Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA

Large-Scale Multiple Testing of Correlations

T. Tony CAI and Weidong LIU

Multiple testing of correlations arises in many applications including gene coexpression network analysis and brain connectivity analysis. In this article, we consider large-scale simultaneous testing for correlations in both the one-sample and two-sample settings. New multiple testing procedures are proposed and a bootstrap method is introduced for estimating the proportion of the nulls falsely rejected among all the true nulls. We investigate the properties of the proposed procedures both theoretically and numerically. It is shown that the procedures asymptotically control the overall false discovery rate and false discovery proportion at the nominal level. Simulation results show that the methods perform well numerically in terms of both the size and power of the test and it significantly outperforms two alternative methods. The two-sample procedure is also illustrated by an analysis of a prostate cancer dataset for the detection of changes in coexpression patterns between gene expression levels. Supplementary materials for this article are available online.

KEY WORDS: Correlation; False discovery proportion; False discovery rate; Multiple testing.

Downloaded by [University of Pennsylvania] at 06:26 06 June 2016

1. INTRODUCTION

Knowledge of the correlation structure is essential for a wide range of statistical methodologies and applications. For example, gene coexpression network plays an important role in genomics and understanding the correlations between the genes is critical for the construction of such a network. See, for example, Kostka and Spang (2004), Carter et al. (2004), Lai et al. (2004), and de la Fuente (2010). In this article, we consider large-scale multiple testing of correlations in both one- and two-sample cases. A particular focus is on the high-dimensional setting where the dimension can be much larger than the sample size.

Multiple testing of correlations arises in many applications, including brain connectivity analysis (Shaw et al. 2006) and gene coexpression network analysis (Zhang, Li, and Deng 2008; de la Fuente 2010), where one tests thousands or millions of hypotheses on the changes of the correlations between genes. Multiple testing of correlations also has important applications in the selection of the significant gene pairs and in correlation analysis of factors that interact to shape children's language development and reading ability; see Lee, Hsu, and Sajdak (2004), Carter et al. (2004), Zhu et al. (2005), Dubois et al. (2010), Hirai et al. (2007), and Raizada et al. (2008).

A common goal in multiple testing is to control the false discovery rate (FDR), which is defined to be the expected proportion of false positives among all rejections. This testing problem has been well studied in the literature, especially in the case where the test statistics are independent. The well-known step-up procedure of Benjamini and Hochberg (1995), which guarantees the control of the FDR, thresholds the p-values of the individual tests. Sun and Cai (2007) developed under a mixture model an optimal and adaptive multiple testing procedure that minimizes the false nondiscovery rate subject to a constraint on the FDR. See also Storey (2002), Genovese and Wasserman (2004), and Efron (2004), among many others. The multiple testing problem is more complicated when the test statistics are dependent. The effects of dependency on FDR procedures have been considered, for example, in Benjamini and Yekutieli (2001), Storey, Taylor, and Siegmund (2004), Qiu, Klebanov, and Yakovlev (2005), Farcomeni (2007), Wu (2008), Efron (2007), and Sun and Cai (2009). In particular, Qiu, Klebanov, and Yakovlev (2005) demonstrated that the dependency effects can significantly deteriorate the performance of many FDR procedures. Farcomeni (2007) and Wu (2008) showed that the FDR is controlled at the nominal level by the Benjamini–Hochberg step-up procedure under some stringent dependency assumptions. The procedure in Benjamini and Yekutieli (2001) allows the general dependency by paying a logarithmic term loss on the FDR, which makes the method very conservative.

For large-scale multiple testing of correlations, a natural starting point is the sample correlation matrix, whose entries are intrinsically dependent even if the original observations are independent. The dependence structure among these sample correlations is rather complicated. The difficulties of this multiple testing problem lie in the construction of suitable test statistics for testing the individual hypotheses and more importantly in constructing a good procedure to account for the multiplicity of the tests so that the overall FDR is controlled. To the best of our knowledge, existing procedures cannot be readily applied to this testing problem to have a solid theoretical guarantee on the FDR level while maintaining good power.

