# AN ADAPTIVE PROCEDURE FOR DETECTING REPLICATED SIGNALS WITH k-FAMILY-WISE ERROR RATE CONTROL

#### NINH TRAN

ABSTRACT. Partial conjunction (PC) hypothesis testing is widely used to assess the replicability of scientific findings across multiple comparable studies. In high-throughput meta-analyses, testing a large number of PC hypotheses with k-family-wise error rate (k-FWER) control often suffers from low statistical power due to the multiplicity burden. The state-of-the-art AdaFilter-Bon procedure by Wang et al. (2022, Ann. Stat., 50(4), 1890-1909) alleviates this problem by filtering out hypotheses unlikely to be false before applying a rejection rule. However, a side effect of filtering is that it renders the rejection rule more stringent than necessary, leading to conservative k-FWER control. In this paper, we mitigate this conservativeness—and thereby improve the power of AdaFilter-Bon—by incorporating a post-filter null proportion estimate into the procedure. The resulting method, AdaFilter-AdaBon, has proven asymptotic k-FWER control under weak dependence and demonstrates empirical finite-sample control with higher power than the original AdaFilter-Bon in simulations.

#### 1. Introduction

Over the last two decades, the partial conjunction (PC) hypothesis testing (Benjamini and Heller, 2008) has become a standard statistical tool for assessing replicability in meta-analyses involving  $n \geq 2$  comparable studies. Within this framework, a scientifically interesting finding (i.e., a signal) is deemed credible if it is exhibited in at least u of the n studies, a notion referred to as u/n replication. The value of  $u \in \{2, ..., n\}$  is freely chosen by the meta-analyst and is often guided by conventions within the field. For example, u = 2 is widely regarded as the gold standard in drug assessment (Zhan et al., 2022).

Identifying u/n replication requires testing the PC null hypothesis that fewer than u of the n studies exhibit the signal. A p-value for testing this null hypothesis is typically obtained by combining the p-values from the n individual studies using the general methodology of Benjamini and Heller (2008), which was later extended by Wang and Owen (2019). Testing many PC nulls with combined p-values under family-wise error rate (FWER) control often suffers from low power due to the multiplicity correction. This issue is especially acute in high-throughput meta-analyses of gene expression, where the number of PC nulls can run into the hundreds or thousands. Wang et al. (2022) address this problem by proposing **AdaFilter-Bon**, a state-of-the-art FWER method

 $<sup>2020\</sup> Mathematics\ Subject\ Classification.\ 62F03.$ 

Key words and phrases. k-FWER, multiple testing, partial conjunction, adaptive, filtering.

This research was supported by The University of Melbourne's Research Computing Services, the Petascale Campus Initiative, and the Australian Government Research Training Program (RTP) Scholarship.

that uses a p-value-adaptive filter to remove PC nulls that are likely to be true, before applying a rejection rule similar to the **Bon**ferroni procedure.

In Section 2.2, we provide an overview of AdaFilter-Bon and show that, with a simple modification, AdaFilter-Bon can control the k-FWER—a generalization of the FWER<sup>1</sup>—for any integer  $k \in \mathbb{N}$ . While filtering reduces the multiplicity burden and increases power, it can also result in overly conservative k-FWER control. For a target k-FWER level  $\alpha \in (0,1)$ , we show in Section 2.3 that AdaFilter-Bon controls the k-FWER at  $\alpha$  multiplied by the expected proportion of true PC nulls among those remaining after filtering. This proportion, which lies in [0,1], decreases as more true PC nulls are filtered out. Consequently, when the filter operates ideally—removing many PC nulls likely to be true—AdaFilter-Bon's k-FWER level can drop well below  $\alpha$ .

In this paper, we introduce AdaFilter-AdaBon, a multiple testing procedure based on AdaFilter-Bon that utilises a p-value-adaptive estimator of the post-filter proportion of true PC nulls to offset the conservative k-FWER control induced by filtering. Under mild assumptions that allow for weak dependence among the data, this new procedure has theoretical asymptotic k-FWER control. In simulation studies, we show that AdaFilter-AdaBon maintains robust finite-sample k-FWER control and yields more true positive discoveries than the original AdaFilter-Bon.

To the best of our knowledge, this is the first work to employ an estimate of the post-filter proportion of true PC nulls for controlling the k-FWER. In particular, we were unable to find existing research addressing k-FWER control in replicability analyses, and found only limited efforts to estimate the post-filter proportion of true PC nulls (Bogomolov and Heller, 2018, Dickhaus et al., 2024, Tran and Leung, 2025). Even in single-study multiple testing, research on k-FWER control has been modest, with notable contributions from Lehmann and Romano (2005), Romano and Shaikh (2006), Sarkar (2007, 2008), none of which incorporate null proportion estimation.

The structure of the paper is as follows. In Section 2, we present the problem formulation, overview the AdaFilter-Bon procedure, and outline the theory behind its conservative k-FWER control. In Section 3, we introduce the new method, AdaFilter-AdaBon, and establish its asymptotic k-FWER control guarantees. Section 4 reports a simulation study comparing the performance of AdaFilter-AdaBon with AdaFilter-Bon and other related methods. Finally, Section 5 concludes the paper by highlighting extensions of AdaFilter-AdaBon and open challenges for future work.

### 2. Problem formulation

2.1. **Problem statement.** Suppose we are conducting a meta-analysis that consists of  $n \geq 2$  comparable studies, each of which examines the same set of m > 1 features. For each feature  $i \in \{1, ..., m\}$  of study  $j \in \{1, ..., n\}$ , suppose there is a null hypothesis  $H_{ij}$  for which we observe a corresponding valid p-value  $P_{ij}$ , i.e.,

$$\Pr(P_{ij} \leq t) \leq t$$
 for all  $t \in [0,1]$  if  $H_{ij}$  is true.

Since a meta-analysis typically consists of independently conducted studies, we assume that the vectors  $(P_{i1})_{i=1}^m, \ldots, (P_{in})_{i=1}^m$  are independent.

For a given replicability level  $u \in \{2, ..., n\}$ , feature i is said to be u/n replicated if at least u of the n hypotheses among  $H_{i1}, ..., H_{in}$  are false. We consider a problem where the objective is to

<sup>&</sup>lt;sup>1</sup>When k = 1, the k-FWER is equivalent to the FWER.

identify u/n replicated features by testing the partial conjunction (PC) null hypothesis (Benjamini and Heller, 2008):

$$H_i^{u/n}$$
: Less than u null hypotheses among  $H_{i1}, \ldots, H_{in}$  are false,

for i = 1, ..., m. This multiple testing problem is often referred to as a *replicability analysis*. Note that whenever  $u'' \ge u'$ , the following nesting property holds:

(2.1) 
$$H_i^{u'/n}$$
 is true  $\implies H_i^{u''/n}$  is true.

Let  $\mathcal{R} \subseteq \{1, ..., m\}$  be the set of PC null hypotheses rejected by a multiple testing procedure. For any given tolerance level  $k \in \mathbb{N}$ , the k-family-wise error rate (k-FWER) of  $\mathcal{R}$  is defined as

$$k ext{-FWER} \equiv \Pr\left(\sum_{i=1}^{m} I\{i \in \mathcal{R}\}I\{H_i^{u/n} \text{ is true}\} \ge k\right),$$

which is the probability that the number of false discoveries is greater than or equal to k. When k = 1, the k-FWER is simply referred to as the FWER. The power of  $\mathcal{R}$  can be evaluated using the true positive rate (TPR),

$$\text{TPR} \equiv \mathbb{E}\left[\frac{\sum_{i=1}^{m} I\{i \in \mathcal{R}\} I\{H_i^{u/n} \text{ is false}\}}{1 \vee \sum_{i=1}^{m} I\{H_i^{u/n} \text{ is false}\}}\right],$$

which is the expected proportion of true discoveries among the false PC nulls. Our goal is to develop a powerful multiple testing procedure for replicability analysis that operates on

$$(2.2) (P_{ij})_{m \times n} \equiv (P_{ij})_{(i,j) \in \{1,\dots,m\} \times \{1,\dots,n\}},$$

and controls the k-FWER below a pre-specified target level  $\alpha \in (0,1)$ .

2.2. k-FWER procedures for replicability analysis. A valid p-value for testing  $H_i^{u/n}$ —referred to as a PC p-value and denoted as  $P_i^{u/n}$ —can be constructed by combining the p-values associated with feature i. That is,

$$P_i^{u/n} \equiv f(P_{i1}, P_{i2}, \dots, P_{in}; u),$$

where  $f:[0,1]^n \to [0,1]$  is a combining function satisfying

$$\Pr(f(P_{i1}, P_{i2}, \dots, P_{in}) \le t) \le t$$
 for all  $t \in [0, 1]$  if  $H_i^{u/n}$  is true.

Benjamini and Heller (2008) proposed a general methodology for constructing a valid combined p-value for testing  $H_i^{u/n}$ , which Wang and Owen (2019) later extended. An example of a  $P_i^{u/n}$  from this methodology is the Bonferroni-combined PC p-value,

$$f_{\text{Bon}}(P_{i1}, \dots, P_{in}; u) = (n - u + 1)P_{i(u)},$$

where  $P_{i(1)} \leq \cdots \leq P_{i(n)}$  are the order statistics of  $P_{i1}, \ldots, P_{in}$ . Further examples and their numerical performance can be found in Hoang and Dickhaus (2022).

Given the problem statement in Section 2.1, the k-FWER can be controlled by applying a multiple testing procedure on the PC p-values  $(P_i^{u/n})_{i=1}^m$ . Arguably, the simplest k-FWER method is the (generalised) Bonferroni procedure introduced in Lehmann and Romano (2005):

**Definition 1** (Bonferroni). Let  $\alpha \in [0,1]$  be a k-FWER target for a given  $k \in \mathbb{N}$ , and  $(P_i^{u/n})_{i=1}^m$  be PC p-values. For  $i=1,\ldots,m$ , reject  $H_i^{u/n}$  if

$$(2.3) P_i^{u/n} \le k \cdot \frac{\alpha}{m}.$$

Although less powerful than the subsequent methods developed by Romano and Shaikh (2006), Sarkar (2007, 2008), the Bonferroni procedure provides a useful illustration of the multiplicity burden: as the number of PC nulls m grows, the rejection threshold in (2.3) becomes increasingly stringent.

