Author/s:
Leung, D;Tran, N

Title:
Adaptive procedures for directional false discovery rate control

Date:
2024-01-01

Citation:
Leung, D. & Tran, N. (2024). Adaptive procedures for directional false discovery rate control. Electronic Journal of Statistics, 18 (1), pp.706-741. https://doi.org/10.1214/24-EJS2213.

Persistent Link:
https://hdl.handle.net/11343/352797

Electronic Journal of Statistics
Vol. 18 (2024) 706–741
ISSN: 1935-7524
https://doi.org/10.1214/24-EJS2213

Adaptive procedures for directional false discovery rate control

Dennis Leung and Ninh Tran

School of Mathematics and Statistics
University of Melbourne, Victoria, Australia
e-mail: dennis.leung@unimelb.edu.au; ninht@student.unimelb.edu.au

Abstract: In multiple hypothesis testing, it is well known that adaptive procedures can enhance power via incorporating information about the number of true nulls present. Under independence, we establish that two adaptive false discovery rate (FDR) methods, upon augmenting sign declarations, also offer *directional* false discovery rate (FDR_{dir}) control in the strong sense. Such FDR_{dir} controlling properties are appealing, because adaptive procedures have the greatest potential to reap substantial gain in power when the underlying parameter configurations contain little to no true nulls, which are precisely settings where the FDR_{dir} is an arguably more meaningful error rate to be controlled than the FDR.

MSC2020 subject classifications: Primary 62F03; secondary 62C12.
Keywords and phrases: Multiple testing, adaptive procedures, false discovery rate, directional inference.

Received April 2023.

1. Introduction

Consider independent observations $z_1, \dots, z_m \in \mathbb{R}$, where each "$z$-value" z_i is a noisy measurement of an effect parameter $\theta_i \in \mathbb{R}$. We suppose m is quite large, and will use the notational shorthand $[m] \equiv \{1, \dots, m\}$ in the sequel. For testing the multiple point null hypotheses

$$H_i : \theta_i = 0, \quad i \in [m], \tag{1.1}$$

Benjamini and Hochberg (1995) proposed their now-celebrated BH procedure to control the *false discovery rate* (FDR),

$$\mathbb{E}\left[\frac{\sum_{i=1}^m \mathbf{1}(H_i \text{ is rejected and } \theta_i = 0)}{1 \vee \sum_{i=1}^m \mathbf{1}(H_i \text{ is rejected})}\right],$$

below a target level $q \in (0, 1)$. However, many statisticians, such as Tukey (1962, 1991) and Gelman and Tuerlinckx (2000), consider testing the point nulls in (1.1) futile, because the effects in reality, however small, are rarely exactly zero. Instead, they argue that one should test the direction/sign of the effect by declaring either $\theta_i > 0$ or $\theta_i < 0$ as a discovery, or making no declaration about θ_i at all if there is insufficient evidence to support either direction. Under this new paradigm, a generic discovery procedure consists of two components:

(i) $\mathcal{R} \subseteq \{1, \dots, m\}$, the set of rejected indices for which sign declarations (discoveries) are made, and
(ii) $\widehat{sgn}_i$, the positive or negative sign declared for each $i \in \mathcal{R}$. (Note that $\widehat{sgn}_i \neq 0$.)

We denote such a procedure, or its associated decisions, by $(\widehat{sgn}_i)_{i\in\mathcal{R}}$. Its error rate analogous to the FDR is the *directional* false discovery rate ($\mathrm{FDR}_{\mathrm{dir}}$), defined as

$$\mathrm{FDR}_{\mathrm{dir}}\left[(\widehat{sgn}_i)_{i\in\mathcal{R}}\right] \equiv \mathbb{E}\left[\frac{\sum_{i\in\mathcal{R}} \mathbf{1}(\mathrm{sgn}(\theta_i) \neq \widehat{sgn}_i)|}{1 \vee |\mathcal{R}|}\right], \tag{1.2}$$

and its power can be measured by the expected number of true discoveries (ETD)

$$\mathrm{ETD}\left[(\widehat{sgn}_i)_{i\in\mathcal{R}}\right] \equiv \mathbb{E}\left[\sum_{i\in\mathcal{R}} \mathbf{1}(\mathrm{sgn}(\theta_i) = \widehat{sgn}_i)\right],$$

where for any $x \in \mathbb{R}$, $\mathrm{sgn}(x) \equiv \mathbf{1}(x > 0) - \mathbf{1}(x < 0)$. This paper treats the control of $\mathrm{FDR}_{\mathrm{dir}}$ for the sign discoveries of the θ_i's; we don't exclude the possibility that some effect parameters can indeed be zero, so a false discovery amounts to declaring $\theta_i < 0$ when the truth is $\theta_i \geqslant 0$, or vice versa.[1]

Methods for controlling the $\mathrm{FDR}_{\mathrm{dir}}$ are surprisingly scant for the simple testing problem above. To our best knowledge, under the independence among $z_1, \dots, z_m$ and some standard assumptions (Assumption 1 and 2 below), the only known procedure that can provably control the $\mathrm{FDR}_{\mathrm{dir}}$ under a target level $q \in (0, 1)$ in the strong sense, i.e. irrespective of the configuration of θ_i's, is what we call the *directional* Benjamini and Hochberg ($\mathrm{BH}_{\mathrm{dir}}$) procedure proposed in Benjamini and Yekutieli (2005, Definition 6). The $\mathrm{BH}_{\mathrm{dir}}$ procedure first decides on the set of rejected indices among $[m]$ by applying the standard BH procedure at level q to the two-sided p-values constructed from $z_1, \dots, z_m$, and then declares the sign of each rejected θ_i as $\mathrm{sgn}(z_i)$. Guo and Romano (2015, Procedure 6) proposed another procedure almost identical to the $\mathrm{BH}_{\mathrm{dir}}$, except that the screening BH step is applied at level $2q$ instead of q. This latter procedure, however, can only control the $\mathrm{FDR}_{\mathrm{dir}}$ under its intended target level q when all $\theta_1, \dots, \theta_m$ are nonzero. Guo and Romano (2015, Procedures 7-9), Zhao and Fung (2018) and Guo, Sarkar and Peddada (2010) consider extensions of the current problem that involve either specific patterns of dependence or multidimensional directional decisions.

We will expand the repertoire of available methods for $\mathrm{FDR}_{\mathrm{dir}}$ inference with strong theoretical guarantee. In the FDR literature, it is known that *adaptive* methods (Benjamini and Hochberg, 2000, Benjamini, Krieger and Yekutieli, 2006) that incorporate a data-driven estimate of the proportion of true nulls

$$\pi \equiv \frac{|\{i : \theta_i = 0\}|}{m} \tag{1.3}$$

[1]Some works call (1.2) the *mixed* directional false discovery rate, to emphasize the two types of errors involved: those from declaring any sign at all for θ_i when $\theta_i = 0$ and those from declaring an opposite sign for θ_i when $\theta_i \neq 0$.

into their procedures have the potential to improve upon the power offered by the vanilla BH procedure. Using martingale arguments, we prove that under independence, two adaptive methods can also provide strong FDR_{dir} control upon the augmentation of sign declarations. The first is Storey, Taylor and Siegmund (2004)'s adaptive FDR procedure, which can be seen as an adaptive variant of the BH procedure (Section 2). The second is a specific procedure belonging to a more recent line of methods driven by a technique called "data masking" first introduced in Lei and Fithian (2018) (Section 3). We also numerically demonstrate their competitive power performances in Section 4.

Adaptive procedures that can offer FDR_{dir} guarantees are particularly important for settings whose underlying parameter configurations contain little to no true nulls. Arguably, if most θ_i's in question are non-zero, the FDR_{dir} is more meaningful as an error measure compared to the FDR because querying about their signs matters more. Moreover, such "non-sparse-signal" settings are precisely those in which adaptive procedures can reap substantial gain in power; see Storey, Taylor and Siegmund (2004, Section 3.1), where their adaptive FDR procedure demonstrates greater improvements in power over the BH procedure as π decreases.

1.1. Notation and assumptions

For $a, b \in \mathbb{R}$, we let $a \wedge b \equiv \min(a, b)$ and $a \vee b \equiv \max(a, b)$. For any two subsets $\mathcal{A}, \mathcal{B} \subset [m]$, $\mathcal{A} \subsetneq (\supsetneq)\mathcal{B}$ means $\mathcal{A}$ is a strict subset (superset) of $\mathcal{B}$. $U(\cdot; a, b)$ denotes a uniform density on the interval $[a, b] \subseteq \mathbb{R}$. $\mathbb{E}_{\boldsymbol{\theta}}[\cdot]$ means a (frequentist) expectation with respect to fixed values of $\theta_1, \ldots, \theta_m$. For each $i \in [m]$ and given $\theta_i = \theta$, $F_{i,\theta}(\cdot)$ denotes the distribution function of z_i with density $f_{i,\theta}(\cdot) \equiv F'_{i,\theta}(\cdot) > 0$ with respect to the Lebesgue measure on $\mathbb{R}$ (so $F_{i,\theta}(\cdot)$ is implicitly assumed to be smooth and strictly increasing). $\Phi(\cdot)$ and $\phi(\cdot)$ denote the standard normal distribution and density functions, respectively. Additionally, the following assumptions will be made for the two main theoretical results in this paper (Theorems 2.1 and 3.1):

Assumption 1. *The null distribution of z_i is known and symmetric around zero, i.e. $F_{i,0}(-z) = 1 - F_{i,0}(z)$ and $f_{i,0}(-z) = f_{i,0}(z)$ for any $z \in \mathbb{R}$.*

Assumption 2. *The family of densities $\{f_{i,\theta}(\cdot)\}_{\theta \in \mathbb{R}}$ satisfies the monotone likelihood ratio (MLR) property, i.e. for any given $\theta < \theta^*$ and $z < z^*$, $\frac{f_{i,\theta^*}(z)}{f_{i,\theta}(z)} \leqslant \frac{f_{i,\theta^*}(z^*)}{f_{i,\theta}(z^*)}$.*

These are not necessarily the weakest assumptions for our theorems to hold, but are standard enough so as not to distract from the key ideas of the proofs. Essentially, Assumption 2 guarantees that z_i becomes "stochastically larger" as θ_i increases, and two examples satisfying Assumption 2 are the normal distributions $N(\theta, \sigma_i^2)$ for a fixed variance σ_i^2 and the noncentral t-distributions $NCT(\theta, v_i)$ for a fixed degree v_i (Kruskal, 1954, Section 3). Moreover, the symmetry condition on $F_{i,0}$ in Assumption 1 is not crucial; it is included primarily to streamline our presentation. If this condition isn't satisfied, instead of z_i, one can alternatively consider the transformed statistic $\Phi^{-1}(F_{i,0}(z_i))$ whose density has the form

$$\frac{f_{i,\theta}\left(F_{i,0}^{-1}\left(\Phi(z)\right)\right)}{f_{i,0}\left(F_{i,0}^{-1}\left(\Phi(z)\right)\right)}\phi(z). \tag{1.4}$$

As a function in z, (1.4) boils down to the symmetric density function $\phi(z)$ when $\theta_i = 0$. The density (1.4) also maintains the MLR property, provided that the base density $f_{i,\theta}(\cdot)$ satisfies the MLR property.

2. Directional control with Storey, Taylor and Siegmund (2004)'s adaptive procedure

Compute the two-sided p-values

$$p_i \equiv 2F_{i,0}(-|z_i|), i \in [m] \tag{2.1}$$

from the null distributions $F_{i,0}$, and for any $t \in [0,1]$, let $\mathcal{R}(t) \equiv \{i : p_i \leqslant t\}$ be the set of rejected indices i defined by $p_i \leqslant t$; Algorithm 1 is a sign-augmented version of Storey, Taylor and Siegmund (2004)'s adaptive procedure for FDR_{dir} control, which we call the "STS_{dir}" for short.

Algorithm 1: The STS_{dir} procedure at target FDR_{dir} level $q \in (0,1)$

Data: $z_1, \dots, z_m$
Input: FDR_{dir} target $q \in (0,1)$, the two-side p-values $\{p_i\}_{i=1}^m$ from (2.1), and $\{\text{sgn}(z_i)\}_{i=1}^m$;

1 For a fixed tuning parameter $\lambda \in (0,1)$, compute $\hat{\pi}(\lambda) \equiv \frac{|\{i : p_i > \lambda\}|+1}{(1-\lambda)m}$ as an estimate for π;

2 Compute $t_q^\lambda \equiv \sup\left\{t \in [0,1] : \widehat{\text{FDR}}_\lambda(t) \leqslant q\right\}$, where

$$\widehat{\text{FDR}}_\lambda(t) \equiv \begin{cases} \frac{\hat{\pi}(\lambda)mt}{|\mathcal{R}(t)| \vee 1}, & \text{if } t \leqslant \lambda \\ 1, & \text{if } t > \lambda \end{cases};$$

3 Compute the rejection set $\mathcal{R}(t_q^\lambda) \equiv \{i : p_i \leqslant t_q^\lambda\}$;
Output: Sign discoveries $(\text{sgn}(z_i))_{i \in \mathcal{R}(t_q^\lambda)}$.

$\mathcal{R}(t_q^\lambda)$ is precisely the rejection set produced by the adaptive procedure in Storey, Taylor and Siegmund (2004, Theorem 3) that was proved to offer strong FDR control for testing the point nulls in (1.1), under the assumptions that the null p-values are independent and uniformly distributed. Roughly speaking, compared to the BH procedure, which essentially uses

$$\frac{mt}{|\mathcal{R}(t)| \vee 1} \tag{2.2}$$

as an estimate for the FDR incurred by rejecting any p_i below a given threshold t, Storey, Taylor and Siegmund (2004)'s FDR estimate $\widehat{\text{FDR}}_\lambda(t)$ adjusts (2.2) by the factor $\hat{\pi}(\lambda)$. This factor serves as a conservative estimate for the unknown null proportion $\pi \in [0,1]$ in (1.3). If π is close to zero, i.e. there are very few true nulls in the problem, Storey, Taylor and Siegmund (2004)'s procedure could produce substantially more rejections compared to the BH procedure. Our first result is that, by simply augmenting the rejections with sign declarations as in the output of Algorithm 1, the procedure also offers strong $\mathrm{FDR}_{\mathrm{dir}}$ control under independence; we note that by the construction of the p-values in (2.1) and Assumption 1, it must be the case that $\mathrm{sgn}(z_i) \neq 0$ for any $i \in \mathcal{R}(t_q^\lambda)$:

Theorem 2.1 ($\mathrm{FDR}_{\mathrm{dir}}$ control of $\mathrm{STS}_{\mathrm{dir}}$)**.** *Under Assumptions 1 and 2, as well as the independence among $z_1, \ldots, z_m$, Algorithm 1 controls the FDR_{dir} at level $q \in (0,1)$, i.e. by letting $\mathcal{S}(t) \equiv \{i : sgn(\theta_i) \neq sgn(z_i) \text{ and } i \in \mathcal{R}(t)\}$.*

$$FDR_{dir}\left[(sgn(z_i))_{i\in\mathcal{R}(t_q^\lambda)}\right] \equiv \mathbb{E}_{\boldsymbol{\theta}}\left[\frac{|\mathcal{S}(t_q^\lambda)|}{|\mathcal{R}(t_q^\lambda)| \vee 1}\right] \leqslant \Big(1 - \mathbb{E}_{\boldsymbol{\theta}}[\lambda^{|\mathcal{S}(1)|}]\Big)q \leqslant q.$$

The proof of Theorem 2.1 in Appendix A.1 relies on the optional stopping time theorem for supermartingales, which extends the original arguments by Storey, Taylor and Siegmund (2004, Theorem 3) based on martingales. The choice of λ entails a bias-variance trade-off in the estimation of π, and Storey, Taylor and Siegmund (2004) fixes $\lambda = 0.5$ in their simulations. Alternatively, similarly to Storey, Taylor and Siegmund (2004, Section 6), we provided a method in Appendix A.2 for automatically selecting λ based on the observed data $z_1, \ldots, z_m$ to achieve a balance between bias and variance. Since Theorem 2.1 is only valid on the premise of a fixed λ, using the provided method of automatically selecting λ may not guarantee $\mathrm{FDR}_{\mathrm{dir}}$ control. Regardless, our simulations in Section 4 demonstrate that this data-driven choice of λ typically leads to empirically robust control of $\mathrm{FDR}_{\mathrm{dir}}$.

