

Università degli Studi del Molise
Dipartimento di Scienze Economiche, Gestionali e Sociali (SEGeS)



ECONOMICS & STATISTICS DISCUSSION PAPER
No. 059/11

FDR Control in the Presence of an Unknown Correlation Structure

Roy Cerqueti

Mauro Costantini

Claudio Lupi

FDR Control in the Presence of an Unknown Correlation Structure

Roy Cerqueti
University of Macerata

Mauro Costantini
University of Vienna

Claudio Lupi
University of Molise

Abstract

The false discovery rate (FDR, [Benjamini and Hochberg 1995](#)) is a powerful approach to multiple testing. However, the original approach developed by [Benjamini and Hochberg \(1995\)](#) applies only to independent tests. [Yekutieli \(2008\)](#) showed that a modification of the Benjamini-Hochberg (BH) approach can be used in the presence of dependent tests and labelled his procedure as *separate subsets* BH (ssBH). However, [Yekutieli \(2008\)](#) left the practical specification of the subsets of p values largely unresolved. In this paper we propose a modification of the ssBH procedure based on a selection of the subsets that guarantees that the dependence properties needed to control the FDR are satisfied. We label this new procedure as the *separate pairs* BH (spBH). An extensive Monte Carlo analysis is presented that compares the properties of the BH and spBH procedures.

Keywords: Multiple testing, False discovery rate, Copulas.

JEL codes: C10, C12.

1. Introduction

When many hypotheses are tested simultaneously, the risk of falsely rejecting truly null hypotheses increases dramatically. In one single test we usually reject the null if the test p value, p , is such that $p < \alpha$, for a pre-specified level α . Since $p \sim U_{(0,1)}$ under the null, we have that $\Pr(p < \alpha | H_0) = \alpha$. But when $m \gg 1$ hypotheses are tested simultaneously it is likely that *at least* one of the p values is less than α even if all the hypotheses are truly null. In general it is possible to derive that

$$\begin{aligned} \Pr(\text{at least one of } \{p_1, \dots, p_m\} < \alpha | H_{01}, \dots, H_{0m}) &\leq \sum_{i=1}^m \Pr(p_i < \alpha | H_{0i}) \\ &= m\alpha \end{aligned} \tag{1}$$

from which we can set the decision rule of rejecting H_{0i} iff $p_i < \alpha/m$, which is the base of the so-called “Bonferroni procedure”. In this way we can control the *Familywise Error Rate* (FWER) in such a way that

$$\Pr\left(\text{at least one of } \{p_1, \dots, p_m\} < \frac{\alpha}{m} \mid H_{01}, \dots, H_{0m}\right) \leq \alpha. \tag{2}$$

It is well known that procedures that control the FWER tend to have low power, with the power decreasing as the number of tested hypotheses increases. However, depending on the

focus of the analysis, it might well be that strict control of the FWER is not even needed. In practice, we need such a strict control only if we want to have $\Pr(\# \text{ of false positives} \geq 1) \leq \alpha$, but we might think that the harm of falsely rejecting the null should be inversely related to the number of hypotheses that are tested. In fact, one would like to identify as many “discoveries” as possible (Sorić 1989), while incurring in a small proportion of false positives. This is the motivation of the concept of *False Discovery Rate* (FDR) introduced by Benjamini and Hochberg (1995). In plain words, the FDR is the expected value of the proportion of errors among the rejected hypotheses (FDP, the *False Discovery Proportion*). It can be shown (Benjamini and Hochberg 1995; Storey, Taylor, and Siegmund 2004) that when all the m hypotheses are truly null the FDR equals the FWER, which implies that controlling the FDR implies (weak) control of the FWER, with substantial gains in power.

More formally, denoting the number of “false discoveries” by F and the number of significant results by R , the FDR can be defined as

$$FDR := \mathbb{E} \left(\frac{F}{R \vee 1} \right) = \mathbb{E} \left(\frac{F}{R} \mid R > 0 \right) \Pr(R > 0). \quad (3)$$

Using $R \vee 1$ in the denominator of (3) sets the FDR to zero when there are no rejections.

Let us introduce a collection of m statistics \mathbf{t} as follows:

$$\mathbf{t} := (\hat{t}_1, \dots, \hat{t}_m). \quad (4)$$

In multiple testing, the False Discovery Rate (FDR) control is attained when the individual test statistics \hat{t} 's in \mathbf{t} are independent (Benjamini and Hochberg 1995), or exhibit a particular dependence structure (Benjamini and Yekutieli 2001). Yekutieli (2008) showed that a modification of the Benjamini-Hochberg (BH) approach can be used in the presence of dependent tests. The *separate subsets* BH (ssBH) procedure proposed by Yekutieli (2008) is based on the idea of using the BH approach not on all the p values simultaneously, but on separate subset of p values, each satisfying certain dependence properties. However, Yekutieli (2008) left the practical specification of the subsets of p values largely unresolved. Our paper fills a gap remained empty in the literature and provides a generalization of the results regarding the relationship between the control of the FDR and the dependence structure of the individual statistics involved. The approach we propose in this paper is a special version of Yekutieli's ssBH that we call the *separate pairs* BH procedure (spBH). We also show how the spBH procedure can be used in practice in the presence of correlated tests.

The rest of the paper is organized as follows: the next Section derives the main theoretical results of the paper and details an operating procedure to carry out multiple testing with FDR control. Section 3 offers a Monte Carlo analysis of the performance of the proposed procedure under different correlation of the test statistics. We also shed new light on the ability of the BH procedure to control the FDR in a rather general setting. The final Section concludes.

2. The statistical model and the main result

Benjamini and Yekutieli (2001) propose a procedure to control the FDR at level q for all joint test statistics, under a particular dependence hypothesis, consisting in the application of the Benjamini-Hochberg (BH) at level $q / \left(\sum_{i=1}^M i^{-1} \right)$. An even more general procedure,

the separate subsets BH (ssBH) procedure, has been introduced recently by Yekutieli (2008) in order to deal with more general forms of dependence. Here it is an explanation on how the ssBH works.

Define the vector \mathbf{p} of the m p values associated to the components of the set \mathbf{t} . Divide \mathbf{p} in S sub-vectors \mathbf{p}^s , for $s = 1, \dots, S$. With a very intuitive notation, the statistics corresponding to \mathbf{p}^s constitute a vector, that will be indicated with \mathbf{t}^s . Assume that the cardinality of \mathbf{p}^s is m^s and denote as \mathbf{p}_0^s the p values corresponding to the true null hypotheses in \mathbf{p}^s . The level q ssBH procedure runs into two steps as follows:

- i. For $s = 1, \dots, S$, apply the BH procedure at level qm^s/m to test \mathbf{p}^s , and denote as \mathbf{r}_{BH}^s the p values corresponding to the rejected hypotheses.
- ii. Reject the null hypothesis corresponding to $\mathbf{r}_{ssBH} = \bigcup_{s=1}^S \mathbf{r}_{BH}^s$.

