb{1}_{\widehat{\theta}_{\mathrm{up}} - \theta \geq 4\sqrt{\log(n)}}\right] \leq \frac{1}{n^2} .$$

Proof of Lemma B.2. The first bound is a slight variation of Proposition 2.3. As in the proof of that proposition, we start from $\widehat{\theta}_{\mathrm{up}} - \theta \succeq \epsilon_{(1)} + 2\sqrt{\log(n)}$ which implies

$$\mathbb{P}_{\theta,\mu}[\widehat{\theta}_{\mathrm{up}} < \theta] \leq \mathbb{P}\left[\epsilon_{(1)} < -2\sqrt{\log(n)}\right] \leq n\overline{\Phi}(2\sqrt{\log(n)}) \leq 1/n \ .$$

where we used an union bound and (118).

Second, we start from $\widehat{\theta}_{\mathrm{up}} - \theta \geq \varepsilon_{(1)} + 2\sqrt{\log n}$ and obtain

$$\begin{aligned}
\mathbb{E}_{\theta,\mu}\left[(\theta - \widehat{\theta}_{\mathrm{up}})_+\right] &= \int_0^\infty \mathbb{P}_{\theta,\mu}(\theta - \widehat{\theta}_{\mathrm{up}} \geq t)dt \\
&\leq \int_0^\infty \mathbb{P}(\varepsilon_{(1)} \leq -(t + 2\sqrt{\log n}))dt = n\int_0^\infty \overline{\Phi}(t + 2\sqrt{\log(n)})dt \\
&\leq \frac{n}{2\sqrt{\log n}} \int_0^\infty \phi(t + 2\sqrt{\log n})dt = \frac{n}{2\sqrt{\log n}}\overline{\Phi}(2\sqrt{\log n}) \leq 1/n \ ,
\end{aligned}$$

where we used (118) and n large enough in the last line.

Finally, since at least one μ_i is zero, we may assume without loss of generality that $\mu_1 = 0$, which implies $\widehat{\theta}_{\mathrm{up}} - \theta \leq \varepsilon_1 + 2\sqrt{\log n}$. This leads us to

$$\begin{aligned}
\mathbb{E}_{\theta,\mu}\left[(\widehat{\theta}_{\mathrm{up}} - \theta)_+ \mathbb{1}_{\widehat{\theta}_{\mathrm{up}} - \theta \geq 4\sqrt{\log n}}\right] &\leq \mathbb{E}_{\theta,\mu}\left[\left[\varepsilon_1 + 2\sqrt{\log n}\right]\mathbb{1}_{\epsilon_1 \geq 2\sqrt{\log n}}\right] \\
&\leq 2\,\mathbb{E}_{\theta,\mu}\left[\varepsilon_1 \mathbb{1}_{\epsilon_1 \geq 2\sqrt{\log n}}\right] = 2\phi(2\sqrt{\log n}) \leq \frac{1}{n^2} \ ,
\end{aligned}$$

by integration. $\qquad\square$

B.3. Proof of Theorem 2.4 (gOSC)

In order to ease the notation, we write λ_k for λ_{q_k}. We first prove the probability bound for $\widehat{\theta}_{q_k}$ and then turn to the moment bound. First, recall that the population function $\psi_{q,\lambda}(u)$ has been defined in such a way that $\psi_{q,\lambda}(u) \in [-1,1]$ for all $u \leq \theta$ and is larger than 1 for u sufficiently large. The following lemma quantifies this phenomenon by providing a lower bound for $\psi_{q,\lambda_q}(\theta + v_k^*)$ with some v_k^* defined by

$$v_k^* := \frac{5k}{(n-k)\lambda_k q_k^2} \text{ if } q_k < q_{\max} \ , \text{ and } v_k^* := \frac{2}{q_k^2 \lambda_k}\log^2\left(\frac{8n}{n-k}\right) \text{ if } q_k = q_{\max} \ . \tag{86}$$

When $q_k < q_{\max}$, we have $k < e^{-2a}n$. As a consequence, we easily check that, for any $k \in [e^{2a}\sqrt{n}, n - 64n^{1-1/(4a)})$, one has

$$v_k^* \leq c\frac{\log^2\left(1 + \sqrt{\frac{k}{n-k}}\right)}{\log^{3/2}\left(\frac{k}{\sqrt{n}}\right)} \ , \tag{87}$$

in both cases, where c is a positive universal constant.

Lemma B.3. *Let us consider the function $\psi_{q,\lambda}$ defined by (15) and any integer $k \in [e^{2a}\sqrt{n}, n - 64n^{1-1/(4a)})$ and v_k^* defined by (86). Assume $\mu \in \mathcal{M}_k$. Then, we have*

$$\psi_{q_k,\lambda_k}(\theta + v_k^*) > 1 + \frac{k}{n}(1 + e^{q_k\lambda_k v_k^*}) \ . \tag{88}$$

Besides, if $q_k < q_{\max}$, we have for any $\omega \geq 1$,

$$\psi_{q_k,\lambda_k}(\theta + \omega v_k^*) > 1 + \omega \frac{k}{n}(1 + e^{q_k \lambda_k v_k^*}) .$$

The second lemma controls the simultaneous deviations of the statistics $\widehat{\psi}_{q_k,\lambda_k}(u)$, $u \in \mathbb{R}$, around their expectations.

Lemma B.4. *Let us consider the functions $\psi_{q,\lambda}$ and $\widehat{\psi}_{q,\lambda}$ defined by (15) and (18), respectively, for some arbitrary $\mu \in \mathcal{M}$. Fix any $t > 0$, any $\lambda > 0$ and any positive even integer q. Then, with probability higher than $1 - 1/t^2$, we have*

$$|\widehat{\psi}_{q,\lambda}(u) - \psi_{q,\lambda}(u)| \leq \frac{t}{\sqrt{n}} q^{3/2} \exp\left[\lambda^2 \frac{q^2}{2} + q \log(3 + 2\sqrt{2}) - \lambda(\theta - u)_+ + \lambda q(u - \theta)_+\right] ,$$

simultaneously over all $u \in \mathbb{R}$.

Let us now define

$$t_k = \frac{e^{aq_k}}{q_k^{3/2}} e^{-q_k^2 \lambda_k^2/2 - q_k \log(3 + 2\sqrt{2})} = \frac{e^{2aq_k/3}}{q_k^{3/2}} \geq \frac{e^{-8a/3}}{q_k^{3/2}}\left(\frac{k}{\sqrt{n}}\right)^{2/3} , \tag{89}$$

by definition of a, and because $\lambda_k = (2/q_k)^{1/2}$ and $q_k \geq a^{-1} \log(k/\sqrt{n}) - 4$. It readily follows from Lemma B.4 that, with probability higher than $1 - t_k^{-2}$, we have

$$\sup_{u \leq \theta} \widehat{\psi}_{q,\lambda_k}(u) \leq \sup_{u \leq \theta} \psi_{q,\lambda_k}(u) + \frac{e^{aq_k}}{n^{1/2}} \leq 1 + \frac{e^{aq_k}}{n^{1/2}} . \tag{90}$$

Together with Lemma B.3, this also leads to (on the same event)

$$\widehat{\psi}_{q_k,\lambda_k}(v_k^* + \theta) \geq \psi_{q_k,\lambda_k}(v_k^* + \theta) - \frac{e^{aq_k}}{n^{1/2}} e^{q_k \lambda_k v_k^*} > 1 + \frac{e^{aq_k}}{n^{1/2}} , \tag{91}$$

since $aq_k \leq \log(k/\sqrt{n})$. Thanks to (90) and (91), we have proved that

$$\mathbb{P}_{\theta,\mu}\left[\widehat{\theta}_{q_k} \in [\theta, \theta + v_k^*]\right] \geq 1 - t_k^{-2} - \mathbb{P}_{\theta,\mu}\left[\widehat{\theta}_{\text{low},q_k} \geq \theta + v_k^*\right] - \mathbb{P}_{\theta,\mu}\left[\widehat{\theta}_{\text{up}} < \theta\right] . \tag{92}$$

Note that, for the probability bound, the preliminary estimators $\widehat{\theta}_{\text{low},q_k}$ and $\widehat{\theta}_{\text{up}}$ do not help at all and we would have obtained a similar result had we simply taken $\widehat{\theta}_{\min} = -\infty$ and $\widehat{\theta}_{\text{up}} = +\infty$ in which case the two last terms in the above bound would be equal to zero. With our choice of preliminary estimators, Lemmas B.1 and B.2 ensure that the two probabilities in the right hand side of (92) are small compared to $1/n$. We have proved that

$$\mathbb{P}_{\theta,\mu}\left[\widehat{\theta}_{q_k} \in [\theta, \theta + v_k^*]\right] \leq c_3 \left(\frac{k}{\sqrt{n}}\right)^{-4/3} \log^3\left(\frac{k}{\sqrt{n}}\right) , \tag{93}$$

for some constant c_3, which in view of the bound (87) of v_k^* leads to the desired probability bound (20).

Let us turn to prove the moment bound (21). We consider separately $\mathbb{E}_{\theta,\mu}[(\widehat{\theta}_{q_k} - \theta)_+]$ and $\mathbb{E}_{\theta,\mu}[(\theta - \widehat{\theta}_{q_k})]_+$.

Step 1: Control of $\mathbb{E}_{\theta,\mu}[(\widehat{\theta}_{q_k} - \theta)_+]$. The analysis is divided into two cases, depending on the value of k.

Case 1: $q_k \geq 0.3a^{-1}\log n$ *which implies* $k \geq n^{4/5}$. Since $\widehat{\theta}_{q_k} \leq \widehat{\theta}_{up}$, we have the following risk decomposition.

$$
\begin{aligned}
\mathbb{E}_{\theta,\mu}[(\widehat{\theta}_{q_k} - \theta)_+] &\leq v_k^* + \mathbb{E}_{\theta,\mu}\left[(\widehat{\theta}_{up} - \theta)_+ \mathbb{1}_{\widehat{\theta}_{q_k} - \theta \geq v_k^*}\right] \\
&\leq v_k^* + 4\sqrt{\log n}\, \mathbb{P}_{\theta,\mu}\left[\widehat{\theta}_{q_k} - \theta \geq v_k^*\right] + \mathbb{E}_{\theta,\mu}\left[(\widehat{\theta}_{up} - \theta)_+ \mathbb{1}_{\widehat{\theta}_{up} - \theta \geq 4\sqrt{\log(n)}}\right].
\end{aligned}
$$

The second term is less than $\log^{7/2}(n)(k/\sqrt{n})^{-4/3}$ which is small in front of v_k^* since $k \geq n^{4/5}$. Finally, the last term is small compared to $1/n$ by Lemma B.2. We have proved that $\mathbb{E}_{\theta,\mu}[(\widehat{\theta}_{q_k} - \theta)_+] \lesssim v_k^*$ (for n large enough).

Case 2: $q_k < 0.3a^{-1}\log n$. Define the event $\mathcal{A} = \{\widehat{\theta}_{low,q_k} \leq \theta\}$. Since $\widehat{\theta}_{q_k} \leq \widehat{\theta}_{up}$, we have the following decomposition.

$$
\begin{aligned}
\mathbb{E}_{\theta,\mu}[(\widehat{\theta}_{q_k} - \theta)_+] &\leq \mathbb{E}_{\theta,\mu}\left[(\widehat{\theta}_{q_k} - \theta)_+ \mathbb{1}_{\mathcal{A}}\right] + \mathbb{E}_{\theta,\mu}\left[(\widehat{\theta}_{up} - \theta)_+ \mathbb{1}_{\mathcal{A}^c}\right] \\
&\leq \mathbb{E}_{\theta,\mu}\left[(\widehat{\theta}_{q_k} - \theta)_+ \mathbb{1}_{\mathcal{A}}\right] + 4\sqrt{\log n}\, \mathbb{P}_{\theta,\mu}[\mathcal{A}^c] + \mathbb{E}_{\theta,\mu}\left[(\widehat{\theta}_{up} - \theta)_+ \mathbb{1}_{\widehat{\theta}_{up} - \theta \geq 4\sqrt{\log n}}\right].
\end{aligned}
$$

By Lemma B.2, the third term in the rhs has been proved to be small compared $1/n$. By Lemma B.1, $\sqrt{\log n}\, \mathbb{P}_{\theta,\mu}[\mathcal{A}^c]$ is small compared to $\sqrt{\log n}/n$ which in turn is smaller than v_k^*. Hence, we only need to prove that $\mathbb{E}_{\theta,\mu}\left[(\widehat{\theta}_{q_k} - \theta)_+ \mathbb{1}_{\mathcal{A}}\right]$ is of order at most v_k^*. By integration, it suffices to prove that, for all $\omega \geq 1$,

$$
\mathbb{P}_{\theta,\mu}[(\widehat{\theta}_{q_k} - \theta)\mathbb{1}_{\mathcal{A}} > \omega v_k^*] \leq (\omega t_k)^{-2}, \tag{94}
$$

Fix some $\omega \geq 1$. Since $q_k < q_{max} - 2$, Lemma B.3 ensures that

$$
\psi_{q_k,\lambda_k}(\theta + \omega v_k^*) > 1 + \omega \frac{k}{n}(1 + e^{q_k \lambda_k v_k^*}).
$$

From Lemma B.4 with $t = \omega t_k$, we deduce as in (91) that, with probability higher than $1 - (\omega t_k)^{-2}$

$$
\widehat{\psi}_{q_k,\lambda_k}(\theta + \omega v_k^*) \geq \psi_{q_k,\lambda_k}(\theta + \omega v_k^*) - \omega \frac{k}{n}e^{q_k \lambda_k v_k^*} > 1 + \frac{k}{n} \geq 1 + \frac{e^{aq_k}}{\sqrt{n}},
$$

Together with $\mathcal{A}$, this event enforces that $\widehat{\theta}_{q_k} \leq \theta + \omega v_k^*$. We have proved (94). This entails $\mathbb{E}_{\theta,\mu}[(\widehat{\theta}_{q_k} - \theta)_+] \leq c v_k^*$ for some universal constant $c > 0$.

Step 2: Control of $\mathbb{E}_{\theta,\mu}[(\theta - \widehat{\theta}_{q_k})_+]$. Define the estimator

$$
\widetilde{\theta}_q = \inf\left\{u \in [\widehat{\theta}_{low,q}, +\infty) \; : \; \widehat{\psi}_{q,\lambda_q}(u) > 1 + \frac{e^{aq}}{\sqrt{n}}\right\}.
$$

It follows from this definition that $\widehat{\theta}_q \geq \widetilde{\theta}_q \wedge \widehat{\theta}_{up}$ and

$$
\mathbb{E}_{\theta,\mu}[(\theta - \widehat{\theta}_{q_k})_+] \leq \mathbb{E}_{\theta,\mu}\left[(\theta - \widehat{\theta}_{up})_+\right] + \mathbb{E}_{\theta,\mu}\left[(\theta - \widetilde{\theta}_{q_k})_+\right].
$$

By Lemma B.2, the first term in the rhs is small compared to $1/n$ and we focus on the second expectation.

Case 1: $q_k > 0.3a^{-1}\log(n)$ which implies $k \geq n^{4/5}$. Fix any $\omega \geq 1$ and define $v_\omega = \log(\omega)/\lambda_k$. It follows from Lemma B.4, that, with probability higher than $1 - (t_k\omega)^{-2}$, we have simultaneously over all $u \geq v_\omega$,

$$\widehat{\psi}_{q_k,\lambda_k}(\theta - u) \leq \psi_{q_k,\lambda_k}(\theta - u) + \omega\frac{e^{aq_k}}{\sqrt{n}}e^{-\lambda_k u} \leq 1 + \omega\frac{e^{aq_k}}{\sqrt{n}}e^{-\lambda_k v_\omega} = 1 + \frac{e^{aq_k}}{\sqrt{n}} \ ,$$

where we used $|\psi_{q_k,\lambda_k}(\theta - u)| \leq 1$ for all $u > 0$. With probability higher than $1 - (t_k\omega)^{-2}$, $\widetilde{\theta}_{q_k}$ is therefore higher than $\theta - \log(\omega)/\lambda_k$. Integrating this last bound leads to

$$\mathbb{E}_{\theta,\mu}[(\theta - \widetilde{\theta}_{q_k})_+] \leq \frac{1}{2t_k^2\lambda_k} \lesssim v_k^* \ , \tag{95}$$

since $k \geq n^{4/5}$ (for n large enough).

Case 2: $q_k < 0.3a^{-1}\log(n)$. Define the event $\mathcal{B}_k = \{\widehat{\theta}_{\text{low},q_k} \geq \theta - v_{\min,k}\}$ with $v_{\min,k} = \sqrt{2}\frac{\pi^2}{72\lambda_k q_k^2} \geq 2\overline{v}$ (where $\overline{v}$ is defined along with $\widehat{\theta}_{\text{low},q}$). Since $\widetilde{\theta}_{q_k} \geq \widehat{\theta}_{\text{low},q_k}$, we have the following decomposition.

$$\mathbb{E}_{\theta,\mu}[(\theta - \widetilde{\theta}_{q_k})_+] \leq \mathbb{E}_{\theta,\mu}\big[(\theta - \widehat{\theta}_{\text{low},q_k})_+\mathbb{1}_{\mathcal{B}_k^c}\big] + \mathbb{E}_{\theta,\mu}\big[(\theta - \widetilde{\theta}_{q_k})_+\mathbb{1}_{\mathcal{B}_k}\big] \ . \tag{96}$$

We start by considering the first term in the right hand side. Since $\overline{v} \leq v_{\min,k}/2$, we rely on Lemma B.1 to derive that

$$\mathbb{E}_{\theta,\mu}\left[(\theta - \widehat{\theta}_{\text{low},q_k})_+\mathbb{1}_{\mathcal{B}_k^c}\right] \leq \mathbb{E}_{\theta,\mu}\left[(\theta - \widehat{\theta}_{\text{low},q_k})_+\mathbb{1}_{\widehat{\theta}_{\text{low},q_k} < \theta - 2\overline{v}}\right] \leq c/\sqrt{n} \ . \tag{97}$$

We now turn to $\mathbb{E}_{\theta,\mu}\big[(\theta - \widetilde{\theta}_{q_k})_+\mathbb{1}_{\mathcal{B}_k}\big]$ in (96). In comparison to the previous case, this bound is slightly more involved and we rely on the explicit expression of Chebychev Polynomials. For any $v > 0$, we have

$$\psi_{q_k,\lambda_k}(\theta - v) \leq \frac{k}{n} + \left(1 - \frac{k}{n}\right)\cos(q_k \arg\cos(2e^{-\lambda_k v} - 1)) \ .$$

Observe that $1 - e^{-t} \in [t/2, t]$ for $t \in [0, \log 2]$, $\cos t \leq 1 - t^2/4$ for all $t \in [0, \pi/2]$ and $\arg\cos(1 - t) \in [\sqrt{2t}, 2\sqrt{t}]$ for $t \in [0, 1]$. As a consequence, for $v \leq v_{\min,k}$, one has

$$\psi_{q_k,\lambda_k}(\theta - v) \leq 1 - \left(1 - \frac{k}{n}\right)\frac{q_k^2\lambda_k v}{2}.$$

Fix any $\omega > 1$. Thanks to deviation bound in Lemma B.4 we derive that, with probability higher than $1 - (\omega t_k)^{-2}$, we have

$$\widehat{\psi}_{q_k,\lambda_k}(\theta - v) \leq \psi_{q_k,\lambda_k}(\theta - v) + \omega\frac{e^{aq_k}}{\sqrt{n}}e^{-\lambda_k v} \leq 1 + \frac{e^{aq_k}}{\sqrt{n}} + (\omega - 1)\frac{k}{n} - \left(1 - \frac{k}{n}\right)\frac{q_k^2\lambda_k v}{2} \ ,$$

simultaneously for all $v \in [0, v_{\min,k}]$. In the second inequality, we used that $aq_k \leq \log(k/\sqrt{n})$. As a consequence, $\widehat{\psi}_{q_k,\lambda_k}(\theta - v) \leq 1 + e^{aq_k}/\sqrt{n}$ for all v in the (possibly empty) interval

$$v \in \left[\frac{2(\omega - 1)k}{(n - k)q_k^2\lambda_k}, v_{\min,k}\right] \ .$$

Since we work under the event $\mathcal{B}_k = \{\widehat{\theta}_{\text{low},q_k} \geq \theta - v_{\text{min},k}\}$, this implies that

$$\mathbb{P}_{\theta,\mu}\left[(\theta - \widetilde{\theta}_{q_k})_+ \mathbb{1}_{\mathcal{B}_k} > \frac{2(\omega - 1)k}{(n-k)q_k^2\lambda_k}\right] \leq \frac{1}{\omega^2 t_k^2} ,$$

for all $\omega \geq 1$. Integrating this deviation bound, we conclude that

$$\mathbb{E}_{\theta,\mu}\left[(\theta - \widetilde{\theta}_{q_k})_+ \mathbb{1}_{\mathcal{B}_k}\right] \leq 2\frac{k}{(n-k)t_k^2 q_k^2\lambda_k} \lesssim v_k^* .$$

since $t_k \gtrsim 1$.