The state-of-the-art AdaFilter-Bon procedure (Wang et al., 2022) reduces the multiplicity burden by using a *filter* to eliminate features unlikely to be u/n replicated, before applying a rejection rule. Although originally developed for FWER control, we describe below a generalisation of AdaFilter-Bon that controls k-FWER control for any  $k \in \mathbb{N}^2$ .

**Definition 2** (AdaFilter-Bon). Let  $\alpha \in [0,1]$  be a k-FWER target for a given  $k \in \mathbb{N}$ , and  $u \in \{2,\ldots,m\}$  be a replicability level. For  $i=1,\ldots,m$ , let

$$(2.4) S_i \equiv f_{Bon}(P_{i1}, \dots, P_{in}; u)$$

be a Bonferroni-combined PC p-value, and let

(2.5) 
$$F_i \equiv \frac{n-u+1}{n-u+2} \cdot f_{Bon}(P_{i1}, \dots, P_{in}; u-1) = (n-u+1)P_{i(u-1)}$$

be a corresponding filtering p-value. For each i, reject  $H_i^{u/n}$  if  $S_i < \hat{t}$ , where

(2.6) 
$$\hat{t} \equiv \sup \left\{ t \in [0, k\alpha] : t \cdot \sum_{i=1}^{m} I\{F_i < t\} \le k\alpha \right\}.$$

Since  $F_i \leq S_i$ , it follows from Definition 2 that  $H_i^{u/n}$  is filtered out (i.e. cannot be rejected) if  $F_i \in [\hat{t}, 1]$ . This is a sensible filtering approach, since  $F_i$  is closely related to  $f_{\text{Bon}}(P_{i1}, \dots, P_{in}; u-1)$ , which is a valid PC p-value for testing  $H_i^{(u-1)/n}$ . Consequently,  $F_i \in [\hat{t}, 1]$  indicates that  $H_i^{(u-1)/n}$  is likely true, making feature i unlikely to be u/n replicated by the nesting property in (2.1).

We also observe that AdaFilter-Bon can only reject  $H_i^{u/n}$  if

(2.7) 
$$S_i < \hat{t} \le k \cdot \frac{\alpha}{\sum_{\ell=1}^m I\{F_{\ell} < \hat{t}\}}.$$

The right-hand side of (2.7) shows that  $\hat{t}$  is inversely proportional to the number of features remaining after filtering,  $\sum_{\ell=1}^{m} I\{F_{\ell} < \hat{t}\}$ , rather than to m. Hence, in comparison to the Bonferroni procedure (Definition 1), there is a reduction in the multiplicity burden due to filtering.

2.3. **Preview of contribution.** By design, features not filtered out by AdaFilter-Bon,  $\{\ell \in \{1,\ldots,m\}: F_{\ell} < \hat{t}\}$ , are more likely to correspond to false PC nulls than those filtered out. Consequently, there is a tendency for the *post-filter null proportion*,

(2.8) 
$$\pi_0(\hat{t}) \equiv \frac{\sum_{i=1}^m I\{F_i < \hat{t}\}I\{H_i^{u/n} \text{ is true}\}}{\sum_{i=1}^m I\{F_i < \hat{t}\}} \in [0, 1],$$

<sup>&</sup>lt;sup>2</sup>When k = 1, Definition 2 recovers the original AdaFilter-Bon procedure of Wang et al. (2022, Definition 3.1).

to be small. This phenomenon, together with the theorem below, reveals that AdaFilter-Bon has a propensity to control the k-FWER conservatively:

**Theorem 2.1** (AdaFilter-Bon k-FWER control). If  $(P_{ij})_{m\times n}$  is a collection of independent valid p-values, then AdaFilter-Bon (Definition 2) has the following k-FWER property:

(2.9) 
$$k\text{-}FWER(\mathcal{R}) \leq \alpha \cdot \mathbb{E}\left[\pi_0(\hat{t})\right] \in [0, \alpha],$$
 where  $\mathcal{R} = \{i \in \{1, \dots, m\} : S_i < \hat{t}\}$  and  $\pi_0(\hat{t})$  is as defined (2.8).

Theorem 2.1 highlights that AdaFilter-Bon's k-FWER level can fall well below  $\alpha$  when  $\hat{\pi}_0(\hat{t})$  is close to zero, i.e., when the filter removes many PC nulls that are unlikely to be false. Hence, filtering effectively comes at the cost of more conservative k-FWER control. The proof of Theorem 2.1 is provided in Appendix A.2; for k=1, the k-FWER upper bound presented in (2.9) is tighter than the one presented in Theorem 4.2 of Wang et al. (2022), who simply only proved that AdaFilter-Bon controls the FWER at level  $\alpha$ .

The thrust of this paper is to estimate  $\pi_0(t)$  for  $t \in [0,1]$  and incorporate it into AdaFilter-Bon's rejection rule to bring its k-FWER control level closer to  $\alpha$ , thereby yielding a more powerful multiple testing procedure. Specifically, we replace the AdaFilter-Bon rejection threshold defined in (2.6) with

(2.10) 
$$\hat{t}_{\theta} \equiv \sup \left\{ t \in [0,1] : \hat{\pi}_0(t) \cdot t \cdot \sum_{i=1}^m I\{F_i < t\} \le k\alpha \right\}$$

instead, where  $\hat{\pi}_0(t) \geq 0$  is a data-adaptive estimate of  $\pi_0(t)$ . Heuristically, we expect the threshold in (2.10) to yield rejections with a k-FWER level bounded above by

$$\alpha \cdot \mathbb{E}\left[\frac{\pi_0(\hat{t}_{\theta})}{\hat{\pi}_0(\hat{t}_{\theta})}\right],$$

rather than the right-hand side of (2.9). Thus, to ensure k-FWER control at level  $\alpha$ ,  $\hat{\pi}_0(\hat{t}_{\theta})$  should ideally satisfy  $\hat{\pi}_0(\hat{t}_{\theta}) \approx \pi_0(\hat{t}_{\theta})$ , or at least  $\hat{\pi}_0(\hat{t}_{\theta}) \gtrsim \pi_0(\hat{t}_{\theta})$ . In light of the data-**ada**ptive post-filter null proportion estimator used in forming (2.10), we refer to this new multiple testing procedure as AdaFilter-**Ada**Bon.

# 3. AdaFilter-AdaBon

3.1. AdaFilter-AdaBon methodology. Let  $\theta \in (0,1)$  be a fixed tuning parameter. By the validity of Bonferroni-combined PC p-values, it holds that  $\Pr(S_i > \theta t) \ge 1 - \theta t$  for all  $t \in [0,1]$  if  $H_i^{u/n}$  is true. Hence,

$$\frac{\sum_{i=1}^{m} I\{F_i < t\} I\{S_i \ge \theta t\}}{\sum_{i=1}^{m} I\{F_i < t\}} \ge \frac{\sum_{i=1}^{m} I\{F_i < t\} I\{H_i^{u/n} \text{ is true}\} I\{S_i \ge \theta t\}}{\sum_{i=1}^{m} I\{F_i < t\}}$$
$$\gtrsim (1 - \theta t) \frac{\sum_{i=1}^{m} I\{F_i < t\} I\{H_i^{u/n} \text{ is true}\}}{\sum_{i=1}^{m} I\{F_i < t\}}.$$

Rearranging the above display yields our estimator for  $\pi_0(t)$ :

(3.1) 
$$\hat{\pi}_0(t) \equiv \hat{\pi}_0(t;\theta) = \frac{\sum_{i=1}^m I\{F_i < t\} I\{S_i \ge \theta t\}}{(1 - \theta t) \sum_{i=1}^m I\{F_i < t\}}.$$

As the value of  $\theta$  increases while t remains constant, the bias of  $\hat{\pi}_0(t)$  decreases and its variance increases. In our experiments, setting  $\theta = 0.5$  generally strikes a good balance in this bias-variance trade-off. Having defined  $\hat{\pi}_0(t)$ , we now formally present the AdaFilter-AdaBon procedure below:

**Definition 3** (AdaFilter-AdaBon). Let  $\alpha \in [0,1]$  be a k-FWER target for a given  $k \in \mathbb{N}$ ,  $u \in \{2,\ldots,m\}$  be a replicability level,  $\theta \in (0,1)$  be a tuning parameter, and  $S_i$  and  $F_i$  be as defined in (2.4) and (2.5), respectively. For  $i=1,\ldots,m$ , reject  $H_i^{u/n}$  if  $S_i < \hat{t}_{\theta}$ , where

(3.2) 
$$\hat{t}_{\theta} \equiv \sup \left\{ t \in [0, 1] : t \cdot \frac{\sum_{i=1}^{m} I\{F_i < t\} I\{S_i \ge \theta t\}}{1 - \theta t} \le k\alpha \right\}.$$

Note that (3.2) is equivalent to the threshold previewed in (2.10) of Section 2.3 when  $\hat{\pi}_0(t)$  is defined as (3.1). To simplify the implementation of AdaFilter-AdaBon in practice, we recommend defining

$$\mathcal{G} = \{G \in \{0, 1\} \cup \{F_i, S_i, S_i/\theta\}_{i=1}^m : G \le 1\}$$

and using the rejection threshold

(3.3) 
$$\check{t}_{\theta} \equiv \max \left\{ t \in \mathcal{G} : t \cdot \frac{\sum_{i=1}^{m} I\{F_i < t\}I\{S_i \ge \theta t\}}{1 - \theta t} \le k\alpha \right\}$$

as a surrogate for  $\hat{t}_{\theta}$ . The exact value of  $\check{t}_{\theta}$  can be computed in O(m) time by evaluating the set-building condition in (3.3) for each element of  $\mathcal{G}$ . Although  $\check{t}_{\theta} \not\equiv \hat{t}_{\theta}$ , the theorem below states that  $\check{t}_{\theta}$  and  $\hat{t}_{\theta}$  yield the same rejection set when thresholding the PC p-values in  $(S_i)_{i=1}^m$ .