This section aims at developing a formal argument to generalize the relationship between the FDR control and the dependence properties of the individual statistics involved by means of the ssBH procedure.

We first notice that the statistics \mathbf{t} defined in (4) are introduced to test the same null hypothesis across different units (experiments, time series). Hence, the individual statistics should follow the same distribution under the null. Therefore, it is not too restrictive to state the next assumption.

Assumption 2.1. *The individual statistics $\hat{t}_1, \dots, \hat{t}_m$ are pair-wise exchangeable, i.e.*

$$F_{i,j}(x, y) = F_{j,i}(y, x), \quad \forall x, y \in \mathbb{R}, \quad (5)$$

where $F_{i,j}$ is the bivariate cumulative joint distribution function of the couple (\hat{t}_i, \hat{t}_j) , with $i, j = 1, \dots, m$.

Assumption 2.1 guarantees that the testing procedure disregards on an eventual cross-sectional index permutation of the experiments.

We stress that the FDR control is strongly related to the stochastic dependence among the individual statistics belonging to \mathbf{t} . Therefore, a detailed discussion on the dependence structure underlying the \hat{t} 's is needed. In this respect and to be self-contained, we recall here the concept of *positive regression dependency on each one from a subset* $I_0 \subseteq \{1, \dots, n\}$ or, briefly, *PRDS on I_0* :

Definition 2.2. *Consider an increasing set¹ D . The vector \mathbf{t} is assumed to satisfy the PRDS on I_0 if, for each $i \in I_0$, the conditional probability $\Pr(\mathbf{t} \in D \mid \hat{t}_i = x)$ is nondecreasing in x . \square*

Benjamini and Yekutieli (2001) prove that the PRDS property on subset of the test statistics \hat{t} 's corresponding to the true null hypothesis assures the control of the FDR at a certain level by the Benjamini and Hochberg procedure. Unfortunately, this result meets severe drawbacks in practice, because of the difficulty in showing the PRDS property. To overcome this problem, the *multivariate total positivity of order 2* or, briefly, *MTP2* property — a stronger dependence structure — has been introduced. We recall here the definition of MTP2:

¹A set D is said to be increasing when, if $x \in D$ and $y \geq x$, then $y \in D$.

Definition 2.3. Let f be the joint density function of the n -variate random variable \mathbf{t} . \mathbf{t} is said to be MTP2 if and only if, for each \mathbf{x} and \mathbf{y} , it results:

$$f(\mathbf{x}) \cdot f(\mathbf{y}) \leq f(\min\{\mathbf{x}, \mathbf{y}\}) \cdot f(\max\{\mathbf{x}, \mathbf{y}\}) \quad (6)$$

where the min and max operators have to be intended componentwise. \square

A well-known statistical result assures that MTP2 \implies PRDS on $I_0 \forall I_0$. Therefore, the dependence described by the MTP2 can be used instead of the PRDS on I_0 , having in mind that the former condition is stronger.

A rather general way to capture the stochastic dependence structure among random variables is the introduction of the concept of *copula*. Since the development of our model will involve only pairs of statistics, we focus on the *bivariate copulas*, leaving a more general discussion on the multivariate case to Nelsen (1999).

Definition 2.4. The function $C : [0, 1]^2 \rightarrow [0, 1]$ is a copula if and only if:

- (i) $C(0, u) = C(v, 0) = 0$;
- (ii) $C(1, u) = C(u, 1) = u$;
- (iii) $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$, for each $0 \leq u_1 \leq u_2 \leq 1$ and $0 \leq v_1 \leq v_2 \leq 1$.

The classical Sklar's Theorem (Sklar 1959) highlights how the bivariate copula introduced in Definition 2.4 models the dependence structure between pairs of random variables. We report here the enunciation of this result adapted to our case, for sake of completeness.

Theorem 2.5 (Sklar, 1959). Let $F_{i,j}$ be the bidimensional joint distribution function of the couple (\hat{t}_i, \hat{t}_j) , with $i, j = 1, \dots, N$. Define the margins as F_i and F_j . Then there exists a bivariate copula $C_{i,j}$ such that, for each $x, y \in \mathbb{R}$,

$$F_{i,j}(x, y) = C_{i,j}(F_i(x), F_j(y)). \quad (7)$$

If the margins F_i and F_j are continuous, then the copula $C_{i,j}$ is unique. Conversely, if $C_{i,j}$ is a bivariate copula and F_i and F_j are distribution functions, then the function $F_{i,j}$ defined in (7) is a bidimensional distribution function with margins F_i and F_j .

Sklar's Theorem points out that, given a couple of random variables, the relationship between joint and marginal distributions is stated through copulas. As an illustrative example, it is worth noting that (7) implies that the product copula $C_{i,j}(u, v) = uv$ represents the case of independence between \hat{t}_i and \hat{t}_j .

Furthermore, the cross-sectional dependence between the individual \hat{t} 's of (4) can also be well represented by a non-diagonal variance/covariance matrix $(\sigma_{i,j})_{i,j=1,\dots,N}$, where the variance is indicated with a unique index as: $\sigma_{i,i} = \sigma_i^2$. Hence, it is natural to guess a derivation of the value of the covariances through copulas. This problem has a formal solution by means of the Sklar's Theorem, in virtue of which we immediately have:

Proposition 2.6.

$$\sigma_{i,j} = \frac{1}{\sigma_i \sigma_j} \int \int_{\mathbb{R}^2} [C_{i,j}(F_i(x), F_j(y)) - F_i(x)F_j(y)] dx dy,$$

where $C_{i,j}$ is the copula defined as in (7).

Proposition 2.6 shows that the introduction of copulas allows us to deal with a dependence structure simply described through a variance/covariance matrix.

The argument on the stochastic dependence developed above can be applied to the FDR control of the multiple statistics \mathbf{t} . In order to proceed, we need a condition on the sets \mathbf{p}^s introduced above:

Condition 2.7. *One of the following assumptions holds:*

- (i) if $p_i \in \mathbf{p}_0$, then there exists a unique $s_i \in \{1, \dots, S\}$ such that $p_i \in \mathbf{p}^{s_i}$. Moreover, for each $s = 1, \dots, S$, it must be:

$$m^s = \begin{cases} 2, & \text{if } \mathbf{p}_0^s \neq \emptyset; \\ \text{arbitrary,} & \text{otherwise.} \end{cases}$$

- (ii) $\mathbf{p}^{s_i} \cap \mathbf{p}^{s_j} = \emptyset$, for $s_i \neq s_j$, and $m^s = 2$, for each $s = 1, \dots, S$.

