

# A central limit theorem for the Benjamini-Hochberg false discovery proportion under a factor model

Dan M. Kluger<sup>1</sup>, and Art B. Owen<sup>1</sup>

<sup>1</sup>*Department of Statistics, Stanford University, e-mail:*  
kluger@stanford.edu; owen@stanford.edu

**Abstract:** The Benjamini-Hochberg (BH) procedure remains widely popular despite having limited theoretical guarantees in the commonly encountered scenario of correlated test statistics. Of particular concern is the possibility that the method could exhibit bursty behavior, meaning that it might typically yield no false discoveries while occasionally yielding both a large number of false discoveries and a false discovery proportion (FDP) that far exceeds its own well controlled mean. In this paper, we investigate which test statistic correlation structures lead to bursty behavior and which ones lead to well controlled FDPs. To this end, we develop a central limit theorem for the FDP in a multiple testing setup where the test statistic correlations can be either short-range or long-range as well as either weak or strong. The theorem and our simulations from a data-driven factor model suggest that the BH procedure exhibits severe burstiness when the test statistics have many strong, long-range correlations, but does not otherwise.

**MSC2020 subject classifications:** Primary 62J15; secondary 60F17.

**Keywords and phrases:** functional central limit theorem, functional delta method, empirical cumulative distribution function, multiple hypothesis testing, Simes line.

## 1. Introduction

The Benjamini-Hochberg (BH) procedure is a widely used method for balancing Type I and Type II errors when testing many hypotheses simultaneously. The procedure is designed to control the False Discovery Rate (FDR), which is the *expected value* of the proportion of discoveries that are false (FDP), below a user specified threshold (Benjamini and Hochberg (1995)). The procedure was originally shown to guarantee FDR control when the test statistics are assumed to be independent, an assumption unlikely to hold in most application settings. The BH procedure was later proven in Benjamini and Yekutieli (2001) to control FDR when there are dependent test statistics satisfying the Positive Regression Dependency (PRDS) property. While PRDS is quite restrictive (for example, it does not hold for two-sided hypothesis tests when the test statistics are correlated or when there are negatively correlated test statistics (Fithian and Lei (2020))), under more general conditions simulation studies have found BH to conservatively control FDR (Farcomeni (2006), Kim and van de Wiel (2008)).

While FDR control is important, the motivation for this paper is our concern that FDR control alone can give investigators who use BH false confidence in a low prevalence of false discoveries among their rejected hypothesis. This can happen if the distribution of the FDP has both a wide right tail and a mean that is still below the user specified threshold. As an example, it would be worrisome in the plausible scenario that an investigator is led to believe that roughly 10 percent of their discoveries are false, when in fact, a majority of them are false. To address such concerns, a number of multiple testing procedures have been proposed to control the tail probability that the FDP exceeds a user specified threshold (Korn et al. (2004), Romano and Shaikh (2006), Romano and Wolf (2007)). Efron (2007) also raised concerns about high variability of FDP due to correlations of the test statistics and proposed an empirical Bayes approach for estimating a dispersion parameter of the test statistics and controlling FDR conditionally on the dispersion parameter. Despite the promise of these methods, the BH procedure remains overwhelmingly popular and the default method of choice for investigators with multiple testing problems. It is therefore important to determine under which conditions are we assured that the distribution of the FDP will be well concentrated about its mean, the FDR, and under which conditions there is a risk that the distribution of the FDP has a wide right tail. Throughout this text we will refer to the former scenario with low variability of the FDP about its mean as the non-bursty regime, and the alarming, latter scenario where occasionally the FDP is much larger than expected as the bursty regime.

The aim of this paper is to understand when burstiness is a concern for BH. We identify dependency structures among test statistics that in conjunction with certain proportions of nonnulls make BH prone to delivering bursts of false discoveries. We find other settings where such bursts must be rare. Our results are asymptotic and hold in a two-group mixture model previously studied by Genovese and Wasserman (2004), Delattre and Roquain (2016) and Izmirlian (2020). In that model, independent  $\text{Bern}(\pi_1)$  variables define which hypotheses are nonnull, the null  $p$ -values have the  $\text{Unif}(0, 1)$  distribution and the nonnull  $p$ -values have some other distribution in common.

The BH procedure and the asymptotic distribution of the FDP is well studied for the setting where the test statistics are independent. Finner and Roters (2001, 2002) study properties of the number of false discoveries under independence both when there are no nonnulls and when the tests of nonnull hypotheses always have  $p$ -values of 0. Using an empirical process approach which we build upon, the limiting distribution of the FDP has been studied by Genovese and Wasserman (2004) under independence of the test statistics and was further extended in Farcomeni (2007) to the setting where the sequence of  $p$ -values is stationary and satisfies some mixing conditions. However, these asymptotic FDP distributions are derived for the “plug-in” method (Benjamini and Hochberg (2000)) rather than the standard BH procedure. For the BH procedure itself, to our knowledge, a central limit theorem (CLT) for the FDP was first found in Neuvial (2008) with a correction to the asymptotic variance formula by Izmirlian (2020). These works showed that in the two group mixture model with Bernoulli

parameter  $\pi_1$  and with  $m$  hypothesis to test, when using the BH procedure at FDR control level  $q$ ,  $\sqrt{m}(\text{FDP} - (1 - \pi_1)q)$  converges as  $m \rightarrow \infty$  to a centered Gaussian with variance that depends on  $q$ ,  $\pi_1$ , and the common nonnull  $p$ -value distribution.

There are fewer results on the limiting distribution of the FDP for dependent test statistics. Using the proof methodology of Neuvial (2008), Delattre and Roquain (2011) derive a CLT for the FDP of the BH procedure for one sided testing, when the test statistics follow an equi-correlated Gaussian model, with correlation parameter  $\rho \rightarrow 0$  as the number of tests  $m \rightarrow \infty$ . Delattre and Roquain (2016) extend this result to settings where the Gaussian test statistics follow arbitrary dependence structures but the average pairwise correlation of the test statistics, and the average 2nd and 4th powers of the pairwise correlations of the test statistics satisfy some constraints.

In Delattre and Roquain (2016), CLTs for the FDP are derived under two distinct regimes. In their first regime, the average pairwise correlation among test statistics is strictly greater than  $\mathcal{O}(1/m)$  for  $m$  tests. Under this regime, the FDP is not  $\sqrt{m}$ -consistent for the product of the FDR control parameter with the limiting proportion of nulls, and the FDP only converges to a Gaussian with scale factors much smaller than  $\sqrt{m}$ . Their other regime considered has an average correlation among test statistics that is at most  $\mathcal{O}(1/m)$ . For this regime Delattre and Roquain (2016) derive a CLT for the FDP with  $\sqrt{m}$  scaling, but they require a restrictive assumption which they call “vanishing-second order”, precluding settings where there are short-range correlations of constant order. Examples of test statistic correlation matrices to which Delattre and Roquain (2016) will not apply include tridiagonal Toeplitz correlation matrices as well as a block correlation matrices of constant block size, both of which are simple models of interest for studying multiple testing under dependence.

With the aim of identifying dependency structures for which the investigator should be concerned about bursty behavior of the BH procedure, in this paper, we introduce a model that allows for a combination of long-range and potentially strong dependence among the test statistics via a factor model, along with additional strongly-mixing noise that has rapidly decaying long-range dependence. The model also allows for the proportion of nonnulls among the  $m$  hypothesis tests to vary as a function of  $m$ . Under some regularity conditions on the factor model and on the noise with rapidly decaying long-range dependence, we prove a CLT for the FDP under more general conditions than prior CLTs. Our CLT proof structure extends that of Delattre and Roquain (2016). We also establish a CLT for the False Positive Ratio (FPR), which is the proportion false discoveries among all tests conducted. The new CLTs hold conditionally on the realized latent variable of the factor model. Applying these new results, we make the following contributions to the literature of asymptotic results for the BH procedure:

1. CLTs of the FDP for simple models, with short-range and constant-order dependency structures, that were not covered by the results of Delattre and Roquain (2016). Examples include block correlation structures with

fixed block size and banded correlation structures (e.g. tridiagonal Toeplitz correlation structure).

2. Conditional CLTs in settings where the long-range dependency is modeled by a factor model which includes scenarios not covered in Delattre and Roquain (2016).
3. CLTs for the FPR rather than just the FDP because the FDP limiting behavior is unilluminating when it converges in probability to 1.
4. CLTs where the expected proportion of nonnulls varies as the number of test statistics grows, allowing for a sparse allocation of nonnulls.
5. A discussion of the dependency regimes under which the investigator should be concerned about BH having bursty behavior, such as the setting where the number of nonnulls is  $o_p(\sqrt{m})$ , and the dependency structure contains a factor model component.

To qualify point 2 above, we note that Delattre and Roquain (2016) include some CLTs for the FDP that our theorems do not. Ours all have a  $\sqrt{m}$  scaling. They include some with a slower than  $\sqrt{m}$  scaling. For instance, they get such a CLT under an equicorrelated Gaussian model with correlation  $\rho \rightarrow 0$  but  $\sqrt{m}\rho \rightarrow \infty$ .

Figure 1 shows some simulations of the FDP under models that we study in this paper. In each case, there are 25,000 Monte Carlo simulations. The BH procedure is used with  $q = 0.1$  on test statistics that are  $\mathcal{N}(0, 1)$  for null hypotheses and  $\mathcal{N}(2, 1)$  for alternative hypotheses. Each hypothesis is independently null with probability 0.9 and nonnull otherwise. The models differ in the correlation among test statistics. For the first histogram,  $m = 22,283$  test statistics were sampled with correlations based on a 3-factor model fit to some Duchenne Muscular Dystrophy data described in Section 6. The second histogram is for the same correlation matrix after dividing the off-diagonal entries by 10. Next are two block correlation models with blocks of size 100 and within-block correlations of 0.05 or 0.5. To keep  $m = 22,283$  one of the blocks had only 83 test statistics in it. In all four settings the FDR is seen to be controlled below 0.1, as desired. The positive False Discovery Rate (pFDR), defined as the expected value of the FDP conditional on there being at least one rejection, does not exceed 0.1 in these simulations either. Of the four histograms, one shows a long tailed distribution for FDP that we consider extremely bursty, two show FDPs that are typically quite close to the target FDR of  $q = 0.1$ , and the fourth one, with downsampled correlations is intermediate.

The organization of this paper is as follows. In Section 2, we describe our multiple testing setup and our model to account for both long-range correlations and short-range correlations among the test statistics. We also introduce our notation, definitions and the conditions under which our results hold. In Section 3, we state our most general CLTs for the FDP and FPR of the BH procedure. These hold conditionally on the common latent factors in our model. The proofs of these theorems are provided in the Appendix. In Section 4, we exploit these theorems to obtain FDP CLTs (that are not conditional on a latent factor) for settings where the long-range dependence is rapidly decaying and no factor

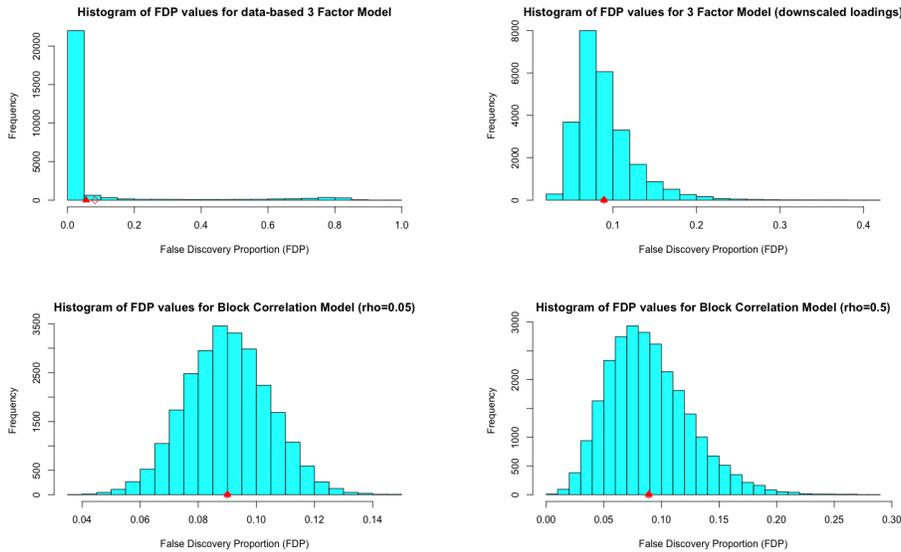


FIG 1. *These are histograms of the false discovery proportion in 25,000 simulations. The data come from the two-group mixture model as described in the text. Each histogram's mean is marked with a triangle and each histogram's mean amongst nonzero FDP values is marked with a hollow diamond, which estimate the FDR and  $pFDR$  respectively. The target FDR control is  $q = 0.1$ .*

model component is needed to account for long-range dependencies. In Section 5, we exploit these theorems to obtain FDP and FPR CLTs conditional on the latent factor that give insight into the burstiness of the BH procedure, and we show that the burstiness of the BH procedure is particularly alarming when the test statistics follow a factor model and the number of nonnulls is sparse (for example, if the the number of nonnulls is  $o_p(\sqrt{m})$  where  $m$  is the number of hypothesis tested). In Section 6, we describe the Duchenne Muscular Dystrophy dataset and the 3 factor model that was fit to it, and we show simulations based on the fitted factor model. In Section 7, we discuss these results and their implications for multiple testing.

## 2. Setup and Definitions

In this section we introduce our notation for the two-group mixture model. Our version relies on a factor analysis model that we also introduce. We also review the BH procedure and state our regularity conditions in this section.

### 2.1. Two-group mixture model with factors

Our setting has  $m$  hypothesis tests indexed by  $i = 1, \dots, m$  and our asymptotics let  $m \rightarrow \infty$ . For  $1 \leq i \leq m < \infty$ , let  $H_{mi} \in \{0, 1\}$  be an indicator

variable with  $H_{mi} = 1$  if and only if hypothesis  $i$  of  $m$  is nonnull. We take  $H_{m1}, \dots, H_{mm} \stackrel{\text{iid}}{\sim} \text{Bern}(\pi_1^{(m)})$  for  $\pi_1^{(m)} \in (0, 1)$ . Our two-group mixture model is thus based on a sequence of nonnull probabilities and letting  $\pi_1^{(m)} \rightarrow 0$  will let us model sparsity of nonnull hypotheses. For instance, with  $\pi_1^{(m)} = \lambda/m$ , the number of nonnulls has constant expectation  $\lambda$  and has an asymptotic Poisson distribution.

To focus on dependency among tests it is convenient to assume Gaussian test statistics  $X_{mi}$  for  $1 \leq i \leq m < \infty$ . The test statistics where the null holds have mean zero and the ones where the alternative hypothesis holds have common mean  $\mu_A > 0$ . We assume that  $X_{mi} = \mu_A H_{mi} + Z_{mi}$  where  $(Z_{m1}, \dots, Z_{mm})$  is multivariate Gaussian, and we induce dependence among our  $p$ -values by introducing correlations among the  $Z_{mi}$ .

We study two kinds of dependence operating simultaneously. One is an  $\alpha$ -mixing dependence that decays rapidly as the distance between hypothesis indices  $i$  increases. This model captures some of the dependence one expects from hypotheses corresponding to a linearly ordered variable such as the position of a single nucleotide polymorphism (SNP) along the genome.

The other form of dependence we include is a factor model. Owen (2005) gives conditions where the correlation matrix for  $m$  test statistics measuring association of a single phenotype with expression levels of  $m$  genes is actually equal to the correlation matrix of the sampled gene measurements. A factor model can capture important aspects of such a matrix. More generally we might suppose that there are possibly sparse factors affecting both genotype and phenotype and then nonnull hypotheses correspond to one or more shared factors. Some other uses of factor analysis models in multiple hypothesis testing include Friguet et al. (2009), Lucas et al. (2010), Sun et al. (2012) and Gerard and Stephens (2020).

We construct the  $k$ -factor model for the array  $\{Z_{mi} : 1 \leq i \leq m < \infty\}$  as follows. We let  $\mathbf{W} \sim \mathcal{N}(0, I_k)$  remain constant as  $m \rightarrow \infty$  and let  $\{\mathbf{L}_{mi} : 1 \leq i \leq m < \infty\}$  be a triangular array of fixed ‘loading’ vectors in  $\mathbb{R}^k$ . The factor model component of  $Z_{mi}$  is  $\mathbf{L}_{mi}^\top \mathbf{W}$ . For our  $\alpha$ -mixing model with possibly strong short-range correlations but rapidly diminishing long-range correlations, we let  $\{\Sigma^{(m)}\}_{m=1}^\infty$  be a sequence of covariance matrices and for each  $m$ , we let  $(\varepsilon_{m1}, \dots, \varepsilon_{mm}) \sim \mathcal{N}(0, \Sigma^{(m)})$ . To combine both dependency structures, we let  $Z_{mi} = \mathbf{L}_{mi}^\top \mathbf{W} + \varepsilon_{mi}$ , giving our correlation structure for the array  $\{Z_{mi} : 1 \leq i \leq m < \infty\}$ .

We suppose that all of the test statistics have the same variance and without loss of generality, we take this common variance to be one. We do not assume the factor model to have a perfect fit, and assume instead that  $\|\mathbf{L}_{mi}\|_2^2 + \Sigma_{ii}^{(m)} = 1$  where  $\Sigma_{ii}^{(m)} > 0$  for all  $i, m$ . Because  $Z_{mi} = \mathbf{L}_{mi}^\top \mathbf{W} + \varepsilon_{mi}$ , these assumptions give  $Z_{m1}, \dots, Z_{mm} \sim \mathcal{N}(0, 1)$  along with the two kinds of dependency discussed above.

We let  $\varphi$  and  $\Phi$  denote the probability density function (PDF) and the cumulative distribution function (CDF), respectively, of  $\mathcal{N}(0, 1)$  and we let  $\bar{\Phi} = 1 - \Phi$  be the complementary CDF. Then our  $p$ -values for one-sided hypothesis tests

are

$$P_{mi} = \bar{\Phi}(X_{mi}) = \bar{\Phi}(\mu_A H_{mi} + \mathbf{L}_{mi}^T \mathbf{W} + \varepsilon_{mi}) \quad (1)$$

for  $1 \leq i \leq m < \infty$ , and so  $P_{mi} \sim \text{Unif}(0,1)$  for the true null hypotheses. Fixing  $q \in (0,1)$ , throughout the text we will let  $\tau_{\text{BH},m}$ ,  $V_m$ ,  $\text{FDP}_m$ , and  $\text{FPR}_m$  denote the rejection threshold, the number of false discoveries, the FDP, and the FPR respectively when applying the Benjamini-Hochberg procedure at level  $q$  to the  $p$ -values  $(P_{m1}, \dots, P_{mm})$ . The formulas for these quantities are given explicitly in the next subsection, where we review the BH procedure. In our main theorems, we state CLTs for the quantities  $\text{FDP}_m$  and  $\text{FPR}_m$  conditionally on the value of the latent factor  $\mathbf{W} = \mathbf{w} \in \mathbb{R}^k$ .

## 2.2. The BH procedure

Here we describe how the BH procedure is conducted at level  $q$  on  $m$  tests with  $p$ -values  $(P_{m1}, \dots, P_{mm})$ . First take the sorted  $p$ -values  $P_{m(1)} \leq \dots \leq P_{m(m)}$  and set  $P_{m(0)} = 0$ . The number of rejected hypothesis will be given by

$$R_m \equiv \max\{j : P_{m(j)} \leq \frac{jq}{m}, j \in \{0, 1, \dots, m\}\}. \quad (2)$$

The BH procedure rejects the hypotheses that correspond to the  $R_m$  smallest  $p$ -values: that is it will reject all hypothesis  $i$  for which  $P_{mi} \leq P_{m(R_m)} \equiv \tau_{\text{BH},m}$ . As noted in Neuvial (2008),  $\tau_{\text{BH},m}$  can equivalently be defined as the largest  $t \in [0, 1]$  at which the empirical CDF (ECDF) of the  $p$ -values is at least as large as  $t/q$ . We leverage this equivalence in our theorem proofs.

Letting  $H_{m1}, \dots, H_{mm}$  be as defined in Section 2.1, the number of false discoveries is

$$V_m \equiv \sum_{i=1}^m I\{P_{mi} \leq \tau_{\text{BH},m}, H_{mi} = 0\}. \quad (3)$$

Then  $\text{FPR}_m \equiv V_m/m$  and  $\text{FDP}_m \equiv V_m/\max\{R_m, 1\}$ .

## 2.3. Definitions and Conditions

Here we present some definitions as well as regularity conditions sufficient for our conditional CLTs to hold. All of the definitions, conditions and formulas in this section are conditional on a fixed value of the latent factor  $\mathbf{W} \in \mathbb{R}^k$ .

### 2.3.1. Long-range dependency conditions

For  $1 \leq i \leq m < \infty$ , define  $\xi_{mi} = (\varepsilon_{mi}, H_{mi})$ . Now let  $\mathcal{A}_1^n(m)$  be the  $\sigma$ -field generated by the variables  $\xi_{mi}$  for  $1 \leq i \leq n$  and  $\mathcal{A}_{n+d}^\infty(m)$  be the  $\sigma$ -field

generated by the variables  $\xi_{mi}$  for  $n + d \leq i \leq m$ . For integers  $d \geq 1$  our  $\alpha$ -mixing parameters  $\alpha(d) \in [0, 1]$  are defined by

$$\alpha(d) \equiv \sup_{n, m \in \mathbb{N}} \sup_{\substack{A_0 \in \mathcal{A}_1^n(m) \\ A_1 \in \mathcal{A}_{n+d}^\infty(m)}} |P(A_0 \cap A_1) - P(A_0)P(A_1)|.$$

**Condition 1.** *There exists an even integer  $Q > 2$  and  $\gamma > 0$  such that both*

$$(i) \quad \frac{\gamma}{2 + \gamma} + \frac{2}{Q} < 1 \quad \text{and} \quad (ii) \quad \sum_{d=1}^{\infty} d^{Q-2} \alpha(d)^{\frac{\gamma}{Q+\gamma}} < \infty$$

**Notes about Condition 1:** Throughout the text we will let  $Q, \gamma$  be such numbers. Note that it is possible that this condition can be loosened to allow  $Q$  to be rational, but then we need to trust a claim in Andrews and Pollard (1994) that their Theorem 2.2 would still hold for  $Q$  not an even integer. Condition 1 will hold when the correlation between  $\varepsilon_{mi}$  and  $\varepsilon_{mj}$  is a rapidly decaying function of  $|i - j|$ . If this correlation is always zero for each  $|i - j| > M$  (making the error sequences  $(\varepsilon_{mi})_{1 \leq i \leq m}$   $M$ -dependent for each  $m$ ), Condition 1 will hold.

For any non-empty  $T \subseteq [m] \equiv \{1, 2, \dots, m\}$ , define  $\mathcal{F}_{mT} = \sigma(\{\xi_{mi}\}_{i \in T})$ . For non-empty  $T, S \subseteq [m]$  let

$$R(T, S, m) \equiv \sup_{\substack{f \in L_2(\mathcal{F}_{mT}) \\ g \in L_2(\mathcal{F}_{mS})}} |\text{corr}(f, g)|$$

be the maximal correlation coefficient between the  $\sigma$ -algebras  $\mathcal{F}_{mT}$  and  $\mathcal{F}_{mS}$ . For non-empty  $T, S \subseteq [m]$ , let  $\text{dist}(T, S) = \min\{|i - j| : i \in T, j \in S\}$ . Define

$$\rho_*(d) \equiv \sup_{m \in \mathbb{N}} \sup_{\substack{T, S \subseteq [m] \\ \text{dist}(T, S) \geq d \\ T, S \neq \emptyset}} R(T, S, m) \quad (4)$$

to be the interlaced  $\rho$ -mixing coefficient of the triangular array  $\{\xi_{mi}, 1 \leq i \leq m < \infty\}$ .

