

A Unified Multiple Testing Framework based on ρ -values

Shenghao Qin[†], Bowen Gang[†], and Yin Xia^{*,†}

[†]*Department of Statistics and Data Science, Fudan University*

Abstract

Multiple testing is an important research area with widespread scientific applications, including in biology and neuroscience. Among popularly adopted multiple testing procedures, many are based on p-values or Local false discovery rate (Lfdr) statistics. However, p-values—often obtained via the probability integral transform of standard test statistics—typically lack information from the alternatives, resulting in suboptimal performance. In contrast, Lfdr-based methods can achieve asymptotic optimality, but their ability to control the false discovery rate (FDR) hinges on accurate estimation of the Lfdr, which can be challenging, especially when incorporating side information. In this article, we introduce a novel and flexible class of statistics, termed ρ -values, and develop a corresponding multiple testing framework that integrates the strengths of both p-values and Lfdr, while addressing their respective limitations. Specifically, the ρ -value framework unifies these two paradigms through a two-step process: ranking and thresholding. The ranking induced by ρ -values closely resembles that of Lfdr-based methods, while the thresholding step aligns with conventional p-value procedures. Therefore, the proposed framework guarantees FDR control under mild assumptions; it maintains the integrity of the structural information encoded by the summary statistics and the auxiliary covariates, and hence can be asymptotically optimal. We demonstrate the advantages of the ρ -value framework through comprehensive simulations and analyses of two real datasets: one from microbiome research and another related to attention deficit hyperactivity disorder (ADHD).

*Corresponding author: xiayin@fudan.edu.cn

Keywords: Asymptotically optimal, False discovery rate, Local false discovery rate, p-value, Side information

1 Introduction

With the advent of big data and increased data availability, multiple testing has become an increasingly critical challenge in modern scientific research. For instance, in microbiome-wide association studies (MWAS), investigating the relationship between microbiome features and complex host traits typically involves testing thousands of variables across numerous microorganisms. Without proper correction for multiplicity, such analyses are prone to inflated false positive rates. Similarly, in magnetic resonance imaging (MRI) studies aimed at identifying functional regions of the human brain for clinical diagnosis or medical research, the massive volume of high-resolution 3D imaging data complicates the task of simultaneous inference. High-dimensional regression settings—such as those encountered in gene association studies—present another example, where thousands of genes are tested for associations with drug sensitivities, particularly in cancer research.

A widely adopted measure of Type I error in multiple testing is the false discovery rate (FDR; [Benjamini and Hochberg \(1995\)](#)), defined as the expected proportion of false positives among all discoveries. Since its introduction, FDR has rapidly become a central concept in modern statistics and a primary tool for large-scale inference across a wide range of scientific disciplines. At a high level, most FDR-controlling procedures that control FDR operate in two steps: first rank all hypotheses according to some significance indices and then reject those with index values less than or equal to some threshold. In this paper, we propose a new multiple testing framework built upon a novel concept called ρ -values. This framework unifies the commonly used p-value and local false discovery rate (Lfdr) approaches while offering several advantages over both. Moreover, the ρ -value framework is closely connected to e-value based methods, providing greater flexibility in handling data dependencies. In what follows, we begin by reviewing conventional multiple testing practices and identifying their limitations, and then introduce the proposed ρ -value

framework, highlighting its theoretical and practical contributions.

Several popular FDR-controlling procedures use p-values as significance indices for ranking hypotheses (e.g., Benjamini and Hochberg, 1995; Genovese et al., 2006; Liu, 2013; Lei and Fithian, 2018; Cai et al., 2022). Typically, p-values are derived by applying a probability integral transform to well-known test statistics. For example, Li and Barber (2019) uses a permutation test, Roquain and Van De Wiel (2009) employs a Mann–Whitney U test and Cai et al. (2022) adopts a t-test. However, the p-value based methods can be inefficient because the conventional p-values do not incorporate information from the alternative distributions (e.g., Sun and Cai, 2007; Leung and Sun, 2022). The celebrated Neyman–Pearson lemma states that the optimal statistic for testing a single hypothesis is the likelihood ratio. In the multiple testing context, the local false discovery rate (Lfdr) serves as the natural analog of the likelihood ratio statistic (e.g., Efron et al., 2001; Efron, 2003; Aubert et al., 2004; Hong et al., 2009; Sarkar and Zhao, 2022). It has been shown that a ranking and thresholding procedure based on Lfdr is asymptotically optimal for FDR control (Sun and Cai, 2007; Xie et al., 2011; Cai et al., 2019). Subsequently, Heller and Rosset (2021) demonstrates that such Lfdr-based procedures are in fact exact optimal among all FDR-controlling rules. Nevertheless, the performance of Lfdr-based methods critically depends on the accurate estimation of the Lfdr, which itself requires integrating information across all test statistics—a task that can be quite challenging in practice (Marandon et al., 2024). This challenge becomes even more pronounced when incorporating side information, a common necessity in real-world applications. For example, in the MWAS dataset analyzed in Section 4.1, the proportion of zeros across samples for each operational taxonomic unit (OTU) serves as side information, capturing both biological and technical variability (Xia, 2020). Similarly, in the MRI dataset analyzed in Section 4.2, spatial coordinate indices act as side information, enabling the use of spatial structures to enhance signal detection and interpretability (Paloyelis et al., 2007). To overcome the Lfdr estimation challenge, several methods have been developed that aim to approximate Lfdr-based procedures using weighted p-values (e.g., Lei and Fithian, 2018; Li and Barber, 2019; Liang et al., 2023). While promising, these Lfdr-mimicking approaches often rely on strong

model assumptions or remain suboptimal in practice. Recent developments in conformal inference offer an appealing alternative by constructing provably valid marginal p-values without requiring explicit knowledge of the null distribution. These conformal p-values use data-driven calibration and provide finite-sample marginal FDR guarantees under relatively mild assumptions (Bates et al., 2023; Marandon et al., 2024). Nevertheless, this added flexibility comes at the cost of weaker FDR guarantees compared to methods that leverage exact null information.

To address the aforementioned challenges, this article introduces a novel concept—the ρ -value—which adopts the form of a likelihood ratio while offering greater flexibility in the choice of density functions compared to the traditional Lfdr. Building on this concept, we propose a new and flexible multiple testing framework that unifies p-value-based and Lfdr-based approaches. The ρ -value-based Benjamini–Hochberg (BH) procedure, including its weighted variant (analogous to weighted p-values), also follows the standard two-step structure: first, all hypotheses are ranked according to their (weighted) ρ -values, with the ranking designed to approximate that of the Lfdr; second, hypotheses with (weighted) ρ -values less than or equal to a data-driven threshold are rejected. The thresholding strategy is analogous to that used in conventional p-value procedures, thereby ensuring both interpretability and control of the FDR.

Compared to existing frameworks, the proposed ρ -value framework offers several notable advantages. First, when carefully constructed, ρ -values produce a ranking of hypotheses that coincides with that of Lfdr statistics. As a result, procedures based on ρ -values can achieve asymptotic optimality. Second, the thresholding strategy in ρ -value-based methods mirrors that of p-value-based procedures. This similarity allows the proposed methods to inherit desirable theoretical properties of p-value-based approaches. Importantly, FDR control in the ρ -value framework does not rely on consistent estimation of Lfdr statistics, making the approach significantly more flexible than traditional Lfdr-based methods. Third, side information can be seamlessly incorporated into ρ -value procedures through a simple weighting scheme, enhancing the ranking of hypotheses and thereby improving power. This integration is often more straightforward than in the Lfdr framework, where side informa-

tion complicates density estimation. Fourth, our proposed ρ -BH procedure demonstrates superior power compared to the e-BH procedure, and the weighted ρ -BH procedure consistently outperforms the weighted BH procedure in terms of detection power. Finally, the proposed framework provides a unified perspective that bridges p-value-based and Lfdr-based methodologies. In particular, we show that these two paradigms are more closely related than previously suggested (e.g., Sun and Cai, 2007; Leung and Sun, 2022).

The paper is structured as follows. Section 2 starts with the problem formulation. It then introduces the ρ -BH procedure and its variations. Sections 3 and 4 present numerical comparisons of the proposed methods and other competing approaches using simulated and real data, respectively. More discussions of the proposed framework are provided in Section 5. The weighted ρ -BH procedures and the corresponding theories, as well as all technical proofs are collected in the Appendix.

2 Methodology

In this section, we begin by introducing the problem formulation and the motivation behind the development of ρ -value-based procedures. Subsequently, in Sections 2.2–2.5, we present a series of ρ -value-based multiple testing methods. Specifically, we introduce both oracle and data-driven versions of the ρ -BH procedure, followed by extensions that incorporate side information into the ρ -value framework.

2.1 Problem formulation

Suppose our goal is to simultaneously test the following m hypotheses:

$$H_{0,i} : \theta_i = 0 \quad \text{versus} \quad H_{1,i} : \theta_i = 1, \quad i = 1, \dots, m.$$

Assume that we observe independent summary statistics $\{X_i\}_{i=1}^m$ arising from the following random mixture model:

$$\theta_i \stackrel{iid}{\sim} \text{Ber}(\pi), \quad X_i | \theta_i \stackrel{ind}{\sim} (1 - \theta_i)f_0 + \theta_i f_1, \quad (1)$$

where $\pi = P(\theta_i = 1)$ and f_0 and f_1 respectively represent the null and alternative density functions of X_i . By convention, the null density f_0 is assumed to be known. Such a model has been widely adopted in many large-scale inference problems (e.g., [Efron, 2004](#); [Efron and Tibshirani, 2007](#); [Jin and Cai, 2007](#)). For simplicity, we assume homogeneous π and f_1 in Model (1) for now, and it will be extended to heterogeneous scenarios in later sections. In addition, Model (1) does not necessarily correspond to the actual data generation process. Instead, it only serves as a hypothetical model to motivate our methodology.

Denote by $\boldsymbol{\delta} = (\delta_1, \dots, \delta_m) \in \{0, 1\}^m$ an m -dimensional decision vector, where $\delta_i = 1$ means we reject $H_{0,i}$, and $\delta_i = 0$ otherwise. In large-scale multiple testing problems, false positives are inevitable if one wishes to discover non-nulls with a reasonable power. Instead of aiming to avoid any false positives, [Benjamini and Hochberg \(1995\)](#) introduces the FDR, i.e., the expected proportion of false positives among all selections, written formally as $\text{FDR}(\boldsymbol{\delta}) = E[\{\sum_{i=1}^m (1 - \theta_i)\delta_i\}/\{\max\{\sum_{i=1}^m \delta_i, 1\}\}]$, and a practical goal is to control the FDR at a pre-specified significance level. A closely related quantity of FDR is the marginal false discovery rate (mFDR), defined by $\text{mFDR}(\boldsymbol{\delta}) = E\{\sum_{i=1}^m (1 - \theta_i)\delta_i\}/E(\sum_{i=1}^m \delta_i)$. Under certain first- and second-order conditions on the number of rejections, the mFDR and the FDR are asymptotically equivalent ([Genovese and Wasserman, 2002](#); [Basu et al., 2018](#); [Cai et al., 2019](#)). The mFDR criterion is often employed to facilitate methodological development and derive optimality results in large-scale testing problems. An ideal multiple testing procedure should both control the FDR (or mFDR) at a pre-specified nominal level α and maximize statistical power, which is quantified by the expected number of true positives: $\text{ETP}(\boldsymbol{\delta}) = E(\sum_{i=1}^m \theta_i \delta_i)$. We call a multiple testing procedure *valid* if it controls the mFDR asymptotically at the nominal level α , and *optimal* if it has the largest ETP among all valid procedures. We call $\boldsymbol{\delta}$ asymptotically optimal if $\text{ETP}(\boldsymbol{\delta})/\text{ETP}(\boldsymbol{\delta}') \geq 1 + o(1)$ for any decision rule $\boldsymbol{\delta}'$ that controls mFDR at the pre-specified level α asymptotically.

2.2 Motivation, ρ -value and the ρ -BH procedure

The classical BH procedure ([Benjamini and Hochberg, 1995](#)) remains one of the most widely used approaches for multiple testing. It defines the decision rule as $\boldsymbol{\delta}_{BH} = (\mathbb{I}(p_1 \leq$

$p_{(k)}, \dots, \mathbb{I}(p_m \leq p_{(k)})$, where p_i is the p-value for $H_{0,i}$ and $k = \max\{i : mp_{(i)} \leq \alpha i\}$. It is demonstrated in (Benjamini and Hochberg, 1995) that the BH procedure in fact controls FDR at level $(1 - \pi)\alpha$. Therefore, a straightforward refinement of BH procedure, which we refer to as the adjusted BH procedure, selects $k = \max\{i : m(1 - \pi)p_{(i)} \leq \alpha i\}$. Intuitively, the adjusted BH procedure ranks hypotheses by their p-values and selects a threshold $p_{(k)}$ such that the estimated FDR, $(1 - \pi)mp_{(k)}/k$, does not exceed α . Though the adjusted BH procedure guarantees FDR control under independence, p-values generally lack information about the alternative distribution, which may result in suboptimal power.

An alternative approach is based on the Lfdr (Efron, 2004), defined as

$$\text{Lfdr}_i \equiv P(\theta_i = 0 | X_i) = \frac{(1 - \pi)f_0(X_i)}{(1 - \pi)f_0(X_i) + \pi f_1(X_i)}.$$

Let $\text{Lfdr}_{(1)} \leq \dots \leq \text{Lfdr}_{(m)}$ be the order statistics of $\text{Lfdr}_1, \dots, \text{Lfdr}_m$. It is shown in Sun and Cai (2007) that the decision rule $\delta_{SC} = (\mathbb{I}(\text{Lfdr}_1 \leq \text{Lfdr}_{(k)}), \dots, \mathbb{I}(\text{Lfdr}_m \leq \text{Lfdr}_{(k)}))$, where $k = \max\{i : i^{-1} \sum_{j=1}^i \text{Lfdr}_{(j)} \leq \alpha\}$, is asymptotically optimal among all mFDR control rules. We refer to this rule as the SC procedure. The intuition behind the SC procedure is twofold: first, the Lfdr statistics provide an optimal ranking of hypotheses based on their likelihood of being null; second, since the mFDR is an increasing function of the threshold, the threshold $\text{Lfdr}_{(k)}$ is chosen such that the estimate of FDR, given by $k^{-1} \sum_{j=1}^k \text{Lfdr}_{(j)}$, is just below α . Therefore, for the SC procedure, consistent estimates of the Lfdr_i 's are essential to ensure that $k^{-1} \sum_{j=1}^k \text{Lfdr}_{(j)}$ approximates the true mFDR when the threshold is set at $\text{Lfdr}_{(k)}$.