Condition 2.7 means that the division of the set \mathbf{p} in the subsets \mathbf{p}^s is such that each p value of a true null hypothesis is contained in one \mathbf{p}^s , and each \mathbf{p}^s containing a p value of a true null hypothesis has cardinality equals to 2. This is not a restrictive hypothesis, since the decomposition of $\{\mathbf{p}^s\}_{s=1, \dots, S}$ to be used for the ssBH procedure can be arbitrarily chosen. It is worth noting that when (ii) of Condition 2.7 holds, then $m^s = 2$, for each $s = 1, \dots, S$; if (i) is true, then $m^s = 2$, for each $s = 1, \dots, \tilde{S}$, with $\tilde{S} \leq S$. In the simulation analysis we will deal with Condition 2.7-(ii). Hence, in our version of the ssBH, all the separate subsets have cardinality equal to 2: for this reason we label the procedure we adopt as the separate pairs BH (spBH).

We are now able to state the main result of this section.

Proposition 2.8. *Assume that Condition 2.7 holds and that the dependence between the statistics in \mathbf{t}^s is described by a copula C_s such that:*

$$C_s(u, v) = uv + \theta \phi(u)\phi(v), \quad (8)$$

for each $s = 1, \dots, S$, with $\theta \in [-1, 1]$ and ϕ convex or concave.

Then the level q spBH procedure controls the FDR at level qm_0/m .

Proof. Denote as X and Y the individual statistics in \mathbf{t}^s . By Assumption 2.1, X and Y are exchangeable. Amblard and Girard (2002) showed that, if the dependence between X and Y is described through the copula C_s in (8), then Y is stochastically increasing in X and X is stochastically increasing in Y , i.e. the following conditions hold:

$$\begin{cases} P(Y > y | X = x) \text{ is nondecreasing in } x, \forall y; \\ P(X > x | Y = y) \text{ is nondecreasing in } y, \forall x. \end{cases} \quad (9)$$

The system (9) is equivalent to the TP2 property for the set \mathbf{t}^s (see Nelsen, 2006). Hence, Condition 2.7 and Proposition 2.2 in Yekutieli (2008) give the thesis. \square

Remark 2.9. *The pairwise dependence introduced in the set up given by family of $\{\mathbf{p}^s\}_{s=1,\dots,S}$, Condition 2.7 and the copulas in (8) allow us to describe a system with both positively and negatively correlated units. Indeed, the positive dependence condition formalized by TP2 is required only for some pairs of statistics in \mathbf{t} (the ones appearing in the \mathbf{t}^s 's), while no assumptions are stated on the remaining couples.*

Remark 2.10. *Copula defined in (8) allows to derive explicitly the correlation between the individual statistics in \mathbf{t}^s . Indeed, some algebra provides that if the stochastic dependence between X and Y is described through copula C_s in (8), then the correlation coefficient $\rho_{X,Y}$ between X and Y can be written as:*

$$\rho_{X,Y} = 12\theta \left(\int_0^1 \phi(\xi) d\xi \right)^2.$$

3. Monte Carlo analysis

In this Section we compare the performance of the BH with that of the spBH procedure, using an extensive simulation study. As far as the formation of the sub-vectors \mathbf{p}^s are concerned, in the simulations we refer to Condition 2.7-(ii), so that $m^s = 2$ for each $s = 1, \dots, S$. In particular, the s -th pair \mathbf{p}^s is formed by the s -th smallest and the s -th largest p values.

We sample m_0 p values under the null and m_1 p values under the one-sided alternative, with $m := m_0 + m_1$ being the number of hypotheses to be tested. We consider one-sided alternatives, because the one-sided p values are co-monotone transformations of the corresponding test statistics, from which they inherit their dependence properties. However, considering one-sided alternatives is not restrictive, given that two-sided tests can be built on two one-sided procedures. We want to allow the p values to come from positively correlated tests (both under the null and the alternative). The kind of experiments we consider are extensions of those carried out e.g. in Benjamini and Hochberg (1995), Brown and Russell (1997) and Verhoeven, Simonsen, and McIntyre (2005).

Dependent p values can be simulated by generating first the m -variate normal $\mathbf{y} \sim N(\mathbf{0}, \Sigma)$. In order to generate m_1 p values under the alternative, we shift the last m_1 components of the multivariate normal \mathbf{y} by a quantity such that the probability of obtaining a p value less than α (the significance level) is equal to π_α (the power at the given significance level α). Then, the m one-sided p values are generated as $p_i = 1 - \Phi(y_i)$ with $i = 1, \dots, m$.

The matrix Σ is also randomly drawn at each replication and is such that the average (off-diagonal) correlation is $\bar{\sigma}_{ij} := (\beta_1 + \beta_2)^2/4$, for given design parameters β_1 and β_2 .²

In order to cover different dependence structures we consider 10 combinations of the β_1 and β_2 parameters as detailed in Table 1. Furthermore, we set $m \in \{16, 32, 64, 128, 256\}$, $m_0 \in \{\frac{1}{4}m, \frac{1}{2}m, \frac{1}{3}m, \frac{2}{3}m, \frac{3}{4}m\}$, $\alpha = 0.05$ and $\pi_\alpha \in \{0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90\}$. Therefore, the experimental design gives rise to 2,250 experiments, each carried out using 10,000 replications.³

²A detailed explanation of the simulation algorithm for the p values is offered in Appendix A.

³Simulations have been carried out using R version 2.12.1 (R Development Core Team 2011).

β_1	β_2	$\bar{\sigma}_{ij}$
0.0	0.0	0.0000
0.0	1.0	0.2500
0.1	1.0	0.3025
0.2	1.0	0.3600
0.3	1.0	0.4225
0.4	1.0	0.4900
0.5	1.0	0.5625
0.6	1.0	0.6400
0.7	1.0	0.7225
0.8	1.0	0.8100

Table 1: Combinations of the β_1 and β_2 parameters. $\bar{\sigma}_{ij}$ represents the average off-diagonal element of Σ .

3.1. Simulation results

The first quantity of interest in our simulations is the *actual* FDR, defined as the percentage of falsely rejected hypotheses over the total number of rejected hypotheses. In our experiments we use a nominal 10% FDR level. The second quantity of interest is the *average power*, defined as the percentage of correctly rejected hypotheses over the total number of false hypotheses (see, e.g., [Benjamini and Hochberg 1995](#), p. 296).

Prior to proceeding to the detailed analysis of the results, it is worth noting that the median FDR for the BH procedure is about 8.1%, but in 41.4% of the experiments the actual FDR was higher than the nominal 10% level with values as high as about 38%. On the contrary, only 3 out of 2,250 experiments resulted in an actual FDR greater than 10% when using the spBH procedure and the maximum FDR was in this case equal to 11.6%. This confirms our conjecture that our spBH procedure is effective in controlling the FDR in the presence of correlated experiments. Of course, we expect to pay the better FDR control performance of the spBH procedure in terms of lower power, as compared to the original BH procedure. Indeed, the median average power of the spBH procedure is 23.6% as opposed to the 58.8% of the BH procedure; the maximum average power is 72.7% and 92.8%, respectively. However, it is fair to say that the power of the two procedures should not be compared *face value*, because the fraction of erroneously rejected null hypotheses in the BH procedure is much higher than when using the spBH approach. It is therefore natural that the BH procedure is “more powerful”.