**Theorem 3.1** ( $\check{t}_{\theta}$  as a surrogate for  $\hat{t}_{\theta}$ ). Let  $\hat{t}_{\theta}$  and  $\check{t}_{\theta}$  be as defined in (3.2) and (3.3), respectively. It holds that

$$\{i \in \{1, \dots, m\} : S_i < \hat{t}_{\theta}\} = \{i \in \{1, \dots, m\} : S_i < \check{t}_{\theta}\}.$$

The proof of Theorem 3.1, given in Appendix B.2, is based on the fact that  $\sum_{i=1}^{m} I\{F_i < t\}I\{S_i \ge \theta t\}$  and  $\{i \in \{1, ..., m\} : S_i < t\}$  are both step functions in t.

3.2. Asymptotic results for AdaFilter-AdaBon. In this subsection, we provide the result showing that AdaFilter-AdaBon (Definition 3) asymptotically controls the k-FWER at level  $\alpha$  as  $m \to \infty$ . The following assumption underpins our asymptotic result:

**Assumption 1.** For any given m, let

$$m_0 \equiv \sum_{i=1}^m I\{H_i^{u/n} \text{ is true}\}$$
 and  $m_1 \equiv \sum_{i=1}^m I\{H_i^{u/n} \text{ is false}\}$ 

be the number of true and false PC nulls, respectively. The following limits exist almost surely for each  $t \in (0, t]$ :

$$\lim_{m \to \infty} \frac{\sum_{i=1}^{m} I\{S_i < t\} I\{H_i^{u/n} \text{ is true}\}}{m_0} = \tilde{S}_0(t), \quad \lim_{m \to \infty} \frac{\sum_{i=1}^{m} I\{F_i < t\} I\{H_i^{u/n} \text{ is true}\}}{m_0} = \tilde{F}_0(t)$$

$$\lim_{m \to \infty} \frac{\sum_{i=1}^{m} I\{S_i < t\} I\{H_i^{u/n} \text{ is false}\}}{m_1} = \tilde{S}_1(t), \quad \lim_{m \to \infty} \frac{\sum_{i=1}^{m} I\{F_i < t\} I\{H_i^{u/n} \text{ is false}\}}{m_1} = \tilde{F}_1(t),$$

where  $\tilde{S}_0, \tilde{F}_0, \tilde{S}_1$ , and  $\tilde{F}_1$  are continuous function. For any  $t', t'' \in (0,1]$  where  $t' \leq t''$ , it holds that

(3.6) 
$$0 < \tilde{S}_0(t') \le t' \tilde{F}_0(t'').$$

Moreover, the following limit also exists:

(3.7) 
$$\lim_{m \to \infty} \frac{m_0}{m} = \pi_0 \in (0, 1).$$

The convergence properties stated in (3.4), (3.5), and (3.7) are typical in the asymptotic analysis literature; for example, see Storey et al. (2003, Sec 2.2), Wang et al. (2022, Sec 4.2), and Zhang and Chen (2022, Sec 3.1). Any type of dependence where (3.4) and (3.5) can hold is what we consider weak dependence. For instance, dependence in finite-dimensional blocks, autoregressive dependence, and mixing distributions are candidates for weak dependence. The properties relating to  $\tilde{S}_0$  and  $\tilde{F}_0$  in (3.6) are not standard in the asymptotic analysis literature but arise naturally from the conditionally validity of Bonferroni-combined PC p-values:

**Lemma 3.2** (Conditional validity). Suppose  $H_i^{u/n}$  is true and the p-values  $P_{i1}, \ldots, P_{in}$  are all valid and mutually independent. Then for any fixed  $t', t'' \in [0, 1]$  where  $t' \leq t''$ , we have that

(3.8) 
$$\Pr(S_i < t' | F_i < t'') = \frac{\Pr(S_i < t', F_i < t'')}{\Pr(F_i < t'')} = \frac{\Pr(S_i < t')}{\Pr(F_i < t'')} \le t'$$

if 
$$\Pr(F_i < t') > 0$$
.

Rearranging (3.8) gives  $\Pr(S_i < t') \le t' \cdot \Pr(F_i < t'')$  when  $H_i^{u/n}$  is true, which is analogous to the property stated in (3.6) when (3.4) and (3.7) hold. The proof for Lemma 3.2 is provided in Appendix B.1.

Below, we present our asymptotic result, which states that if k remains a constant proportion of m as  $m \to \infty$ , then AdaFilter-AdaBon asymptotically controls the k-FWER below  $\alpha$ .

**Theorem 3.3** (AdaFilter-AdaBon k-FWER control). Suppose Assumption 1 holds and

$$k \equiv k(m) = \omega \cdot m$$

for a constant  $\omega \in (0,1)$ . If there exists a threshold  $t \in (0,1]$  such that

(3.9) 
$$\lim_{m \to \infty} t \cdot \frac{\sum_{i=1}^{m} I\{F_i < t\} I\{S_i \ge \theta t\}}{m(1 - \theta t)} < \omega \alpha$$

with probability 1, then AdaFilter-AdaBon (Definition 3) has the following asymptotic k-FWER property:

$$\limsup_{m \to \infty} (k - FWER(\mathcal{R})) \le \alpha,$$

where 
$$\mathcal{R} = \{i \in \{1, \dots, m\} : S_i \leq \hat{t}_{\theta}\}.$$

The technical condition stated in (3.9) ensures that  $\hat{t}_{\theta}$  is well behaved in the limit. The proof of Theorem 3.3, given in Appendix B.4, first establishes that the expected number of false discoveries

is asymptotically bounded by  $k\alpha$ , and then applies Markov's inequality to show that the k-FWER is below  $\alpha$ .

# 4. Simulation studies

4.1. **Simulated data.** Consider a meta-analysis consisting of m = 500 features across n = 4 studies. For each feature i and study j, let  $\mu_{ij} \in \mathbb{R}$  be an effect parameter, and consider the null hypothesis  $H_{ij}: \mu_{ij} = 0$  with the corresponding alternative  $\mu_{ij} \neq 0$ . The data for this meta-analysis are simulated according to the steps provided below.

For a given correlation parameter  $\rho \in [-1, 1]$ , we arbitrarily partition the 500 features into b blocks of size 500/b each, where

$$b = \begin{cases} 5 & \text{if } \rho \ge 0, \\ 250 & \text{if } \rho < 0. \end{cases}$$

We then generate standard normal variables  $\epsilon_{ij} \sim N(0,1)$  for each pair (i,j), subject to the following dependence structure:

- The sequences  $(\epsilon_{i1})_{i=1}^{500}$ ,  $(\epsilon_{i2})_{i=1}^{500}$ ,  $(\epsilon_{i3})_{i=1}^{500}$ , and  $(\epsilon_{i4})_{i=1}^{500}$  are mutually independent;
- Within each study j,

$$\operatorname{cor}(\epsilon_{ij}, \epsilon_{i'j}) = \begin{cases} \rho, & \text{if } i \text{ and } i' \text{ belong to the same block and } i \neq i'; \\ 1, & \text{if } i = i'; \\ 0, & \text{if } i \text{ and } i' \text{ do not belong to the same block.} \end{cases}$$

The following process is then used to generate the p-values  $(P_{ij})_{500\times4}$ . For each  $i=1,\ldots,500$ :

- (1) Generate  $c_i \sim \text{Bernoulli}(\pi_1)$ , where  $\pi_1 \in [0,1]$  is a given signal density parameter.
  - (i) If  $c_i = 0$ , set  $(\mu_{i1}, \mu_{i2}, \mu_{i3}, \mu_{i4}) = (0, 0, 0, 0)$ ;
  - (ii) If  $c_i = 1$ , sample  $(\mu_{i1}, \mu_{i2}, \mu_{i3}, \mu_{i4})$  uniformly from the set  $\{0, 4\}^4$ .
- (2) For j = 1, ..., 4, compute  $P_{ij} = 1 \Phi(\mu_{ij} + \epsilon_{ij})$ .

Hence, by construction, the p-values in  $(P_{ij})_{500\times4}$  are valid and independent across studies, while exhibiting block-wise equicorrelation within each study.

- 4.2. **Compared methods.** The following multiple testing procedures, representing a mix of state-of-the-art and well-established approaches, are considered in our simulations:
  - (a) AdaFilter-AdaBon: Definition 3 with  $\theta = 0.5$ .
  - (b) AdaFilter-Bon: Definition 2.
  - (c) Bonferroni: Definition 1.
  - (d) Hochberg: The step-up k-FWER procedure described in Theorem 4.2 of Sarkar (2007).
  - (e) Adaptive Bonferroni: Definition 1 of Guo (2009) with tuning parameter  $\lambda = 0.5$ .
  - (f) Adaptive Hochberg: Definition 4 of Sarkar et al. (2012) with  $\hat{n}(\kappa)$  where  $\kappa = 490$ .

Let  $W_u(\cdot)$  denote the cdf of a chi-squared distribution with  $2 \cdot (n-u+1)$  degrees of freedom. Methods (c)—(f) are implemented in our simulations with the PC p-value inputs  $(P_i^{u/n})_{i=1}^m$ , where each  $P_i^{u/n}$  is constructed as the Fisher-combined PC p-value (Benjamini and Heller, 2008, Eq. 5):

$$f_{\text{Fisher}}(P_{i1}, \dots, P_{in}; u) \equiv 1 - W_u \left(-2 \sum_{j=u}^n \log P_{i(j)}\right).$$

This choice is motivated by the fact that Fisher-combined PC p-values are less conservative than their Bonferroni-combined PC p-values (Bogomolov and Heller, 2023, Sec. 2.1). By contrast, methods (a) and (b), by definition, must operate on Bonferroni-combined PC p-values.