Our proposal aims to combine the strengths of the SC and the adjusted BH procedures. To motivate this, consider the idealized setting where f_1 is known. Define $\rho_i \equiv f_0(X_i)/f_1(X_i)$. Since f_0 is known, the null cumulative distribution function for ρ_i is also known; we denote it by $c(\cdot)$. By definition, $c(\rho_i)$ is a valid p-value for testing $H_{0,i}$. Consequently, we can apply the adjusted BH procedure directly to the transformed values $(c(\rho_1), \dots, c(\rho_m))$, yielding a procedure we denote by δ_ρ .

The new procedure δ_ρ can be conceptualized in two steps. First, ranking null hypotheses

according to $c(\rho_i)$. Second, thresholding using the adjusted BH rule. Since $c(\cdot)$ is an increasing function, it is clear that ranking by $c(\cdot)$ is equivalent to ranking by Lfdr_i . Moreover, the threshold chosen by the adjusted BH procedure is sharp in the sense that increasing the threshold to $c(\rho_{(k+1)})$, where $k = \max\{i : m(1 - \pi)c(\rho_{(i)}) \leq \alpha i\}$, would violate the FDR control constraint at level α . Thus, it is intuitively clear that δ_ρ is equivalent to δ_{SC} asymptotically. The power equivalence is formalized in the next theorem, while the FDR validity for the general ρ -value defined in Definition 1 below is deferred to Theorem 2.

Theorem 1. *Let $\rho_i = f_0(X_i)/f_1(X_i)$ and suppose X_i 's are independent. Let δ be any testing rule with $mFDR(\delta) \leq \alpha$ asymptotically. Suppose*

$$(A1) \quad mP\left(\rho_i \leq \frac{\alpha\pi}{(1-\pi)(1-\alpha)}\right) \rightarrow \infty.$$

Then we have $ETP(\delta_\rho)/ETP(\delta) \geq 1 + o(1)$.

Remark 1. *If the SC procedure results in at least one rejection with probability tending to 1, it implies that $mP(\text{Lfdr}_i \leq \alpha) \rightarrow \infty$ as $m \rightarrow \infty$. This serves as an equivalent condition for Assumption (A1), which thus represents a mild condition.*

At first glance, the procedure δ_ρ may appear to offer no new advantages over the SC procedure. However, a crucial distinction lies in its *robustness*: the validity of δ_ρ remains intact even if the alternative density $f_1(\cdot)$ is replaced by an arbitrary function $g(\cdot)$. That is, even when $\rho_i = f_0(X_i)/g(X_i)$ is computed using a misspecified or surrogate alternative, the resulting adjusted BH procedure applying to p-values $c(\rho_i)$ still controls the FDR under independence.

This robustness implies that δ_ρ can be viewed as a generalization and improvement of the SC procedure. On one hand, it retains FDR control under independence without requiring consistent estimation of the true Lfdr or the true f_1 . On the other hand, when π and Lfdr_i are known, the procedure δ_ρ is asymptotically optimal.

Generally, we refer to $\rho_i = f_0(X_i)/g(X_i)$, for any density function $g(\cdot)$, as a ρ -value of X_i , and the corresponding procedure δ_ρ as the ρ -BH procedure.

Definition 1. Suppose X is a summary statistic and $f_0(\cdot)$ is the density of X under the null. A ρ -value of X is defined as

$$\rho \equiv f_0(X)/g(X),$$

where $g(\cdot)$ is any density function satisfying $g(X) \neq 0$.

A summary of the ρ -BH algorithm is provided in Algorithm 1, and its theoretical validity is established in Theorem 2.

Algorithm 1 The ρ -BH procedure

Input: $\{X_i\}_{i=1}^m$; a predetermined density function $g(\cdot)$; non-null proportion π ; desired FDR level α .

1. Calculate the ρ -values $\rho_i = f_0(X_i)/g(X_i)$, for $i = 1, \dots, m$.
2. Sort the ρ -values from smallest to largest $\rho_{(1)} \leq \dots \leq \rho_{(m)}$.
3. Compute the null distribution function of ρ_i 's, and denote it by $c(\cdot)$.
4. Let $k = \max_{1 \leq j \leq m} [c(\rho_{(j)}) \leq (\alpha j)/\{m(1 - \pi)\}]$.

Output: The rejection set $\{i = 1, \dots, m: \rho_i \leq \rho_{(k)}\}$.

Theorem 2. Assume that the null ρ -values are mutually independent and are independent of the non-null ρ -values, then $FDR_{Algorithm\ 1} \leq \alpha$.

2.3 The data-driven ρ -BH procedure

In practice, f_1 and π are usually unknown and need to be estimated from the data. The problem of estimating non-null proportion has been discussed extensively in the literatures (e.g., Storey, 2002; Meinshausen and Rice, 2006; Jin and Cai, 2007; Chen, 2019). To ensure valid mFDR control, we require the estimator $\hat{\pi}$ to be conservative consistent, defined as follows.

Definition 2. An estimator $\hat{\pi}$ is a conservative consistent estimator of π , if $|\hat{\pi} - \tilde{\pi}| \xrightarrow{P} 0$ as $m \rightarrow \infty$, for some $\tilde{\pi}$ satisfying $0 \leq \tilde{\pi} \leq \pi$.

One possible choice of such $\hat{\pi}$ is the Storey estimator as provided by the following proposition.

Proposition 1. *The estimator $\hat{\pi}^\tau = 1 - \#\{i : c(\rho_i) \geq \tau\}/\{m(1 - \tau)\}$ proposed in [Storey \(2002\)](#) is conservative consistent for any τ satisfying $0 \leq \tau \leq 1$.*

The problem of estimating f_1 is more complicated. If we use the entire sample $\{X_i\}_{i=1}^m$ to construct \hat{f}_1 and let $\rho_i = f_0(X_i)/\hat{f}_1(X_i)$, then ρ_i 's are no longer independent even if X_i 's are. One possible strategy to circumvent this dependence problem is to use sample splitting. More specifically, we can randomly split the data into two disjoint halves and use the first half of the data to estimate the alternative density for the second half, i.e., $\hat{f}_1^{(2)}$ (e.g., we can use the estimator proposed in [Fu et al. \(2022\)](#)), then the ρ -values for the second half can be calculated by $f_0(X_i)/\hat{f}_1^{(2)}(X_i)$. Hence, when testing the second half of the data, $\hat{f}_1^{(2)}$ can be regarded as predetermined and independent of the data being tested. The decisions on the first half of the data can be obtained by switching the roles of the first and the second halves and repeating the above steps. If the FDR is controlled at level α for each half, then the overall mFDR is also controlled at level α asymptotically. We summarize the above discussions in Algorithm 2.

Algorithm 2 The data-driven ρ -BH procedure

Input: $\{X_i\}_{i=1}^m$; desired FDR level α .

1. Randomly split the data into two disjoint halves $\{X_i\}_{i=1}^m = \{X_{1,i}\}_{i=1}^{m_1} \cup \{X_{2,i}\}_{i=1}^{m_2}$, where $m_1 = \lfloor m/2 \rfloor$.
2. Use $\{X_{1,i}\}_{i=1}^{m_1}$ to construct the second half alternative estimate $\hat{f}_1^{(2)}$ and a conservative consistent estimate $\hat{\pi}_2$.
3. Run Algorithm 1 with $\{X_{2,i}\}_{i=1}^{m_2}$, $\hat{f}_1^{(2)}$, $\hat{\pi}_2$, α as inputs.
4. Switch the roles of $\{X_{1,i}\}_{i=1}^{m_1}$ and $\{X_{2,i}\}_{i=1}^{m_2}$. Repeat Steps 2 and 3, and combine rejections.

Output: The combined rejection set.

Remark 2. *A natural question for the data-splitting approach is whether it will negatively impact the power. Suppose that $\hat{f}_1^{(1)}, \hat{f}_1^{(2)}$ are consistent estimators for some function g , and*

$\hat{\pi}_1, \hat{\pi}_2$ are consistent estimators for some constant $\tilde{\pi}$. Denote by t_α the threshold selected by Algorithm 1 with g and $\tilde{\pi}$ as inputs, on full data. Then it is expected that the thresholds \hat{t}_1 and \hat{t}_2 selected by Algorithm 2 for each half of the data will both converge to t_α , provided that the empirical distributions of the two halves are similar. Therefore, the decision rule by data-splitting tends to be as powerful as the rule based on the full data.

Next we provide the theoretical guarantee for Algorithm 2 in the following theorem.

Theorem 3. Assume that X_i 's are independent. Denote by $\{\hat{\rho}_{d,i}\}_{i=1}^{m_d}$, $\hat{\rho}_{d,(k_d)}$ and $\hat{\pi}_d$ the p -values, selected thresholds and the estimated alternative proportions obtained from Algorithm 2, for the first and second halves of the data respectively, $d = 1, 2$. Denote by \hat{c}_d the null distribution function for $\hat{\rho}_{d,i}$. Suppose $\hat{\pi}_d > 0$ and $|\hat{\pi}_d - \tilde{\pi}_d| \xrightarrow{P} 0$ for some $\tilde{\pi}_d$ satisfying $0 \leq \tilde{\pi}_d \leq \pi$, and let $\tilde{Q}_d(t) = (1 - \tilde{\pi}_d)\hat{c}_d(t)/P(\hat{\rho}_{d,i} \leq t)$ and $t_{d,L} = \sup\{t > 0 : \tilde{Q}_d(t) \leq \alpha\}$, $d = 1, 2$. Assume the following hold

$$(A2) \quad \hat{\rho}_{d,(k_d)} \geq \nu \frac{\hat{\pi}_d}{1 - \hat{\pi}_d} \text{ and } P(\hat{\rho}_{d,i} \leq \nu \frac{\hat{\pi}_d}{1 - \hat{\pi}_d}) > c, \text{ for some constants } \nu, c > 0;$$

$$(A3) \quad \limsup_{t \rightarrow 0^+} \tilde{Q}_d(t) < \alpha, \quad \liminf_{t \rightarrow \infty} \tilde{Q}_d(t) > \alpha;$$

$$(A4) \quad \inf_{t \geq t_{d,L} + \epsilon_t} \tilde{Q}_d(t) \geq \alpha + \epsilon_\alpha, \text{ and } \tilde{Q}_d(t) \text{ is strictly increasing in } t \in (t_{d,L} - \epsilon_t, t_{d,L} + \epsilon_t), \\ \text{for some constants } \epsilon_\alpha, \epsilon_t > 0.$$

Then we have $\lim_{m \rightarrow \infty} mFDR_{Algorithm 2} \leq \alpha$.

Remark 3. Theorem 1 and the oracle rule in Sun and Cai (2007) imply that, when the alternative density and the non-null proportion are estimated by the truths, the threshold of the p -values should be at least $\alpha\pi/\{(1-\pi)(1-\alpha)\}$. Since $\hat{\pi}_d$'s are conservative consistent, we have $\hat{\pi}_d/(1 - \hat{\pi}_d)$ converges in probability to a number less than $\pi/(1 - \pi)$. Therefore, the first part of (A2) is mild. Moreover, by setting ν equal to some fixed number, say $\alpha/(1 - \alpha)$, the first part of (A2) can be easily checked numerically. The second part of (A2) is only slightly stronger than the condition that the total number of rejections for each half of the data is of order m . It is a sufficient condition to show that the estimated FDP, $m_d(1 - \hat{\pi}_d)\hat{c}_d(t)/\sum_{i=1}^{m_d} \mathbb{I}(\hat{\rho}_{d,i} \leq t)$, is close to $\tilde{Q}_d(t)$, and it can be easily relaxed if

$\hat{\pi}$ satisfies certain convergence rate. (A3) is also a reasonable condition, it excludes the trivial cases where no null hypothesis can be rejected or all null hypotheses are rejected. If \tilde{Q}_d 's are continuous, then the first part of (A4) is implied by (A3) and the definition of $t_{d,L}$. The second part of (A4) can be easily verified numerically and it is also mild under the continuity of \tilde{Q}_d . Finally, all of the above conditions are automatically satisfied in the oracle case where π and $Lfdr_i$ are known.

2.4 ρ -BH under dependence

So far, we have assumed that the summary statistics X_i 's are independent. However, in many applications, the X_i 's are observed across related groups, spatial locations, or time points, leading to inherent dependencies among the observations. To account for arbitrary dependence structures, Benjamini and Yekutieli (2001) proposed a conservative adjustment to the BH procedure. Specifically, they showed that the adjusted BH procedure with nominal level α controls the FDR at level $\alpha S(m)$, where $S(m) = \sum_{i=1}^m \frac{1}{i} \approx \log m$ is known as the Benjamini–Yekutieli (BY) correction factor. Accordingly, we obtain the following result for Algorithm 1 under the BY correction.

Theorem 4. *By setting the target FDR level to $\alpha/S(m)$, we have $FDR_{Algorithm\ 1} \leq \alpha$ under arbitrary dependence.*

An alternative approach for handling arbitrary dependence involves the use of e-values (Wang and Ramdas, 2022; Vovk and Wang, 2021). A non-negative random variable e is called an e-value if $\mathbb{E}(e) \leq 1$, where the expectation is taken under the null hypothesis. Using Markov's inequality, it can be shown that the reciprocal of an e-value is a valid p-value (Wang and Ramdas, 2022). Since

$$\mathbb{E}(1/\rho) = \int \frac{g(x)}{f_0(x)} f_0(x) dx = \int g(x) dx = 1,$$

it follows that $1/\rho$ is an e-value. It is further shown in (Wang and Ramdas, 2022) that if the reciprocals of e-values are used as input for the BH procedure, the resulting method—known as the e-BH procedure—controls the FDR at the target level under arbitrary dependence.

If one opts for either the BY correction or the e-BH procedure, then the data-splitting step in data-driven procedures such as Algorithm 2 is no longer necessary.

We further note that our earlier FDR control results under independence can be extended to settings with weakly dependent ρ -values. Specifically, the technical tools developed in (Liu, 2013; Chen and Liu, 2018; Cai et al., 2022) can be employed to establish asymptotic control of error rates under weak dependence. However, to maintain focus on the introduction of the proposed ρ -value-based testing framework, we defer detailed discussions of weak dependence to future work.