**Condition 2.**  $\lim_{d \rightarrow \infty} \rho_*(d) < 1$ .

**Notes about Condition 2:**  $\{\rho_*(d)\}_{d=1}^{\infty}$  is a monotone non-increasing sequence of elements in  $[0, 1]$ . In checking whether the condition holds, note that the limit will always exist. While we have not proven a relationship between Conditions 1 and 2, we do not expect there to be any common situations where Condition 1 holds but Condition 2 does not. If for some  $M$ ,  $\text{corr}(\varepsilon_{mi}, \varepsilon_{mj}) = 0$  whenever  $|i - j| > M$ , then by  $M$ -dependence Condition 2 will hold.

We now mention another condition that forces the variance of all  $\varepsilon_{mi}$  terms to be bounded away from zero.

**Condition 3.**  $S_L \equiv \sup_{1 \leq i \leq m < \infty} \|\mathbf{L}_{mi}\|_2^2 < 1$ .

**Note about Condition 3:** recalling that  $\|\mathbf{L}_{mi}\|_2^2 + \text{var}(\varepsilon_{mi}) = 1$  in our model, this condition provides a uniform bound  $\text{var}(\varepsilon_{mi}) \geq 1 - S_L > 0$  for all  $1 \leq i \leq m < \infty$ .

### 2.3.2. Definitions of some subdistributions of $p$ -values and their condition

For any  $m \in \mathbb{N}_+$ , define  $\pi_0^{(m)} \equiv 1 - \pi_1^{(m)}$ , and then write

$$H_{mi0} \equiv 1 - H_{mi} \quad \text{and} \quad H_{mi1} \equiv H_{mi}.$$

Our definitions use  $r = 0$  for quantities based on the null hypotheses and  $r = 1$  for quantities from the nonnull hypotheses. For  $t \in [0, 1]$  and  $r \in \{0, 1\}$  let

$$\hat{F}_{m,r}(t) \equiv \frac{1}{m} \sum_{i=1}^m H_{mir} I\{P_{mi} \leq t\} = \frac{1}{m} \sum_{i=1}^m H_{mir} I\{\bar{\Phi}(\mu_{Ar} + \varepsilon_{mi} + \mathbf{L}_{mi}^\top \mathbf{W}) \leq t\}.$$

We call  $\hat{F}_{m,0}$  and  $\hat{F}_{m,1}$  the empirical subdistribution functions of the null and nonnull  $p$ -values respectively. These empirical subdistribution functions sum to the ECDF of the  $p$ -values. Let  $\gamma_{mir} : [0, 1] \rightarrow [0, 1]$  be the monotone increasing bijection given by

$$\gamma_{mir}(t) \equiv \Pr(P_{mi} \leq t | H_{mi} = r, \mathbf{W} = \mathbf{w}) = \bar{\Phi}\left(\frac{\bar{\Phi}^{-1}(t) - \mu_{Ar} - \mathbf{L}_{mi}^\top \mathbf{w}}{\sqrt{1 - \|\mathbf{L}_{mi}\|_2^2}}\right).$$

We aggregate  $\gamma_{mir}$  in the following subdistribution functions

$$F_{m,r}(t) \equiv \mathbb{E}(\hat{F}_{m,r}(t)) = \frac{\pi_r^{(m)}}{m} \sum_{i=1}^m \gamma_{mir}(t)$$

and then let

$$F_r(t) \equiv \lim_{m \rightarrow \infty} F_{m,r}(t) = \lim_{m \rightarrow \infty} \frac{\pi_r^{(m)}}{m} \sum_{i=1}^m \gamma_{mir}(t). \quad (5)$$

Condition 4 ensures that these quantities are well defined.

**Condition 4.** For all  $t \in [0, 1]$  and  $r \in \{0, 1\}$ ,  $F_r(t) \equiv \lim_{m \rightarrow \infty} F_{m,r}(t)$  exists.

### 2.3.3. Defining the asymptotic ECDF and the Simes point

Now define  $\hat{G}_m, G : [0, 1] \rightarrow [0, 1]$  via

$$\hat{G}_m(t) \equiv \hat{F}_{m,0}(t) + \hat{F}_{m,1}(t) \quad \text{and} \quad G(t) \equiv F_0(t) + F_1(t) \quad (6)$$

for  $t \in [0, 1]$ . Note that  $\hat{G}_m$  is the ECDF of the  $p$ -values and  $G$  is the limiting expected ECDF of the  $p$ -values. Throughout the text we will refer to  $G$  as the

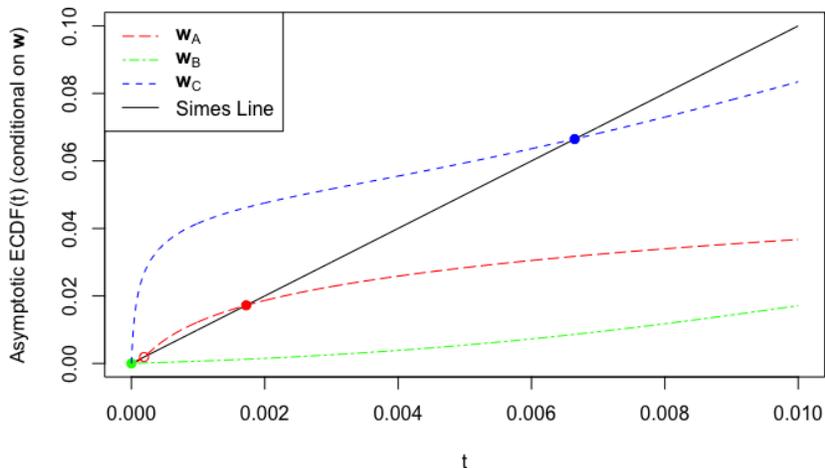


FIG 2. The curves are conditional asymptotic ECDFs of  $p$ -values in a 3-factor model based on some Duchenne Muscular Dystrophy data described in Section 6. The three draws satisfy  $\Phi(\mathbf{w}_A) = (0.8, 0.4, 0.9)$ ,  $\Phi(\mathbf{w}_B) = (0.45, 0.56, 0.62)$  and  $\Phi(\mathbf{w}_C) = (0.02, 0.85, 0.78)$ . Filled circles show the Simes points. An open circle for  $\mathbf{w}_A$  shows a crossing of the Simes line that is not the Simes point because it is not the final crossing.

asymptotic ECDF because under most dependency structures, we expect  $G$  to be the point-wise limit in probability of  $\hat{G}_m$ .

The rejection threshold for the BH procedure at level  $q$  is the largest point  $t$  such that the ECDF of the  $p$ -values evaluated at  $t$  lies above the line through the origin of slope  $1/q$ , called the Simes line. It is reasonable to expect the limiting  $p$ -value rejection threshold for the BH procedure at level  $q$  to be the largest  $t$  at which  $(t, G(t))$  intersects the Simes line. We use the term *Simes point* to describe the largest point where the asymptotic ECDF intersects the Simes line. More precisely, the Simes point is

$$\tau_* \equiv \sup\{t \in (0, 1) : G(t) \geq t/q\}, \quad (7)$$

interpreting the supremum of the empty set to be zero. The Simes point satisfies  $0 \leq \tau_* \leq q$ . The upper limit follows from  $G(t) \leq 1$ . Both  $G$  and  $\tau_*$  depend on the specific realization of latent factor  $\mathbf{W} \in \mathbb{R}^k$  on which we condition.

Figure 2 illustrates the Simes points. The setting has  $\mu_A = 2$ ,  $\pi_0 = 0.9$  and  $q = 0.1$ . The horizontal axis has putative  $p$ -values over the range  $t \in [0, 0.01]$ . The Simes line is  $t/q$ . There are  $m = 22,283$  hypotheses corresponding to genes in the GDS 3027 Duchenne Muscular Dystrophy data described in Section 6. For three draws  $\mathbf{W} \sim \mathcal{N}(0, I_3)$  we show the asymptotic ECDF curves. One of them crosses the Simes line twice and the Simes point is the last crossing. One crosses it only once and one has Simes point  $\tau_* = 0$  because the Simes line is never crossed.

We will need continuity of  $G(\cdot)$  on  $(0, 1)$  under Conditions 3 and 4. We do not know whether  $G$  must be continuous at 0 or 1, but our results do not depend on that.

**Proposition 1.** *Under Conditions 3 and 4,  $G$  is continuous on  $(0, 1)$ .*

*Proof.* It is sufficient to show that  $G$  is Lipschitz continuous on  $(\epsilon, 1-\epsilon)$  whenever  $0 < \epsilon < 1/2$ . For any such  $\epsilon$ , observe that for  $r \in \{0, 1\}$  and integers  $1 \leq i \leq m < \infty$

$$\gamma'_{mir}(t) = \frac{1}{\varphi(\Phi^{-1}(t))\sqrt{1 - \|\mathbf{L}_{mi}\|_2^2}} \varphi\left(\frac{\bar{\Phi}^{-1}(t) - \mathbf{L}_{mi}^\top \mathbf{w} - \mu_{Ar}}{\sqrt{1 - \|\mathbf{L}_{mi}\|_2^2}}\right).$$

Now  $\varphi(\cdot) \leq 1/\sqrt{2\pi}$  and then using Condition 3 it follows that for any  $t \in (\epsilon, 1-\epsilon)$ ,

$$|\gamma'_{mir}(t)| \leq \frac{1/\sqrt{2\pi}}{\varphi(\Phi^{-1}(t))\sqrt{1 - S_L}} \leq \frac{1/\sqrt{2\pi}}{\varphi(\Phi^{-1}(\epsilon))\sqrt{1 - S_L}} \equiv C_\epsilon.$$

Since  $\sup_{t \in (\epsilon, 1-\epsilon)} |\gamma'_{mir}(t)| \leq C_\epsilon < \infty$ ,  $|\gamma_{mir}(t) - \gamma_{mir}(s)| \leq C_\epsilon |t - s|$  for any  $t, s \in (\epsilon, 1-\epsilon)$ . This argument holds for any  $1 \leq i \leq m < \infty$  and  $r \in \{0, 1\}$ , and so for any  $t, s \in (\epsilon, 1-\epsilon)$  and integer  $m$  and  $r \in \{0, 1\}$ ,

$$|F_{m,r}(t) - F_{m,r}(s)| \leq \frac{\pi_r^{(m)}}{m} \sum_{i=1}^m |\gamma_{mir}(t) - \gamma_{mir}(s)| \leq C_\epsilon |t - s|.$$

Taking the limit as  $m \rightarrow \infty$  of the left side of the above inequality, which exists by Condition 4, we get  $|F_r(t) - F_r(s)| \leq C_\epsilon |t - s|$  for all  $t, s \in (\epsilon, 1-\epsilon)$  and for both  $r \in \{0, 1\}$ . Thus,  $F_r$  is Lipschitz continuous on  $(\epsilon, 1-\epsilon)$  for  $r \in \{0, 1\}$  which implies  $G = F_0 + F_1$  is Lipschitz continuous on  $(\epsilon, 1-\epsilon)$ .  $\square$

#### 2.3.4. Defining a focal interval $[a, b] \subset [0, 1]$ for our processes

We are going to work with an interval  $[a, b]$  of positive length for which the Simes point  $\tau_*$  is the unique element  $t \in (a, b)$  with  $G(t) = t/q$ . First we need a technical condition to rule out some pathological behavior. Under this condition there will be a unique point in  $[a, b]$  where  $G$  crosses the Simes line.

**Condition 5.** *The Simes point is positive, is the largest point where  $G$  actually crosses the Simes line, and is not an accumulation point for points of intersection of  $G$  and the Simes line. That is,*

- (i)  $\tau_* > 0$ ,
- (ii)  $\tau_* = \sup\{t \in (0, 1) : G(t) > t/q\}$ , and
- (iii)  $\tau_*$  is not an accumulation point of  $\{t \in (0, 1) : G(t) = t/q\}$ .

**Note about Condition 5:** For many factor model choices, Condition 5 will hold with some probability in  $(0, 1)$  depending on the specific realization of  $\mathbf{W} \sim \mathcal{N}(0, I_k)$ . This is due to the dependence of  $\tau_*$  and  $G$  on the specific realization of the latent factor  $\mathbf{W} \in \mathbb{R}^k$  on which we condition. For example, see Figure 2.

**Proposition 2.** *If Condition 5 holds, then for any  $b \in (q, 1)$  there exists a point  $a \in (0, q)$  such that*

- (i)  $G(a) > a/q$ , and
- (ii) the Simes point  $\tau_*$  is the unique  $t \in (a, b)$  solving  $G(t) = t/q$ .

*Proof.* Pick any  $b \in (q, 1)$  and suppose that Condition 5 holds. Then

$$\tau_* \equiv \sup\{t \in (0, 1) : G(t) \geq t/q\} = \sup\{t \in (0, 1) : G(t) > t/q\} > 0.$$

Now  $\tau_* \leq q$  because  $G(t) \leq 1$  for all  $t \in (0, 1)$ . Also,  $G(\tau_*) = \tau_*/q$  by continuity of  $G$  (see Proposition 1). Hence, because  $\tau_* = \sup\{t \in (0, 1) : G(t) > t/q\}$  but  $G(\tau_*) = \tau_*/q$ , there exists a sequence  $a_n \uparrow \tau_*$  such that for all  $n$ ,  $G(a_n) > a_n/q$  and  $a_n < \tau_*$ . Since  $\{t \in (0, 1) : G(t) = t/q\}$  does not have an accumulation point at  $\tau_*$  and since  $G(t) - t/q$  is continuous, there is a sufficiently large  $N_*$  with  $G(t) > t/q$  for all  $t \in [a_{N_*}, \tau_*)$ . We choose  $a = a_{N_*} \in (0, \tau_*)$  and then property (i) holds by our definition of  $a_n$ . Also,  $a \in (0, q)$  because  $a < \tau_* \leq q$ . Turning to property (ii),  $G(t) > t/q$  for all  $t \in [a, \tau_*)$  by the choice of  $N_*$  and  $a$ , while for all  $t \in (\tau_*, b)$ ,  $G(t) < t/q$  by the definition of  $\tau_*$ .  $\square$

Throughout the text, when conditioning on  $\mathbf{W} = \mathbf{w} \in \mathbb{R}^k$ , if Condition 5 holds we will let  $[a, b]$  be an interval satisfying the properties (i) and (ii) with  $a \in (0, q)$  and  $b \in (q, 1)$  that are guaranteed by Proposition 2.

### 2.3.5. Defining our stochastic process and its Gaussian process limit

Our stochastic processes of interest are two jointly distributed random càdlàg functions on  $[a, b]$ . Such functions are continuous from the right and have limits from the left. We will show convergence to a pair of Gaussian processes with continuous sample paths on  $[a, b]$ . The expressions  $C([a, b] \times \{0, 1\})$  and  $(C[a, b])^2$  are both awkward, while  $C[a, b]^2$  denotes functions on a square region. Therefore, we use the symbol  $[a, b]_2$  to denote  $[a, b] \times \{0, 1\}$  and study random elements in  $C[a, b]_2$  and  $D[a, b]_2$ . Explicitly,  $C[a, b]_2$  is the collection of all pairs of real valued continuous functions on  $[a, b]$  while  $D[a, b]_2$  is the collection of all pairs of real valued càdlàg functions on  $[a, b]$ . We study the following processes in  $D[a, b]_2$ :

$$\begin{aligned} W_{m,r}(t) &\equiv \sqrt{m}(\hat{F}_{m,r}(t) - F_{m,r}(t)) \quad \text{and} \\ \hat{W}_{m,r}(t) &\equiv \sqrt{m}(\hat{F}_{m,r}(t) - F_r(t)). \end{aligned} \tag{8}$$

We are ultimately interested in a functional central limit theorem (FCLT) for the joint process  $(\hat{W}_{m,0}(\cdot), \hat{W}_{m,1}(\cdot))$ , so we must find the limiting joint covariance kernel of this pair of processes. To describe this limiting covariance kernel we introduce some convenient definitions and notation.

For convenience, throughout the text we will define  $\{\Gamma^{(m)}\}_{m=1}^\infty$  to be the sequence of correlation matrices corresponding to  $\{\Sigma^{(m)}\}_{m=1}^\infty$  and for each  $m, i$  define  $\tilde{\varepsilon}_{mi} \equiv \varepsilon_{mi}/\sqrt{1 - \|\mathbf{L}_{mi}\|_2^2}$ . Note that  $(\tilde{\varepsilon}_{m1}, \dots, \tilde{\varepsilon}_{mm}) \sim \mathcal{N}(0, \Gamma^{(m)})$  and that each  $\tilde{\varepsilon}_{mi}$  has unit variance. For any  $t, s \in [0, 1]$  and  $|\rho| \leq 1$ , define

$$\tilde{\rho}(t, s, \rho) \equiv \Pr\left(\tilde{\varepsilon}_1 \geq \bar{\Phi}^{-1}(t), \tilde{\varepsilon}_2 \geq \bar{\Phi}^{-1}(s) \mid \begin{pmatrix} \tilde{\varepsilon}_1 \\ \tilde{\varepsilon}_2 \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)\right) - ts. \quad (9)$$

Given a bivariate Gaussian with unit variance and correlation  $\rho$ , the above quantity is the covariance between the indicator that the first coordinate of this bivariate Gaussian exceeds its  $1 - t$  quantile and the indicator that the second coordinate of this bivariate Gaussian exceeds its  $1 - s$  quantile.

Now for any  $s, t \in (a, b)$  and  $r_0, r_1 \in \{0, 1\}$  and  $m \in \mathbb{N}_+$  define

$$c_m^{(r_0, r_1)}(t, s) \equiv \text{cov}(W_{m, r_0}(t), W_{m, r_1}(s)).$$

It is convenient to break up the expression of  $c_m^{(r_0, r_1)}(t, s)$  into two terms. Define

$$c_{m, \text{diag}}^{(r_0, r_1)}(t, s) \equiv \frac{1}{m} \sum_{i=1}^m \left( \pi_{r_0}^{(m)} \gamma_{mir_0}(t \wedge s) I\{r_0 = r_1\} - \pi_{r_0}^{(m)} \pi_{r_1}^{(m)} \gamma_{mir_0}(t) \gamma_{mir_1}(s) \right),$$

and define

$$c_{m, \text{cross}}^{(r_0, r_1)}(t, s) \equiv \frac{\pi_{r_0}^{(m)} \pi_{r_1}^{(m)}}{m} \sum_{i \neq j} \tilde{\rho}\left(\gamma_{mir_0}(t), \gamma_{mjr_1}(s), \Gamma_{ij}^{(m)}\right).$$

In the following proposition we show that  $c_m = c_{m, \text{diag}} + c_{m, \text{cross}}$ .

**Proposition 3.** *For any  $s, t \in [a, b]$  and  $r_0, r_1 \in \{0, 1\}$  and  $m \geq 2$*

$$c_m^{(r_0, r_1)}(t, s) = c_{m, \text{diag}}^{(r_0, r_1)}(t, s) + c_{m, \text{cross}}^{(r_0, r_1)}(t, s).$$

*Proof.* For  $i, j \in [m]$  define

$$\begin{aligned} C_{i, j, m}^{(r_0, r_1)}(t, s) &\equiv \text{cov}(H_{mir_0} I\{P_{mi} \leq t\}, H_{mjr_1} I\{P_{mj} \leq s\}) \\ &= \text{cov}(H_{mir_0} I\{\bar{\Phi}(\tilde{\varepsilon}_{mi}) \leq \gamma_{mir_0}(t)\}, H_{mjr_1} I\{\bar{\Phi}(\tilde{\varepsilon}_{mj}) \leq \gamma_{mjr_1}(s)\}). \end{aligned}$$

For  $i \neq j$ ,  $H_{mir_0}$ ,  $H_{mjr_1}$  and  $\tilde{\varepsilon}$  are all independent, so

$$\begin{aligned} C_{i, j, m}^{(r_0, r_1)}(t, s) &= \pi_{r_0}^{(m)} \pi_{r_1}^{(m)} \text{cov}(I\{\bar{\Phi}(\tilde{\varepsilon}_{mi}) \leq \gamma_{mir_0}(t)\}, I\{\bar{\Phi}(\tilde{\varepsilon}_{mj}) \leq \gamma_{mjr_1}(s)\}) \\ &= \pi_{r_0}^{(m)} \pi_{r_1}^{(m)} \tilde{\rho}(\gamma_{mir_0}(t), \gamma_{mjr_1}(s), \Gamma_{ij}^{(m)}). \end{aligned}$$

When  $i = j$ ,  $H_{mir_0} H_{mjr_1} = H_{mir_0} I\{r_0 = r_1\}$  and  $\tilde{\varepsilon}_{mi} = \tilde{\varepsilon}_{mj}$ , so that

$$C_{i, j, m}^{(r_0, r_1)}(t, s) = \pi_{r_0}^{(m)} I\{r_0 = r_1\} \gamma_{mir_0}(t \wedge s) - \pi_{r_0}^{(m)} \pi_{r_1}^{(m)} \gamma_{mir_0}(t) \gamma_{mir_1}(s).$$

Since the above expressions hold for any  $i, j \in [m]$ ,

$$\begin{aligned}
c_m^{(r_0, r_1)}(t, s) &= \text{cov}(W_{m, r_0}(t), W_{m, r_1}(s)) \\
&= m \text{cov}(\hat{F}_{m, r_0}(t), \hat{F}_{m, r_1}(s)) \\
&= \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^m \text{cov}(H_{mir_0} I\{P_{mi} \leq t\}, H_{mjr_1} I\{P_{mj} \leq s\}) \\
&= \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^m C_{i, j, m}^{(r_0, r_1)}(t, s) \\
&= \frac{1}{m} \sum_{i=1}^m C_{i, i, m}^{(r_0, r_1)}(t, s) + \frac{1}{m} \sum_{i \neq j} C_{i, j, m}^{(r_0, r_1)}(t, s) \\
&= c_{m, \text{diag}}^{(r_0, r_1)}(t, s) + c_{m, \text{cross}}^{(r_0, r_1)}(t, s). \quad \square
\end{aligned}$$

Now define

$$c^{(r_0, r_1)}(t, s) \equiv \lim_{m \rightarrow \infty} c_m^{(r_0, r_1)}(t, s). \quad (10)$$

By the simplified formula for  $c_m^{(r_0, r_1)}(t, s)$ , the above limit exists by Condition 4 if we further impose Condition 6 below.

**Condition 6.** For any  $(s, t) \in [a, b]$  and  $r_0, r_1 \in \{0, 1\}$ ,

$$\begin{aligned}
\lim_{m \rightarrow \infty} \frac{\pi_{r_0}^{(m)} \pi_{r_1}^{(m)}}{m} \sum_{i=1}^m \gamma_{mir_0}(t) \gamma_{mir_1}(s) \quad \text{and} \\
\lim_{m \rightarrow \infty} \frac{\pi_{r_0}^{(m)} \pi_{r_1}^{(m)}}{m} \sum_{i \neq j} \tilde{\rho}(\gamma_{mir_0}(t), \gamma_{mjr_1}(s), \Gamma_{ij}^{(m)})
\end{aligned}$$

both exist.

It is easy to see that the functions  $c(\cdot, \cdot)$  defined above gives a joint covariance kernel that is symmetric and positive semidefinite because it is the limit of symmetric and positive semidefinite joint covariance kernels  $c_m$ .