- 4.3. Simulation results. We apply methods (a)—(f) to the simulated data described in Section 4.1 to test the PC nulls  $(H_i^{u/4})_{i=1}^{500}$  for  $u \in \{2,3,4\}$ , targeting FWER control at level  $\alpha = 0.05$ . The following parameter ranges are considered for  $\pi_1$  and  $\rho$ :
  - $\pi_1 \in \{0.025, 0.05, 0.075, 0.10, 0.125, 0.15\}.$
  - $\rho \in \{-0.8, -0.2, 0.2, 0.8\}.$

By design, each  $H_i^{u/n}$  becomes more likely to be false as  $\pi_1$  increases from 0.025 to 0.15. Moreover, the block-wise correlation strength among the p-values in  $(P_{ij})_{500\times4}$  shifts from strongly negative to strongly positive as  $\rho$  increases from -0.8 to 0.8. The simulation results are presented in Figure 4.1, where the empirical FWER and TPR level of the compared methods are computed based on 1,000 repetitions. The following observations can be made:

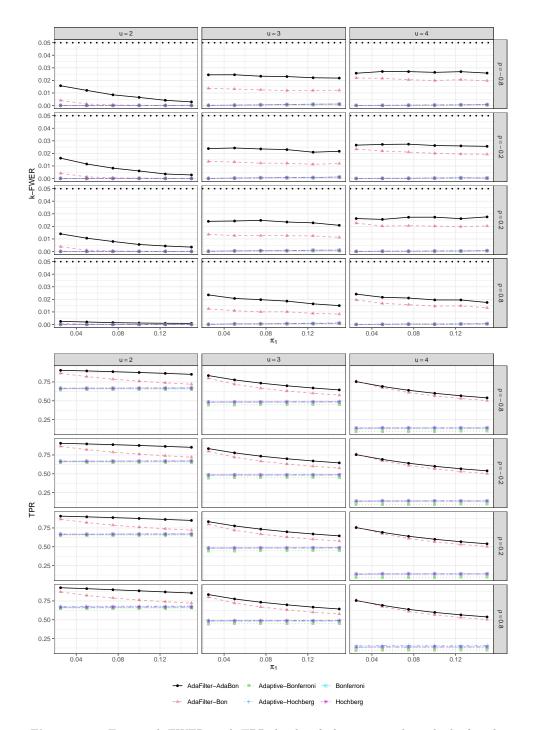
- AdaFilter-AdaBon demonstrated finite-sample FWER control across all settings of  $\pi_1$ , u, and  $\rho$ . This is reassuring, given that we only established its theoretical control guarantee asymptotically as  $m \to \infty$  (Theorem 3.3). We also observe that AdaFilter-AdaBon's FWER level is uniformly greater than that of AdaFilter-Bon, indicating that the former is less conservative than the latter. Outside of AdaFilter-AdaBon and AdaFilter-Bon, the remaining methods generally exhibited extremely conservative FWER control. As  $\rho$  increases from -0.8 to 0.8, the FWER level for all methods decreases slightly.
- AdaFilter-AdaBon outperforms AdaFilter-Bon under most parameter settings, with notably higher power when  $\pi_1 \geq 0.10$  and u = 2. Only when  $\pi_1 = 0.025$  and u = 4 does AdaFilter-Bon match the power of AdaFilter-AdaBon. The remaining methods exhibited substantially lower power than both AdaFilter-AdaBon and AdaFilter-Bon, particularly when u = 4, where their power was close to zero.

Hence, our simulations demonstrate that AdaFilter-AdaBon delivers state-of-the-art performance, especially when the replicability requirement (u) is small and the signal density  $(\pi_1)$  is large. Even in less favorable settings, its power remains comparable to that of its predecessor, AdaFilter-Bon.

Additional simulation results comparing the methods under k-FWER control for k=5 and 10, using the same  $\pi_1$  and  $\rho$  settings as above, are provided in Appendix C. The simulation results presented in this section and in Appendix C can be reproduced in R via the steps provided in https://github.com/ninhtran02/AdaFilterAdaBon.

# 5. Discussion

We proposed AdaFilter-AdaBon, a p-value-based procedure for testing partial conjunction hypotheses with asymptotic k-FWER control. By combining filtering with a post-filter null proportion estimator,



**Figure 4.1.** Empirical FWER and TPR levels of the compared methods for the parameter settings  $\alpha = 0.05, \ u \in \{2,3,4\}, \ \pi_1 \in \{0.025,0.05,0.075,0.10,0.125,0.15\},$  and  $\rho \in \{-0.8,-0.2,0.2,0.8\}.$ 

our method improves power over its predecessor, AdaFilter-Bon (Wang et al., 2022), while maintaining robust finite-sample k-FWER control, as shown in our simulation studies.

AdaFilter-AdaBon can be extended to control the false exceedance rate (FDX) and the false discovery rate (FDR). For a set  $\mathcal{R} \subseteq \{1, \dots, m\}$  of rejected PC nulls and tolerance level  $\gamma \in (0, 1)$ , the FDX is defined as

$$FDX \equiv FDX(\mathcal{R}; \gamma) = \Pr\left(\frac{\sum_{i=1}^{m} I\{i \in \mathcal{R}\} I\{H_i^{u/n} \text{ is true}\}}{1 \vee \sum_{i=1}^{m} I\{i \in \mathcal{R}\}} \ge \gamma\right),$$

which is the probability that the proportion of false discoveries among the total number of discoveries exceeds  $\gamma$ . The false discovery rate is defined as

$$FDR \equiv FDR(\mathcal{R}) = \Pr\left[\frac{\sum_{i=1}^{m} I\{i \in \mathcal{R}\} I\{H_i^{u/n} \text{ is true}\}}{1 \vee \sum_{i=1}^{m} I\{i \in \mathcal{R}\}}\right] \in (0,1),$$

which is the expected proportion of false discoveries among the total number of discoveries. The following procedure, based on AdaFilter-AdaBon, controls both FDX and FDR asymptotically:

**Definition 4.** Let  $\alpha \in [0,1]$  be a k-FWER target for a given  $k \in \mathbb{N}$ ,  $u \in \{2,\ldots,m\}$  be a replicability level,  $\theta \in (0,1)$  be a tuning parameter, and  $F_i$  and  $S_i$  be as defined in (2.5) and (2.4), respectively. For  $i = 1,\ldots,m$ , reject  $H_i^{u/n}$  if  $S_i \leq \hat{\tau}$ , where

(5.1) 
$$\hat{\tau} \equiv \sup \left\{ \tau \in [0, 1] : \frac{\sum_{i=1}^{m} I\{\hat{t}_{\theta} \le S_i \le \tau\} + k}{1 \vee \sum_{i=1}^{m} I\{S_i \le \tau\}} \le \gamma \right\}$$

and  $\hat{t}_{\theta}$  is AdaFilter-AdaBon's rejection threshold as defined in (3.2).

**Corollary 5.1** (FDX and FDR control). If Theorem 3.3 holds, then the procedure in Definition 4 has the following asymptotic error properties:

$$\limsup_{m \to \infty} \mathit{FDX}(\mathcal{R}) \leq \alpha \quad and \quad \limsup_{m \to \infty} \mathit{FDR}(\mathcal{R}) \leq \alpha + \gamma,$$

where 
$$\mathcal{R} = \{i \in \{1, ..., m\} : S_i \leq \hat{\tau}\}.$$

The proof of Corollary 5.1 is given in Appendix D.1. Within the proof, it is shown that the procedure in Definition 4 controls FDX and FDR indirectly through k-FWER control. Hence, an avenue for future work is to develop more direct approaches for controlling FDX and FDR that incorporate filtering and our post-filter null proportion estimator. This may lead to improved power over the procedure presented in Definition 4.

### References

Benjamini, Y. and Heller, R. (2008). Screening for partial conjunction hypotheses. <u>Biometrics</u>, 64(4):1215–1222.

Bogomolov, M. and Heller, R. (2018). Assessing replicability of findings across two studies of multiple features. Biometrika, 105(3):505–516.

Bogomolov, M. and Heller, R. (2023). Replicability Across Multiple Studies. <u>Statistical Science</u>, 38(4):602 – 620.

Dickhaus, T., Heller, R., Hoang, A.-T., and Rinott, Y. (2024). A procedure for multiple testing of partial conjunction hypotheses based on a hazard rate inequality.

- Guo, W. (2009). A note on adaptive bonferroni and holm procedures under dependence. <u>Biometrika</u>, 96(4):1012–1018.
- Hoang, A.-T. and Dickhaus, T. (2022). Combining independent p-values in replicability analysis: a comparative study. Journal of Statistical Computation and Simulation, 92(10):2184–2204.
- Lehmann, E. L. and Romano, J. P. (2005). Generalizations of the familywise error rate. <u>The Annals of Statistics</u>, 33(3):1138 1154.
- Romano, J. P. and Shaikh, A. M. (2006). Stepup procedures for control of generalizations of the familywise error rate. The Annals of Statistics, 34(4):1850 1873.
- Sarkar, S. K. (2007). Stepup procedures controlling generalized fwer and generalized fdr. The Annals of Statistics, 35(6):2405–2420.
- Sarkar, S. K. (2008). Generalizing simes' test and hochberg's stepup procedure. <u>The Annals of Statistics</u>, 36(1):337–363.
- Sarkar, S. K., Guo, W., and Finner, H. (2012). On adaptive procedures controlling the familywise error rate. Journal of Statistical Planning and Inference, 142(1):65–78.
- Storey, J. D., Taylor, J. E., and Siegmund, D. (2003). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach. <u>Journal of the Royal</u> Statistical Society Series B: Statistical Methodology, 66(1):187–205.
- Tran, N. and Leung, D. (2025). A covariate-adaptive test for replicability across multiple studies with false discovery rate control.
- Wang, J., Gui, L., Su, W. J., Sabatti, C., and Owen, A. B. (2022). Detecting multiple replicating signals using adaptive filtering procedures. The Annals of Statistics, 50(4):1890 1909.
- Wang, J. and Owen, A. B. (2019). Admissibility in partial conjunction testing. <u>Journal of the American</u> Statistical Association, 114(525):158–168.
- Zhan, S. J., Kunz, C. U., and Stallard, N. (2022). Should the two-trial paradigm still be the gold standard in drug assessment? Pharm Stat, 22(1):96–111.
- Zhang, X. and Chen, J. (2022). Covariate adaptive false discovery rate control with applications to omics-wide multiple testing. Journal of the American Statistical Association, 117(537):411–427.