2.5 ρ -BH with side information

In many scientific applications, additional covariate information—such as patterns of the signals—is often available. Consequently, the problem of multiple testing with side information has garnered significant attention and has emerged as an active area of research in recent years (e.g., Du and Zhang, 2014; Lei and Fithian, 2018; Ramdas et al., 2019; Li and Barber, 2019; Ignatiadis and Huber, 2021; Cao et al., 2022; Zhang and Chen, 2022; Liang et al., 2023). As demonstrated in these studies, appropriately incorporating such side information can substantially improve both the power and the interpretability of simultaneous inference procedures. Let X_i denote the primary statistic and $s_i \in \mathbb{R}^l$ the side information. Upon observing $\{(X_i, s_i)\}_{i=1}^m$, we test $H_{0,i} : \theta_i = 0$ versus $H_{1,i} : \theta_i = 1$ for $i = 1, \dots, m$. To motivate our analysis, we model the data generation process as follows.

$$\theta_i | s_i \stackrel{\text{ind}}{\sim} \text{Ber}(\pi(s_i)), \quad X_i | s_i, \theta_i \stackrel{\text{ind}}{\sim} (1 - \theta_i)f_0(\cdot | s_i) + \theta_i f_1(\cdot | s_i), \quad i = 1, \dots, m. \quad (2)$$

Again, similarly as Cai et al. (2019); Liang et al. (2023), we do not assume that the data are generated exactly as described in Model (2). Such model only serves as a tool to motivate methodological development. As before, we assume the null distributions $f_0(\cdot | s_i)$ are known. While this assumption may appear strong at first glance, in practice the null distribution frequently does not depend on the auxiliary or side information variables (Cai et al., 2019; Leung and Sun, 2022). The ρ -value framework can be easily adapted to incorporate such

side information by introducing a weighting scheme that leverages the additional side information associated with each hypothesis. We define the ρ -values by $\rho_i = f_0(X_i|s_i)/g(X_i|s_i)$ for some density function $g(\cdot|s_i)$. Let $\eta : \mathbb{R}^l \rightarrow (0, 1)$ denote a predetermined function, and $c_i(\cdot)$ represent the null distribution of ρ_i . Then we incorporate the side information through a ρ -value weighting scheme by choosing an appropriate function $\eta(\cdot)$ to determine the weights. The detailed steps are summarized in Algorithm 3.

Algorithm 3 The ρ -BH procedure with side information

Input: $\{X_i\}_{i=1}^m$; $\{s_i\}_{i=1}^m$; predetermined density functions $\{g(\cdot|s_i)\}_{i=1}^m$; non-null proportions $\{\pi(s_i)\}_{i=1}^m$; predetermined $\{\eta(s_i)\}_{i=1}^m$; desired FDR level α .

1. for $i = 1$ **to** m **do:**

 Calculate the ρ -values $\rho_i = f_0(X_i|s_i)/g(X_i|s_i)$.

 Compute the null distribution functions of each ρ_i , and denote them by $\{c_i(\cdot)\}_{i=1}^m$.

 Let $w_i = \frac{\eta(s_i)}{1-\eta(s_i)}$, and compute the weighted ρ -values $q_i = \rho_i/w_i$.

end for

2. Sort the weighted ρ -values from smallest to largest $q_{(1)} \leq \dots \leq q_{(m)}$.

3. Let $k = \max_{1 \leq j \leq m} \left\{ \sum_{i=1}^m \{1 - \pi(s_i)\} c_i(q_{(j)} w_i) \leq \alpha j \right\}$.

Output: The rejection set $\{i = 1, \dots, m : q_i \leq q_{(k)}\}$.

In contrast to the ρ -BH procedure, Algorithm 3 is no longer equivalent to the adjusted BH procedure with $\{c_i(\rho_i)/w_i\}$'s as the inputs since the rankings produced by $\{c_i(\rho_i)/w_i\}$'s and $\{\rho_i/w_i\}$'s are different. The ideal choice of $g(\cdot|s_i)$ is again $f_1(\cdot|s_i)$, while the ideal choice of $\eta(\cdot)$ is $\pi(\cdot)$, and the rationale is provided as follows. Define the conditional local false discovery rate (Clfdr, Fu et al. (2022); Cai et al. (2019)) as

$$\text{Clfdr}_i \equiv \frac{\{1 - \pi(s_i)\} f_0(X_i|s_i)}{\{1 - \pi(s_i)\} f_0(X_i|s_i) + \pi(s_i) f_1(X_i|s_i)}.$$

Cai et al. (2019) shows that a ranking and thresholding procedure based on Clfdr is asymptotically optimal for FDR control. If we take $g(\cdot|s_i)$ to be $f_1(\cdot|s_i)$ and $\eta(\cdot)$ to be $\pi(\cdot)$, then the ranking produced by ρ_i/w_i 's is identical to that produced by Clfdr statistics. However,

the validity of the data-driven methods proposed in Cai et al. (2019) and Fu et al. (2022) relies on the consistent estimation of Clfdr_i 's. In many real applications, it is extremely difficult to accurately estimate Clfdr even when the dimension of s_i is moderate (Cai et al., 2022). In contrast, the mFDR guarantee of Algorithm 3 does not rely on any of such Clfdr consistency results and our proposal is valid under much weaker conditions as demonstrated by the next theorem.

Theorem 5. *Assume that $\{X_i, \theta_i\}_{i=1}^m$ are independent. Let $Q(t) = \frac{\sum_{i=1}^m \{1 - \pi(s_i)\} c_i(w_i t)}{E\{\sum_{i=1}^m \mathbb{I}(q_i \leq t)\}}$ and $t_L = \sup\{t > 0 : Q(t) \leq \alpha\}$. Based on the notations from Algorithm 3 and suppose*

$$(A5) \quad q_{(k)} \geq \nu \text{ and } \sum_{i=1}^m P(q_i \leq \nu) \rightarrow \infty \text{ as } m \rightarrow \infty, \text{ for some } \nu > 0;$$

$$(A6) \quad \limsup_{t \rightarrow 0^+} Q(t) < \alpha; \liminf_{t \rightarrow \infty} Q(t) > \alpha;$$

$$(A7) \quad \inf_{t \geq t_L + \epsilon_t} Q(t) \geq \alpha + \epsilon_\alpha, \text{ } Q(t) \text{ is strictly increasing in } t \in (t_L - \epsilon_t, t_L + \epsilon_t), \text{ for some constants } \epsilon_\alpha, \epsilon_t > 0.$$

Then we have $\lim_{m \rightarrow \infty} mFDR_{\text{Algorithm 3}} \leq \alpha$.

The validity of the above theorem allows flexible choices of functions $g(\cdot | s_i)$ and the weights w_i . Hence, similarly as the comparison between ρ -value and Lfdr, the ρ -value framework with side information is again much more flexible than the Clfdr framework that requires the consistent estimation of the Clfdr statistics.

We also remark that Cai et al. (2022) recommends using $\pi(s_i)/1 - \pi(s_i)$ as weights for p-values derived from two-sample t-statistics. However, their justification is primarily heuristic, and the advantage of using $\pi(s_i)/1 - \pi(s_i)$ over the alternative $1/1 - \pi(s_i)$ is not rigorously established. In contrast, we present the following optimality result, which offers a more principled justification for the use of $\pi(s_i)/1 - \pi(s_i)$.

Theorem 6. *Assume that $\{X_i, \theta_i\}_{i=1}^m$ are independent. Denote by δ_ρ the rule described in Algorithm 3 with $\eta(\cdot) = \pi(\cdot)$ and $g(\cdot | s_i) = f_1(\cdot | s_i)$, and let δ be any other rule that controls mFDR at level α asymptotically. Based on the notations from Algorithm 3 and suppose*

$$(A8) \quad \sum_{i=1}^m P(q_i \leq \frac{\alpha}{1-\alpha}) \rightarrow \infty \text{ as } m \rightarrow \infty.$$

Then we have $ETP(\boldsymbol{\delta}_\rho)/ETP(\boldsymbol{\delta}) \geq 1 + o(1)$.

Remark 4. Assumptions (A5)-(A7) are automatically satisfied under the conditions assumed by Theorem 6. Therefore, in such ideal setting, Algorithm 3 is optimal among all testing rules that asymptotically control mFDR at level α . In addition, Theorem 6 implies that the weighted BH procedure (Genovese et al., 2006) based on the ranking of $\{c_i(\rho_i)/w_i\}$ is suboptimal.

In practice, we need to choose $\eta(\cdot)$ and $g(\cdot|s_i)$ based on the available data $\{(X_i, s_i)\}_{i=1}^m$. Again, if the entire sample is used to construct $\eta(\cdot)$ and $g(\cdot|s_i)$, then the dependence among w_i 's and ρ_i 's is complicated. Similar to Algorithm 2, we can use sample splitting to circumvent this problem, and the details are provided in Algorithm 4. To ensure a valid mFDR control, we require a uniformly conservative consistent estimator of $\pi(\cdot)$, whose definition is given below.

Definition 3. An estimator $\hat{\pi}(\cdot)$ is a uniformly conservative consistent estimator of $\pi(\cdot)$, if $\sup_i E\{\hat{\pi}(s_i) - \tilde{\pi}(s_i)\}^2 \rightarrow 0$ as $m \rightarrow \infty$, where $0 \leq \tilde{\pi}(s_i) \leq \pi(s_i)$ for $i = 1, \dots, m$.

The problem of constructing such uniformly conservative consistent estimator $\hat{\pi}(\cdot)$ has been discussed in the literatures; see for example, Cai et al. (2019).

Algorithm 4 The data-driven ρ -BH procedure with side information

Input: $\{X_i\}_{i=1}^m$; $\{s_i\}_{i=1}^m$; desired mFDR level α .

1. Randomly split the data into two disjoint halves $\{X_i\}_{i=1}^m = \{X_{1,i}\}_{i=1}^{m_1} \cup \{X_{2,i}\}_{i=1}^{m_2}$, and $\{s_i\}_{i=1}^m = \{s_{1,i}\}_{i=1}^{m_1} \cup \{s_{2,i}\}_{i=1}^{m_2}$, where $m_1 = \lfloor m/2 \rfloor$.
2. Use $\{X_{1,i}\}_{i=1}^{m_1}$ and $\{s_{1,i}\}_{i=1}^{m_1}$ to construct the second half alternatives estimates $\{\hat{f}_1^{(2)}(\cdot|s_{2,i})\}_{i=1}^{m_2}$ and a uniformly conservative consistent estimate $\hat{\pi}_2(\cdot)$.
3. Run Algorithm 3 with $\{X_{2,i}\}_{i=1}^{m_2}$, $\{s_{2,i}\}_{i=1}^{m_2}$, $\{\hat{f}_1^{(2)}(\cdot|s_{2,i})\}_{i=1}^{m_2}$, $\{\hat{\pi}_2(s_{2,i})\}_{i=1}^{m_2}$, $\{\eta(s_{2,i})\}_{i=1}^{m_2}$, α as inputs, where $\eta(s_{2,i}) = \hat{\pi}_2(s_{2,i})$.
4. Switch the roles of $\{X_{1,i}\}_{i=1}^{m_1}$, $\{s_{1,i}\}_{i=1}^{m_1}$ and $\{X_{2,i}\}_{i=1}^{m_2}$, $\{s_{2,i}\}_{i=1}^{m_2}$. Repeat Steps 2 and 3, and combine the rejections.

Output: The combined set of rejections.

The next theorem shows that Algorithm 4 indeed controls mFDR at the target level asymptotically under conditions analogous to those assumed in Theorem 3.

Theorem 7. *Assume that $\{X_i, \theta_i\}_{i=1}^m$ are independent. Denote by $\{\hat{q}_{d,i}\}_{i=1}^{m_d}$, $\hat{q}_{d,(k_d)}$ and $\hat{\pi}_d$ the weighted ρ -values, selected thresholds and the estimated alternation proportions obtained from Algorithm 4, for the first and second halves of the data respectively, $d = 1, 2$. Denote by $\hat{c}_{d,i}$ the null distribution function for $\hat{\rho}_{d,i}$. Suppose $\hat{\pi}_d(s_i) > 0$ and $\sup_i E\{\hat{\pi}_d(s_i) - \tilde{\pi}_d(s_i)\}^2 \rightarrow 0$ for some $\tilde{\pi}_d(\cdot)$ satisfying $0 \leq \tilde{\pi}_d(\cdot) \leq \pi(\cdot)$, and let $\tilde{Q}_d(t) = \frac{\sum_{i=1}^{m_d} \{1 - \tilde{\pi}_d(s_{d,i})\} \hat{c}_{d,i}(w_{d,i}, t)}{E\{\sum_{i=1}^{m_d} \mathbb{I}(\hat{q}_{d,i} \leq t)\}}$ and $t_{d,L} = \sup\{t > 0 : \tilde{Q}_d(t) \leq \alpha\}$, $d = 1, 2$. Based on the notations from Algorithm 4 and suppose*

$$(A9) \quad \hat{q}_{d,(k_d)} \geq \nu, \quad \sum_{i=1}^{m_d} P(\hat{q}_{d,i} \leq \nu) \geq cm, \text{ for some constants } \nu, c > 0;$$

$$(A10) \quad \limsup_{t \rightarrow 0^+} \tilde{Q}_d(t) < \alpha, \quad \liminf_{t \rightarrow \infty} \tilde{Q}_d(t) > \alpha;$$

$$(A11) \quad \inf_{t \geq t_{d,L} + \epsilon_t} \tilde{Q}_d(t) \geq \alpha + \epsilon_\alpha, \quad \tilde{Q}_d(t) \text{ is strictly increasing in } t \in (t_{d,L} - \epsilon_t, t_{d,L} + \epsilon_t), \\ \text{for some constants } \epsilon_\alpha, \epsilon_t > 0.$$

Then we have $\lim_{m \rightarrow \infty} mFDR_{Algorithm 4} \leq \alpha$.

3 Numerical Experiments

In this section, we conduct several numerical experiments to compare our proposed procedures with some state-of-the-art methods. In all experiments, we study the general case where side information is available, and generate data according to the following hierarchical model:

$$\theta_i \stackrel{\text{ind}}{\sim} \text{Ber}\{\pi(s_i)\}, \quad X_i | s_i, \theta_i \stackrel{\text{ind}}{\sim} (1 - \theta_i)N(0, 1) + \theta_i f_1(\cdot | s_i), \quad (3)$$

where $\theta_i \in \mathbb{R}$, $X_i \in \mathbb{R}$ and $s_i \in \mathbb{R}^l$ for $i = 1, \dots, m$. Again, we test $H_{0,i} : \theta_i = 0$ versus $H_{1,i} : \theta_i = 1$ for $i = 1, \dots, m$. To implement our proposed data-driven procedure with side information, i.e., Algorithm 4, we use the following variation of the Storey estimator to estimate $\pi(s_i)$:

$$\hat{\pi}_2(s_{2,i}) = 1 - \frac{\sum_{j=1}^{m_1} K(s_{2,i}, s_{1,j}) \mathbb{I}(p_{1,j} \geq \tau)}{(1 - \tau) \sum_{j=1}^{m_1} K(s_{2,i}, s_{1,j})}, \quad i = 1, \dots, m_2, \quad (4)$$

where $p_{1,j} = 2\{1 - \Phi(|X_{1,j}|)\}$ is the two-sided p-value with Φ being the cumulative distribution function (cdf) of standard normal variable, and τ is chosen as the p-value threshold of the BH procedure at $\alpha = 0.9$; this ensures that the null cases are dominant in the set $\{j : p_{1,j} \geq \tau\}$. Let $K(s_{2,i}, s_{1,j}) = \phi_H(s_{2,i} - s_{1,j})$, where $\phi_H(\cdot)$ is the density of multivariate normal distribution with mean zero and covariance matrix H . We use the function `Hns` in the R package `ks` to chose H . Similar strategies for choosing τ and H are employed in [Cai et al. \(2022\)](#) and [Ma et al. \(2023\)](#). We construct $\hat{f}_1^{(2)}(\cdot | s_{2,i})$ using a modified version of the two-step approach proposed in [Fu et al. \(2022\)](#) as follows.