Together with (95) and (97), we have proved that $\mathbb{E}_{\theta,\mu}[(\theta - \widehat{\theta}_{q_k})_+] \lesssim v_k^*$, which concludes the proof of the theorem.

Proof of Lemma B.3. Let us first prove the following inequality:

$$g_q(t) = \cosh(q \arg\cosh(2e^t - 1) \geq \max\left[1 + q^2 t, \frac{1}{2}e^{q\sqrt{2t}}\right]. \tag{98}$$

Since $\cosh(t) \leq e^{t^2/2}$ for all $t > 0$ (compare the power expansions), we have $\cosh(\sqrt{2}t) \leq e^t \leq 2e^t - 1$, implying that $g_q(t) \geq \cosh(q\sqrt{2t})$. Then, we use that $\cosh(t) \geq 1 + t^2/2$ and $\cosh(t) \geq e^t/2$ to conclude.

For a k-sparse vector μ, we have already observed in (16) that, for all $t > 0$,

$$\psi_{q,\lambda}(\theta + t) \geq -\frac{k}{n} + \frac{n-k}{n}g_q(\lambda t) .$$

The analysis is divided into two cases, depending on the value of k.

Case 1: $q_k < q_{\max}$. For any $t > 0$, it follows from (98) that $g_q(t) \geq 1 + q^2 t$

$$\psi_{q_k,\lambda_k}(\theta + t) \geq -\frac{k}{n} + \frac{n-k}{n}(1 + q_k^2\lambda_k t) = 1 - 2\frac{k}{n} + \frac{k}{n}\frac{(n-k)q_k^2\lambda_k t}{k} .$$

If we choose $t = \omega v_k^*$ with $\omega \geq 1$, we have

$$\psi_{q_k,\lambda_k}(\theta + t) \geq 1 + (5\omega - 2)\frac{k}{n} \geq 1 + 3\omega\frac{k}{n} .$$

Finally, we have

$$\exp(\lambda_k v_k^* q_k) = \exp\left[\frac{5k}{(n-k)q_k}\right] \leq \exp\left[\frac{5k}{2(n-k)}\right] < 2 ,$$

where we used in the last inequality the fact that $q_k < q_{\max}$ which implies $k < e^{-2a}n \leq n/6$. This concludes the first part of the proof.

Case 2: $q_k = q_{\max}$, that is $k \in [\sqrt{n}e^{aq_{\max}}, n - 64n^{1-1/(4a)})$. Recall $v_k^* = \frac{2}{q_k^2\lambda_k}\log^2(8n/(n-k))$. Together with (98), this yields

$$\begin{aligned}
\psi_{q_k,\lambda_k}(\theta + v_k^*) &\geq -\frac{k}{n} + \frac{n-k}{2n}\exp\left[q_k\sqrt{2\lambda_k v_k^*}\right] \geq -1 + \frac{n-k}{2n}\exp\left[q_k\sqrt{2\lambda_k v_k^*}\right] \\
&\geq -1 + 4\exp\left[\frac{1}{2}q_k\sqrt{2\lambda_k v_k^*}\right] \geq 2 + \exp\left[\frac{1}{2}q_k\sqrt{2\lambda_k v_k^*}\right] ,
\end{aligned}$$

because $\exp\left[\frac{1}{2}q_k\sqrt{2\lambda_k v_k^*}\right] \geq 8n/(n-k)$ by definition of v_k^*. Next, note that

$$q_k = q_{\max} = \lfloor\frac{1}{2a}\log(n)\rfloor_{even} - 2 \geq \frac{1}{2a}\log(n) - 4 .$$

Furthermore, the condition $k \leq n - 64n^{1-1/(4a)}$ has been chosen in such that a way that

$$q_k \geq 2\log\left(\frac{8n}{n-k}\right) ,$$

which implies that $\lambda_k v_k^* \leq 1/2$. This allows to conclude that

$$\psi_{q_k,\lambda_k}(\theta + v_k^*) > 1 + \frac{k}{n}(1 + e^{q_k\lambda_k v_k^*}) .$$

$\square$

Proof of Lemma B.4. At $u = \theta$, simple computations lead to $\mathrm{Var}\,\widehat{\eta}_\lambda(\theta) \leq e^{\lambda^2}/n$. It then follows from Chebychev's inequality that

$$\mathbb{P}_{\theta,\mu}\left[|\widehat{\eta}_\lambda(\theta) - \eta_\lambda(\theta)| \geq \frac{t}{\sqrt{n}}e^{\lambda^2/2}\right] \leq \frac{1}{t^2} ,$$

for all $t > 0$. For a general $u \in \mathbb{R}$, observe that $\eta_\lambda(u)$ (resp. $\widehat{\eta}_\lambda(u)$) is a simple transformation of $\eta_\lambda(\theta)$ (resp. $\widehat{\eta}_\lambda(\theta)$):

$$\widehat{\eta}_\lambda(u) = \widehat{\eta}_\lambda(\theta)e^{\lambda(u-\theta)} \quad\text{and}\quad \eta_\lambda(u) = \eta_\lambda(\theta)e^{\lambda(u-\theta)} .$$

This entails

$$\mathbb{P}_{\theta,\mu}\left[\exists u \in \mathbb{R}, \ |\widehat{\eta}_\lambda(u) - \eta_\lambda(u)| \geq \frac{t}{\sqrt{n}}e^{\lambda^2/2}e^{\lambda(u-\theta)}\right] \leq \frac{1}{t^2} .$$

Then, taking an union bound over all $j = 1, \ldots, q$, we obtain that, for any $t > 0$, we have

$$|\widehat{\eta}_{\lambda j}(u) - \eta_{\lambda j}(u)| \leq t\sqrt{\frac{q}{n}}e^{(\lambda j)^2/2+\lambda j(u-\theta)} , \quad \text{for all } u \in \mathbb{R} \text{ and } j = 1, \ldots, q , \quad (99)$$

with probability higher than $1 - 1/t^2$. Then, we rely on the upper bound (116) of the coefficients $|a_{j,q}|$ in the definition of $\widehat{\psi}_{q,\lambda}$ to obtain

$$|\widehat{\psi}_{q,\lambda}(u) - \psi_{q,\lambda}(u)| \leq \frac{t}{\sqrt{n}}q^{3/2}\exp\left[\lambda^2\frac{q^2}{2} + q\log(3 + 2\sqrt{2}) - \lambda(\theta - u)_+ + \lambda q(u - \theta)_+\right] ,$$

simultaneously over all $u \in \mathbb{R}$ with probability higher than $1 - 1/t^2$. $\square$

B.4. Proof of Theorem 2.7 (gOSC)

Consider any $(\theta, \mu) \in \mathbb{R} \times \mathcal{M}$ and denote $k = \|\mu\|_0$, which is such that $k \leq n-1$ by assumption. Let us denote

$$q_* = \begin{cases} 0 & \text{if } k < e^{2a}\sqrt{n} ; \\ q_k & \text{if } k \in [e^{2a}\sqrt{n}, n - 64n^{1-1/4a}) ; \\ q_{\max} + 2 & \text{else} . \end{cases}$$

We call $\widehat{\theta}_{q*}$ the oracle estimator because $\widehat{\theta}_{q*}$ has been shown to achieve the desired risk bounds (see Propositions 2.2 and 2.3 and Theorem 2.4). Let us also underline that $q \in \{0, \ldots, q_{\max} + 2\} \mapsto \delta_q$ is increasing, because $x \in [1, \infty) \mapsto ax - (3/2)\log x$ is also increasing.

We start by proving the probability bound (23). We first assume that $q_* < q_{\max}$ and consider afterwards the case $q_* \geq q_{\max}$. Consider the event $\mathcal{A} = \cap_{q \geq q_*}\{|\widehat{\theta}_q - \theta| \leq \delta_q/2\}$. Under the event $\mathcal{A}$, it follows from triangular inequality and the fact that the sequence δ_q is increasing that $\widehat{q} \leq q_*$. Relying again on triangular inequality and the definition of $\widehat{q}$, we obtain

$$|\widehat{\theta}_{\widehat{q}} - \theta| \leq |\widehat{\theta}_{\widehat{q}} - \widehat{\theta}_{q*}| + |\widehat{\theta}_{q*} - \theta| \leq \frac{3}{2}\delta_{q*}.$$

We deduce

$$\mathbb{P}_{\theta,\mu}\left[|\widehat{\theta}_{\widehat{q}} - \theta| > \frac{3}{2}\delta_{q*}\right] \leq \mathbb{P}_{\theta,\mu}[\mathcal{A}^c] \leq \sum_{q \geq q_*} \mathbb{P}_{\theta,\mu}\left[|\widehat{\theta}_q - \theta| > \delta_q/2\right].$$

Since for any $k < k'$, a k-sparse vector is also a k'-sparse vector, we can apply the deviation bounds (92) and (93) in the proof of Theorem 2.4 (and the definition (89) of t_k) to all estimators $\widehat{\theta}_q$ with $q = q_*, \ldots, q_{\max}$. For such q, we obtain $\mathbb{P}_{\theta,\mu}[|\widehat{\theta}_q - \theta| > \delta_q/2] \lesssim e^{-4aq/3}q^3$.

Proposition 2.3 also enforces that $\mathbb{P}_{\theta,\mu}[|\widehat{\theta}_{q_{\max}+2} - \theta| > \delta_{q_{\max}+2}/2] \lesssim 1/n$. We conclude that

$$\mathbb{P}_{\theta,\mu}\left[|\widehat{\theta}_{\mathrm{ad}} - \theta| \geq \frac{3}{2}\delta_{q*}\right] \lesssim e^{-4aq_*/3}q_*^3 + \frac{1}{n},$$

which leads to the desired result. For $q_* = q_{\max}$, it follows again from (93) in the proof of Theorem 2.7 that

$$\mathbb{P}_{\theta,\mu}\left[|\widehat{\theta}_{q_{\max}} - \theta| \geq v_k^*\right] \lesssim \left(\frac{k}{\sqrt{n}}\right)^{4/3}\log^3(\frac{k}{\sqrt{n}}), \text{ where } v_k^* = \frac{4}{\sqrt{2}q_{\max}^{3/2}}\log^2\left(\frac{8n}{n-k}\right).$$

We consider two subcases: (i) $v_k^* \leq 2\sqrt{2\log(n)} = \delta_{q_{\max}+2}/2$ and (ii) $v_k^* > 2\sqrt{2\log(n)}$. Under (i), the event $\mathcal{A}' = \{|\widehat{\theta}_{q_{\max}} - \theta| \vee |\widehat{\theta}_{q_{\max}+2} - \theta| \leq \delta_{q_{\max}+2}/2\}$ has large probability and ensures that $\widehat{q} \leq q_{\max}$, which in turn implies that $|\widehat{\theta}_{\mathrm{ad}} - \theta|$ is smaller than $v_k^* + \delta_{q_{\max}} \lesssim v_k^*$. Under (ii), we simple use $|\widehat{\theta}_{\mathrm{ad}} - \theta| \leq \delta_{q_{\max}+2} + |\widehat{\theta}_{q_{\max}+2} - \theta|$ which is less than $3\delta_{q_{\max}+2}/2 \lesssim v_k^*$ with probability higher than $1 - c/n$ by Proposition 2.3. Finally, the case $q_* = q_{\max} + 2$ is handled similarly: we use $|\widehat{\theta}_{\mathrm{ad}} - \theta| \leq \delta_{q_{\max}+2} + |\widehat{\theta}_{q_{\max}+2} - \theta|$, which is less than $3\delta_{q_{\max}+2}/2$ with probability higher than $1 - c/n$ by Proposition 2.3.

Let us turn to the moment bound. We decompose the risk in a sum of two terms depending on the value of $\widehat{q}$.

$$\mathbb{E}_{\theta,\mu}\left[|\widehat{\theta}_{\mathrm{ad}} - \theta|\right] = \mathbb{E}_{\theta,\mu}\left[|\widehat{\theta}_q - \theta|\mathbb{1}_{\widehat{q} \leq q_*}\right] + \mathbb{E}_{\theta,\mu}\left[|\widehat{\theta}_q - \theta|\mathbb{1}_{\widehat{q} > q_*}\right]. \tag{100}$$

For $\widehat{q} < q_*$, it follows from triangular inequality and the definition of $\widehat{q}$ that

$$|\widehat{\theta}_q - \theta|\mathbb{1}_{\widehat{q}=q} \leq |\widehat{\theta}_{q*} - \widehat{\theta}_q|\mathbb{1}_{\widehat{q}=q} + |\widehat{\theta}_{q*} - \theta|\mathbb{1}_{\widehat{q}=q} \leq \delta_{q*}\mathbb{1}_{\widehat{q}=q} + |\widehat{\theta}_{q*} - \theta|\mathbb{1}_{\widehat{q}=q}.$$

Summing all these terms, we arrive at

$$\sum_{q=0}^{q_*} \mathbb{E}_{\theta,\mu}\left[|\widehat{\theta}_q - \theta|\mathbb{1}_{\widehat{q}=q}\right] \leq \delta_{q*} + \mathbb{E}_{\theta,\mu}[|\widehat{\theta}_{q*} - \theta|].$$

This expectation has been studied in Proposition 2.3 and Theorem 2.4, which leads us to

$$\mathbb{E}_{\theta,\mu}\left[|\widehat{\theta}_q - \theta|\mathbb{1}_{\widehat{q}\leq q_*}\right] \lesssim \frac{\log^2\left(1 + \sqrt{\frac{k}{n-k}}\right)}{\log^{3/2}\left(1 + (\frac{k}{\sqrt{n}})^{2/3}\right)} \ . \tag{101}$$

Turning to the second sum in the decomposition (100), we first assume that $q_* < q_{\max}$ define $\widetilde{q}_+ = \max\{q' : |\widehat{\theta}_{q'} - \theta| \geq \delta_{q'}/2\}$. By definition of δ_q, one has $\widetilde{q}_+ \geq \widehat{q} - 2$ under the event $\widehat{q} > q_*$. Then, we deduce that

$$\begin{aligned}
|\widehat{\theta}_q - \theta|\mathbb{1}_{\widehat{q}=q} &\leq \sum_{q'=q-2}^{q_{\max}+2} |\widehat{\theta}_q - \theta|\mathbb{1}_{\widehat{q}=q}\mathbb{1}_{\widetilde{q}_+=q'} \\
&\leq |\widehat{\theta}_q - \theta|\mathbb{1}_{\widehat{q}=q}\mathbb{1}_{\widetilde{q}_+=q-2} + \sum_{q'=q}^{q_{\max}} \left[|\widehat{\theta}_q - \widehat{\theta}_{q'+2}| + |\widehat{\theta}_{q'+2} - \theta|\right]\mathbb{1}_{\widehat{q}=q}\mathbb{1}_{\widetilde{q}_+=q'} \\
&\quad + \left[|\widehat{\theta}_q - \widehat{\theta}_{q_{\max}+2}| + |\widehat{\theta}_{q_{\max}+2} - \theta|\right]\mathbb{1}_{\widehat{q}=q}\mathbb{1}_{\widetilde{q}_+=q_{\max}+2} \\
&\leq \frac{3}{2}\sum_{q'=q-2}^{q_{\max}} \delta_{q'+2}\mathbb{1}_{\widehat{q}=q}\mathbb{1}_{\widetilde{q}_+=q'} + \left[\delta_{q_{\max}+2} + |\widehat{\theta}_{q_{\max}+2} - \theta|\right]\mathbb{1}_{\widehat{q}=q}\mathbb{1}_{\widetilde{q}_+=q_{\max}+2} \ ,
\end{aligned}$$

where we used again the definition of $\widetilde{q}_+$ and $\widehat{q}$. Summing the above bound over all even $q > q_*$ leads to

$$\begin{aligned}
\sum_{q=q_*+2}^{q_{\max}} \mathbb{E}_{\theta,\mu}\left[|\widehat{\theta}_q - \theta|\mathbb{1}_{\widehat{q}=q}\right] &\leq \frac{3}{2}\sum_{q=q_*+2}^{q_{\max}} \delta_{q+2}\mathbb{P}_{\theta,\mu}[\widetilde{q}_+ = q] + \mathbb{E}_{\theta,\mu}\left[(\delta_{q_{\max}+2} + |\widehat{\theta}_{q_{\max}+2} - \theta|)\mathbb{1}_{\widetilde{q}_+=q_{\max}+2}\right] \\
&\leq \frac{3}{2}\sum_{q=q_*}^{q_{\max}} \delta_{q+2}\mathbb{P}_{\theta,\mu}[|\widehat{\theta}_q - \theta| \geq \frac{\delta_q}{2}] + 3\mathbb{E}_{\theta,\mu}\left[|\widehat{\theta}_{q_{\max}+2} - \theta|\mathbb{1}_{|\widehat{\theta}_{q_{\max}+2}-\theta|\geq\delta_{q_{\max}+2}/2}\right] \ .
\end{aligned}$$