### 2.3.6. Regularity conditions on $F_0, F_1, F_{m,0}$ and $F_{m,1}$

Before introducing the main theorems, we introduce another two conditions that will be used in their proof.

**Condition 7.** Both  $F_0$  and  $F_1$  are differentiable at  $\tau_*$ .

The final condition is needed to derive a  $(\hat{W}_{m,0}(\cdot), \hat{W}_{m,1}(\cdot))$  FCLT from a  $(W_{m,0}(\cdot), W_{m,1}(\cdot))$  FCLT. We would like to hold the subdistribution functions  $F_{m,0}$  and  $F_{m,1}$  constant as  $m$  changes but this is not possible outside of trivial cases as they have step sizes  $1/m$ . Instead we assume that they approach limits  $F_0$  and  $F_1$  at a fast rate.

**Condition 8.** For  $r \in \{0, 1\}$ ,  $\lim_{m \rightarrow \infty} \sup_{t \in [a, b]} |\sqrt{m}(F_{m,r}(t) - F_r(t))| = 0$ .

**Notes about Condition 8:** While the distance between  $F_{m,r}$  and  $F_r$  could potentially be as small as  $O(1/m)$  we do not need that level of precision. We only need the distance to be small compared to random sampling of  $m$  IID random variables. In addition, a version of Theorem 1 below will still hold if we slightly loosen Condition 8 to say that there exist continuous functions  $z_0 : [a, b] \rightarrow \mathbb{R}$  and  $z_1 : [a, b] \rightarrow \mathbb{R}$  such that for both  $r \in \{0, 1\}$ ,

$$\lim_{m \rightarrow \infty} \sup_{t \in [a, b]} |\sqrt{m}(F_{m,r}(t) - F_r(t)) - z_r(t)| = 0.$$

However, if we use this slightly looser condition, the resulting theorem statement will be messier.

### 3. Statement of the Theorems

**Theorem 1.** *For the model of Section 2.1, suppose that conditionally on a specific value of the latent factor  $\mathbf{W} = \mathbf{w} \in \mathbb{R}^k$  that Conditions 1–8 all hold. Then*

$$\sqrt{m} \left( \text{FDP}_m - \frac{qF_0(\tau_*)}{\tau_*} \mid \mathbf{W} = \mathbf{w} \right) \xrightarrow{d} \mathcal{N}(0, \sigma_L^2) \quad (11)$$

as  $m \rightarrow \infty$  where

$$\sigma_L^2 \equiv \frac{q^2}{\tau_*^2} \left( (1 + \alpha)^2 c^{(0,0)}(\tau_*, \tau_*) + \alpha^2 c^{(1,1)}(\tau_*, \tau_*) + 2\alpha(1 + \alpha) c^{(1,0)}(\tau_*, \tau_*) \right) \quad (12)$$

for the function  $F_0$  given at (5), the asymptotic ECDF  $G$  given at (6), the Simes point  $\tau_*$  from (7), the covariances  $c^{(r_0, r_1)}(\cdot, \cdot)$  given by (10) and

$$\alpha \equiv \frac{F_0'(\tau_*) - F_0(\tau_*)/\tau_*}{1/q - G'(\tau_*)}. \quad (13)$$

*Proof.* See the appendix for a proof of this theorem.  $\square$

The proof is quite long but to summarize we first derive an FCLT for the joint process  $(\hat{F}_{m,0}, \hat{F}_{m,1})$  by proving finite dimensional distribution convergence using a result from Peligrad (1996) and then extend to an FCLT by using a result from Andrews and Pollard (1994). We then define  $\Psi^{(\text{FDP})} : D[a, b]_2 \rightarrow \mathbb{R}$  to be a particular function satisfying  $\text{FDP}_m = \Psi^{(\text{FDP})}(\hat{F}_{m,0}, \hat{F}_{m,1})$  with probability converging to 1 as  $m \rightarrow \infty$ . Then we argue that  $\Psi^{(\text{FDP})}$  is Hadamard differentiable at  $(F_0, F_1)$  tangentially to  $C[a, b]_2$  and compute the Hadamard derivative by mimicking the approach in Neuvial (2008). To complete the proof of the CLT given in (11), we tie these results together with the functional delta method in Chapter 20.2 of van der Vaart (1998). Using the same proof technique we obtain the following conditional CLT for the ratio  $V_m/m$ .

**Theorem 2.** *Under the conditions of Theorem 1,*

$$\sqrt{m} \left( \frac{V_m}{m} - F_0(\tau_*) \mid \mathbf{W} = \mathbf{w} \right) \xrightarrow{d} \mathcal{N}(0, \sigma_R^2) \quad \text{as } m \rightarrow \infty \quad (14)$$

where

$$\sigma_R^2 \equiv (1 + \beta)^2 c^{(0,0)}(\tau_*, \tau_*) + \beta^2 c^{(1,1)}(\tau_*, \tau_*) + 2\beta(1 + \beta)c^{(1,0)}(\tau_*, \tau_*) \quad (15)$$

for

$$\beta \equiv \frac{F'_0(\tau_*)}{1/q - G'(\tau_*)}. \quad (16)$$

*Proof.* See the appendix for a proof of this theorem.  $\square$

It is often the case that for a particular factor and noise model, Conditions 1–8 will not hold for all values of the latent factor  $\mathbf{W} = \mathbf{w} \in \mathbb{R}^k$ . Condition 5 will often be violated when drawing  $\mathbf{W} \sim \mathcal{N}(0, I_k)$ . We do not expect a positive probability that  $\tau_*$  will be an accumulation point of  $\{t : G(t) = t/q\}$ , nor do we expect a positive probability that  $\sup\{t \in (0, 1) : G(t) \geq t/q\} \neq \sup\{t \in (0, 1) : G(t) > t/q\}$ , but we do expect a positive probability that  $\tau_* = 0$ . Therefore, in the next theorem we describe the asymptotic behavior of the BH procedure, conditional on  $\mathbf{W} = \mathbf{w} \in \mathbb{R}^k$  when  $\tau_* = 0$ .

**Theorem 3.** *For the model of Section 2.1, when conditioning on the latent factor  $\mathbf{W} = \mathbf{w} \in \mathbb{R}^k$ , suppose that Conditions 1–4 hold, that  $\tau_* = 0$ , and that Conditions 6, 8 hold when setting  $[a, b] = [0, 1]$ . Then*

$$\tau_{\text{BH},m} | \mathbf{W} = \mathbf{w} \xrightarrow{p} 0 \quad \text{and} \quad \frac{V_m}{m} | \mathbf{W} = \mathbf{w} \xrightarrow{p} 0. \quad (17)$$

*Proof.* See the appendix material for a proof of this theorem.  $\square$

**Remark 1.** Theorem 3 will still hold if we loosen the mixing Condition 1 and simply require that the error array  $\{\varepsilon_{mi} : 1 \leq i \leq m < \infty\}$  is strongly mixing. This is because the proof of Proposition B.1 does not require Condition 1 to hold and merely requires that the error array is strongly mixing.

**Remark 2.** When  $\tau_* = 0$  and  $\tau_{\text{BH},m} \xrightarrow{p} 0$ , it is not guaranteed that  $\text{FDP}_m \xrightarrow{p} 0$ . For example, consider the scenario where  $\mathbf{L}_{mi} = 0$  for all  $m, i$ , the errors  $(\varepsilon_{m1}, \dots, \varepsilon_{mm})$  are independent and  $\pi_1^{(m)} = 1/m$ . Then  $G(t) = F_0(t) = t$ ,  $\tau_* = 0$  and by Theorem 3,  $\tau_{\text{BH},m} \xrightarrow{p} 0$ . Meanwhile, since all the null test statistics are independent, when we condition on  $\sum_{i=1}^m H_{mi1} = 1$ , then  $\text{FDP}_m \geq 0.5$  any time there is at least one false discovery, which is guaranteed to happen when at least one of the  $m - 1$  IID  $\text{Unif}(0, 1)$  null  $p$ -values is less than  $q/m$ . Therefore, letting  $U_1, \dots, U_{m-1} \stackrel{\text{iid}}{\sim} \text{Unif}(0, 1)$  be generic uniform random variables,

$$\begin{aligned} \Pr(\text{FDP}_m \geq 0.5) &\geq \Pr\left(\text{FDP}_m \geq 0.5 \mid \sum_{i=1}^m H_{mi1} = 1\right) \Pr\left(\sum_{i=1}^m H_{mi1} = 1\right) \\ &\geq \Pr\left(\cup_{i=1}^{m-1} \{U_i < q/m\}\right) (1 - 1/m)^{m-1} \\ &= (1 - (1 - q/m)^{m-1}) (1 - 1/m)^{m-1}. \end{aligned}$$

Taking the  $\liminf$  as  $m \rightarrow \infty$  of each side,  $\liminf_{m \rightarrow \infty} \Pr(\text{FDP}_m \geq 0.5) > 0$ . Hence, in this scenario,  $\tau_* = 0$  and  $\tau_{\text{BH},m} \xrightarrow{p} 0$ , but  $\text{FDP}_m \xrightarrow{p} 0$  does not hold.

#### 4. Corollaries when there is no factor model component

The simplest applications of Theorems 1 and 2 are to settings where there is no factor model component. That is  $k = 0$ , or equivalently  $\mathbf{L}_{mi} = 0$  for  $1 \leq i \leq m$ . Then the test statistics are  $X_{mi} = \mu_A H_{mi} + \varepsilon_{mi}$  where  $(\varepsilon_{m1}, \dots, \varepsilon_{mm}) \sim \mathcal{N}(0, \Gamma^{(m)})$  for a correlation matrix  $\Gamma^{(m)} \in \mathbb{R}^{m \times m}$ . Next suppose, as is usual in the two-group mixture model that  $\pi_1^{(m)} = \pi_1 \in (0, 1)$  for all  $m \geq 1$ . Finally, we will assume that the errors  $(\varepsilon_{m1}, \dots, \varepsilon_{mm})$  are  $M$ -dependent.

Many of the 8 conditions in our theorems hold trivially in this setting. Most trivially,  $S_L = 0$  making Condition 3 hold. The mixing Conditions 1 and 2 will hold by the assumed  $M$ -dependence. Also in this setting, because for each  $m$ ,  $F_0(t) = F_{m,0}(t) = (1 - \pi_1)t$  and  $F_1(t) = F_{m,1}(t) = \pi_1 \bar{\Phi}(\bar{\Phi}^{-1}(t) - \mu_A)$ , Conditions 4 and 8 on the subdistributions can be seen to hold.

To check that Condition 5 ruling out pathologies about  $\tau_*$  holds note that  $G(t) = (1 - \pi_1)t + \pi_1 \bar{\Phi}(\bar{\Phi}^{-1}(t) - \mu_A)$ . A simple calculation shows that  $G'(t) = (1 - \pi_1) + \exp(\mu_A \bar{\Phi}^{-1}(t) - \mu_A^2/2)$  and

$$G''(t) = -\pi_1 \exp(\mu_A \bar{\Phi}^{-1}(t) - \mu_A^2/2) / \varphi(\bar{\Phi}^{-1}(t)),$$

implying that  $G$  is concave on  $(0, 1)$  and that  $G'(t) \rightarrow \infty$  as  $t \downarrow 0$ . By concavity of  $G$ , and since both  $G$  and the Simes line intersect the origin,  $\{t > 0 : G(t) = t/q\}$  contains at most one point. It remains to show existence of a point  $t > 0$  with  $G(t) = t/q$ . Since  $G'(t) \rightarrow \infty$  as  $t \downarrow 0$  and  $G(0) = 0$ , there must be an  $\epsilon > 0$  such that  $G(\epsilon) > \epsilon/q$ . Also  $G(1) = 1 < 1/q$ , so the continuous function  $t \mapsto G(t) - t/q$  must cross 0 at some unique  $t_* \in (0, 1)$ . By uniqueness and continuity of  $G$  this unique  $t_*$  is the Simes point  $\tau_*$  defined in (7) and further Condition 5 will be satisfied. Because  $\tau_* \in (0, 1)$  and because  $F_0$  and  $F_1$  are differentiable on  $(0, 1)$ , Condition 7 also holds.

The only remaining Condition to check is Condition 6 on convergence of the covariance kernels. Let  $(a, b) \subset (0, 1)$  be any small open interval containing  $\tau_*$ . Since for  $r \in \{0, 1\}$ ,  $\gamma_{mir}$  does not vary with  $m$ , the first limit in Condition 6 always holds. Therefore, Condition 6 holds whenever

$$\varrho_{r_0 r_1}(t, s) \equiv \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i \neq j} \tilde{\rho}(t, s, \Gamma_{ij}^{(m)}) \quad (18)$$

exists for all  $s, t \in [a, b]$  and  $r_0, r_1 \in \{0, 1\}$  with  $\tilde{\rho}$  defined at (9). We summarize this along with an application of Theorem 1 in Corollary 1.

**Corollary 1.** *Let the testing problem satisfy the conditions of Theorem 1 with the following modifications:*

- 1) the factor loadings  $\mathbf{L}_{mi}$  are all zero,
- 2) the errors  $\varepsilon_{mi}$  are  $M$ -dependent for all  $m$  and some  $M < \infty$ , and
- 3) the probability  $\pi_1 \in (0, 1)$  of nonnull hypotheses does not depend on  $m$ .

Let  $\tau_*$  be the unique  $t \in (0, 1)$  satisfying  $t/q = \pi_0 t + \pi_1 \bar{\Phi}(\bar{\Phi}^{-1}(t) - \mu_A)$  and let  $(a, b) \subset (0, 1)$  be a small open interval containing  $\tau_*$ . Let the correlations

among  $\varepsilon_{mi}$  be such that  $\varrho_{r_0 r_1}(t, s)$  defined at (18) exists for all  $t, s \in [a, b]$  and  $r_0, r_1 \in \{0, 1\}$ . Then

$$\sqrt{m}(\text{FDP}_m - \pi_0 q) \xrightarrow{d} N\left(0, \frac{\pi_0 q^2}{\tau_*^2} \left(\tau_* - \pi_0 \tau_*^2 + \pi_0 \varrho_{00}(\tau_*, \tau_*)\right)\right) \quad (19)$$

where  $\pi_0 \equiv 1 - \pi_1$ .

*Proof.* As discussed before the statement of the corollary, Conditions 1–8 hold in this setting. Noting that in this setting  $\alpha = 0$  and there is no dependency on the latent factor  $\mathbf{W}$ , the result holds by Theorem 1.  $\square$

We can specialize Corollary 1 to settings with block diagonal correlations and with Toeplitz correlations. The conclusion simplifies for these cases.

**Corollary 2** (Block diagonal correlations). *Under the conditions of Corollary 1 suppose that  $(\varepsilon_{m1}, \dots, \varepsilon_{mm})$  has a block diagonal correlation matrix with blocks of fixed size  $s_B$  in which the off diagonal correlations are to be  $\rho_B$ . Let  $\tau_*$  be the unique  $t \in (0, 1)$  satisfying  $t/q = \pi_0 t + \pi_1 \bar{\Phi}(\bar{\Phi}^{-1}(t) - \mu_A)$ . Then*

$$\sqrt{m}(\text{FDP}_m - \pi_0 q) \xrightarrow{d} \mathcal{N}\left(0, \frac{\pi_0^2 q^2}{\tau_*^2} \left(\frac{\tau_*}{\pi_0} - \tau_*^2 + (s_B - 1)\tilde{\rho}(\tau_*, \tau_*, \rho_B)\right)\right) \quad (20)$$

where  $\pi_0 \equiv 1 - \pi_1$  and  $\tilde{\rho}$  is defined at (9).

*Proof.* This follows from a direct application of Corollary 1.  $\square$

The corollary as written requires  $m$  to be a multiple of  $s_B$  but it extends easily to  $m \rightarrow \infty$  through an arbitrary sequence of  $m$ . One can let the “last” block be smaller than the others if necessary.

Another simple correlation structure we can consider has banded Toeplitz correlation matrices for  $\varepsilon_{m1}, \dots, \varepsilon_{mm}$ .

**Corollary 3** (Toeplitz correlation). *Under the conditions of Corollary 1 suppose that  $(\varepsilon_{m1}, \dots, \varepsilon_{mm})$  has a Toeplitz correlation matrix*

$$\Sigma_{ij}^{(m)} = I\{i = j\} + \sum_{\ell=1}^M \rho_\ell I\{|i - j| = \ell\}$$

where  $\rho_1, \dots, \rho_M \in (-1, 1)$  are such that  $\Sigma^{(m)}$  is positive semi-definite for all  $m > M$ . Let  $\tau_*$  be the unique  $t \in (0, 1)$  satisfying  $t/q = \pi_0 t + \pi_1 \bar{\Phi}(\bar{\Phi}^{-1}(t) - \mu_A)$ . Then,

$$\sqrt{m}(\text{FDP}_m - \pi_0 q) \xrightarrow{d} \mathcal{N}\left(0, \frac{\pi_0^2 q^2}{\tau_*^2} \left(\frac{\tau_*}{\pi_0} - \tau_*^2 + 2 \sum_{\ell=1}^M \tilde{\rho}(\tau_*, \tau_*, \rho_\ell)\right)\right) \quad (21)$$

where  $\pi_0 \equiv 1 - \pi_1$  and  $\tilde{\rho}$  is defined at (9).

*Proof.* This follows from a direct application of Corollary 1.  $\square$

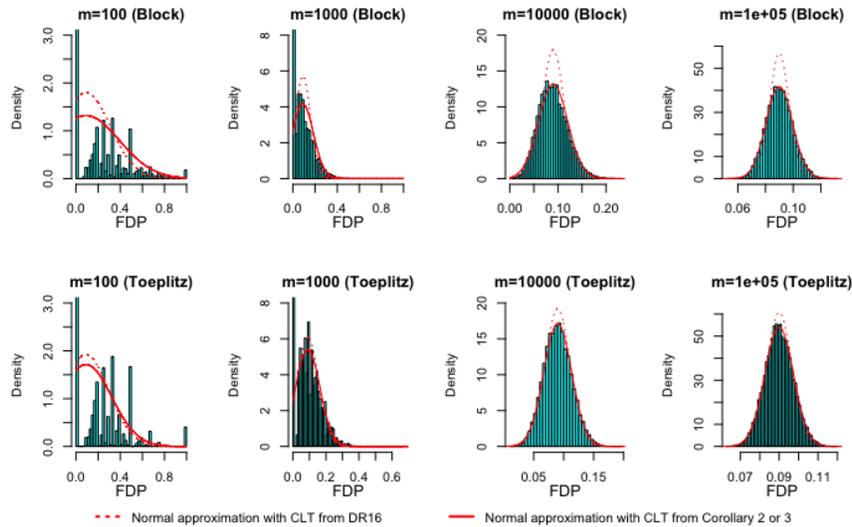


FIG 3. This figure compares the histograms of FDP to our normal approximation and an earlier one by Delattre and Roquain (2016) that does not necessarily cover these cases. Each panel is based on 25,000 Monte Carlo simulations as described in the text.

We check the CLTs provided by Corollaries 2 and 3 via simulation in Figure 3. We compare the normal approximation given by these CLTs to the normal approximation given by Corollary 4.2 in Delattre and Roquain (2016). We simulate block and banded correlation structures that do not satisfy the sufficient conditions in their Corollary 4.2.

For each correlation structure we ran 25,000 Monte Carlo simulations with  $\mu_A = 2$ ,  $\pi_0 = 0.9$ ,  $q = 0.1$  and  $m = 10,000$ . The block correlation matrices we considered had block size  $s_B = 20$ , and within block correlations  $\rho_B = 0.5$ . The banded Toeplitz correlation matrix had  $M = 2$  with  $\rho = \rho_1 = 0.65$  above and below the diagonal and  $\rho = \rho_2 = 0.3$  two rows above and below the diagonal. Our normal approximation fits very well for the larger values of  $m$ . For large  $m$  the normal approximation from Delattre and Roquain (2016) appears to accurately estimate the mean but not the variance of the FDP. From this we believe that something in their sufficient conditions must also have been necessary.

## 5. Burstiness in a factor model

The results in the previous section are CLTs that do not require conditioning on the latent factor as they assumed no long-range correlations modeled via a factor model. The CLTs are messier when factor model components are introduced, so we present two examples for factor model settings where the formulas for the asymptotic distribution of the FDP have some simplifications.

### 5.1. 1-factor model for long-range equicorrelated Gaussian noise

Suppose that for each  $m$ ,  $H_{m1}, \dots, H_{mm} \stackrel{\text{iid}}{\sim} \text{Bern}(\pi_1)$  for a fixed  $\pi_1 \in (0, 1)$ , but we now have a one dimensional latent factor; that is  $\mathbf{W} \sim \mathcal{N}(0, 1)$ . For simplicity, we consider the simplest factor model structure: an equicorrelated Gaussian model. In particular, we let  $\mathbf{L}_{mi} = \sqrt{\rho_1}$  where  $\rho_1 \in [0, 1)$  for all  $m, i$ . We will also allow for errors with shorter range correlations to be added to the model by supposing that  $(\tilde{\varepsilon}_{m1}, \dots, \tilde{\varepsilon}_{mm}) \sim \mathcal{N}(0, \Gamma^{(m)})$  where  $\Gamma^{(m)}$  is a correlation matrix with blocks of size  $s_B$  and off diagonal within-block correlations of  $\rho_2$ . We assume that the blocks are of equal size, except the last one if  $m$  does not divide  $s_B$ . In this model the test statistics are  $X_{mi} = \mu_A H_{mi} + \sqrt{\rho_1} \mathbf{W} + \sqrt{1 - \rho_1} \tilde{\varepsilon}_{mi}$  and the correlation structure of the errors (not related to the indicators  $H_{mi}$  of whether the hypotheses are true) follows a matrix  $\Sigma_{B_2}$  where

$$(\Sigma_{B_2})_{ij} = \begin{cases} 1, & \text{if } i = j, \\ \rho_1, & \text{if } i \text{ and } j \text{ are in different blocks,} \\ \rho_1 + (1 - \rho_1)\rho_2, & \text{if } i \neq j \text{ and if } i \text{ and } j \text{ are in the same block.} \end{cases}$$

In such a setting it is easy to show that all of our conditions, except possibly Condition 5 will hold. Condition 5 ruling out pathologies involving  $\tau_*$  will hold depending on the value of  $\mathbf{W}$  drawn from  $\mathcal{N}(0, 1)$ . Some  $\mathbf{w}$  may give  $\tau_* = 0$  though we do not expect a positive probability that  $\tau_*$  will be an accumulation point of  $\{t : G(t) = t/q\}$ .