# APPENDIX A. RESULTS AND PROOFS RELATING TO SECTION 2

A.1. Technical lemmas for proving Theorem 2.1. For all  $i \in \{1, ..., m\}$ , let

(A.1) 
$$\hat{t}_i = \sup \left\{ t \in [0, 1] : t \cdot \left( 1 + \sum_{\ell \neq i} I\{F_\ell < t\} \right) \le \alpha \right\}$$

be a threshold similar to  $\hat{t}$  in (2.6), but computed without data from feature i. We now state and prove two key lemmas involving  $\hat{t}_i$ , which are required for the proof of Theorem 2.1 in Appendix A.2.

**Lemma A.1**  $(\hat{t}_i \leq \hat{t})$ . Let  $\hat{t}$  and  $\hat{t}_i$  be as defined in (2.6) and (A.1), respectively. Then  $\hat{t}_i \leq \hat{t}$ .

Proof of Lemma A.1. Note that

$$t \cdot \sum_{\ell=1}^{m} I\{F_{\ell} < t\} = t \cdot \left( I\{F_{i} < t\} + \sum_{\ell \neq i} I\{F_{\ell} < t\} \right) \le t \cdot \left( 1 + \sum_{\ell \neq i} I\{F_{\ell} < t\} \right).$$

As a consequence of the display above, we have that

$$\hat{t} \equiv \sup \left\{ t \in [0, 1] : t \cdot \sum_{\ell=1}^{m} I\{F_{\ell} < t\} \le \alpha \right\}$$

$$\geq \sup \left\{ t \in [0, 1] : t \cdot \left( 1 + \sum_{\ell \neq i} I\{F_{\ell} < t\} \right) \le \alpha \right\} \equiv \hat{t}_{i}.$$

**Lemma A.2** ( $\hat{t}_i = \hat{t}$  if  $F_i < \hat{t}$ ). Let  $\hat{t}$  and  $\hat{t}_i$  be as defined in (2.6) and (A.1), respectively. If  $F_i < \hat{t}$ , then  $\hat{t}_i = \hat{t}$ .

Proof of Lemma A.2. If  $F_i < \hat{t}$ , it must be the case that

$$\hat{t} \equiv \sup \left\{ t \in [0, 1] : t \cdot \sum_{\ell=1}^{m} I\{F_{\ell} < t\} \le \alpha \right\}$$

$$= \sup \left\{ t \in (F_{i}, 1] : t \cdot \sum_{\ell=1}^{m} I\{F_{\ell} < t\} \le \alpha \right\}$$

$$= \sup \left\{ t \in (F_{i}, 1] : t \cdot \left( 1 + \sum_{\ell \neq i} I\{F_{\ell} < t\} \right) \le \alpha \right\}$$

where

(A.2) 
$$\exists t^* \in (F_i, 1] \text{ such that } t^* \cdot \left(1 + \sum_{\ell \neq i} I\{F_\ell < t^*\}\right) \leq \alpha.$$

From (A.1), we have that

$$\hat{t}_i \equiv \sup \left\{ t \in [0, 1] : t \cdot \left( 1 + \sum_{\ell \neq i} I\{F_\ell < t\} \right) \le \alpha \right\}$$

$$= \sup \left\{ t \in (F_i, 1] : t \cdot \left( 1 + \sum_{\ell \neq i} I\{F_\ell < t\} \right) \le \alpha \right\}$$

where the second equality is a consequence of (A.2). Hence, we can conclude that  $\hat{t} = \hat{t}_i$ .

# A.2. Proof of Theorem 2.1.

Let  $V = \sum_{i=1}^{m} I\{S_i < \hat{t}\}I\{H_i^{u/n} \text{ is true}\}$  denote the number of false discoveries made by AdaFilter-Bon (Definition 2), and  $\hat{t}_i$  be the rejection threshold defined in (A.1). We have that

$$\mathbb{E}[V] = \mathbb{E}\left[\sum_{i=1}^{m} I\{S_i < \hat{t}\}I\{H_i^{u/n} \text{ is true}\}\right]$$
$$= \mathbb{E}\left[\sum_{i=1}^{m} I\{S_i < \hat{t}\}I\{F_i < \hat{t}\}I\{H_i^{u/n} \text{ is true}\}\right]$$

$$= \mathbb{E}\left[\sum_{i=1}^{m} I\{S_i < \hat{t}_i\}I\{F_i < \hat{t}_i\}I\{H_i^{u/n} \text{ is true}\}\right]$$

where the second equality is a consequence of  $F_i \leq S_i$ , and the third equality is a result of Lemma A.2. Continuing on from the previous display, we have that

$$\mathbb{E}[V] = \sum_{i=1}^{m} \mathbb{E}\left[\mathbb{E}\left[I\left\{S_{i} < \hat{t}_{i}\right\} \middle| \hat{t}_{i}, F_{i} < \hat{t}_{i}\right] I\left\{F_{i} < \hat{t}_{i}\right\} I\left\{H_{i}^{u/n} \text{ is true}\right\}\right]$$

$$\leq \sum_{i=1}^{m} \mathbb{E}\left[\hat{t}_{i} \cdot I\left\{F_{i} < \hat{t}_{i}\right\} I\left\{H_{i}^{u/n} \text{ is true}\right\}\right]$$

$$\leq \sum_{i=1}^{m} \mathbb{E}\left[k\alpha \cdot \frac{1}{1 + \sum_{\ell \neq i} I\left\{F_{\ell} \leq \hat{t}_{i}\right\}} \cdot I\left\{F_{i} \leq \hat{t}_{i}\right\} I\left\{H_{i}^{u/n} \text{ is true}\right\}\right]$$

$$\leq \sum_{i=1}^{m} \mathbb{E}\left[\frac{k\alpha}{\sum_{\ell=1}^{m} I\left\{F_{\ell} \leq \hat{t}_{i}\right\}} I\left\{F_{i} \leq \hat{t}_{i}\right\} I\left\{H_{i}^{u/n} \text{ is true}\right\}\right]$$

where the first inequality is a consequence of Lemma 3.2 and the independence between  $S_i$  and  $\hat{t}_i$ , and the second inequality follows from (A.1). Continuing on from the display above, we have that

$$\mathbb{E}[V] \leq \sum_{i=1}^{m} \mathbb{E}\left[\frac{k\alpha}{\sum_{\ell=1}^{m} I\{F_{\ell} \leq \hat{t}_{i}\}} I\{F_{i} \leq \hat{t}\} I\{H_{i}^{u/n} \text{ is true}\}\right]$$

$$= \sum_{i=1}^{m} \mathbb{E}\left[\frac{k\alpha}{\sum_{\ell=1}^{m} I\{F_{\ell} \leq \hat{t}\}} I\{F_{i} \leq \hat{t}\} I\{H_{i}^{u/n} \text{ is true}\}\right]$$

$$= k\alpha \cdot \mathbb{E}\left[\frac{\sum_{i=1}^{m} I\{F_{i} \leq \hat{t}\} I\{H_{i}^{u/n} \text{ is true}\}}{\sum_{\ell=1}^{m} I\{F_{\ell} \leq \hat{t}\}}\right]$$

where the inequality is a result of Lemma A.1 and the first equality is a consequence of Lemma A.2. By Markov's inequality, it follows that

$$k\text{-FWER}(\mathcal{R}) = \Pr(V \ge k) \le \alpha \mathbb{E}\left[\frac{\sum_{i=1}^m I\{F_i \le \hat{t}\}I\{H_i^{u/n} \text{ is true}\}}{\sum_{\ell=1}^m I\{F_\ell \le \hat{t}\}}\right].$$

Appendix B. Results and proofs relating to Section 3

# B.1. Proof of Lemma 3.2.

We have that

$$\Pr(S_i < t' | F_i < t'') = \frac{\Pr(S_i < t', F_i < t'')}{\Pr(F_i < t'')}$$

$$\leq \frac{\Pr(S_i < t')}{\Pr(F_i < t'')}$$
14

$$\leq \frac{\Pr(S_i < t')}{\Pr(F_i < t')}$$

$$= \frac{\Pr(S_i < t', F_i < t')}{\Pr(F_i < t')}$$

$$= \Pr(S_i < t' | F_i < t')$$

$$\leq t'$$

where the second equality is a result of  $F_i \leq S_i$ , and the last inequality is a consequence of Lemma 4.1 from Wang et al. (2022).