As explained earlier, we know that

$$\mathbb{P}_{\theta,\mu}[|\widehat{\theta}_q - \theta| \geq \frac{\delta_q}{2}] \lesssim e^{-4aq/3}q^3 \quad \text{for } q = q_*, \ldots, q_{\max} \ ,$$

and $\mathbb{P}_{\theta,\mu}[|\widehat{\theta}_{q_{\max}+2} - \theta| \geq (\delta_{q_{\max}+2})/2] \lesssim 1/n$. Since $\delta_{q_{\max}+2} \lesssim \sqrt{\log(n)}$

$$\frac{3}{2}\sum_{q=q_*}^{q_{\max}} \delta_{q+2}\mathbb{P}_{\theta,\mu}[|\widehat{\theta}_q - \theta| \geq \frac{\delta_q}{2}] \lesssim \frac{1}{\sqrt{n}}\sum_{q=q_*}^{q_{\max}-2} q^{3/2}e^{-aq/3} + \frac{1}{\sqrt{n}} \lesssim \frac{1}{\sqrt{n}} \ .$$

Arguing as in the proof of Lemma B.2, we obtain that the second term

$$\mathbb{E}_{\theta,\mu}\left[|\widehat{\theta}_{q_{\max}+2} - \theta|\mathbb{1}_{|\widehat{\theta}_{q_{\max}+2}-\theta|\geq\delta_{q_{\max}+2}/2}\right] \lesssim \frac{1}{n} \ .$$

We have proved

$$\mathbb{E}_{\theta,\mu}\left[|\widehat{\theta}_q - \theta|\mathbb{1}_{\widehat{q}>q_*}\right] \lesssim \frac{1}{\sqrt{n}} \ .$$

We now turn to the case $q_* = q_{\max}$. We only need to bound $\mathbb{E}_{\theta,\mu}[|\widehat{\theta}_{q_{\max}+2} - \theta|\mathbb{1}_{\widehat{q}=q_{\max}+2}]$. The event $\widehat{q} = q_{\max} + 2$ only occurs if either $|\widehat{\theta}_{q_{\max}+2} - \theta| \geq \delta_{q_{\max}+2}/2$ or if $|\widehat{\theta}_{q_{\max}} - \theta| \geq \delta_{q_{\max}+2}/2$. This leads to

$$
\begin{aligned}
\mathbb{E}_{\theta,\mu}\left[|\widehat{\theta}_{q_{\max}+2} - \theta|\mathbb{1}_{\widehat{q}=q_{\max}+2}\right] &\leq \mathbb{E}_{\theta,\mu}\left[|\widehat{\theta}_{q_{\max}+2} - \theta|\mathbb{1}_{|\widehat{\theta}_{q_{\max}+2}-\theta|\geq\delta_{q_{\max}+2}}\right] \\
&\quad + 2\sqrt{2\log(n)}\mathbb{P}_{\theta,\mu}\left[|\widehat{\theta}_{q_{\max}} - \theta| \geq 2\sqrt{2\log(n)}\right] \\
&\lesssim \frac{1}{\sqrt{n}} + \sqrt{\log(n)}\mathbb{P}_{\theta,\mu}\left[|\widehat{\theta}_{q_*} - \theta| \geq 2\sqrt{2\log(n)}\right] .
\end{aligned}
$$

As previously, we consider two subcases: (i) $v_k^* < 2\sqrt{2\log(n)}$, in which case the deviation bound (93) implies that

$$
\begin{aligned}
\sqrt{\log(n)}\mathbb{P}_{\theta,\mu}\left[|\widehat{\theta}_{q_*} - \theta| \geq 2\sqrt{2\log(n)}\right] &\lesssim (\frac{k}{\sqrt{n}})^{-4/3}\log^{7/2}(n) \lesssim n^{-2/3}\log^{7/2}(n) \\
&\lesssim \frac{\log^2\left(1 + \sqrt{\frac{k}{n-k}}\right)}{\log^{3/2}\left(1 + (\frac{k}{\sqrt{n}})^{2/3}\right)} ,
\end{aligned}
$$

since $q_* = q_{\max}$. If (ii) $v_k^* \geq 2\sqrt{2\log(n)}$, we straightforwardly derive the rough bound

$$
\mathbb{E}_{\theta,\mu}\left[|\widehat{\theta}_{q_{\max}+2} - \theta|\mathbb{1}_{\widehat{q}=q_{\max}+2}\right] \lesssim \sqrt{\log(n)} ,
$$

which is nevertheless optimal. Together with (100) and (101), we have proved the desired risk bound.

B.5. Proofs for the quantile estimators (OSC)

Proof of Theorem 3.2. The proof is based on Lemmas D.6 and D.8. First recall that the following holds:

$$
\begin{aligned}
\xi_{(q)} + \overline{\Phi}^{-1}(q/n) \preceq \widetilde{\theta}_q - \theta &\preceq \xi_{(q:n-k)} + \overline{\Phi}^{-1}(q/n) \\
&\preceq \left[\xi_{(q:n-k)} + \overline{\Phi}^{-1}(q/(n-k))\right]_+ + \overline{\Phi}^{-1}(q/n) - \overline{\Phi}^{-1}(q/(n-k)) .
\end{aligned}
$$

Let us prove (28). It follows from the above decomposition that, for any $x > 0$,

$$
-x \leq \widetilde{\theta}_q - \theta \leq \overline{\Phi}^{-1}(q/n) - \overline{\Phi}^{-1}(q/(n-k)) + x ,
$$

with probability higher than $1 - \mathbb{P}[\xi_{(q:n-k)} + \overline{\Phi}^{-1}(\frac{q}{n-k}) \geq x] - \mathbb{P}[\xi_{(q)} + \overline{\Phi}^{-1}(\frac{q}{n}) \leq -x]$. Then, Lemmas D.6 and D.8 yield the desired bound (28) for all $x \leq c_3 q^2$.

Let us now prove (29). Define the event

$$
\mathcal{A} = \left\{|\widetilde{\theta}_q - \theta| \leq c_1 \frac{\log\left(\frac{n}{n-k}\right)}{\sqrt{\log\left(\frac{n-k}{q}\right)_+ \vee 1}} + c_2\sqrt{\frac{c_3}{[\log(\frac{n-k}{q}) \vee 1]}}\right\} .
$$

---

[2]Actually, c_3 corresponds to c_1 in the statement of Lemma D.8.

From above, the random variable $|\widetilde{\theta}_q - \theta|\mathbb{1}_{\mathcal{A}}$ satisfies for all $x > 0$,

$$\mathbb{P}_{\theta,\pi}\left[|\widetilde{\theta}_q - \theta|\mathbb{1}_{\mathcal{A}} \le c_1 \frac{\log\left(\frac{n}{n-k}\right)}{\sqrt{\log\left(\frac{n-k}{q}\right)_+ \vee 1}} + c_2\sqrt{\frac{x}{q[\log(\frac{n-k}{q}) \vee 1]}}\right] \ge 1 - 2e^{-x} . \tag{102}$$

Integrating this deviation inequality yields

$$\mathbb{E}_{\theta,\pi}\left[|\widetilde{\theta}_q - \theta|\mathbb{1}_{\mathcal{A}}\right] \le c_1 \frac{\log\left(\frac{n}{n-k}\right)}{\sqrt{\log\left(\frac{n-k}{q}\right)_+ \vee 1}} + c_2'\sqrt{\frac{1}{q[\log(\frac{n-k}{q}) \vee 1]}} .$$

Let us control the remaining term $\mathbb{E}_{\theta,\pi}[|\widetilde{\theta}_q - \theta|\mathbb{1}_{\mathcal{A}^c}]$. By Cauchy-Schwarz inequality, we have

$$\mathbb{E}_{\theta,\pi}[|\widetilde{\theta}_q - \theta|\mathbb{1}_{\mathcal{A}^c}] \le \mathbb{E}_{\theta,\pi}^{1/2}[(\widetilde{\theta}_q - \theta)^2]\mathbb{P}_{\theta,\pi}^{1/2}[\mathcal{A}^c] \le \sqrt{2}e^{-c_3q/2}\left[\mathbb{E}_{\theta,\pi}^{1/2}[(\widetilde{\theta}_q - \theta)_-^2] + \mathbb{E}_{\theta,\pi}^{1/2}[(\widetilde{\theta}_q - \theta)_+^2]\right] .$$

We can use a crude stochastic bound $\xi_{(1)} - \overline{\Phi}^{-1}(1/n) \preceq \widetilde{\theta}_q - \theta \preceq \xi_{(n)} + \overline{\Phi}^{-1}(1/n)$. By an union bound together with integration, we arrive at $\mathbb{E}_{\theta,\pi}[(\widetilde{\theta}_q - \theta)^2] \le c\log(n)$. Putting everything together, we obtain

$$\mathbb{E}_{\theta,\pi}\left[|\widetilde{\theta}_q - \theta|\right] \le c_1 \frac{\log\left(\frac{n}{n-k}\right)}{\sqrt{\log\left(\frac{n-k}{q}\right)_+ \vee 1}} + c_2'\sqrt{\frac{1}{q[\log(\frac{n-k}{q}) \vee 1]}} + c_4'e^{-c_3q/2}\sqrt{\log(n)} .$$

Taking $c_4 = 4/c_3$ in the statement of the theorem, we have $e^{-c_3q/2} \le n^{-2}$ and (29) follows.
$\square$

Proof of Corollary 3.3 . For $k \le n-n^{4/5}$, this bound is a straightforward consequence of (29). In the proof of Proposition 2.3 (see Section B.1), we have shown that, for any $\pi \in \overline{\mathcal{M}}_{n-1}$, $\mathbb{E}_{\theta,\pi}\left[|\widetilde{\theta}_1 - \theta|\right] \lesssim \sqrt{\log(n)}$, which is (up to multiple constants) smaller than $\frac{\log(\frac{n}{n-k})}{\log^{1/2}(1+\frac{k^2}{n})}$ for all $k \ge n - n^{4/5}$.
$\square$

Proof of Proposition 3.4. Consider any $(\theta, \pi) \in \mathbb{R} \times \overline{\mathcal{M}}$ and denote k the number of contaminations in π. In the sequel, we write $(\overline{q}_1, \ldots, \overline{q}_{\max})$ for the ordered values in $\mathcal{Q}$ so that $\overline{q}_1 = 1$, $\overline{q}_{\max} = \lceil n/2 \rceil$ and in between, the $\overline{q}_i$ form a dyadic sequence. For short, we write $q_* = q_k$ so that $\widetilde{\theta}_{q_*}$ achieves minimax performances. Besides, we let i^* be the indice such that $q_* = \overline{q}_{i^*}$. We shall prove that $\widetilde{\theta}_{\mathrm{ad}}$ performs almost as well $\widetilde{\theta}_{q_*}$. The general strategy is the same as in Theorem 2.7.

As in the previous proof, we decompose the risk as a sum of two terms depending on the value of $\widehat{q}$.

$$\mathbb{E}_{\theta,\pi}\left[|\widetilde{\theta} - \theta|\right] = \mathbb{E}_{\theta,\pi}\left[|\widetilde{\theta}_q - \theta|\mathbb{1}_{\widehat{q}\ge q_*}\right] + \mathbb{E}_{\theta,\pi}\left[|\widetilde{\theta}_q - \theta|\mathbb{1}_{\widehat{q}<q_*}\right] .$$

For $q \ge q_*$, it follows from triangular inequality and the definition of $\widehat{q}$ that

$$|\widetilde{\theta}_q - \theta|\mathbb{1}_{\widehat{q}=q} \le |\widetilde{\theta}_{q^*} - \widetilde{\theta}_q|\mathbb{1}_{\widehat{q}=q} + |\widetilde{\theta}_{q_*} - \theta|\mathbb{1}_{\widehat{q}=q} \le \delta_{q_*}\mathbb{1}_{\widehat{q}=q} + |\widetilde{\theta}_{q_*} - \theta|\mathbb{1}_{\widehat{q}=q} .$$

Summing all these terms over all $q \ge q_*$, we arrive at $\mathbb{E}_{\theta,\pi}\left[|\widetilde{\theta}_q - \theta|\mathbb{1}_{\widehat{q}\ge q_*}\right] \le \delta_{q_*} + \mathbb{E}_{\theta,\pi}[|\widetilde{\theta}_{q_*} - \theta|]$, which, by Corollary 3.3 together with the definition (33) of δ_q, leads to

$$\mathbb{E}_{\theta,\pi}\left[|\widetilde{\theta}_q - \theta|\mathbb{1}_{\widehat{q}\ge q_*}\right] \lesssim \frac{\log\left(\frac{n}{n-k}\right)}{\log^{1/2}\left(1 + \frac{k^2}{n}\right)} . \tag{103}$$

Turning to the second expression $\mathbb{E}_{\theta,\pi}\big[|\widetilde{\theta}_q - \theta|\mathbb{1}_{\widehat{q}<q_*}\big]$, we first assume either that $i^* = 1$ or $i^* > 3$, the case $i^* = 2, 3$ being deferred to the end of the proof. Define $\widetilde{q}_+ = \min\{q' : |\widetilde{\theta}_{q'} - \theta| \geq \delta_{q'}/2\}$. By definition of $\widehat{q}$ and by monotonicity of δ_q, one has $\widetilde{q}_+ \leq \overline{q}_{\widehat{i}+1}$ where $\widehat{i}$ is such that $\overline{q}_{\widehat{i}} = \widehat{q}$. Then, we deduce that, for $q < q_*$,

$$
\begin{aligned}
|\widetilde{\theta}_q - \theta|\mathbb{1}_{\widehat{q}=q} \ &\leq \ \sum_{j=1}^{i^*} |\widetilde{\theta}_q - \theta|\mathbb{1}_{\widehat{q}=q}\mathbb{1}_{\widetilde{q}_+=\overline{q}_j} \\
&\leq \ \sum_{j=2}^{i^*} \Big[|\widetilde{\theta}_q - \widetilde{\theta}_{\overline{q}_{j-1}}| + |\widetilde{\theta}_{\overline{q}_{j-1}} - \theta|\Big]\mathbb{1}_{\widehat{q}=q}\mathbb{1}_{\widetilde{q}_+=\overline{q}_j} + \Big[|\widetilde{\theta}_q - \widetilde{\theta}_1| + |\widetilde{\theta}_1 - \theta|\Big]\mathbb{1}_{\widehat{q}=q}\mathbb{1}_{\widetilde{q}_+=1} \\
&\leq \ \frac{3}{2}\sum_{j=2}^{i^*} \delta_{\overline{q}_{j-1}}\mathbb{1}_{\widehat{q}=q}\mathbb{1}_{\widetilde{q}_+=\overline{q}_j} + \Big[\delta_1 + |\widetilde{\theta}_1 - \theta|\Big]\mathbb{1}_{\widehat{q}=q}\mathbb{1}_{\widetilde{q}_+=1} \ ,
\end{aligned}
$$

where we used again the definition of $\widetilde{q}_+$ and $\widehat{q}$. Summing the above bound over all $q < q_*$ leads to

$$
\begin{aligned}
\sum_{q<q_*} \mathbb{E}_{\theta,\pi}\Big[|\widetilde{\theta}_q - \theta|\mathbb{1}_{\widehat{q}=q}\Big] \ &\leq \ \frac{3}{2}\sum_{j=2}^{i^*} \delta_{\overline{q}_{j-1}}\mathbb{P}_{\theta,\pi}[\widetilde{q}_+ = \overline{q}_j] + \mathbb{E}_{\theta,\pi}\Big[\big(\delta_1 + |\widetilde{\theta}_1 - \theta|\big)\mathbb{1}_{\widetilde{q}_+=1}\Big] \\
&\leq \ \frac{3}{2}\sum_{j=2}^{i^*} \delta_{\overline{q}_{j-1}}\mathbb{P}_{\theta,\pi}\Big[|\widetilde{\theta}_{\overline{q}_j} - \theta| \geq \frac{\delta_{\overline{q}_j}}{2}\Big] + 3\mathbb{E}_{\theta,\pi}\Big[|\widetilde{\theta}_1 - \theta|\mathbb{1}_{|\widetilde{\theta}_1 - \theta|\geq\delta_1/2}\Big] \ .
\end{aligned}
$$

For any $\overline{q}_4 \leq q \leq q_*$, we apply Theorem 3.2 and it follows from the choice (33) of δ_q with c_0 large enough that $\mathbb{P}_{\theta,\pi}[|\widetilde{\theta}_q-\theta| \geq \frac{\delta_q}{2}] \leq \exp[-c(n/q)^{1/3}]$. For $q \leq \overline{q}_3 \wedge q_*$, it follows from Theorem 3.2 and the proof of Proposition 2.3 that $\mathbb{P}_{\theta,\pi}[|\widetilde{\theta}_q - \theta| \geq \frac{\delta_q}{2}] \leq 1/n$. Since $\delta_{q_1} \lesssim \sqrt{\log(n)}$, we obtain

$$
\frac{3}{2}\sum_{j=2}^{i^*} \delta_{\overline{q}_j-1}\mathbb{P}_{\theta,\pi}\Big[|\widetilde{\theta}_{\overline{q}_j} - \theta| \geq \frac{\delta_{\overline{q}_j}}{2}\Big] \lesssim \frac{\sqrt{\log(n)}}{n} + \sum_{j=4}^{i^*} e^{-c(n/\overline{q}_j)^{1/3}}\frac{n^{1/6}}{\overline{q}_j^{2/3}\sqrt{\log(\frac{n}{\overline{q}_j}) \vee 1}} \lesssim \frac{1}{\sqrt{n}} \ .
$$

Finally, arguing as in the proof of Theorem 2.7, we observe that $\mathbb{E}_{\theta,\pi}[|\widetilde{\theta}_1 - \theta|\mathbb{1}_{|\widetilde{\theta}_1-\theta|\geq\delta_1/2}] \lesssim 1/\sqrt{n}$. Putting everything together we have proved

$$
\mathbb{E}_{\theta,\pi}\Big[|\widetilde{\theta}_q - \theta|\mathbb{1}_{\widehat{q}<q_*}\Big] \lesssim \frac{1}{\sqrt{n}} \ , \tag{104}
$$

as long as $i^* = 1$ or $i^* > 3$.