3. Directional control with data masking: ZDIRECT

Early FDR methods, such as Storey, Taylor and Siegmund (2004)'s adaptive procedure covered in Section 2, process data in a pre-determined manner to decide on the set of rejected hypotheses. In contrast to the conventional belief that the design of a testing procedure should not be affected by the observed data's patterns to avoid data snooping, Lei and Fithian (2018) recently showed that, as long as the data are initially "masked" as a trade-off, one can iteratively interact with the gradually revealed data in a legitimate way to devise valid FDR procedures. The flexibility of this masking technique not only allows one to adapt to information about the null proportion in (1.3) implied by the data, but also side information provided by suitable external covariates that are present; see (Chao and Fithian, 2021, Lei, Ramdas and Fithian, 2021, Leung and Sun, 2022, Tian, Liang and Li, 2021, Yurko et al., 2020) for a series of follow-up works, including the ZAP (finite) algorithm proposed by one of the present authors (Leung and Sun, 2022, Algorithm 2).

Motivated by the construct of ZAP, we propose a masking algorithm that is augmented with sign declarations for the rejected θ_i's, and is proven to provide strong FDR_{dir} control. To facilitate our presentation, for each $i \in [m]$, we first define

$$u_i \equiv F_{i,0}(z_i), \tag{3.1}$$

the probability integral transform of z_i under the null that takes values in the unit interval $(0, 1)$, as well as its *reflection*

$$\breve{u}_i \equiv (0.5 - u_i)I(u_i \leqslant 0.5) + (1.5 - u_i)I(u_i > 0.5) \tag{3.2}$$

about the midpoint 0.25 of the left sub-interval $(0, 0.5)$, or about the midpoint at 0.75 of the right sub-interval $(0.5, 1)$, depending on whether $u_i \leqslant 0.5$ or $u > 0.5$. Moreover, we let

$$u_i' \equiv \begin{cases} u_i \wedge \breve{u}_i & \text{if } u_i, \breve{u}_i \in [0, 0.5] \\ u_i \vee \breve{u}_i & \text{if } u_i, \breve{u}_i \in (0.5, 1) \end{cases}, \tag{3.3}$$

which, for any given i, is the value between u_i and $\breve{u}_i$ closer to the endpoints of the unit interval. In particular, u_i' only provides partial knowledge about the value of u_i: If $u_i' \leqslant 0.5$, u_i may be either u_i' or $0.5 - u_i'$; if $u_i' > 0.5$, u_i may be either u_i' or $1.5 - u_i'$.

With these preparations, we are now in a position to describe our iterative testing procedure for the directional inference of $\theta_1, \dots, \theta_m$ with a target FDR_{dir} level $q \in (0, 1)$, whose steps are listed out in Algorithm 2; we call it "ZDIRECT" as it interacts with the z-values via their one-to-one transformations u_i's for FDR_{dir} control. Essentially, the algorithm iterates through steps $t = 0, 1, \dots$ to construct a strictly decreasing sequence of subsets

$$[m] = \mathcal{M}_0 \supsetneq \mathcal{M}_1 \supsetneq \mathcal{M}_2 \cdots$$

based on the data, and from each $\mathcal{M}_t$, it forms the candidate "acceptance" and "rejection" sets

$$\mathcal{A}_t \equiv \{i : i \in \mathcal{M}_t \text{ and } u_i \in (0.25, 0.75)\} \text{ and}$$
$$\mathcal{R}_t \equiv \{i : i \in \mathcal{M}_t \text{ and } u_i \in (0, 0.25] \cup [0.75, 1)\};$$

the algorithm terminates at step $\hat{t} = \min\{t : \widehat{\text{FDR}}_{\text{dir}}(t) \leqslant q\}$ as soon as the FDR_{dir} estimate

$$\widehat{\text{FDR}}_{\text{dir}}(t) \equiv \frac{1 + |\mathcal{A}_t|}{|\mathcal{R}_t| \vee 1}$$

falls below q, and declares the sign discoveries $(\text{sgn}(z_i))_{i \in \mathcal{R}_{\hat{t}}}$. Since $\mathcal{R}_{\hat{t}} \subset \{i : u_i \in (0, 0.25] \cup [0.75, 1)\}$, it must be the case that $\text{sgn}(z_i) \neq 0$ for any $i \in \mathcal{R}_{\hat{t}}$, under Assumption 1. Moreover, the sets $\mathcal{M}_t$ are shrunk in such a way that these two conditions must be respected (line 6 in Algorithm 2):

(C1) $\mathcal{M}_{t+1}$ must be constructed based only on the partial data $\{\tilde{u}_{i,t}\}_{i\in[m]}$ available at step t, where for each $i \in [m]$, we define

$$\tilde{u}_{i,t} \equiv \begin{cases} u_i & \text{if } i \notin \mathcal{M}_t \\ u_i' & \text{if } i \in \mathcal{M}_t \end{cases}, \tag{3.4}$$

which may reveal the true value of u_i depending on whether $u_i \in \mathcal{M}_t$. Essentially, any human/computer routine who shrinks $\mathcal{M}_t$ to $\mathcal{M}_{t+1}$ can only know that the true value of u_i is one of two possibilities if i is still in the "masked" set $\mathcal{M}_t$.

(C2) $\mathcal{M}_{t+1}$ must be a strict subset of $\mathcal{M}_t$ to ensure the algorithm does terminate.

Algorithm 2: The ZDIRECT procedure at target FDR_{dir} level $q \in (0,1)$

Data: $z_1, \dots, z_m$
Input: FDR_{dir} target $q \in (0,1)$, the initial set $\mathcal{M}_0 = [m]$;
for t= 0,1 . . . , **do**
  Find the candidate "acceptance set" $\mathcal{A}_t \equiv \{i : i \in \mathcal{M}_t \text{ and } u_i \in (0.25, 0.75)\}$;
  Find the candidate "rejection set" $\mathcal{R}_t \equiv \{i : i \in \mathcal{M}_t \text{ and } u_i \in (0, 0.25] \cup [0.75, 1)\}$;
  Compute $\widehat{\text{FDR}}_{\text{dir}}(t) \equiv \frac{1+|\mathcal{A}_t|}{|\mathcal{R}_t| \vee 1}$;
  if $\widehat{FDR}_{dir}(t) > q$ **then**
    construct $\mathcal{M}_{t+1} \subsetneq \mathcal{M}_t$ using only the partially masked data $\{\tilde{u}_{i,t}\}_{i\in[m]}$ from (3.4);
  else
    Set $\hat{t} = t$; break;
  end
end
Output: Sign discoveries $(\text{sgn}(z_i))_{i\in\mathcal{R}_{\hat{t}}}$.

Theorem 3.1, which is proven in Appendix B.2, states that ZDIRECT provides strong FDR_{dir} control under the assumptions in this paper:

Theorem 3.1 (FDR_{dir} control of ZDIRECT)**.** *Under Assumptions 1 and 2, as well as the independence between $z_1, \dots, z_m$, Algorithm 2 controls the FDR_{dir} at level $q \in (0,1)$. Specifically, if the algorithm terminates at step $\hat{t} = \min\{t : \widehat{FDR}_{dir}(t) \leqslant q\}$, we have*

$$\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{|\{i : sgn(\theta_i) \neq sgn(z_i) \ and \ i \in \mathcal{R}_{\hat{t}}\}|}{1 \vee |\mathcal{R}_{\hat{t}}|}\right] \leqslant q. \tag{3.5}$$

Apart from the final sign declarations, Algorithm 2 inherits much of its structure from the ZAP (finite) algorithm in Leung and Sun (2022), but is situated in the more general "without-threshold" framework (Lei and Fithian, 2018, Section 6.1) that does not explicitly involve any thresholding functions. In fact, a quantity like $\widehat{\text{FDR}}_{\text{dir}}(t)$ has been used as an FDR estimate in Leung and Sun (2022). To also make sense of it as a suitable FDR_{dir} estimate for the sign discoveries of $(\mathrm{sgn}(z_i))_{i\in\mathcal{R}_t}$, note that an $i \in \mathcal{R}_t$ constitutes a false discovery if either

$$u_i \in (0, 0.25] \text{ and } \mathrm{sgn}(\theta_i) \geqslant 0, \tag{3.6}$$

or

$$u_i \in [0.75, 1) \text{ and } \mathrm{sgn}(\theta_i) \leqslant 0. \tag{3.7}$$

By the continuity of $F_{i,0}$, each u_i is uniformly distributed when $\theta_i = 0$. Suppose we are in an error-prone scenario where all θ_i's are either exactly or very close to zero, so that all u_i's stochastically behave much like uniform random variables. For a given $i \in \mathcal{M}_t$, the event $\{u_i \in (0, 0.25]\}$ and $\{u_i \in (0.25, 0.5]\}$ should be approximately equally likely, so the set size $|\{i : i \in \mathcal{M}_t \text{ and } u_i \in (0.25, 0.5]\}|$ is a reasonable estimate of the number of false discoveries by way of (3.6). Analogously, $|\{i : i \in \mathcal{M}_t \text{ and } u_i \in [0.5, 0.75)\}|$ serves as an estimate of the number of false discoveries by way of (3.7). As such, $|\mathcal{A}_t|$ makes sense as an estimate of the number of false discoveries in $\mathcal{R}_t$, where the additive "1" in the numerator of $\widehat{\mathrm{FDR}}_{\mathrm{dir}}(t)$ is a theoretical adjustment factor to make it conservative enough. In fact, this concept is akin to how the FDR estimate for the *knockoff filter* for variable selection in linear regressions (Barber et al., 2015) can also serve as an $\mathrm{FDR}_{\mathrm{dir}}$ estimate when sign declarations are augmented (Barber et al., 2019).

However, our specific methodology for updating $\mathcal{M}_t$, as described next, considerably differs from existing data masking algorithms. Notably, we rely solely on $z_1, \dots, z_m$ as the available data for our problem, without harnessing external covariate information. This poses a greater challenge for boosting power, but our simulations in Section 4 demonstrate that ZDIRECT remains competitive in terms of power for $\mathrm{FDR}_{\mathrm{dir}}$ control when compared to other existing methods.

3.1. Shrinking the masked sets $\mathcal{M}_t$

To achieve power, we aim to shrink $\mathcal{M}_t$ in accordance with conditions (C1)–(C2) in such a way that Algorithm 2 can mimic the *optimal discovery procedure* (ODP) under a Bayesian formulation[2] of the problem, which imposes the additional assumption that the effects are random and independently generated by a common prior distribution, i.e.

$$\theta_i \sim_{i.i.d.} G(\theta), \quad i \in [m], \tag{3.8}$$

for an unknown distribution function $G(\cdot)$. For a given target level $q \in (0, 1)$, the ODP, denoted by $(\widehat{sgn}_i^{ODP})_{i\in\mathcal{R}^{ODP}}$, is defined to be the procedure with the properties that

$$\mathrm{FDR}_{\mathrm{dir}}\Big[(\widehat{sgn}_i^{ODP})_{i\in\mathcal{R}^{ODP}}\Big] \leqslant q$$

and

$$\mathrm{ETD}\Big[(\widehat{sgn}_i^{ODP})_{i\in\mathcal{R}^{ODP}}\Big] \geqslant \mathrm{ETD}\Big[(\widehat{sgn}_i)_{i\in\mathcal{R}}\Big]$$

[2]More precisely, it is an *empirical Bayes* formulation, because $G(\cdot)$ is assumed to be unknown.

TABLE 1
Optimal sign declaration strategy for a given i.

If	Then
$P(\theta_i < 0\|z_i) < P(\theta_i > 0\|z_i)$	declare $\theta_i > 0$
$P(\theta_i > 0\|z_i) < P(\theta_i < 0\|z_i)$	declare $\theta_i < 0$
$P(\theta_i > 0\|z_i) = P(\theta_i < 0\|z_i)$	declare either $\theta_i > 0$ or $\theta_i < 0$

for any procedure $(\widehat{sgn}_i)_{i\in\mathcal{R}}$ with $\text{FDR}_{\text{dir}}\big[(\widehat{sgn}_i)_{i\in\mathcal{R}}\big] \leqslant q$, where the expectation operator defining all the FDR_{dir} and ETD quantities just mentioned is with respect to *both* the randomness of the data $\{z_i\}_{i\in[m]}$ and that of the parameters $\{\theta_i\}_{i\in[m]}$ under (3.8) (i.e. different from the frequentist $\mathbb{E}_{\boldsymbol{\theta}}[\cdot]$ operator). In other words, the ODP is the best procedure in terms of maximizing ETD, among all that can control FDR_{dir} under a target level.

In order to shrink $\mathcal{M}_t$ in a manner that Algorithm 2 can mimic the ODP, we have to better understand the latter's operational characteristics; to that end, we shall first intuitively grasp what the best course of action for a directional decision maker should be under the Bayesian assumption (3.8). If s/he were obliged to unambivalently declare a non-zero sign for a *specific* θ_i based on z_i, the optimal strategy is clearly the one outlined in Table 1 based on the posterior probabilities $P(\theta_i < 0|z_i)$ and $P(\theta_i > 0|z_i)$, whose associated probability of making a false discovery can be calculated as

$$lfsr_i = P(\theta_i \leqslant 0|z_i) \wedge P(\theta_i \geqslant 0|z_i) \tag{3.9}$$

and must be no larger than the probability of false discovery made by any other strategy. In the literature, the quantity in (3.9) is known as the *local false sign rate* for i (Stephens, 2017, p. 279), and a smaller $lfsr_i$ suggests higher confidence in the sign declaration for θ_i prescribed by Table 1. Now, if s/he were to make non-zero sign declarations for the largest possible subset of parameters from $\{\theta_i\}_{i\in[m]}$ with FDR_{dir} control in mind, the intuition would be to prioritize making sign declarations for those i's with the smallest local false sign rates, each using Table 1's strategy. This is in fact what the ODP does, as stated in Theorem 3.2 below. We note that the ODP is not implementable in practice as the underlying prior $G(\cdot)$ is, by assumption, unknown for computing the local false sign rates.

Theorem 3.2 (Operational characteristics of the ODP under Bayesian formulation)**.** *Assume the prior in* (3.8)*, and that conditional on $\{\theta_i\}_{i\in[m]}$, $z_1, \dots, z_m$ are independent with respective distributions $F_{1,\theta_1}, \dots, F_{m,\theta_m}$. For a given level $q \in (0,1)$, the optimal discovery procedure $(\widehat{sgn}_i^{ODP})_{i\in\mathcal{R}^{ODP}}$ must be such that*

(i) $lfsr_i \leqslant lfsr_j$ for any $i \in \mathcal{R}^{ODP}$ and $j \in [m]\backslash\mathcal{R}^{ODP}$, and
(ii) For each $i \in \mathcal{R}^{ODP}$, $\widehat{sgn}_i^{ODP}$ is declared in accordance with Table 1.

The proof of Theorem 3.2 is in Appendix B.3, which extends the arguments in Heller and Rosset (2021, Theorem 2.1) on the optimal FDR procedure for the point null testing problem in (1.1). While the ODP cannot be operationalized in practice, one can make ZDIRECT mimic its characteristic that indices with the smallest local false sign rates get rejected first: At each step t, we aim to get rid of exactly one element from the masked set $\mathcal{M}_t$ that has potentially the largest local false sign rate, since only elements remaining in the next $\mathcal{M}_{t+1}$ may be rejected. In what follows, let

$$z_i' \equiv F_{i,0}^{-1}(u_i'), \quad \breve{z}_i \equiv F_{i,0}^{-1}(\breve{u}_i) \quad \text{and} \quad \widetilde{z}_{i,t} \equiv \begin{cases} z_i & \text{if } i \notin \mathcal{M}_t \\ z_i' & \text{if } i \in \mathcal{M}_t \end{cases},$$

which convert u_i', $\breve{u}_i$ and $\widetilde{u}_{i,t}$ back onto their original z-value scale. Specifically, we estimate the local false sign rates as

$$\widehat{lfsr}_{i,t} = \min\left(\frac{\int_{\theta \leqslant 0} f_{i,\theta}(z_i') d\hat{G}_t(\theta)}{\int f_{i,\theta}(z_i') d\hat{G}_t(\theta)}, \frac{\int_{\theta \geqslant 0} f_{i,\theta}(z_i') d\hat{G}_t(\theta)}{\int f_{i,\theta}(z_i') d\hat{G}_t(\theta)} \right) \text{ for each } i \in \mathcal{M}_t, \tag{3.10}$$

where $\hat{G}_t(\cdot)$ is an estimate of $G(\cdot)$ based on the partial dataset $\{\tilde{z}_{i,t}\}_{i\in[m]}$, or equivalently, $\{\tilde{u}_{i,t}\}_{i\in[m]}$ from (3.4). The index to be unmasked from $\mathcal{M}_t$ is then

$$\hat{i}_t = \arg\max_{i \in \mathcal{M}_t} \widehat{lfsr}_{i,t}, \tag{3.11}$$

which has the largest estimated local false sign rate; that the estimates in (3.10) are evaluated at the z_i''s presumes that any given masked element in $\mathcal{M}_t$ may come from the rejection set $\mathcal{R}_t$.