We now proceed by synthesising the 2,250 simulation outcomes for the FDR using response surfaces (see e.g., [Hendry 1984](#)). First, we transform the actual FDR using the logit transform $x^* := \log[x/(1-x)]$; then we fit a linear model to the transformed outcomes, using m_0/m , π_α and $\bar{\sigma}_{ij}$ as explanatory variables. We use heteroskedasticity-consistent standard errors ([Cribari-Neto 2004](#)). The response surfaces for the FDR are reported in [Tables 2 and 3](#).

Actual BH’s FDR increases significantly with the proportion of null hypotheses as well as with the correlation among the tests, while it tends to decrease with the power of the single test. These paths are present, but much less pronounced, also in the spBH procedure.

Direct interpretation of the coefficients reported in [Tables 2 and 3](#) is somewhat cumbersome

	Estimate	Std. Error	t value	Pr(> t)
Intercept	-4.190389	0.040366	-103.8093	<2e-16
m_0/m	4.008254	0.073114	54.8218	<2e-16
π_α	-0.004207	0.049951	-0.0842	0.9329
$\bar{\sigma}_{ij}$	1.556458	0.064716	24.0507	<2e-16
$m_0/m:\pi_\alpha$	-1.729597	0.074707	-23.1516	<2e-16
$m_0/m:\bar{\sigma}_{ij}$	1.441892	0.061784	23.3375	<2e-16
$\pi_\alpha:\bar{\sigma}_{ij}$	-1.951205	0.087856	-22.2091	<2e-16
Residual standard error: 0.07867 on 2243 degrees of freedom				
Multiple R-squared: 0.9887, Adjusted R-squared: 0.9887				
F-statistic: 3.276e+04 on 6 and 2243 DF, p-value: < 2.2e-16				

Table 2: Estimated response surface for the FDR of the BH procedure. The terms indicated by “:” are interaction terms.

	Estimate	Std. Error	t value	Pr(> t)
Intercept	-4.71734	0.25170	-18.7416	< 2.2e-16
m_0/m	4.63742	0.37809	12.2655	< 2.2e-16
π_α	-3.00331	0.36362	-8.2594	2.470e-16
$\bar{\sigma}_{ij}$	1.19758	0.33000	3.6290	0.0002909
$m_0/m:\pi_\alpha$	-1.15899	0.52702	-2.1991	0.0279708
$m_0/m:\bar{\sigma}_{ij}$	-0.69754	0.28967	-2.4081	0.0161165
$\pi_\alpha:\bar{\sigma}_{ij}$	-0.27562	0.44988	-0.6126	0.5401715
Residual standard error: 0.6316 on 2243 degrees of freedom				
Multiple R-squared: 0.6388, Adjusted R-squared: 0.6379				
F-statistic: 661.3 on 6 and 2243 DF, p-value: < 2.2e-16				

Table 3: Estimated response surface for the FDR of the spBH procedure. The terms indicated by “:” are interaction terms.

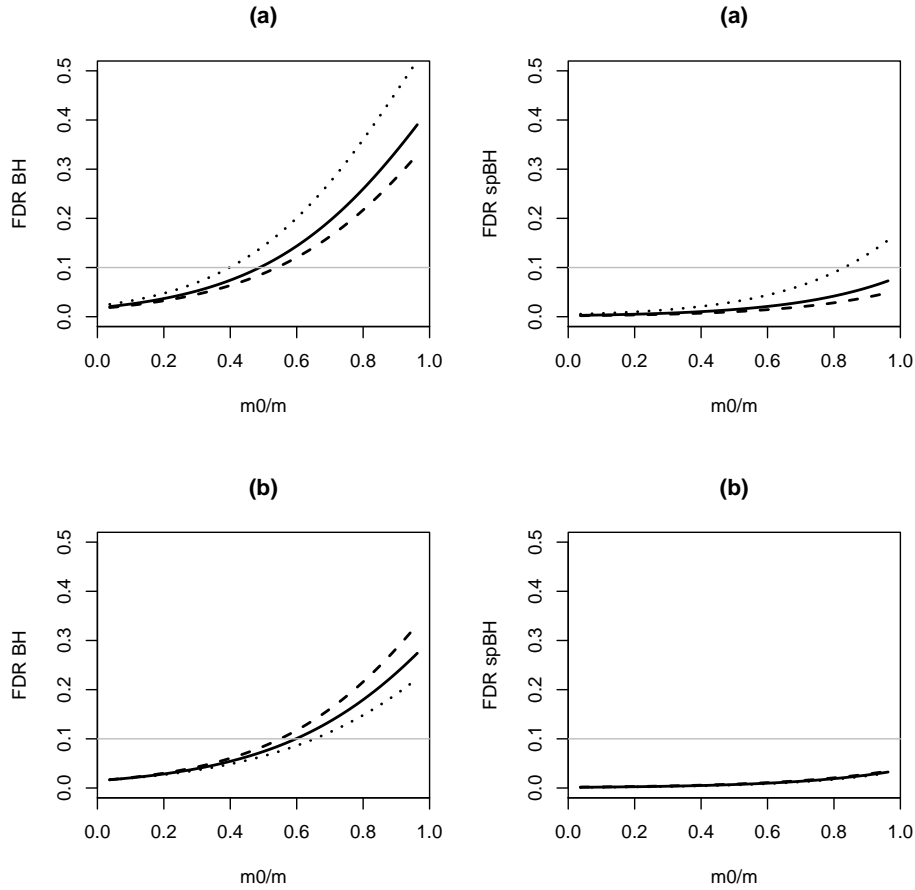


Figure 1: Fitted response surfaces (in original FDR untransformed scale) as functions of m_0/m for various values of π_α and $\bar{\sigma}_{ij}$. In panels (a), solid lines are used to represent $\pi_\alpha = 0.6$ and $\bar{\sigma}_{ij} = 0.5$; dashed lines for $\pi_\alpha = 0.7$ and $\bar{\sigma}_{ij} = 0.5$; dotted lines for $\pi_\alpha = 0.4$ and $\bar{\sigma}_{ij} = 0.5$. In panels (b), solid lines are used to represent $\pi_\alpha = 0.8$ and $\bar{\sigma}_{ij} = 0.5$; dashed lines for $\pi_\alpha = 0.8$ and $\bar{\sigma}_{ij} = 0.7$; dotted lines for $\pi_\alpha = 0.8$ and $\bar{\sigma}_{ij} = 0.3$. The thin horizontal lines represent the nominal 10% FDR.