It remains to consider the case $i_* = 2, 3$. In that situation, observe that $\log(n/(n - k)) \asymp \log(n)$. From Theorem 3.2, we derive that, for $q = \overline{q}_1, \overline{q}_2$,

$$
\mathbb{E}_{\theta,\pi}\big[|\widetilde{\theta}_q - \theta|\big] \lesssim \frac{\log\left(\frac{n}{n-k}\right)}{\log^{1/2}\left(1 + \frac{k^2}{n}\right)} \ .
$$

This leads us to

$$
\mathbb{E}_{\theta,\pi}\Big[|\widetilde{\theta}_q - \theta|\mathbb{1}_{\widehat{q}<q_*}|\Big] \leq \sum_{i=1}^{2} \mathbb{E}_{\theta,\pi}\big[|\widetilde{\theta}_{\overline{q}_i} - \theta|\big] \lesssim \frac{\log\left(\frac{n}{n-k}\right)}{\log^{1/2}\left(1 + \frac{k^2}{n}\right)} \ .
$$

Together with (103) and (104), this concludes the proof. □

Proof of Proposition 4.2. First consider the variance estimator $\widetilde{\sigma}_{q_k,q'_k}$. We only deal with the case where $k \leq n - n^{4/5}$, the other case being trivial. We start from the decomposition

$$\frac{\left|\widetilde{\sigma}_{q,q'} - \sigma\right|}{\sigma} \leq \left|\frac{Y_{(q)}/\sigma - \theta/\sigma + \overline{\Phi}^{-1}(\frac{q}{n})}{\overline{\Phi}^{-1}(q'/n) - \overline{\Phi}^{-1}(q/n)}\right| + \left|\frac{Y_{(q')}/\sigma - \theta/\sigma + \overline{\Phi}^{-1}(\frac{q'}{n})}{\overline{\Phi}^{-1}(q'/n) - \overline{\Phi}^{-1}(q/n)}\right| .$$

Since the rescaled non contaminated observations Y_i/σ have variance 1, we can apply Theorem 3.2 to control the expectations of the above rhs term. This leads us to

$$\mathbb{E}_{\theta,\pi,\sigma}\left[\frac{\left|\widetilde{\sigma}_{q_k,q'_k} - \sigma\right|}{\sigma}\right] \lesssim \frac{\log\left(\frac{n}{n-k}\right)}{\log^{1/2}\left(1 + \frac{k^2}{n}\right)} \cdot \frac{1}{\overline{\Phi}^{-1}(q'_k/n) - \overline{\Phi}^{-1}(q_k/n)} . \tag{105}$$

It remains to derive a lower bound of the difference in the denominator. We claim that

$$\overline{\Phi}^{-1}(q'_k/n) - \overline{\Phi}^{-1}(q_k/n) \gtrsim \sqrt{\log\left(\frac{k}{\sqrt{n}}\right) \vee 1} , \tag{106}$$

which together with previous bound leads to (39). Let us show this claim. When $k/\sqrt{n}$ is smaller than some constant $\bar{c}$ that will be fixed later, the difference is lower bounded by an absolute constant (depending on $\bar{c}$) by the first inequality in (123) in Lemma D.5. If $\bar{c}$ is chosen large enough, we have, $q'_k \leq q_k \leq 0.004\,n$ for $k \geq \bar{c}\sqrt{n}$. The ratio q_k/q'_k is larger than $k/\sqrt{n}$. The third inequality in (123) together with (121) then implies that

$$\begin{aligned}
\overline{\Phi}^{-1}(q'_k/n) - \overline{\Phi}^{-1}(q_k/n) &\geq \frac{1}{\overline{\Phi}^{-1}(q'_k/n)}\left[\log\left(\frac{q_k}{eq'_k}\right) + \frac{1}{2}\log\log\left(\frac{n}{q_k}\right) - \frac{1}{2}\log\log\left(\frac{n}{q'_k}\right)\right] \\
&\gtrsim \frac{1}{\sqrt{\log(\frac{k}{\sqrt{n}})_+}}\left[\log\left(\frac{k}{\sqrt{n}}\right) - c' - \log\log\left(\frac{k}{\sqrt{n}}\right)\right] ,
\end{aligned}$$

where c' is some constant. Since $k/\sqrt{n} \geq \bar{c}$, the first logarithmic term is larger than the remaining expressions in the rhs, and we obtain (106).

We now consider the estimator $\widetilde{\theta}_{q_k,q'_k}$. We start from the decomposition

$$\mathbb{E}_{\theta,\pi,\sigma}\left[\frac{|\widetilde{\theta}_{q_k,q'_k} - \theta|}{\sigma}\right] \leq \mathbb{E}_{\theta,\pi,\sigma}\left[\frac{|\widetilde{\theta}_{q_k} - \theta|}{\sigma}\right] + \mathbb{E}_{\theta,\pi,\sigma}\left[\frac{|\widetilde{\sigma}_{q_k,q'_k} - \sigma|}{\sigma}\right]\left|\overline{\Phi}^{-1}\left(\frac{q_k}{n}\right)\right| .$$

The first expectation in the rhs has been controlled in Corollary 3.3 whereas the second expectation has been handled in the first part of this proof. We deduce from Lemma D.4 that $\overline{\Phi}^{-1}\left(\frac{q_k}{n}\right) \lesssim \sqrt{\log(n/q_k)_+ \vee 1} \lesssim \sqrt{\log(\frac{k^2}{n})_+ \vee 1}$. Putting everything together leads to the desired result.

□

B.6. Proof of Proposition 5.1 (OSC)

Proof. First note that $q_n/n \asymp n^{-1/4}$ while $q'_n/(n-k_0) \asymp n^{-3/4}$, because we assume $k_0 \leq 0.9n$. This implies that $\overline{\Phi}^{-1}(q'_n/(n-k_0)) \geq \overline{\Phi}^{-1}(q_n/n)$ for n large enough. Also by Lemma D.5, we have

$$\overline{\Phi}^{-1}(q'_n/(n-k_0)) - \overline{\Phi}^{-1}(q_n/n) \asymp \log^{1/2}(n) \ .$$

Now use the following decomposition:

$$\frac{\widetilde{\sigma}_+ - \sigma}{\sigma} = \frac{Y_{(q_n)}/\sigma - \theta/\sigma + \overline{\Phi}^{-1}(q_n/n)}{\overline{\Phi}^{-1}(q'_n/(n-k_0)) - \overline{\Phi}^{-1}(q_n/n)} - \frac{Y_{(q'_n)}/\sigma - \theta/\sigma + \overline{\Phi}^{-1}(q'_n/(n-k_0))}{\overline{\Phi}^{-1}(q'_n/(n-k_0)) - \overline{\Phi}^{-1}(q_n/n)}$$

$$= \qquad\qquad T_1 \qquad\qquad\qquad - \qquad\qquad T_2 \ ,$$

We consider separately the deviations of T_1 and T_2. We apply Theorem 3.2 to T_1. Hence, for some constant $c > 0$ and for all $x \in (0, c_3 q_n)$, we have

$$\mathbb{P}_{\theta,\pi,\sigma}\left(T_1 < -c\frac{\sqrt{x}}{n^{3/8}\log(n)}\right) \leq \mathbb{P}_{\theta,\pi,\sigma}\left([\overline{\Phi}^{-1}(q'_n/(n-k_0)) - \overline{\Phi}^{-1}(q_n/n)]T_1 < -c'\sqrt{\frac{x}{q_n\log(n)}}\right)$$

$$\leq 2e^{-x} \ .$$

For T_2, we start from

$$Y_{(q'_n)}/\sigma - \theta/\sigma + \overline{\Phi}^{-1}(q'_n/(n-k_0)) \leq Y_{(q'_n)}/\sigma - \theta/\sigma + \overline{\Phi}^{-1}(q'_n/n_0)$$

$$\preceq \xi_{(q'_n:n_0)} + \overline{\Phi}^{-1}(q'_n/n_0) \ .$$

Then, we use (127) to derive that there exists constants c' and $\overline{c}'$ such that, for all $x \in (0, n^{1/4})$,

$$\mathbb{P}_{\theta,\pi,\sigma}\left(T_2 > c'\frac{\sqrt{x}}{n^{1/8}\log(n)}\right) \leq \mathbb{P}\left(\xi_{(q'_n:n_0)} + \overline{\Phi}^{-1}(q'_n/n_0) > \overline{c}'\sqrt{\frac{x}{q'_n\log(n)}}\right) \leq e^{-x} \ ,$$

Combining the two bounds leads to the following deviation inequality,

$$\mathbb{P}_{\theta,\pi,\sigma}\left(\frac{\widetilde{\sigma}_+ - \sigma}{\sigma} < -c''\frac{\sqrt{x}}{n^{1/8}\log(n)}\right) \leq 3e^{-x} \ , \tag{107}$$

holding for all $x \in (0, n^{1/4})$. We obtain (48) by taking $x = (c'')^{-2}n^{1/8}\log^2(n)$ in (107).

Relation (47) is obtained similarly by using the decomposition:

$$\frac{\widetilde{\theta}_+ - \theta}{\sigma} = Y_{(q_n)}/\sigma - \theta/\sigma + \overline{\Phi}^{-1}(q_n/n) + \frac{\widetilde{\sigma}_+ - \sigma}{\sigma}\overline{\Phi}^{-1}\left(\frac{q_n}{n}\right) \ . \tag{108}$$

This gives that for some constant $c''' > 0$, for all $x \in (0, n^{1/4})$,

$$\mathbb{P}_{\theta,\pi,\sigma}\left(\frac{\widetilde{\theta}_+ - \theta}{\sigma} < -c'''\sqrt{x}n^{-1/8}\log^{-1/2}(n)\right) \leq e^{-x} \ , \tag{109}$$

which, for $x = (c''')^{-2}n^{1/8}\log(n)$ leads to (47).

Let us now establish (50). By (107), we only have to study the probabilities of overestimation. As in the first part of the proof, we consider separately T_1 and T_2. First, Theorem 3.2 (used with $k = k_0$), gives that for some constant $c > 0$, for all $x \in (0, c_3 q_n)$,

$$
\mathbb{P}_{\theta,\pi,\sigma} \left(T_1 > c \frac{k_0}{n \log(n)} + c \frac{\sqrt{x}}{n^{3/8} \log(n)} \right)
$$
$$
\leq \mathbb{P}_{\theta,\pi,\sigma} \left(Y_{(q_n)}/\sigma - \theta/\sigma + \overline{\Phi}^{-1}(q_n/n) > \frac{\overline{c}\, k_0}{n\sqrt{\log(n)}} + \overline{c} \sqrt{\frac{x}{n^{3/4} \log(n)}} \right) \leq 2 e^{-x} .
$$

Turning to T_2, we start by controlling the difference of quantiles with (122):

$$
\overline{\Phi}^{-1}(q_n'/n) - \overline{\Phi}^{-1}(q_n'/(n - k_0)) \lesssim \frac{q_n'/(n - k_0) - q_n'/n}{\frac{q_n'}{n} \log^{1/2}(n)} \lesssim \frac{k_0}{n \log^{1/2}(n)} .
$$

Then, by stochastic domination , we have for some constant $c_0 > 0$,

$$
Y_{(q_n')}/\sigma - \theta/\sigma + \overline{\Phi}^{-1}(q_n'/(n - k_0)) \geq Y_{(q_n')}/\sigma - \theta/\sigma + \overline{\Phi}^{-1}(q_n'/n) - c_0 \frac{k_0}{n \log^{1/2}(n)}
$$
$$
\succeq \xi_{(q_n')} + \overline{\Phi}^{-1}(q_n'/n) - c_0 \frac{k_0}{n \log^{1/2}(n)} .
$$

Putting the above inequalities together and relying on the deviation bound (128) leads to

$$
\mathbb{P}_{\theta,\pi,\sigma} \left(T_2 < -c' \frac{k_0}{n \log(n)} - c_0 \frac{\sqrt{x}}{n^{1/8} \log(n)} \right)
$$
$$
\leq \mathbb{P} \left(\frac{\xi_{(q_n')} + \overline{\Phi}^{-1}(q_n'/n)}{\overline{\Phi}^{-1}(q_n'/(n - k_0)) - \overline{\Phi}^{-1}(q_n/n)} < -c' \frac{\sqrt{x}}{n^{1/8} \log(n)} \right) \leq e^{-x} ,
$$

for all $x \in (0, q_n'/8)$. Combining the two deviation inequalities for T_1 and T_2 gives that for some constants $c'', c_4 > 0$, for all $x \in (0, c_4 n^{1/4})$,

$$
\mathbb{P}_{\theta,\pi,\sigma} \left(\frac{\widetilde{\sigma}_+ - \sigma}{\sigma} > c'' \frac{k_0}{n \log(n)} + c'' \frac{\sqrt{x}}{n^{1/8} \log(n)} \right) \leq 3 e^{-x} .
$$

Choosing $x = (c'')^{-2} n^{1/8} \log^2(n)$ leads to (50).

Finally, (49) follows from the decomposition (108), the relation on T_1 and (50). □

Appendix C: Proofs for multiple testing and post hoc bounds

In these proofs, to lighten the notation, the subscript α will be sometimes dropped in $\widehat{\ell}_\alpha(u, s), \widehat{t}_\alpha(u, s)$; the parameters θ, π, σ are removed in $\mathbb{P}_{\theta,\pi,\sigma}$ and $\mathbb{E}_{\theta,\pi,\sigma}$ and $\widetilde{\theta}_+$ (resp. $\widetilde{\sigma}_+$) are denoted by $\hat{\theta}$ (resp. $\hat{\sigma}$). We also let $\delta_n = n^{-1/16}$, so that, by Proposition 5.1, we have $\mathbb{P}(\hat{\theta} - \theta \leq -\sigma \delta_n) \leq c/n$ and $\mathbb{P}(\hat{\sigma} - \sigma \leq -\sigma \delta_n) \leq c/n$, for some constant $c > 0$.

C.1. Proof of Theorem 5.2

We start with a key observation. For $i \in \{1, \ldots, n\}$, the quantity $Y^{(i)}_{(q)}$ denotes the q-smallest element of $\{Y_j, 1 \le j \le n, j \ne i\}$ and

$$
\begin{cases}
\hat{\theta}^{(i)} = Y^{(i)}_{(q_n)} + \hat{\sigma}^{(i)} \, \overline{\Phi}^{-1}(q_n/n) \; ; \\
\hat{\sigma}^{(i)} = \dfrac{Y^{(i)}_{(q_n)} - Y^{(i)}_{(q'_n)}}{\overline{\Phi}^{-1}(q'_n/(0.1n)) - \overline{\Phi}^{-1}(q_n/n)} \; ,
\end{cases}
$$

so that the estimators $\hat{\theta}^{(i)}$ and $\hat{\sigma}^{(i)}$ are independent of Y_i.

Claim: For any n large enough and for any $t \in (0, \alpha]$, we have

$$
\left\{ p_i(\hat{\theta}, \hat{\sigma}) \le t \right\} = \left\{ p_i(\hat{\theta}, \hat{\sigma}) \le t, \hat{\theta} = \hat{\theta}^{(i)}, \hat{\sigma} = \hat{\sigma}^{(i)} \right\} = \left\{ p_i(\hat{\theta}^{(i)}, \hat{\sigma}^{(i)}) \le t \right\} . \tag{110}
$$

Proof of the claim. Consider any i such that $p_i(\hat{\theta}, \hat{\sigma}) \le t$. By definition (41) of the p-values, we have $Y_i - Y_{(q_n)} \ge \hat{\sigma} \big[\overline{\Phi}^{-1}(q_n/n) + \overline{\Phi}^{-1}(\alpha) \big]$ which is positive for n large enough. This entails $Y_{(q_n)} = Y^{(i)}_{(q_n)}$, $Y_{(q'_n)} = Y^{(i)}_{(q'_n)}$ and therefore $\hat{\theta} = \hat{\theta}^{(i)}$, $\hat{\sigma} = \hat{\sigma}^{(i)}$.

Conversely, if $p_i(\hat{\theta}^{(i)}, \hat{\sigma}^{(i)}) \le t$, we have $Y_i - Y^{(i)}_{(q_n)} \ge \hat{\sigma}^{(i)} \big[\overline{\Phi}^{-1}(q_n/n) + \overline{\Phi}^{-1}(\alpha) \big] > 0$ for n large enough. This also leads to $Y_{(q_n)} = Y^{(i)}_{(q_n)}$, $Y_{(q'_n)} = Y^{(i)}_{(q'_n)}$. We have proved (110). $\qquad\square$

The latter property can be suitably combined with Lemma C.1 (see the notation therein) to give the following equalities:

$$
\begin{aligned}
\left\{ p_i(\hat{\theta}, \hat{\sigma}) \le \alpha \widehat{\ell}(\hat{\theta}, \hat{\sigma})/n \right\} &= \left\{ p_i(\hat{\theta}, \hat{\sigma}) \le \alpha \widehat{\ell}^{(i)}(\hat{\theta}, \hat{\sigma})/n \right\} \\
&= \left\{ p_i(\hat{\theta}^{(i)}, \hat{\sigma}^{(i)}) \le \alpha \widehat{\ell}^{(i)}(\hat{\theta}, \hat{\sigma})/n, \hat{\theta} = \hat{\theta}^{(i)}, \hat{\sigma} = \hat{\sigma}^{(i)} \right\} \\
&= \left\{ p_i(\hat{\theta}^{(i)}, \hat{\sigma}^{(i)}) \le \alpha \widehat{\ell}^{(i)}(\hat{\theta}^{(i)}, \hat{\sigma}^{(i)})/n, \hat{\theta} = \hat{\theta}^{(i)}, \hat{\sigma} = \hat{\sigma}^{(i)} \right\} \\
&= \left\{ p_i(\hat{\theta}^{(i)}, \hat{\sigma}^{(i)}) \le \alpha \widehat{\ell}^{(i)}(\hat{\theta}^{(i)}, \hat{\sigma}^{(i)})/n \right\} \\
&\subset \left\{ \widehat{\ell}^{(i)}(\hat{\theta}^{(i)}, \hat{\sigma}^{(i)}) = \widehat{\ell}(\hat{\theta}, \hat{\sigma}) \right\} .
\end{aligned} \tag{111}
$$