Now we describe how we get $\hat{G}_t(\cdot)$ based on the partial data. Since the prior $G(\cdot)$ is unknown, we will model it to have a *unimodal* mixture density

$$g(\cdot; \mathbf{w}) = w_0 \delta_0(\cdot) + \sum_{k=-K,\dots,-1,1,\dots,K} w_k h_k(\cdot) \tag{3.12}$$

proposed in Stephens (2017), where $\mathbf{w} \equiv (w_{-K}, \dots, w_{-1}, w_0, w_1, \dots, w_K)$ are mixing probabilities that sum to 1, $\delta_0(\cdot)$ denotes the delta function at zero, and h_k's are uniform densities of the forms

$$h_k(\cdot) = \begin{cases} U(\cdot; 0, a_k) & \text{if } k = 1, \dots, K \\ U(\cdot; a_k, 0) & \text{if } k = -1, \dots, -K \end{cases}$$

for predetermined endpoints $a_1, \dots, a_K > 0$ and $a_{-1}, \dots, a_{-K} < 0$. The log-likelihood of (3.12) with respect to the partial data $\{\tilde{z}_{i,t}\}_{i\in[m]}$ at step t can then be computed as

$$L_t(\mathbf{w}) = \sum_{i \in [m]} \log \left[w_0 l_{0,i,t} + \sum_{k=-K,\dots,-1,1,\dots,K} w_k l_{k,i,t} \right] \tag{3.13}$$

where $l_{k,i,t}$ are the likelihoods of the mixture components of the forms

$$l_{0,i,t} = \begin{cases} f_{i,0}(z_i) + f_{i,0}(\breve{z}_i) & \text{if } i \in \mathcal{M}_t \\ f_{i,0}(z_i) & \text{if } i \in [m] \backslash \mathcal{M}_t \end{cases}$$

and

$$l_{k,i,t} = \begin{cases} \int [f_{i,\theta}(z_i) + f_{i,\theta}(\breve{z}_i)]\, h_k(\theta) d\theta & \text{if } i \in \mathcal{M}_t \\ \int f_{i,\theta}(z_i) \cdot h_k(\theta) d\theta & \text{if } i \in [m] \backslash \mathcal{M}_t \end{cases} \quad (3.14)$$
$$\text{for} \quad k = -K, \ldots, -1, 1, \ldots, K.$$

When $f_{i,\theta}$ is from the family of normal distributions $N(\theta, \sigma_i^2)$, the quantities in (3.14) have closed-form expressions; if $f_{i,\theta}$ is from the family of noncentral t-distributions $NCT(\theta, \nu_i)$, methods for approximating the quantities in (3.14) are discussed in Appendix B.1. From (3.13), the density of $\hat{G}_t$ is taken as $\hat{g}_t(\cdot) = g(\cdot; \hat{\mathbf{w}}_t)$, where $\hat{\mathbf{w}}_t$ solves the penalized maximum likelihood estimation:

$$\max_{\substack{\mathbf{w}: \\ \sum_k w_k = 1, \\ w_k \geqslant 0 \forall k}} \left[L_t(\mathbf{w}) + \sum_{k=-K}^{K} (\lambda_k - 1) \log w_k \right]. \quad (3.15)$$

Above, the last term is a Dirichlet penalty with tuning parameters $\lambda_k > 0$.

Remarks Adaptivity is implicitly built into our algorithm, since the null probability w_0 in our modeling density (3.12) is the Bayesian analogue of the frequentist null proportion in (1.3). By striving to mimic the operational characteristic of the ODP, it also allows adaptivity to other features, such as the asymmetry in the distribution of the z_i's; in the FDR literature, it is well known that *local false discovery rate* approach based on z-values can further boosts testing power by leveraging distributional asymmetry (Storey, Dai and Leek, 2007, Sun and Cai, 2007), and the same discussion can carry over to FDR_{dir} control with local false sign rates. We stress that although (3.12) may well be misspecified as a density for the hypothetical prior $G(\cdot)$, strong frequentist FDR_{dir} control is guaranteed by (3.5) in Theorem 3.1. Moreover, our choice of it as a working model carries two main advantages:

(a) *Speed*: The iterative updates of $\mathcal{M}_t$ for most existing FDR data-masking algorithms are computationally expensive, as they usually employ beta mixture models that require the EM algorithm (Dempster, Laird and Rubin, 1977) for estimation. On the contrary, as explained in Stephens (2017, Supplementary material), an optimization problem with the form in (3.15) is convex and can be solved by fast and reliable interior point methods.

(b) *Appropriate flexibility*: Stephens (2017) argues for the plausibility of unimodality in many real applications since most effects are close to zero while larger effects are decreasingly likely; by increasing the number of components K and expanding the supporting intervals defined by a_k, (3.12) can approximate *any unimodal distribution about zero* (Feller, 1971, p. 158). On the other hand, as heuristically discussed in Leung and Sun (2022, Section 5), having a too-flexible model could overfit the partial data $\{\tilde{z}_{i,t}\}_{i \in [m]}$, which one is constrained to work with to adhere to data masking, only to underfit the original data $\{z_{i,t}\}_{i \in [m]}$. Stephens (2017, Section 3.1.4) also advocates for unimodality as a form of "regularization" because it prevents density estimates from concentrating in small pockets. Our simulation results in Section 4 show that ZDIRECT can still be competitive against other practical methods even when the unimodality is misspecified.

4. Simulations

4.1. Simulation setups

We simulate $m = 1000$ independent $z_i \sim N(\theta_i, 1)$ values where $\theta_1, \ldots, \theta_m$ are independently generated from a mixture density of the form

$$g(\theta) = w\delta_0(\theta) + (1-w)g_1(\theta), \tag{4.1}$$

where $\delta_0(\cdot)$ is the delta function at zero for the nulls, and $g_1(\theta)$ is an "alternative" density which is itself a mixture of two normal components of the form

$$g_1(\theta) = (1-v)\phi(\theta+\xi) + v\phi(\theta-\xi).$$

The simulation parameters controlling different aspects are chosen as follows.

(a) w (*signal sparsity parameter*): Takes one of the values in $\{0.8, 0.5, 0.2, 0\}$. Letting $w = 0.8$ renders a setting with approximately 80% of the θ_i's equal to 0, and taking $w = 0$ renders a setting with all θ_i's being non-zero.
(b) ξ (*signal size parameter*): Takes one of the values in $\{0.5, 1, 1.5, 2, 2.5\}$. It influences the absolute value of an effect θ_i if it is non-zero.
(c) v (*asymmetry parameter*): Takes one of the values in $\{0.5, 0.75, 1\}$. It controls the proportion of the alternative θ_i's generated from the positively centered normal component $\phi(\theta-\xi)$; a larger v makes $g(\theta)$ more asymmetric.

How the different combinations of w, ξ and v change the shape of $g(\theta)$ is illustrated in Figure 4.1; note that many of these $g(\theta)$ are evidently not unimodal.

4.2. Methods compared

We compare the following seven methods for FDR_{dir} control:

(a) BH_{dir}: The directional BH procedure proposed by Benjamini and Yekutieli (2005). We note that the validity of BH_{dir} is originally proved under the assumption that the family $\{F_{i,\theta}(\cdot)\}_{\theta\in\mathbb{R}}$ is stochastically increasing, which is implied by the MLR property in Assumption 2 (Lehmann, Romano and Casella, 2005, Lemma 3.4.2 (*ii*)).
(b) LFSR: A computationally simpler substitute for the ODP based on the oracle lfsr_i's in (3.9); inspired by Sun and Cai (2007)'s earlier work on optimal FDR control. Its implementation details are deferred to Appendix C.1. The implementation of the ODP described in Theorem 3.2 entails solving a complex infinite integer problem to compute a rejection threshold

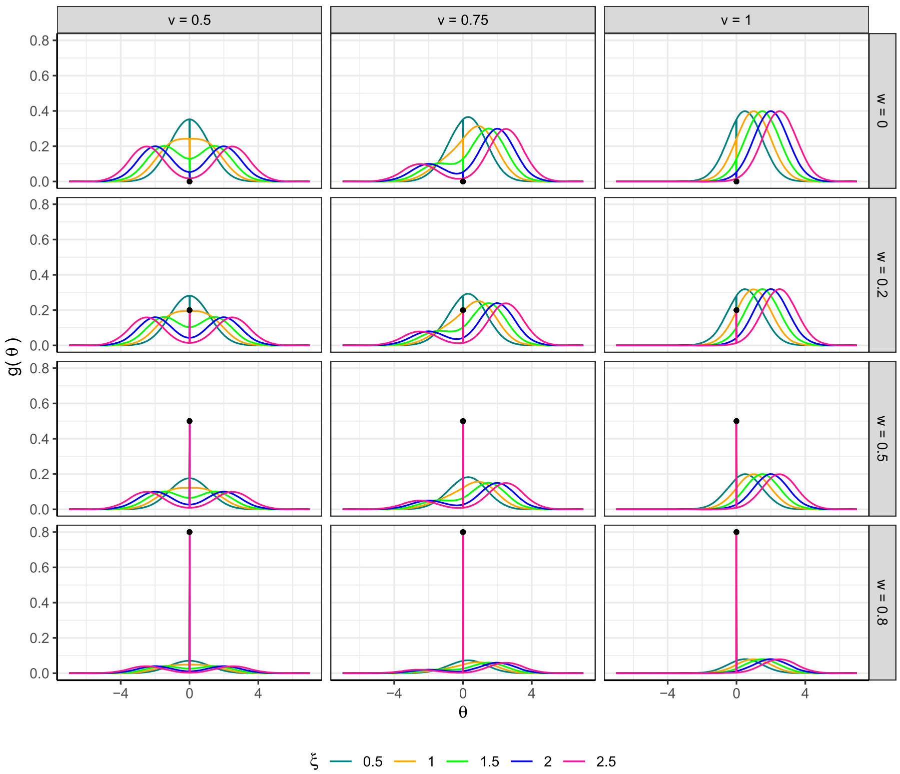

FIG 4.1. *Plots of the effect generating density $g(\theta)$ for every possible combination of w, ξ and v considered in Section 4.*

for the $lfsr_i$'s (Heller and Rosset, 2021). In comparison, LFSR is simpler to compute whilst often having comparable power. Like the ODP, LFSR offers FDR_{dir} control under the Bayesian formulation in (3.8), but it can only serve as a hypothetical benchmark for other methods since it also requires oracle knowledge of the true generating prior in (4.1).

(c) ASH ("**a**daptive **sh**rinkage", Stephens (2017)): A procedure that is almost the same as LFSR in its implementation, except that the oracle $g(\cdot)$ from (4.1) is replaced by an estimated $g(\cdot; \hat{\mathbf{w}})$ based on the unimodal model in (3.12), where $\hat{\mathbf{w}}$ is a penalized maximum likelihood estimate of $\mathbf{w}$ with respect to the full data $\{z_i\}_{i\in[m]}$ obtained by the R package `ashr`; all tuning parameters involved are chosen to be the default described in Stephens (2017, Supplementary information). This procedure may risk violating the desired FDR_{dir} target if the unimodal density (3.12) is too far from the true prior density of the θ's.

(d) GR (Guo and Romano (2015, Procedure 6)): The FDR_{dir} testing procedure mentioned in Section 1 that provides (frequentist) control of the FDR_{dir} under the target level q when all the θ_i's are non-zero, i.e. $\pi = 0$; see Guo and Romano (2015, Theorem 5 and its proof).

(e) STS_{dir}: Algorithm 1 by setting $\lambda = 0.5$.
(f) aSTS_{dir}: The STS_{dir} procedure, except with an automatic (data-driven) choice for λ as described in Appendix A.2. It has no proven theoretical control of the FDR_{dir}.
(g) ZDIRECT: Algorithm 2, by initializing $\mathcal{M}_1 = \{i : u'_i \leqslant 0.2 \text{ or } u'_i \geqslant 0.8\}$ based on the $\{\tilde{u}_{i,0}\}_{i=1}^m = \{\tilde{u}'_i\}_{i=1}^m$, which respects condition (C1). Thereafter, $\mathcal{M}_t$ is updated as in Section 3.1, where the optimization in (3.15) is performed using a solver in the R package `Rmosek` (ApS, 2022). The points a_k are picked to give a large and dense grid; in particular, for the positive supports, the minimum and maximum are set as $a_1 = 10^{-1}$ and $a_K = 2\sqrt{\max_i z_i'^2 - 1}$, with the rest set as $a_{k+1} = \sqrt{2}a_k \leqslant a_K$ based on the multiplicative factor $\sqrt{2}$ (and so, K is implicitly determined). The negative supports are set by taking $a_{-1} = -a_1, \dots, a_{-K} = -a_K$. This grid follows the recommendation of Stephens (2017) except a_K is determined with z'_i's instead of z_i to observe the masking condition (C1). Moreover, we set $\lambda_k = 0.8$ for all $k = -K, \dots, 0, \dots, K$. This further regularizes the estimation by encouraging sparsity in the estimates of the mixing proportions $\mathbf{w}$, and provides consistently good performance. Lastly, to speed up the algorithm, $\hat{g}_t(\cdot)$ is only re-estimated by (3.15) for every $\lceil m/200 \rceil$ steps, i.e., the same $\hat{g}_t(\cdot)$ is used $\lceil m/200 \rceil$ times to update the candidate rejection and acceptance sets before the algorithm terminates.

4.3. Results

The empirical FDR_{dir} and power of the different methods implemented for the target FDR_{dir} level $q = 0.1$ are evaluated with 1000 sets of repeatedly generated $\{z_i, \theta_i\}_{i \in [m]}$. The results are shown in Figure 4.2, where the power is shown as the *true positive rate*, defined as $\mathbb{E}\left[\frac{\sum_{i \in \mathcal{R}} \mathbf{1}(\text{sgn}(\mu_i) = \widehat{sgn}_i)|}{1 \vee |\mathcal{R}|}\right]$ for a generic procedure $(\widehat{sgn}_i)_{i \in \mathcal{R}}$, which some consider to be more illustrative than the ETD. The following observations can be made:

(a) Throughout, the only methods that can visibly control FDR_{dir} under the target $q = 0.1$ in all settings are LSFR, BH_{dir}, ZDIRECT and STS_{dir}, precisely the ones with theoretical guarantees. However, LSFR is not implementable in practice and only has FDR_{dir} guarantee under the Bayesian formulation in (3.8). While ZDIRECT and STS_{dir} still lag considerably behind in power compared to LFSR in settings with small w (or small π, as a frequentist analogue), their power advantage over BH_{dir} becomes substantial as w approaches 0. Across the board, aSTS_{dir}, which is calibrated with a data-driven λ and has no theoretical guarantee, displays slightly better power than STS_{dir}, but also violates the FDR_{dir} target ever so slightly for large ξ when $w = 0.8$.
(b) ASH is the one practical method that overall matches LFSR closest in power, but severely violates the desired FDR_{dir} level in some settings when $w \in \{0.2, 0.5\}$. This is not surprising because the unimodal working

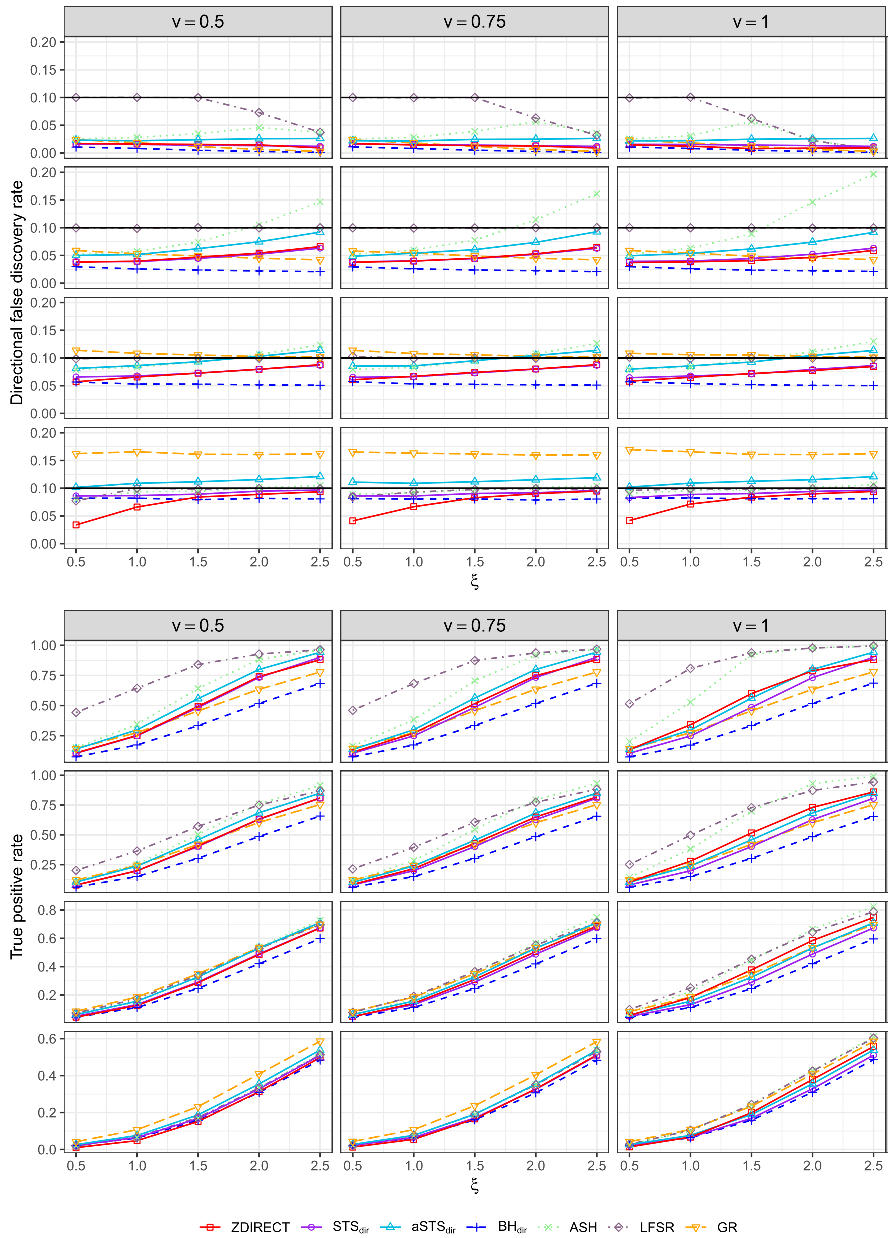

Fig 4.2. *Empirical directional false discovery rate and true positive rates of the seven compared methods for the simulations in Section 4; each method was implemented at a target* FDR_{dir} *level* $q = 0.1$ *(black horizontal lines).*

model in (3.12), which ASH is based on, is misspecified for many of the multimodal data generating $g(\cdot)$ in Figure 4.1.