because of the presence of interaction terms and because the dependent variable is transformed to the logit scale. Therefore, making use of the estimated response surfaces, in Figure 1 we offer a more intuitive approach. In fact, Figure 1 represents the fitted response surfaces as functions of m_0/m for different values of the power of the individual tests, π_α , and of the average correlation among the test statistics, $\bar{\sigma}_{ij}$. In particular, in panels (a) $\bar{\sigma}_{ij}$ is held constant and equal to 0.5, while π_α is set to 0.5 (solid lines), 0.7 (dashed lines) and 0.4 (dotted lines). It is clear that, for the same proportion of null hypotheses and constant correlation among tests, the FDR increases for lower values of the power of the individual tests. While the increase in the spBH procedure is not such to jeopardise FDR control, the influence of π_α on the ability of the BH procedure to control the FDR can be extremely important. In panels (b) we consider quite powerful individual tests ($\pi_\alpha = 0.8$) and let the correlation among tests free to vary by setting $\bar{\sigma}_{ij} = 0.5$ (solid lines), $\bar{\sigma}_{ij} = 0.7$ (dashed lines) and $\bar{\sigma}_{ij} = 0.3$ (dotted lines). The results show that the FDR of the BH procedure increases for increasing values of the dependence of the individual tests, as expected, even in the presence of rather powerful individual tests. On the contrary, the FDR of the spBH procedure does not change significantly with $\bar{\sigma}_{ij}$. From Figure 1 it could be argued that the BH procedure could safely be used when the individual tests have good power and the expected number of null hypothesis is small relative to the total number of tested hypothesis ($m_0/m < 0.20$, say), but in many practical circumstances the situation is quite the opposite, where the investigator has to identify a small number of alternative hypothesis within a large number of tested hypotheses. In this situation, the use of the BH procedure can give misleading results, as far as the tests are dependent.

Figure 2 represents the fitted response surfaces for the FDR as functions of the power of the individual tests, π_α . The FDR decreases as π_α increases, especially for the BH procedure, as anticipated. In panels (a) the correlation among tests is held constant at $\bar{\sigma}_{ij} = 0.5$ and the proportion of null hypothesis is set to $m_0/m = 0.6$ (solid lines), $m_0/m = 0.8$ (dashed lines) and $m_0/m = 0.4$ (dotted lines). Of course, for constant values of π_α the FDR is larger for larger m_0/m . The indication that emerges is that, in order to use FDR procedures successfully, they must be based on powerful individual tests. The relation between the FDR and the power of the underlying test is a novel result of our paper. In panels (b) of Figure 2 it is the correlation among the test statistics, $\bar{\sigma}_{ij}$, that is made to vary between 0.5 (solid lines), 0.7 (dashed lines) and 0.3 (dotted lines), while m_0/m is held constant at 0.5. It is again very well visible that the correlation influences the FDR in the BH procedure, but not in the spBH one.

Finally, the response surfaces as functions of the average correlation among the test statistics are plotted in Figure 3. Looking at the figure, it is again clear how the FDR in the BH increases as the correlation across tests increases, while the FDR in the spBH procedure is largely independent of the correlation existing across the test statistics, again confirming our theoretical results. Furthermore, the FDR depends positively on the proportion of null hypotheses (panels a in Figure 3) and negatively on the power of the individual tests (panels b in Figure 3), as already highlighted. This dependence is considerably smaller in the spBH procedure, as compared to the BH one.

We turn now to examine the power of the two procedures. Given that the spBH procedure is based on the repeated application of the BH procedure on *pairs* of p values, we expect it to have lower power than the BH procedure (see e.g., Yekutieli 2008, p. 407).

The estimated response surfaces are reported in Tables 4 and 5. For both procedures the average power increases with the power of the single tests (π_α , as expected), but BH's power

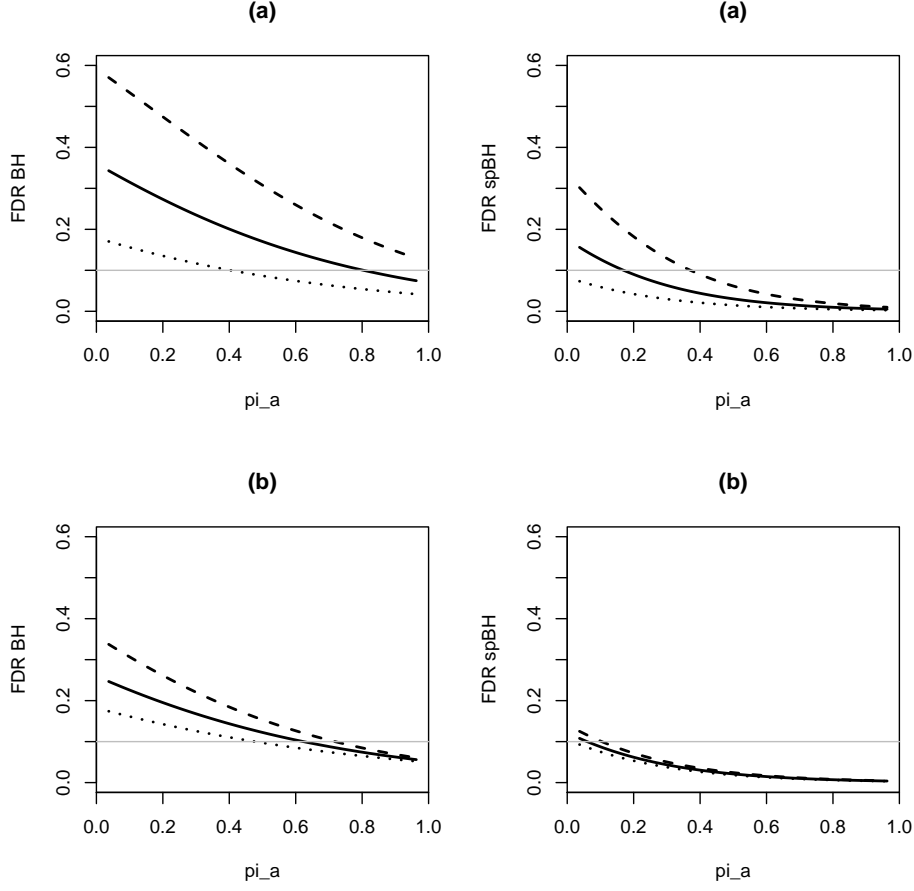


Figure 2: Fitted response surfaces (in original FDR untransformed scale) as functions of π_α for various values of m_0/m and $\bar{\sigma}_{ij}$. In panels (a), solid lines are used to represent $m_0/m = 0.6$ and $\bar{\sigma}_{ij} = 0.5$; dashed lines for $m_0/m = 0.8$ and $\bar{\sigma}_{ij} = 0.5$; dotted lines for $m_0/m = 0.4$ and $\bar{\sigma}_{ij} = 0.5$. In panels (b), solid lines are used to represent $m_0/m = 0.5$ and $\bar{\sigma}_{ij} = 0.5$; dashed lines for $m_0/m = 0.5$ and $\bar{\sigma}_{ij} = 0.7$; dotted lines for $m_0/m = 0.5$ and $\bar{\sigma}_{ij} = 0.3$. The thin horizontal lines represent the nominal 10% FDR.