The first equality is a direct application of Lemma C.1, used with $u = \hat{\theta}$ and $s = \hat{\sigma}$. The second equality comes from (110) used with $t = \alpha \widehat{\ell}^{(i)}(\hat{\theta}, \hat{\sigma})/n$. The third equality is trivial. The fourth equality comes from (110) used with $t = \alpha \widehat{\ell}^{(i)}(\hat{\theta}^{(i)}, \hat{\sigma}^{(i)})/n$. Now, since $\mathrm{BH}_\alpha(\hat{\theta}, \hat{\sigma}) = \{1 \le i \le n : p_i(\hat{\theta}, \hat{\sigma}) \le \alpha \hat{\ell}(\hat{\theta}, \hat{\sigma})/n\}$, we have

$$
\begin{aligned}
&\mathrm{FDP}(\pi, \mathrm{BH}_\alpha(\hat{\theta}, \hat{\sigma})) \, \mathbb{1}\{\hat{\theta} - \theta > -\sigma\delta_n, \hat{\sigma} - \sigma > -\sigma\delta_n\} \\
&= \sum_{i \in \mathcal{H}_0} \frac{\mathbb{1}\{p_i(\hat{\theta}, \hat{\sigma}) \le \alpha \hat{\ell}(\hat{\theta}, \hat{\sigma})/n\}}{\hat{\ell}(\hat{\theta}, \hat{\sigma}) \vee 1} \, \mathbb{1}\{\hat{\theta}^{(i)} - \theta > -\sigma\delta_n, \hat{\sigma}^{(i)} - \sigma > -\sigma\delta_n\} \\
&= \sum_{i \in \mathcal{H}_0} \frac{\mathbb{1}\{p_i(\hat{\theta}^{(i)}, \hat{\sigma}^{(i)}) \le \alpha \hat{\ell}^{(i)}(\hat{\theta}^{(i)}, \hat{\sigma}^{(i)})/n\}}{\hat{\ell}^{(i)}(\hat{\theta}^{(i)}, \hat{\sigma}^{(i)})} \, \mathbb{1}\{\hat{\theta}^{(i)} - \theta > -\sigma\delta_n, \hat{\sigma}^{(i)} - \sigma > -\sigma\delta_n\} \, ,
\end{aligned}
$$

by applying (110) and (111). By integration,

$$\mathbb{E}\left[\mathrm{FDP}(\pi,\mathrm{BH}_\alpha(\hat{\theta},\hat{\sigma}))\,\mathbb{1}\{\hat{\theta}-\theta>-\sigma\delta_n,\hat{\sigma}-\sigma>-\sigma\delta_n\}\right]$$

$$=\sum_{i\in\mathcal{H}_0}\mathbb{E}\left[\frac{\mathbb{1}\{p_i(\hat{\theta}^{(i)},\hat{\sigma}^{(i)})\leq\alpha\hat{\ell}^{(i)}(\hat{\theta}^{(i)},\hat{\sigma}^{(i)})/n\}}{\hat{\ell}^{(i)}(\hat{\theta}^{(i)},\hat{\sigma}^{(i)})}\,\mathbb{1}\{\hat{\theta}^{(i)}-\theta>-\sigma\delta_n,\hat{\sigma}^{(i)}-\sigma>-\sigma\delta_n\}\right]$$

$$=\sum_{i\in\mathcal{H}_0}\mathbb{E}\left[\frac{\mathbb{1}\{\hat{\theta}^{(i)}-\theta>-\sigma\delta_n,\hat{\sigma}^{(i)}-\sigma>-\sigma\delta_n\}}{\hat{\ell}^{(i)}(\hat{\theta}^{(i)},\hat{\sigma}^{(i)})}\mathbb{P}\left[p_i(\hat{\theta}^{(i)},\hat{\sigma}^{(i)})\leq\alpha\hat{\ell}^{(i)}(\hat{\theta}^{(i)},\hat{\sigma}^{(i)})/n\mid Y_j,j\neq i\right]\right]$$

$$=\sum_{i\in\mathcal{H}_0}\mathbb{E}\left[\frac{\mathbb{1}\{\hat{\theta}^{(i)}-\theta>-\sigma\delta_n,\hat{\sigma}^{(i)}-\sigma>-\sigma\delta_n\}}{\hat{\ell}^{(i)}(\hat{\theta}^{(i)},\hat{\sigma}^{(i)})}U_{\hat{\theta}^{(i)},\hat{\sigma}^{(i)}}(\alpha\hat{\ell}^{(i)}(\hat{\theta}^{(i)},\hat{\sigma}^{(i)})/n)\right]\,,$$

by independence between the Y_i's and by (43), because the perfectly corrected p-values (42) are uniformly distributed on $(0,1)$. Now, using that $U_{u,s}(t)$ is nonincreasing both in u and s for $t<1/2$, the last display is smaller than

$$\sum_{i\in\mathcal{H}_0}\mathbb{E}\left[\frac{U_{\theta-\sigma\delta_n,\sigma-\sigma\delta_n}(\alpha\hat{\ell}^{(i)}(\hat{\theta}^{(i)},\hat{\sigma}^{(i)})/n)}{\hat{\ell}^{(i)}(\hat{\theta}^{(i)},\hat{\sigma}^{(i)})}\right]\leq\frac{\alpha}{n}\sum_{i\in\mathcal{H}_0}\left[\max_{\alpha/n\leq t\leq\alpha}\frac{U_{\theta-\sigma\delta_n,\sigma-\sigma\delta_n}(t)}{t}\right]$$

$$=\alpha\frac{n_0}{n}\left(1+\max_{\alpha/n\leq t\leq\alpha}\left\{\frac{U_{\theta-\sigma\delta_n,\sigma-\sigma\delta_n}(t)-t}{t}\right\}\right)\,.$$

The first inequality comes from $\hat{\ell}^{(i)}(\hat{\theta}^{(i)},\hat{\sigma}^{(i)})\geq1$, which holds because the BH procedure always rejects a null hypothesis corresponding to a zero p-value (see the notation of Lemma C.1). The result (52) by is then a consequence of Lemma C.4 and Proposition 5.1.

Let us now turn to the second statement (53) and assume $n_1\geq1$ (otherwise the result is trivial). Remember that we have $n_1/n\asymp k_0/n$. Thus, Proposition 5.1 implies, for some constant $c>0$, $x_n=c\big((n_1/n)\log^{-1/2}(n)+n^{-1/16}\big)$ and $y_n=c\big((n_1/n)\log^{-1}(n)+n^{-1/16}\big)$, the deviation inequalities $\mathbb{P}\left(|\hat{\theta}-\theta|\geq\sigma x_n\right)\leq c/n$ and $\mathbb{P}\left(|\hat{\sigma}-\sigma|\geq\sigma y_n\right)\leq c/n$. Denote the event

$$\mathcal{A}=\left\{\hat{\theta}-\theta\leq\sigma x_n;\quad\hat{\sigma}-\sigma\leq\sigma y_n;\quad\hat{\sigma}\geq\sigma/2\right\}\,,$$

so that $\mathbb{P}(\mathcal{A}^c)\lesssim1/n$. For any $\eta>0$, we have

$$\mathbb{E}\left(\mathrm{TDP}(\pi,\mathrm{BH}_\alpha^\star)\right)\leq\eta+\int_\eta^1\mathbb{P}\left(\mathrm{TDP}(\pi,\mathrm{BH}_\alpha(\theta,\sigma))\geq u\right)du$$

$$\leq\eta+\int_\eta^1\mathbb{P}\left(\mathrm{TDP}(\pi,\mathrm{BH}_\alpha(\theta,\sigma))\geq u,\mathcal{A}\right)du+\mathbb{P}(\mathcal{A}^c)\,.\qquad(112)$$

Consider the event $\mathcal{A}\cap\{\mathrm{TDP}(\pi,\mathrm{BH}_\alpha(\theta,\sigma))\geq\eta\}$. By definition (10) of the TDP, when this event holds, we have $\hat{\ell}_\alpha(\theta,\sigma)\geq\eta n_1$. Write $t_0=\alpha\eta n_1/n$. Invoking Lemma C.3, we obtain $\hat{\ell}_{\alpha_0}(\hat{\theta},\hat{\sigma})\geq\hat{\ell}_\alpha(\theta,\sigma)\geq1$, for $\alpha_0>0$ such that

$$\frac{\alpha_0}{\alpha}=\frac{U_{\hat{\theta},\hat{\sigma}}^{-1}(\hat{t}_\alpha(\theta,\sigma))}{\hat{t}_\alpha(\theta,\sigma)}=\frac{\overline{\Phi}\left(\overline{\Phi}^{-1}\left(\hat{t}_\alpha(\theta,\sigma)\right)-\frac{\hat{\sigma}-\sigma}{\hat{\sigma}}\overline{\Phi}^{-1}\left(\hat{t}_\alpha(\theta,\sigma)\right)+\frac{\theta-\hat{\theta}}{\hat{\sigma}}\right)}{\hat{t}_\alpha(\theta,\sigma)}$$

$$\leq\sup_{t\in[t_0,\alpha]}\left\{\frac{\overline{\Phi}\left(\overline{\Phi}^{-1}(t)-\frac{2y_n\sigma}{\sigma}\overline{\Phi}^{-1}(t)-\frac{2x_n\sigma}{\sigma}\right)}{t}\right\}=1+\sup_{t\in[t_0,\alpha]}\left\{\frac{U_{\theta-2x_n\sigma,\sigma-2y_n\sigma}(t)-t}{t}\right\}\,.$$

Now using Lemma C.4, we get

$$\frac{\alpha_0 - \alpha}{\alpha} \lesssim x_n \log^{1/2}(\tfrac{1}{t_0}) + y_n \log(\tfrac{1}{t_0}) \ ,$$

as soon as this upper-bound is smaller than some constant small enough. But now

$$x_n \log^{1/2}(\tfrac{1}{t_0}) + y_n \log(\tfrac{1}{t_0}) \lesssim n^{-1/16} \log\big(\frac{n}{\alpha\eta n_1}\big) + \frac{n_1}{n} \log^{-1/2}(n) \log\big(\frac{n}{n_1\alpha\eta}\big) \ .$$

Since $\sup_{x\in(0,1)}(x\log(1/x)) = e^{-1}$ and since $\epsilon_n \gg \log^{-1/2}(n)$, we have for n large enough (as a function of α and η),

$$\frac{\alpha_0 - \alpha}{\alpha} \leq \epsilon_n \ .$$

As a consequence, on the event $\mathcal{A} \cap \{\mathrm{TDP}(\pi, \mathrm{BH}_\alpha(\theta, \sigma)) \geq \eta\}$, we have $\widehat{\ell}_{\alpha(1+\epsilon_n)}(\widehat{\theta}, \widehat{\sigma}) \geq \widehat{\ell}_{\alpha_0}(\widehat{\theta}, \widehat{\sigma}) \geq \widehat{\ell}_\alpha(\theta, \sigma)$, and thus $\mathrm{TDP}(\pi, \mathrm{BH}_{\alpha(1+\epsilon_n)}(\widehat{\theta}, \widehat{\sigma})) \geq \mathrm{TDP}(\pi, \mathrm{BH}_\alpha(\theta, \sigma))$ for n large enough. Coming back to (112), we obtain for n large enough,

$$\mathbb{E}\left(\mathrm{TDP}(\pi, \mathrm{BH}_\alpha(\theta))\right) \leq \eta + \int_\eta^1 \mathbb{P}\left(\mathrm{TDP}(\pi, \mathrm{BH}_{\alpha(1+\epsilon_n)}(\widehat{\theta}, \widehat{\sigma})) \geq u\right) du + \mathbb{P}(\mathcal{A}^c)$$

$$\leq \eta + \mathbb{E}\left(\mathrm{TDP}(\pi, \mathrm{BH}_{\alpha(1+\epsilon_n)}(\widehat{\theta}, \widehat{\sigma}))\right) + \mathbb{P}(\mathcal{A}^c) \ .$$

As a result,

$$\limsup_n \{\mathbb{E}\left(\mathrm{TDP}(\pi, \mathrm{BH}_\alpha(\theta))\right) - \mathbb{E}\left(\mathrm{TDP}(\pi, \mathrm{BH}_{\alpha(1+\epsilon_n)}(\widehat{\theta}, \widehat{\sigma}))\right)\} \leq \eta.$$

This gives the result by making η tends to 0.

C.2. Proof of Theorem 5.3

By using (43), we obtain

$$\mathbb{P}\left[\exists \ell \in \{1, \ldots, n_0\} \ : \ p_{(\ell:\mathcal{H}_0)}(\widehat{\theta}, \widehat{\sigma}) \leq \alpha\ell/n, \ \widehat{\theta} - \theta > -\sigma\delta_n, \ \widehat{\sigma} - \sigma > -\sigma\delta_n\right]$$

$$= \mathbb{P}\left[\exists \ell \in \{1, \ldots, n_0\} \ : \ p^\star_{(\ell:\mathcal{H}_0)} \leq U_{\widehat{\theta}, \widehat{\sigma}}(\alpha\ell/n), \ \widehat{\theta} - \theta > -\sigma\delta_n, \ \widehat{\sigma} - \sigma > -\sigma\delta_n\right]$$

$$\leq \mathbb{P}\left[\exists \ell \in \{1, \ldots, n_0\} \ : \ p^\star_{(\ell:\mathcal{H}_0)} \leq U_{\theta-\sigma\delta_n, \sigma-\sigma\delta_n}(\alpha\ell/n)\right] \ .$$

because the quantity $U_{u,s}(\alpha\ell/n)$ is nonincreasing both in u and s (since $\overline{\Phi}^{-1}(\alpha) \geq 0$). Now using the classical Simes inequality (54), the last display is upper-bounded by

$$\mathbb{P}\left[\exists \ell \in \{1, \ldots, n_0\} \ : \ p^\star_{(\ell:\mathcal{H}_0)} \leq \max_{\alpha/n \leq t \leq \alpha}\left\{\frac{U_{\theta-\sigma\delta_n, \sigma-\sigma\delta_n}(t)}{t}\right\} \alpha\ell/n\right]$$

$$\leq \max_{\alpha/n \leq t \leq \alpha}\left\{\frac{U_{\theta-\sigma\delta_n, \sigma-\sigma\delta_n}(t)}{t}\right\} \alpha = \alpha + \alpha \max_{\alpha/n \leq t \leq \alpha}\left\{\frac{U_{\theta-\sigma\delta_n, \sigma-\sigma\delta_n}(t) - t}{t}\right\}$$

$$\leq \alpha + c\delta_n \log(n) \ ,$$

for some constant $c > 0$, by applying Lemma C.4 and Proposition 5.1.

C.3. Proof of Corollary 5.4

Denote

$$R_\ell = \{i \in \{1, \ldots, n\} \, : \, p_i(\hat\theta, \hat\sigma) \leq \alpha\ell/n\}, \ 1 \leq \ell \leq n \ .$$

From (55), with probability at least $1 - \alpha - c\log(n)/n^{1/16}$, the following event holds true:

$$\mathcal{E} = \{\forall \ell \in \{1, \ldots, n_0\} \, , \, p_{(\ell:\mathcal{H}_0)}(\hat\theta, \hat\sigma) > \alpha\ell/n\}$$
$$= \{\forall \ell \in \{1, \ldots, n_0\} \, , \, |R_\ell \cap \mathcal{H}_0| \leq \ell - 1\} \ ,$$

Now, on the event $\mathcal{E}$, we have for any $S \subset \{1, \ldots, n\}$, and for any $\ell \in \{1, \ldots, n\}$,

$$|S \cap \mathcal{H}_0| = |S \cap \mathcal{H}_0 \cap R_\ell| + |S \cap \mathcal{H}_0 \cap R_\ell^c|$$
$$\leq |\mathcal{H}_0 \cap R_\ell| + |S \cap R_\ell^c|$$
$$\leq \ell - 1 + |S \cap R_\ell^c| \ .$$

By taking the minimum in ℓ in the latter relation, we have

$$|S \cap \mathcal{H}_0| \leq \min_{\ell \in \{1, \ldots, n\}} \{\ell - 1 + |S \cap R_\ell^c|\},$$

so that

$$\mathrm{FDP}(\pi, S) = \frac{|S \cap \mathcal{H}_0|}{|S| \vee 1} \leq \frac{\min_{\ell \in \{1, \ldots, n\}} \{\ell - 1 + |S \cap R_\ell^c|\}}{|S| \vee 1},$$

which yields (56).

C.4. Auxiliary lemmas

The next lemma is a well known property of step-up procedure in the multiple testing theory, see, e.g., [31] and Lemma 7.1 in [69].

Lemma C.1. *Consider arbitrary $u \in \mathbb{R}$, $s > 0$ and $\alpha \in (0, 1)$. The rejection number $\widehat\ell(u, s)$ of the Benjamini-Hochberg procedure $BH_\alpha(u, s)$ (defined in Section 1.3.1) satisfies the following: for all $i \in \{1, \ldots, n\}$,*

$$\left\{p_i(u, s) \leq \alpha\widehat\ell(u, s)/n\right\} = \left\{p_i(u, s) \leq \alpha\widehat\ell^{(i)}(u, s)/n\right\} = \left\{\widehat\ell^{(i)}(u, s) = \widehat\ell(u, s)\right\} \ ,$$

where $\widehat\ell^{(i)}(u, s)$ is the rejection number of the Benjamini-Hochberg procedure applied to the p-value set $\{0, p_j(u, s), j \neq i\}$, that is, to the p-value set where $p_i(u, s)$ has been replaced by 0.

Lemma C.2. *Fix $\theta \in \mathbb{R}$ and $\sigma > 0$ and $U.(\cdot)$ as in (44). Consider arbitrary $u \in \mathbb{R}$, $s > 0$ and $\alpha \in (0, 1)$. Then*

$$\widehat t_\alpha(u, s) = \max\left\{t \in [0, 1] \, : \, \widehat G_{u,s}(t) \geq t/\alpha\right\} \ ,$$

where $\widehat G_{u,s}(t) = n^{-1} \sum_{i=1}^n \mathbb{1}_{\{p_i(u,s) \leq t\}} = n^{-1} \sum_{i=1}^n \mathbb{1}_{\{p_i(\theta,\sigma) \leq U_{u,s}(t)\}} = \widehat G_{\theta,\sigma}(U_{u,s}(t))$

Lemma C.3. *Fix $\theta \in \mathbb{R}$ and $\sigma > 0$ and $U.(\cdot)$ as in (44). Consider arbitrary $u \in \mathbb{R}$, $s > 0$ and $\alpha \in (0,1)$. Then*

$$\widehat{\ell}_{\alpha_0}(u,s) \geq \widehat{\ell}_{\alpha}(\theta,\sigma), \quad \text{where} \quad \alpha_0 = \alpha \frac{U_{u,s}^{-1}(\widehat{t}_{\alpha}(\theta,\sigma))}{\widehat{t}_{\alpha}(\theta,\sigma)} .$$

Proof. Denoting $t_0 = \widehat{t}_{\alpha}(\theta,\sigma)$ and invoking Lemma C.2, we have

$$\widehat{G}_{u,s}\left(U_{u,s}^{-1}(t_0)\right) = \widehat{G}_{\theta,\sigma}(t_0) \geq t_0/\alpha = \frac{t_0}{U_{u,s}^{-1}(t_0)} \frac{U_{u,s}^{-1}(t_0)}{\alpha} = \frac{U_{u,s}^{-1}(t_0)}{\alpha_0} .$$

By using again Lemma C.2, this gives $\widehat{t}_{\alpha_0}(u,s) \geq U_{u,s}^{-1}(t_0)$. Hence, $\widehat{t}_{\alpha_0}(u,s) \geq \frac{\alpha_0}{\alpha} \widehat{t}_{\alpha}(\theta,\sigma)$, which gives the result. $\square$

Lemma C.4. *There exists a universal constant $c > 0$ such that the following holds. For all $\alpha \in (0,0.4)$, for all $x, y \geq 0$ and $t_0 \in (0,\alpha)$, we have*

$$\max_{t_0 \leq t \leq \alpha} \left\{ \frac{U_{\theta-x,\sigma-y}(t) - t}{t} \right\} \leq c \left(\frac{x}{\sigma} (2\log(1/t_0))^{1/2} + \frac{y}{\sigma} 2\log(1/t_0) \right) . \tag{113}$$

provided that $(x/\sigma)(2\log(1/t_0))^{1/2} + (y/\sigma)2\log(1/t_0) \leq 0.05$, and where $U_{\theta-x,\sigma-y}(\cdot)$ is defined by (44).