(c) GR visibly violates the FDR_{dir} target when $w \in \{0.5, 0.8\}$, and severely so for $w = 0.8$. When $w = 0$, the only case where GR can provably control the FDR_{dir}, GR's power is at best comparable to STS_{dir} and ZDIRECT when the signal size ξ is small, but inferior to them when ξ is large.

(d) ZDIRECT's power is considerably better than STS_{dir} when $v = 1$, the most asymmetric setting for g. As discussed in the remarks of Section 3.1, by attempting to mimic the operational characteristics of the ODP via estimating the $lfsr_i$ quantities in (3.10), ZDIRECT has the potential to leverage asymmetry in the distribution of the z-values to boost testing power, just like the ODP does. This is even more remarkable, considering that the working model (3.12) is obviously misspecified for the true θ-generating prior in (4.1), which attests to the practical usefulness of Stephens (2017)'s unimodal assumption when combined with ZDIRECT's data masking mechanism to ensure strong FDR_{dir} control (Theorem 3.1).

5. Discussion

We have proved that, under independence and upon augmenting sign declarations, Storey, Taylor and Siegmund (2004)'s adaptive procedure and ZDIRECT, a particular implementation of the recently introduced line of adaptive "data masking" algorithms, can offer FDR_{dir} control in the strong sense. These results are particularly important when the parameter configurations contain little to no true nulls because adaptive procedures precisely reap the most power benefit in such scenarios. Moreover, under "non-sparse-signal" settings, FDR_{dir} is arguably a more meaningful error rate to be controlled than the FDR. Both methods require tuning parameters; in our experience, the simple choice of $\lambda = 0.5$ for STS_{dir} and $\lambda_k = 0.8$ for ZDIRECT have consistently given us competitive power performance, even for some less instructive simulation setups considered in earlier versions of this paper. For ZDIRECT, while other working models that may further boost the power can be deployed, we find the current implementation based on Stephens (2017)'s unimodal model to be attractive, as the optimization under the hood is numerically very stable and fast. While our theory doesn't cover settings where the z-values can be dependent, additional simulation results in this vein are included in Appendix C.2.

An interesting dual problem to sign declarations is to construct, for each i in a data-dependent selected subset $\mathcal{R} \subseteq [m]$, a confidence interval $CI_i \subseteq \mathbb{R}$ such that

(a) each CI_i is *sign-determining*, i.e. $CI_i \subseteq (-\infty, 0]$ or $CI_i \subseteq (0, \infty)$, and

(b) the false coverage rate (FCR)

$$\mathbb{E}\left[\frac{|\{i : \theta_i \in CI_i\}|}{1 \vee |\mathcal{R}|}\right]$$

is controlled under a desired level $q \in (0, 1)$.

This paradigm of inference has been recently suggested in Weinstein and Yekutieli (2020), and they proposed a first procedure that constructs such selective sign-determining confidence intervals. The adjective "selective" indicates that the set $\mathcal{R}$ is also chosen based on the same data $\{z_i\}_{i=1}^m$ the CI_i's are constructed with. Note, STS_{dir} and ZDIRECT do correspond to such procedures that construct trivially long intervals

$$CI_i = \begin{cases} (0, \infty) & \text{if } z_i > 0 \\ (-\infty, 0) & \text{if } z_i < 0 \end{cases}$$

for each i in their rejection sets. It will be a challenging but meaningful task to devise non-trivial selective sign-determining confidence intervals where the target set $\mathcal{R}$ is chosen in a more adaptive manner akin to STS_{dir} and ZDIRECT, to offer more powerful alternative procedures to Weinstein and Yekutieli (2020, Definition 2)'s procedure.

Appendix A: Additional content for Section 2

A.1. Proof of Theorem 2.1

We shall first state four intermediate results, which allow us to extend the arguments in Storey, Taylor and Siegmund (2004, Section 4.3) to prove the FDR_{dir} controlling properties of STS_{dir} under Assumptions 1 and 2.

Lemma A.1. *Under Assumption 2, for $z \in \mathbb{R}$,*

(a) if $\theta_i < 0$, then $\left[\frac{f_{i,\theta_i}(z)}{f_{i,0}(z)}\right][1 - F_{i,0}(z)] - [1 - F_{i,\theta_i}(z))] \geqslant 0$;
(b) if $\theta_i > 0$, then $\left[\frac{f_{i,\theta_i}(z)}{f_{i,0}(z)}\right] F_{i,0}(z) - F_{i,\theta_i}(z) \geqslant 0$.

Proof of Lemma A.1. To prove statement (a), let $\theta_i < 0$ and let $z_0, z_1 \in \mathbb{R}$ such that $z_0 < z_1$. By Assumption 2, we have that

$$\frac{f_{i,\theta_i}(z_0)}{f_{i,0}(z_0)} \geqslant \frac{f_{i,\theta_i}(z_1)}{f_{i,0}(z_1)}.$$

Multiplying both sides by $f_{i,0}(z_1)$ and integrating over z_1 from z_0 to ∞ yields

$$\left[\frac{f_{i,\theta_i}(z_0)}{f_{i,0}(z_0)}\right][1 - F_{i,0}(z_0)] \geqslant [1 - F_{i,\theta_i}(z_0)],$$

which proves statement (a). The proof of statement (b) follows analogously to that of (a). ☐

In the lemma below, the probability operator $P_{\theta_i=0}(\cdot)$ emphasizes the law is driven by a value of θ_i equal to zero; the operators $P_{\theta_i<0}(\cdot)$ and $P_{\theta_i>0}(\cdot)$ have similar meanings.

Lemma A.2. *Under Assumptions 1 and 2, for $0 < s \leqslant t \leqslant 1$,*

(i) $P_{\theta_i=0}\left(p_i \leqslant s | p_i \leqslant t, z_i \neq 0\right) = s/t$;
(ii) $P_{\theta_i<0}\left(p_i \leqslant s | p_i \leqslant t, z_i \geqslant 0\right) \leqslant s/t$;
(iii) $P_{\theta_i>0}\left(p_i \leqslant s | p_i \leqslant t, z_i \leqslant 0\right) \leqslant s/t$.

Proof of Lemma A.2. For any $x \in (0, 1]$, we first rewrite

$$P_{\theta_i=0}\left(p_i \leqslant x, z_i \neq 0\right) = \left[1 - F_{i,0}\left(-F_{i,0}^{-1}(x/2)\right)\right] + F_{i,0}\left(F_{i,0}^{-1}(x/2)\right) = x; \quad \text{(A.1)}$$

$$P_{\theta_i<0}\left(p_i \leqslant x, z_i \geqslant 0\right) = 1 - F_{i,\theta_i}\left(-F_{i,0}^{-1}(x/2)\right); \quad \text{(A.2)}$$

$$P_{\theta_i>0}\left(p_i \leqslant x, z_i \leqslant 0\right) = F_{i,\theta_i}\left(F_{i,0}^{-1}(x/2)\right). \quad \text{(A.3)}$$

(i) follows from (A.1) since $P_{\theta_i=0}\left(p_i \leqslant s | p_i \leqslant t, z_i \neq 0\right) = \frac{P_{\theta_i=0}(p_i \leqslant s, z_i \neq 0)}{P_{\theta_i=0}(p_i \leqslant t, z_i \neq 0)} = \frac{s}{t}$.

(ii) is obvious when $s = t$; when $s < t$, by applying the mean value theorem on $\frac{1 - F_{i,\theta}\left(-F_{i,0}^{-1}(x/2)\right)}{x}$ as a function in x in light of (A.2), there exists $c \in (s, t)$ such that, for $y \equiv F_{i,0}^{-1}(c/2)$,

$$\begin{aligned}
&\frac{\left[\frac{P_{\theta_i<0}(p_i \leqslant t, z_i \geqslant 0)}{t}\right] - \left[\frac{P_{\theta_i<0}(p_i \leqslant s, z_i \geqslant 0)}{s}\right]}{t - s} \\
&= \frac{\left[\frac{f_{i,\theta_i}(-y)}{f_{i,0}(y)}\right] F_{i,0}(y) - \left[1 - F_{i,\theta_i}(-y)\right]}{4F_{i,0}(y)^2} \\
&= \frac{\left[\frac{f_{i,\theta_i}(-y)}{f_{i,0}(-y)}\right] \left[1 - F_{i,0}(-y)\right] - \left[1 - F_{i,\theta_i}(-y))\right]}{4F_{i,0}(y)^2}, \qquad \text{(A.4)}
\end{aligned}$$

where the last equality follows from the symmetry of $F_{i,0}$ in Assumption 1. Since $\theta_i < 0$, by applying Lemma A.1 (a) to the numerator in (A.4), we get that

$$\left[\frac{P_{\theta_i<0}\left(p_i \leqslant t, z_i \geqslant 0\right)}{t}\right] - \left[\frac{P_{\theta_i<0}\left(p_i \leqslant s, z_i \geqslant 0\right)}{s}\right] \geqslant 0,$$

which is equivalent to $\frac{P_{\theta_i<0}(p_i \leqslant s, z_i \geqslant 0)}{P_{\theta_i<0}(p_i \leqslant t, z_i \geqslant 0)} \leqslant \frac{s}{t}$, and (ii) is proved. The proof for (iii) is analogous to that of (ii), using the mean value theorem on $\frac{F_{i,\theta_i}\left(F_{i,0}^{-1}(x/2)\right)}{x}$, (A.3) and Lemma A.1 (b). □

Lemma A.3. *Let $\mathcal{S}(t)$ be defined as in Theorem 2.1. Under Assumptions 1 and 2, as well as the independence among $z_1, \ldots, z_m$, $|\mathcal{S}(t)|/t$ for $0 \leqslant t < 1$ is a supermartingale with time running backward, with respect to the filtration*

$$\mathcal{F}_t = \sigma\left(\{\mathbf{1}(p_i \leqslant x), \mathbf{1}(p_i \leqslant x, sgn(z_i) \neq sgn(\theta_i))\}_{i \in [m]} : 0 < t \leqslant x \leqslant 1\right).$$

That is, for $0 < s \leqslant t \leqslant 1$, $\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{|\mathcal{S}(s)|}{s} \middle| \mathcal{F}_t\right] \leqslant \frac{|\mathcal{S}(t)|}{t}$.

Proof of Lemma A.3. Since $|\mathcal{S}(s)|$ can be written as

$$|\mathcal{S}(s)| = \sum_{i:\theta_i=0} \mathbf{1}(p_i \leqslant s, z_i \neq 0) + \sum_{i:\theta_i>0} \mathbf{1}(p_i \leqslant s, z_i \leqslant 0) + \sum_{i:\theta_i<0} \mathbf{1}(p_i \leqslant s, z_i \geqslant 0), \tag{A.5}$$

it is not difficult to observe from statements (i)–(iii) of Lemma A.2 that $|\mathcal{S}(s)|$ given $\mathcal{F}_t$ is stochastically dominated by the Binomial$(|\mathcal{S}(t)|, s/t)$ distribution. Hence, $\mathbb{E}_{\boldsymbol{\theta}}[|\mathcal{S}(s)||\mathcal{F}_t] \leqslant |\mathcal{S}(t)| \cdot (s/t)$. □

Lemma A.4. *If $Y \sim$ Binomial(n, λ), then for any $\lambda \in (0, 1)$ and $n \in \mathbb{N}$,*

$$D_n(\lambda) \equiv \mathbb{E}\left[\frac{Y}{n - Y + 1}\right] = \frac{(1 - \lambda^n)\lambda}{1 - \lambda},$$

where $D_n(\lambda)$ is strictly increasing in λ.

Proof of Lemma A.4. By direct computation,

$$\begin{aligned} D_n(\lambda) &= \sum_{i=1}^{n} \binom{n}{i} (1 - \lambda)^{n-i} \lambda^i \cdot \frac{i}{n - i + 1} \\ &= \frac{\lambda}{1 - \lambda} \sum_{i=1}^{n} \binom{n}{i-1} (1 - \lambda)^{n-i+1} \lambda^{i-1} = \frac{\lambda}{1 - \lambda} \cdot (1 - \lambda^n) \end{aligned}$$

which proves the expectation result. Taking the derivative of $D_n(\lambda)$ with respect to λ yields

$$D'_n(\lambda) = \frac{\lambda^n(n\lambda - n - 1) + 1}{(\lambda - 1)^2}.$$

Let the numerator of $D'_n(\lambda)$ be denoted as $\dot{D}'_n(\lambda) \equiv \lambda^n(n\lambda - n - 1) + 1$. To prove that $D_n(\lambda)$ is strictly increasing in λ, we will show that $\dot{D}'_n(\lambda) > 0$ by induction. Suppose $\dot{D}'_N(\lambda) > 0$ is true for *some* fixed $N \in \mathbb{N}$. Then

$$\begin{aligned} \dot{D}'_{N+1}(\lambda) &= \lambda^{N+1}((N + 1)\lambda - (N + 1) - 1) + 1 \\ &= \lambda(\dot{D}'_N(\lambda) + \lambda^N(\lambda - 1) - 1) + 1 \\ &> \lambda(\lambda^N(\lambda - 1) - 1) + 1 \\ &= (1 - \lambda)(1 - \lambda^{N+1}) \\ &> 0 \end{aligned}$$

where the first inequality stems from the inductive condition $\dot{D}'_N(\lambda) > 0$. Since $\dot{D}'_1(\lambda) = (\lambda - 1)^2 > 0$, it follows that $\dot{D}'_n(\lambda) > 0$ for any $\lambda \in (0, 1)$ and $n \in \mathbb{N}$. □

Now we can begin the proof of Theorem 2.1. First, we can write

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\theta}}\left[\frac{|\mathcal{S}(t_q^\lambda)|}{|\mathcal{R}(t_q^\lambda)| \vee 1}\right] &= \mathbb{E}_{\boldsymbol{\theta}}\left[\frac{|\mathcal{S}(t_q^\lambda)|}{|\mathcal{R}(t_q^\lambda)| \vee 1}; \widehat{\mathrm{FDR}}_\lambda(\lambda) \geqslant q\right] \\ &\quad + \mathbb{E}_{\boldsymbol{\theta}}\left[\frac{|\mathcal{S}(t_q^\lambda)|}{|\mathcal{R}(t_q^\lambda)| \vee 1}; \widehat{\mathrm{FDR}}_\lambda(\lambda) < q\right]. \end{aligned}$$