	Estimate	Std. Error	t value	Pr(> t)
Intercept	-3.612765	0.052739	-68.5029	< 2.2e-16
m_0/m	-1.346034	0.088237	-15.2548	< 2.2e-16
π_α	7.282354	0.075146	96.9098	< 2.2e-16
$\bar{\sigma}_{ij}$	0.675278	0.075907	8.8961	< 2.2e-16
$m_0/m:\pi_\alpha$	-0.800207	0.117893	-6.7875	1.455e-11
$m_0/m:\bar{\sigma}_{ij}$	0.547344	0.057548	9.5111	< 2.2e-16
$\pi_\alpha:\bar{\sigma}_{ij}$	-1.677226	0.105609	-15.8815	< 2.2e-16
Residual standard error: 0.1233 on 2243 degrees of freedom				
Multiple R-squared: 0.9797, Adjusted R-squared: 0.9796				
F-statistic: 1.801e+04 on 6 and 2243 DF, p-value: < 2.2e-16				

Table 4: Estimated response surface for the power of the BH procedure. The terms indicated by “:” are interaction terms.

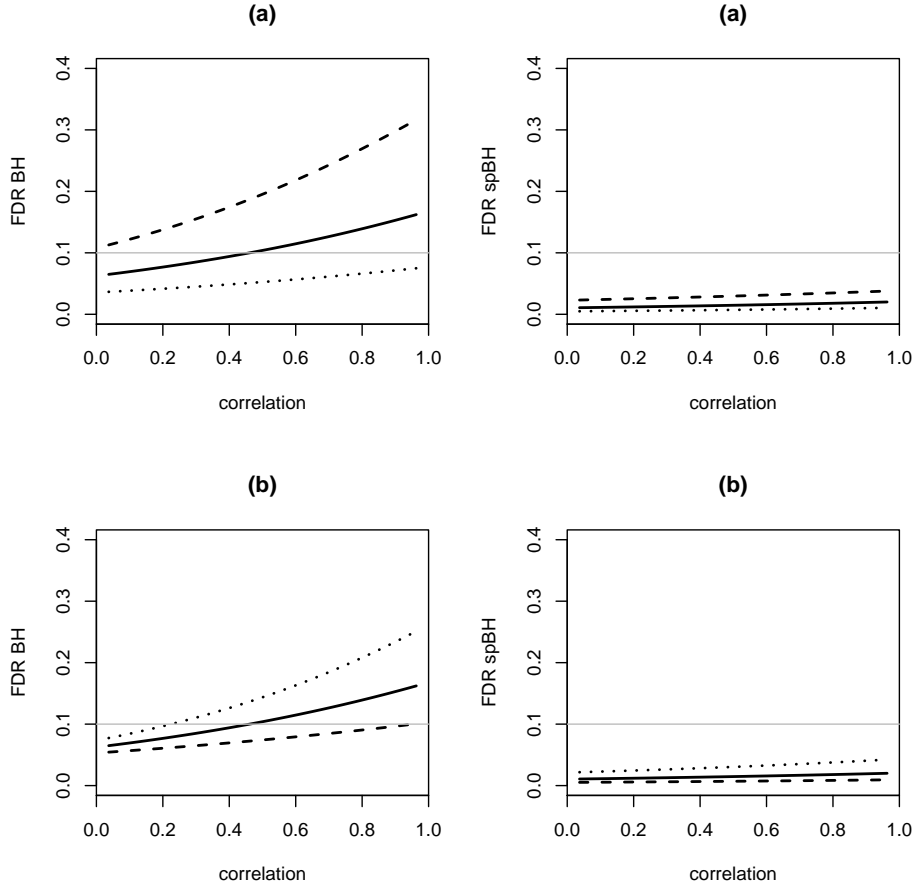


Figure 3: Fitted response surfaces (in original FDR untransformed scale) as functions of $\bar{\sigma}_{ij}$ for various values of m_0/m and π_α . In panels (a), solid lines are used to represent $m_0/m = 0.5$ and $\pi_\alpha = 0.6$; dashed lines for $m_0/m = 0.7$ and $\pi_\alpha = 0.6$; dotted lines for $m_0/m = 0.3$ and $\pi_\alpha = 0.6$. In panels (b), solid lines are used to represent $m_0/m = 0.5$ and $\pi_\alpha = 0.6$; dashed lines for $m_0/m = 0.5$ and $\pi_\alpha = 0.8$; dotted lines for $m_0/m = 0.5$ and $\pi_\alpha = 0.4$. The thin horizontal lines represent the nominal 10% FDR.

	Estimate	Std. Error	t value	Pr(> t)
Intercept	-5.196950	0.241581	-21.5123	<2e-16
m_0/m	0.255360	0.375858	0.6794	0.4970
π_α	5.696634	0.325764	17.4870	<2e-16
$\bar{\sigma}_{ij}$	0.181832	0.320364	0.5676	0.5704
$m_0/m:\pi_\alpha$	-0.405152	0.493091	-0.8217	0.4114
$m_0/m:\bar{\sigma}_{ij}$	-0.235477	0.271061	-0.8687	0.3851
$\pi_\alpha:\bar{\sigma}_{ij}$	-0.033556	0.403644	-0.0831	0.9338
Residual standard error: 0.5559 on 2243 degrees of freedom				
Multiple R-squared: 0.6194, Adjusted R-squared: 0.6183				
F-statistic: 608.3 on 6 and 2243 DF, p-value: < 2.2e-16				

Table 5: Estimated response surface for the power of the spBH procedure. The terms indicated by “:” are interaction terms.