**Corollary 4.** *In the multiple hypothesis testing setting of Section 5.1, condition on  $\mathbf{W} = \mathbf{w} \in \mathbb{R}$  and for  $t \in (0, 1)$  let  $G(t) = (1 - \pi_1)\gamma_0(t) + \pi_1\gamma_1(t)$  with  $\gamma_r(t) = \bar{\Phi}((\bar{\Phi}^{-1}(t) - \mu_A r - \sqrt{\rho_1}\mathbf{w})/\sqrt{1 - \rho_1})$  for  $r \in \{0, 1\}$ . If  $\mathbf{w}$  is such that  $G$  satisfies Condition 5 that the Simes point is positive with no pathologies, then*

$$\sqrt{m} \left( \text{FDP}_m - \frac{q\pi_0\gamma_0(\tau_*)}{\tau_*} \mid \mathbf{W} = \mathbf{w} \right) \xrightarrow{d} \mathcal{N}(0, \sigma_{L,2}^2) \quad \text{as } m \rightarrow \infty$$

where  $\pi_0 \equiv 1 - \pi_1$  and

$$\sigma_{L,2}^2 = \frac{q^2}{\tau_*^2} \left( (1 + \alpha)^2 c^{(0,0)}(\tau_*, \tau_*) + \alpha^2 c^{(1,1)}(\tau_*, \tau_*) + 2\alpha(1 + \alpha) c^{(1,0)}(\tau_*, \tau_*) \right)$$

where

$$\alpha = \frac{\gamma'_0(\tau_*) - \gamma_0(\tau_*)/\tau_*}{1/q - G'(\tau_*)}$$

and for  $r \in \{0, 1\}$ ,

$$c^{(r,r)}(\tau_*, \tau_*) = \pi_r \gamma_r(\tau_*) - \pi_r^2 \gamma_r(\tau_*) \gamma_r(\tau_*) + \pi_r^2 (s_B - 1) \tilde{\rho}(\gamma_r(\tau_*), \gamma_r(\tau_*), \rho_2)$$

and

$$c^{(1,0)}(\tau_*, \tau_*) = -\pi_0 \pi_1 \gamma_0(\tau_*) \gamma_1(\tau_*) + \pi_0 \pi_1 (s_B - 1) \tilde{\rho}(\gamma_0(\tau_*), \gamma_1(\tau_*), \rho_2)$$

where  $\tilde{\rho}$  is defined in Equation (9).

*Proof.* As mentioned earlier in this section Conditions 1, 2, 3, 4, 6, 7, and 8 will hold for any value of  $\mathbf{W}$  drawn. Therefore the above result holds from applying Theorem 1 in the setting where Condition 5 also holds.  $\square$

**Remark 3.** In the setting of Corollary 4, if  $m/s_B$  is large, then perhaps the test statistic correlations can be modeled with a factor model with a bit more than  $m/s_B$  factors but asymptotically such an approach would require adding infinitely many factors in the model. In our setup, the equicorrelations  $\rho_1$  are long-range and persist as  $m \rightarrow \infty$  and hence they are modeled with a factor model whereas the additional noise with correlation blocks of size  $s_B$  involves short-range correlations and are therefore not modeled as a factor.

### 5.2. Setting where number of nonnulls is $o_p(\sqrt{m})$

Here we consider sparse nonnulls with  $\pi_1^{(m)} = o(1/\sqrt{m})$ . It can be shown with a Chernoff bound for the binomial that in this case the number of nonnulls is  $o_p(\sqrt{m})$ . Also suppose that under the two group mixture model for test statistics of Section 2.1, Conditions 1- 8 hold. Then it will follow that  $F_1 = 0$  and moreover  $W_{m,r} = \sqrt{m}(\hat{F}_{m,1} - F_{m,1}) \xrightarrow{P} 0$ . If this is the case, then we will have  $c^{(1,1)} = 0$ ,  $c^{(1,0)} = 0$ ,  $F_0(\tau_*) = \tau_*/q$  and  $\alpha = -1$ .

**Corollary 5.** *Suppose that we are in the two group mixture model for test statistics of Section 2.1 and that  $\pi_1^{(m)} = o(1/\sqrt{m})$ . If, conditionally on a specific value of the latent factor  $\mathbf{W} = \mathbf{w} \in \mathbb{R}^k$ , Conditions 1-8 hold, then*

$$\sqrt{m}(\text{FDP}_m - 1 | \mathbf{W} = \mathbf{w}) \xrightarrow{P} 0$$

and

$$\sqrt{m}\left(\frac{V_m}{m} - \frac{\tau_*}{q} | \mathbf{W} = \mathbf{w}\right) \xrightarrow{d} \mathcal{N}(0, \sigma_{R,3}^2) \quad \text{as } m \rightarrow \infty$$

where

$$\sigma_{R,3}^2 \equiv (1 - qG'(\tau_*))^{-2} c^{(0,0)}(\tau_*, \tau_*).$$

*Proof.* Apply Theorems 1 and 2 noting that  $F_0(\tau_*) = \tau_*/q$ ,  $\alpha = -1$ ,  $c^{(1,1)} = 0$ ,  $c^{(1,0)} = 0$ .  $\square$

The corollary indicates that severe bursts can occur;  $V_m/m$  can converge to a positive number even while the proportion of hypotheses that are nonnull converges to 0.

We check the result of Corollary 5 via simulation in Figure 4. We simulate from the 1-factor model described in Section 5.1, except now the proportion of nonnulls  $\pi_1^{(m)}$  is not fixed in  $m$ . Instead we set  $\pi_1^{(m)} = 5m^{-2/3}$ . For each  $m \in \{10^2, 10^3, 10^4, 10^5, 10^6\}$ , we conditioned on  $\mathbf{w} = 2.5$  and ran 25,000 Monte Carlo simulations with  $\mu_A = 2$ ,  $\pi_0^{(m)} = 1 - 5m^{-2/3}$ ,  $q = 0.1$ ,  $\rho_1 = 0.3$ , and  $\rho_2 = 0.6$ . For these choices of parameters, the conditions of Corollary 5 are met.

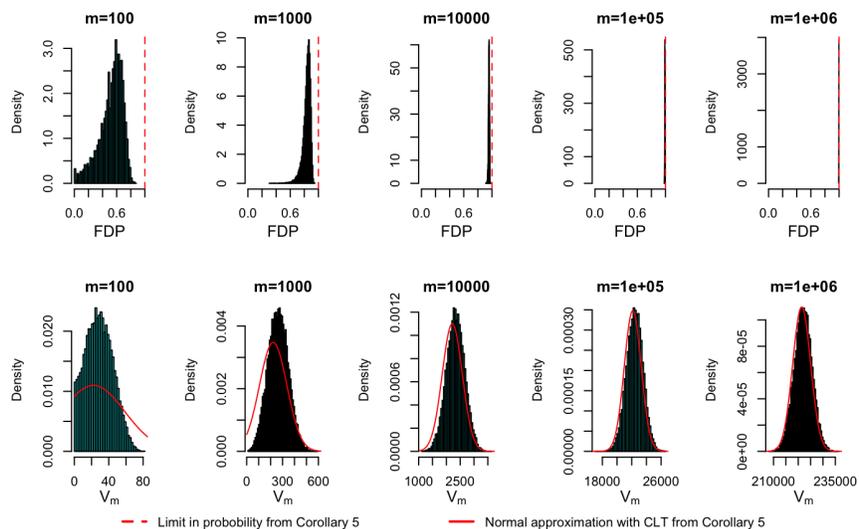


FIG 4. This figure compares the histograms of the FDP and the number of discoveries  $V_m$  to the asymptotic estimates of their distributions given by Corollary 5. Each panel is based on 25,000 Monte Carlo simulations as described in the text.

## 6. Data Driven Factor Model Example

We fit a 3-factor model to the GDS 3027 Duchenne Muscular Dystrophy (DMD) data, which can be found on the Gene Expression Omnibus. This data set was analyzed in Kotelnikova et al. (2012) and Wang et al. (2020) and had  $m = 22,283$  genes and  $n = 37$  subjects. Of these subjects 23 had DMD and 14 did not. We centered the data for each gene and stored it in a matrix  $Y \in \mathbb{R}^{37 \times 22,283}$ . To fit a homoskedastic factor model, we looked at the plot of the singular values of  $Y$  and chose to work with the largest three of them for illustrative purposes. We then computed  $Y_3$ , the singular value decomposition-based rank 3 approximation to  $Y$ , and estimated the homoskedastic noise,  $\sigma_E$  as the standard deviation of the entries in  $Y - Y_3$ . We let  $\tilde{L} \in \mathbb{R}^{m \times 3}$  be the matrix whose columns consist of the first 3 right singular vectors of  $Y$  scaled by their corresponding singular values. We subsequently treat  $\tilde{L}$  and  $\sigma_E$  as fixed quantities and then assume the following factor model under the global null:  $Y^\top = \tilde{L}F + \sigma_E E$ , where the entries of  $F \in \mathbb{R}^{3 \times 37}$  and  $E \in \mathbb{R}^{m \times 3}$  are IID standard Gaussians. Under the alternative we suppose that for each nonnull gene, the values of  $Y$  for that gene are shifted by a fixed constant for DMD subjects and a different fixed constant, maintaining centering of the columns of  $Y$ , for the control subjects.

Since the dataset is from a case-control study, to compute the test statistics we condition on DMD status and assume that the stochasticity in our observations  $Y$  comes from the random matrices  $E$  and  $F$ . The unstandardized test statistics  $X_{\text{ustd}} \in \mathbb{R}^m$  are simply the difference-in-means between the DMD

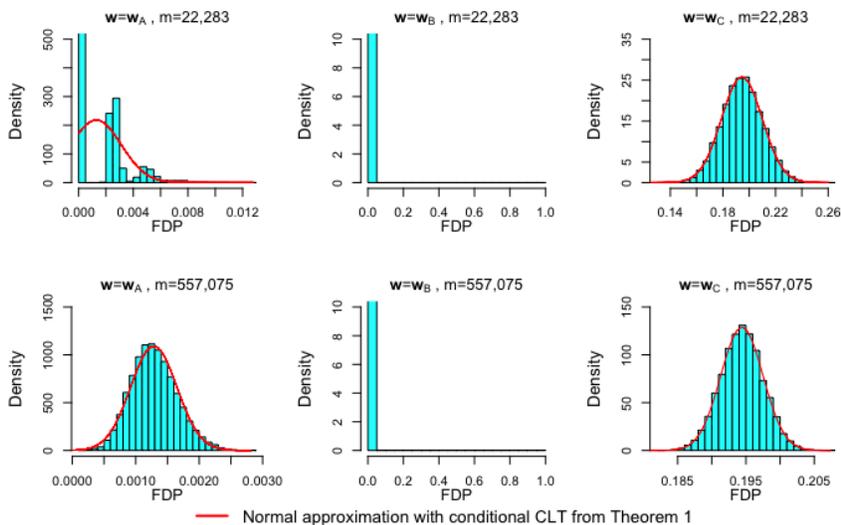


FIG 5. The histograms show 25,000 samples of the number of false discoveries in the two group model with a data-driven three factor model for dependence. The 3-factor model has draws  $A$ ,  $B$  and  $C$  as described in the text. The top row has  $m = 22,283$  hypotheses and the bottom has 25 times as many hypotheses. The nominal FDR control threshold is  $q = 0.1$ . Draw  $C$  yields a higher FDP while draw  $A$  yields a lower one. Draw  $B$  has a heavy tail, corresponding to one or two simulations where the  $FDP = 1$  due to a single false discovery. The red curves for  $A$  and  $C$  are asymptotic Gaussians from Theorem 1. Case  $B$  does not satisfy the sufficient conditions for the conditional CLT, but satisfies the conditions for Theorem 3.

group and the control group for each gene and this unstandardized test statistics vector has covariance matrix proportional to  $\sigma_B^2 I_m + \tilde{L}\tilde{L}^\top$ . The standardized test statistics  $X \in \mathbb{R}^m$  are given by dividing each entry of  $X_{\text{ustd}}$  by the square-root of the corresponding diagonal entry of  $\sigma_B^2 I_m + \tilde{L}\tilde{L}^\top$ . The vector of test statistics then satisfy  $X = \tilde{\mu} + \mathbf{L}\mathbf{W} + \varepsilon$  where  $\tilde{\mu}$  is a vector of constant means (which are zero for the null genes),  $\mathbf{L}$  is a matrix of factor loadings similar to  $\tilde{L}$  but with appropriately rescaled rows,  $\mathbf{W} \sim \mathcal{N}(0, I_3)$  and  $\varepsilon$  is heteroskedastic, independent, and centered Gaussian noise. Assuming that each standardized nonnull test statistic has the same mean  $\mu_A > 0$ , that we conduct one-sided testing, and that the nonnulls are determined by IID  $\text{Bern}(\pi_1)$  draws, using the test statistics  $X$  we are in the multiple hypothesis testing setting of Section 2.1.

Figure 2 in Section 2 shows the asymptotic ECDF of the  $p$ -values for three specific realizations of the latent factor  $\mathbf{w}$  in the data-driven 3-factor model and multiple testing setting described above, with  $\mu_A = 2$ ,  $\pi_1 = 0.1$ , and  $q = 0.1$ . Figure 5 shows histograms of the FDP based on 25,000 Monte Carlo simulations for the same data-driven 3-factor model, multiple testing setup, factor outcomes, and parameters as Figure 2.

In the top panel of Figure 5,  $m = 22,283$  as is the case in the original 3-factor model fit to the GDS 3027 dataset. In the bottom panel, to increase the number

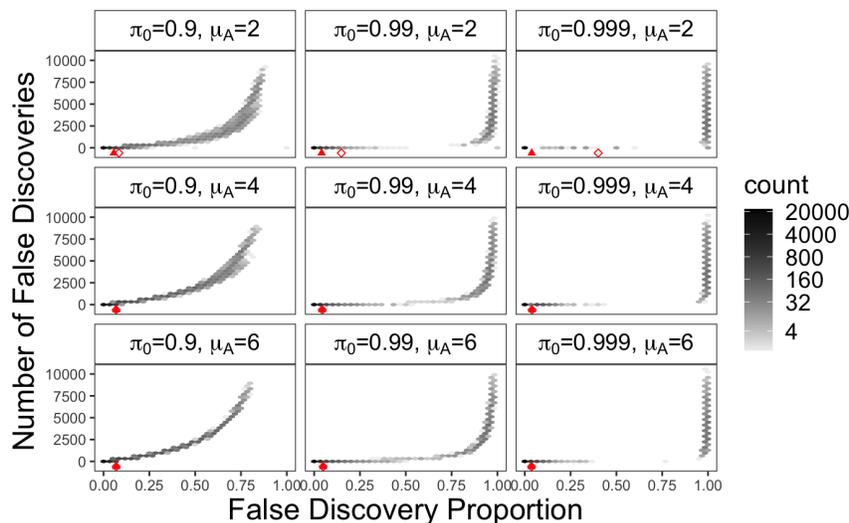


FIG 6. The hexagonal-binned heatmaps each show the results of 25,000 Monte Carlo simulations for the data-driven 3-factor model described in the text, using various nonnull effect size  $\mu_A$  and proportions of nulls  $\pi_0$ . In each simulation, the BH method is applied at level  $q = 0.1$  to  $m = 22,283$  test statistics. For each plot, the mean of the FDP is marked with a solid triangle and the mean amongst nonzero FDP values is marked with a hollow diamond, which estimate the FDR and pFDR respectively.

of tests and check asymptotic behavior, we copy each row of factor loadings in the original factor model 25 times to get a distribution of the FDP when  $m = 22,283 \times 25$  tests are conducted. That is much larger than we would need for gene expression and approaches the range we would encounter for SNPs. In case A, the CLT is reasonable for the larger but not the smaller sample size. The CLT fits well for both sample sizes for case C. In Case B, the sufficient conditions for the conditional CLT do not hold and Theorem 3 holds instead.

This simulation shows some bursty behavior for BH as follows. Cases A and C are both covered by the conditional CLT and there we see that even in cases covered by the conditional CLT, the FDP can vary greatly, being nearly Gaussian with means varying by nearly 100-fold. When cases like case B arise there is no conditional CLT and by Theorem 3, the BH rejection threshold converges to 0 in probability. In case B, we observe a very heavy tail to the FDP distribution, although fewer than 1 out of every 5,000 Monte Carlo simulations yields a nonzero FDP, and no simulation yields more than 1 false discovery. In conclusion, the simulations are consistent with the results of Theorems 1 and 3.

A large FDP tail is not necessarily indicative of alarming bursty behavior for BH, as the FDP can be equal to 1 in scenarios where there is only one false discovery. Looking at the FDP simultaneously with the number of false discoveries  $V_m$  gives a clearer sense of whether the bursty behavior is alarming. In Figure 6, we run 25,000 Monte Carlo simulations using the previously de-

scribed data-driven 3-factor model. In contrast to Figure 5, we do not condition on specific realizations of the latent factor  $\mathbf{w}$  and we also plot the joint distribution of the FDP and the number of false discoveries rather than the marginal distribution of the FDP. In the simulations, we set the FDR control parameter  $q = 0.1$  and repeat the simulations for the nonnull effect size  $\mu_A \in \{2, 4, 6\}$  and for Bernoulli mixture null parameter  $\pi_0 \in \{0.9, 0.99, 0.999\}$ .

**Remark 4.** Controlling pFDR using Storey’s  $q$ -value is another popular multiple testing approach that is heralded for avoiding a flood of false positives (Storey and Tibshirani (2003)). Our simulations show that when the effect sizes are small and the number of nonnulls is sparse the pFDR will be high, implying that controlling the pFDR would mitigate the issue of burstiness in such cases. When there are many nonnulls or when the effect sizes of the nonnulls are large, the pFDR is nearly the same as the FDR (due to few simulations with no discoveries), implying that controlling the pFDR would not mitigate the burstiness in such cases.

## 7. Discussion

Here we discuss the conclusions that can be drawn from our theorems and simulations about when BH exhibits alarming burstiness and when BH is safe from burstiness concerns. We start with bursty settings.

Burstiness occurs when there are many strong, long-range correlations between the test statistics. When we model the long-range correlations via a factor model, this phenomenon can be explained by Theorem 1. By Theorem 1, the asymptotic limit of  $\text{FDP}_m \mid \mathbf{W}$  is  $qF_0(\tau_*)/\tau_*$ , a quantity that can vary drastically for different realizations of  $\mathbf{W} \sim \mathcal{N}(0, I_k)$ . The variation in  $qF_0(\tau_*)/\tau_*$  is greater when the long-range correlations are stronger (or equivalently, when the factor model loading vectors  $\mathbf{L}_{mi}$  have larger magnitude). Therefore, the FDP has high variability when there are many strong, long-range correlations between the test statistics. Meanwhile, by Theorem 2, there could be a flood of false discoveries, making the bursts severe. Our simulations from the 3-factor model fit to the DMD dataset indicate a wide right tail of the FDP distribution as well as severe bursts (see the top-left panel of Figure 1 and Figure 6). Notably, we find that sparsity of the number of nonnulls exacerbates burstiness issues. This can be explained by Corollary 5 and is observed in Figures 4 and 6.

Conversely, our theorems and simulations indicate that there are many settings where the test statistics are correlated, but the BH procedure is free of burstiness concerns. When there are no long-range correlations, no factor model is needed to model the correlations, so the variance of the FDP will decrease rapidly as the number of tests increases, even when the short-range correlations are strong. For example, in the setting of Corollary 1,  $\text{FDP}_m$  converges to a quantity less than the desired FDR control  $q$  and has variance of order  $1/m$ , even with strong short-range correlations. The simulations in the bottom of Figure 1 and all the panels of Figure 3 involve strong short-range correlations and still demonstrate this desirable behavior (the desirable behavior is not seen in

the panels of Figure 3 where  $m$  is small because, in that case, the “short-range” correlations are actually long-range relative to the number of tests). Even when there are long-range correlations but the long-range correlations are weak, modeled by a factor model with loading vectors  $\mathbf{L}_{mi}$  of small magnitude, the BH procedure will not exhibit worrisome bursts. With loading vectors  $\mathbf{L}_{mi}$  with small magnitude, the asymptotic limit of  $\text{FDP}_m | \mathbf{W}$  given in Theorem 1 will not oscillate much as  $\mathbf{W} \sim \mathcal{N}(0, I_k)$  varies. Indeed, in Figure 1, when the long-range correlations are all reduced by a factor of 10 (as we move from the top-left panel to the top-right panel), alarming burstiness is no longer observed.

### Acknowledgments

This work was supported by the National Science Foundation under grant IIS-1837931. We thank Kenneth Tay and Marius Tirlea for discussions and Jingshu Wang for the DMD data.

### References

- Andrews, D. W. and Pollard, D. (1994). An introduction to functional central limit theorems for dependent stochastic processes. *International Statistical Review*, 62(1):119–132.
- 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*, 57(1):289–300.
- 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(1):60–83.
- Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188.
- Delattre, S. and Roquain, E. (2011). On the false discovery proportion convergence under Gaussian equi-correlation. *Statistics & Probability Letters*, 81(1):111–115.
- Delattre, S. and Roquain, E. (2016). On empirical distribution function of high-dimensional Gaussian vector components with an application to multiple testing. *Bernoulli*, 22(1):302–324.
- Dudley, R. M. (1999). *Uniform Central Limit Theorems*, chapter 1, page 1–22. Cambridge Studies in Advanced Mathematics. Cambridge University Press.
- Efron, B. (2007). Correlation and large-scale simultaneous significance testing. *Journal of the American Statistical Association*, 102(477):93–103.
- Farcomeni, A. (2006). More powerful control of the false discovery rate under dependence. *Statistical Methods & Applications*, 15:43–73.
- Farcomeni, A. (2007). Some results on the control of the false discovery rate under dependence. *Scandinavian Journal of Statistics*, 34:275–297.
- Finner, H. and Roters, M. (2001). On the false discovery rate and expected type I errors. *Biometrical Journal*, 43(8):985–1005.

- Finner, H. and Roters, M. (2002). Multiple hypotheses testing and expected number of type I. errors. *The Annals of Statistics*, 30(1):220 – 238.
- Fithian, W. and Lei, L. (2020). Conditional calibration for false discovery rate control under dependence.
- Friguet, C., Kloareg, M., and Causeur, D. (2009). A factor model approach to multiple testing under dependence. *Journal of the American Statistical Association*, 104:1406–1415.
- Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control. *Annals of Statistics*, 32(3):1035–1061.
- Gerard, D. and Stephens, M. (2020). Empirical Bayes shrinkage and false discovery rate estimation, allowing for unwanted variation. *Biostatistics*, 21(1):15–32.
- Izmirlian, G. (2020). Strong consistency and asymptotic normality for quantities related to the Benjamini–Hochberg false discovery rate procedure. *Statistics and Probability Letters*, 160.
- Kim, K. I. and van de Wiel, M. A. (2008). Effects of dependence in high-dimensional multiple testing problems. *BMC Bioinformatics*, 9:114.
- Korn, E. L., Troendle, J. F., McShane, L. M., and Simon, R. (2004). Controlling the number of false discoveries: application to high-dimensional genomic data. *Journal of of Statistical Planning and Inference*, 124:379–398.
- Kotelnikova, E., Shkrob, M. A., Pyatnitskiy, M. A., Ferlini, A., and Daraselia, N. (2012). Novel approach to meta-analysis of microarray datasets reveals muscle remodeling-related drug targets and biomarkers in duchenne muscular dystrophy. *PLOS Computational Biology*, 8(2):1–10.
- Lucas, J. E., Kung, H.-N., and Chi, J.-T. A. (2010). Latent factor analysis to discover pathway-associated putative segmental aneuploidies in human cancers. *PLOS Computational Biology*, 6:e100920:1–15.
- Neuville, P. (2008). Asymptotic properties of false discovery rate controlling procedures under independence. *Electronic Journal of Statistics*, 2:1065–1110.
- Owen, A. B. (2005). Variance of the number of false discoveries. *Journal of the Royal Statistical Society, Series B*, 67(3):411–426.
- Peligrad, M. (1996). On the asymptotic normality of sequences of weak dependent random variables. *Journal of Theoretical Probability*, 9:703–715.
- Pollard, D. (1990). Empirical processes: Theory and applications. *NSF-CBMS Regional Conference Series in Probability and Statistics*, 2:i–86.
- Romano, J. P. and Shaikh, A. M. (2006). On stepdown control of the false discovery proportion. *Institute of Mathematical Statistics Lecture Notes - Monograph Series*, 49:33–50.
- Romano, J. P. and Wolf, M. (2007). Control of generalized error rates in multiple testing. *The Annals of Statistics*, 35(4):1378–1408.
- Storey, J. D. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 100(16):9440–9445.
- Sun, Y., Zhang, N. R., and Owen, A. B. (2012). Multiple hypothesis testing adjusted for latent variables, with an application to the AGEMAP gene expression data. *The Annals of Applied Statistics*, 6(4):1664–1688.
- van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press,

New York, NY.