Proof. First note that by (44), we have

$$U_{\theta-x,\sigma-y}(t) = \overline{\Phi}\left(\overline{\Phi}^{-1}(t) - z(t)\right) , \quad z(t) = \frac{y}{\sigma}\overline{\Phi}^{-1}(t) + \frac{x}{\sigma}.$$

By Lemma D.4, we have for all $t \in [t_0, \alpha]$,

$$z(t) \leq \frac{y}{\sigma}(2\log(1/t))^{1/2} + \frac{x}{\sigma} \leq 0.05(2\log(1/t))^{-1/2} \leq 0.05/\overline{\Phi}^{-1}(t) ,$$

where we used the assumption of the lemma. Now using that $\overline{\Phi}(\sqrt{0.05}) \geq 0.4 \geq t$, we deduce $z(t) \leq \overline{\Phi}^{-1}(t)$ for all $t \in [t_0, \alpha]$. Also deduce that for such a value of t,

$$\frac{\phi\left(\overline{\Phi}^{-1}(t) - z(t)\right)}{\phi\left(\overline{\Phi}^{-1}(t)\right)} = e^{-z^2(t)/2}e^{z(t)\overline{\Phi}^{-1}(t)} \leq e^{z(t)\overline{\Phi}^{-1}(t)} \leq e^{0.05} \leq 2.$$

Now, since $\overline{\Phi}$ is decreasing and its derivative is $-\phi$, we have for all $t \in [t_0, \alpha]$

$$\overline{\Phi}(\overline{\Phi}^{-1}(t) - z(t)) - \overline{\Phi}(\overline{\Phi}^{-1}(t)) \leq z(t)\,\phi\left(\overline{\Phi}^{-1}(t) - z(t)\right)$$

$$\leq z(t)\frac{\phi\left(\overline{\Phi}^{-1}(t) - z(t)\right)}{\phi\left(\overline{\Phi}^{-1}(t)\right)}\phi\left(\overline{\Phi}^{-1}(t)\right)$$

$$\leq 2z(t)\phi\left(\overline{\Phi}^{-1}(t)\right)$$

$$\leq 2z(t)\left(1 + \left(\overline{\Phi}^{-1}(t)\right)^{-2}\right) t\,\overline{\Phi}^{-1}(t),$$

by using Lemma D.4. Finally, the last display is smaller than

$$z(t_0)\,t\,\overline{\Phi}^{-1}(t_0)2\left(1 + \left(\overline{\Phi}^{-1}(0.4)\right)^{-2}\right),$$

which gives (113). $\square$

Appendix D: Auxiliary results

D.1. Chebychev polynomials

In this subsection, we first remind the reader of the definition and important properties of Chebychev polynomials. For any $k \geq 1$, the k-th Chebychev polynomial is defined by

$$T_k(x) = (k/2) \sum_{j=0}^{\lfloor k/2 \rfloor} (-1)^j \frac{(k-j-1)!}{j!(k-2j)!} (2x)^{k-2j} .$$

It satisfies the following inequalities.

Proposition D.1. *For $k \geq 1$, the polynomial T_k satisfies the following properties:*

(i) for all $y \in [0,1]$, $T_k(1-2y) = T_{2k}(\sqrt{1-y})$, and thus for all $y \in \mathbb{R}$,

$$T_k(1-2y) = \sum_{j=0}^{k} (-4)^j \frac{k(k+j-1)!}{(k-j)!(2j)!} y^j . \tag{114}$$

(ii) for $x \in \mathbb{R}$,

$$T_k(x) = \begin{cases} \cos(k \arccos x) & \text{if } x \in [-1,1] \; ; \\ \cosh(k \operatorname{arccosh} x) & \text{if } x \geq 1 \; ; \\ (-1)^k \cosh(k \operatorname{arccosh}(-x)) & \text{if } x \leq -1 \; . \end{cases}$$

Next, we remind the reader of extremal properties satisfied by the Chebychev polynomial. The first inequality is a consequence of Chebychev's Theorem whereas the second inequality is a consequence of Markov's theorem. Both may be found in [36], Page 119.

Lemma D.2. *Denote by $\mathcal{P}_k$ the set of polynomials of degree smaller than or equal to k. Let $a, b, c \in \mathbb{R}$ with $a < b < c$. Then we have,*

$$\sup_{\substack{P \in \mathcal{P}_k \\ \|P\|_{\infty,[b,c]} \leq 1}} |P(a)| = \left| T_k \left(1 + 2\frac{b-a}{c-b} \right) \right| ,$$

$$\sup_{x \in [a,c]} |P'(x)| \leq \frac{2k^2}{c-a} \sup_{x \in [a,c]} |P(x)| \qquad \forall P \in \mathcal{P}_k .$$

The coefficient of the polynomial defined in (14) are given explicitly using the explicit expression of Chebychev polynomials:

$$a_{j,q} = (-4)^j \frac{q(q+j-1)!}{(q-j)!(2j)!}, \quad 0 \leq j \leq q , \tag{115}$$

Lemma D.3. *For any even integer q and any integer $j \in [0,q]$, we have*

$$|a_{j,q}| \leq (3+2\sqrt{2})^q \tag{116}$$

Proof. This upper bound is obviously true for $j = 0$ and $j = q$. Henceforth, we restrict ourselves to the case $j \in [1, q-1]$ (and therefore $q \geq 2$). For any positive integer n, Stirling's inequalities ensure that $n! e^n n^{-n-1/2} \in [\sqrt{2\pi}, e]$. This leads us to

$$|a_{j,q}| \leq 4^j \frac{e}{2\pi} \frac{(q+j)^{q+j+1/2}}{(q-j)^{q-j+1/2}(2j)^{2j+1/2}} \leq \frac{e}{2\pi} \sqrt{\frac{q+j}{2(q-j)j}} \frac{(q+j)^{(q+j)}}{j^{2j}(q-j)^{q-j}} .$$

Since we assume that $j \in [1, q-1]$ and since the function $x \mapsto x(1-x)$ is increasing on $(0, 1/2)$ and decreasing on $(1/2, 1)$, we obtain

$$\sqrt{\frac{q+j}{2(q-j)j}} \leq \sqrt{\frac{2q}{2q(1-1/q)}} \leq \sqrt{\frac{1}{1-1/q}} \leq \sqrt{2} \ ,$$

which leads us to

$$|a_{j,q}| \leq \frac{e}{\sqrt{2\pi}} \exp\left[q\left((1+\frac{j}{q})\log(1+\frac{j}{q}) - 2\frac{j}{q}\log(\frac{j}{q}) - (1-\frac{j}{q})\log(1-\frac{j}{q})\right)\right] \ . \tag{117}$$

Consider the function $h : x \mapsto (1+x)\log(1+x) - (1-x)\log(1-x) - 2x\log(x)$ defined on $(0, 1)$. Relying on standard derivation arguments, we observe that h achieves its maximum $x_0 = 1/\sqrt{2}$ and that $h(x_0) = \log(3 + 2\sqrt{2})$. Coming back to (117), we have proved that

$$|a_{j,q}| \leq (3 + 2\sqrt{2})^q$$

$\square$

D.2. Inequalities for Gaussian quantile and upper-tail functions

Lemma D.4 (Quantile function of the normal distribution). *We have*

$$\max\left(\frac{t\phi(t)}{1+t^2}, \frac{1}{2} - \frac{t}{\sqrt{2\pi}}\right) \leq \overline{\Phi}(t) \leq \phi(t)\min\left(\frac{1}{t}, \sqrt{\frac{\pi}{2}}\right), \quad \text{for all } t > 0 \ . \tag{118}$$

As a consequence, for any $x < 0.5$, we have

$$\sqrt{2\pi}(1/2 - x) \leq \overline{\Phi}^{-1}(x) \leq \sqrt{2\log\left(\frac{1}{2x}\right)} \ , \tag{119}$$

$$\log\left(\frac{[\overline{\Phi}^{-1}(x)]^2}{[\overline{\Phi}^{-1}(x)]^2 + 1}\right) \leq \frac{[\overline{\Phi}^{-1}(x)]^2}{2} - \log\left(\frac{1}{x}\right) + \log\left(\sqrt{2\pi}\overline{\Phi}^{-1}(x)\right) \leq 0 \ , \tag{120}$$

and if additionally $x \leq 0.004$, we have

$$\overline{\Phi}^{-1}(x) \geq \sqrt{\log\left(\frac{1}{x}\right)} \ . \tag{121}$$

Proof of Lemma D.4. Inequality (118) is standard. Relation (119) (resp. (120)) is a consequence of $1/2 - t/\sqrt{2\pi} \leq \overline{\Phi}(t) \leq \phi(t)\sqrt{\frac{\pi}{2}}$ (resp. $(t^2/(1+t^2))\phi(t)/t \leq \overline{\Phi}(t) \leq \phi(t)/t$). The last relation (121) comes from (120), because for $x \leq 0.004$ (thus $\overline{\Phi}^{-1}(x) \geq 1$), we have

$$[\overline{\Phi}^{-1}(x)]^2 \geq 2\log\left(\frac{1}{x}\right) - 2\log\left(\sqrt{2\pi}\overline{\Phi}^{-1}(x)\frac{[\overline{\Phi}^{-1}(x)]^2 + 1}{[\overline{\Phi}^{-1}(x)]^2}\right)$$

$$\geq 2\log\left(\frac{1}{x}\right) - 2\log\left(2\sqrt{2\pi}\overline{\Phi}^{-1}(x)\right) \ ,$$

which is larger than $\log\left(\frac{1}{x}\right)$ provided that $16\pi x \log\left(\frac{1}{2x}\right) \leq 1$ by (119). This last bound holds for $x \leq 0.004$ by monotonicity. $\square$

Lemma D.5. *We have*

$$\overline{\Phi}^{-1}(x) - \overline{\Phi}^{-1}(y) \leq \begin{cases} 3|y-x| & \text{if } 0.3 \leq x \leq y \leq 0.7 \text{ ;} \\ \frac{|y-x|}{x\overline{\Phi}^{-1}(x)} & \text{if } x < \min(y, 1-y) \text{ ;} \\ \frac{1}{\overline{\Phi}^{-1}(y)}\left[\log\left(\frac{y}{x}\right) + \frac{1}{[\overline{\Phi}^{-1}(y)]^2}\right] & \text{if } x \leq y < 0.5 \text{ .} \end{cases} \quad (122)$$

Besides, we also have

$$\overline{\Phi}^{-1}(x) - \overline{\Phi}^{-1}(y) \geq \begin{cases} 2.5|y-x| & \text{if } x \leq y \text{ ;} \\ \left(\frac{|\overline{\Phi}^{-1}(y)|^2}{1+|\overline{\Phi}^{-1}(y)|^2}\right)\frac{|y-x|}{y\overline{\Phi}^{-1}(y)} & \text{if } x \leq y < 0.5 \text{ ;} \\ \frac{1}{\overline{\Phi}^{-1}(x)}\left[\log\left(\frac{y}{ex}\right) + \frac{1}{2}\log\log\left(\frac{1}{y}\right) - \frac{1}{2}\log\log\left(\frac{1}{x}\right)\right] & \text{if } x \leq y \leq 0.004 \text{ .} \end{cases} \quad (123)$$

Proof of Lemma D.5. We start by proving the two first inequalities of each bound (122) and (123). By the mean-value theorem, we have

$$\frac{y-x}{\sup_{z\in[x,y]}\phi(\overline{\Phi}^{-1}(z))} \leq \overline{\Phi}^{-1}(x) - \overline{\Phi}^{-1}(y) \leq \frac{y-x}{\inf_{z\in[x,y]}\phi(\overline{\Phi}^{-1}(z))} \text{ .} \quad (124)$$

The function $t \mapsto \phi(\overline{\Phi}^{-1}(t+1/2))$ defined on $[-1/2, 1/2]$ is symmetric and increasing on $[-1/2, 0]$. Thus if $0.3 \leq x \leq y \leq 0.7$, the above infimum equals $\phi(\overline{\Phi}^{-1}(0.3))$ which is larger than $1/3$. This proves the first inequality. Turning to the second inequality ($x \leq \min(y, 1-y)$), the above infimum is achieved at $z = x$ which yields.

$$0 \leq \overline{\Phi}^{-1}(x) - \overline{\Phi}^{-1}(y) \leq \frac{|y-x|}{\phi(\overline{\Phi}^{-1}(x))} \text{ .}$$

It follows from (120), that $\phi(\overline{\Phi}^{-1}(x)) \geq x|\overline{\Phi}^{-1}(x)|$, which yields the second result. Turning to the lower bounds, we still apply the mean theorem (124) and observe that $\phi(\overline{\Phi}^{-1}(z)) \leq 1/\sqrt{2\pi} \leq (2.5)^{-1}$, which proves the first bound in (123). As for the second bound in (123), the maximum of $\phi(\overline{\Phi}^{-1}(z))$ is achieved at $z = y$ and the result follows from (120) in Lemma D.4.

Finally, we consider the last bounds in (122) and (123). We first apply the mean value theorem to the square root function.

$$\frac{[\overline{\Phi}^{-1}(x)]^2 - [\overline{\Phi}^{-1}(y)]^2}{2\overline{\Phi}^{-1}(x)} \leq \overline{\Phi}^{-1}(x) - \overline{\Phi}^{-1}(y) \leq \frac{[\overline{\Phi}^{-1}(x)]^2 - [\overline{\Phi}^{-1}(y)]^2}{2\overline{\Phi}^{-1}(y)} \text{ .}$$

For the upper bound, we use Lemma D.4 for x and y to get

$$\begin{aligned} \overline{\Phi}^{-1}(x) - \overline{\Phi}^{-1}(y) &\leq \frac{1}{\overline{\Phi}^{-1}(y)}\left[\log\left(\frac{y}{x}\right) + \log\left(\frac{\overline{\Phi}^{-1}(y)}{\overline{\Phi}^{-1}(x)}\right) - \log\left(\frac{[\overline{\Phi}^{-1}(y)]^2}{1+[\overline{\Phi}^{-1}(y)]^2}\right)\right] \\ &\leq \frac{1}{\overline{\Phi}^{-1}(y)}\left[\log\left(\frac{y}{x}\right) + \frac{1}{[\overline{\Phi}^{-1}(y)]^2}\right] \text{ .} \end{aligned}$$

For the lower bound, we use again Lemma D.4 together with $\overline{\Phi}^{-1}(x) \geq \overline{\Phi}^{-1}(0.004)$, to get

$$
\begin{aligned}
\overline{\Phi}^{-1}(x) - \overline{\Phi}^{-1}(y) &\geq \frac{1}{\overline{\Phi}^{-1}(x)} \left[\log\left(\frac{y}{x}\right) + \log\left(\frac{\overline{\Phi}^{-1}(y)}{\overline{\Phi}^{-1}(x)}\right) + \log\left(\frac{[\overline{\Phi}^{-1}(x)]^2}{1 + [\overline{\Phi}^{-1}(x)]^2}\right) \right] \\
&\geq \frac{1}{\overline{\Phi}^{-1}(x)} \left[\log\left(\frac{y}{x}\right) + \frac{1}{2}\log\left(\frac{\log(1/y)}{2\log(1/x)}\right) + \log\left(\frac{[\overline{\Phi}^{-1}(0.004)]^2}{1 + [\overline{\Phi}^{-1}(0.004)]^2}\right) \right] \\
&\geq \frac{1}{\overline{\Phi}^{-1}(x)} \left[\log\left(\frac{y}{x}\right) + \frac{1}{2}\log\log\left(\frac{1}{y}\right) - \frac{1}{2}\log\log\left(\frac{1}{x}\right) - 1 \right] ,
\end{aligned}
$$

which concludes the proof. $\qquad\square$

Lemma D.6. *There exists some universal constant c such that the following holds for all positive integers n, $k \leq n-1$ and all positive integers $q \leq 0.7(n-k)$:*

$$
\overline{\Phi}^{-1}(q/n) - \overline{\Phi}^{-1}(q/(n-k)) \leq c \frac{\log\left(\frac{n}{n-k}\right)}{\sqrt{\log\left(\frac{n-k}{q}\right)_+ \vee 1}} .
$$

Proof. We first consider the case $q \geq n/3$. Since $q \leq 0.7(n-k)$, it follows that $n \leq 2.1(n-k)$. We then deduce from (122) that

$$
\overline{\Phi}^{-1}\left(\frac{q}{n}\right) - \overline{\Phi}^{-1}\left(\frac{q}{n-k}\right) \leq 3\frac{qk}{n(n-k)} \leq 3\frac{k}{n-k} \lesssim \log\left(\frac{n}{n-k}\right) \lesssim \frac{\log\left(\frac{n}{n-k}\right)}{\sqrt{\log\left(\frac{n-k}{q}\right)_+ \vee 1}} ,
$$

because $n/(n-k) \leq 2.1$ and $(n-k)/q \leq 3$.

Now assume $q \leq n/3$ and $k < n/2$. It follows from the second inequality in (122) (which can be used because $q/n < q/(n-k)$ and $q/n \leq 1 - 2q/n < 1 - q/(n-k)$) that

$$
\overline{\Phi}^{-1}\left(\frac{q}{n}\right) - \overline{\Phi}^{-1}\left(\frac{q}{n-k}\right) \lesssim \frac{k}{(n-k)\overline{\Phi}^{-1}(q/n)} \lesssim \frac{\log\left(\frac{n}{n-k}\right)}{\overline{\Phi}^{-1}(q/n)} ,
$$

because $k/(n-k) \leq 1$. Also, for $x \leq 0.004$, we have $\overline{\Phi}^{-1}(x) \geq \sqrt{\log(1/x)}$ by (121). This implies $\overline{\Phi}^{-1}(x) \gtrsim \sqrt{\log(1/x)}$ for all $x \leq 1/3$, and we obtain

$$
\overline{\Phi}^{-1}\left(\frac{q}{n}\right) - \overline{\Phi}^{-1}\left(\frac{q}{n-k}\right) \lesssim \frac{\log\left(\frac{n}{n-k}\right)}{\sqrt{\log\left(\frac{n}{q}\right)_+ \vee 1}} \lesssim \frac{\log\left(\frac{n}{n-k}\right)}{\sqrt{\log\left(\frac{n-k}{q}\right)_+ \vee 1}} .
$$

Then, we consider the case where $q \leq n/3$, $k \geq n/2$, and $q \leq 0.4(n-k)$. It follows from the last inequality in (122) that

$$
\overline{\Phi}^{-1}\left(\frac{q}{n}\right) - \overline{\Phi}^{-1}\left(\frac{q}{n-k}\right) \leq \frac{\log\left(\frac{n}{n-k}\right) + \frac{1}{[\overline{\Phi}^{-1}(\frac{q}{n-k})]^2}}{\overline{\Phi}^{-1}(\frac{q}{n-k})} \lesssim \frac{\log\left(\frac{n}{n-k}\right)}{\sqrt{\log\left(\frac{n-k}{q}\right)_+ \vee 1}} ,
$$

where we used in the last inequality that $\overline{\Phi}^{-1}(\frac{q}{n-k}) \gtrsim \sqrt{\log\left(\frac{n-k}{q}\right)_+}$ for $q/(n-k) \leq 0.4$.