In the one-sample case, let $\boldsymbol{X} = (X_1, \ldots, X_p)'$ be a p-dimensional random vector with mean $\boldsymbol{\mu}$ and correlation matrix $\boldsymbol{R} = (\rho_{ij})_{p\times p}$, and one wishes to simultaneously test the hypotheses

$$H_{0ij}: \quad \rho_{ij} = 0 \quad \text{versus} \quad H_{1ij}: \quad \rho_{ij} \neq 0, \quad \text{for} \quad 1 \le i < j \le p, \tag{1}$$

based on a random sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ from the distribution of $\boldsymbol{X}$. In the two-sample case, let $\boldsymbol{X} = (X_1, \ldots, X_p)'$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_p)'$ be two p-dimensional random vectors with means

T. Tony Cai is Dorothy Silberberg Professor of Statistics, Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104 (E-mail: *tcai@wharton.upenn.edu*). Weidong Liu is Professor, Department of Mathematics, Institute of Natural Sciences and MOE-LSC, Shanghai Jiao Tong University, Shanghai, China (E-mail: *liuweidong99@gmail.com*, *weidongl@sjtu.edu.cn*). Tony Cai's research was supported in part by NSF Grant DMS-1208982 and DMS-1403708, and NIH Grant R01 CA-127334. Weidong Liu's research was supported by NSFC, grants no. 11201298, no. 11322107, and no. 11431006, Program for New Century Excellent Talents in University, Shanghai Pujiang Program, 973 Program (2015CB856004) and a grant from Australian Research Council.

Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/jasa*.

© 2016 American Statistical Association
Journal of the American Statistical Association
March 2016, Vol. 111, No. 513, Theory and Methods
DOI: 10.1080/01621459.2014.999157

Downloaded by [University of Pennsylvania] at 06:26 06 June 2016

$\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and correlation matrices $\boldsymbol{R}_1 = (\rho_{ij1})_{p\times p}$ and $\boldsymbol{R}_2 = (\rho_{ij2})_{p\times p}$, respectively, and we are interested in the simultaneous testing of correlation changes,

$$H_{0ij}: \quad \rho_{ij1} = \rho_{ij2} \quad \text{versus} \quad H_{1ij}: \quad \rho_{ij1} \neq \rho_{ij2}, \quad \text{for} \quad 1 \le i < j \le p, \tag{2}$$

based on two independent random samples, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{n_1}$ from the distribution of $\boldsymbol{X}$ and $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{n_2}$ from the distribution of $\boldsymbol{Y}$, where $c_1 \le n_1/n_2 \le c_2$ for some $c_1, c_2 > 0$.

We shall focus on the two-sample case in the following discussion. The one-sample case is slightly simpler and will be considered in Section 4. The classical statistics for correlation detection are based on the sample correlations. For the two independent and identically distributed random samples $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{n_1}\}$ and $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_{n_2}\}$, denote by $\boldsymbol{X}_k = (X_{k,1}, \ldots, X_{k,p})'$ and $\boldsymbol{Y}_k = (Y_{k,1}, \ldots, Y_{k,p})'$. The sample correlations are defined by

$$\hat{\rho}_{ij1} = \frac{\sum_{k=1}^{n_1}(X_{k,i} - \bar{X}_i)(X_{k,j} - \bar{X}_j)}{\sqrt{\sum_{k=1}^{n_1}(X_{k,i} - \bar{X}_i)^2 \sum_{k=1}^{n_1}(X_{k,j} - \bar{X}_j)^2}},$$

and

$$\hat{\rho}_{ij2} = \frac{\sum_{k=1}^{n_2}(Y_{k,i} - \bar{Y}_i)(Y_{k,j} - \bar{Y}_j)}{\sqrt{\sum_{k=1}^{n_2}(Y_{k,i} - \bar{Y}_i)^2 \sum_{k=1}^{n_2}(Y_{k,j} - \bar{Y}_j)^2}},$$

where $\bar{X}_i = \frac{1}{n_1}\sum_{k=1}^{n_1} X_{k,i}$ and $\bar{Y}_i = \frac{1}{n_2}\sum_{k=1}^{n_2} Y_{k,i}$. The sample correlations $\hat{\rho}_{ij1}$ and $\hat{\rho}_{ij2}$ are heteroscedastic and the null distribution of $\hat{\rho}_{ij1}$ and $\hat{\rho}_{ij2}$ depends on unknown parameters. A well-known variance stabilization method is Fisher's z-transformation,