Wang, J., Gui, L., Su, W. J., Sabatti, C., and Owen, A. B. (2020). Detecting multiple replicating signals using adaptive filtering procedures.

## Appendix A: Background needed to prove Theorems 1, 2 and 3

This appendix contains results and definitions that we draw upon in order to prove our theorems. They come from Peligrad (1996), Andrews and Pollard (1994) and van der Vaart (1998).

### A.1. Results and definitions from Peligrad (1996)

Peligrad (1996) gives a CLT for triangular arrays of dependent random variables. While that CLT does not assume stationarity, it does assume that the array is strongly mixing (i.e., that the  $\alpha$ -mixing coefficients converge to 0) and a constraint on the interlaced  $\rho$ -mixing coefficient of the array. Peligrad (1996) defines the interlaced  $\rho$ -mixing coefficients in the same way we do in Equation (4) (although we remark that there is typo in the definition originally presented in Peligrad (1996), as their equation (1.7) should involve a supremum over  $\text{dist}(T, S) \geq k$  rather than a supremum over  $k$ ).

We restate results from Theorem 2.1 in Peligrad (1996) with notation convenient for us.

**Theorem A.1.** *Let  $\{Y_{mi} : 1 \leq i \leq k_m\}$  with  $\lim_{m \rightarrow \infty} k_m = \infty$  be a triangular array of centered random variables, which is strongly mixing and has finite second moments. Assume that  $\lim_{d \rightarrow \infty} \rho_*(d) < 1$  and let  $\sigma_m^2 = \text{var}(\sum_{i=1}^{k_m} Y_{mi})$ . Suppose further that the following two conditions are satisfied*

$$\sup_m \frac{1}{\sigma_m^2} \sum_{i=1}^{k_m} \mathbb{E}(Y_{mi}^2) < \infty \quad (22)$$

and for every  $\epsilon > 0$ ,

$$\frac{1}{\sigma_m^2} \sum_{i=1}^{k_m} \mathbb{E}\left(Y_{mi}^2 I\{|Y_{mi}| > \epsilon \sigma_m\}\right) \rightarrow 0 \quad \text{as } m \rightarrow \infty. \quad (23)$$

Then  $\sigma_m^{-1} \sum_{i=1}^{k_m} Y_{mi} \xrightarrow{d} \mathcal{N}(0, 1)$  as  $m \rightarrow \infty$ .

This theorem, in conjunction with the Cramér Wold device, will later be used to establish f.d.d. convergence of our empirical process  $(W_{m,0}(\cdot), W_{m,1}(\cdot))$  from (8) to a joint Gaussian process. In order to obtain an FCLT from f.d.d. convergence we will need to introduce some definitions and results from Andrews and Pollard (1994).

## A.2. Results and definitions from Andrews and Pollard (1994)

Andrews and Pollard (1994) use a chaining argument to provide an FCLT for bounded function classes. Their FCLT does not require independence.

We first introduce some definitions from their paper. Let  $\{\Xi_{mi} : 1 \leq i \leq m < \infty\}$  be a triangular array of  $\mathcal{Y}$ -valued random elements of a measurable space. Define  $\{\tilde{\alpha}(d)\}_{d=1}^{\infty}$  to be the sequence of  $\alpha$ -mixing coefficients for this array. We use  $\mathcal{F}$  to denote a collection of bounded functions from  $\mathcal{Y}$  to  $\mathbb{R}$ .

**Definition A.1.** The seminorm function for the array  $\{\Xi_{mi} : 1 \leq i \leq m < \infty\}$  of  $\mathcal{Y}$ -valued random variables is the function  $p$  that maps any real valued function  $f$  on  $\mathcal{Y}$ , to a real number via the following equation

$$p(f) = \sup_{m,i} \sqrt{\mathbb{E}(f(\Xi_{mi})^2)}. \quad (24)$$

This  $p(\cdot)$  can be used to define a seminorm on a collection of functions on  $\mathcal{Y}$ .

Before stating the theorem we also restate Definition 2.1 of Andrews and Pollard (1994) which gives the bracketing number.

**Definition A.2.** Let  $\mathcal{F}$  be a function class and  $p$  be the seminorm function for an array of random variables  $\{\Xi_{mi} : 1 \leq i \leq m < \infty\}$ , as in Definition A.1. For  $\delta > 0$  the bracketing number  $N(\delta) = N(\delta, \mathcal{F}, p)$  is the smallest natural number  $N$  for which there exists functions  $f_1, \dots, f_N \in \mathcal{F}$  and functions  $b_1, \dots, b_N : \mathcal{Y} \rightarrow \mathbb{R}$  with  $p(b_i) \leq \delta$  for all  $i \in \{1, \dots, N\}$  such that for each  $f \in \mathcal{F}$ , there exists a  $j \in \{1, \dots, N\}$  such that  $|f - f_j| \leq b_j$ . Here,  $|f - f_j| \leq b_j$  means that  $|f(\Xi) - f_j(\Xi)| \leq b_j(\Xi)$  for all  $\Xi \in \mathcal{Y}$ , and we will use this shorthand notation in the proof of Theorem B.1.

To relate the function class  $\mathcal{F}$  to an empirical process we introduce the following operator  $\nu_m$ .

**Definition A.3.** Let  $\mathcal{F}$  be a class of real valued functions whose domain contains the support of each individual entry in the array  $\{\Xi_{mi} : 1 \leq i \leq m < \infty\}$ . For any  $m \in \mathbb{Z}_+$  and any  $f \in \mathcal{F}$ , define

$$\nu_m f \equiv \frac{1}{\sqrt{m}} \sum_{i=1}^m (f(\Xi_{mi}) - \mathbb{E}(f(\Xi_{mi}))). \quad (25)$$

This operator  $\nu_m$  maps individual functions in  $\mathcal{F}$  to centered and renormalized random variables. It defines an empirical process indexed by the elements of  $\mathcal{F}$ . We are now ready to restate the FCLT given in Corollary 2.3 of Andrews and Pollard (1994).

**Theorem A.2.** Let  $\{\Xi_{mi} : 1 \leq i \leq m < \infty\}$  be a strongly mixing triangular array whose  $\alpha$ -mixing coefficients  $\{\tilde{\alpha}(d)\}_{d=1}^{\infty}$  satisfy

$$\sum_{d=1}^{\infty} d^{Q-2} \tilde{\alpha}(d)^{\gamma/(Q+\gamma)} < \infty \quad (26)$$

for some even integer  $Q \geq 2$  and some  $\gamma > 0$ . Let  $p$  be the seminorm function for the array  $\{\Xi_{mi} : 1 \leq i \leq m < \infty\}$  as given in Definition A.1, and let  $\mathcal{F}$  be a uniformly bounded class of real-valued functions whose bracketing numbers (Definition A.2) satisfy

$$\int_0^1 x^{-\gamma/(2+\gamma)} N(x, \mathcal{F}, p)^{1/Q} dx < \infty \quad (27)$$

for the same  $Q$  and  $\gamma$ . Finally, let  $\nu_m$  be the operator from Definition A.3. If  $(\nu_m f_1, \dots, \nu_m f_k)$  converges to a multivariate Gaussian distribution for all  $k \in \mathbb{Z}_+$  and  $f_1, \dots, f_k \in \mathcal{F}$ , then  $\{\nu_m f : f \in \mathcal{F}\}$  converges in distribution to a Gaussian process indexed by  $\mathcal{F}$  with  $p$ -continuous sample paths.

We will now clarify what  $p$ -continuous sample paths and “convergence in distribution” in the above theorem mean. We start by clarifying what  $p$ -continuity of sample paths for a Gaussian process means.

**Definition A.4.** Let  $\mathcal{F}$  be a function class, let  $p$  be a seminorm, and let  $W_{\mathcal{F}}$  be a Gaussian process indexed by  $\mathcal{F}$ , defined on the probability space  $\Omega$ . Note that for each  $f \in \mathcal{F}$ ,  $W_{\mathcal{F}}(f)$  is a real valued random variable with a univariate normal distribution, and for a fixed  $\omega \in \Omega$ ,  $W_{\mathcal{F}}(f)(\omega) \in \mathbb{R}$  is a specific realization of that random variable. Fixing  $\omega \in \Omega$ , we say that  $W_{\mathcal{F}}(\cdot)(\omega)$  has a  $p$ -continuous sample path at  $f$  if for all  $\epsilon > 0$ , there exists an  $\eta > 0$  such that  $|W_{\mathcal{F}}(f)(\omega) - W_{\mathcal{F}}(g)(\omega)| < \epsilon$  holds for any  $g \in \mathcal{F}$  with  $p(f - g) < \eta$ . We say  $W_{\mathcal{F}}$  has  $p$ -continuous sample paths, if for every  $\omega \in \Omega$  and  $f \in \mathcal{F}$ ,  $W_{\mathcal{F}}(\cdot)(\omega)$  has a  $p$ -continuous sample path at  $f$ .

For a fixed  $m$ , the process  $\{\nu_m f : f \in \mathcal{F}\}$  is not guaranteed to be Borel measurable. Chapter 9 of Pollard (1990) raises caution about measurability issues of such empirical process and page 9 of Dudley (1999) provides an example showing that the empirical process of IID  $\text{Unif}[0, 1]$  random variables is not guaranteed to be measurable with respect to the sup-norm topology. To skirt around these measurability issues, we will use a modified notion of convergence provided Definition 9.1 of Pollard (1990). Introducing the modified notion of convergence allows us to use the FCLT from Andrews and Pollard (1994) (reproduced in Theorem A.2) and ultimately allows us to extend upon the results of Delattre and Roquain (2016), which use a different FCLT.

We reproduce the modified definition of convergence in distribution below. Some readers might prefer to skip this definition and to instead treat everything as measurable so that the outer expectation and the expectation are the same and so that the standard notion of convergence in distribution is interchangeable with the modified definition.

**Definition A.5.** Let  $\{X_n\}_{n=1}^{\infty}$  be a sequence of functions from a probability space  $\Omega$  into a metric space  $\mathcal{X}$  (not necessarily Borel measurable), and let  $X$  be a Borel measurable map from  $\Omega$  into  $\mathcal{X}$ . For any bounded real valued function  $f$ ,  $\mathbb{E}(f(X_n))$  is not necessarily well defined, so we define  $\mathbb{E}^*$  to be an operator that gives the outer expectation of a real valued function on  $\Omega$ . In particular, if

$H$  is a real valued function on  $\Omega$ , then

$$\mathbb{E}^*(H) \equiv \inf\{\mathbb{E}(U) : H \leq U, \text{ and } U \text{ is measurable and integrable}\}.$$

We say that  $X_n$  converges in distribution to  $X$ , if for every bounded, uniformly continuous, real valued function  $f$  on  $\mathcal{X}$ ,  $\lim_{n \rightarrow \infty} \mathbb{E}^*(f(X_n)) = \mathbb{E}(f(X))$ .

Throughout the text we will use the operator  $\xrightarrow{\mathcal{D}}$  in the statements such as  $X_n \xrightarrow{\mathcal{D}} X$  to denote convergence in distribution in the sense of Definition A.5.

### A.3. Results and definitions from van der Vaart (1998)

In the proof of Theorems 1 and 2, we will rely on some definitions and theorems from van der Vaart (1998). Most notably, we need to introduce the notion of Hadamard differentiability and the functional delta method from van der Vaart (1998). The functional delta method is defined for an alternate definition of convergence in distribution, provided in Chapter 18.2 of van der Vaart (1998), which turns out to be equivalent to Pollard's (1990) notion of convergence in distribution in Definition A.5. We reproduce the alternate definition of convergence in distribution provided in Chapter 18.2 of van der Vaart (1998) below.

**Definition A.6.** Given a sequence  $\{X_n\}_{n=1}^{\infty}$  of random variables  $X_n$  converges in distribution to  $X$ , in the sense of Chapter 18.2 in van der Vaart (1998), if  $\lim_{n \rightarrow \infty} \mathbb{E}^*(f(X_n)) = \mathbb{E}(f(X))$  for all bounded, uniformly continuous functions  $f$ , where  $\mathbb{E}^*$  denotes outer expectation.

**Remark A.1.** We can equivalently define  $\xrightarrow{\mathcal{D}}$  to denote convergence in distribution in the sense of Definition A.6 because this notion of convergence in distribution is equivalent to the one provided in Definition A.5.

*Proof.* Definition A.5 states that  $X_n$  converges in distribution to  $X$  if and only if  $\lim_{n \rightarrow \infty} \mathbb{E}^*(f(X_n)) = \mathbb{E}(f(X))$  for all bounded, uniformly continuous functions  $f$ . Definition A.6 states that  $X_n$  converges in distribution to  $X$  if and only if  $\lim_{n \rightarrow \infty} \mathbb{E}^*(f(X_n)) = \mathbb{E}(f(X))$  for all bounded, continuous functions  $f$  (which need not be uniformly continuous). Clearly, if convergence in distribution holds in the sense of Definition A.6, it will hold in the sense of Definition A.5. To show the converse note that if  $\lim_{n \rightarrow \infty} \mathbb{E}^*(f(X_n)) = \mathbb{E}(f(X))$  for all bounded, uniformly continuous functions  $f$ , then  $\lim_{n \rightarrow \infty} \mathbb{E}^*(f(X_n)) = \mathbb{E}(f(X))$  for all bounded, Lipschitz functions  $f$ , but by the Portmanteau Lemma (Lemma 18.9 in van der Vaart (1998)) this is equivalent to  $\lim_{n \rightarrow \infty} \mathbb{E}^*(f(X_n)) = \mathbb{E}(f(X))$  for all bounded, continuous functions  $f$ . The notions of convergence of distribution in Definitions A.5 and A.6 are therefore equivalent.  $\square$

In our proofs of Theorems 1 and 2, it will also be helpful to introduce the alternate definition for convergence in probability of non-measurable real valued functions from a probability space, seen in Chapter 18.2 of van der Vaart

(1998), and to derive Slutsky's lemma for our alternate notions of convergence in distribution and convergence in probability.

Below, we reproduce the alternate notion of convergence in probability seen in Chapter 18.2 of van der Vaart (1998).

**Definition A.7.** If  $\{X_n\}_{n=1}^\infty$  is a sequence of functions from a probability space  $\Omega$  into a metric space  $(\mathcal{X}, d)$ , we say that  $X_n$  converges in probability to  $X$  if for all  $\epsilon > 0$ ,  $\lim_{n \rightarrow \infty} \mathbb{P}^*(d(X_n, X) > \epsilon) = 0$ . Here,  $\mathbb{P}^*$  denotes outer probability measure since neither  $X_n$  nor  $d(X_n, X)$  needs to be Borel-measurable. Throughout the text, we will use the operator  $\xrightarrow{\mathcal{P}}$  to denote convergence in probability in the sense of this definition.

We will need to use Slutsky's lemma for our alternate definitions of convergence in probability and convergence in distribution, which we state and prove below by directly applying Theorems 18.10(v) and 18.11(i) from Chapter 18.2 of van der Vaart (1998).

**Lemma A.1.** (*Slutsky's lemma*). If  $X_n \xrightarrow{\mathcal{D}} X$  and  $Y_n \xrightarrow{\mathcal{P}} c$  for some constant  $c$ , then  $X_n + Y_n \xrightarrow{\mathcal{D}} X + c$ .

*Proof.* Suppose  $X_n \xrightarrow{\mathcal{D}} X$  and  $Y_n \xrightarrow{\mathcal{P}} c$ . Then, by Theorem 18.10(v) in van der Vaart (1998),  $(X_n, Y_n) \xrightarrow{\mathcal{D}} (X, c)$ . By the continuous mapping theorem (Theorem 18.11(i) in van der Vaart (1998)) since  $(Z_1, Z_2) \mapsto Z_1 + Z_2$  is continuous, it follows that  $X_n + Y_n \xrightarrow{\mathcal{D}} X + c$ .  $\square$

Since our proofs of Theorems 1 and 2 rely heavily on the concept of Hadamard differentiability and on the functional delta method, we reproduce the definition of Hadamard differentiability as well as the functional delta method from Chapter 20.2 of van der Vaart (1998) below.

**Definition A.8.** (Hadamard differentiability). Let  $\mathbb{D}, \mathbb{F}$  be normed spaces, let  $\phi : \mathbb{D}_\phi \rightarrow \mathbb{F}$  be a map defined on a subset  $\mathbb{D}_\phi \subseteq \mathbb{D}$ , and consider an element  $\theta \in \mathbb{D}_\phi$ . Further, let  $\mathbb{D}_0$  be a different subset of  $\mathbb{D}$ . We say that  $\phi$  is Hadamard differentiable at  $\theta$  tangentially to the subset  $\mathbb{D}_0$  if there exists a linear map  $\phi'_\theta : \mathbb{D}_0 \rightarrow \mathbb{F}$  such that for any  $h \in \mathbb{D}_0$  and collection of elements of  $\mathbb{D}$   $(h_t)_{t>0}$  satisfying both  $\lim_{t \downarrow 0} \|h_t - h\|_{\mathbb{D}} = 0$  and  $\{\theta + th_t : t > 0\} \subseteq \mathbb{D}_\phi$ ,

$$\left\| \frac{\phi(\theta + th_t) - \phi(\theta)}{t} - \phi'_\theta(h) \right\|_{\mathbb{F}} \rightarrow 0 \quad \text{as } t \downarrow 0.$$

This map  $\phi'_\theta : \mathbb{D}_0 \rightarrow \mathbb{F}$ , if it exists, is said to be the Hadamard derivative of the function  $\phi$  at  $\theta$  tangentially to the subset  $\mathbb{D}_0$ .

**Theorem A.3.** (*Functional Delta Method*) Let  $\mathbb{D}$  and  $\mathbb{F}$  be normed linear spaces and let  $\phi : \mathbb{D}_\phi \subseteq \mathbb{D} \rightarrow \mathbb{F}$  be Hadamard differentiable at  $\theta$  tangentially to  $\mathbb{D}_0 \subseteq \mathbb{D}$  with Hadamard derivative at  $\theta$  tangentially to  $\mathbb{D}_0$  given by the linear map  $\phi'_\theta : \mathbb{D}_0 \rightarrow \mathbb{F}$ . Let  $\{T_n\}_{n=1}^\infty$  be a sequence of maps from our probability space to  $\mathbb{D}_\phi$  such that  $r_n(T_n - \theta) \xrightarrow{\mathcal{D}} T$  for some sequence of numbers  $r_n \rightarrow \infty$  and a  $\mathbb{D}_0$  valued random variable  $T$ . Then  $r_n(\phi(T_n) - \phi(\theta)) \xrightarrow{\mathcal{D}} \phi'_\theta(T)$ .

*Proof.* See Theorem 20.8 in van der Vaart (1998).  $\square$

## Appendix B: Proof of Theorems 1, 2, and 3

### B.1. Helpful results for proving Theorems 1 and 2

In this section of the appendix we develop results to prove Theorems 1, 2 and 3. We start with a helpful lemma.

**Lemma B.1.** *Under the conditions of Theorem 1, there is a constant  $C < \infty$  such that for  $r \in \{0, 1\}$  and  $1 \leq i \leq m < \infty$*

$$\sup_{(t,s) \in [a,b]^2} \left| \gamma_{mir}(t) - \gamma_{mir}(s) \right| \leq C|t - s|. \quad (28)$$

*Proof.* For any  $r \in \{0, 1\}$  and  $1 \leq i \leq m < \infty$ , note that the derivative of  $\gamma_{mir}(t)$  with respect to  $t$  satisfies

$$|\gamma'_{mir}(t)| \leq \frac{\varphi(0)}{\varphi(\Phi^{-1}(t))\sqrt{1-S_L}} \leq \frac{\varphi(0)}{\sqrt{1-S_L}} \max\left\{ \frac{1}{\varphi(\Phi^{-1}(a))}, \frac{1}{\varphi(\Phi^{-1}(b))} \right\} \equiv C$$

by the argument used in the proof of Proposition 1. Since  $C$  does not depend on  $r$ ,  $i$  or  $m$ , this completes the proof.  $\square$

We now define the joint Gaussian process on  $[a, b]_2$  to which  $(W_{m,0}(\cdot), W_{m,1}(\cdot))$  converges in distribution and start by proving f.d.d. convergence.

**Definition B.1.** We define  $(\tilde{W}_0, \tilde{W}_1)$  be a joint Gaussian process on  $[a, b]_2$  with mean zero and joint covariance kernel  $c$ .

We note that  $(\tilde{W}_0, \tilde{W}_1)$  is well defined because  $c$  is a symmetric and positive semidefinite joint covariance kernel.

**Proposition B.1.** *Under the conditions of Theorem 1,  $(W_{m,0}, W_{m,1}) \xrightarrow{\text{f.d.d.}} (\tilde{W}_0, \tilde{W}_1)$ .*

*Proof.* Fix any  $k \in \mathbb{N}$  and fix any  $t_1, \dots, t_k \in [a, b]$  and  $r_1, \dots, r_k \in \{0, 1\}$ . Define  $\Sigma \in \mathbb{R}^{k \times k}$  to be the covariance matrix given by  $\Sigma_{ll'} = c^{(r_l, r_{l'})}(t_l, t_{l'})$  for  $l, l' \in \{1, \dots, k\}$ . To show f.d.d. convergence we must show that  $(W_{m,r_1}(t_1), \dots, W_{m,r_k}(t_k)) \xrightarrow{d} N(0, \Sigma)$ . By the Cramér-Wold device, it suffices to show that  $\sum_{l=1}^k a_l W_{m,r_l}(t_l) \xrightarrow{d} N(0, \mathbf{a}^\top \Sigma \mathbf{a})$  for any  $\mathbf{a} \in \mathbb{R}^k$ .

In the case where  $\mathbf{a}^\top \Sigma \mathbf{a} = 0$ ,

$$\text{var} \left( \sum_{l=1}^k a_l W_{m,r_l}(t_l) \right) = \sum_{l=1}^k \sum_{l'=1}^k a_l a_{l'} c_m^{(r_l, r_{l'})}(t_l, t_{l'}) \rightarrow \mathbf{a}^\top \Sigma \mathbf{a} = 0,$$

as  $m \rightarrow \infty$ . Then by a trivial application of Chebyshev's inequality we have  $\sum_{l=1}^k a_l W_{m,r_l}(t_l) \xrightarrow{d} N(0, \mathbf{a}^\top \Sigma \mathbf{a})$  for this case.

Now fix  $\mathbf{a} \in \mathbb{R}^k$  satisfying  $\mathbf{a}^\top \Sigma \mathbf{a} > 0$  because it remains to show  $\sum_{l=1}^k a_l W_{m,r_l}(t_l) \xrightarrow{d} \mathcal{N}(0, \mathbf{a}^\top \Sigma \mathbf{a})$  in this case. To show this, we will use a result from Peligrad (1996). For  $1 \leq i \leq m < \infty$  define

$$Y_{mi} \equiv \sum_{l=1}^k \left( H_{mir_l} I\{\bar{\Phi}(\varepsilon_{mi} + \mathbf{L}_{mi}^\top \mathbf{w} + \mu_{Ar_l}) \leq t_l\} - \pi_{r_l} \gamma_{mir_l}(t_l) \right) a_l.$$

Clearly  $\mathbb{E}(Y_{mi}) = 0$  for all  $m, i$ . Note that  $|Y_{mi}| \leq k \|\mathbf{a}\|_\infty$  which implies that  $\mathbb{E}(Y_{mi}^2) \leq k^2 \|\mathbf{a}\|_\infty^2 < \infty$ . Hence,  $\{Y_{mi} : 1 \leq i \leq m < \infty\}$  is a triangular array of centered variables with finite second moments.