#### B.2. Proof of Theorem 3.1.

Without loss of generality, suppose  $\mathcal{G}$  is ordered and is of size  $|\mathcal{G}| = M$ . To denote the ordered elements of  $\mathcal{G}$ , we let

$$\mathcal{G} \equiv \{G_1, G_2, \dots, G_M\}$$

where  $0 = G_1 \le G_2 \le \cdots \le G_M = 1$ . Consider the functions  $f: [0,1] \longrightarrow [0,\infty)$  and  $g: [0,1] \longrightarrow 2^{\{1,\dots,m\}}$  defined below:

$$f(t) \equiv \sum_{i=1}^{m} I\{F_i < t\}I\{S_i \ge \theta t\}$$
 and  $g(t) \equiv \{i \in \{1, \dots, m\} : S_i < t\}.$ 

It is not difficult to see that both f(t) and g(t) are step functions on the following intervals for t:

$$\mathcal{I}_1 = [G_1, G_2), \quad \mathcal{I}_2 = [G_2, G_3), \quad \dots, \quad \mathcal{I}_{M-1} = [G_{M-1}, G_M), \quad \mathcal{I}_M = \{G_M\},$$

where  $\bigcup_{\ell=1}^{M} \mathcal{I}_{\ell} = [0,1]$ . That is, for any  $t \in [0,1]$ , f(t) and g(t) can be written as

$$f(t) = \sum_{i=1}^{m} I\{F_i < G_\ell\}I\{S_i \ge \theta \cdot G_\ell\}$$
 and  $g(t) = \{i \in \{1, \dots, m\} : S_i < G_\ell\}$ 

respectively, where

$$t \in \mathcal{I}_{\ell} = \begin{cases} [G_{\ell}, G_{\ell+1}), & \ell \neq M \\ \{G_M\}, & \ell = M \end{cases}$$

Let  $\hat{\ell}_{\theta} \in \{1, \dots, M\}$  be the index such that  $\hat{t}_{\theta} \in \mathcal{I}_{\hat{\ell}_{\theta}}$ . It follows that

(B.1) 
$$\sum_{i=1}^{m} I\{F_i < \hat{t}_{\theta}\}I\{S_i \ge \theta \cdot \hat{t}_{\theta}\} = f(\hat{t}_{\theta}) = \sum_{i=1}^{m} I\{F_i < G_{\hat{\ell}_{\theta}}\}I\{S_i \ge \theta \cdot G_{\hat{\ell}_{\theta}}\}$$

and

(B.2) 
$$\left\{ i \in \{1, \dots, m\} : S_i < \hat{t}_{\theta} \right\} = g(\hat{t}_{\theta}) = \left\{ i \in \{1, \dots, m\} : S_i < G_{\hat{\ell}_{\theta}} \right\}.$$

Since  $\mathcal{G} \subseteq [0, 1]$ , it follows from the definition of  $\hat{t}_{\theta}$  and  $\check{t}_{\theta}$  in (3.2) and (3.3), respectively, that  $\check{t}_{\theta} \leq \hat{t}_{\theta}$ . Since  $G_{\hat{\ell}_{\theta}}$  is the element in  $\mathcal{G}$  that is closest to  $\hat{t}_{\theta}$  without being strictly greater than  $\hat{t}_{\theta}$ , it must hold from the inequality  $\check{t}_{\theta} \leq \hat{t}_{\theta}$  that  $\check{t}_{\theta} \leq G_{\hat{\ell}_{\theta}}$ . Now consider the following result:

$$G_{\hat{\ell}_{\theta}} \cdot \frac{\sum_{i=1}^{m} I\{F_{i} < G_{\hat{\ell}_{\theta}}\} I\{S_{i} \ge \theta \cdot G_{\hat{\ell}_{\theta}}\}}{1 - \theta \cdot G_{\hat{\ell}_{\theta}}} = \frac{G_{\hat{\ell}_{\theta}}}{1 - \theta \cdot G_{\hat{\ell}_{\theta}}} \cdot \sum_{i=1}^{m} I\{F_{i} < \hat{t}_{\theta}\} I\{S_{i} \ge \theta \cdot \hat{t}_{\theta}\}$$

$$\leq \frac{\hat{t}_{\theta}}{1 - \theta \cdot \hat{t}_{\theta}} \cdot \sum_{i=1}^{m} I\{F_i < \hat{t}_{\theta}\}I\{S_i \geq \theta \cdot \hat{t}_{\theta}\}$$
  
$$\leq k\alpha$$

where the equality is a result of (B.1), the first inequality is a result of  $\hat{t}_{\theta} \geq G_{\hat{\ell}_{\theta}}$ , and the last inequality is a result of the definition of  $\hat{t}_{\theta}$  in (3.2). The display directly above implies that  $\check{t}_{\theta} \geq G_{\hat{\ell}_{\theta}}$  by the definition of  $\check{t}_{\theta}$  in (3.3). However, since we have already established that  $\check{t}_{\theta} \leq G_{\hat{\ell}_{\theta}}$ , it therefore must be that  $\check{t}_{\theta} = G_{\hat{\ell}_{\theta}}$ . Hence, it follows from (B.2) that

$$\{i \in \{1, \dots, m\} : S_i < \hat{t}_{\theta}\} = \{i \in \{1, \dots, m\} : S_i < G_{\hat{\ell}_{\theta}}\} = \{i \in \{1, \dots, m\} : S_i < \check{t}_{\theta}\}.$$

B.3. **Technical lemmas for proving Theorem 3.3.** We provide in this section technical results analogous to those of Section 5 in Storey et al. (2003). They serve as essential ingredients in proving Theorem 3.3, presented in Appendix B.4.

**Lemma B.1.** Under Assumption 1, we have for fixed constants  $\theta \in (0,1)$  and  $\delta \in (0,1]$  that

$$\lim_{m \to \infty} \inf_{t \ge \delta} \left( t \cdot \frac{\sum_{i=1}^m I\{F_i < t\} I\{S_i \ge \theta t\}}{(1 - \theta t) \sum_{i=1}^m I\{F_i < t\}} - \frac{\pi_0 \tilde{S}_0(t) m}{\sum_{i=1}^m I\{F_i < t\}} \right) \ge 0$$

with probability 1.

Proof of Lemma B.1. Note that  $I\{F_i < t\} = 1$  whenever  $I\{S_i < \theta t\} = 1$ , since  $F_i \le S_i$  and  $\theta t < t$ . This gives us the following result:

$$\frac{\sum_{i=1}^{m} I\{F_{i} < t\} I\{S_{i} \ge \theta t\}}{(1 - \theta t) \sum_{i=1}^{m} I\{F_{i} < t\}} = \frac{\sum_{i=1}^{m} I\{F_{i} < t\} (1 - I\{S_{i} < \theta t\})}{(1 - \theta t) \sum_{i=1}^{m} I\{F_{i} < t\}}$$

$$= \frac{\sum_{i=1}^{m} I\{F_{i} < t\} - \sum_{i=1}^{m} I\{F_{i} < t\} I\{F_{i} < t\}}{(1 - \theta t) \sum_{i=1}^{m} I\{F_{i} < t\}}$$

$$= \frac{\sum_{i=1}^{m} I\{F_{i} < t\} - \sum_{i=1}^{m} I\{S_{i} < \theta t\}}{(1 - \theta t) \sum_{i=1}^{m} I\{F_{i} < t\}}$$

$$= \frac{(1/m) \sum_{i=1}^{m} I\{F_{i} < t\} - (1/m) \sum_{i=1}^{m} I\{S_{i} < \theta t\}}{(1 - \theta t) (1/m) \sum_{i=1}^{m} I\{F_{i} < t\}}.$$
(B.3)

Since  $F_i \leq S_i$  and  $\tilde{F}_1$  is a non-decreasing function by construction, it follows from (3.5) in Assumption 1 that

(B.4) 
$$\tilde{S}_1(\theta t) \leq \tilde{F}_1(\theta t) \leq \tilde{F}_1(t)$$

for all  $t \in (0,1]$ . We also have by (3.6) in Assumption 1 that

(B.5) 
$$\tilde{S}_0(\theta t) \le \theta t \tilde{F}_0(\theta t) \le \theta t \tilde{F}_0(t)$$

for all  $t \in (0,1]$ .

For the next part of this proof, we adopt the following notation for convenience:

(B.6) 
$$\tilde{F}(t) \equiv \pi_0 \tilde{F}_0(t) + \pi_1 \tilde{F}_1(t) \text{ and } \tilde{S}(t) \equiv \pi_0 \tilde{S}_0(t) + \pi_1 \tilde{S}_1(t).$$

By (3.4), (3.5), and (3.7) in Assumption 1, and by (B.3), we have for all  $t \in (0,1]$  that

$$\lim_{m \to \infty} \frac{\sum_{i=1}^{m} I\{F_{i} < t\} I\{S_{i} \ge \theta t\}}{(1 - \theta t) \sum_{i=1}^{m} I\{F_{i} < t\}} = \frac{\tilde{F}(t) - \tilde{S}(\theta t)}{(1 - \theta t) \tilde{F}(t)}$$

$$= \frac{\pi_{0} \tilde{F}_{0}(t) + \pi_{1} \tilde{F}_{1}(t) - \pi_{0} \tilde{S}_{0}(\theta t) - \pi_{1} \tilde{S}_{1}(\theta t)}{(1 - \theta t) \tilde{F}(t)}$$

$$\ge \frac{\pi_{0} \tilde{F}_{0}(t) + \pi_{1} \tilde{F}_{1}(t) - \pi_{0} \theta t \tilde{F}_{0}(t) - \pi_{1} \tilde{F}_{1}(t)}{(1 - \theta t) \tilde{F}(t)}$$

$$= \frac{\pi_{0} (1 - \theta t) \tilde{F}_{0}(t)}{(1 - \theta t) \tilde{F}(t)}$$

$$= \frac{\pi_{0} \tilde{F}_{0}(t)}{\tilde{F}(t)},$$
(B.7)

where the inequality is a consequence of (B.5) and (B.4). Since  $\tilde{S}_0(t) \leq t\tilde{F}_0(t)$  holds by (3.6) in Assumption 1, it follows from (B.7) that

(B.8) 
$$\lim_{m \to \infty} \left( t \cdot \frac{\sum_{i=1}^{m} I\{F_i < t\} I\{S_i \ge \theta t\}}{(1 - \theta t) \sum_{i=1}^{m} I\{F_i < t\}} - \frac{\pi_0 \tilde{S}_0(t)}{\tilde{F}(t)} \right) \ge 0.$$

By (3.4), (3.5), and (3.7) in Assumption 1, we also have that

(B.9) 
$$\lim_{m \to \infty} \frac{\pi_0 \tilde{S}_0(t) m}{\sum_{i=1}^m I\{F_i < t\}} = \frac{\pi_0 \tilde{S}_0(t)}{\tilde{F}(t)}.$$

By combining (B.8) and (B.9), it holds th

$$\lim_{m \to \infty} \inf_{t \ge \delta} \left( t \cdot \frac{\sum_{i=1}^{m} I\{F_i < t\} I\{S_i \ge \theta t\}}{(1 - \theta t) \sum_{i=1}^{m} I\{F_i < t\}} - \frac{\pi_0 \tilde{S}_0(t) m}{\sum_{i=1}^{m} I\{F_i < t\}} \right) \ge 0$$

**Lemma B.2.** Under Assumption 1, we have for fixed constants  $\theta \in (0,1)$  and  $\delta \in (0,1]$  that

$$\lim_{m \to \infty} \sup_{t \ge \delta} \left| \frac{\sum_{i=1}^{m} I\{S_i < t\} I\{H_i^{u/n} \text{ is true}\}}{\sum_{i=1}^{m} I\{F_i < t\}} - \frac{\pi_0 m \tilde{S}_0(t)}{\sum_{i=1}^{m} I\{F_i < t\}} \right| = 0$$

with probability 1.