If $\widehat{\mathrm{FDR}}_\lambda(\lambda) \geqslant q$, then $t_q^\lambda \leqslant \lambda$. Hence, $|\mathcal{R}(t_q^\lambda)| \vee 1 \geqslant \hat{\pi}_0(\lambda) t_q^\lambda m/q$ and so

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{|\mathcal{S}(t_q^\lambda)|}{|\mathcal{R}(t_q^\lambda)| \vee 1}; \widehat{\mathrm{FDR}}_\lambda(\lambda) \geqslant q\right] \\
&\quad \leqslant \mathbb{E}_{\boldsymbol{\theta}}\left[q \frac{1-\lambda}{|\{i : p_i > \lambda\}| + 1} \frac{|\mathcal{S}(t_q^\lambda)|}{t_q^\lambda}; \widehat{\mathrm{FDR}}_\lambda(\lambda) \geqslant q\right] \\
&\quad = \mathbb{E}_{\boldsymbol{\theta}}\left[q \frac{1-\lambda}{|\{i : p_i > \lambda\}| + 1} \mathbb{E}_{\boldsymbol{\theta}}\left[\left.\frac{|\mathcal{S}(t_q^\lambda)|}{t_q^\lambda}\right| \mathcal{F}_\lambda\right]; \widehat{\mathrm{FDR}}_\lambda(\lambda) \geqslant q\right] \\
&\quad \leqslant \mathbb{E}_{\boldsymbol{\theta}}\left[q \frac{1-\lambda}{|\{i : p_i > \lambda\}| + 1} \frac{|\mathcal{S}(\lambda)|}{\lambda}; \widehat{\mathrm{FDR}}_\lambda(\lambda) \geqslant q\right]
\end{aligned}$$

where the last step follows by Lemma A.3 and the optional stopping theorem since t_q^λ is a stopping time with respect to $\mathcal{F}_t$ with time running backward. If $\widehat{\mathrm{FDR}}_\lambda(\lambda) < q$, then $t_q^\lambda = \lambda$ and so

$$\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{|\mathcal{S}(t_q^\lambda)|}{|\mathcal{R}(t_q^\lambda)| \vee 1}; \widehat{\mathrm{FDR}}_\lambda(\lambda) < q\right] \leqslant \mathbb{E}_{\boldsymbol{\theta}}\left[q \frac{1-\lambda}{|\{i : p_i > \lambda\}| + 1} \frac{|\mathcal{S}(\lambda)|}{\lambda}; \widehat{\mathrm{FDR}}_\lambda(\lambda) < q\right].$$

Hence,

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{|\mathcal{S}(t_q^\lambda)|}{|\mathcal{R}(t_q^\lambda)| \vee 1}\right] &\leqslant \mathbb{E}_{\boldsymbol{\theta}}\left[q \frac{1-\lambda}{|\{i : p_i > \lambda\}| + 1} \frac{|\mathcal{S}(\lambda)|}{\lambda}\right] \\
&\leqslant \mathbb{E}_{\boldsymbol{\theta}}\left[q \frac{|\mathcal{S}(\lambda)|}{|\mathcal{S}(1)| - |\mathcal{S}(\lambda)| + 1} \frac{1-\lambda}{\lambda}\right].
\end{aligned}$$

By taking $s = \lambda$ and $t = 1$ in statements (i)–(iii) of Lemma A.2, in light of the equality in (A.5), it is not difficult to see that $|\mathcal{S}(\lambda)|$ given $|\mathcal{S}(1)|$ is stochastically dominated by the Binomial$(|\mathcal{S}(1)|, \lambda)$ distribution. Hence,

$$\mathbb{E}_{\boldsymbol{\theta}}\left[q \frac{|\mathcal{S}(\lambda)|}{|\mathcal{S}(1)| - |\mathcal{S}(\lambda)| + 1} \frac{1-\lambda}{\lambda}\right] \leqslant \left(1 - \mathbb{E}_{\boldsymbol{\theta}}[\lambda^{|\mathcal{S}(1)|}]\right) q \tag{A.6}$$

by Lemma A.4. Combining (A.6) with $1 - \mathbb{E}_{\boldsymbol{\theta}}[\lambda^{|\mathcal{S}(1)|}] \leqslant 1 - \lambda^{\mathbb{E}_{\boldsymbol{\theta}}[|\mathcal{S}(1)|]} \leqslant 1$, a consequence of Jensen's inequality, Theorem 2.1 is proved.

A.2. Automatic $\boldsymbol{\lambda}$ selection procedure

Two inputs are required for this procedure: B, the number of bootstrap samples; and Λ, a set of candidate values for λ. Our recommendations are $B = 1000$ and $\Lambda = \{0.05, 0.10, \ldots, 0.95\}$. The procedure is summarized in the following algorithm.

(1) Compute $\hat{\pi}(\lambda')$ for each $\lambda' \in \Lambda$.

(2) For each $\lambda' \in \Lambda$, construct B bootstrap estimates $\{\hat{\pi}^b(\lambda')\}_{b=1}^B$ by bootstrap sampling the p-values.
(3) Compute $\hat{\pi}^{\min} \equiv \min_{\lambda' \in \Lambda} \ \hat{\pi}(\lambda')$.
(4) For each $\lambda' \in \Lambda$, compute

$$\widehat{MSE}(\lambda') = \frac{1}{B} \sum_{b=1}^{B} [\hat{\pi}^b(\lambda') - \hat{\pi}^{\min}]^2.$$

(5) Output the estimated optimal tuning parameter $\hat{\lambda} = \mathrm{argmin}_{\lambda' \in \Lambda} \{\widehat{MSE}(\lambda')\}$.

The above algorithm is nearly identical to that of Storey, Taylor and Siegmund (2004)'s automatic λ selection algorithm (Section 6), except that Storey, Taylor and Siegmund (2004) omit the additive factor "1" in the numerator of all the estimators for π involved, but we retain it to robustify the $\mathrm{FDR}_{\mathrm{dir}}$ controlling property of the resulting procedure. Regardless, the intuition behind is the same, i.e. choose a λ which minimizes an estimated mean square error.

Appendix B: Additional content for Section 3

B.1. Computation of the component loglikelihoods

We discuss computations of the likelihoods in (3.14) when $f_{i,\theta}$ belongs to the normal family $N(\theta, \sigma_i^2)$ or the noncentral t-distributional family $NCT(\theta, \nu_i)$.

(i) $N(\theta, \sigma_i)$: In this case, each $l_{k,i,t}$ in (3.14) has the explicit analytic form

$$l_{k,i,t} = \begin{cases} \frac{\Phi(z_i/\sigma_i) - \Phi((z_i - a_k)/\sigma_i)}{a_k} + \frac{\Phi(\check{z}_i/\sigma_i) - \Phi((\check{z}_i - a_k)/\sigma_i)}{a_k} \\ \quad \text{if } i \in \mathcal{M}_t \text{ and } k \geqslant 1 \\ \frac{\Phi(z_i/\sigma_i) - \Phi((z_i - a_k)/\sigma_i)}{a_k} \\ \quad \text{if } i \in [m] \backslash \mathcal{M}_t \text{ and } k \geqslant 1 \\ \frac{\Phi((z_i - a_k)/\sigma_i) - \Phi(z_i/\sigma_i)}{-a_k} + \frac{(\Phi(\check{z}_i - a_k)/\sigma_i) - \Phi(\check{z}_i/\sigma_i)}{-a_k} \\ \quad \text{if } i \in \mathcal{M}_t \text{ and } k \leqslant -1 \\ \frac{\Phi((z_i - a_k)/\sigma_i) - \Phi(z_i/\sigma_i)}{-a_k} \\ \quad \text{if } i \in [m] \backslash \mathcal{M}_t \text{ and } k \leqslant -1 \end{cases}.$$

(ii) $NCT(\theta, \nu_i)$: Without sophisticated numerical integration methods, it may be hard to obtain good numerical values for the quantities in (3.14). However, approximation methods can be potentially leveraged; in what follows we assume the common use case where $\nu_i = \nu$ for all $i \in [m]$. In a variance-stabilizing manner, Laubscher (1960, Section 2) suggests that, if z_i is a noncentral t-distributed random variable with noncentrality parameter θ and degree $\nu \geqslant 4$, by *bijectively* transforming z_i to the variable

$$\xi_i \equiv \alpha \sinh^{-1}(\beta z_i),$$

where $\alpha = \alpha(\nu)$ and $\beta = \beta(\nu)$ are positive numbers depending only on ν,

$$\xi_i \text{ is approximately distributed as } N(\gamma, 1)$$

for the mean

$$\gamma \equiv \alpha \sinh^{-1}\left(\beta \cdot \theta \cdot \frac{\Gamma(\nu/2 - 1/2)\sqrt{\nu/2}}{\Gamma(\nu/2)}\right),$$

which is also strictly increasing in θ. Hence, by also letting $\check{\xi}_i \equiv \alpha \sinh^{-1}(\beta \check{z}_i)$, one can alternatively implement ZDIRECT by replacing the component likelihoods in (3.14) with

$$l'_{k,i,t} = \begin{cases} \frac{\Phi(\xi_i)-\Phi(\xi_i-a_k)}{a_k} + \frac{\Phi(\check{\xi}_i)-\Phi(\check{\xi}_i-a_k)}{a_k} & \text{if } i \in \mathcal{M}_t \text{ and } k \geqslant 1 \\ \frac{\Phi(\xi_i)-\Phi(\xi_i-a_k)}{a_k} & \text{if } i \in [m]\backslash\mathcal{M}_t \text{ and } k \geqslant 1 \\ \frac{\Phi(\xi_i-a_k)-\Phi(\xi_i)}{-a_k} + \frac{\Phi(\check{\xi}_i-a_k)-\Phi(\check{\xi}_i)}{-a_k} & \text{if } i \in \mathcal{M}_t \text{ and } k \leqslant -1 \\ \frac{\Phi(\xi_i-a_k)-\Phi(\xi_i)}{-a_k} & \text{if } i \in [m]\backslash\mathcal{M}_t \text{ and } k \leqslant -1 \end{cases}.$$

In doing so, we have essentially imposed the prior $g(\cdot, \mathbf{w})$ in (3.12) on γ instead of θ, but it doesn't change things in the grand scheme as it still approximates a unimodal density about zero on θ; note that $\sinh^{-1}(0) = 0$. Importantly, we are still working with the partial data $\{\tilde{z}_{i,t}\}_{i \in [m]}$ so strong FDR_{dir} control is guaranteed by virtue of Theorem 3.1. Other such strategies may also be explored, possibly based on ideas from Kraemer and Paik (1979) and other references therein.

B.2. Proof of Theorem 3.1

We first quote Lei and Fithian (2018, Lemma 2), a fundamental tool for developing data-masking algorithms.

Lemma B.1. *Suppose that, conditionally on the σ-field $\mathcal{G}_{-1}$, $b_1, \ldots, b_n$ are independent Bernoulli random variables with $P(b_i = 1 \mid \mathcal{G}_{-1}) = \rho_i \geqslant \rho > 0$, almost surely. Also suppose that $[n] \supseteq \mathcal{C}_0 \supseteq \mathcal{C}_1 \supseteq \cdots$, with each subset $\mathcal{C}_{t+1}$ measurable with respect to*

$$\mathcal{G}_t = \sigma\left\{\mathcal{G}_{-1}, \mathcal{C}_t, (b_i)_{i \notin \mathcal{C}_t}, \sum_{i \in \mathcal{C}_t} b_i\right\}.$$

If $\hat{t}$ is an almost-surely finite stopping time with respect to the filtration $(\mathcal{G}_t)_{t \geqslant 0}$, then

$$\mathbb{E}\left[\frac{1 + |\mathcal{C}_{\hat{t}}|}{1 + \sum_{i \in \mathcal{C}_{\hat{t}}} b_i} \mid \mathcal{G}_{-1}\right] \leqslant \rho^{-1}.$$

In Lemma B.1, we remark that $(\mathcal{G}_t)_{t \geqslant 0}$ defines a filtration precisely because $\mathcal{C}_{t+1} \in \mathcal{G}_t$.

Proof of Theorem 3.1. The arguments below are inspired by those from Barber et al. (2019) for establishing the FDR_{dir} controlling property of the knockoff filter for variable selection in linear regressions. To begin our proof, write the directional false discovery proportion as

$$\begin{aligned}\text{FDP}_{\text{dir}}(\hat{t}) &= \frac{|\{i : \text{sgn}(\theta_i) \neq \text{sgn}(z_i) \text{ and } i \in \mathcal{R}_{\hat{t}}\}|}{1 \vee |\mathcal{R}_{\hat{t}}|} \\ &= \frac{|\{i : \text{sgn}(\theta_i) \neq \text{sgn}(z_i) \text{ and } i \in \mathcal{R}_{\hat{t}}\}|}{1 + |\mathcal{A}_{\hat{t}}|} \frac{1 + |\mathcal{A}_{\hat{t}}|}{1 \vee |\mathcal{R}_{\hat{t}}|}.\end{aligned}$$

Since $\widehat{\text{FDR}}_{\text{dir}}(\hat{t}) = \frac{1+|\mathcal{A}_{\hat{t}}|}{1\vee|\mathcal{R}_{\hat{t}}|} \leqslant q$ by definition, it suffices to show that

$$\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{|\{i : \text{sgn}(\theta_i) \neq \text{sgn}(z_i) \text{ and } i \in \mathcal{R}_{\hat{t}}\}|}{1 + |\mathcal{A}_{\hat{t}}|}\right] \leqslant 1. \tag{B.1}$$

For each $i \in [m]$, we define the variables

$$b_i \equiv \mathbf{1}\Big(u_i \in (0.25, 0.75)\Big)$$

and

$$E_i \equiv 1 - 2b_i = \begin{cases} +1 & \text{if } |z_i| \geqslant |\breve{z}_i| \\ -1 & \text{if } |\breve{z}_i| > |z_i| \end{cases},$$

where, by the symmetry of $f_{i,0}(\cdot)$ from Assumption 1, the latter is equal to $+1$ if u_i is at least as close as $\breve{u}_i$ to the endpoints of the unit interval $[0,1]$, or equal to -1 otherwise. In particular, since

$$\mathcal{R}_{\hat{t}} \equiv \{i : i \in \mathcal{M}_{\hat{t}} \text{ and } u_i \in (0, 0.25] \cup [0.75, 1)\},$$

any $i \in \mathcal{R}_{\hat{t}}$ must have its corresponding u_i at least as close to the two endpoints of $[0,1]$ as its reflection $\breve{u}_i \in [0.25, 0.75]$. As such, it must always be that

$$\mathcal{R}_{\hat{t}} \subseteq \{i : E_i = +1\}, \tag{B.2}$$

i.e., an element i can possibly be a discovery only if $E_i = +1$. Moreover, define

$$S_i = \text{sgn}(z_i)E_i,$$

which will take on the same sign as z_i if $E_i = +1$ and $z_i \neq 0$, as well as the set

$$\hat{\mathcal{H}}_0 \equiv \{i : S_i \neq \text{sgn}(\theta_i)\}.$$

Here, $\hat{\mathcal{H}}_0$ can act like a "random null set" since a false discovery precisely amounts to declaring a non-zero sign for any $i \in \hat{\mathcal{H}}_0 \cap \{i : E_i = +1\}$, in light of (B.2) being always true. Hence,

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{|\{i : \text{sgn}(\theta_i) \neq \text{sgn}(z_i) \text{ and } i \in \mathcal{R}_{\hat{t}}\}|}{1 + |\mathcal{A}_{\hat{t}}|}\right] &= \mathbb{E}_{\boldsymbol{\theta}}\left[\frac{|\mathcal{R}_{\hat{t}} \cap \hat{\mathcal{H}}_0|}{1 + |\mathcal{A}_{\hat{t}}|}\right] \\ &\leqslant \mathbb{E}_{\boldsymbol{\theta}}\left[\frac{|\mathcal{R}_{\hat{t}} \cap \hat{\mathcal{H}}_0|}{1 + |\mathcal{A}_{\hat{t}} \cap \hat{\mathcal{H}}_0|}\right].\end{aligned}$$

The last inequality in the preceding display implies that (B.1) can be proved if it is true that $\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{|\mathcal{R}_{\hat{t}}\cap\hat{\mathcal{H}}_0|}{1+|\mathcal{A}_{\hat{t}}\cap\hat{\mathcal{H}}_0|}\right]\leqslant 1$, which, in turn, we will prove by showing

$$\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{|\mathcal{R}_{\hat{t}}\cap\hat{\mathcal{H}}_0|}{1+|\mathcal{A}_{\hat{t}}\cap\hat{\mathcal{H}}_0|}\middle|\sigma\{\hat{\mathcal{H}}_0\}\right]\leqslant 1. \tag{B.3}$$