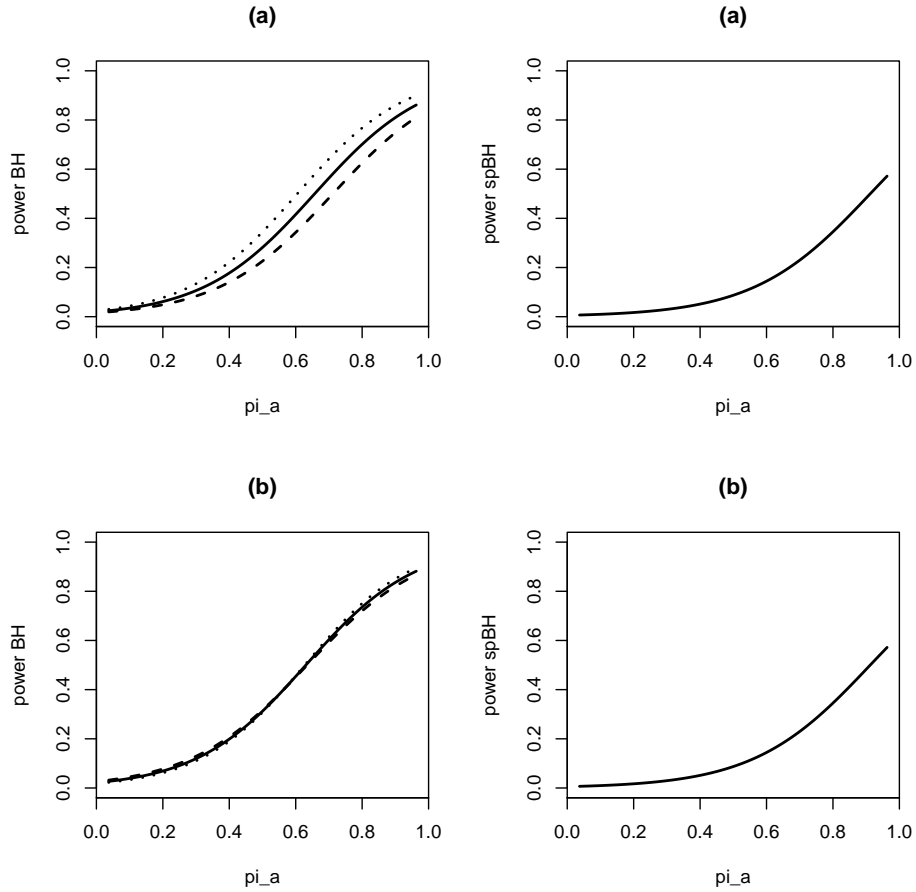


Figure 4: Fitted response surfaces for power as functions of π_α for various values of m_0/m and $\bar{\sigma}_{ij}$. In panels (a), solid lines are used to represent $m_0/m = 0.6$ and $\bar{\sigma}_{ij} = 0.5$; dashed lines for $m_0/m = 0.8$ and $\bar{\sigma}_{ij} = 0.5$; dotted lines for $m_0/m = 0.4$ and $\bar{\sigma}_{ij} = 0.5$. In panels (b), solid lines are used to represent $m_0/m = 0.5$ and $\bar{\sigma}_{ij} = 0.5$; dashed lines for $m_0/m = 0.5$ and $\bar{\sigma}_{ij} = 0.7$; dotted lines for $m_0/m = 0.5$ and $\bar{\sigma}_{ij} = 0.3$. spBH figures present a single line, because in the spBH procedure power depends only on π_α .

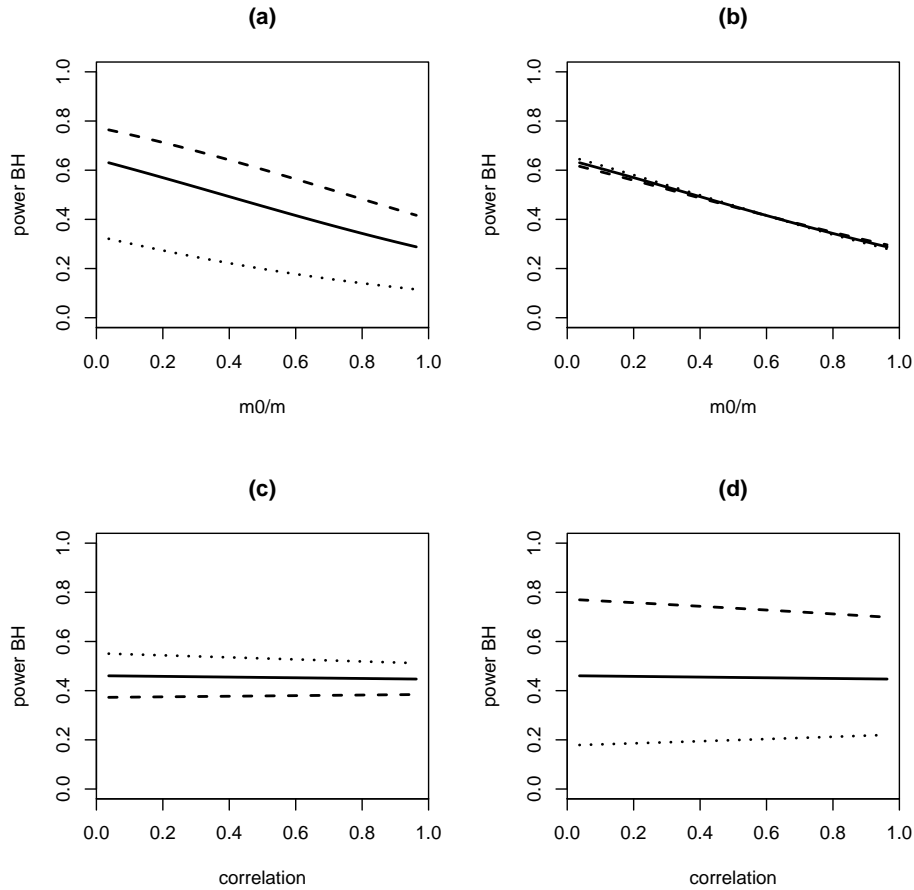


Figure 5: Fitted response surfaces for power of the BH procedure. In panels (a) and (b) the response surfaces are represented as functions of m_0/m for various values of π_α and $\bar{\sigma}_{ij}$. In panel (a) $\bar{\sigma}_{ij}$ is held fixed at 0.5 and the solid, dashed and dotted lines are used to represent $\pi_\alpha = 0.6$, $\pi_\alpha = 0.7$, and $\pi_\alpha = 0.4$, respectively: in panel (b) π_α is held fixed at 0.6 and $\bar{\sigma}_{ij}$ is set to 0.5 (solid), 0.7 (dashed) and 0.3 (dotted). In panels (c) and (d) the response surfaces are represented as functions of $\bar{\sigma}_{ij}$. In panel (c), π_α is held fixed at 0.6 and m_0/m is set to 0.5 (solid), 0.7 (dashed) and 0.3 (dotted): in panel (d) m_0/m is held fixed at 0.5 and π_α is set to 0.6 (solid), 0.8 (dashed) and 0.4 (dotted).

decreases with the fraction of null hypotheses, m_0/m . Benjamini and Hochberg (1995) already found that the power decreases when the number of tested hypothesis increases. Here we emphasize that is the fraction of null hypotheses that really matters most. Furthermore, the power of the BH procedure depends also on the average correlation among the tests, $\bar{\sigma}_{ij}$, although the effect of $\bar{\sigma}_{ij}$ is small. Quite on the contrary, the power of the spBH procedure depends exclusively on the power of the single tests.

Again, in order to have a more direct interpretation of the results, we report the estimated response surfaces in graphical form in Figures 4 and 5.