Proof of Lemma B.2. By the Glivenko-Cantelli theorem, it follows from Assumption 1 that

(B.10) 
$$\lim_{m \to \infty} \sup_{0 \le t \le 1} \left| \frac{\sum_{i=1}^{m} I\{S_i < t\} I\{H_i^{u/n} \text{ is true}\}}{m} - \pi_0 \tilde{S}_0(t) \right| = 0$$

with probability 1. To prove the lemma, note that

$$\begin{split} & \lim_{m \to \infty} \sup_{t \ge \delta} \left| \frac{\sum_{i=1}^{m} I\{S_i < t\} I\{H_i^{u/n} \text{ is true}\}}{\sum_{i=1}^{m} I\{F_i < t\}} - \frac{\pi_0 m \tilde{S}_0(t)}{\sum_{i=1}^{m} I\{F_i < t\}} \right| \\ & \le \lim_{m \to \infty} \sup_{t \ge \delta} \left| \frac{m}{\sum_{i=1}^{m} I\{F_i < t\}} \right| \left| \frac{\sum_{i=1}^{m} I\{S_i < t\} I\{H_i^{u/n} \text{ is true}\}}{m} - \pi_0 \tilde{S}_0(t) \right| \end{split}$$

$$\leq \lim_{m \to \infty} \left| \frac{m}{\sum_{i=1}^{m} I\{F_i < \delta\}} \right| \sup_{t > \delta} \left| \frac{\sum_{i=1}^{m} I\{S_i < t\} I\{H_i^{u/n} \text{ is true}\}}{m} - \pi_0 \tilde{S}_0(t) \right| = 0$$

where the equality is a consequence of (B.10).

**Lemma B.3.** Under Assumption 1, we have for fixed constants  $\theta \in (0,1)$  and  $\delta \in (0,1]$  that

$$\lim_{m \to \infty} \inf_{t \ge \delta} \left( \frac{\sum_{i=1}^{m} I\{F_i < t\} I\{S_i \ge \theta t\}}{m(1 - \theta t)} \cdot t - \frac{\sum_{i=1}^{m} I\{S_i < t\} I\{H_i^{u/n} \text{ is true}\}}{m} \right) \ge 0$$

with probability 1.

Proof of Lemma B.3. We have that

$$\lim_{m \to \infty} \inf_{t \ge \delta} \left( t \cdot \frac{\sum_{i=1}^{m} I\{F_i < t\} I\{S_i \ge \theta t\}}{m(1 - \theta t)} - \frac{\sum_{i=1}^{m} I\{S_i < t\} I\{H_i^{u/n} \text{ is true}\}}{m} \right)$$

$$= \lim_{m \to \infty} \inf_{t \ge \delta} \left( \frac{\sum_{i=1}^{m} I\{F_i < t\}}{m} \right) \left( t \cdot \frac{\sum_{i=1}^{m} I\{F_i < t\} I\{S_i \ge \theta t\}}{(1 - \theta t) \sum_{i=1}^{m} I\{F_i < t\}} - \frac{\sum_{i=1}^{m} I\{S_i < t\} I\{H_i^{u/n} \text{ is true}\}}{\sum_{i=1}^{m} I\{F_i < t\}} \right)$$

$$\geq \lim_{m \to \infty} \left( \frac{\sum_{i=1}^{m} I\{F_i < \delta\}}{m} \right) \inf_{t \ge \delta} \left( t \cdot \frac{\sum_{i=1}^{m} I\{F_i < t\} I\{S_i \ge \theta t\}}{(1 - \theta t) \sum_{i=1}^{m} I\{F_i < t\}} - \frac{\sum_{i=1}^{m} I\{S_i < t\} I\{H_i^{u/n} \text{ is true}\}}{\sum_{i=1}^{m} I\{F_i < t\}} \right)$$

$$\geq \lim_{m \to \infty} \left( \frac{\sum_{i=1}^{m} I\{F_i < \delta\}}{m} \right) \inf_{t \ge \delta} \left( \frac{\pi_0 \tilde{S}_0(t) m}{\sum_{i=1}^{m} I\{F_i < t\}} - \frac{\sum_{i=1}^{m} I\{S_i < t\} I\{H_i^{u/n} \text{ is true}\}}{\sum_{i=1}^{m} I\{F_i < t\}} \right)$$

$$\geq \lim_{m \to \infty} \left( -\frac{\sum_{i=1}^{m} I\{F_i < \delta\}}{m} \right) \sup_{t \ge \delta} \left| \frac{\pi_0 \tilde{S}_0(t) m}{\sum_{i=1}^{m} I\{F_i < t\}} - \frac{\sum_{i=1}^{m} I\{S_i < t\} I\{H_i^{u/n} \text{ is true}\}}{\sum_{i=1}^{m} I\{F_i < t\}} \right|$$

$$= 0$$

where the second inequality is a result of Lemma B.1 and the last equality is a consequence of Lemma B.2.  $\Box$ 

### B.4. Proof of Theorem 3.3.

For any  $t \in (0,1]$ , we have by Assumption 1 that

$$\lim_{m \to \infty} t \cdot \frac{\sum_{i=1}^{m} I\{F_i < t\} I\{S_i \ge \theta t\}}{m(1 - \theta t)} = \lim_{m \to \infty} t \cdot \frac{\sum_{i=1}^{m} I\{F_i < t\} (1 - I\{S_i < \theta t\})}{m(1 - \theta t)}$$

$$= \lim_{m \to \infty} t \cdot \frac{\sum_{i=1}^{m} I\{F_i < t\} - \sum_{i=1}^{m} I\{F_i < t\} I\{S_i < \theta t\}}{m(1 - \theta t)}$$

$$= \lim_{m \to \infty} t \cdot \frac{\sum_{i=1}^{m} I\{F_i < t\} - \sum_{i=1}^{m} I\{S_i < \theta t\}}{m(1 - \theta t)}$$

$$= t \cdot \frac{(\pi_0 \tilde{F}_0(t) + \pi_1 \tilde{F}_1(t)) - (\pi_0 \tilde{S}_0(\theta t) + \pi_1 \tilde{S}_1(\theta t))}{1 - \theta t}$$

$$= t \cdot \frac{\tilde{F}(t) - \tilde{S}(\theta t)}{1 - \theta t}$$
(B.11)

with probability 1, where  $\tilde{F}(t)$  and  $\tilde{S}(t)$  are as defined in (B.6). By (3.9) in Assumption 1 and (B.11), there exists a t' > 0 such that

(B.12) 
$$\alpha \cdot \omega - t' \cdot \frac{\tilde{F}(t') - \tilde{S}(\theta t')}{1 - \theta t'} = \frac{\epsilon}{2}$$

for some constant  $\epsilon > 0$ . For a sufficiently large m, we also have by (3.9) in Assumption 1 and (B.11) that

$$\left| t' \cdot \frac{\tilde{F}(t') - \tilde{S}(\theta t')}{1 - \theta t'} - t' \cdot \frac{\sum_{i=1}^{m} I\{F_i < t'\} I\{S_i \ge \theta t'\}}{m(1 - \theta t')} \right| < \frac{\epsilon}{2},$$

which then implies from (B.12) that

$$t' \cdot \frac{\sum_{i=1}^{m} I\{F_i < t'\}I\{S_i \ge \theta t'\}}{m(1-\theta t')} < \alpha \cdot \omega \implies t' \cdot \frac{\sum_{i=1}^{m} I\{F_i < t'\}I\{S_i \ge \theta t'\}}{1-\theta t'} < \alpha \cdot k.$$

Then by the definition of  $\hat{t}_{\theta}$  in (3.2), it follows that

$$\lim_{m\to\infty} \hat{t}_{\theta} \geq t'$$

with probability 1. The display above and Lemma B.3 then gives the following result:

$$\lim_{m \to \infty} \inf \left( \hat{t}_{\theta} \cdot \frac{\sum_{i=1}^{m} I\{F_{i} < \hat{t}_{\theta}\} I\{S_{i} \ge \theta \hat{t}_{\theta}\}}{m(1 - \theta \hat{t}_{\theta})} - \frac{\sum_{i=1}^{m} I\{S_{i} < \hat{t}_{\theta}\} I\{H_{i}^{u/n} \text{ is true}\}}{m} \right)$$

$$(B.13) \qquad \geq \lim_{m \to \infty} \inf_{t \ge \delta} \left( t \cdot \frac{\sum_{i=1}^{m} I\{F_{i} < t\} I\{S_{i} \ge \theta t\}}{m(1 - \theta t)} - \frac{\sum_{i=1}^{m} I\{S_{i} < t\} I\{H_{i}^{u/n} \text{ is true}\}}{m} \right) \ge 0$$

for  $\delta = t'/2$ . Since it holds for any m that

$$\hat{t}_{\theta} \cdot \frac{\sum_{i=1}^{m} I\{F_i < \hat{t}_{\theta}\} I\{S_i \ge \theta \hat{t}_{\theta}\}}{1 - \theta \hat{t}_{\theta}} \le k \cdot \alpha,$$

by the definition of  $\hat{t}_{\theta}$  in (3.2), it must follow from (B.13) that

$$\lim_{m \to \infty} \inf \left( \frac{k \cdot \alpha}{m} - \frac{\sum_{i=1}^{m} I\{S_i < \hat{t}_{\theta}\} I\{H_i^{u/n} \text{ is true}\}}{m} \right) \ge 0.$$