Finally, we assume that $q \leq n/3$, $k \geq n/2$ and $q/(n-k) \in (0.4, 0.7]$. Then, it follows from (119) that

$$\overline{\Phi}^{-1}\left(\frac{q}{n}\right) - \overline{\Phi}^{-1}\left(\frac{q}{n-k}\right) \leq \overline{\Phi}^{-1}\left(\frac{q}{n}\right) + \overline{\Phi}^{-1}(0.3) \leq \sqrt{2\log(n/q)} + \overline{\Phi}^{-1}(0.3) \lesssim \sqrt{\log(n/q)}$$

$$\lesssim \sqrt{\log\left(\frac{n}{n-k}\right)} \lesssim \frac{\log\left(\frac{n}{n-k}\right)}{\sqrt{\log\left(\frac{n-k}{q}\right)_+ \vee 1}} \; .$$

$\square$

D.3. Deviation inequalities for Gaussian empirical quantiles

Lemma D.7. *Let $\xi = (\xi_1, \ldots, \xi_n)$ be a standard Gaussian vector of size n. For any integer $q \in (0.3n, 0.7n)$, we have for all $0 < x \leq \frac{8}{225}q \wedge \left(\frac{n^2}{18q}[\overline{\Phi}^{-1}(q/n) - \overline{\Phi}^{-1}(0.7)]^2\right)$,*

$$\mathbb{P}\left[\xi_{(q)} + \overline{\Phi}^{-1}(q/n) \geq 3\frac{\sqrt{2qx}}{n}\right] \leq e^{-x} \; , \tag{125}$$

and for all $0 < x \leq \frac{n^2}{18q}[\overline{\Phi}^{-1}(0.3) - \overline{\Phi}^{-1}(q/n)]^2$,

$$\mathbb{P}\left[\xi_{(q)} + \overline{\Phi}^{-1}(q/n) \leq -3\frac{\sqrt{2qx}}{n}\right] \leq e^{-x} \; . \tag{126}$$

Now consider any integer $q \leq 0.4n$. For all $x \leq \frac{1}{8}q[(\overline{\Phi}^{-1}(q/n))^2 \wedge (\overline{\Phi}^{-1}(q/n))^4]$, we have

$$\mathbb{P}\left[\xi_{(q)} + \overline{\Phi}^{-1}(q/n) \geq \frac{1}{\overline{\Phi}^{-1}(q/n)}\sqrt{\frac{2x}{q}} + \frac{8}{3}\frac{x}{q\overline{\Phi}^{-1}(q/n)}\right] \leq e^{-x} \; . \tag{127}$$

For all $x \leq q/8$, we have

$$\mathbb{P}\left[\xi_{(q)} + \overline{\Phi}^{-1}(q/n) \leq -\frac{2}{\overline{\Phi}^{-1}(q/n)}\sqrt{\frac{2x}{q}}\right] \leq e^{-x} \; . \tag{128}$$

Proof of Lemma D.7. Consider any $t \geq 0$ and denote $p = \overline{\Phi}[\overline{\Phi}^{-1}(q/n) - t]$ which belongs to $(q/n, 1)$. We have

$$\mathbb{P}\left[\xi_{(q)} \geq -\overline{\Phi}^{-1}(q/n) + t\right] = \mathbb{P}\left[\mathcal{B}(n, p) \leq q - 1\right] \leq \mathbb{P}\left[\mathcal{B}(n, p) \leq q\right] \; . \tag{129}$$

By the mean value theorem, we have $p - q/n \geq t \inf_{x \in [0,t]} \phi[\overline{\Phi}^{-1}(q/n) - x]$. Assume first that q/n belongs to $(0.3, 0.7)$ and that $\overline{\Phi}^{-1}(q/n) - t \geq \overline{\Phi}^{-1}(0.7)$. Then, it follows from the previous inequality that $p - q/n \geq t\phi[\overline{\Phi}^{-1}(0.3)] \geq t/3$. Together with Bernstein's inequality, we obtain

$$\mathbb{P}\left[\xi_{(q)} \geq -\overline{\Phi}^{-1}(q/n) + t\right] \leq \mathbb{P}\left[\mathcal{B}(n, q/n + t/3) \leq q\right]$$

$$\leq \exp\left[-\frac{n^2 t^2/9}{2(q + nt/3)(1 - (q/n + t/3)) + 2nt/9}\right]$$

$$\leq \exp\left[-\frac{n^2 t^2/9}{1.4q + nt(1.4/3 + 2/9)}\right] \; ,$$

where we used that $q/n \geq 0.3$ in the last line. If we further assume that $t \leq 0.8q/n$, we obtain

$$\mathbb{P}\Big[\xi_{(q)} \geq -\overline{\Phi}^{-1}(q/n) + t\Big] \leq \exp\left[-n^2 t^2/(18q)\right] \ .$$

In view of the conditions $t \leq 0.8q/n$ and $t \leq \overline{\Phi}^{-1}(q/n) - \overline{\Phi}^{-1}(0.7)$, we have proved (125).

Let us now prove (127). Assume that $q/n \leq 0.4$ and $t \leq \overline{\Phi}^{-1}(q/n)$. This implies $p \leq 1/2$ and we have $p - q/n \geq t\phi[\overline{\Phi}^{-1}(q/n)] \geq t(q/n)\overline{\Phi}^{-1}(q/n)$ by Inequality (120) in Lemma D.4. Then, (129) together with Bernstein's inequality yields

$$\mathbb{P}\Big[\xi_{(q)} \geq -\overline{\Phi}^{-1}(q/n) + t\Big] \leq \exp\left[-\frac{t^2 q^2 (\overline{\Phi}^{-1}(q/n))^2}{2q + \frac{8}{3} tq\overline{\Phi}^{-1}(q/n)}\right] \ ,$$

which implies (127) by simple algebraic manipulations.

Next, we consider the left deviations. For any $t > 0$, we write $p = \overline{\Phi}[\overline{\Phi}^{-1}(q/n) + t]$. We have

$$\mathbb{P}[\xi_{(q)} \leq -\overline{\Phi}^{-1}(q/n) - t] = \mathbb{P}[\mathcal{B}(n,p) \geq q].$$

Then, Bernstein's inequality yields

$$\mathbb{P}\Big[\xi_{(q)} \leq -\overline{\Phi}^{-1}(q/n) - t\Big] \leq \exp\left[-\frac{(q-np)^2}{2np(1-p) + 2(q-np)/3}\right] \leq \exp\left[-\frac{(q-np)^2}{2q}\right] \ , \quad (130)$$

because $2np(1-p) + 2(q-np)/3 \leq (2 - 2/3)np + 2q/3 \leq 2q$ since $p \leq q/n$. First, assume that $q/n \in (0.3, 0.7)$ and that $\overline{\Phi}^{-1}(q/n) + t \leq \overline{\Phi}^{-1}(0.3)$. Then, it follows from the mean value theorem that $q/n - p \geq t \inf_{x \in [0,t]} \phi\Big[\overline{\Phi}^{-1}(q/n) + x\Big] \geq t/3$, which implies

$$\mathbb{P}\Big[\xi_{(q)} \leq -\overline{\Phi}^{-1}(q/n) - t\Big] \leq \exp\left[-\frac{n^2 t^2}{18q}\right] \ .$$

We have shown (126).

Now assume that $q/n \leq 0.4$ and consider any $0 < t < (\overline{\Phi}^{-1}(q/n))^{-1}$. It follows again from the mean value theorem and Lemma D.4 that

$$q/n - p \geq t\phi\Big[\overline{\Phi}^{-1}(p)\Big] \geq tp\,\overline{\Phi}^{-1}(p) \geq tp\,\overline{\Phi}^{-1}(q/n) \ ,$$

which implies $np \leq q/(1 + t\,\overline{\Phi}^{-1}(q/n))$ and thus

$$q - np \geq q\left(1 - \frac{1}{1 + t\,\overline{\Phi}^{-1}(q/n)}\right) = \frac{tq\,\overline{\Phi}^{-1}(q/n)}{1 + t\,\overline{\Phi}^{-1}(q/n)} \geq \tfrac{1}{2} tq\,\overline{\Phi}^{-1}(q/n) \ .$$

Coming back to (130), we get

$$\mathbb{P}\Big[\xi_{(q)} \leq -\overline{\Phi}^{-1}(q/n) - t\Big] \leq \exp\left[-\frac{t^2 q(\overline{\Phi}^{-1}(q/n))^2}{8}\right] \ ,$$

which shows (128). $\qquad\square$

Lemma D.8. *Let $\xi = (\xi_1, \ldots, \xi_n)$ be a standard Gaussian vector of size n. There exist two positive constants c_1 and c_2 such that the following holds: for any integer $1 \leq q \leq n - 1$ and any $x \leq c_1 q$, we have*

$$\mathbb{P}\left[\xi_{(q)} + \overline{\Phi}^{-1}(q/n) \geq c_2 \sqrt{\frac{x}{q[\log(\frac{n}{q}) \vee 1]}}\right] \leq e^{-x}, \tag{131}$$

$$\mathbb{P}\left[\xi_{(q)} + \overline{\Phi}^{-1}(q/n) \leq -c_2 \sqrt{\frac{x}{q[\log(\frac{n}{q}) \vee 1]}}\right] \leq e^{-x}. \tag{132}$$

Proof. First consider the case $q \leq 0.6n$. It follows from Lemma D.4 that $|\overline{\Phi}^{-1}(q/n)| \vee 1 \lesssim \sqrt{\log(\frac{n}{q})} \vee 1$. Then, the result is a straightforward consequence of Lemma D.7 by gathering all the deviation bounds and taking c_1 small enough and c_2 large enough. For $q \geq 0.6n$, we use the symmetry of the normal distribution and observe that $\xi_{(q)}$ is distributed as $-\xi_{(n-q)}$ while $\overline{\Phi}^{-1}(q/n) = -\overline{\Phi}^{-1}(1 - q/n)$. $\square$

Appendix E: Additional numerical experiments

In this section, we provide numerical experiments for two scenarios for the alternatives:

- the alternatives m_i, $1 \leq i \leq n_1$, are linearly increasing from 0.01 to 2Δ, that is,

$$m_i = 0.01 + (2\Delta - 0.01)(i - 1)/n_1, \quad 1 \leq i \leq n_1;$$

- the alternatives m_i, $1 \leq i \leq n_1$, are generated as n_1 i.i.d. uniform variables in $(0.01, 2\Delta)$ (previously and independently from the Monte-Carlo loop).

References

[1] David J. Aldous. Exchangeability and related topics. In École d'été de probabilités de Saint-Flour, XIII—1983, volume 1117 of Lecture Notes in Math., pages 1–198. Springer, Berlin, 1985.

[2] Ery Arias-Castro and Shiyun Chen. Distribution-free multiple testing. Electron. J. Stat., 11(1):1983–2001, 2017.

[3] Yannick Baraud. Non-asymptotic minimax rates of testing in signal detection. Bernoulli, 8(5):577–606, 2002.

[4] Rina Foygel Barber and Emmanuel J. Candès. Controlling the false discovery rate via knockoffs. Ann. Statist., 43(5):2055–2085, 2015.

[5] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Statist. Soc. Ser. B, 57(1):289–300, 1995.

[6] Yoav Benjamini and Daniel Yekutieli. The control of the false discovery rate in multiple testing under dependency. Ann. Statist., 29(4):1165–1188, 2001.

[7] G. Blanchard, P. Neuvial, and E. Roquain. Post hoc inference via joint family-wise error rate control. ArXiv e-prints, March 2017.
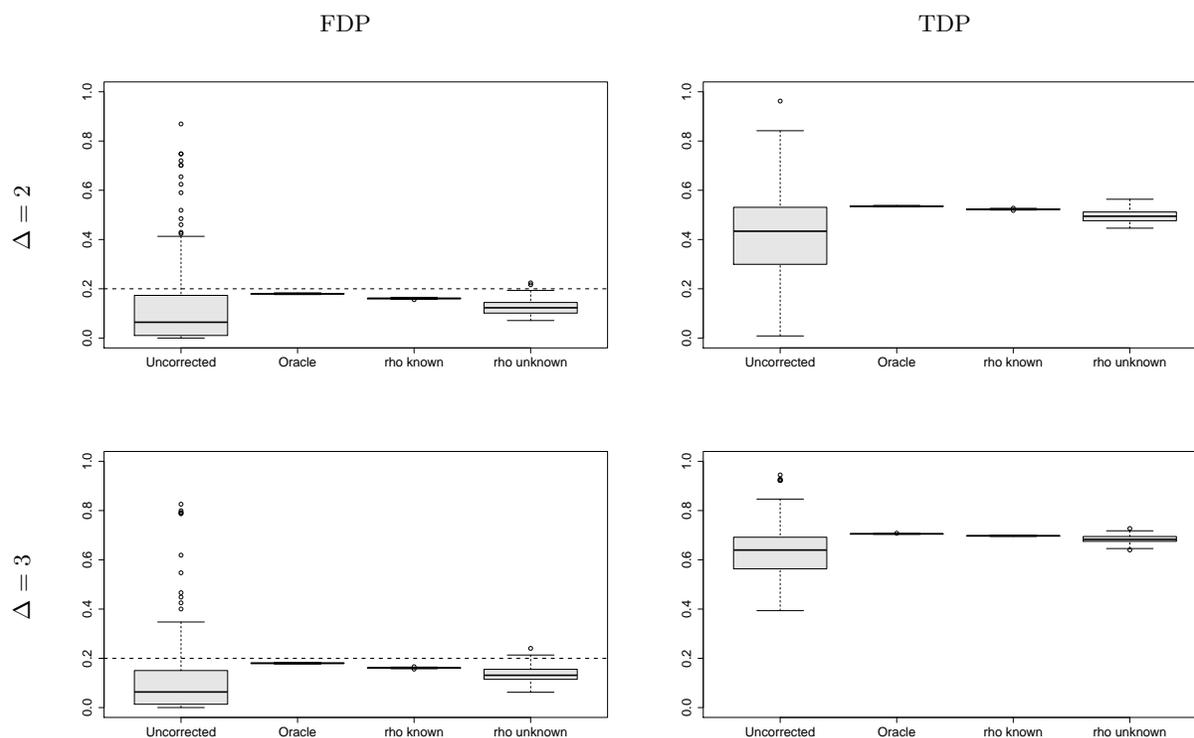
FDP TDP

FIG 4. *Same as Figure 2 but for alternative m_i linearly increasing from* 0.01 *to* 2Δ.

[8] Taras Bodnar and Thorsten Dickhaus. On the Simes inequality in elliptical models. Ann. Inst. Stat. Math., 69(1):215–230, 2017.

[9] Malgorzata Bogdan, Ewout van den Berg, Chiara Sabatti, Weijie Su, and Emmanuel J. Candès. SLOPE—adaptive variable selection via convex optimization. Ann. Appl. Stat., 9(3):1103–1140, 2015.

[10] T. Tony Cai and Jiashun Jin. Optimal rates of convergence for estimating the null density and proportion of nonnull effects in large-scale multiple testing. Ann. Statist., 38(1):100–145, 2010.

[11] T. Tony Cai and Mark G. Low. Nonquadratic estimators of a quadratic functional. Ann. Statist., 33(6):2930–2956, 2005.

[12] T. Tony Cai and Mark G. Low. Testing composite hypotheses, Hermite polynomials and optimal estimation of a nonsmooth functional. Ann. Statist., 39(2):1012–1041, 2011.

[13] Alexandra Carpentier and Nicolas Verzelen. Adaptive estimation of the sparsity in the gaussian vector model. arXiv preprint arXiv:1703.00167, 2017.

[14] Mengjie Chen, Chao Gao, and Zhao Ren. Robust covariance and scatter matrix estimation under huber's contamination model. The Annals of Statistics, 46(5):1932–1960, 2018.

[15] O. Collier, L. Comminges, and A. B. Tsybakov. On estimation of nonsmooth functionals of sparse normal means. ArXiv e-prints, May 2018.

[16] Olivier Collier, Laëtitia Comminges, and Alexandre B Tsybakov. Minimax estimation of linear and quadratic functionals on sparsity classes. The Annals of Statistics, 45(3):923–958, 2017.
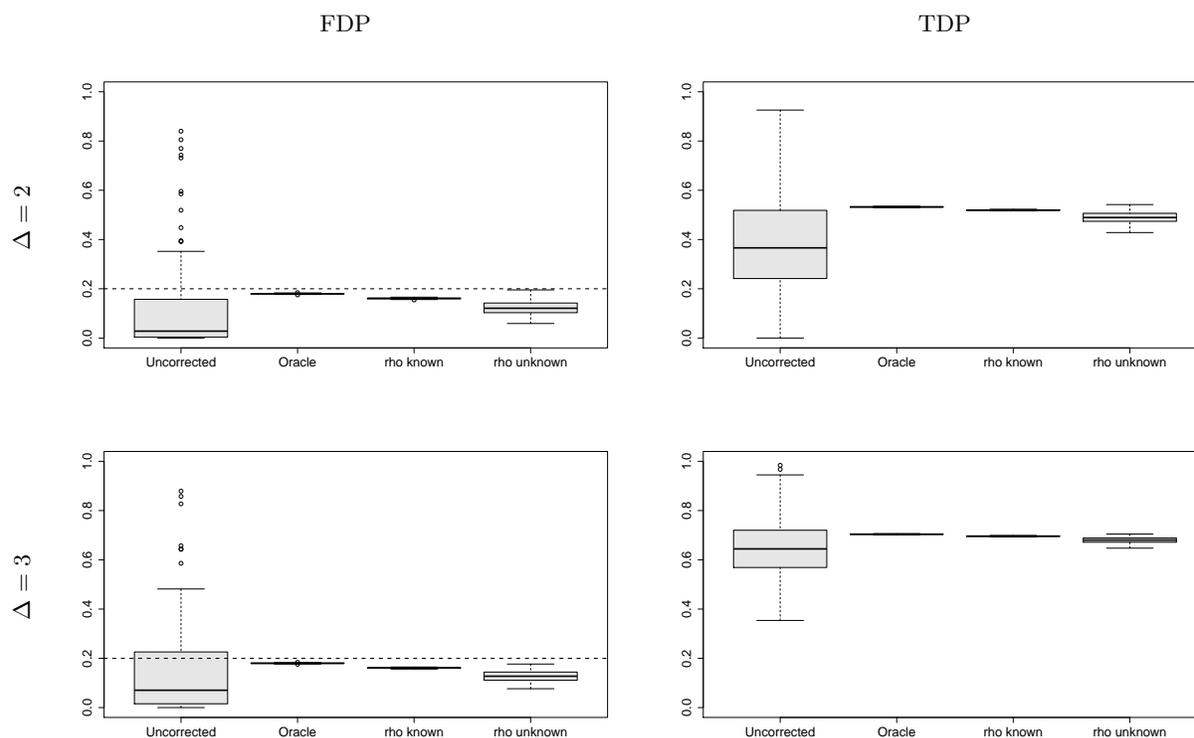
FIG 5. *Same as Figure 2 but for alternative m_i i.i.d. uniform in $(0.01, 2\Delta)$.*

[17] Olivier Collier, Laëtitia Comminges, Alexandre B Tsybakov, and Nicolas Verzelen. Optimal adaptive estimation of linear functionals under sparsity. arXiv preprint arXiv:1611.09744, 2016.