The rest of the proof proceeds by setting the scene to apply Lemma B.1. First, recall $\mathcal{M}_0=[m]$ and let $\mathcal{G}_{-1}\equiv\sigma\{\hat{\mathcal{H}}_0\}$. For $t=0,1,\dots,$ define

$$\mathcal{C}_t\equiv\hat{\mathcal{H}}_0\cap\mathcal{M}_t,$$

and the filtrations

$$\mathcal{G}_t\equiv\sigma\left\{\mathcal{G}_{-1},\mathcal{C}_t,(b_i)_{i\notin\mathcal{C}_t},\sum_{i\in\mathcal{C}_t}b_i\right\};$$

note that $\mathcal{C}_{t+1}\in\mathcal{G}_t$ since $\hat{\mathcal{H}}_0\in\mathcal{G}_{-1}$ and $\mathcal{M}_{t+1}\in\mathcal{G}_t$ by respecting the condition (C1). By the definitions of $\mathcal{A}_t$ and $\mathcal{R}_t$, we must have that

$$|\mathcal{A}_t\cap\hat{\mathcal{H}}_0|=\sum_{i\in\mathcal{C}_t}b_i \text{ and } |\mathcal{R}_t\cap\hat{\mathcal{H}}_0|=|\mathcal{C}_t|-\sum_{i\in\mathcal{C}_t}b_i. \tag{B.4}$$

Writing

$$\begin{aligned}|\mathcal{A}_t|&=|\mathcal{A}_t\cap\hat{\mathcal{H}}_0|+|\{i:i\notin\hat{\mathcal{H}}_0,i\in\mathcal{M}_t\text{ and }u_i\in(0.25,0.75)\}|\\&=|\mathcal{A}_t\cap\hat{\mathcal{H}}_0|+|\{i:i\in([m]\backslash\hat{\mathcal{H}}_0)\cap\mathcal{M}_t\text{ and }b_i=1\}|\end{aligned}$$

and

$$\begin{aligned}|\mathcal{R}_t|&=|\mathcal{R}_t\cap\hat{\mathcal{H}}_0|+|\{i:i\notin\hat{\mathcal{H}}_0,i\in\mathcal{M}_t\text{ and }u_i\in(0,0.25]\cup[0.75,1)\}|\\&=|\mathcal{R}_t\cap\hat{\mathcal{H}}_0|+|\{i:i\in([m]\backslash\hat{\mathcal{H}}_0)\cap\mathcal{M}_t\text{ and }b_i=0\}|,\end{aligned}$$

one can see that $|\mathcal{A}_t|,|\mathcal{R}_t|\in\mathcal{G}_t$ for two reasons: First, $|\mathcal{A}_t\cap\hat{\mathcal{H}}_0|$ and $|\mathcal{R}_t\cap\hat{\mathcal{H}}_0|$ belong to $\mathcal{G}_t$ because of (B.4). Second, for any $i\in([m]\backslash\hat{\mathcal{H}}_0)\cap\mathcal{M}_t$, it must also be true that $i\notin\mathcal{C}_t$ (by the definition of $\mathcal{C}_t$), which implies that b_i belongs to $\mathcal{G}_t$; as such, both $|\{i:i\in([m]\backslash\hat{\mathcal{H}}_0)\cap\mathcal{M}_t$ and $b_i=1\}|$ and $|\{i:i\in([m]\backslash\hat{\mathcal{H}}_0)\cap\mathcal{M}_t$ and $b_i=0\}|$ are measurable with respect to $\mathcal{G}_t$. Hence, $\hat{t}$ is a stopping time with respect to $(\mathcal{G}_t)_{t\geqslant 0}$, and is almost surely finite because ZDIRECT guarantees to terminate in light of the condition (C2).

Lastly, by writing

$$\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{|\mathcal{R}_{\hat{t}}\cap\hat{\mathcal{H}}_0|}{1+|\mathcal{A}_{\hat{t}}\cap\hat{\mathcal{H}}_0|}\middle|\mathcal{G}_{-1}\right]=\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{|\mathcal{C}_t|-\sum_{i\in\mathcal{C}_t}b_i}{1+\sum_{i\in\mathcal{C}_t}b_i}\middle|\mathcal{G}_{-1}\right]=\mathbb{E}_{\boldsymbol{\theta}}\left[\frac{1+|\mathcal{C}_t|}{1+\sum_{i\in\mathcal{C}_t}b_i}\middle|\mathcal{G}_{-1}\right]-1,$$

in light of Lemma B.1, one only need to show that

$$P(b_i=1\mid\mathcal{G}_{-1})\geqslant 0.5 \text{ for each } i\in\hat{\mathcal{H}}_0$$

to wrap up the proof of (B.3). We can break this into three cases; in what follows we also use the operator symbols $P_{\theta_i=0}(\cdot)$, $P_{\theta_i>0}(\cdot)$ and $P_{\theta_i<0}(\cdot)$ defined immediately before Lemma A.2 to emphasize the underlying value of θ_i driving the law:

- Case 1, $\theta_i > 0$: Since $i \in \hat{\mathcal{H}}_0$, under $\theta_i > 0$ it must be that $S_i = -1$ or 0. This can be true with either $u_i \in [0.5, 0.75)$ or $u_i \in (0, 0.25]$, only the former of which can give $b_i = 1$. These two events respectively have probabilities

$$P_{\theta_i>0}(u_i \in [0.5, 0.75)) = \int_{[F_{i,0}^{-1}(0.5), F_{i,0}^{-1}(0.75))} f_{i,\theta_i}(z)dz = \int_{[F_{i,0}^{-1}(0.5), F_{i,0}^{-1}(0.75))} \frac{f_{i,\theta_i}(z)}{f_{i,0}(z)} f_{i,0}(z)dz$$

and

$$P_{\theta_i>0}(u_i \in (0, 0.25]) = \int_{(-\infty, F_{i,0}^{-1}(0.25)]} f_{i,\theta_i}(z)dz = \int_{(-\infty, F_{i,0}^{-1}(0.25)]} \frac{f_{i,\theta_i}(z)}{f_{i,0}(z)} f_{i,0}(z)dz.$$

By the MLR property in Assumption 2, $P_{\theta_i>0}(u_i \in [0.5, 0.75)) \geqslant P_{\theta_i>0}(u_i \in (0, 0.25])$ and hence

$$P_{\theta_i>0}(b_i = 1 \mid \mathcal{G}_{-1}) = \frac{P_{\theta_i>0}(u_i \in [0.5, 0.75))}{P_{\theta_i>0}(u_i \in [0.5, 0.75)) + P_{\theta_i>0}(u_i \in (0, 0.25])} \geqslant 0.5.$$

- Case 2, $\theta_i < 0$: The derivations are completely analogous to that of Case 1.
- Case 3, $\theta_i = 0$: In that case, S_i can be $+1$ or -1. Since u_i is uniformly distributed under $\theta_i = 0$, it is easy to see that $P_{\theta_i=0}(b_i = 1 \mid \mathcal{G}_{-1}) = 0.5$.

(We remark that the arguments above work precisely because $\mathcal{G}_{-1}$ only provides the meager knowledge of $\hat{\mathcal{H}}_0$, without any other knowledge about the data $\{z_i\}_{i=1}^m$.) □

B.3. Proof of Theorem 3.2

For two procedures $(\widehat{sgn}_i^{(1)})_{i\in\mathcal{R}^{(1)}}$ and $(\widehat{sgn}_i^{(2)})_{i\in\mathcal{R}^{(2)}}$, $(\widehat{sgn}_i^{(2)})_{i\in\mathcal{R}^{(2)}}$ is said to improve upon $(\widehat{sgn}_i^{(1)})_{i\in\mathcal{R}^{(1)}}$ if $\text{ETD}[(\widehat{sgn}_i^{(2)})_{i\in\mathcal{R}^{(2)}}] \geqslant \text{ETD}[(\widehat{sgn}_i^{(1)})_{i\in\mathcal{R}^{(1)}}]$ and $\text{FDR}_{\text{dir}}[(\widehat{sgn}_i^{(2)})_{i\in\mathcal{R}^{(2)}}] \leqslant \text{FDR}_{\text{dir}}[(\widehat{sgn}_i^{(1)})_{i\in\mathcal{R}^{(1)}}]$.

Let $\mathbf{z} = (z_1, \dots, z_m)$, and let $(\widehat{sgn}_i)_{i\in\mathcal{R}}$ be a certain procedure with $\text{FDR}_{\text{dir}}[(\widehat{sgn}_i)_{i\in\mathcal{R}}] \leqslant q$. We also write $\mathcal{R} = \mathcal{R}(\mathbf{z})$ and $\widehat{sgn}_i = \widehat{sgn}_i(\mathbf{z})$ to emphasize that both are functions in $\mathbf{z}$. It suffices to show that the two statements below are true:

- Statement 1: If there exists $j \in [m]$ such that one or both of the disjoint events

$$\mathcal{Z}_j^{(1)} \equiv \left\{ \mathbf{z} \,\middle|\, \begin{array}{l} j \in \mathcal{R}(\mathbf{z}); \\ P(\theta_j \leqslant 0|z_j) < P(\theta_j \geqslant 0|z_j); \\ \widehat{sgn}_j(\mathbf{z}) = -1. \end{array} \right\} \textit{ and}$$
$$\mathcal{Z}_j^{(2)} \equiv \left\{ \mathbf{z} \,\middle|\, \begin{array}{l} j \in \mathcal{R}(\mathbf{z}); \\ P(\theta_j \leqslant 0|z_j) > P(\theta_j \geqslant 0|z_j); \\ \widehat{sgn}_j(\mathbf{z}) = 1. \end{array} \right\}$$

  have non-zero probabilities, the procedure $(\widehat{sgn}'_i)_{i \in \mathcal{R}'}$ defined by

$$\mathcal{R}'(\mathbf{z}) = \mathcal{R}(\mathbf{z}) \text{ for all } \mathbf{z}$$

  and, for $i \in \mathcal{R} = \mathcal{R}'$,

$$\widehat{sgn}'_i(\mathbf{z}) = \begin{cases} 1 & \text{if } i = j \text{ and } \mathbf{z} \in \mathcal{Z}_j^{(1)} \\ -1 & \text{if } i = j \text{ and } \mathbf{z} \in \mathcal{Z}_j^{(2)} \\ \widehat{sgn}_i(\mathbf{z}) & \text{if otherwise} \end{cases}$$

  improves upon $(\widehat{sgn}_i)_{i \in \mathcal{R}}$.
- Statement 2: If there exist two distinct $j, l \in [m]$ such that the event

$$\mathcal{Z}_{jl} = \{\mathbf{z} : \mathit{lfsr}_j < \mathit{lfsr}_l, l \in \mathcal{R} \text{ and } j \notin \mathcal{R}\}$$

  has non-zero probability, then it is possible to construct an improved procedure $(\widehat{sgn}'_i)_{i \in \mathcal{R}'}$ with the property that

$$\mathcal{R}'(\mathbf{z}) = \begin{cases} (\mathcal{R}(\mathbf{z}) \backslash \{l\}) \cup \{j\} & \text{if } \mathbf{z} \in \mathcal{Z}_{jl} \\ \mathcal{R}(\mathbf{z}) & \text{if } \mathbf{z} \notin \mathcal{Z}_{jl} \end{cases}.$$

Suppose both statements can be shown. Then any procedure can be improved by repeatedly applying Statement 2, until we end up with a procedure for which $P(\mathcal{Z}_{jl}) = 0$ for all (j, l) pairs. We can then further improve this procedure by applying Statement 1, and end up with a procedure for which $P(\mathcal{Z}_j^{(1)}) = P(\mathcal{Z}_j^{(2)}) = 0$ for all j, and hence satisfying the conditions (i) and (ii) in Theorem 3.2; since the ODP cannot be improved, it must have the latter two conditions satisfied.

Proof of Statement 1. We write

$$\begin{aligned} &\mathrm{ETD}[(\widehat{sgn}'_i)_{i \in \mathcal{R}'}] - \mathrm{ETD}[(\widehat{sgn}_i)_{i \in \mathcal{R}}] \\ &= \int \sum_{i \in \mathcal{R}'(\mathbf{z})} P(\widehat{sgn}'_i(\mathbf{z}) = \operatorname{sgn}(\theta_i)|\mathbf{z}) P(\mathbf{z}) d\mathbf{z} \\ &\quad - \int \sum_{i \in \mathcal{R}(\mathbf{z})} P(\widehat{sgn}_i(\mathbf{z}) = \operatorname{sgn}(\theta_i)|\mathbf{z}) P(\mathbf{z}) d\mathbf{z} \end{aligned}$$

$$
\begin{aligned}
&= \int_{\mathcal{Z}_j^{(1)} \cup \mathcal{Z}_j^{(2)}} \sum_{i \in \mathcal{R}'(\mathbf{z})} [P(\widehat{sgn}'_i(\mathbf{z}) = \mathrm{sgn}(\theta_i)|\mathbf{z}) - P(\widehat{sgn}_i(\mathbf{z}) = \mathrm{sgn}(\theta_i)|\mathbf{z})] P(\mathbf{z}) d\mathbf{z} \\
&= \int_{\mathcal{Z}_j^{(1)} \cup \mathcal{Z}_j^{(2)}} [P(\widehat{sgn}'_j(\mathbf{z}) = \mathrm{sgn}(\theta_j)|\mathbf{z}) - P(\widehat{sgn}_j(\mathbf{z}) = \mathrm{sgn}(\theta_j)|\mathbf{z})] P(\mathbf{z}) d\mathbf{z} \\
&= \int_{\mathcal{Z}_j^{(1)}} [P(\widehat{sgn}'_j(\mathbf{z}) = \mathrm{sgn}(\theta_j)|\mathbf{z}) - P(\widehat{sgn}_j(\mathbf{z}) = \mathrm{sgn}(\theta_j)|\mathbf{z})] P(\mathbf{z}) d\mathbf{z} \\
&\quad + \int_{\mathcal{Z}_j^{(2)}} [P(\widehat{sgn}'_j(\mathbf{z}) = \mathrm{sgn}(\theta_j)|\mathbf{z}) - P(\widehat{sgn}_j(\mathbf{z}) = \mathrm{sgn}(\theta_j)|\mathbf{z})] P(\mathbf{z}) d\mathbf{z} \\
&= \int_{\mathcal{Z}_j^{(1)}} [P(\theta_j > 0|z_j) - P(\theta_j < 0|z_j)] P(\mathbf{z}) d\mathbf{z} \\
&\quad + \int_{\mathcal{Z}_j^{(2)}} [P(\theta_j < 0|z_j)) - P(\theta_j > 0|z_j)] P(\mathbf{z}) d\mathbf{z} > 0,
\end{aligned}
$$

where the second and third equalities come from the fact that $(\widehat{sgn}'_i)_{i \in \mathcal{R}'}$ and $(\widehat{sgn}_i)_{i \in \mathcal{R}}$ differ only on $\mathcal{Z}_j^{(1)} \cup \mathcal{Z}_j^{(2)}$ and for j, the fourth equality is from the disjointness of $\mathcal{Z}_j^{(1)}$ and $\mathcal{Z}_j^{(2)}$, and the last equality is from how $\widehat{sgn}'_j$ is defined on $\mathcal{Z}_j^{(1)}$ and $\mathcal{Z}_j^{(2)}$ as well as the independence across all $i = 1, \dots, m$. Similarly,

$$
\begin{aligned}
&\mathrm{FDR}_{\mathrm{dir}}[(\widehat{sgn}_i)_{i \in \mathcal{R}}] - \mathrm{FDR}_{\mathrm{dir}}[(\widehat{sgn}'_i)_{i \in \mathcal{R}'}] \\
&\quad = \int \frac{\sum_{i \in \mathcal{R}(\mathbf{z})} P(\widehat{sgn}_i(\mathbf{z}) \neq \mathrm{sgn}(\theta_i)|\mathbf{z})}{1 \vee |\mathcal{R}(\mathbf{z})|} P(\mathbf{z}) d\mathbf{z} \\
&\qquad - \int \frac{\sum_{i \in \mathcal{R}'(\mathbf{z})} P(\widehat{sgn}'_i(\mathbf{z}) \neq \mathrm{sgn}(\theta_i)|\mathbf{z})}{1 \vee |\mathcal{R}'(\mathbf{z})|} P(\mathbf{z}) d\mathbf{z} \\
&\quad = \int_{\mathcal{Z}_j^{(1)}} \left[\frac{P(\widehat{sgn}_j(\mathbf{z}) \neq \mathrm{sgn}(\theta_j)|\mathbf{z})}{1 \vee |\mathcal{R}(\mathbf{z})|} - \frac{P(\widehat{sgn}'_j(\mathbf{z}) \neq \mathrm{sgn}(\theta_j)|\mathbf{z})}{1 \vee |\mathcal{R}'(\mathbf{z})|} \right] P(\mathbf{z}) d\mathbf{z} \\
&\qquad + \int_{\mathcal{Z}_j^{(2)}} \left[\frac{P(\widehat{sgn}_j(\mathbf{z}) \neq \mathrm{sgn}(\theta_j)|\mathbf{z})}{1 \vee |\mathcal{R}(\mathbf{z})|} - \frac{P(\widehat{sgn}'_j(\mathbf{z}) \neq \mathrm{sgn}(\theta_j)|\mathbf{z})}{1 \vee |\mathcal{R}'(\mathbf{z})|} \right] P(\mathbf{z}) d\mathbf{z} \\
&\quad = \int_{\mathcal{Z}_j^{(1)}} \left[\frac{P(\theta_j \geqslant 0|z_j)}{1 \vee |\mathcal{R}(\mathbf{z})|} - \frac{P(\theta_j \leqslant 0|z_j)}{1 \vee |\mathcal{R}'(\mathbf{z})|} \right] P(\mathbf{z}) d\mathbf{z} \\
&\qquad + \int_{\mathcal{Z}_j^{(2)}} \left[\frac{P(\theta_j \leqslant 0|z_j)}{1 \vee |\mathcal{R}(\mathbf{z})|} - \frac{P(\theta_j \geqslant 0|z_j)}{1 \vee |\mathcal{R}'(\mathbf{z})|} \right] P(\mathbf{z}) d\mathbf{z} \geqslant 0.
\end{aligned}
$$

so $(\widehat{sgn}'_i)_{i \in \mathcal{R}'}$ improves upon $(\widehat{sgn}_i)_{i \in \mathcal{R}}$. □