We can therefore conclude that

$$\lim_{m \to \infty} \sup \left( \frac{\sum_{i=1}^m I\{S_i < \hat{t}_\theta\} I\{H_i^{u/n} \text{ is true}\}}{m} \right) \leq \omega \cdot \alpha.$$

By Fatou's lemma, we have

$$\limsup_{m \to \infty} \left( \mathbb{E}\left[ \frac{\sum_{i=1}^{m} I\{S_i < \hat{t}_{\theta}\} I\{H_i^{u/n} \text{ is true}\}}{m} \right] \right) \leq \mathbb{E}\left[ \limsup_{m \to \infty} \left( \frac{\sum_{i=1}^{m} I\{S_i < \hat{t}_{\theta}\} I\{H_i^{u/n} \text{ is true}\}}{m} \right) \right] \leq \omega \cdot \alpha.$$

It then follows from the above display and Markov's inequality that

$$\limsup_{m \to \infty} \Pr\left(k\text{-FWER}(\mathcal{R}) \ge k\right) = \limsup_{m \to \infty} \Pr\left(\sum_{i=1}^{m} I\{S_i < \hat{t}_{\theta}\} I\{H_i^{u/n} \text{ is true}\} \ge k\right)$$

$$\leq \limsup_{m \to \infty} \left(\mathbb{E}\left[\frac{\sum_{i=1}^{m} I\{S_i < \hat{t}_{\theta}\} I\{H_i^{u/n} \text{ is true}\}}{k}\right]\right)$$

$$= \limsup_{m \to \infty} \left(\mathbb{E}\left[\frac{\sum_{i=1}^{m} I\{S_i < \hat{t}_{\theta}\} I\{H_i^{u/n} \text{ is true}\}}{m \cdot \omega}\right]\right)$$

$$\leq \frac{\omega \cdot \alpha}{\omega}$$

$$= \alpha.$$

# APPENDIX C. ADDITIONAL SIMULATION RESULTS

Additional simulation results comparing the methods under k-FWER control for k=5 and 10, using the same  $\pi_1$  and  $\rho$  settings in Section 4.3, are provided in Figure C.1 and Figure C.2 respectively. Note that methods (e) and (f) are designed for FWER control and are therefore omitted from these comparisons.

In these additional simulations, the observed k-FWER levels for k = 5 and k = 10 across all methods remain close to zero. AdaFilter-AdaBon exhibits the least conservative control, most prominently when  $\rho = 0.8$  (large positive correlation) and k = 5. Regarding power, AdaFilter-AdaBon surpasses the other methods by a clear margin across all simulation settings, with the improvements over the Bonferroni and Hochberg procedures being especially pronounced when u = 4.

### APPENDIX D. ADDITIONAL ERROR CONTROL PROOFS

#### D.1. Proof of Corollary 5.1.

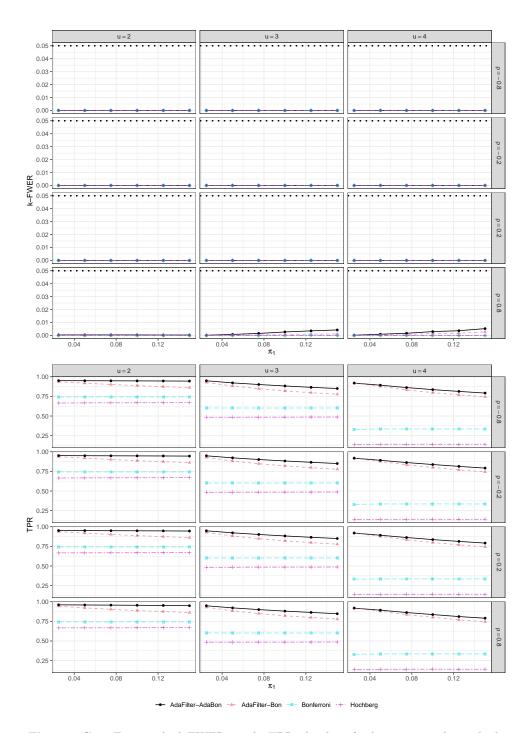
We have that

$$FDX(\mathcal{R}) = \Pr\left(\frac{\sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\}I\{H_{i}^{u/n} \text{ is true}\}}{1 \vee \sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\}} \geq \gamma\right)$$

$$\leq \Pr\left(\frac{\sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\}I\{H_{i}^{u/n} \text{ is true}\}}{1 \vee \sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\}} \geq \frac{\sum_{i=1}^{m} I\{\hat{t}_{\theta} \leq S_{i} \leq \hat{\tau}\} + k}{1 \vee \sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\}}\right)$$

$$= \Pr\left(\sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\}I\{H_{i}^{u/n} \text{ is true}\} \geq \sum_{i=1}^{m} I\{\hat{t}_{\theta} \leq S_{i} \leq \hat{\tau}\} + k\right)$$

$$\leq \Pr\left(\sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\}I\{H_{i}^{u/n} \text{ is true}\} \geq \sum_{i=1}^{m} I\{\hat{t}_{\theta} \leq S_{i} \leq \hat{\tau}\}I\{H_{i}^{u/n} \text{ is true}\} + k\right)$$



 $\textbf{Figure} \quad \textbf{C.1.} \ \, \textbf{Empirical} \quad k\text{-FWER} \quad \text{and} \quad \textbf{TPR} \quad \textbf{levels} \quad \textbf{of} \quad \textbf{the} \quad \textbf{compared} \quad \textbf{methods}$ for the parameter settings  $k=5,~\alpha=0.05,~u\in\{2,3,4\},~\pi_1\in$  $\{0.025, 0.05, 0.075, 0.10, 0.125, 0.15\}, \text{ and } \rho \in \{-0.8, -0.2, 0.2, 0.8\}.$ 

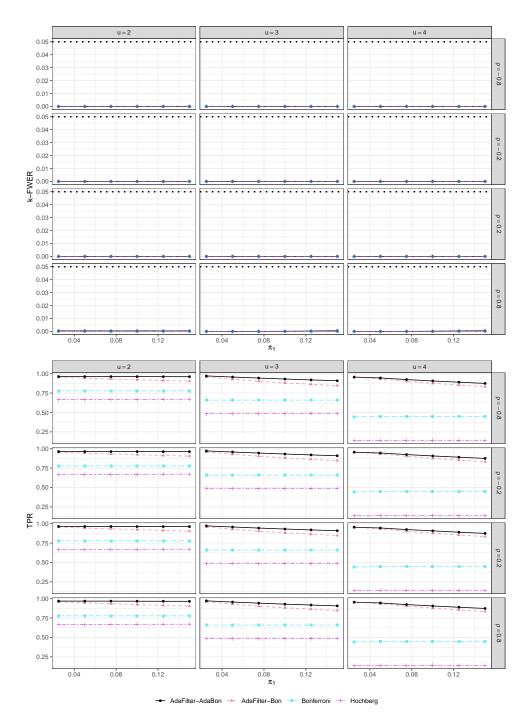


Figure C.2. Empirical k-FWER and TPR levels of the compared methods for the parameter settings  $k=10,~\alpha=0.05,~u\in\{2,3,4\},~\pi_1\in$  $\{0.025, 0.05, 0.075, 0.10, 0.125, 0.15\}, \text{ and } \rho \in \{-0.8, -0.2, 0.2, 0.8\}.$ 

$$\leq \Pr\left(\sum_{i=1}^{m} I\{S_i < \hat{t}_{\theta}\}I\{H_i^{u/n} \text{ is true}\} \geq k\right)$$
  
=  $k\text{-FWER}(\mathcal{R})$ 

where the first inequality follows from (5.1). Given the display above and Theorem 3.3, we have that

(D.1) 
$$\limsup_{m \longrightarrow \infty} \mathrm{FDX}(\mathcal{R}) \leq \limsup_{m \longrightarrow \infty} k\text{-FWER}(\mathcal{R}) \leq \alpha,$$

which proves FDX control. To prove FDR control, note that

$$\begin{aligned} \operatorname{FDR}(\mathcal{R}) &= \mathbb{E}\left[\frac{\sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\} I\{H_{i}^{u/n} \text{ is true}\}}{1 \vee \sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\}}\right] \\ &\leq \mathbb{E}\left[\frac{\sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\} I\{H_{i}^{u/n} \text{ is true}\}}{1 \vee \sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\}} \cdot I\left\{\frac{\sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\} I\{H_{i}^{u/n} \text{ is true}\}}{1 \vee \sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\}} \leq \gamma\right\}\right] \\ &+ \mathbb{E}\left[\frac{\sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\} I\{H_{i}^{u/n} \text{ is true}\}}{1 \vee \sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\}} \cdot I\left\{\frac{\sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\} I\{H_{i}^{u/n} \text{ is true}\}}{1 \vee \sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\}} \geq \gamma\right\}\right] \\ &\leq \gamma + \mathbb{E}\left[I\left\{\frac{\sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\} I\{H_{i}^{u/n} \text{ is true}\}}{1 \vee \sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\}}\right\}\right] \\ &= \gamma + \operatorname{Pr}\left(\frac{\sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\} I\{H_{i}^{u/n} \text{ is true}\}}{1 \vee \sum_{i=1}^{m} I\{S_{i} \leq \hat{\tau}\}} \geq \gamma\right) \\ &= \gamma + \operatorname{FDX}(\mathcal{R}). \end{aligned}$$

The result then immediately follows from (D.1).