1. Let $\hat{\pi}'_1(s_{1,i}) = 1 - \frac{\sum_{j=1}^{m_1} K(s_{1,i}, s_{1,j}) \mathbb{I}(p_{1,j} \geq \tau)}{(1-\tau) \sum_{j=1}^{m_1} K(s_{1,i}, s_{1,j})}$ for $i = 1, \dots, m_1$.
2. Calculate $\tilde{f}_{1,2,j}(x_{1,j}) = \sum_{l=1}^{m_1} \frac{K(s_{1,j}, s_{1,l}) \phi_{h_x}(x_{1,j} - x_{1,l})}{\sum_{l=1}^{m_1} K(s_{1,j}, s_{1,l})}$,
and the weights $\hat{w}_{1,j} = 1 - \min \left\{ \frac{\{1 - \hat{\pi}'_1(s_{1,j})\} f_0(x_{1,j})}{\tilde{f}_{2,j}(x_{1,j})}, 1 \right\}$ for $j = 1, \dots, m_1$.
3. Obtain the non-null density estimate $\hat{f}_1^{(2)}(x | s_{2,i}) = \sum_{j=1}^{m_1} \frac{\hat{w}_{1,j} K(s_{2,i}, s_{1,j}) \phi_{h_x}(x - x_{1,j})}{\sum_{j=1}^{m_1} \hat{w}_{1,j} K(s_{2,i}, s_{1,j})}$ for $i = 1, \dots, m_2$.

Here, the kernel function K is the same as the one in Equation (4), and the bandwidth h_x is chosen automatically using the function `density` in the R package `stats`. We note that the distribution of null ρ_i can be complicated, making the analytical form of $c_i(\cdot)$ intractable. Nonetheless, since we can sample from the null hypothesis $H_{0,i}$ to generate as many null copies of ρ_i as needed, we are able to approximate $c_i(\cdot)$ to arbitrary accuracy. Particularly, to estimate the null densities $c_i(\cdot)$'s, i.e., the distribution functions of $f_0(\cdot | s_{2,i})/\hat{f}_1^{(2)}(\cdot | s_{2,i})$ under $H_{0,i}$ with f_0 being the density function of $N(0, 1)$, $i = 1, \dots, m_2$, we independently generate 1000 samples Y_j 's from $f_0(\cdot | s_{2,i})$ for each i and estimate $c_i(\cdot)$ through the empirical distribution of $f_0(Y_j | s_{2,i})/\hat{f}_1^{(2)}(Y_j | s_{2,i})$'s. The estimations on the first half of the data can be obtained by switching the roles of the first and the second halves.

We compare the performance of the following seven methods throughout the section:

- ρ -BH.OR: Algorithm 3 with $\rho_i = f_0(X_i | s_i)/f_1(X_i | s_i)$, $c_i(t) = P_{H_{0,i}}(\rho_i \leq t)$, $\eta(\cdot) = \pi(\cdot)$.

- ρ -BH.DD: Algorithm 4 with implementation details described above.
- LAWS: data-driven LAWS procedure (Cai et al., 2022) with p-value equals to $2\{1 - \Phi(|X_i|)\}$.
- CAMT: the CAMT procedure (Zhang and Chen, 2022) with the same p-values used in LAWS.
- BH: the BH Procedure (Benjamini and Hochberg, 1995) with the same p-values used in LAWS.
- e-BH: the e-BH Procedure (Wang and Ramdas, 2022) with reciprocal of ρ -values used in ρ -BH.DD as e-values.
- Clfdr: Clfdr based method (Fu et al., 2022) with $\text{Clfdr}_{d,i} = \hat{q}_{d,i}/(1 + \hat{q}_{d,i})$, where $d = 1, 2$ and $\hat{q}_{d,i}$'s are the weighted ρ -values used in ρ -BH.DD. Specifically, we calculate the threshold $k_d = \max_{1 \leq i \leq m_d} \{\sum_{j=1}^i \text{Clfdr}_{d,(j)}/i \leq \alpha\}$ and reject those with $\text{Clfdr}_{d,i} \leq \text{Clfdr}_{d,(k_d)}$.

All simulation results are based on 100 independent replications with target level $\alpha = 0.05$. The FDR is estimated by the average of the FDP, $\sum_{i=1}^m \{(1 - \theta_i \delta_i)/(\sum_{i=1}^m \delta_i \vee 1)\}$, and the average power is estimated by the average proportion of the true positives that are correctly identified, $\sum_{i=1}^m (\theta_i \delta_i)/\sum_{i=1}^m \theta_i$, both over the number of repetitions.

3.1 Bivariate side information

We first consider a similar setting as Setup S2 in Zhang and Chen (2022), where the non-null proportions and non-null distributions are closely related to a two dimensional covariate. Specifically, the parameters in Equation (3) are determined by the following equations.

$$s_i = (s_i^{(1)}, s_i^{(2)}) \stackrel{iid}{\sim} N(0, I_2), \quad \pi(s_i) = \frac{1}{1 + e^{k_{e,i}}}, \quad k_{e,i} = k_c + k_d s_i^{(1)}, \\ f_1(\cdot | s_i) \sim N\left(e^{k_t} 2e^{k_f s_i^{(2)}} / \{1 + e^{k_f s_i^{(2)}}\}, 1\right), \quad i = 1, \dots, 5000,$$

where I_2 is the 2×2 identity matrix, k_c , k_d , k_f and k_t are hyper-parameters that determine the impact of s_i on π and f_1 . In the experiments, we fix k_c at 2 or 1 (denoted as “Medium”

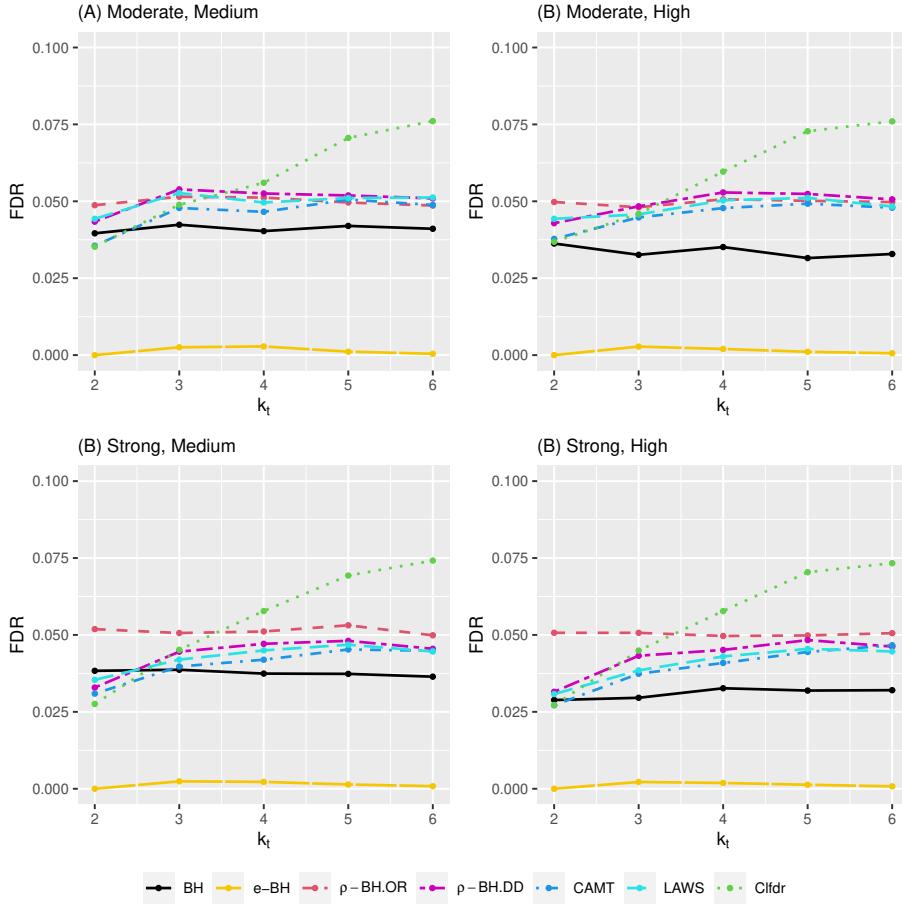


Figure 1: The empirical FDR of BH (black solid), e-BH (yellow extra-long dash), oracle ρ -BH (red dashed), data-driven ρ -BH (purple twodash), CAMT (dark blue dotdash), LAWS (light blue longdash) and Clfdr (green dotted) for the settings described in the bivariate scenario; $\alpha = 0.05$.

and ‘‘High’’, respectively), (k_d, f_f) at $(1.5, 0.4)$ or $(2.5, 0.6)$ (denoted as ‘‘Moderate’’ and ‘‘Strong’’, respectively), and vary k_t from 2 to 6. [Zhang and Chen \(2022\)](#) assumes it is known that $\pi(\cdot)$ and $f_1(\cdot|s_i)$ each depends on one coordinate of the covariate when implementing their procedure. Hence, for a fair comparison, we employ the same assumption, substitute $s_{d,i}$ by $s_{d,i}^{(1)}$ for the estimations of $\hat{\pi}(\cdot)$ (as defined in (4)) and $\hat{\pi}'(\cdot)$ (as defined in Step 1 of constructing $\hat{f}_1^{(2)}(\cdot|s_{2,i})$), and substitute $s_{d,i}$ by $s_{d,i}^{(2)}$ in the rest steps of obtaining $\hat{f}_1^{(2)}(\cdot|s_{2,i})$, for $d = 1, 2$.

It can be seen from Figure 1 that, except the Clfdr procedure, all other methods successfully control the FDR at the target level. Figure 2 shows that, the empirical powers

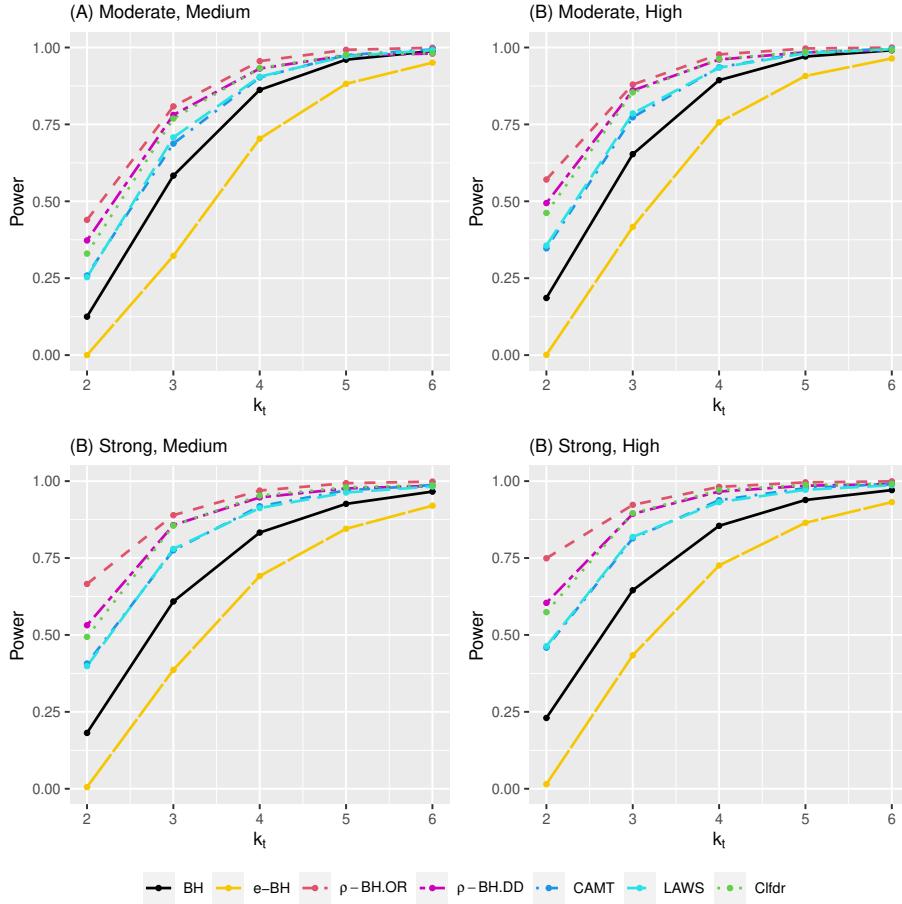


Figure 2: The empirical power comparison, with same legend as Figure 1.

of ρ -BH.OR and ρ -BH.DD are significantly higher than all other FDR controlled methods. It is not surprising that ρ -BH.OR and ρ -BH.DD outperform LAWS and BH, because the p -values only rely on the null distribution, whereas the ρ -values mimic the likelihood ratio statistics and encode the information from the alternative distribution. Both ρ -BH.OR and ρ -BH.DD outperforms CAMT as well, because CAMT uses a parametric model to estimate the likelihood ratio, while ρ -BH.DD employs a more flexible non-parametric approach that can better capture the structural information from the alternative distribution. Finally, as discussed in the previous sections, the Clfdr based approaches strongly rely on the estimation accuracy of $\pi(\cdot)$ and $f_1(\cdot|\cdot)$, which can be difficult in practice. Hence as expected, we observe severe FDR distortion of Clfdr method. Such phenomenon reflects the advantage of the proposed ρ -value framework because its FDR control can still be guaranteed even if

$\hat{f}_1(\cdot|\cdot)$ is far from the ground truth.