Now since  $Y_{mi}$  can be written as a measurable function of  $\xi_{mi}$  for each  $m$  and  $i$ , the  $\alpha$ -mixing coefficients and the  $\rho_*$ -mixing coefficients of the array  $\{Y_{mi} : 1 \leq i \leq m < \infty\}$  must be at most as large as the  $\alpha$ -mixing and  $\rho_*$ -mixing coefficients of the array  $\{\xi_{mi} : 1 \leq i \leq m < \infty\}$  respectively. Combining this observation with Condition 1, it is clear that the array  $\{Y_{mi} : 1 \leq i \leq m < \infty\}$  is  $\alpha$ -mixing (strongly mixing) and combining the observation with Condition 2 it is clear that the  $\rho_*$  mixing coefficients of  $\{Y_{mi} : 1 \leq i \leq m < \infty\}$  have a limit that is strictly less than 1.

Therefore, to apply Theorem A.1 to the strongly mixing array  $\{Y_{mi} : 1 \leq i \leq m < \infty\}$ , it remains to check (22) and (23). To check them note that because  $\sum_{i=1}^m Y_{mi} = \sqrt{m} \sum_{l=1}^k a_l W_{m,r_l}(t_l)$ ,

$$\sigma_m^2 \equiv \text{var} \left( \sum_{i=1}^m Y_{mi} \right) = m \sum_{l=1}^k \sum_{l'=1}^k a_l a_{l'} c_m^{(r_l, r_{l'})}(t_l, t_{l'}) = m(\mathbf{a}^\top \Sigma \mathbf{a} + o(1)). \quad (29)$$

Since clearly  $\sum_{i=1}^m Y_{mi}$  takes on different values with positive probability it follows that  $\sigma_m^2/m > 0$  for all  $m$ . Combining this with (29) and  $\mathbf{a}^\top \Sigma \mathbf{a} > 0$ , it follows that  $\inf_m \sigma_m^2/m > 0$ . Inequality (22) holds for our array  $\{Y_{mi} : 1 \leq i \leq m < \infty\}$  because

$$\sup_m \frac{1}{\sigma_m^2} \sum_{i=1}^m E(Y_{mi}^2) \leq \sup_m \frac{mk^2 \|\mathbf{a}\|_\infty^2}{\sigma_m^2} = \frac{k^2 \|\mathbf{a}\|_\infty^2}{\inf_m \sigma_m^2/m} < \infty.$$

It is easy to see that (23) holds because  $\lim_{m \rightarrow \infty} \sigma_m = \infty$  while for all  $m, i$ ,  $|Y_{mi}|$  is bounded above by the finite value  $k \|\mathbf{a}\|_\infty$ .

Since (22) and (23) both hold, by Theorem A.1  $\sigma_m^{-1} \sum_{i=1}^m Y_{mi} \xrightarrow{d} \mathcal{N}(0, 1)$  as  $m \rightarrow \infty$ . Therefore,

$$\frac{\sum_{l=1}^k a_l W_{m,r_l}(t_l)}{\sqrt{\sigma_m^2/m}} = \frac{\sum_{i=1}^m Y_{mi}}{\sigma_m} \xrightarrow{d} \mathcal{N}(0, 1).$$

By (29),  $\lim_{m \rightarrow \infty} \sqrt{\sigma_m^2/m}/\sqrt{\mathbf{a}^\top \Sigma \mathbf{a}} = 1$ , so multiplying each side of the above stochastic convergence by  $\sqrt{\sigma_m^2/m}/\sqrt{\mathbf{a}^\top \Sigma \mathbf{a}}$  and applying Slutsky's lemma shows that  $\sum_{l=1}^k a_l W_{m,r_l}(t_l) \xrightarrow{d} \mathcal{N}(0, \mathbf{a}^\top \Sigma \mathbf{a})$  as desired.  $\square$

Now that we have established  $(W_{m,0}, W_{m,1}) \xrightarrow{\text{f.d.d.}} (\tilde{W}_0, \tilde{W}_1)$  we can derive an FCLT for  $(W_{m,0}, W_{m,1})$  by applying a result from Andrews and Pollard (1994), which is based on a chaining argument. The FCLT uses a slightly modified definition of weak convergence which can be applied to a sequence of non-Borel measurable  $D[a, b]_2$ -valued random variables (see Definition A.5 or equivalently Definition A.6.)

**Theorem B.1.** *Under the conditions of Theorem 1,  $(W_{m,0}, W_{m,1}) \xrightarrow{D} (W_0, W_1)$  where  $(W_0, W_1) \in C[a, b]_2$  almost surely and  $(W_0, W_1)$  is a joint Gaussian process with mean zero and joint covariance kernel  $c$ .*

*Proof.* The proof is lengthy and so we split it into a sequences of steps as follows:

- 1) Defining the function class  $\mathcal{F}$  and seminorm  $p$  that we need in order to use Theorem A.2, the FCLT from Andrews and Pollard (1994),
- 2) computing the bracketing number  $N(x, \mathcal{F}, p)$ ,
- 3) checking the FCLT conditions,
- 4) applying the FCLT,
- 5) defining limiting processes  $W_0$  and  $W_1$ , and finally
- 6) verifying that  $(W_0, W_1) \in C[a, b]_2$ .

**Defining the relevant function class,  $\mathcal{F}$ , and seminorm,  $p$ .** First, for each  $m, i$  recall that  $\tilde{\varepsilon}_{mi} = \varepsilon_{mi} / \sqrt{1 - \|\mathbf{L}_{mi}\|_2^2}$  has unit variance. Let  $\{\tilde{\xi}_{mi} : 1 \leq i \leq m < \infty\}$  be an expansion of our triangular array where  $\xi_{mi} = (\tilde{\varepsilon}_{mi}, H_{mi}, \mathbf{L}_{mi}, \mathbf{w})$ . Because  $\mathbf{L}_{mi}$  and  $\mathbf{w}$  are deterministic and  $\|\mathbf{L}_{mi}\|_2^2$  is bounded away from 1 by Condition 3, the arrays  $\{\tilde{\xi}_{mi} : 1 \leq i \leq m < \infty\}$  and  $\{(H_{mi}, \varepsilon_{mi}) : 1 \leq i \leq m < \infty\}$  have the same  $\alpha$ -mixing coefficients. These coefficients  $\{\alpha(d)\}_{d=1}^\infty$  satisfy Condition 1. Now define  $\mathcal{Y} = \mathbb{R} \times \{0, 1\} \times \mathbb{R}^k \times \mathbb{R}^k$ , and define  $\tilde{f} : \mathcal{Y} \times ([a, b] \times \{0, 1\}) \rightarrow [0, 1]$  via

$$\tilde{f}((\tilde{\varepsilon}, H, \mathbf{L}, \mathbf{w}), (t, r)) = I\{H = r\} I\left\{\bar{\Phi}(\tilde{\varepsilon}) \leq \bar{\Phi}\left(\frac{\bar{\Phi}^{-1}(t) - \mathbf{L}^\top \mathbf{w} - \mu_A r}{\sqrt{1 - \|\mathbf{L}\|_2^2}}\right)\right\}.$$

For  $(t, r) \in [a, b] \times \{0, 1\}$ , note that  $\tilde{f}(\tilde{\xi}_{mi}, (t, r)) = H_{mir} I\{\bar{\Phi}(\tilde{\varepsilon}_{mi}) \leq \gamma_{mir}(t)\}$ .

The function class we will use includes the various restrictions of  $\tilde{f}$  to a fixed  $t \in [a, b]$  and  $r \in \{0, 1\}$  while  $\tilde{\varepsilon}, H, \mathbf{L}$  and  $\mathbf{w}$  vary:

$$\mathcal{F} \equiv \{\tilde{f}(\cdot, (t, r)) : (t, r) \in [a, b] \times \{0, 1\}\}. \quad (30)$$

Letting  $p$  be the seminorm function for the array  $\{\tilde{\xi}_{mi} : 1 \leq i \leq m < \infty\}$  given by Definition A.1, observe that for  $t, s \in [a, b]$  and  $r \in \{0, 1\}$  a simple computation gives

$$p(|\tilde{f}(\cdot, (t, r)) - \tilde{f}(\cdot, (s, r))|) = \sup_{m,i} \sqrt{\pi_r^{(m)}} |\gamma_{mij}(t) - \gamma_{mij}(s)| \leq \sqrt{C|t - s|}, \quad (31)$$

where the inequality follows from Lemma B.1.

**Computing the bracketing number  $N(x, \mathcal{F}, p)$ .** To apply the FCLT (Theorem A.2), we must first compute an upper bound on the bracketing number  $N(x, \mathcal{F}, p)$  for each  $x > 0$ . Fixing  $x > 0$ , take  $N_* \equiv \lceil C(b-a)/x^2 \rceil$ . For each  $l \in \{1, \dots, N_*, N_* + 1\}$  and  $r \in \{0, 1\}$  define

$$f_{l,r} = \tilde{f}(\cdot, (a + (b-a)(l-1)/N_*, r)),$$

and for  $1 \leq l \leq N_*$  define  $b_{l,r} = f_{l+1,r} - f_{l,r}$ . Note for a given  $f \in \mathcal{F} \setminus \{f_{1,0}, f_{1,1}\}$ , since  $f = \tilde{f}(\cdot, (t_*, r_*))$  for some  $(t_*, r_*) \in (a, b] \times \{0, 1\}$  we can choose  $r = r_*$  and  $l \in \{1, \dots, N_*\}$  such that  $t_* \in (a + (b-a)(l-1)/N_*, a + (b-a)l/N_*]$ , and it is easy to check that  $|f - f_{l,r}| \leq b_{l,r}$ . Adding the trivial cases of  $f \in \{f_{1,0}, f_{1,1}\}$  it follows that for any  $f \in \mathcal{F}$ , there exists an  $l \in \{1, \dots, N_*\}$  and  $r \in \{0, 1\}$  such that  $|f - f_{l,r}| \leq b_{l,r}$ . Furthermore, for any  $l \in \{1, \dots, N_*\}$  and  $r \in \{0, 1\}$ , by applying Inequality (31),

$$p(b_{l,r}) = p(f_{l+1,r} - f_{l,r}) \leq \sqrt{C(b-a)/N_*} \leq x.$$

Thus, we showed that there exists functions  $f_{1,0}, \dots, f_{N_*,0}, f_{1,1}, \dots, f_{N_*,1} \in \mathcal{F}$  and functions  $b_{1,0}, \dots, b_{N_*,0}, b_{1,1}, \dots, b_{N_*,1} : \mathcal{Y} \rightarrow \mathbb{R}$ , satisfying  $p(b_{l,r}) \leq x$  for all  $(l, r) \in \{1, \dots, N_*\} \times \{0, 1\}$ , such that for any  $f \in \mathcal{F}$  there exists an  $(l, r) \in \{1, \dots, N_*\} \times \{0, 1\}$  for which  $|f - f_{l,r}| \leq b_{l,r}$ . Hence, by Definition A.2,  $N(x, \mathcal{F}, p) \leq 2N_* = 2\lceil C(b-a)/x^2 \rceil$ , and this argument holds for any  $x > 0$ .

**Checking the conditions for the FCLT of Andrews and Pollard (1994) (Theorem A.2.)** Letting  $Q > 2$  be the even integer and  $\gamma > 0$  be the number guaranteed by Condition 1, we find that

$$\int_0^1 x^{-\frac{\gamma}{2+\gamma}} N(x, \mathcal{F}, p)^{\frac{1}{Q}} dx \leq \int_0^1 x^{-\frac{\gamma}{2+\gamma}} \left( 2 \left\lceil \frac{C(b-a)}{x^2} \right\rceil \right)^{\frac{1}{Q}} dx < \infty. \quad (32)$$

The integral above is finite because  $\gamma/(2+\gamma) + 2/Q < 1$ . The  $\alpha$ -mixing coefficients of  $\{\tilde{\xi}_{mi}, 1 \leq i \leq m < \infty\}$  are  $\alpha(d)$  which satisfy Condition 1. Therefore,

$$\sum_{d=1}^{\infty} d^{Q-2} \alpha(d)^{\frac{\gamma}{Q+\gamma}} < \infty. \quad (33)$$

**Applying the FCLT of Andrews and Pollard (1994) (Theorem A.2.)**

By (32) and (33), since  $\mathcal{F}$  is a uniformly bounded class of real-valued functions, and since  $\{\tilde{\xi}_{mi}, 1 \leq i \leq m < \infty\}$  is strongly mixing triangular array, we are in the setting of Theorem A.2. To obtain an FCLT it therefore remains to check that for any  $f_1, \dots, f_k \in \mathcal{F}$ ,  $(\nu_m f_1, \dots, \nu_m f_k)$  converges in distribution to a multivariate Gaussian as  $m \rightarrow \infty$ , where  $\nu_m$  is the operator from Definition A.3. This is indeed the case because for any  $f \in \mathcal{F}$  there exists a  $(t, r) \in [a, b] \times \{0, 1\}$  such that  $\nu_m f = W_{m,r}(t)$  and because by Proposition B.1  $(W_{m,0}, W_{m,1})$  converges in f.d.d to a Gaussian process on  $[a, b] \times \{0, 1\}$ . Thus applying Theorem A.2, it follows that the stochastic process  $\{\nu_m f : f \in \mathcal{F}\}$  converges in distribution to a Gaussian process indexed by  $\mathcal{F}$  which has  $p$ -continuous sample paths in the sense of Definition A.4, where convergence in distribution is in the sense of Definition A.5.

**Defining  $(W_0, W_1)$  in terms of  $W_{\mathcal{F}}$ .** Let  $W_{\mathcal{F}}$  be the Gaussian process on  $\mathcal{F}$  to which  $\{\nu_m f : f \in \mathcal{F}\}$  converges. For  $(t, r) \in [a, b] \times \{0, 1\}$  we define  $W_r(t) \equiv W_{\mathcal{F}}(\tilde{f}(\cdot, (t, r)))$ . By considering the correspondence between  $\mathcal{F}$  and  $[a, b] \times \{0, 1\}$  where for any  $f \in \mathcal{F}$ ,  $\nu_m f = W_{m,r}(t)$  for some unique  $(t, r) \in [a, b] \times \{0, 1\}$  because  $\{\nu_m f : f \in \mathcal{F}\} \xrightarrow{\mathcal{D}} W_{\mathcal{F}}$  it is clear that  $(W_{m,0}, W_{m,1}) \xrightarrow{\mathcal{D}} (W_0, W_1)$ . This of course implies that  $(W_{m,0}, W_{m,1}) \xrightarrow{\text{f.d.d.}} (W_0, W_1)$ . But since by Proposition B.1,  $(W_{m,0}, W_{m,1}) \xrightarrow{\text{f.d.d.}} (\tilde{W}_0, \tilde{W}_1)$  it follows that  $(W_0, W_1)$  has the same finite dimensional distribution as  $(\tilde{W}_0, \tilde{W}_1)$ , so  $(W_0, W_1)$  is indeed a joint Gaussian process with mean zero and joint covariance kernel  $c$ .

**Checking that  $(W_0, W_1) \in C[a, b]_2$  always, using  $p$ -continuity of  $W_{\mathcal{F}}$ .** It remains to show that  $(W_0, W_1) \in C[a, b]_2$  always. To check this let  $\Omega$  be the probability space on which  $\{\xi_{mi}\}$  are defined. Fix any  $\omega \in \Omega$ , and we will show that  $(W_0(\cdot)(\omega), W_1(\cdot)(\omega)) \in C[a, b]_2$ . To do this, further fix any  $t \in [a, b]$ ,  $r \in \{0, 1\}$  and  $\epsilon > 0$ . Let  $g = \tilde{f}(\cdot, (t, r))$ . Since  $W_{\mathcal{F}}(\cdot)(\omega)$  is  $p$ -continuous at  $g$ , by Definition A.4, there exists an  $\eta > 0$  such that if  $f \in \mathcal{F}$  with  $p(f - g) < \eta$ , then  $|W_{\mathcal{F}}(f)(\omega) - W_{\mathcal{F}}(g)(\omega)| < \epsilon$ . Taking such an  $\eta$ , define  $\eta_* = \eta^2/C$ . We must show that  $|W_r(t)(\omega) - W_r(s)(\omega)| < \epsilon$  for any  $s \in [a, b]$  with  $|t - s| < \eta_*$ . Let  $s \in [a, b]$  with  $|t - s| < \eta_*$ . Define  $h = \tilde{f}(\cdot, (s, r)) \in \mathcal{F}$ . By (31),

$$p(h - g) \leq \sqrt{C|t - s|} < \sqrt{C\eta_*} = \eta$$

which, by a previous claim, further implies  $|W_{\mathcal{F}}(h)(\omega) - W_{\mathcal{F}}(g)(\omega)| < \epsilon$ . Now from definition of  $W_r$ , we get  $|W_r(s)(\omega) - W_r(t)(\omega)| < \epsilon$ . Thus, we have shown that for any  $\epsilon > 0$  there exists  $\eta_* > 0$  such that if  $s \in [a, b]$  and  $|t - s| < \eta_*$  then  $|W_r(t)(\omega) - W_r(s)(\omega)| < \epsilon$ . Hence,  $W_r(\cdot)(\omega)$  is continuous at  $t$ . Since this argument holds for any  $(t, r) \in [a, b] \times \{0, 1\}$ ,  $(W_0(\cdot)(\omega), W_1(\cdot)(\omega)) \in C[a, b]_2$ . Since this holds for any  $\omega \in \Omega$ ,  $(W_0, W_1) \in C[a, b]_2$  always.  $\square$

With some manipulation of the previous result, we can obtain the following result, which sets us up for using the functional delta method (Theorem A.3).

**Corollary B.1.** *Under the conditions of Theorem 1,  $(\hat{W}_{m,0}, \hat{W}_{m,1}) \xrightarrow{\mathcal{D}} (W_0, W_1)$  where  $(W_0, W_1) \in C[a, b]_2$  almost surely and  $(W_0, W_1)$  is a joint Gaussian process with mean zero and joint covariance kernel  $c$ .*

*Proof.* Observe that

$$(\hat{W}_{m,0}, \hat{W}_{m,1}) = (W_{m,0}, W_{m,1}) + (\sqrt{m}(F_{m,0} - F_0), \sqrt{m}(F_{m,1} - F_1))$$

where by Condition 8, the right hand term deterministically converges in  $C[a, b]_2$  with respect to the sup-norm to  $(0, 0) \in C[a, b]_2$ , and hence, it also converges in probability to  $(0, 0)$ . Therefore, by Slutsky's lemma (Lemma A.1) and Theorem B.1,  $(\hat{W}_{m,0}, \hat{W}_{m,1}) \xrightarrow{\mathcal{D}} (W_0, W_1)$ . Also by Theorem B.1,  $(W_0, W_1) \in C[a, b]_2$  almost surely and  $(W_0, W_1)$  is a joint Gaussian process with mean zero and joint covariance kernel  $c$ .  $\square$

To derive CLTs for  $FDP_m$  and  $V_m/m$  from an FCLT for  $(\hat{W}_{m,0}, \hat{W}_{m,1})$  we must apply the functional delta method. To make this precise, we define some functions. First, we define  $\mathcal{T} : D[a, b] \rightarrow [a, b]$  via

$$\mathcal{T}(H) = \sup\{u \in [a, b] : H(u) \geq u/q\} \quad (34)$$

for all  $H \in D[a, b]$ . The Simes point is  $\tau_* = \mathcal{T}(G)$ . Next, define  $\mathcal{R} : D[a, b]_2 \rightarrow \mathbb{R}$  as

$$\mathcal{R}(H_0, H_1) = (H_0 + H_1)(\mathcal{T}(H_0 + H_1)) \quad (35)$$

for all  $(H_0, H_1) \in D[a, b]_2$ . Now for  $(H_0, H_1) \in D[a, b]_2$ , define real-valued FPR and FDP functions

$$\Psi^{(\text{FPR})}(H_0, H_1) = H_0(\mathcal{T}(H_0 + H_1)), \quad \text{and} \quad (36)$$

$$\Psi^{(\text{FDP})}(H_0, H_1) = \frac{H_0(\mathcal{T}(H_0 + H_1))}{(H_0 + H_1)(\mathcal{T}(H_0 + H_1))} = \frac{\Psi^{(\text{FPR})}(H_0, H_1)}{\mathcal{R}(H_0, H_1)}. \quad (37)$$

Our goal is to compute the Hadamard derivative of both  $\Psi^{(\text{FPR})}$  and  $\Psi^{(\text{FDP})}$  at the point  $(F_0, F_1) \in C[a, b]_2$  tangentially to  $C[a, b]_2$ . While the supplement of Delattre and Roquain (2016) calculates the Hadamard derivative of  $\Psi^{(\text{FDP})}$  at  $(F_0, F_1)$  in the case where  $F_0(t) = \pi_0 t$ , our Hadamard derivative calculation comes out differently because, in our setting,  $F_0$  is not necessarily linear with slope  $\pi_0$ . To compute the Hadamard derivative for  $\Psi^{(\text{FPR})}$  and  $\Psi^{(\text{FDP})}$  at the point  $(F_0, F_1)$  tangentially to  $C[a, b]_2$ , it is helpful to first calculate the Hadamard derivative of  $\mathcal{T}$  at  $(F_0, F_1)$  tangentially to  $C[a, b]_2$ . To compute the latter, we mimic the calculations seen in Neuvial (2008), reproving some of their results in order to generalize from functions defined on  $D[0, 1]$  to functions defined on  $D[a, b]$  and to situations where  $F_0$  is not linear with slope  $\pi_0$ . We chose to reproduce the proof for Hadamard differentiability of  $\mathcal{T}$  at  $(F_0, F_1)$  tangentially to  $C[a, b]_2$  rather than reference to the result from Neuvial (2008) in order to avoid generalization and notation misinterpretation errors and in order to familiarize the reader with Hadamard derivative calculations which, regardless, are needed for  $\Psi^{(\text{FPR})}$  in our setting. The proofs of Hadamard differentiability and the calculations of the Hadamard derivative of both  $\Psi^{(\text{FPR})}$  and  $\Psi^{(\text{FDP})}$  at the point  $(F_0, F_1) \in C[a, b]_2$  tangentially to  $C[a, b]_2$  are carried out in Appendix C. Some readers may prefer to skip Appendix C because it only contains Hadamard differentiability proofs and because the proof methodology is similar to that in Neuvial (2008).

With the Hadamard derivatives of  $\Psi^{(\text{FDP})}$  and  $\Psi^{(\text{FPR})}$  at  $(F_0, F_1)$ , we can apply the functional delta method (Theorem A.3) to the result of Theorem B.1 to get a CLT for  $\Psi^{(\text{FDP})}(\hat{F}_{m,0}, \hat{F}_{m,1})$  and  $\Psi^{(\text{FPR})}(\hat{F}_{m,0}, \hat{F}_{m,1})$ . However, we are primarily interested in CLTs for  $FDP_m$  and  $V_m/m$ , so we prove the following lemma to help us derive CLTs for  $FDP_m$  and  $V_m/m$  using CLTs for  $\Psi^{(\text{FDP})}(\hat{F}_{m,0}, \hat{F}_{m,1})$  and  $\Psi^{(\text{FPR})}(\hat{F}_{m,0}, \hat{F}_{m,1})$ .