From Figure 4 it is clear that both procedures are less powerful than the individual tests they are based on: this is the price of correcting for multiplicity. Therefore, it is recommended that these procedures are applied to rather powerful tests. Figure 5 shows that the power of the BH procedure decreases quite sharply when m_0/m increases. Therefore, in a typical situation where $m_1 := m - m_0$ is small relative to m , in the presence of dependent tests the BH procedure is likely to have FDR larger than the nominal level, and reduced power. The adverse effect of m_0/m on power is not present in the spBH procedure whose power depends only on the power of the individual tests. As anticipated, the effect of the average correlation on the power of the BH procedure is instead very small indeed.

4. Concluding remarks

The false discovery rate (FDR, Benjamini and Hochberg 1995) is a powerful approach to multiple testing, where the expected proportion of falsely rejected hypotheses is controlled instead of the familywise error rate (FWER), i.e. the probability of making at least one Type I error. Yekutieli (2008) showed that a modification of the Benjamini-Hochberg (BH) approach can be used in the presence of dependent tests and labelled his procedure as *separate subsets* BH (ssBH). The ssBH procedure is based on the idea of using the BH approach not on all the p values simultaneously, but on separate subset of p values, each satisfying certain dependence properties. However, Yekutieli (2008) left the practical specification of the subsets of p values largely unresolved. In this paper we propose a modification of the ssBH procedure based on a selection of the subsets that guarantees that the dependence properties needed to control the FDR are satisfied. In our version of the ssBH all the separate subsets have cardinality equal to 2: for this reason we label it as the *separate pairs* BH (spBH). We prove that the spBH procedure is indeed able to control the FDR in the presence of correlated tests. However, given that the spBH procedure is based on the application of the BH procedure on pairs of p value, we expect its power to be lower than that of both the BH and ssBH procedure. In order to compare the FDR and power properties of the BH and spBH procedures, we run an extensive Monte Carlo analysis over 2,250 experiments. The results are summarised via response surfaces (Hendry 1984). We show that, contrary to the BH procedure, the spBH procedure is indeed able to control the FDR in the presence of correlated tests. Furthermore, based on our simulations we show that *actual* FDR in the BH procedure increases significantly with the proportion of null hypotheses, with the average correlation among the tests and decreases with the power of the individual tests. As expected, the power of the spBH procedure is lower than the power of the BH one which, however, can be used only in the presence of independent tests. All in all, the spBH procedure gives a very reliable, but rather conservative, test that should be used whenever the cost of falsely rejecting is high.

References

- Amblard C, Girard S (2002). “Symmetry and Dependence Properties within a Semiparametric Family of Bivariate Copulas.” *Nonparametric Statistics*, **14**(6), 715–727.
- Benjamini Y, Hochberg Y (1995). “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society, Series B*, **57**(1), 289–300.
- Benjamini Y, Yekutieli D (2001). “The Control of the False Discovery Rate in Multiple Testing under Dependency.” *Annals of Statistics*, **29**(4), 1165–1188.
- Brown BW, Russell K (1997). “Methods Correcting for Multiple Testing: Operating Characteristics.” *Statistics in Medicine*, **16**(22), 2511–2528.
- Cribari-Neto F (2004). “Asymptotic Inference Under Heteroskedasticity of Unknown Form.” *Computational Statistics & Data Analysis*, **45**(2), 215–233.
- Hendry DF (1984). “Monte Carlo Experimentation in Econometrics.” In Z Griliches, M Intriligator (eds.), *Handbook of Econometrics*, volume 2, chapter 16, pp. 937–976. Elsevier Science Publishers, Amsterdam.
- R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://R-project.org>.
- Sklar A (1959). “Fonctions de répartition à n dimensions et leurs marges.” *Publications de l’Institut de Statistique de L’Université de Paris*, **8**, 229–231.
- Sorić B (1989). “Statistical ‘Discoveries’ and Effect-Size Estimation.” *Journal of the American Statistical Association*, **84**(406), 608–610.
- Storey JD, Taylor JE, Siegmund D (2004). “Strong Control, Conservative Point Estimation and Simultaneous Conservative Consistency of False Discovery Rates: A Unified Approach.” *Journal of the Royal Statistical Society, Series B*, **66**(1), 187–205.
- Verhoeven KJF, Simonsen KL, McIntyre LM (2005). “Implementing false discovery rate control: increasing your power.” *Oikos*, **108**(3), 643–647.
- Yekutieli D (2008). “False Discovery Rate Control for Non-positively Regression Dependent Test Statistics.” *Journal of Statistical Planning and Inference*, **138**(2), 405–415.

Appendix A: Monte Carlo details

This Appendix provides some details about the algorithm used for the the generation of the p values utilized in the Monte Carlo experiments.

In all Monte Carlo experiments m_0 out of m p values are generated under the null; the remaining $m_1 := m - m_0$ p values are generated under the alternative:

1. Generate a random m -vector $\boldsymbol{\rho}$ with generic i -th element $\rho_i \sim U_{(\beta_1, \beta_2)}$ ($0 < \beta_1 < \beta_2 < 1$) and compute $\boldsymbol{\Sigma}^* = \boldsymbol{\rho}\boldsymbol{\rho}'$. Set $\text{diag}(\boldsymbol{\Sigma}^*) = \boldsymbol{\nu}$, with $\boldsymbol{\nu} = (1, 1, \dots, 1)$, and call $\boldsymbol{\Sigma}$ the resulting matrix. $\boldsymbol{\Sigma}$ is a proper symmetric positive definite correlation matrix. Furthermore, given that ρ_i and ρ_j are independent, $\mathbf{E}(\rho_i\rho_j) = \mathbf{E}(\rho_i)\mathbf{E}(\rho_j)$ so that the expected value of the off-diagonal elements of $\boldsymbol{\Sigma}$ is $\bar{\sigma}_{ij} := \mathbf{E}(\sigma_{ij}) = (\beta_1 + \beta_2)^2 / 4$.
2. Generate the m -variate normal $\mathbf{y} \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma})$.
3. Shift the mean of the last m_1 normals in \mathbf{y} of a quantity $\mu_1 := \Phi^{-1}(\pi_\alpha) - \Phi^{-1}(\alpha)$. This ensures that the simulated power at the significance level α is π_α .
4. The i -th p value (with $i = 1, \dots, m$) is $1 - \Phi(y_i)$.

Affiliation:

Roy Cerqueti
University of Macerata
Dept. of Economic and Financial Institutions
Via Crescimbeni, 20. I-62100 Macerata, Italy
E-mail: roy.cerqueti@unimc.it

Mauro Costantini
University of Vienna
Dept. of Economics
Brünner Straße, 72. A-1210 Vienna, Austria
E-mail: mauro.costantini@univie.ac.at

Claudio Lupi (corresponding author)
University of Molise
Dept. of Economics, Management and Social Sciences (SEGeS)
Via De Sanctis. I-86100 Campobasso, Italy
E-mail: lupi@unimol.it