3.2 Univariate side information

Next, we consider the univariate covariate case and generate data as follows

$$\theta_i \stackrel{\text{ind}}{\sim} \text{Bernoulli}\{\pi(i)\}, \quad X_i | \theta_i \stackrel{\text{ind}}{\sim} (1 - \theta_i)N(0, 1) + \theta_i N(\mu, 1), \quad i = 1, \dots, 5000.$$

Two settings are considered. In Setting 1, the signals appear with elevated frequencies in the following blocks: $\pi(i) = 0.9$ for $i \in [1001, 1200] \cup [2001, 2200]$; $\pi(i) = 0.6$ for $i \in [3001, 3200] \cup [4001, 4200]$. For the rest of the locations we set $\pi(i) = 0.01$. We vary μ from 2 to 4 to study the impact of signal strength. In Setting 2, we set $\pi(i) = \pi_0$ in the above specified blocks and $\pi(i) = 0.01$ elsewhere. We fix $\mu = 3$ and vary π_0 from 0.5 to 0.9 to study the influence of sparsity levels. In these two cases, the side information s_i can be interpreted as the signal location i . When implementing CAMT, we use a spline basis with six equiquantile knots for $\pi(i)$ and $f_1(\cdot|i)$ to account for potential complex nonlinear effects as suggested in [Zhang and Chen \(2022\)](#) and [Lei and Fithian \(2018\)](#).

We compare the seven procedures as in Section 3.1, and the results of Settings 1 and 2 are summarized in the first and second rows of Figure 3, respectively. We can see from the first column of Figure 3 that, in both settings all methods control FDR appropriately at the target level. From the second column of Figure 3, it can be seen that both ρ -BH.OR and ρ -BH.DD outperform the other five methods. This is due to the fact that, besides the ability in incorporating the sparsity information, the ρ -value statistic also adopts other structural knowledge and is henceforth more informative than the p-value based methods. In addition, the nonparametric approach employed by ρ -BH.DD is better at capturing nonlinear information than the parametric model used in CAMT, leads to a more powerful procedure.

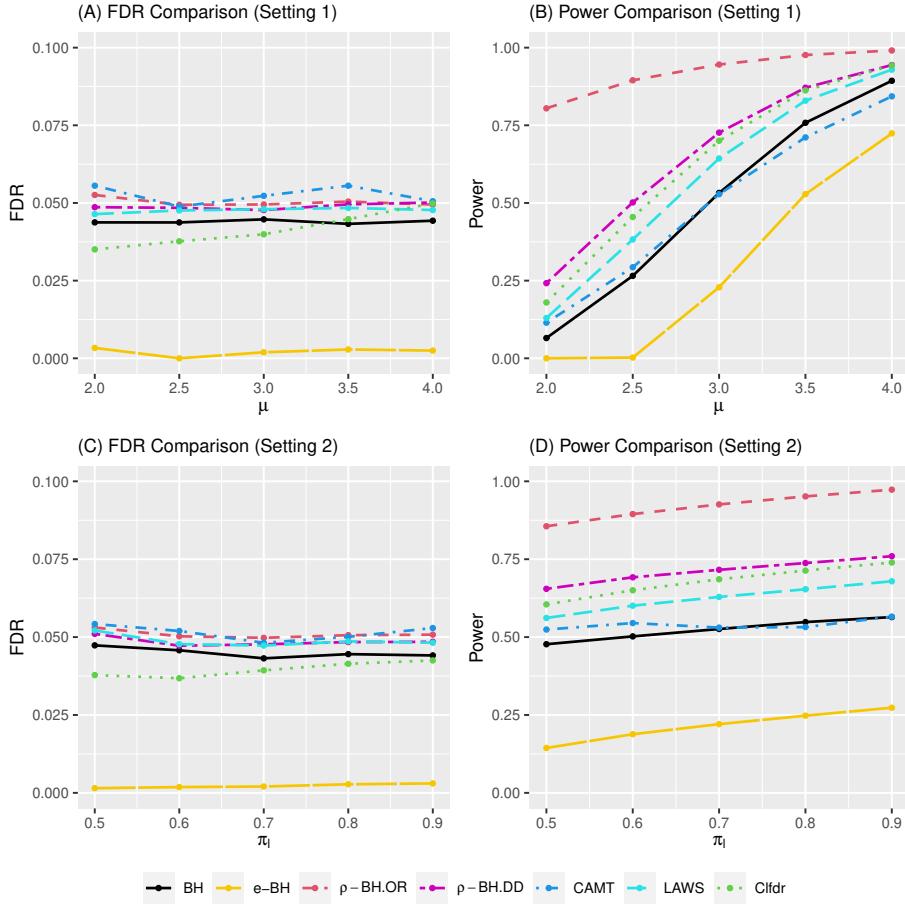


Figure 3: The empirical FDR and power for univariate scenario, with the same legends as in Figure 1.

4 Data Analysis

In this section, we compare the performance of ρ -BH.DD with Clfdr, CAMT, LAWS, BH, and e-BH on two real datasets.

4.1 MWAS data

We first analyze a dataset from a microbiome-wide association study (MWAS) of sex effect (McDonald et al., 2018), which is available at <https://github.com/knightlab-analyses/american-gut-analyses>. The aim of the study is to distinguish the abundant bacteria in the gut microbiome between males and females by the sequencing of a fingerprint gene in the bacteria 16S rRNA gene. This dataset is also analyzed in Zhang and Chen (2022). We

follow their preprocessing procedure to obtain 2492 p-values from Wilcoxon rank sum test for different OTUs, and the percentages of zeros across samples for the OTUs are considered as the univariate side information. Because a direct estimation of the non-null distributions of the original Wilcoxon rank sum test statistics is difficult, we construct pseudo z-values by $z_i = \Phi^{-1}(p_i) \times (2B_i - 1)$, where B_i 's are independent Bernoulli(0.5) random variables and Φ^{-1} is the inverse of standard normal cdf. Then we run ρ -BH.DD on those pseudo z-values by employing the same estimation methods of $\pi(\cdot)$ and $f_1(\cdot|\cdot)$ as described in Section 3. When implementing CAMT, we use the spline basis with six equiquantile knots as the covariates as recommended in [Zhang and Chen \(2022\)](#). The results are summarized in Figure 4 (A). We can see that ρ -BH.DD rejects significantly more hypotheses than LAWS and BH across all FDR levels, while e-BH procedure fails to reject any hypotheses. ρ -BH.DD also rejects slightly more tests than Clfdr under most FDR levels, and is more stable than CAMT. Because Clfdr may suffer from possible FDR inflation as shown in the simulations, we conclude that ρ -BH.DD enjoys the best performance on this dataset.

4.2 ADHD data

We next analyze a preprocessed magnetic resonance imaging (MRI) data for a study of attention deficit hyperactivity disorder (ADHD). The dataset is available at <http://neurobureau.projects.nitrc.org/ADHD200/Data.html>. We adopt the Gaussian filter-blurred skullstripped gray matter probability images from the Athena Pipeline, which are MRI images with a resolution of $197 \times 233 \times 189$. We pool the 776 training samples and 197 testing samples together, remove 26 samples with no ADHD index, and split the pooled data into one ADHD sample of size 585 and one normal sample of size 362. Then we downsize the resolution of images to $30 \times 36 \times 30$ by taking the means of pixels within blocks, and then obtain $30 \times 36 \times 30$ two-sample t-test statistics. Similar data preprocessing strategy is also used in [Cai et al. \(2022\)](#). In such dataset, the 3-dimensional coordinate indices can be employed as the side information. The results of the five methods are summarized in Figure 4 (B). Again, we see that ρ -BH.DD rejects more hypotheses than CAMT, LAWS, Clfdr, BH and e-BH across all FDR levels.

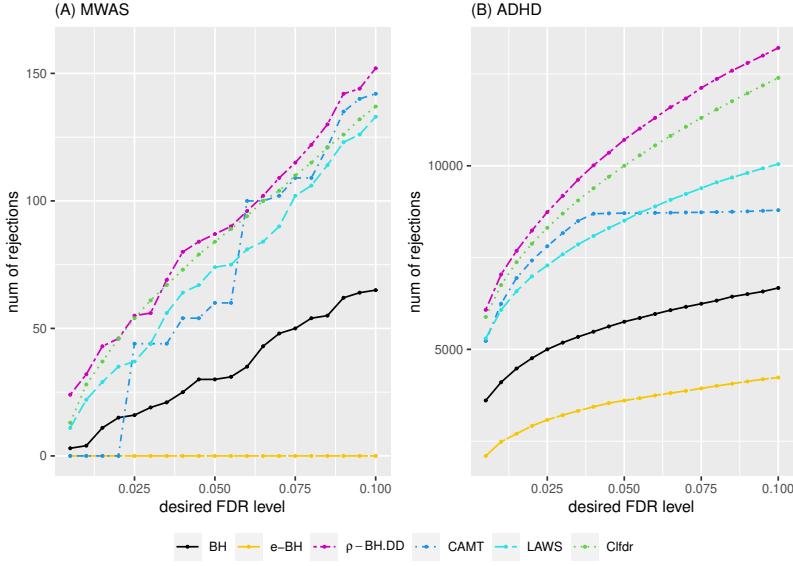


Figure 4: Total numbers of rejections for BH (black solid), e-BH (yellow extra-long dash), data-driven ρ -BH (purple twodash), CAMT (dark blue dotdash), LAWS (light blue long-dash) and Clfdr (green dotted) at various FDR levels, respectively for MWAS (Left) and ADHD (Right) datasets.

5 Discussions

This article introduces a novel multiple testing framework based on the newly proposed ρ -values. The strengths of this framework lie in its ability to unify existing procedures based on p-values and local false discovery rate (Lfdr) statistics, while requiring substantially weaker conditions than those typically imposed by Lfdr-based methods. Moreover, the framework naturally extends to incorporate side information through appropriate weighting schemes, under which asymptotic optimality can still be achieved.

As a concluding remark, we emphasize that the frameworks based on p-values and Lfdr statistics are not as fundamentally different as often portrayed in the literature. A central message of [Storey et al. \(2007\)](#), [Sun and Cai \(2007\)](#), and [Leung and Sun \(2022\)](#) is that reducing z-values to p-values can result in substantial information loss, implicitly framing Lfdr and p-values as two distinct statistical paradigms. However, as we have shown in Section 2.2, the $c(\rho)$ is a special case of a p-value but in the meantime it can yield the same ranking as the Lfdr. This suggests a more accurate interpretation of their message: Statistics that incorporate information from the alternative distribution outperform those

that do not.

To be more concrete, we show below that a Lfdr based procedure proposed in [Leung and Sun \(2022\)](#) is actually a special variation of Algorithm 1 under Model (1). Suppose f_1 and f_0 are known but π is not. As mentioned in Section 2.3, a natural choice of $\hat{\pi}$ is the Storey estimator. Note that, the Storey estimator requires a predetermined tuning parameter τ . By replacing π with $\hat{\pi}$, the threshold in the ρ -BH procedure becomes $\rho_{(k)}$ where $k = \max_j \{(1 - \hat{\pi})c(\rho_{(j)}) \leq \alpha j/m\}$. In a special case when we allow varying τ for different j and let $\tau = c(\rho_{(j)})$, then it yields that $k = \max_j \{\#\{i : c(\rho_i) \geq 1 - c(\rho_{(j)})\} \leq \alpha j\}$. Now if we add 1 to the numerator and let $k = \max_j \{1 + \#\{i : c(\rho_i) \geq 1 - c(\rho_{(j)})\} \leq \alpha j\}$, then the decision rule $\boldsymbol{\delta} = \{\mathbb{I}(\rho_i \leq \rho_{(k)})\}_{i=1}^m$ is equivalent to the rule given by the ZAP procedure ([Leung and Sun, 2022](#)) that is based on Lfdr. Hence, the ZAP procedure can be viewed as a special case of the ρ -BH procedure under Model (1), and can be unified into the proposed ρ -BH framework.

Appendix A Weighted ρ -BH procedures

In this section, we present the weighted ρ -BH procedures and the corresponding theoretical results.

Similar to incorporating prior information via a p-value weighting scheme (e.g., Benjamini and Hochberg, 1997; Genovese et al., 2006; Dobriban et al., 2015), we can also employ such weighting strategy in the current ρ -value framework. Let $\{w_i\}_{i=1}^m$ be a set of positive weights such that $\sum_{i=1}^m w_i = m$. The weighted BH procedure proposed in Genovese et al. (2006) uses p_i/w_i 's as the inputs of the original BH procedure. Genovese et al. (2006) proves that, if p_i 's are independent and $\{w_i\}_{i=1}^m$ are independent of $\{p_i\}_{i=1}^m$ conditional on $\{\theta_i\}_{i=1}^m$, then the weighted BH procedure controls FDR at level less than or equal to $(1 - \pi)\alpha$.

Following their strategy, we can apply the adjusted weighted BH procedure (with $1 - \pi$ adjustment) to $c(\rho_i)$'s and obtain the same FDR control result. However, such procedure might be suboptimal as explained in the paper. Alternatively, we derive a weighted ρ -BH procedure (without requiring $\sum_{i=1}^m w_i = m$) and the details are presented in Algorithm 5.

Algorithm 5 Weighted ρ -BH procedure

Input: $\{X_i\}_{i=1}^m$; a predetermined density function $g(\cdot)$; non-null proportion π ; predetermined weights $\{w_i\}_{i=1}^m$; desired FDR level α .

1. for $i = 1$ **to** m :

 Calculate the ρ -values $\rho_i = f_0(X_i)/g(X_i)$.

 Compute the weighted ρ -values $q_i = \rho_i/w_i$.

end for

2. Sort the weighted ρ -values from smallest to largest $q_{(1)} \leq \dots \leq q_{(m)}$.

3. Compute the null distribution function of ρ_i 's, and denote it by $c(\cdot)$.

4. Let $k = \max_{1 \leq j \leq m} \{\sum_{i=1}^m (1 - \pi)c(q_{(j)}w_i) \leq \alpha j\}$.

Output: The rejection set $\{i = 1, \dots, m : q_i \leq q_{(k)}\}$.

Note that $\{\rho_i/w_i\}_{i=1}^m$ in Algorithm 5 produces a different ranking than $\{c(\rho_i)/w_i\}_{i=1}^m$, which may improve the power of the weighted p-value procedure with proper choices of ρ -

values and weights. On the other hand, the non-linearity of $c(\cdot)$ imposes challenges on the theoretical guarantee for the mFDR control of Algorithm 5 compared to that of Genovese et al. (2006), and we derive the following result based on similar assumptions as in Theorem 3.