**Lemma B.2.** *Under the conditions of Theorem 1,*

$$\begin{aligned}\sqrt{m}(\text{FDP}_m - \Psi^{(\text{FDP})}(\hat{F}_{m,0}, \hat{F}_{m,1})) &\xrightarrow{\mathcal{P}} 0 \quad \text{and} \\ \sqrt{m}(V_m/m - \Psi^{(\text{FPR})}(\hat{F}_{m,0}, \hat{F}_{m,1})) &\xrightarrow{\mathcal{P}} 0\end{aligned}$$

as  $m \rightarrow \infty$ , where  $\xrightarrow{\mathcal{P}}$  denotes convergence in probability in the sense of Definition A.7.

*Proof.* To prove this, define  $\eta \equiv (G(a) - a/q)/2 > 0$ . We will show that whenever  $|\hat{G}_m(a) - G(a)| < \eta$ , both  $\text{FDP}_m = \Psi^{(\text{FDP})}(\hat{F}_{m,0}, \hat{F}_{m,1})$  and  $V_m/m = \Psi^{(\text{FPR})}(\hat{F}_{m,0}, \hat{F}_{m,1})$ . To do this, note that  $|\hat{G}_m(a) - G(a)| < \eta \Rightarrow \hat{G}_m(a) > a/q$  because if  $|\hat{G}_m(a) - G(a)| < \eta$ , then

$$\hat{G}_m(a) - \frac{a}{q} = \hat{G}_m(a) - G(a) + G(a) - \frac{a}{q} = \hat{G}_m(a) - G(a) + 2\eta > \eta > 0.$$

Thus, letting  $\tau_{\text{BH},m}$  be the Benjamini-Hochberg threshold when applying the procedure at level- $q$  to the first  $m$   $p$ -values (with ECDF  $\hat{G}_m$ ),

$$|\hat{G}_m(a) - G(a)| < \eta \Rightarrow \tau_{\text{BH},m} = \sup\left\{t \in [0, 1] : \hat{G}_m(t) \geq \frac{t}{q}\right\} \geq a.$$

Noting that  $\tau_{\text{BH},m} \leq b$  always (because  $b > q$  and  $\hat{G}_m \leq 1$ ), it follows that  $\tau_{\text{BH},m} = \sup\{t \in [a, b] : \hat{G}_m(t) \geq t/q\} = \mathcal{T}(\hat{G}_m)$  whenever  $|\hat{G}_m(a) - G(a)| < \eta$ . Now, since

$$\text{FDP}_m = \frac{\hat{F}_{m,0}(\tau_{\text{BH},m})}{\hat{G}_m(\tau_{\text{BH},m})} \quad \text{and} \quad \Psi^{(\text{FDP})}(\hat{F}_{m,0}, \hat{F}_{m,1}) = \frac{\hat{F}_{m,0}(\mathcal{T}(\hat{G}_m))}{\hat{G}_m(\mathcal{T}(\hat{G}_m))},$$

while

$$\frac{V_m}{m} = \hat{F}_{m,0}(\tau_{\text{BH},m}) \quad \text{and} \quad \Psi^{(\text{FPR})}(\hat{F}_{m,0}, \hat{F}_{m,1}) = \hat{F}_{m,0}(\mathcal{T}(\hat{G}_m)),$$

it is easy to see that whenever  $|\hat{G}_m(a) - G(a)| < \eta$ , both  $\text{FDP}_m = \Psi^{(\text{FDP})}(\hat{F}_{m,0}, \hat{F}_{m,1})$  and  $V_m/m = \Psi^{(\text{FPR})}(\hat{F}_{m,0}, \hat{F}_{m,1})$ .

Now note that by Proposition B.1,  $(W_{m,0}(a), W_{m,1}(a)) \xrightarrow{d} (\tilde{W}_0(a), \tilde{W}_1(a))$ . Since Condition 8 implies  $\sqrt{m}(F_{m,0}(a) - F_0(a), F_{m,1}(a) - F_1(a)) \xrightarrow{\mathcal{P}} 0$ , by Slutsky's lemma we get

$$\begin{bmatrix} \hat{W}_{m,0}(a) \\ \hat{W}_{m,1}(a) \end{bmatrix} = \begin{bmatrix} W_{m,0}(a) \\ W_{m,1}(a) \end{bmatrix} + \sqrt{m} \begin{bmatrix} F_{m,0}(a) - F_0(a) \\ F_{m,1}(a) - F_1(a) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \tilde{W}_0(a) \\ \tilde{W}_1(a) \end{bmatrix}.$$

Then the continuous mapping theorem implies that  $\hat{W}_{m,0}(a) + \hat{W}_{m,1}(a) \xrightarrow{d} \tilde{W}_0(a) + \tilde{W}_1(a)$  or, equivalently, that  $\sqrt{m}(\hat{G}_m(a) - G(a)) \xrightarrow{d} \tilde{W}_0(a) + \tilde{W}_1(a)$  where  $\tilde{W}_0(a) + \tilde{W}_1(a)$  is a Gaussian random variable with finite variance. It

follows that  $|\hat{G}_m(a) - G(a)| \xrightarrow{\mathcal{P}} 0$  which, by the definition of convergence in probability, implies that  $\lim_{m \rightarrow \infty} P(|\hat{G}_m(a) - G(a)| \geq \eta) = 0$ . Now fixing  $\epsilon > 0$  and letting  $P^*$  denote outer probability measure, it follows that

$$\begin{aligned} 0 &= \limsup_{m \rightarrow \infty} P(|\hat{G}_m(a) - G(a)| \geq \eta) \\ &\geq \limsup_{m \rightarrow \infty} P^*\left(\text{FDP}_m \neq \Psi^{(\text{FDP})}(\hat{F}_{m,0}, \hat{F}_{m,1})\right) \\ &\geq \limsup_{m \rightarrow \infty} P^*\left(\sqrt{m}(\text{FDP}_m - \Psi^{(\text{FDP})}(\hat{F}_{m,0}, \hat{F}_{m,1})) > \epsilon\right) \end{aligned}$$

by the fact that  $\text{FDP}_m \neq \Psi^{(\text{FDP})}(\hat{F}_{m,0}, \hat{F}_{m,1}) \Rightarrow |\hat{G}_m(a) - G(a)| \geq \eta$  and by monotonicity of outer probability measure. A similar argument shows that,

$$0 \geq \limsup_{m \rightarrow \infty} P^*\left(\sqrt{m}(V_m/m - \Psi^{(\text{FPR})}(\hat{F}_{m,0}, \hat{F}_{m,1})) > \epsilon\right).$$

Since this holds for any  $\epsilon > 0$  and since outer measure is nonnegative, by the definition of convergence in probability in the sense of Definition A.7 we get the desired result.  $\square$

## B.2. Proofs of Theorem 1 and 2

Assume that all of the conditions of Theorem 1 hold. By our definition for  $\hat{W}_{m,0}$  and  $\hat{W}_{m,1}$  in (8) and by Corollary B.1,

$$\sqrt{m}((\hat{F}_{m,0}, \hat{F}_{m,1}) - (F_0, F_1)) \xrightarrow{\mathcal{D}} (W_0, W_1) \quad (38)$$

where  $(W_0, W_1) \in C[a, b]_2$  and is a joint Gaussian process with mean zero and joint covariance kernel  $c$ .

Since according to Propositions C.4 and C.2, both  $\Psi^{(\text{FDP})}$  and  $\Psi^{(\text{FPR})}$  are Hadamard differentiable at  $(F_0, F_1)$  tangentially to  $C[a, b]_2$ , by applying the functional delta method (Theorem A.3) to (38) we get

$$\sqrt{m}\left(\Psi^{(\text{FDP})}(\hat{F}_{m,0}, \hat{F}_{m,1}) - \Psi^{(\text{FDP})}(F_0, F_1)\right) \xrightarrow{\mathcal{D}} \dot{\Psi}_{(F_0, F_1)}^{(\text{FDP})}(W_0, W_1) \quad (39)$$

and

$$\sqrt{m}\left(\Psi^{(\text{FPR})}(\hat{F}_{m,0}, \hat{F}_{m,1}) - \Psi^{(\text{FPR})}(F_0, F_1)\right) \xrightarrow{\mathcal{D}} \dot{\Psi}_{(F_0, F_1)}^{(\text{FPR})}(W_0, W_1). \quad (40)$$

Now, by Lemma B.2,  $\sqrt{m}(\text{FDP}_m - \Psi^{(\text{FDP})}(\hat{F}_{m,0}, \hat{F}_{m,1})) \xrightarrow{\mathcal{P}} 0$  and  $\sqrt{m}(V_m/m - \Psi^{(\text{FPR})}(\hat{F}_{m,0}, \hat{F}_{m,1})) \xrightarrow{\mathcal{P}} 0$ , so adding these terms that converge to 0 in probability to the left hand sides of (39) and (40), respectively, and applying Slutsky's lemma (Lemma A.1), we find that

$$\sqrt{m}\left(\text{FDP}_m - \Psi^{(\text{FDP})}(F_0, F_1)\right) \xrightarrow{\mathcal{D}} \dot{\Psi}_{(F_0, F_1)}^{(\text{FDP})}(W_0, W_1) \quad (41)$$

and

$$\sqrt{m} \left( \frac{V_m}{m} - \Psi^{(\text{FPR})}(F_0, F_1) \right) \xrightarrow{\mathcal{D}} \dot{\Psi}_{(F_0, F_1)}^{(\text{FPR})}(W_0, W_1). \quad (42)$$

Now noting that  $\Psi^{(\text{FDP})}(F_0, F_1) = F_0(\tau_*)/G(\tau_*) = qF_0(\tau_*)/\tau_*$  and that  $\Psi^{(\text{FPR})}(F_0, F_1) = F_0(\tau_*)$  and recalling the formulas for  $\dot{\Psi}_{(F_0, F_1)}^{(\text{FDP})}$  and  $\dot{\Psi}_{(F_0, F_1)}^{(\text{FPR})}$  from Propositions C.2 and C.4, (41) and (42) simplify to

$$\sqrt{m} \left( \text{FDP}_m - \frac{qF_0(\tau_*)}{\tau_*} \right) \xrightarrow{\mathcal{D}} \frac{q}{\tau_*} \left( (1 + \alpha)W_0(\tau_*) + \alpha W_1(\tau_*) \right) \quad (43)$$

and

$$\sqrt{m} \left( \frac{V_m}{m} - F_0(\tau_*) \right) \xrightarrow{\mathcal{D}} W_0(\tau_*) + \frac{F'_0(\tau_*)}{c_G} (W_0(\tau_*) + W_1(\tau_*)). \quad (44)$$

We now remark that for a finite  $m$ ,  $\text{FDP}_m$  and  $V_m$  are Borel measurable random variables, because they take on finitely many possible values and each value they take on can be written as finite unions and intersections of measurable sets. Because  $\text{FDP}_m$  and  $V_m$  are Borel measurable random variables, the notion of convergence in distribution in Definition A.6 that is referenced in (43) and (44) is interchangeable with the standard notion of convergence in distribution, which we denote with the symbol  $\xrightarrow{d}$ . Combining this with the fact that  $(W_0, W_1)$  is a joint Gaussian process with mean zero and joint covariance kernel  $c$ , we get that (43) and (44) simplify to

$$\sqrt{m} \left( \text{FDP}_m - \frac{qF_0(\tau_*)}{\tau_*} \right) \xrightarrow{d} \mathcal{N}(0, \sigma_L^2) \quad (45)$$

and

$$\sqrt{m} \left( \frac{V_m}{m} - F_0(\tau_*) \right) \xrightarrow{d} \mathcal{N}(0, \sigma_R^2) \quad (46)$$

where

$$\sigma_L^2 \equiv \frac{q^2}{\tau_*^2} \left( (1 + \alpha)^2 c^{(0,0)}(\tau_*, \tau_*) + 2\alpha(1 + \alpha)c^{(1,0)}(\tau_*, \tau_*) + \alpha^2 c^{(1,1)}(\tau_*, \tau_*) \right)$$

and where, defining  $\beta \equiv F'_0(\tau_*)/c_G = F'_0(\tau_*)/(1/q - G'(\tau_*))$ ,

$$\sigma_R^2 \equiv (1 + \beta)^2 c^{(0,0)}(\tau_*, \tau_*) + 2\beta(1 + \beta)c^{(1,0)}(\tau_*, \tau_*) + \beta^2 c^{(1,1)}(\tau_*, \tau_*).$$

Since equations (45) and (46) hold under the conditions of Theorem 1, we have thus proved Theorems 1 and 2.  $\square$

### B.3. Proof of Theorem 3

Set  $[a, b] = [0, 1]$  and fix the  $\mathbf{W} = \mathbf{w} \in \mathbb{R}^k$  on which we condition and suppose that the conditions of Theorem 3 hold. Fix  $\epsilon \in (0, q)$ , and we will show that  $\lim_{m \rightarrow \infty} \Pr(\tau_{\text{BH}, m} \leq \epsilon) = 1$ . To do this, define  $\psi_G : [0, 1] \rightarrow \mathbb{R}$ , to be the function given by  $\psi_G(t) = G(t) - t/q$ . Since  $\tau_* = 0$ ,  $\psi_G(t) < 0$  for all  $t \in (0, 1)$ .

By Proposition 1,  $\psi_G$  is continuous on  $[\epsilon, q]$ , so by the extreme value theorem,  $\psi_G$  attains a maximum at some  $t_{\max} \in [\epsilon, q]$ . Because  $\psi_G(t_{\max}) < 0$ , it follows that letting  $\delta = -\psi_G(t_{\max}) > 0$ ,  $\psi_G(t) \leq -\delta$  for all  $t \in [\epsilon, q]$ .

Now for  $l \in \mathbb{N}$ , define  $t_l \equiv lq\delta/2$ , and let  $L_\delta = \{l : t_l \in [\epsilon, q]\}$ . Let  $T_\delta = \{q\} \cap \{t_l : l \in L_\delta\}$ .  $T_\delta$  is a finite set. Let

$$A_m = \bigcap_{t \in T_\delta} \{\hat{G}_m(t) - t/q < -\delta/2\}.$$

We will first show that  $A_m \subseteq \{\tau_{\text{BH},m} \leq \epsilon\}$ , and then we will use this to show that  $\lim_{m \rightarrow \infty} \Pr(\tau_{\text{BH},m} \leq \epsilon) = 1$ . To prove the former, suppose the event  $A_m$  occurs and take any  $t \in [\epsilon, 1]$ . If  $t > q$ , then  $\hat{G}_m(t) - t/q < 0$ , since  $\hat{G}_m(t) \leq 1$ . If instead  $t \in [\epsilon, q]$ , there exists a  $t_\uparrow \in T_\delta$  such that  $t_\uparrow - t \in [0, q\delta/2]$ . Therefore, since  $\hat{G}_m$  is non-decreasing,

$$\hat{G}_m(t) - \frac{t}{q} \leq \hat{G}_m(t_\uparrow) - \frac{t_\uparrow}{q} + \frac{t_\uparrow - t}{q} < -\frac{\delta}{2} + \frac{\delta}{2} < 0,$$

where the second to last inequality holds under the event  $A_m$ . Thus under the event  $A_m$ ,  $\hat{G}_m(t) - t/q < 0$  for all  $t \in [\epsilon, 1]$ , implying that  $A_m \subseteq \{\tau_{\text{BH},m} \leq \epsilon\}$ .

To show that  $\lim_{m \rightarrow \infty} \Pr(\tau_{\text{BH},m} \leq \epsilon) = 1$ , note that the proof of Proposition B.1 holds when Conditions 1, 2, 3, 4, and 6 hold, implying that  $(W_{m,0}, W_{m,1}) \xrightarrow{\text{f.d.d.}} (\tilde{W}_0, \tilde{W}_1)$  where  $(\tilde{W}_0, \tilde{W}_1)$  is a joint Gaussian process on  $[0, 1]$  with mean zero and joint covariance kernel  $c$ . Since Condition 8 holds for  $[a, b] = [0, 1]$ , by Slutsky's lemma,  $(\hat{W}_{m,0}, \hat{W}_{m,1}) \xrightarrow{\text{f.d.d.}} (\tilde{W}_0, \tilde{W}_1)$ . By the continuous mapping theorem,  $\hat{W}_{m,0} + \hat{W}_{m,1} \xrightarrow{\text{f.d.d.}} \tilde{W}_0 + \tilde{W}_1$  which implies that  $(\hat{W}_{m,0}(t) + \hat{W}_{m,1}(t))_{t \in T_\delta}$  converges in distribution to a multivariate Gaussian. Since  $\hat{W}_{m,0}(t) + \hat{W}_{m,1}(t) = \sqrt{m}(\hat{G}_m(t) - G(t))$  and since  $T_\delta$  is a finite set, if we let  $B_m$  be the event that  $|\hat{G}_m(t) - G(t)| < \delta/2$  for all  $t \in T_\delta$ , it is clear that  $\lim_{m \rightarrow \infty} \Pr(B_m) = 1$ . Now note that if  $B_m$  occurs  $A_m$  occur, since  $G(t) - t/q \leq -\delta$  for all  $t \in [\epsilon, q]$ . Because  $B_m \subseteq A_m \subseteq \{\tau_{\text{BH},m} \leq \epsilon\}$  and  $\lim_{m \rightarrow \infty} \Pr(B_m) = 1$ ,  $\lim_{m \rightarrow \infty} \Pr(\tau_{\text{BH},m} \leq \epsilon) = 1$ . The above argument holds for any  $\epsilon > 0$ , so by nonnegativity of  $\tau_{\text{BH},m}$ ,  $\tau_{\text{BH},m} \xrightarrow{p} 0$ .

To show that  $V_m/m \xrightarrow{p} 0$ , fix  $\epsilon > 0$ . Note that if  $\tau_{\text{BH},m} \leq q\epsilon/2$ , then

$$\frac{V_m}{m} = \hat{F}_{m,0}(\tau_{\text{BH},m}) \leq \hat{G}_m(\tau_{\text{BH},m}) \leq \hat{G}_m(q\epsilon) < \frac{q\epsilon}{q} = \epsilon$$

where the last inequality holds because, if it did not hold, there would be a contradiction to  $\tau_{\text{BH},m} \leq q\epsilon/2$ . Therefore,  $\tau_{\text{BH},m} \leq q\epsilon/2 \Rightarrow V_m/m \leq \epsilon$ , and so by the result of the previous paragraph,  $\lim_{m \rightarrow \infty} \Pr(V_m/m \leq \epsilon) = 1$ . By nonnegativity of  $V_m/m$  it follows that  $V_m/m \xrightarrow{p} 0$ .  $\square$

### Appendix C: Hadamard derivative calculations for proofs of Theorems 1 and 2

Recall that our goal is to compute the Hadamard derivative of both  $\Psi^{(\text{FPR})}$  and  $\Psi^{(\text{FDP})}$  at the point  $(F_0, F_1) \in C[a, b]_2$  tangentially to  $C[a, b]_2$ . We start by computing the Hadamard derivative of  $\mathcal{T}$  (defined in Equation (34)) at  $G$  tangentially to  $C[a, b]$ , but to do this, it first helps to prove the following modified version of Lemma 7.7 and Proposition 7.8 in Neuvial (2008).

**Lemma C.1.** *For any  $F \in D[a, b]$  such that  $\mathcal{T}(F) \in (a, b)$ , one of the following must hold:*

- (i)  $F(\mathcal{T}(F)) = \mathcal{T}(F)/q$ , or
- (ii)  $F(\mathcal{T}(F)) \leq \mathcal{T}(F)/q \leq \lim_{t \uparrow \mathcal{T}(F)} F(t)$ ,

where case (i) will hold whenever  $F$  is monotone nondecreasing.

*Proof.* Fix  $F \in D[a, b]$  such that  $\mathcal{T}(F) \in (a, b)$ . Assume by way of contradiction that  $F(\mathcal{T}(F)) > \mathcal{T}(F)/q$ . By right continuity of  $t \mapsto F(t) - t/q$ , there exists a sequence of  $t_n \downarrow \mathcal{T}(F)$  (where  $t_n \in (a, b)$  for all  $n$ ) such that  $\lim_{n \rightarrow \infty} F(t_n) - t_n/q = F(\mathcal{T}(F)) - \mathcal{T}(F)/q > 0$ , implying that for some  $n \in \mathbb{N}$ ,  $F(t_n) - t_n/q > 0$ . But that implies that for some  $t_n \in (a, b)$  and  $t_n > \mathcal{T}(F)$ ,  $F(t_n) \geq t_n/q$ , contradicting the definition of  $\mathcal{T}(F)$ . Hence it follows that  $F(\mathcal{T}(F)) \leq \mathcal{T}(F)/q$  always. If  $F(\mathcal{T}(F)) = \mathcal{T}(F)/q$ , then we are in case (i), and otherwise,  $F(\mathcal{T}(F)) < \mathcal{T}(F)/q$ . Now if  $F(\mathcal{T}(F)) < \mathcal{T}(F)/q$ , by definition of  $\mathcal{T}(F)$ , there exists a sequence of  $t_n \uparrow \mathcal{T}(F)$  such that  $F(t_n) - t_n/q \geq 0$ . So taking the limit as  $n \rightarrow \infty$  of each side, since  $F$  always has existing left limits,  $\lim_{t \uparrow \mathcal{T}(F)} F(t) - \mathcal{T}(F)/q \geq 0$ . Hence whenever  $F(\mathcal{T}(F)) < \mathcal{T}(F)/q$ ,  $\mathcal{T}(F)/q \leq \lim_{t \uparrow \mathcal{T}(F)} F(t)$ , so case (ii) will hold. When  $F$  is monotone nondecreasing case (i) holds because the RHS in case (ii) is bounded above by  $F(\mathcal{T}(F))$  for monotone nondecreasing  $F$ .  $\square$

Using this Lemma and an approach similar to Neuvial (2008), we can calculate the Hadamard derivative of  $\mathcal{T}$  at  $G$  tangentially to  $C[a, b]$ . The proposition below involves the following constant which will be convenient to define for the remainder of the appendix:

$$c_G \equiv 1/q - G'(\tau_*) > 0 \tag{47}$$

To see why this constant is positive, note that by Proposition 2 and our definitions of  $\tau_*$  and the interval  $[a, b]$ ,  $\tau_*$  is the unique point in  $[a, b]$  where the function  $t \mapsto G(t) - t/q$  crosses from a positive to negative value (and does not merely touch zero).