[18] S. Delattre and E. Roquain. On the false discovery proportion convergence under Gaussian equi-correlation. Statist. Probab. Lett., 81(1):111–115, 2011.

[19] Sylvain Delattre and Etienne Roquain. New procedures controlling the false discovery proportion via Romano-Wolf's heuristic. Ann. Statist., 43(3):1141–1177, 2015.

[20] Sylvain Delattre and Etienne Roquain. On empirical distribution function of high-dimensional Gaussian vector components with an application to multiple testing. Bernoulli, 22(1):302–324, 2016.

[21] David Donoho and Jiashun Jin. Higher criticism for detecting sparse heterogeneous mixtures. Ann. Statist., 32(3):962–994, 2004.

[22] David L Donoho and Michael Nussbaum. Minimax quadratic estimation of a quadratic functional. Journal of Complexity, 6(3):290–323, 1990.

[23] Sandrine Dudoit and Mark J. van der Laan. Multiple testing procedures with applications to genomics. Springer Series in Statistics. Springer, New York, 2008.

[24] Bradley Efron. Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. J. Am. Stat. Assoc., 99(465):96–104, 2004.

[25] Bradley Efron. Correlation and large-scale simultaneous significance testing. J. Amer. Statist. Assoc., 102(477):93–103, 2007.

[26] Bradley Efron. Doing thousands of hypothesis tests at the same time. Metron - International Journal of Statistics, LXV(1):3–21, 2007.
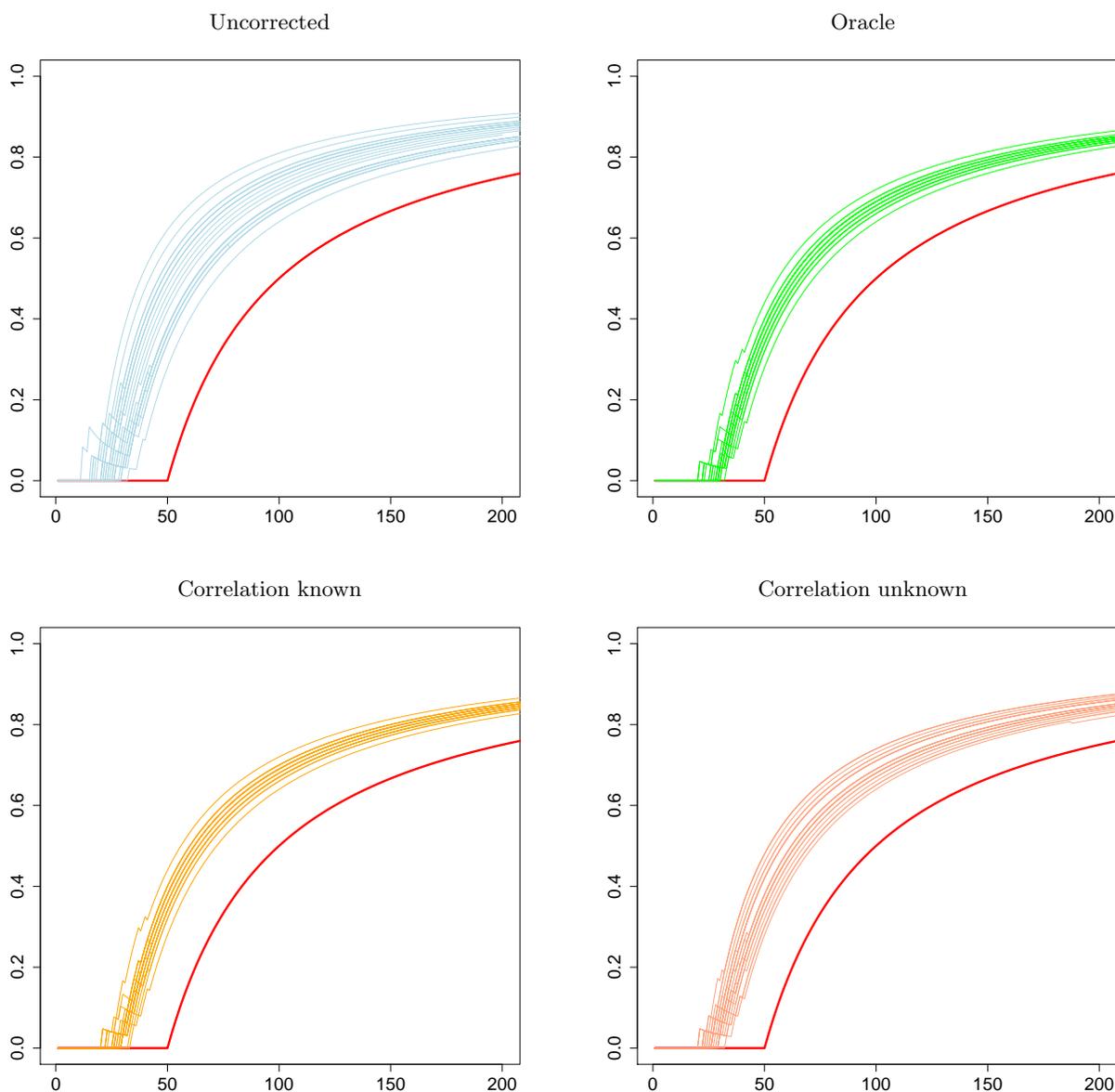
Uncorrected

Oracle

Correlation known

Correlation unknown

FIG 6. *Same as Figure 3 but for alternative m_i linearly increasing from* 0.01 *to* 2Δ.

[27] Bradley Efron. Empirical Bayes estimates for large-scale prediction problems. *J. Am. Stat. Assoc.*, 104(487):1015–1028, 2009.

[28] Bradley Efron. Correlated z -values and the accuracy of large-scale statistical estimates. *J. Amer. Statist. Assoc.*, 105(491):1042–1055, 2010.

[29] Jianqing Fan and Xu Han. Estimation of the false discovery proportion with unknown dependence. *J. R. Stat. Soc., Ser. B, Stat. Methodol.*, 79(4):1143–1164, 2017.

[30] Jianqing Fan, Xu Han, and Weijie Gu. Estimating false discovery proportion under arbitrary covariance dependence. *J. Am. Stat. Assoc.*, 107(499):1019–1035, 2012.

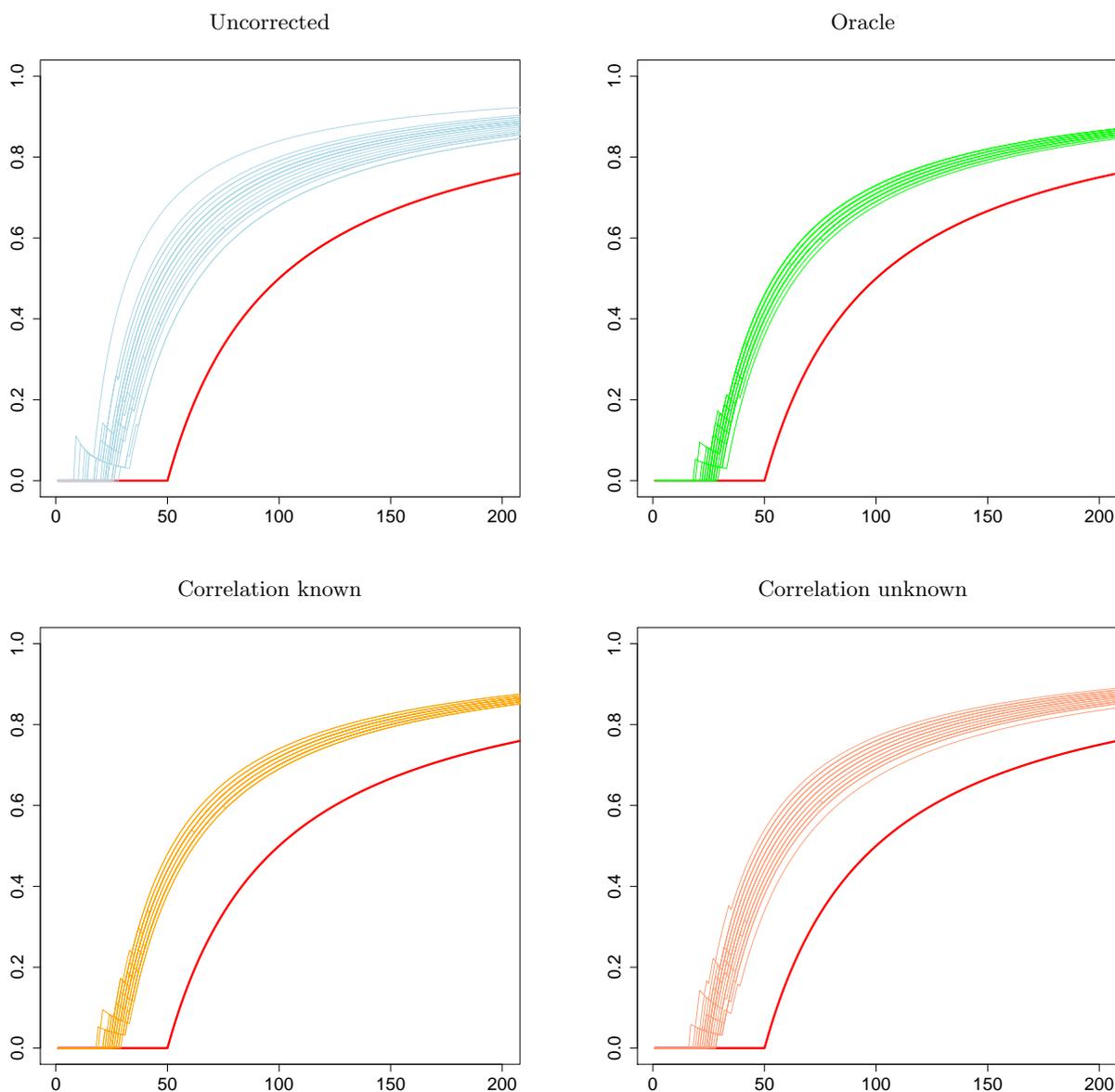[31] J. A. Ferreira and A. H. Zwinderman. On the Benjamini-Hochberg method. *Ann. Statist.*,

FIG 7. *Same as Figure 3 but for alternative m_i i.i.d. uniform in $(0.01, 2\Delta)$.*

34(4):1827–1849, 2006.

[32] H. Finner, M. Roters, and K. Strassburger. On the Simes test under dependence. Stat. Pap., 58(3):775–789, 2017.

[33] Helmut Finner, Thorsten Dickhaus, and Markus Roters. Dependency and false discovery rate: asymptotics. Ann. Statist., 35(4):1432–1455, 2007.

[34] Dominique Foata. Some Hermite polynomial identities and their combinatorics. Adv. in Appl. Math., 2(3):250–259, 1981.

[35] Chloé Friguet, Maela Kloareg, and David Causeur. A factor model approach to multiple testing under dependence. J. Amer. Statist. Assoc., 104(488):1406–1415, 2009.

[36] R. V. Gamkrelidze. Analysis II, volume 14 of Encyclopaedia of Mathematical Sciences. Springer, Berlin, 1990.

[37] Y. Gavrilov, Y. Benjamini, and Sanat K. Sarkar. An adaptive step-down procedure with proven FDR control under independence. Ann. Statist., 37(2):619–629, 2009.

[38] Christopher Genovese and Larry Wasserman. A stochastic process approach to false discovery control. Ann. Statist., 32(3):1035–1061, 2004.

[39] Christopher R. Genovese and Larry Wasserman. Exceedance control of the false discovery proportion. J. Amer. Statist. Assoc., 101(476):1408–1417, 2006.

[40] J. Goeman, R. Meijer, T. Krebs, and A. Solari. Simultaneous Control of All False Discovery Proportions in Large-Scale Multiple Hypothesis Testing. ArXiv e-prints, November 2016.

[41] J. Goeman and A. Solari. The sequential rejection principle of familywise error control. Ann. Statist., 38(6):3782–3810, 2010.

[42] Jelle J. Goeman and Aldo Solari. Multiple testing for exploratory research. Statist. Sci., 26(4):584–597, 2011.

[43] Alexander Goldenshluger and Oleg Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. Ann. Statist., 39(3):1608–1632, 2011.

[44] Wenge Guo, Li He, and Sanat K. Sarkar. Further results on controlling the false discovery proportion. The Annals of Statistics, 42(3):1070–1101, 06 2014.

[45] Yanjun Han, Jiantao Jiao, and Tsachy Weissman. Minimax estimation of kl divergence between discrete distributions. arXiv preprint arXiv:1605.09124, 2016.

[46] Peter J Huber. Robust estimation of a location parameter. The annals of mathematical statistics, pages 73–101, 1964.

[47] Peter J Huber. Robust statistics. In International Encyclopedia of Statistical Science, pages 1248–1251. Springer, 2011.

[48] Ildar Abdullovich Ibragimov and Rafail Zalmanovich Khasminskii. On nonparametric estimation of the value of a linear functional in gaussian white noise. Theory of Probability & Its Applications, 29(1):18–32, 1985.

[49] N. Ignatiadis and W. Huber. Covariate powered cross-weighted multiple testing. ArXiv e-prints, January 2017.

[50] Yuri Ingster and Irina A Suslina. Nonparametric goodness-of-fit testing under Gaussian models, volume 169. Springer Science & Business Media, 2012.

[51] Jiantao Jiao, Yanjun Han, and Tsachy Weissman. Minimax estimation of the l1 distance. In Information Theory (ISIT), 2016 IEEE International Symposium on, pages 750–754. IEEE, 2016.

[52] Jiashun Jin. Proportion of non-zero normal means: universal oracle equivalences and uniformly consistent estimators. J. R. Stat. Soc. Ser. B Stat. Methodol., 70(3):461–493, 2008.

[53] Jiashun Jin and T. Tony Cai. Estimating the null and the proportional of nonnull effects in large-scale multiple comparisons. J. Amer. Statist. Assoc., 102(478):495–506, 2007.

[54] Anatoli Juditsky and Arkadi Nemirovski. On nonparametric tests of positivity/monotonicity/convexity. Ann. Statist., 30(2):498–527, 2002.

[55] Jana Jurečková, Jan Picek, and Pranab Kumar Sen. Methodology in robust and nonparametric statistics. CRC Press, 2012.

[56] Edward L. Korn, James F. Troendle, Lisa M. McShane, and Richard Simon. Controlling the number of false discoveries: application to high-dimensional genomic data. J. Statist.

Plann. Inference, 124(2):379–398, 2004.

[57] C. Lacour and P. Massart. Minimal penalty for Goldenshluger-Lepski method. Stochastic Process. Appl., 126(12):3774–3789, 2016.

[58] Tony Lancaster. The incidental parameter problem since 1948. Journal of econometrics, 95(2):391–413, 2000.

[59] Jeffrey T. Leek and John D. Storey. A general framework for multiple testing dependence. Proceedings of the National Academy of Sciences, 105(48):18718–18723, 2008.

[60] Oleg Lepski, Arkady Nemirovski, and Vladimir Spokoiny. On estimation of the l_r norm of a regression function. Probability theory and related fields, 113(2):221–253, 1999.

[61] O. V. Lepskiĭ. A problem of adaptive estimation in Gaussian white noise. Teor. Veroyatnost. i Primenen., 35(3):459–470, 1990.

[62] A. Li and R. Foygel Barber. Multiple testing with the structure adaptive Benjamini-Hochberg algorithm. ArXiv e-prints, June 2016.

[63] Jerzy Neyman and Elizabeth L Scott. Consistent estimates based on partially consistent observations. Econometrica: Journal of the Econometric Society, pages 1–32, 1948.

[64] M. Rabinovich, A. Ramdas, M. I. Jordan, and M. J. Wainwright. Optimal rates and tradeoffs in multiple testing. ArXiv e-prints, May 2017.

[65] Joseph P. Romano, Azeem M. Shaikh, and Michael Wolf. Control of the false discovery rate under dependence using the bootstrap and subsampling. TEST: An Official Journal of the Spanish Society of Statistics and Operations Research, 17(3):417–442, December 2008.

[66] Joseph P. Romano and Michael Wolf. Exact and approximate stepdown methods for multiple hypothesis testing. J. Amer. Statist. Assoc., 100(469):94–108, 2005.

[67] Joseph P. Romano and Michael Wolf. Control of generalized error rates in multiple testing. Ann. Statist., 35(4):1378–1408, 2007.

[68] E. Roquain and M. van de Wiel. Optimal weighting for false discovery rate control. Electron. J. Stat., 3:678–711, 2009.

[69] Etienne Roquain and Fanny Villers. Exact calculations for false discovery proportion with application to least favorable configurations. Ann. Statist., 39(1):584–612, 2011.

[70] Sanat K. Sarkar. Rejoinder: On methods controlling the false discovery rate [mr2551809; mr2551810; mr2551811]. Sankhyā, 70(2, Ser. A):183–185, 2008.

[71] R. J. Simes. An improved Bonferroni procedure for multiple tests of significance. Biometrika, 73(3):751–754, 1986.

[72] Alexandre B. Tsybakov. Introduction to nonparametric estimation. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

[73] Nicolas Verzelen. Minimax risks for sparse regressions: ultra-high dimensional phenomenons. Electron. J. Stat., 6:38–90, 2012.

[74] P. H. Westfall and S. S. Young. Resampling-Based Multiple Testing. Wiley, 1993. Examples and Methods for P- Value Adjustment.

[75] Yihong Wu and Pengkun Yang. Chebyshev polynomials, moment matching, and optimal estimation of the unseen. arXiv preprint arXiv:1504.01227, 2015.