Proof of Statement 2. The proof for Statement 2 follows in a similar vein. We first define

$$
\mathcal{Z}_{jl}^{(1)} \equiv \left\{ \mathbf{z} \;\middle|\; \begin{array}{l} \mathbf{z} \in \mathcal{Z}_{jl}; \\ P(\theta_j \leqslant 0|z_j) < P(\theta_j \geqslant 0|z_j). \end{array} \right\} \quad \text{and}
$$

$$
\mathcal{Z}_{jl}^{(2)} \equiv \left\{ \mathbf{z} \;\middle|\; \begin{array}{l} \mathbf{z} \in \mathcal{Z}_{jl}; \\ P(\theta_j \leqslant 0|z_j) > P(\theta_j \geqslant 0|z_j). \end{array} \right\}.
$$

Since $\mathcal{R}'$ has been defined, we only have to define $\widehat{sgn}_i'$ for each $i \in \mathcal{R}'$, as

$$\widehat{sgn}_i'(\mathbf{z}) = \begin{cases} 1 & \text{if } i = j \text{ and } \mathbf{z} \in \mathcal{Z}_{jl}^{(1)} \\ -1 & \text{if } i = j \text{ and } \mathbf{z} \in \mathcal{Z}_{jl}^{(2)} \\ \widehat{sgn}_i(\mathbf{z}) & \text{if otherwise} \end{cases}.$$

One can then write

$$\begin{aligned} &\text{ETD}[(\widehat{sgn}_i')_{i\in\mathcal{R}'}] - \text{ETD}[(\widehat{sgn}_i)_{i\in\mathcal{R}}] \\ &= \int_{\mathcal{Z}_{jl}^{(1)}} [P(\widehat{sgn}_j'(\mathbf{z}) = \text{sgn}(\theta_j)|\mathbf{z}) - P(\widehat{sgn}_l(\mathbf{z}) = \text{sgn}(\theta_l)|\mathbf{z})]P(\mathbf{z})d\mathbf{z} \\ &\quad + \int_{\mathcal{Z}_{jl}^{(2)}} [P(\widehat{sgn}_j'(\mathbf{z}) = \text{sgn}(\theta_j)|\mathbf{z}) - P(\widehat{sgn}_l(\mathbf{z}) = \text{sgn}(\theta_l)|\mathbf{z})]P(\mathbf{z})d\mathbf{z} \\ &= \int_{\mathcal{Z}_{jl}^{(1)}} [P(\theta_j > 0|\mathbf{z}) - P(\widehat{sgn}_l(\mathbf{z}) = \text{sgn}(\theta_l)|\mathbf{z})]P(\mathbf{z})d\mathbf{z} \quad \text{(B.5)} \\ &\quad + \int_{\mathcal{Z}_{jl}^{(2)}} [P(\theta_j < 0|\mathbf{z}) - P(\widehat{sgn}_l(\mathbf{z}) = \text{sgn}(\theta_l)|\mathbf{z})]P(\mathbf{z})d\mathbf{z}. \quad \text{(B.6)} \end{aligned}$$

by a similar train of equalities as in the proof for Statement 1. For $\mathbf{z} \in \mathcal{Z}_{jl}^{(1)}$, $P(\theta_j \leqslant 0|\mathbf{z}) = \mathit{lfsr}_j < \mathit{lfsr}_l \leqslant P(\widehat{sgn}_l(\mathbf{z}) \neq \text{sgn}(\theta_l)|\mathbf{z})$, where the last inequality comes from the fact that $\mathit{lfsr}_l = P(\theta_l \leqslant 0|\mathbf{z}) \wedge P(\theta_l \geqslant 0|\mathbf{z})$ is the smallest conditional probability of making a false discovery that can possibly be achieved by $\widehat{sgn}_l(\mathbf{z})$. This in turns implies $P(\theta_j > 0|\mathbf{z}) > P(\widehat{sgn}_l(\mathbf{z}) = \text{sgn}(\theta_l)|\mathbf{z})$, which means that the term in (B.5) is greater than 0. Similarly one can show that the term in (B.6) is also greater than 0, which gives $\text{ETD}[(\widehat{sgn}_i')_{i\in\mathcal{R}'}] - \text{ETD}[(\widehat{sgn}_i)_{i\in\mathcal{R}}] > 0$.

We can also show that $\text{FDR}_{\text{dir}}[(\widehat{sgn}_i')_{i\in\mathcal{R}'}] \leqslant \text{FDR}_{\text{dir}}[(\widehat{sgn}_i)_{i\in\mathcal{R}}]$ similarly; to avoid repetitions, we leave the details to the reader. □

Appendix C: Additional content for Section 4

C.1. Implementation of the LSFR procedure

An exact implementation of the ODP described in Theorem 3.2 involves solving a rather complex infinite integer programming problem (Heller and Rosset, 2021) to determine a threshold for the local false sign rates. As an alternative, in Section 4, LFSR is a similar oracle procedure with an attractively simpler implementation first suggested by Sun and Cai (2007), and it suffices to serve as an oracle benchmark for the power of our compared procedures. Suppose we denote this procedure as $(\widehat{sgn}_i^{SC})_{i\in\mathcal{R}_{SC}}$ in our notation. Then the signs $\widehat{sgn}_i^{SC}$ are declared as in Table 1, i.e., $\widehat{sgn}_i^{SC} = \widehat{sgn}_i^{ODP}$, and the discovery set is defined by

$$\mathcal{R}_{SC} \equiv \{i : \mathit{lfsr}_i \leqslant \mathit{lfsr}_{(j)}\}$$

with the index

$$j = j(q) \equiv \max \left\{ i' \in [m] : \frac{\sum_{i=1}^{i'} \mathit{lfsr}_{(i)}}{i'} \leqslant q \right\},$$

where $\mathit{lfsr}_{(1)} \leqslant \cdots \leqslant \mathit{lfsr}_{(m)}$ are the order statistics of true local false sign rates, and $\mathcal{R}_{SC}$ is the empty set if j is not well-defined. The ratio $\frac{\sum_{i=1}^{i'} \mathit{lfsr}_{(i)}}{i'}$ in the definition of j is precisely the conditional $\mathrm{FDR}_{\mathrm{dir}}$

$$\mathbb{E}\left[\frac{|\{i : \widehat{sgn}_i^{SC} \neq \mathrm{sgn}(\theta_i) \text{ and } \mathit{lfsr}_i \leqslant \mathit{lfsr}_{(i')}\}|}{i'} \Big| \{z_i\}_{i \in [m]} \right]$$

of optimally declaring the signs for the subset $\{i : \mathit{lfsr}_i \leqslant \mathit{lfsr}_{(i')}\}$ given the data, which also implies the $\mathrm{FDR}_{\mathrm{dir}}$ of $(\widehat{sgn}_i^{SC})_{i \in \mathcal{R}_{SC}}$,

$$\begin{aligned} \mathbb{E}\left[\frac{\widehat{sgn}_i^{SC} \neq \mathrm{sgn}(\theta_i)}{1 \vee |\mathcal{R}_{SC}|} \right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{\widehat{sgn}_i^{SC} \neq \mathrm{sgn}(\theta_i)}{1 \vee |\mathcal{R}_{SC}|} \Big| \{z_i\}_{i \in [m]} \right] \right] \\ &= \mathbb{E}\left[\frac{\sum_{i=1}^{j} \mathit{lfsr}_{(i)}}{j} \Big| \mathcal{R}_{SC} \neq \varnothing \right] P(\mathcal{R}_{SC} \neq \varnothing), \end{aligned}$$

is less than q by how j was defined, under the Bayesian formulation (3.8).

C.2. Other simulation results

An R package for our methods is available at `https://github.com/ninhtran02/zdirect`. We also conducted additional simulation studies to evaluate the performance of different methods under dependent z-values. Specifically, given $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ generated exactly as described in Section 4.1 using the same simulation parameters, we generate $\mathbf{z} = (z_1, \ldots, z_m)$ with a multivariate normal distribution $N(\boldsymbol{\theta}, \Sigma)$ and an autoregressive covariance structure $\Sigma_{ij} = \rho^{|i-j|}$ for $1 \leqslant i, j, \leqslant m$. The following values of ρ are experimented with:

C.2.1. Positive autoregressive dependence

Weak and strong positive dependence with $\rho = 0.5$ and $\rho = 0.8$.

C.2.2. Negative autoregressive dependence

Weak and strong negative dependence with $\rho = -0.5$ and $\rho = -0.8$.

C.2.3. Brief summary

The performances of different methods are included in Figures C.1 and C.2 for the positively dependent settings in Appendix C.2.1, and Figures C.3 and C.4 for the negatively dependent settings in Appendix C.2.2. Note that, in addition to the existing methods from Section 4.2, we have included an extra method called "$\mathrm{dBH}_{\mathrm{dir}}$" in our results. The term "dBH" refers to the *dependence-adjusted BH procedure*, a recent advancement in FDR testing under arbitrary dependence proposed by Fithian and Lei (2022). It serves as a theoretically valid and more powerful alternative to the widely recognized but very conservative BY procedure (Benjamini and Yekutieli, 2001). We note that dBH requires prior knowledge of the underlying dependence structure, distinguishing it from the BY procedure. $\mathrm{dBH}_{\mathrm{dir}}$ is a variant of dBH designed specifically for multivariate normal z-values and $\mathrm{FDR}_{\mathrm{dir}}$ control. It is implemented through the function `dBH_mvgauss` in the R package `dbh`, where the `gamma` parameter is set to 0.9 following the recommendation by Fithian and Lei (2022) for two-sided testing. Its exact $\mathrm{FDR}_{\mathrm{dir}}$ control for our settings with dependent z-values is established by Fithian and Lei (2022, Corollary 7).

The $\mathrm{FDR}_{\mathrm{dir}}$ and power of each method under the dependent settings outlined in Appendix C.2.1 and Appendix C.2.2 are similar to those under the independent setting outlined in Section 4. Nevertheless, subtle differences emerge. For strong autoregressive dependence $\rho \in \{-0.8, 0.8\}$, when $w = 0.8$ and $\xi \in \{0.5, 1\}$, the methods $\mathrm{STS}_{\mathrm{dir}}$, $\mathrm{aSTS}_{\mathrm{dir}}$, $\mathrm{BH}_{\mathrm{dir}}$, LFSR, GR and $\mathrm{dBH}_{\mathrm{dir}}$ exhibited slight $\mathrm{FDR}_{\mathrm{dir}}$ decreases by approximately 0.01 to 0.02. Conversely, ZDIRECT displayed slight $\mathrm{FDR}_{\mathrm{dir}}$ increases by approximately 0.01. Despite these increases in $\mathrm{FDR}_{\mathrm{dir}}$ under strong autoregressive dependence, ZDIRECT consistently maintained empirical control of $\mathrm{FDR}_{\mathrm{dir}}$ below the designated level of $q = 0.10$ throughout our additional simulations.

We observed that $\mathrm{dBH}_{\mathrm{dir}}$ is slightly more conservative in $\mathrm{FDR}_{\mathrm{dir}}$ control and less powerful than $\mathrm{BH}_{\mathrm{dir}}$ across our additional simulations. This difference in performance may be attributed to the recommended `gamma` parameter choice of 0.9 by Fithian and Lei (2022), chosen to reduce the likelihood of obtaining a randomly "pruned" rejection set; see Fithian and Lei (2022, Section 2.2) for an explanation of why it is preferable to avoid the randomized pruning step built into dBH-type methods. This cautious parameter choice may compromise any potential power advantage $\mathrm{dBH}_{\mathrm{dir}}$ could have over $\mathrm{BH}_{\mathrm{dir}}$ in the presence of autoregressive dependence.

Acknowledgments

The authors would like to thank the anonymous referees and an Associate Editor for their constructive comments that improved the quality of this paper.

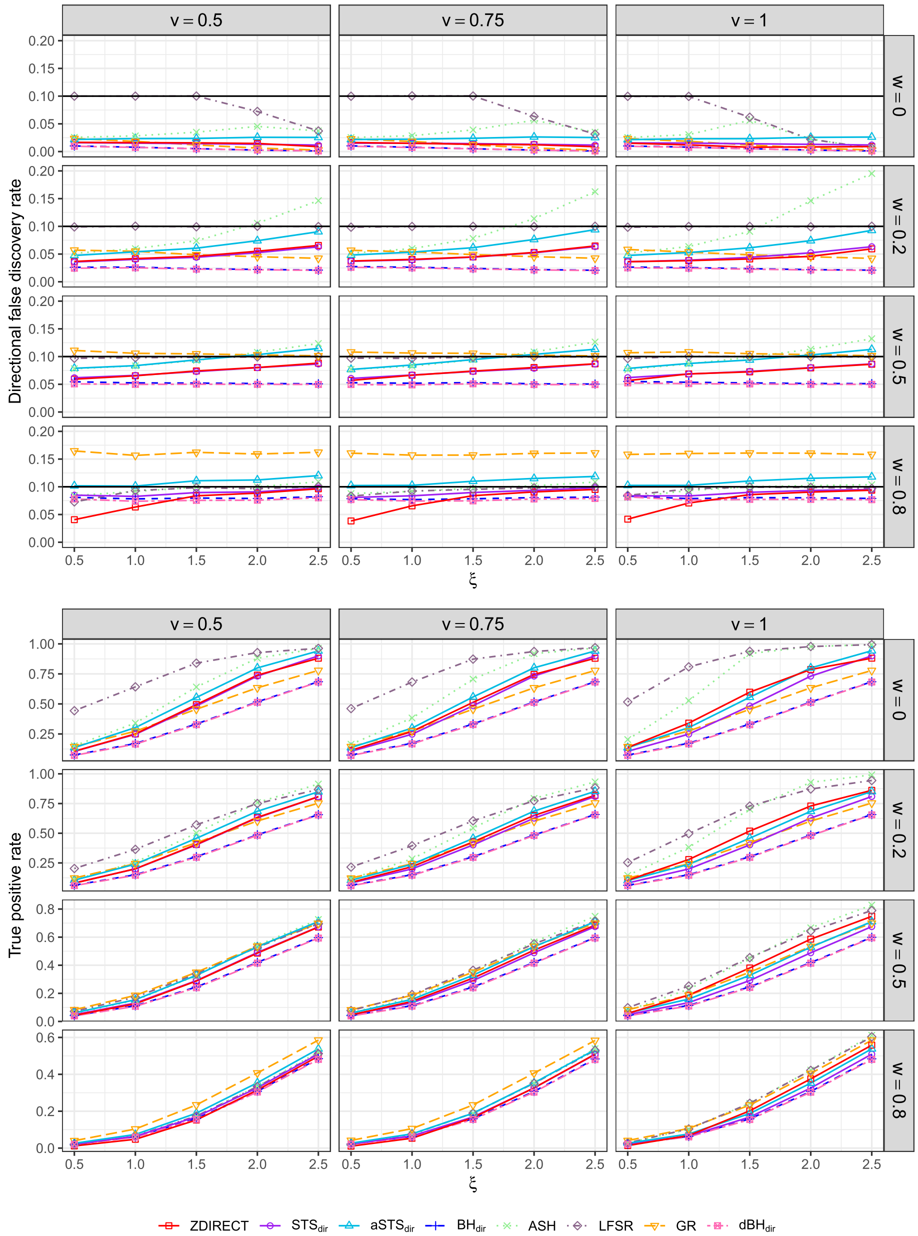

FIG C.1. *Empirical directional false discovery rate and true positive rates of the eight compared methods for the simulations in Appendix C.2.1 with $\rho = 0.5$; each method was implemented at a target FDR_{dir} level $q = 0.1$ (black horizontal lines).*