**Proposition C.1.**  *$\mathcal{T}$  is Hadamard differentiable at  $G$ , tangentially to  $C[a, b]$  with Hadamard derivative at  $G$  given by  $\dot{\mathcal{T}}_G(H) = H(\tau_*)/c_G$  for  $H \in C[a, b]$ .*

*Proof.* To prove that the Hadamard derivative of  $\mathcal{T}$  at  $G$  tangentially to  $C[a, b]$  is given by  $\dot{\mathcal{T}}_G(H) = H(\tau_*)/c_G$ , fix any  $H \in C[a, b]$  and fix any collection  $(H_t)_{t>0}$

where  $H_t \in D[a, b]$  for all  $t$  and  $\lim_{t \downarrow 0} \|H_t - H\|_\infty = 0$ . We must show that

$$\lim_{t \downarrow 0} \left| \frac{\mathcal{T}(G + tH_t) - \mathcal{T}(G)}{t} - \frac{H(\tau_*)}{c_G} \right| = 0. \quad (48)$$

To do this, define  $G_t \equiv G + tH_t$  and  $\tau_t \equiv \mathcal{T}(G_t)$  for  $t > 0$ . Also define for any  $F \in D[a, b]$ ,  $\psi_F : [a, b] \rightarrow \mathbb{R}$  to be given by  $\psi_F(u) = u/q - F(u)$ . Note  $\mathcal{T}(F) = \sup\{u \in [a, b], \psi_F(u) \leq 0\}$ . We must first show that there exists a  $\delta > 0$  such that for all  $t \in (0, \delta)$ ,  $\tau_t \in (a, b)$ . To see this, note that by our choice of interval  $(a, b)$  (see Section 2.3.4, and more specifically, see Proposition 2),  $G(a) > a/q \Rightarrow \psi_G(a) < 0$ ,  $G(b) < b/q \Rightarrow \psi_G(b) > 0$ , and  $\psi_G(\cdot)$  crosses 0 once on the interval  $[a, b]$ .  $\psi_G$  is continuous by continuity of  $G$  (see Proposition 1), so there exists a  $b_- < b$  such that  $\psi_G(u) > \frac{1}{2}\psi_G(b)$  for all  $u \in [b_-, b]$ . Also note that

$$\|G - G_t\|_\infty = \|tH_t\|_\infty \leq t\|H\|_\infty + t\|H - H_t\|_\infty.$$

Since  $H \in C[a, b]$ ,  $\|H\|_\infty < \infty$ , so  $\lim_{t \downarrow 0} t\|H\|_\infty = 0$ . Hence, taking the limit as  $t \downarrow 0$  of the above result we get that  $\lim_{t \downarrow 0} \|G - G_t\|_\infty = 0$ . Thus, we can pick  $\delta > 0$  such that for all  $t \in (0, \delta)$ ,  $\|G - G_t\|_\infty < \min\{|\psi_G(a)|, |\psi_G(b)|/2\}$ . Since for all  $u \in [a, b]$ ,  $|\psi_{G_t}(u) - \psi_G(u)| \leq \|G - G_t\|_\infty$ , we get that for all  $t \in (0, \delta)$  and  $u \in [b_-, b]$ ,  $\psi_{G_t}(a) < \psi_G(a) + |\psi_G(a)| = 0$  and  $\psi_{G_t}(u) > \psi_G(u) - \|G - G_t\|_\infty > \frac{1}{2}\psi_G(b) > 0$ . Thus, for all  $t \in (0, \delta)$  and  $u \in [b_-, b]$ , both  $\psi_{G_t}(a) < 0$  and  $\psi_{G_t}(u) > 0$ , where  $\psi_{G_t}$  is right continuous with left limits. This implies that  $\tau_t = \mathcal{T}(G_t) \in (a, b)$  for all  $t \in (0, \delta)$ .

For any  $t \in (0, \delta)$ , we will next apply Lemma C.1 to  $G_t$  to show that  $|\psi_G(\tau_t)| \leq \|G - G_t\|_\infty$ . If case (ii) of Lemma C.1 holds for  $G_t$ , then

$$-\lim_{u \uparrow \mathcal{T}(G_t)} G_t(u) \leq -\frac{\mathcal{T}(G_t)}{q} \leq -G_t(\mathcal{T}(G_t)),$$

so adding  $G(\tau_t) = G(\mathcal{T}(G_t))$  to each side of the equality and by continuity of  $G$ , we get

$$\lim_{u \uparrow \mathcal{T}(G_t)} ((G - G_t)(u)) \leq G(\tau_t) - \frac{\tau_t}{q} \leq (G - G_t)(\tau_t),$$

which implies that

$$|\psi_G(\tau_t)| = \left| G(\tau_t) - \frac{\tau_t}{q} \right| \leq \|G - G_t\|_\infty.$$

Alternatively, if case (i) from Lemma C.1 holds for  $G_t$ , then  $-\tau_t/q = -G_t(\tau_t)$  implying  $G(\tau_t) - \tau_t/q = (G - G_t)(\tau_t)$ , further implying  $|\psi_G(\tau_t)| \leq \|G - G_t\|_\infty$ . Thus, in either case,  $|\psi_G(\tau_t)| \leq \|G - G_t\|_\infty$  for all  $t \in (0, \delta)$ . Because  $\lim_{t \downarrow 0} \|G - G_t\|_\infty = 0$ ,  $\lim_{t \downarrow 0} \psi_G(\tau_t) = 0 = \psi_G(\tau_*)$ . Now since  $\psi_G$  is continuous (because  $G$  is continuous) it follows that  $\psi_G(\limsup_{t \downarrow 0} \tau_t) = \psi_G(\tau_*) = 0$  and  $\psi_G(\liminf_{t \downarrow 0} \tau_t) = \psi_G(\tau_*) = 0$ . Because  $\tau_*$  is the unique and existing  $u \in [a, b]$  such that  $\psi_G(u) = 0$  (see the definition of  $[a, b]$  and Proposition 2), it follows that  $\liminf_{t \downarrow 0} \tau_t = \tau_* = \limsup_{t \downarrow 0} \tau_t$ . Hence,  $\lim_{t \downarrow 0} \tau_t = \tau_*$ .

We would also like to prove that  $\lim_{t \downarrow 0} \psi_G(\tau_t)/t = H(\tau_*)$ . To show this, we start by showing that for  $t \in (0, \delta)$ ,

$$\left| \frac{\psi_G(\tau_t)}{t} - H(\tau_t) \right| \leq \max \left\{ \left| H_t(\tau_t) - H(\tau_t) \right|, \left| \lim_{u \uparrow \tau_t} (H_t(u) - H(u)) \right| \right\}. \quad (49)$$

Let  $T_{(i)}$  be the set of  $t \in (0, \delta)$  such that  $G_t$  is in case (i) of Lemma C.1 and let  $T_{(ii)}$  be the set of  $t \in (0, \delta)$  such that  $G_t$  is in case (ii) of Lemma C.1. Note that if  $t \in T_{(i)}$ , then  $\psi_G(\tau_t) = \tau_t/q - G(\tau_t) = (G_t - G)(\tau_t) = tH_t(\tau_t)$  which implies that  $\psi_G(\tau_t)/t - H(\tau_t) = H_t(\tau_t) - H(\tau_t)$  and hence (49) holds for  $t \in T_{(i)}$ . If instead  $t \in T_{(ii)}$ , then

$$G_t(\tau_t) \leq \frac{\tau_t}{q} \leq \lim_{u \uparrow \tau_t} G_t(u),$$

so subtracting  $G(\tau_t)$  from each side and using the fact that  $G$  is continuous, we get that

$$(G_t - G)(\tau_t) \leq \psi_G(\tau_t) \leq \lim_{u \uparrow \tau_t} (G_t - G)(u).$$

Recalling that  $G_t - G = tH_t$ , we can divide each side of the above inequality by  $t$  and subtract  $H(\tau_t)$  to get

$$H_t(\tau_t) - H(\tau_t) \leq \frac{\psi_G(\tau_t)}{t} - H(\tau_t) \leq \lim_{u \uparrow \tau_t} H_t(u) - H(\tau_t),$$

where the RHS can be rewritten as  $\lim_{u \uparrow \tau_t} (H_t(u) - H(u))$  by continuity of  $H$ . Thus, (49) holds for any  $t \in T_{(ii)}$ . Since  $T_{(i)} \cup T_{(ii)} = (0, \delta)$  it follows that (49) holds for all  $t \in (0, \delta)$ .

Now observe that both  $|H_t(\tau_t) - H(\tau_t)| \leq \|H_t - H\|_\infty$  and  $|\lim_{u \uparrow \tau_t} (H_t(u) - H(u))| < \|H_t - H\|_\infty$  hold. Therefore, (49) implies that for any  $t \in (0, \delta)$ ,  $|\psi_G(\tau_t)/t - H(\tau_t)| \leq \|H_t - H\|_\infty$ , and so

$$\begin{aligned} \left| \frac{\psi_G(\tau_t)}{t} - H(\tau_*) \right| &\leq \left| \frac{\psi_G(\tau_t)}{t} - H(\tau_t) \right| + |H(\tau_t) - H(\tau_*)| \\ &\leq \|H_t - H\|_\infty + |H(\tau_t) - H(\tau_*)|. \end{aligned}$$

Taking the limit as  $t \downarrow 0$  of each side above completes the proof that  $\lim_{t \downarrow 0} \psi_G(\tau_t)/t = H(\tau_*)$ , where the fact that  $\lim_{t \downarrow 0} |H(\tau_t) - H(\tau_*)| = 0$  holds by the composite limit theorem and our earlier result that  $\lim_{t \downarrow 0} \tau_t = \tau_*$ .

To complete the proof of (48), note that  $c_G = 1/q - G'(\tau_*) = \psi'_G(\tau_*)$ , and hence by the alternative definition of a limit

$$c_G = \psi'_G(\tau_*) = \lim_{u \rightarrow \tau_*} \frac{\psi_G(u) - \psi_G(\tau_*)}{u - \tau_*} = \lim_{u \rightarrow \tau_*} \frac{\psi_G(u)}{u - \tau_*} = \lim_{t \downarrow 0} \frac{\psi_G(\tau_t)}{\tau_t - \tau_*},$$

where the last equality holds by the composite limit theorem because  $\lim_{t \downarrow 0} \tau_t = \tau_*$ . Therefore, for  $t > 0$ ,

$$\frac{\mathcal{T}(G + tH_t) - \mathcal{T}(G)}{t} = \frac{\tau_t - \tau_*}{t} = \frac{\psi_G(\tau_t)}{t} \left( \frac{\psi_G(\tau_t)}{\tau_t - \tau_*} \right)^{-1}$$

and hence,

$$\lim_{t \downarrow 0} \frac{\mathcal{T}(G + tH_t) - \mathcal{T}(G)}{t} = \left( \lim_{t \downarrow 0} \frac{\psi_G(\tau_t)}{t} \right) \left( \lim_{t \downarrow 0} \frac{\psi_G(\tau_t)}{\tau_t - \tau_*} \right)^{-1} = \frac{H(\tau_*)}{c_G}$$

proving (48). The above proof of (48) holds for any fixed  $H \in C[a, b]$  and fixed collection  $(H_t)_{t>0}$  of  $H_t \in D[a, b]$  such that  $\lim_{t \downarrow 0} \|H_t - H\|_\infty = 0$ . Thus, the Hadamard derivative of  $\mathcal{T}$  at  $G$  tangentially to  $C[a, b]$  is given by  $\dot{\mathcal{T}}_G(H) = H(\tau_*)/c_G$ .  $\square$

Using Proposition C.1 and the chain rule for Hadamard differentiability we can compute the Hadamard derivatives of  $\Psi^{(\text{FPR})}$ ,  $\mathcal{R}$ , and  $\Psi^{(\text{FDP})}$  at  $(F_0, F_1)$  tangentially to  $C[a, b]_2$ . We start by computing the Hadamard derivative of  $\Psi^{(\text{FPR})}$ .

**Proposition C.2.**  $\Psi^{(\text{FPR})}$  (defined in Equation (36)) is Hadamard differentiable at  $(F_0, F_1)$  tangentially to  $C[a, b]_2$  with

$$\dot{\Psi}_{(F_0, F_1)}^{(\text{FPR})}(H_0, H_1) = H_0(\tau_*) + F_0'(\tau_*)(H_0(\tau_*) + H_1(\tau_*))/c_G$$

for  $(H_0, H_1) \in C[a, b]_2$  being the Hadamard derivative at  $(F_0, F_1)$ .

*Proof.* To compute the Hadamard derivative of  $\Psi^{(\text{FPR})}$  at  $(F_0, F_1)$  tangentially to  $C[a, b]_2$  fix  $(H_0, H_1) \in C[a, b]_2$  and a collection  $((H_{0t}, H_{1t}))_{t>0}$  of elements of  $D[a, b]_2$  and such that  $\lim_{t \downarrow 0} \|(H_{0t}, H_{1t}) - (H_0, H_1)\|_\infty = 0$ . Let  $F_{0t} \equiv F_0 + tH_{0t}$ ,  $F_{1t} \equiv F_1 + tH_{1t}$ , and  $G_t \equiv F_{0t} + F_{1t} = (F_0 + F_1) + t(H_{0t} + H_{1t})$  where  $H_{0t} + H_{1t} \rightarrow H_0 + H_1$  as  $t \downarrow 0$ . Letting  $\tau_t \equiv \mathcal{T}(G_t)$  and recalling  $\tau_* = \mathcal{T}(G)$ , note that

$$\frac{\Psi^{(\text{FPR})}(F_{0t}, F_{1t}) - \Psi^{(\text{FPR})}(F_0, F_1)}{t} = H_{0t}(\tau_t) + \frac{F_0(\tau_t) - F_0(\tau_*)}{t}. \quad (50)$$

Since  $\mathcal{T}$  is Hadamard differentiable at  $G$  tangentially to  $C[a, b]$  (see Proposition C.1) and  $F_0$  is differentiable at  $\mathcal{T}(G)$ , by the chain rule for Hadamard differentiability (see, for example, Theorem 20.9 in van der Vaart (1998)), it follows that  $F_0 \circ \mathcal{T} : D[a, b] \rightarrow [0, 1]$  is Hadamard differentiable at  $G$  tangentially to  $C[a, b]$  with Hadamard derivative given by  $F_0'(\mathcal{T}(G))\dot{\mathcal{T}}_G(H)$  for  $H \in C[a, b]$ . By the definition of Hadamard differentiability and since  $G_t = (F_0 + F_1) + t(H_{0t} + H_{1t})$  where  $H_{0t} + H_{1t} \rightarrow H_0 + H_1$  as  $t \downarrow 0$ , it follows that

$$\lim_{t \downarrow 0} \frac{F_0(\tau_t) - F_0(\tau_*)}{t} = \lim_{t \downarrow 0} \frac{F_0(\mathcal{T}(G_t)) - F_0(\mathcal{T}(G))}{t} = F_0'(\mathcal{T}(G))\dot{\mathcal{T}}_G(H_0 + H_1).$$

By combining the above result with the formula from Proposition C.1,

$$\lim_{t \downarrow 0} \frac{F_0(\tau_t) - F_0(\tau_*)}{t} = \frac{F_0'(\tau_*)(H_0(\tau_*) + H_1(\tau_*))}{c_G}. \quad (51)$$

Also note that

$$\begin{aligned} |H_{0t}(\tau_t) - H_0(\tau_*)| &= |H_{0t}(\tau_t) - H_0(\tau_t)| + |H_0(\tau_t) - H_0(\tau_*)| \\ &\leq \|H_{0t} - H_0\|_\infty + |H_0(\tau_t) - H_0(\tau_*)|. \end{aligned}$$

We know that  $\lim_{t \downarrow 0} \|H_{0t} - H_0\|_\infty = 0$ . Further since  $H_0$  is continuous and we showed in the proof of Proposition C.1 that  $\lim_{t \downarrow 0} \tau_t = \tau_*$  (in a setting where  $\tau_t = \mathcal{T}(G_t)$  and  $G_t = G + tH_t \in D[a, b]$  for some  $H_t \rightarrow H \in C[a, b]$  as  $t \downarrow 0$ ) we get by the composite limit theorem that  $\lim_{t \downarrow 0} |H_0(\tau_t) - H_0(\tau_*)| = 0$ . Hence by an above inequality it follows that  $\lim_{t \downarrow 0} H_{0t}(\tau_t) = H_0(\tau_*)$ .

Combining this with (50) and (51), we get that

$$\lim_{t \downarrow 0} \frac{\Psi^{(\text{FPR})}(F_{0t}, F_{1t}) - \Psi^{(\text{FPR})}(F_0, F_1)}{t} = H_0(\tau_*) + \frac{F'_0(\tau_*)(H_0(\tau_*) + H_1(\tau_*))}{c_G}.$$

Since this argument holds for any fixed  $(H_0, H_1) \in C[a, b]_2$  and collection  $((H_{0t}, H_{1t}))_{t > 0}$  of elements of  $D[a, b]_2$  such that  $\lim_{t \downarrow 0} \|(H_{0t}, H_{1t}) - (H_0, H_1)\|_\infty = 0$  where  $F_{0t} \equiv F_0 + tH_{0t}$  and  $F_{1t} \equiv F_1 + tH_{1t}$ , it follows that  $\Psi^{(\text{FPR})}$  is Hadamard differentiable at  $(F_0, F_1)$  tangentially to  $C[a, b]_2$  with Hadamard derivative given by

$$\dot{\Psi}_{(F_0, F_1)}^{(\text{FPR})}(H_0, H_1) = H_0(\tau_*) + F'_0(\tau_*)(H_0(\tau_*) + H_1(\tau_*))/c_G. \quad \square$$

**Proposition C.3.**  $\mathcal{R}$  (defined in Equation (35)) is Hadamard differentiable at  $(F_0, F_1)$  tangentially to  $C[a, b]_2$  with

$$\dot{\mathcal{R}}_{(F_0, F_1)}(H_0, H_1) = (H_0(\tau_*) + H_1(\tau_*))/(qc_G)$$

for  $(H_0, H_1) \in C[a, b]_2$  being the Hadamard derivative at  $(F_0, F_1)$ .

*Proof.* Define  $\Psi^{(\text{A})} : D[a, b]_2 \rightarrow \mathbb{R}$  via  $\Psi^{(\text{A})}(H_0, H_1) = H_1(\mathcal{T}(H_0 + H_1))$  for any  $(H_0, H_1) \in D[a, b]_2$ . Swapping the indices of the two  $D[a, b]$  functions in the input of  $\Psi^{(\text{A})}$  gives exactly the same function as  $\Psi^{(\text{FPR})}$ . Therefore, by the argument in Proposition C.2,  $\Psi^{(\text{A})}$  is Hadamard differentiable at  $(F_0, F_1)$  tangentially to  $C[a, b]_2$  with Hadamard derivative given by  $\dot{\Psi}_{(F_0, F_1)}^{(\text{A})}(H_0, H_1) = H_1(\tau_*) + F'_1(\tau_*)(H_0(\tau_*) + H_1(\tau_*))/c_G$ . Because  $\mathcal{R} = \Psi^{(\text{FPR})} + \Psi^{(\text{A})}$  we find that  $\mathcal{R}$  is Hadamard differentiable at  $(F_0, F_1)$  tangentially to  $C[a, b]_2$  with Hadamard derivative  $\dot{\mathcal{R}}_{(F_0, F_1)} = \dot{\Psi}_{(F_0, F_1)}^{(\text{FPR})} + \dot{\Psi}_{(F_0, F_1)}^{(\text{A})}$ . Combining our formula for  $\dot{\Psi}_{(F_0, F_1)}^{(\text{A})}$  with Proposition C.2 we get that for any  $(H_0, H_1) \in C[a, b]_2$ ,

$$\begin{aligned} \dot{\mathcal{R}}_{(F_0, F_1)}(H_0, H_1) &= (H_0(\tau_*) + H_1(\tau_*)) \left( 1 + \frac{F'_0(\tau_*) + F'_1(\tau_*)}{c_G} \right) \\ &= \frac{H_0(\tau_*) + H_1(\tau_*)}{qc_G}. \end{aligned} \quad \square$$

**Proposition C.4.**  $\Psi^{(\text{FDP})}$  (defined in Equation (37)) is Hadamard differentiable at  $(F_0, F_1)$  tangentially to  $C[a, b]_2$  with

$$\dot{\Psi}_{(F_0, F_1)}^{(\text{FDP})}(H_0, H_1) = q((1 + \alpha)H_0(\tau_*) + \alpha H_1(\tau_*))/\tau_*$$

for  $(H_0, H_1) \in C[a, b]_2$  being the formula for the Hadamard derivative at  $(F_0, F_1)$ , and where  $\alpha$  is defined in (13).

*Proof.* Define  $\Theta : D[a, b]_2 \rightarrow \mathbb{R}^2$  as

$$\Theta(H_0, H_1) = (\Psi^{(\text{FPR})}(H_0, H_1), \mathcal{R}(H_0, H_1))$$

for  $(H_0, H_1) \in D[a, b]_2$ . It is easy to see that the Hadamard derivative of  $\Theta$  at  $(F_0, F_1)$  tangentially to  $C[a, b]_2$  is given by  $\dot{\Theta}_{(F_0, F_1)}(H_0, H_1) = (\dot{\Psi}_{(F_0, F_1)}^{(\text{FPR})}(H_0, H_1), \dot{\mathcal{R}}_{(F_0, F_1)}(H_0, H_1))$  and by Propositions C.2 and C.3,  $\Theta$  is indeed Hadamard differentiable at  $(F_0, F_1)$  tangentially to  $C[a, b]_2$ . Now let  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  be given by  $f(x, y) = x/y$ . Note that  $\nabla f_{(x, y)} = (1/y, -x/y^2)$  for  $y \neq 0$ . Note that for any  $(H_0, H_1) \in D[a, b]_2$ ,  $\Psi^{(\text{FDP})}(H_0, H_1) = f(\Theta(H_0, H_1))$ . Hence by the chain rule for Hadamard differentiability (Theorem 20.9 in van der Vaart) we get that if  $\mathcal{R}(F_0, F_1) \neq 0$ , then  $\Psi^{(\text{FDP})}$  is Hadamard differentiable at  $(F_0, F_1)$  tangentially to  $C[a, b]_2$  with Hadamard derivative given by  $(\nabla f_{\Theta(F_0, F_1)})^\top \dot{\Theta}_{(F_0, F_1)}(H_0, H_1)$  for  $(H_0, H_1) \in C[a, b]_2$ . But since  $\mathcal{R}(F_0, F_1) = (F_0 + F_1)(\mathcal{T}(F_0 + F_1)) = G(\tau_*) = \tau_*/q \neq 0$ ,  $\Psi^{(\text{FDP})}$  is indeed Hadamard differentiable at  $(F_0, F_1)$  tangentially to  $C[a, b]_2$  with Hadamard derivative given by  $(\nabla f_{\Theta(F_0, F_1)})^\top \dot{\Theta}_{(F_0, F_1)}(H_0, H_1)$ . We compute this value more specifically for  $(H_0, H_1) \in C[a, b]_2$  using Propositions C.2 and C.3:

$$\begin{aligned} \dot{\Psi}_{(F_0, F_1)}^{(\text{FDP})}(H_0, H_1) &= \frac{\dot{\Psi}_{(F_0, F_1)}^{(\text{FPR})}(H_0, H_1)}{\mathcal{R}(F_0, F_1)} - \frac{\Psi^{(\text{FPR})}(F_0, F_1) \dot{\mathcal{R}}_{(F_0, F_1)}(H_0, H_1)}{(\mathcal{R}(F_0, F_1))^2} \\ &= \frac{q}{\tau_*} \left( \dot{\Psi}_{(F_0, F_1)}^{(\text{FPR})}(H_0, H_1) - \frac{q}{\tau_*} F_0(\tau_*) \dot{\mathcal{R}}_{(F_0, F_1)}(H_0, H_1) \right) \\ &= \frac{q}{\tau_*} \left( H_0(\tau_*) + \frac{(F_0'(\tau_*) - F_0(\tau_*)/\tau_*)(H_0(\tau_*) + H_1(\tau_*))}{c_G} \right) \\ &= \frac{q}{\tau_*} ((1 + \alpha)H_0(\tau_*) + \alpha H_1(\tau_*)). \end{aligned}$$

The last step above uses our previous definition of  $c_G$  in (47) and the definition of  $\alpha$  in (13).  $\square$