Theorem 8. Assume that $\{X_i, \theta_i\}_{i=1}^m$ are independent. Denote by $Q(t) = \frac{\sum_{i=1}^m (1-\pi)c(w_i t)}{E\{\sum_{i=1}^m \mathbb{I}(q_i \leq t)\}}$ and $t_L = \sup\{t > 0 : Q(t) \leq \alpha\}$. Based on the notations from Algorithm 5 and suppose

- (A12) $q_{(k)} \geq \nu$ and $\sum_{i=1}^m P(q_i \leq \nu) \rightarrow \infty$ as $m \rightarrow \infty$, for some $\nu > 0$;
- (A13) $\limsup_{t \rightarrow 0^+} Q(t) < \alpha$; $\liminf_{t \rightarrow \infty} Q(t) > \alpha$;
- (A14) $\inf_{t \geq t_L + \epsilon_t} Q(t) \geq \alpha + \epsilon_\alpha$, $Q(t)$ is strictly increasing in $t \in (t_L - \epsilon_t, t_L + \epsilon_t)$, for some constants $\epsilon_\alpha, \epsilon_t > 0$.

Then we have $\lim_{m \rightarrow \infty} mFDR_{Algorithm 5} \leq \alpha$.

It is worthwhile to note that, Genovese et al. (2006) requires $\sum_{i=1}^m w_i = m$, which makes the weighted p-value procedure conservative. In comparison, Algorithm 5 no longer requires such condition, and it employs a tight estimate of the FDP that leads to a more powerful testing procedure.

When the oracle parameters are unknown, we can similarly construct a data-driven weighted ρ -BH procedure with an additional data splitting step as in Algorithm 2. We describe it in Algorithm 6.

The next theorem provides the theoretical guarantee for the asymptotic mFDR control of Algorithm 6.

Theorem 9. Assume that $\{X_i, \theta_i\}_{i=1}^m$ are independent. Denote by $\{\hat{q}_{d,i}\}_{i=1}^{m_d}$, $\hat{q}_{d,(k_d)}$ and $\hat{\pi}_d$ the weighted ρ -values, selected thresholds and the estimated alternative proportions obtained from Algorithm 6, for the first and second halves of the data respectively, $d = 1, 2$. Denote by \hat{c}_d the null distribution function for $\hat{\rho}_{d,i}$. Suppose $\hat{\pi}_d > 0$ and $|\hat{\pi}_d - \tilde{\pi}_d| \xrightarrow{P} 0$ for some $\tilde{\pi}_d$ satisfying $0 \leq \tilde{\pi}_d \leq \pi$, and let $\tilde{Q}_d(t) = \frac{\sum_{i=1}^{m_d} (1-\tilde{\pi}_d)\hat{c}_d(w_{d,i} t)}{E\{\sum_{i=1}^{m_d} \mathbb{I}(\hat{q}_{d,i} \leq t)\}}$ and $t_{d,L} = \sup\{t > 0 : \tilde{Q}_d(t) \leq \alpha\}$, $d = 1, 2$. Based on the notations from Algorithm 6 and suppose

Algorithm 6 The data-driven weighted ρ -BH procedure

Input: $\{X_i\}_{i=1}^m$; predetermined weights $\{w_i\}_{i=1}^m$; desired FDR level α .

1. Randomly split the data into two disjoint halves $\{X_i\}_{i=1}^m = \{X_{1,i}\}_{i=1}^{m_1} \cup \{X_{2,i}\}_{i=1}^{m_2}$, $\{w_i\}_{i=1}^m = \{w_{1,i}\}_{i=1}^{m_1} \cup \{w_{2,i}\}_{i=1}^{m_2}$, where $m_1 = \lfloor m/2 \rfloor$.
2. Use $\{X_{1,i}\}_{i=1}^{m_1}$ to construct the second half alternative estimate $\hat{f}_1^{(2)}$ and a conservative consistent estimate $\hat{\pi}_2$.
3. Run Algorithm 5 with $\{X_{2,i}\}_{i=1}^{m_2}$, $\hat{f}_1^{(2)}$, $\hat{\pi}_2$, $\{w_{2,i}\}_{i=1}^{m_2}$, α as inputs.
4. Switch the roles of $\{X_{1,i}\}_{i=1}^{m_1}$, $\{w_{1,i}\}_{i=1}^{m_1}$ and $\{X_{2,i}\}_{i=1}^{m_2}$, $\{w_{2,i}\}_{i=1}^{m_2}$. Repeat Steps 2 and 3, and combine the rejections.

Output: The combined rejection set.

$$(A15) \quad \hat{q}_{d,(k_d)} \geq \nu, \sum_{i=1}^{m_d} P(\hat{q}_{d,i} \leq \nu) \geq cm, \text{ for some constants } \nu, c > 0;$$

$$(A16) \quad \limsup_{t \rightarrow 0^+} \tilde{Q}_d(t) < \alpha, \liminf_{t \rightarrow \infty} \tilde{Q}_d(t) > \alpha;$$

$$(A17) \quad \inf_{t \geq t_{d,L} + \epsilon_t} \tilde{Q}_d(t) \geq \alpha + \epsilon_\alpha, \tilde{Q}_d(t) \text{ is strictly increasing in } t \in (t_{d,L} - \epsilon_t, t_{d,L} + \epsilon_t), \\ \text{for some constants } \epsilon_\alpha, \epsilon_t > 0.$$

Then we have $\lim_{m \rightarrow \infty} mFDR_{Algorithm 6} \leq \alpha$.

Appendix B Proofs of Main Theorems and Propositions

Note that Theorems 2 and 4 follow directly from the proof of the original BH procedure as discussed in the main text. Theorem 1 is a special case of Theorem 6. Theorems 3 and 9 are special cases of Theorem 7 with slight modifications. Theorem 8 is a special case of Theorem 5. Hence, we focus on the proofs of Proposition 1, Theorem 5, Theorem 6 and Theorem 7 in this section.

To simplify notation, we define a procedure equivalent to Algorithm 3. This equivalence is stated in Lemma 1, whose proof will be given later in C.

Lemma 1. *Algorithm 3 and Algorithm 7 are equivalent in the sense that they reject the same set of hypotheses.*

Algorithm 7 A procedure equivalent to Algorithm 3

Input: $\{X_i\}_{i=1}^m$; $\{s_i\}_{i=1}^m$; predetermined density functions $\{g(\cdot|s_i)\}_{i=1}^m$; non-null proportions $\{\pi(s_i)\}_{i=1}^m$; predetermined $\{\eta(s_i)\}_{i=1}^m$; desired FDR level α .

1. for $i = 1$ **to** m :

Calculate the ρ -values $\rho_i = f_0(X_i|s_i)/g(X_i|s_i)$.

Compute the null distribution functions of each ρ_i , and denote them by $\{c_i(\cdot)\}_{i=1}^m$.

Let $w_i = \frac{\eta(s_i)}{1-\eta(s_i)}$, and compute the weighted ρ -values $q_i = \rho_i/w_i$, $i = 1, \dots, m$.

end for

2. Let $t^* = \max_{t \geq 0} \left\{ \frac{\sum_{i=1}^m \{1-\pi(s_i)\} c_i(w_i t)}{\{\sum_{i=1}^m \mathbb{I}(q_i \leq t)\} \vee 1} \leq \alpha \right\}$.

Output: The set of rejections $\{i = 1, \dots, m : q_i \leq t^*\}$.

B.1 Proof of Proposition 1

Proof. Denote by $\hat{P}\{c(\rho_i) > \tau\} := \frac{\sum_{i=1}^m \mathbb{I}\{c(\rho_i) > \tau\}}{m}$. Since $\sum_{i=1}^m \mathbb{I}\{c(\rho_i) > \tau\}$ follows Binomial(m, p) where $p = P\{c(\rho_i) > \tau\}$, we have that

$$\hat{P}\{c(\rho_i) > \tau\} \xrightarrow{P} p.$$

Let $p_0 = P\{c(\rho_i) > \tau | H_{0,i}\}$ and $p_1 = P\{c(\rho_i) > \tau | H_{1,i}\}$, then $p = (1-\pi)p_0 + \pi p_1$. Since $c(\rho_i) \sim \text{Unif}(0, 1)$ under $H_{0,i}$, it follows that $p = (1-\pi)(1-\tau) + \pi p_1$. Hence, $1-p/(1-\tau) < \pi$, and the proposition follows. \square

B.2 Proof of Theorem 5

Proof. By Lemma 1, we only need to prove the mFDR control for Algorithm 7. Assumption (A5) ensures that $Q(t)$ is well defined when $t \geq \nu$. Note that, by Assumption (A5) and standard Chernoff bound for independent Bernoulli random variables, we have uniformly

for $t \geq \nu$ and any $\epsilon > 0$

$$\begin{aligned} & P\left(\left|\frac{\sum_{i=1}^m \mathbb{I}(q_i \leq t)}{E\{\sum_{i=1}^m \mathbb{I}(q_i \leq t)\}} - 1\right| \geq \epsilon\right) \\ & \leq 2e^{-\epsilon^2 \sum_{i=1}^m P(q_i \leq t)/3} \\ & \leq 2e^{-\epsilon^2 \sum_{i=1}^m P(q_i \leq \nu)/3} \rightarrow 0, \end{aligned}$$

which implies

$$\sup_{t \geq \nu} |Q(t) - \widehat{\text{FDP}}(t)| \xrightarrow{P} 0 \quad (5)$$

as $m \rightarrow \infty$, where $\widehat{\text{FDP}}(t) = \frac{\sum_{i=1}^m \{1 - \pi(s_i)\} c_i(w_i t)}{\{\sum_{i=1}^m \mathbb{I}(q_i \leq t)\} \vee 1}$.

Assumption (A6) implies $t_L < \infty$. Moreover, combining Equation (5) with Assumption (A7), we have $\widehat{\text{FDP}}(t) > \alpha$ for any $t \geq t_L + \epsilon_t$ with probability going to 1. Thus, we only have to consider $t < t_L + \epsilon_t$. Specifically, we consider $t \in (t_L - \epsilon_t, t_L + \epsilon_t)$. As $Q(t)$ is strictly increasing within this range by Assumption (A7), we have

$$t^* = Q^{-1}\{Q(t^*)\} \xrightarrow{P} Q^{-1}\{\widehat{\text{FDP}}(t^*)\} = Q^{-1}(\alpha) = t_L.$$

Therefore, we have

$$\begin{aligned} \text{mFDR}_{\text{Algorithm 7}} &= \frac{\sum_{i=1}^m P(q_i \leq t^*, \theta_i = 0)}{E\{\sum_{i=1}^m \mathbb{I}(q_i \leq t^*)\}} \\ &= \frac{\sum_{i=1}^m P(q_i \leq t_L, \theta_i = 0)}{E\{\sum_{i=1}^m \mathbb{I}(q_i \leq t_L)\}} + o(1) \\ &= Q(t_L) + o(1) \leq \alpha + o(1). \end{aligned} \quad (6)$$

□

B.3 Proof of Theorem 6

We first state a useful lemma whose proof will be given later in C.

Lemma 2. *Let $g(\cdot | s_i) \equiv f_1(\cdot | s_i)$, $i = 1, \dots, m$, $\eta(\cdot) \equiv \pi(\cdot)$. For any $t > 0$, let*

$$Q(t) = \frac{\sum_{i=1}^m \{1 - \pi(s_i)\} c_i(w_i t)}{E\{\sum_{i=1}^m \mathbb{I}(q_i \leq t)\}}, \quad t_L = \sup\{t \in (0, \infty) : Q(t) \leq \alpha\}.$$

Suppose Assumption (A8) holds. Then we have

1. $Q(t) < \frac{t}{1+t}$;
2. $Q(t)$ is strictly increasing;
3. $\lim_{m \rightarrow \infty} (ETP_{\delta^L} - ETP_{\delta'}) \geq 0$, for any testing rule δ' based on $\{X_i\}_{i=1}^m$ and $\{s_i\}_{i=1}^m$ such that $\lim_{m \rightarrow \infty} mFDR_{\delta'} \leq \alpha$, where $\delta^L = \{\mathbb{I}(q_i \leq t_L)\}_{i=1}^m$.

Next we prove Theorem 6.

Proof. By Lemma 1, Algorithm 3 is equivalent to reject all hypotheses that satisfying $q_i \leq t^*$, where t^* is the threshold defined in Algorithm 7. To simplify notations, let $\nu = \frac{\alpha}{1-\alpha}$ and we next show that $t^* \geq \nu$ in probability.

By the standard Chernoff bound for independent Bernoulli random variables, we have

$$P\left(\left|\frac{\sum_{i=1}^m \mathbb{I}(q_i \leq \nu)}{\sum_{i=1}^m P(q_i \leq \nu)} - 1\right| \geq \epsilon\right) \leq 2e^{-\epsilon^2 \sum_{i=1}^m P(q_i \leq \nu)/3}$$

for all $0 < \epsilon < 1$. By Assumption (A8), the above implies

$$\left|\frac{\sum_{i=1}^m \mathbb{I}(q_i \leq \nu)}{\sum_{i=1}^m P(q_i \leq \nu)} - 1\right| = o_P(1). \quad (7)$$

Combining Equation (7) and the first part of Lemma 2, we have

$$\begin{aligned} \frac{\sum_{i=1}^m \{1 - \pi(s_i)\} c_i(w_i \nu)}{\{\sum_{i=1}^m \mathbb{I}(q_i \leq \nu)\} \vee 1} &= \frac{\sum_{i=1}^m \{1 - \pi(s_i)\} c_i(w_i \nu)}{\sum_{i=1}^m P(q_i \leq \nu)} + o_P(1) \\ &= Q(\nu) + o_P(1) < \frac{\nu}{1+\nu} + o_P(1) \\ &= \alpha + o_P(1), \end{aligned} \quad (8)$$

which implies $P(t^* \geq \nu) \rightarrow 1$. Therefore, we will only focus the event $\{t^* \geq \nu\}$ in the following proof.

For any $t > 0$, we let

$$Q(t) = \frac{\sum_{i=1}^m \{1 - \pi(s_i)\} c_i(w_i t)}{E\{\sum_{i=1}^m \mathbb{I}(q_i \leq t)\}},$$

$$\widehat{\text{FDP}}(t) = \frac{\sum_{i=1}^m \{1 - \pi(s_i)\} c_i(w_i t)}{\{\sum_{i=1}^m \mathbb{I}(q_i \leq t)\} \vee 1},$$

and

$$t_L = \sup\{t \in (0, \infty) : Q(t) \leq \alpha\},$$

$$t^* = \sup\{t \in (0, \infty) : \widehat{\text{FDP}}(t) \leq \alpha\}.$$

Following the proof of the third part of Lemma 2, we consider two cases: $\lim_{m \rightarrow \infty} \frac{\pi(s_i)}{m} \leq 1 - \alpha$ and $\lim_{m \rightarrow \infty} \frac{\pi(s_i)}{m} > 1 - \alpha$.