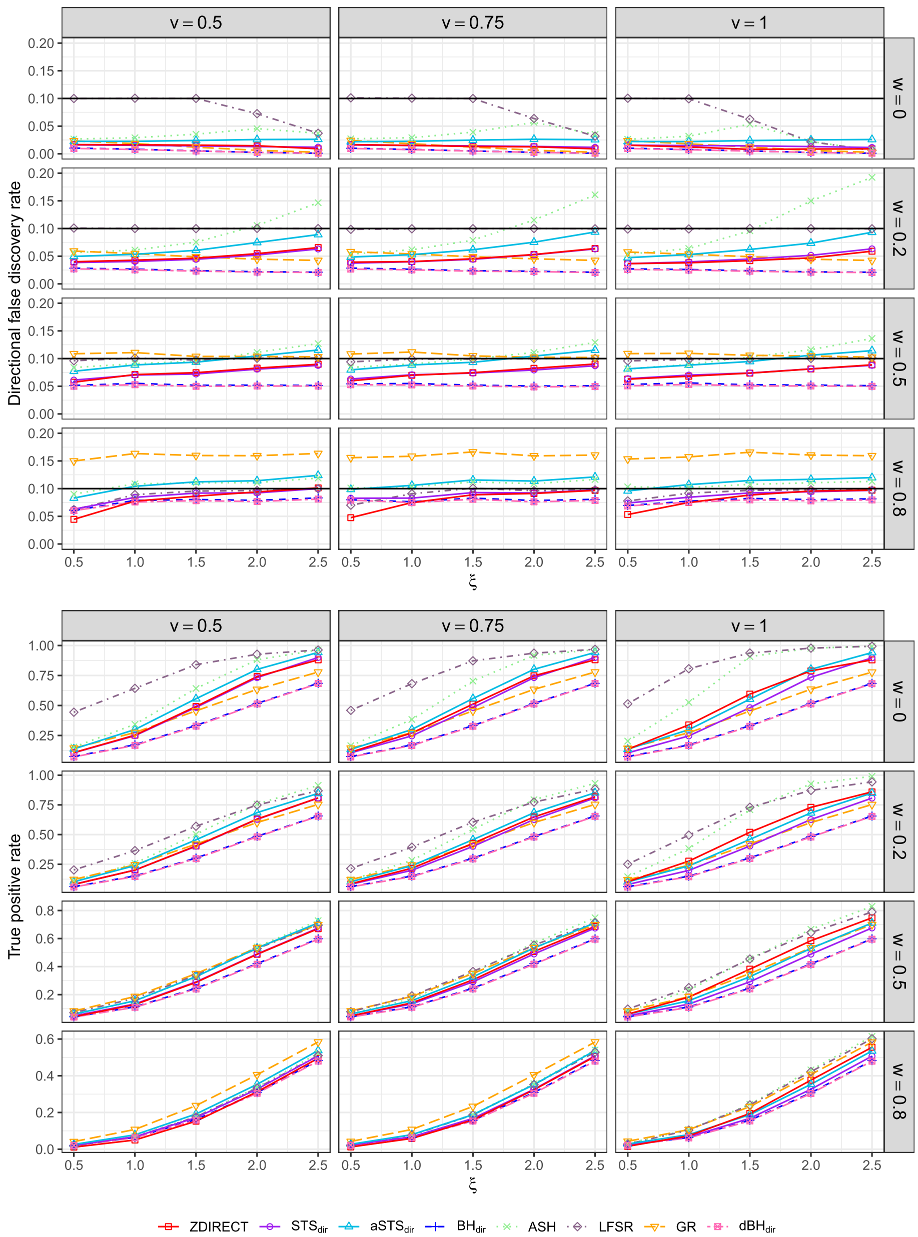

FIG C.2. *Empirical directional false discovery rate and true positive rates of the eight compared methods for the simulations in Appendix C.2.1 with* $\rho = 0.8$*; each method was implemented at a target* FDR_{dir} *level* $q = 0.1$ *(black horizontal lines).*

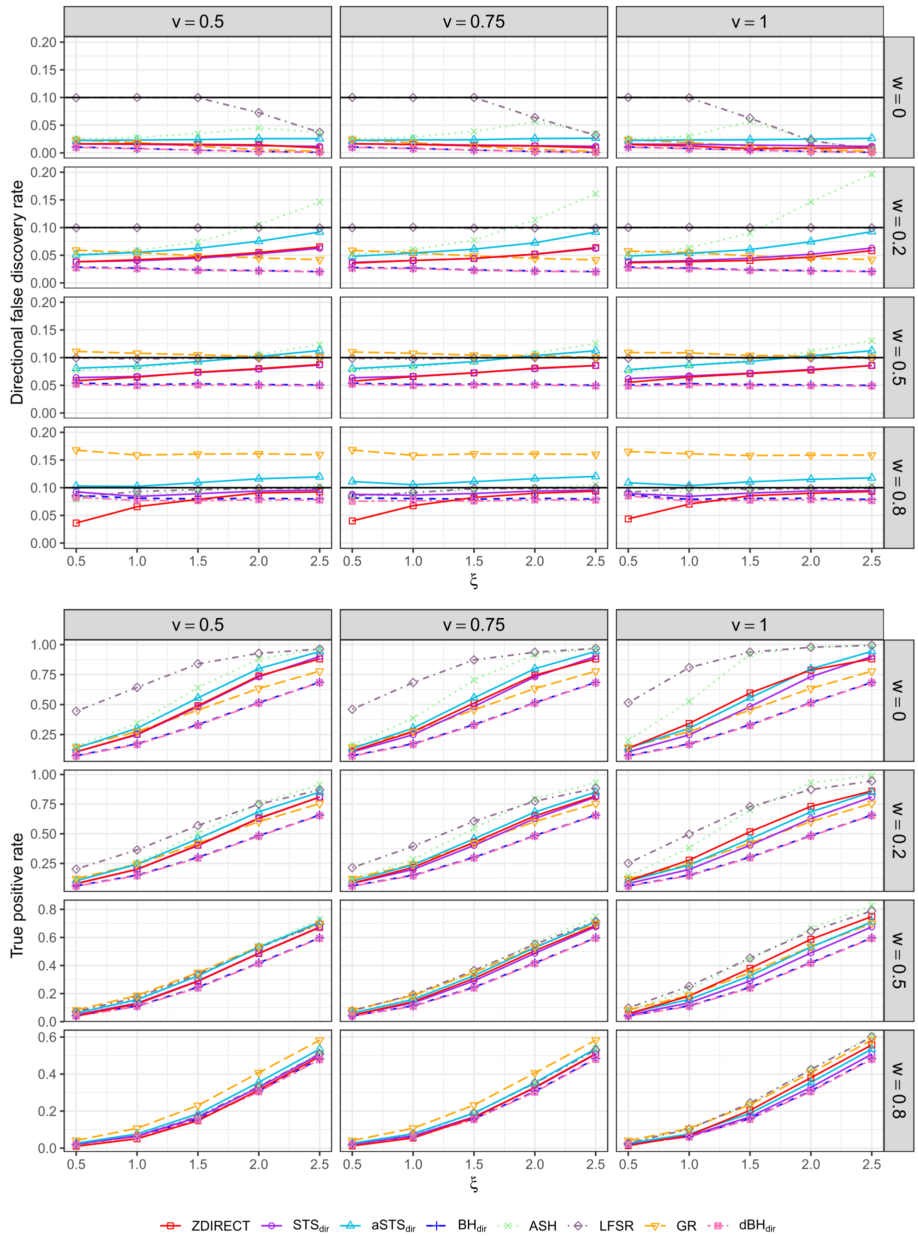

Fig C.3. *Empirical directional false discovery rate and true positive rates of the eight compared methods for the simulations in Appendix C.2.2 with* $\rho = -0.5$*; each method was implemented at a target* FDR_{dir} *level* $q = 0.1$ *(black horizontal lines).*

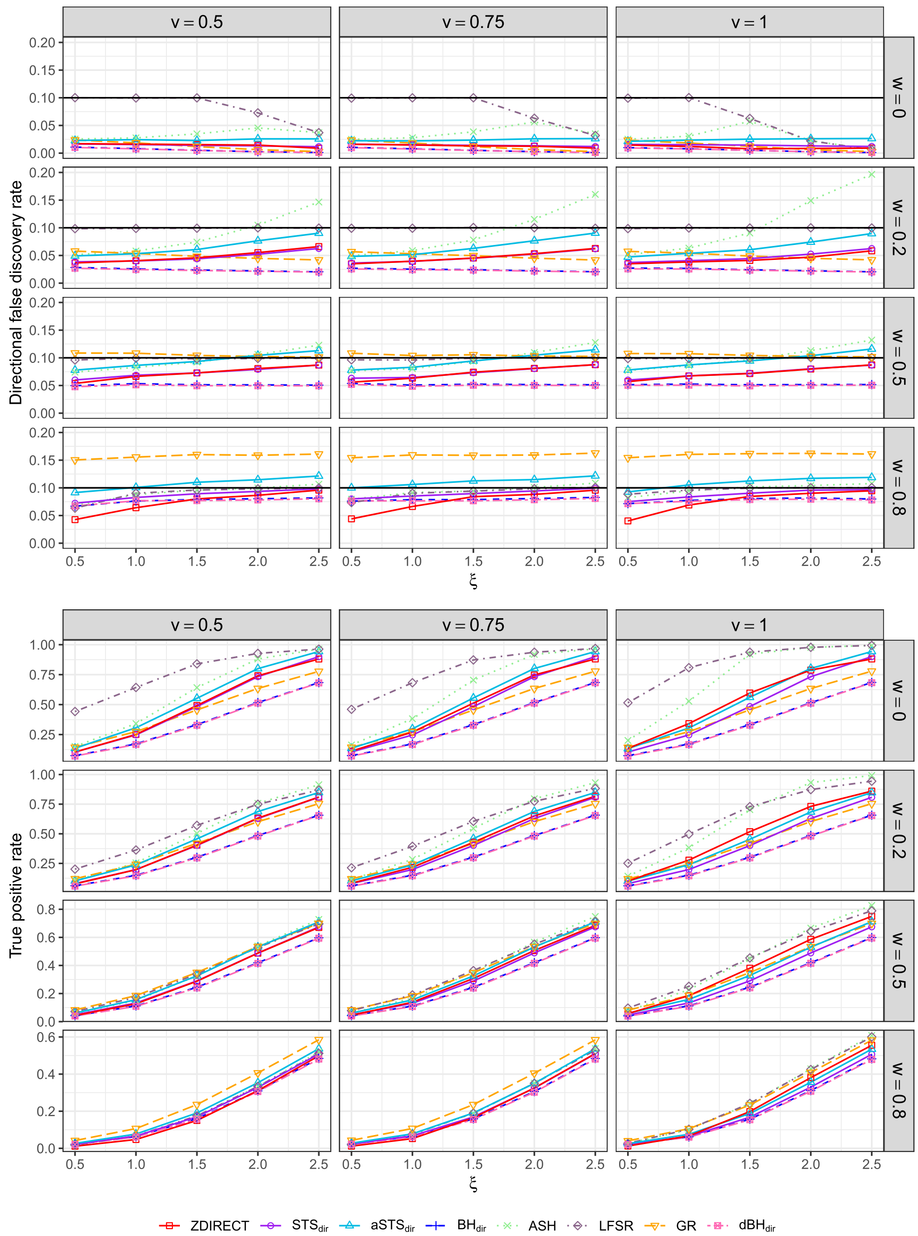

Fig C.4. *Empirical directional false discovery rate and true positive rates of the eight compared methods for the simulations in Appendix C.2.2 with* $\rho = -0.8$*; each method was implemented at a target* FDR_{dir} *level* $q = 0.1$ *(black horizontal lines).*

References

ApS, M. (2022). The MOSEK Rmosek package. Version 9.3.20.

Barber, R. F., Candès, E. J. et al. (2015). Controlling the false discovery rate via knockoffs. *The Annals of Statistics* **43** 2055–2085. MR3375876

Barber, R. F., Candès, E. J. et al. (2019). A knockoff filter for high-dimensional selective inference. *The Annals of Statistics* **47** 2504–2537. MR3988764

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57** 289–300. MR1325392

Benjamini, Y. and Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics* **25** 60–83. MR1869245

Benjamini, Y., Krieger, A. M. and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93** 491–507. MR2261438

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* 1165–1188. MR1869245

Benjamini, Y. and Yekutieli, D. (2005). False discovery rate–adjusted multiple confidence intervals for selected parameters. *Journal of the American Statistical Association* **100** 71–81. MR2156820

Chao, P. and Fithian, W. (2021). AdaPT-GMM: Powerful and robust covariate-assisted multiple testing. *arXiv preprint arXiv:2106.15812*.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* **39** 1–22. MR0501537

Feller, W. (1971). *An introduction to probability theory and its applications. Vol. II*, Second ed. John Wiley & Sons, Inc., New York-London-Sydney. MR0270403

Fithian, W. and Lei, L. (2022). Conditional calibration for false discovery rate control under dependence. *Ann. Statist.* **50** 3091–3118. MR4524490

Gelman, A. and Tuerlinckx, F. (2000). Type S error rates for classical and Bayesian single and multiple comparison procedures. *Computational Statistics* **15** 373–390.

Guo, W. and Romano, J. P. (2015). On stepwise control of directional errors under independence and some dependence. *Journal of Statistical Planning and Inference* **163** 21–33. MR3339038

Guo, W., Sarkar, S. K. and Peddada, S. D. (2010). Controlling false discoveries in multidimensional directional decisions, with applications to gene expression data on ordered categories. *Biometrics* **66** 485–492. MR2758828

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** 133–155. MR4220987

Kraemer, H. C. and Paik, M. (1979). A central t approximation to the noncentral t distribution. *Technometrics* **21** 357–360. MR0538440

Kruskal, W. (1954). The Monotonicity of the Ratio of Two Noncentral t Density Functions. *Journal of Educational and Behavioral Statistics* **25** 162–165. MR0061320

Laubscher, N. F. (1960). Normalizing the Noncentral t and F Distributions. *The Annals of Mathematical Statistics* **31** 1105–1112. MR0117815

Lehmann, E. L., Romano, J. P. and Casella, G. (2005). *Testing statistical hypotheses* **3**. Springer. MR2135927

Lei, L. and Fithian, W. (2018). AdaPT: an interactive procedure for multiple testing with side information. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80** 649–679. MR3849338

Lei, L., Ramdas, A. and Fithian, W. (2021). A general interactive framework for false discovery rate control under structural constraints. *Biometrika* **108** 253–267. MR4259130

Leung, D. and Sun, W. (2022). ZAP: Z-value Adaptive Procedures for False Discovery Rate Control with Side Information. *Journal of the Royal Statistical Society Series B* **84** 1886–1946. MR4515561

Stephens, M. (2017). False discovery rates: a new deal. *Biostatistics* **18** 275–294. MR3824755

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** 414–432. MR2323757

Storey, J. D., Taylor, J. E. and Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66** 187–205. MR2035766

Sun, W. and Cai, T. T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *Journal of the American Statistical Association* **102** 901–912. MR2411657

Tian, Z., Liang, K. and Li, P. (2021). A powerful procedure that controls the false discovery rate with directional information. *Biometrics* **77** 212–222. MR4229733

Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics* **33** 1–67. MR0133937

Tukey, J. W. (1991). The philosophy of multiple comparisons. *Statistical Science* 100–116.

Weinstein, A. and Yekutieli, D. (2020). Selective sign-determining multiple confidence intervals with FCR control. *Statistica Sinica* **30** 531–555. MR4285505

Yurko, R., G'Sell, M., Roeder, K. and Devlin, B. (2020). A selective inference approach for false discovery rate control using multiomics covariates yields insights into disease risk. *Proceedings of the National Academy of Sciences* **117** 15028–15035.

Zhao, H. and Fung, W. K. (2018). Controlling mixed directional false discovery rate in multidimensional decisions with applications to microarray studies. *Test* **27** 316–337. MR3799142