$$\hat{Z} = \frac{1}{2}\ln\frac{1+\hat{\rho}}{1-\hat{\rho}},$$

where $\hat{\rho}$ is a sample correlation coefficient. In the two-sample case, it is easy to see that under the null hypothesis $H_{0ij}: \rho_{ij1} = \rho_{ij2}$ and the bivariate normal assumptions on (X_i, X_j) and (Y_i, Y_j),

$$F_{ij} \equiv \frac{\sqrt{n_1 n_2}}{2\sqrt{n_1+n_2}}\left[\ln\left(\frac{1+\hat{\rho}_{ij1}}{1-\hat{\rho}_{ij1}}\right) - \ln\left(\frac{1+\hat{\rho}_{ij2}}{1-\hat{\rho}_{ij2}}\right)\right] \to N(0,1). \tag{3}$$

See, for example, Anderson (2003). To perform multiple testing (2), a natural approach is to use F_{ij} as the test statistics and then apply a multiple testing method such as the Benjamini–Hochberg procedure or the Benjamini–Yekutieli procedure to the p-values calculated from F_{ij}. See, for example, Shaw et al. (2006) and Zhang, Li, and Deng (2008). However, the asymptotic normality result in (3) heavily depends on the bivariate normality assumptions on (X_i, X_j) and (Y_i, Y_j). The behavior of F_{ij} in the nonnormal case is complicated with the asymptotic variance of F_{ij} depending on $\mathrm{E}X_i^2X_j^2$ and $\mathrm{E}Y_i^2Y_j^2$ even when $\rho_{ij1} = \rho_{ij2} = 0$; see Hawkins (1989). As will be seen in Section 5, the combination of Fisher's z-transformation with either the Benjamini–Hochberg procedure or the Benjamini–Yekutieli procedure does not in general perform well numerically.

In this article, we propose a large-scale multiple testing procedure for correlations that controls the FDR and the false discovery proportion (FDP) asymptotically at any prespecified level $0 < \alpha < 1$. The multiple testing procedure is developed in two stages. We first construct a test statistic for testing the equality of each individual pair of correlations, $H_{0ij}: \rho_{ij1} = \rho_{ij2}$. It is shown that the test statistic has standard normal distribution asymptotically under the null hypothesis H_{0ij} and it is robust against a class of nonnormal population distributions of $\boldsymbol{X}$ and $\boldsymbol{Y}$. We then develop a procedure to account for the multiplicity in testing a large number of hypotheses so that the overall FDR and FDP levels are under control. A key step is the estimation of the proportion of the nulls falsely rejected by the procedure among all the true nulls at any given threshold level. A bootstrap method is introduced for estimating this proportion.

The properties of the proposed procedure are investigated both theoretically and numerically. It is shown that, under regularity conditions, the multiple testing procedure controls the overall FDR and FDP at the prespecified level asymptotically. The proposed procedure works well even when the components of the random vectors are strongly dependent and hence provides theoretical guarantees for a large class of correlation matrices.

In addition to the theoretical properties, the numerical performance of the proposed multiple testing procedure is also studied using both simulated and real data. A simulation study is carried out in Section 5.1, which shows that this procedure performs well numerically in terms of both the size and power of the test. In particular, the procedure significantly outperforms the methods using Fisher's z-transformation together with either the Benjamini–Hochberg procedure or the Benjamini–Yekutieli procedure, especially in the nonnormal case. The simulation study also shows that the numerical performance of the proposed procedure is not sensitive to the choice of the bootstrap replication number. We also illustrate our procedure with an analysis of a prostate cancer dataset for the detection of changes in the coexpression patterns between gene expression levels. The procedure identifies 1341 pairs of coexpression genes (out of a total of 124,750 pairs) and 1.07% nonzero entries of the coexpression matrix. Our method leads to a clear and easily interpretable coexpression network.