The first case is trivial by noting that mFDR can be controlled even if we reject all null hypotheses. For the second case, we need to show that $t^* \xrightarrow{P} t_L$. Similar to the proof of Equation (7), we have uniformly for $t \geq \nu$ and any $\epsilon > 0$,

$$P\left(\left|\frac{\sum_{i=1}^m \mathbb{I}(q_i \leq t)}{E\{\sum_{i=1}^m \mathbb{I}(q_i \leq t)\}} - 1\right| \geq \epsilon\right)$$

$$\leq 2e^{-\epsilon^2 \sum_{i=1}^m P(q_i \leq t)/3}$$

$$\leq 2e^{-\epsilon^2 \sum_{i=1}^m P(q_i \leq \nu)/3} \rightarrow 0,$$

which implies $|\widehat{\text{FDP}}(t) - Q(t)| \xrightarrow{P} 0$ uniformly in $t \geq \nu$. Thus, $\widehat{\text{FDP}}(t^*) \xrightarrow{P} Q(t^*)$. Moreover, by Lemma 2, we know that $Q(t)$ is continuous and strictly increasing. Therefore, we can define the inverse function $Q^{-1}(\cdot)$ of $Q(\cdot)$. Thus, by the continuous mapping theorem, we have

$$t^* = Q^{-1}\{Q(t^*)\} \xrightarrow{P} Q^{-1}\{\widehat{\text{FDP}}(t^*)\} = Q^{-1}(\alpha) = t_L.$$

By the third part of Lemma 2, we have $\lim_{m \rightarrow \infty} (\text{ETP}_{\delta_L} - \text{ETP}_{\delta}) \geq 0$ and therefore,

$$\begin{aligned} \lim_{m \rightarrow \infty} \frac{\text{ETP}_{\delta_\rho}}{\text{ETP}_{\delta}} &= \lim_{m \rightarrow \infty} \frac{\text{ETP}_{\delta_\rho}}{\text{ETP}_{\delta_L}} \frac{\text{ETP}_{\delta_L}}{\text{ETP}_{\delta}} \geq \lim_{m \rightarrow \infty} \frac{\text{ETP}_{\delta_\rho}}{\text{ETP}_{\delta_L}} \\ &= \lim_{m \rightarrow \infty} \frac{E\{\sum_{i=1}^m \theta_i \mathbb{I}(\rho_i \leq t^*)\}}{E\{\sum_{i=1}^m \theta_i \mathbb{I}(\rho_i \leq t_L)\}} \\ &\geq 1 + \lim_{m \rightarrow \infty} \frac{\sum_{i=1}^m \{1 - \pi(s_i)\} o(1)}{\sum_{i=1}^m \{1 - \pi(s_i)\} c_i(w_i t_L)} \geq 1. \end{aligned}$$

□

B.4 Proof of Theorem 7

We first introduce Lemma 3, whose proof will be given later in C.

Lemma 3. Denote Steps 2 to 3 of Algorithm 4 as ‘Half-procedure’ and we inherit all other notations from Theorem 7. Suppose Assumptions (A9)-(A11) hold for $d = 2$. Then we have

$$\lim_{m \rightarrow \infty} mFDR_{\text{Half-procedure}} \leq \alpha.$$

Next we prove Theorem 7.

Proof. Without loss of generality, we assume $\{X_{1,i}\}_{i=1}^{m_1} = \{X_i\}_{i=1}^{m_1}$ and $\{X_{2,i}\}_{i=1}^{m_2} = \{X_i\}_{i=m_1+1}^m$.

By Lemma 1 and Lemma 3, we have that

$$\begin{aligned} \frac{E\{\sum_{i=1}^{m_1} (1 - \theta_i) \delta_i\}}{E\{\sum_{i=1}^{m_1} \delta_i\}} &\leq \alpha + o(1), \\ \frac{E\{\sum_{i=m_1+1}^m (1 - \theta_i) \delta_i\}}{E\{\sum_{i=m_1+1}^m \delta_i\}} &\leq \alpha + o(1). \end{aligned} \tag{9}$$

On the other hand, we can decompose mFDR_δ as

$$\begin{aligned}
\text{mFDR}_\delta &= \frac{E\{\sum_{i=1}^m (1 - \theta_i)\delta_i\}}{E\{\sum_{i=1}^m \delta_i\}} \\
&= \frac{E\{\sum_{i=1}^{m_1} (1 - \theta_i)\delta_i\}}{E\{\sum_{i=1}^m \delta_i\}} + \frac{E\{\sum_{i=m_1+1}^m (1 - \theta_i)\delta_i\}}{E\{\sum_{i=1}^m \delta_i\}} \\
&= \frac{E\{\sum_{i=1}^{m_1} (1 - \theta_i)\delta_i\}}{E\{\sum_{i=1}^{m_1} \delta_i\}} \frac{E\{\sum_{i=1}^{m_1} \delta_i\}}{E\{\sum_{i=1}^m \delta_i\}} + \\
&\quad \frac{E\{\sum_{i=m_1+1}^m (1 - \theta_i)\delta_i\}}{E\{\sum_{i=m_1+1}^m \delta_i\}} \frac{E\{\sum_{i=m_1+1}^m \delta_i\}}{E\{\sum_{i=1}^m \delta_i\}}.
\end{aligned} \tag{10}$$

Therefore, by Equations (9) and (10), we conclude that

$$\lim_{m \rightarrow \infty} \text{mFDR}_\delta \leq \alpha \left\{ \frac{E(\sum_{i=1}^{m_1} \delta_i)}{E(\sum_{i=1}^m \delta_i)} + \frac{E(\sum_{i=m_1+1}^m \delta_i)}{E(\sum_{i=1}^m \delta_i)} \right\} = \alpha.$$

□

Appendix C Proofs of Lemmas

C.1 Proof of Lemma 1

Proof. It is easy to see that $t^* \geq q_{(k)}$ as

$$\frac{\sum_{i=1}^m \{1 - \pi(s_i)\} c_i(w_i q_{(k)})}{\sum_{i=1}^m \mathbb{I}(q_i \leq q_{(k)})} \leq \alpha.$$

Now it suffices to show that, for any $t \geq q_{(k+1)}$, we have

$$\frac{\sum_{i=1}^m \{1 - \pi(s_i)\} c_i(w_i t)}{\sum_{i=1}^m \mathbb{I}(q_i \leq t)} > \alpha. \tag{11}$$

By the definition of k , for any $l \geq k + 1$, we have

$$\frac{\sum_{i=1}^m \{1 - \pi(s_i)\} c_i(w_i q_{(l)})}{\sum_{i=1}^m \mathbb{I}(q_i \leq q_{(l)})} > \alpha.$$

Then for any $l \geq k + 1$, for any $t \in [q_{(l)}, q_{(l+1)}]$ where $q_{(m+1)} = \infty$, we have

$$\begin{aligned} \frac{\sum_{i=1}^m \{1 - \pi(s_i)\} c_i(w_i t)}{\sum_{i=1}^m \mathbb{I}(q_i \leq t)} &= \frac{\sum_{i=1}^m \{1 - \pi(s_i)\} c_i(w_i t)}{l} \\ &\geq \frac{\sum_{i=1}^m \{1 - \pi(s_i)\} c_i(w_i q_{(l)})}{l} \\ &= \frac{\sum_{i=1}^m \{1 - \pi(s_i)\} c_i(w_i q_{(l)})}{\sum_{i=1}^m \mathbb{I}(q_i \leq q_{(l)})} > \alpha. \end{aligned}$$

This proves Equation (11) and concludes the proof. \square

C.2 Proof of Lemma 2

Proof. First of all, by Assumption (A8), we have that $Q(t)$ is well defined for $t \geq \nu$. For any t such that $E\{\sum_{i=1}^m \mathbb{I}(q_i \leq t)\} = 0$, we set $Q(t) = 0$ for simplicity and it will not affect the results. We can rewrite $Q(t)$ as

$$\begin{aligned} Q(t) &= \frac{E\{\sum_{i=1}^m (1 - \theta_i) \delta_i\}}{E(\sum_{i=1}^m \delta_i)} = \frac{E[\sum_{i=1}^m E\{(1 - \theta_i) \delta_i | X_i\}]}{E(\sum_{i=1}^m \delta_i)} \\ &= \frac{E[\sum_{i=1}^m \delta_i E\{(1 - \theta_i) | X_i\}]}{E(\sum_{i=1}^m \delta_i)} = \frac{E\left\{\sum_{i=1}^m \mathbb{I}(q_i \leq t) \frac{q_i}{1+q_i}\right\}}{E\{\sum_{i=1}^m \mathbb{I}(q_i \leq t)\}}. \end{aligned} \tag{12}$$

For the first part of this lemma, note that

$$\begin{aligned} &E\left\{\sum_{i=1}^m \mathbb{I}(q_i \leq t) \frac{q_i}{1+q_i}\right\} - \frac{t}{1+t} E\left\{\sum_{i=1}^m \mathbb{I}(q_i \leq t)\right\} \\ &= E\left\{\sum_{i=1}^m \mathbb{I}(q_i \leq t) \left(\frac{q_i}{1+q_i} - \frac{t}{1+t}\right)\right\} \\ &= E\left\{\sum_{i=1}^m \mathbb{I}(q_i \leq t) \frac{q_i - t}{(1+q_i)(1+t)}\right\} \leq 0. \end{aligned}$$

The equality holds if and only if $P(q_i < t | q_i \leq t) = 0$. Therefore, by Equation (12), we have

$$Q(t) = \frac{E\left\{\sum_{i=1}^m \mathbb{I}(q_i \leq t) \frac{q_i}{1+q_i}\right\}}{E\{\sum_{i=1}^m \mathbb{I}(q_i \leq t)\}} < \frac{t}{1+t}. \tag{13}$$

Denote by $\nu = \frac{\alpha}{1-\alpha}$. By Equation (13), we immediately have $t_L \geq \nu$. Therefore, we only consider $t \geq \nu$ in the following proof.

For the second part, let $\nu \leq t_1 < t_2 < \infty$, $Q(t_1) = \alpha_1$ and $Q(t_2) = \alpha_2$. From the first part, we learn that $\alpha_1 < \frac{t_1}{1+t_1}$. Therefore,

$$\begin{aligned}
Q(t_2) &= \frac{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_2) \frac{q_i}{1+q_i} \right\}}{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_2) \right\}} \\
&= \frac{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_1) \frac{q_i}{1+q_i} \right\}}{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_2) \right\}} + \frac{E \left\{ \sum_{i=1}^m \mathbb{I}(t_1 < q_i \leq t_2) \frac{q_i}{1+q_i} \right\}}{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_2) \right\}} \\
&= \frac{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_1) \frac{q_i}{1+q_i} \right\}}{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_1) \right\}} \frac{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_1) \right\}}{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_2) \right\}} + \\
&\quad \frac{E \left\{ \sum_{i=1}^m \mathbb{I}(t_1 < q_i \leq t_2) \frac{q_i}{1+q_i} \right\}}{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_2) \right\}} \\
&= \alpha_1 \frac{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_1) \right\}}{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_2) \right\}} + \frac{E \left\{ \sum_{i=1}^m \mathbb{I}(t_1 < q_i \leq t_2) \frac{q_i}{1+q_i} \right\}}{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_2) \right\}} \\
&\geq \alpha_1 \frac{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_1) \right\}}{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_2) \right\}} + \frac{t_1}{1+t_1} \frac{E \left\{ \sum_{i=1}^m \mathbb{I}(t_1 < q_i \leq t_2) \right\}}{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_2) \right\}} \\
&> \alpha_1 \frac{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_1) \right\}}{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_2) \right\}} + \alpha_1 \frac{E \left\{ \sum_{i=1}^m \mathbb{I}(t_1 < q_i \leq t_2) \right\}}{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_2) \right\}} \\
&= \alpha_1 = Q(t_1).
\end{aligned}$$

For the third part, note that $Q(t)$ here is continuous and increasing when $m \rightarrow \infty$. We consider two cases: $\lim_{m \rightarrow \infty} \frac{\pi(s_i)}{m} \leq 1 - \alpha$ and $\lim_{m \rightarrow \infty} \frac{\pi(s_i)}{m} > 1 - \alpha$.

The first case is trivial since it implies $\lim_{t \rightarrow \infty} Q(t) \leq \alpha$ and $t_L = \infty$. The procedure rejects all hypotheses and is obviously most powerful. For the second case, we have $\lim_{t \rightarrow \infty} Q(t) = \frac{\sum_{i=1}^m \{1 - \pi(s_i)\}}{m} > \alpha$. Combining this with the fact that $Q(\nu) < \alpha$, we can always find a unique t_L such that $Q(t_L) = \alpha$. Note that, by

$$\begin{aligned}
\lim_{m \rightarrow \infty} \text{mFDR}_{\delta_L} &= \lim_{m \rightarrow \infty} \frac{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_L) \frac{q_i}{1+q_i} \right\}}{E \left\{ \sum_{i=1}^m \mathbb{I}(q_i \leq t_L) \right\}} = \alpha, \\
\text{and } \lim_{m \rightarrow \infty} \text{mFDR}_{\delta'} &= \lim_{m \rightarrow \infty} \frac{E \left\{ \sum_{i=1}^m \delta'_i \frac{q_i}{1+q_i} \right\}}{\{E \sum_{i=1}^m \delta'_i\}} \leq \alpha,
\end{aligned}$$

we have

$$\lim_{m \rightarrow \infty} E \left\{ \sum_{i=1}^m \delta_i^L \left(\frac{q_i}{1+q_i} - \alpha \right) \right\} = 0$$

and $\lim_{m \rightarrow \infty} E \left\{ \sum_{i=1}^m \delta_i' \left(\frac{q_i}{1+q_i} - \alpha \right) \right\} \leq 0,$

which implies

$$\lim_{m \rightarrow \infty} E \left\{ \sum_{i=1}^m (\delta_i^L - \delta_i') \left(\frac{q_i}{1+q_i} - \alpha \right) \right\} \geq 0. \quad (14)$$

Note that, by the law of total expectation as in Equation (12), we have

$$\begin{aligned} & \lim_{m \rightarrow \infty} \{E(\sum_{i=1}^m \delta_i^L \theta_i) - E(\sum_{i=1}^m \delta_i' \theta_i)\} \geq 0 \\ \Leftrightarrow & \lim_{m \rightarrow \infty} E \left\{ \sum_{i=1}^m (\delta_i^L - \delta_i') \frac{1}{1+q_i} \right\} \geq 0. \end{aligned} \quad (15)$$

Hence, it suffices to show $\lim_{m \rightarrow \infty} E \left\{ \sum_{i=1}^m (\delta_i^L - \delta_i') \frac{1}{1+q_i} \right\} \geq 0$. By Equation (14), it suffices to show that there exists some $\lambda \geq 0$ such that $(\delta_i^L - \delta_i') \frac{1}{1+q_i} \geq \lambda (\delta_i^L - \delta_i') \left(\frac{q_i}{1+q_i} - \alpha \right)$ for every i , i.e.,

$$(\delta_i^L - \delta_i') \left\{ \frac{1}{1+q_i} - \lambda \left(\frac{q_i}{1+q_i} - \alpha \right) \right\} \geq 0. \quad (16)$$

By the first part of this lemma, we have $\alpha = Q(t_L) < \frac{t_L}{1+t_L}$ and thus $\frac{1}{t_L - \alpha(1+t_L)} > 0$.

Let $\lambda = \frac{1}{t_L - \alpha(1+t_L)}$, then for each i :

1. If $\delta_i^L = 0$, we have $\delta_i^L - \delta_i' \leq 0$ and $q_i > t_L$. Therefore, $\left\{ \frac{1}{1+q_i} - \lambda \left(\frac{q_i}{1+q_i} - \alpha \right) \right\} < \left\{ \frac{1}{1+t_L} - \lambda \left(\frac{t_L}{1+t_L} - \alpha \right) \right\} = 0$.
2. If $\delta_i^L = 1$, we have $\delta_i^L - \delta_i' \geq 0$ and $q_i \leq t_L$. Therefore, $\left\{ \frac{1}{1+q_i} - \lambda \left(\frac{q_i}{1+q_i} - \alpha \right) \right\} \geq \left\{ \frac{1}{1+t_L} - \lambda \left(\frac{t_L}{1+t_L} - \alpha \right) \right\} = 0$.

This proves Equation (16) and concludes the proof. \square

C.3 Proof of Lemma 3

Proof. For $t \geq 0$, we let

$$\begin{aligned} Q_2(t) &= \frac{\sum_{i=1}^{m_2} \{1 - \pi(s_{2,i})\} \hat{c}_{2,i}(w_{2,i}t)}{E\{\sum_{i=1}^{m_2} \mathbb{I}(\hat{q}_{2,i} \leq t)\}}, \\ \hat{Q}_2(t) &= \frac{\sum_{i=1}^{m_2} \{1 - \hat{\pi}_2(s_{2,i})\} \hat{c}_{2,i}(w_{2,i}t)}{E\{\sum_{i=1}^{m_2} \mathbb{I}(\hat{q}_{2,i} \leq t)\}}, \\ \widehat{\text{FDP}}_2(t) &= \frac{\sum_{i=1}^{m_2} \{1 - \hat{\pi}_2(s_{2,i})\} \hat{c}_{2,i}(w_{2,i}t)}{\{\sum_{i=1}^{m_2} \mathbb{I}(\hat{q}_{2,i} \leq t)\} \vee 1}, \\ t_2^* &= \sup\{t \in [0, \infty) : \widehat{\text{FDP}}(t) \leq \alpha\}. \end{aligned}$$

As $t_2^* \geq \hat{q}_{2,(k)} \geq \nu$ by the first part of Assumption (A9) and Lemma 1, we only consider $t \geq \nu$ in the following proof. The second part of Assumption (A9) implies $E\{\sum_{i=1}^{m_2} \mathbb{I}(\hat{q}_{2,i} \leq t)\} \rightarrow \infty$ when $m \rightarrow \infty$ for $t \geq \nu$, which makes $Q_2(t)$, $\hat{Q}_2(t)$, $\tilde{Q}_2(t)$ well defined when $t \geq \nu$.

Note that, by Assumption (A9) and the standard Chernoff bound for independent Bernoulli random variables, we have uniformly for $t \geq \nu$ and any $\epsilon > 0$

$$\begin{aligned} &P\left(\left|\frac{\sum_{i=1}^{m_2} \mathbb{I}(\hat{q}_{2,i} \leq t)}{E\{\sum_{i=1}^{m_2} \mathbb{I}(\hat{q}_{2,i} \leq t)\}} - 1\right| \geq \epsilon\right) \\ &\leq 2e^{-\epsilon^2 \sum_{i=1}^{m_2} P(\hat{q}_{2,i} \leq t)/3} \\ &\leq 2e^{-\epsilon^2 \sum_{i=1}^{m_2} P(\hat{q}_{2,i} \leq \nu)/3} \rightarrow 0, \end{aligned}$$

which implies

$$\sup_{t \geq \nu} |\hat{Q}_2(t) - \widehat{\text{FDP}}_2(t)| \xrightarrow{P} 0 \quad (17)$$

as $m_2 \rightarrow \infty$. On the other hand, we have uniformly for $t \geq \nu$,

$$\begin{aligned} |\hat{Q}_2(t) - \tilde{Q}_2(t)| &= \left| \frac{\sum_{i=1}^{m_2} \{\tilde{\pi}_2(s_{2,i}) - \hat{\pi}_2(s_{2,i})\} \hat{c}_{2,i}(w_{2,i}t)}{E\{\sum_{i=1}^{m_2} \mathbb{I}(\hat{q}_{2,i} \leq t)\}} \right| \\ &\leq \frac{|\sum_{i=1}^{m_2} \{\tilde{\pi}_2(s_{2,i}) - \hat{\pi}_2(s_{2,i})\}|}{E\{\sum_{i=1}^{m_2} \mathbb{I}(\hat{q}_{2,i} \leq \nu)\}} \\ &= \frac{m_2 \times o_P(1)}{m_2} = o_P(1), \end{aligned}$$

where the first $o_P(1)$ is with regard to $m_1 \rightarrow \infty$ by the uniformly conservative consistency

of $\hat{\pi}_2(\cdot)$, and the term m_2 in the denominator comes from the first part of Assumption (A9). As the data splitting strategy ensures $m_1 \approx m_2$, we obtain the second $o_P(1)$ with regard to $m \rightarrow \infty$. Thus, we have

$$\sup_{t \geq \nu} |\hat{Q}_2(t) - \tilde{Q}_2(t)| \rightarrow 0 \quad (18)$$

in probability as $m \rightarrow \infty$. We note that Equation (18) holds in a similar manner for Theorems 3 and 9, where their conservative consistency is defined in terms of convergence in probability.

Combining Equations (17) and (18), we have

$$\sup_{t \geq \nu} |\widehat{\text{FDP}}_2(t) - \tilde{Q}_2(t)| \rightarrow 0$$

in probability as $m \rightarrow \infty$. Then, following the proof of Theorem 5, we can similarly obtain $t_2^* \rightarrow t_{2,L}$ in probability by Assumptions (A10) and (A11). Finally, we have

$$\begin{aligned} \text{mFDR}_{\text{Half-procedure}} &= \frac{\sum_{i=1}^{m_2} P(\hat{q}_{2,i} \leq t_2^*, \theta_i = 0)}{E\{\sum_{i=1}^{m_2} \mathbb{I}(\hat{q}_{2,i} \leq t_2^*)\}} \\ &= \frac{\sum_{i=1}^{m_2} P(\hat{q}_{2,i} \leq t_{2,L}, \theta_i = 0)}{E\{\sum_{i=1}^{m_2} \mathbb{I}(\hat{q}_{2,i} \leq t_{2,L})\}} + o(1) \\ &= Q_2(t_{2,L}) + o(1) \leq \tilde{Q}_2(t_{2,L}) + o(1) \\ &\leq \alpha + o(1). \end{aligned}$$

□

References

- Aubert, J., Bar-Hen, A., Daudin, J.-J., and Robin, S. (2004). Determination of the differentially expressed genes in microarray experiments using local fdr. *BMC bioinformatics*, 5(1):1–9.
- Basu, P., Cai, T. T., Das, K., and Sun, W. (2018). Weighted false discovery rate control in large-scale multiple testing. *J. Am. Statist. Assoc.*, 113(523):1172–1183.
- Bates, S., Candès, E., Lei, L., Romano, Y., and Sesia, M. (2023). Testing for outliers with conformal p-values. *The Annals of Statistics*, 51(1):149–178.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B*, 57(1):289–300.
- Benjamini, Y. and Hochberg, Y. (1997). Multiple hypotheses testing with weights. *Scand. J. Stat.*, 24(3):407–418.
- Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.*, 29(4):1165–1188.
- Cai, T. T., Sun, W., and Wang, W. (2019). Covariate-assisted ranking and screening for large-scale two-sample inference. *J. R. Statist. Soc. B*, 81(2):187–234.
- Cai, T. T., Sun, W., and Xia, Y. (2022). Laws: A locally adaptive weighting and screening approach to spatial multiple testing. *J. Am. Statist. Assoc.*, 117(539):1370–1383.
- Cao, H., Chen, J., and Zhang, X. (2022). Optimal false discovery rate control for large scale multiple testing with auxiliary information. *Ann. Statist.*, 50(2):807–857.
- Chen, X. (2019). Uniformly consistently estimating the proportion of false null hypotheses via lebesgue–stieltjes integral equations. *J. Multivar. Anal.*, 173:724–744.
- Chen, X. and Liu, W. (2018). Testing independence with high-dimensional correlated samples. *Ann. Statist.*, 46(2):866–894.

- Dobriban, E., Fortney, K., Kim, S. K., and Owen, A. B. (2015). Optimal multiple testing under a gaussian prior on the effect sizes. *Biometrika*, 102(4):753–766.
- Du, L. and Zhang, C. (2014). Single-index modulated multiple testing. *Ann. Statist.*, 42(3):1262–1311.
- Efron, B. (2003). Robbins, empirical bayes and microarrays. *Ann. Statist.*, 31(2):366–378.
- Efron, B. (2004). Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *J. Am. Statist. Assoc.*, 99(465):96–104.
- Efron, B. and Tibshirani, R. (2007). On testing the significance of sets of genes. *Ann. Appl. Stat.*, 1(1):107–129.
- Efron, B., Tibshirani, R., Storey, J. D., and Tusher, V. (2001). Empirical bayes analysis of a microarray experiment. *J. Am. Statist. Assoc.*, 96(456):1151–1160.
- Fu, L., Gang, B., James, G. M., and Sun, W. (2022). Heteroscedasticity-adjusted ranking and thresholding for large-scale multiple testing. *J. Am. Statist. Assoc.*, 117(538):1028–1040.
- Genovese, C. and Wasserman, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *J. R. Statist. Soc. B*, 64(3):499–517.
- Genovese, C. R., Roeder, K., and Wasserman, L. (2006). False discovery control with p-value weighting. *Biometrika*, 93(3):509–524.
- Heller, R. and Rosset, S. (2021). Optimal control of false discovery criteria in the two-group model. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 83(1):133–155.
- Hong, W.-J., Tibshirani, R., and Chu, G. (2009). Local false discovery rate facilitates comparison of different microarray experiments. *Nucleic Acids Res.*, 37(22):7483–7497.
- Ignatiadis, N. and Huber, W. (2021). Covariate powered cross-weighted multiple testing. *J. R. Statist. Soc. B*, 83(4):720–751.

- Jin, J. and Cai, T. T. (2007). Estimating the null and the proportion of nonnull effects in large-scale multiple comparisons. *J. Am. Statist. Assoc.*, 102(478):495–506.
- Lei, L. and Fithian, W. (2018). Adapt: an interactive procedure for multiple testing with side information. *J. R. Statist. Soc. B*, 80(4):649–679.
- Leung, D. and Sun, W. (2022). Zap: z-value adaptive procedures for false discovery rate control with side information. *J. R. Statist. Soc. B*, 84(5):1886–1946.
- Li, A. and Barber, R. F. (2019). Multiple testing with the structure-adaptive benjamini–hochberg algorithm. *J. R. Statist. Soc. B*, 81(1):45–74.
- Liang, Z., Cai, T. T., Sun, W., and Xia, Y. (2023). Locally adaptive algorithms for multiple testing with network structure, with application to genome-wide association studies. *arXiv preprint arXiv:2203.11461*.
- Liu, W. (2013). Gaussian graphical model estimation with false discovery rate control. *Ann. Statist.*, 41(6):2948–2978.
- Ma, L., Xia, Y., and Li, L. (2023). NAPA: neighborhood-assisted and posterior-adjusted two-sample inference. *Stat. Sin.*, (just-accepted):1–42.
- Marandon, A., Lei, L., Mary, D., and Roquain, E. (2024). Adaptive novelty detection with false discovery rate guarantee. *The Annals of Statistics*, 52(1):157–183.
- McDonald, D., Hyde, E., Debelius, J. W., Morton, J. T., Gonzalez, A., Ackermann, G., Aksenenov, A. A., Behsaz, B., Brennan, C., Chen, Y., et al. (2018). American gut: an open platform for citizen science microbiome research. *Msystems*, 3(3):10–1128.
- Meinshausen, N. and Rice, J. (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Statist.*, 34(1):373–393.
- Paloyelis, Y., Mehta, M. A., Kuntsi, J., and Asherson, P. (2007). Functional mri in adhd: a systematic literature review. *Expert review of neurotherapeutics*, 7(10):1337–1356.

- Ramdas, A. K., Barber, R. F., Wainwright, M. J., and Jordan, M. I. (2019). A unified treatment of multiple testing with prior knowledge using the p-filter. *Ann. Statist.*, 47(5):2790–2821.
- Roquain, E. and Van De Wiel, M. A. (2009). Optimal weighting for false discovery rate control. *Electron. J. Stat.*, 3:678–711.
- Sarkar, S. K. and Zhao, Z. (2022). Local false discovery rate based methods for multiple testing of one-way classified hypotheses. *Electron. J. Stat.*, 16:6043–6085.
- Storey, J. D. (2002). A direct approach to false discovery rates. *J. R. Statist. Soc. B*, 64(3):479–498.
- Storey, J. D., Dai, J. Y., and Leek, J. T. (2007). The optimal discovery procedure for large-scale significance testing, with applications to comparative microarray experiments. *Biostatistics*, 8(2):414–432.
- Sun, W. and Cai, T. T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *J. Am. Statist. Assoc.*, 102(479):901–912.
- Vovk, V. and Wang, R. (2021). E-values: Calibration, combination and applications. *Ann. Statist.*, 49(3):1736–1754.
- Wang, R. and Ramdas, A. (2022). False discovery rate control with e-values. *J. R. Statist. Soc. B*, 84(3):822–852.
- Xia, Y. (2020). Correlation and association analyses in microbiome study integrating multiomics in health and disease. *Progress in molecular biology and translational science*, 171:309–491.
- Xie, J., Cai, T. T., Maris, J., and Li, H. (2011). Optimal false discovery rate control for dependent data. *Stat. Interface*, 4(4):417.
- Zhang, X. and Chen, J. (2022). Covariate adaptive false discovery rate control with applications to omics-wide multiple testing. *J. Am. Statist. Assoc.*, 117(537):411–427.