The rest of the article is organized as follows. Section 2 gives a detailed description of the proposed multiple testing procedure. Theoretical properties of the procedure are investigated in Section 3. It is shown that, under some regularity conditions, the procedure controls the FDR and FDP at the nominal level asymptotically. Section 4 discusses the one-sample case. Numerical properties of the proposed testing procedure are studied in Section 5. The performance of the procedure is compared to that of the methods based on the combination of Fisher's z-transformation with either the Benjamini–Hochberg procedure or the Benjamini–Yekutieli procedure. A real dataset is analyzed in Section 5.2. A discussion on extensions and related problems is given in Section 6 and all the proofs are contained in the supplementary material.

2. FDR CONTROL PROCEDURE

In this section, we present a detailed description of the multiple testing procedure for correlations in the two-sample case. The theoretical results given in Section 3 show that the procedure

Downloaded by [University of Pennsylvania] at 06:26 06 June 2016

controls the FDR and FDP at the prespecified level asymptotically.

We begin by constructing a test statistic for testing each individual pair of correlations, $H_{0ij} : \rho_{ij1} = \rho_{ij2}$. In this article, we shall focus on the class of populations with the elliptically contoured distributions (see Condition (C2) in Section 3), which is more general than the multivariate normal distributions. The test statistic for general population distributions is introduced in Section 6.3. Under Condition (C2) and the null hypothesis $H_{0ij} : \rho_{ij1} = \rho_{ij2}$, as $(n_1, n_2) \to \infty$,

$$\frac{\hat{\rho}_{ij1} - \hat{\rho}_{ij2}}{\sqrt{\frac{\kappa_1}{n_1}(1-\rho_{ij1}^2)^2 + \frac{\kappa_2}{n_2}(1-\rho_{ij2}^2)^2}} \to N(0,1) \tag{4}$$

with

$$\kappa_1 \equiv \frac{1}{3}\frac{\mathrm{E}(X_i-\mu_{i1})^4}{[\mathrm{E}(X_i-\mu_{i1})^2]^2} \quad \text{and} \quad \kappa_2 \equiv \frac{1}{3}\frac{\mathrm{E}(Y_i-\mu_{i2})^4}{[\mathrm{E}(Y_i-\mu_{i2})^2]^2},$$

where $(\mu_{11}, \ldots, \mu_{p1})' = \boldsymbol{\mu}_1$ and $(\mu_{12}, \ldots, \mu_{p2})' = \boldsymbol{\mu}_2$. Note that $\kappa_i \geq \frac{1}{3}$ for $i = 1, 2$ and they are related to the kurtosis with $\kappa_1 = \frac{1}{3}\kappa_X + 1$, where $\kappa_X = \frac{\mathrm{E}(X_i-\mu_{i1})^4}{[\mathrm{E}(X_i-\mu_{i1})^2]^2} - 3$ is the kurtosis of X. For multivariate normal distributions, $\kappa_1 = \kappa_2 = 1$.

In general, the parameters ρ_{ij1}, ρ_{ij2}, κ_1, and κ_2 in the denominator are unknown and need to be estimated. In this article, we estimated κ_1 and κ_2, respectively, by

$$\hat{\kappa}_1 = \frac{1}{3p}\sum_{i=1}^{p}\frac{n_1\sum_{k=1}^{n_1}(X_{k,i}-\bar{X}_i)^4}{\left[\sum_{k=1}^{n_1}(X_{k,i}-\bar{X}_i)^2\right]^2}$$

and

$$\hat{\kappa}_2 = \frac{1}{3p}\sum_{i=1}^{p}\frac{n_2\sum_{k=1}^{n_2}(Y_{k,i}-\bar{Y}_i)^4}{\left[\sum_{k=1}^{n_2}(Y_{k,i}-\bar{Y}_i)^2\right]^2}.$$

To estimate ρ_{ij1} and ρ_{ij2}, taking into account of possible sparsity of the correlation matrices, we use the thresholded version of the sample correlation coefficients

$$\tilde{\rho}_{ijl} = \hat{\rho}_{ijl} I\left\{\frac{|\hat{\rho}_{ijl}|}{\sqrt{\frac{\hat{\kappa}_l}{n_l}\left(1-\hat{\rho}_{ijl}^2\right)^2}} \geq 2\sqrt{\frac{\log p}{n_l}}\right\}, \quad l = 1, 2,$$

where $I\{\cdot\}$ denotes the indicator function. Let $\tilde{\rho}_{ij}^2 = \max\{\tilde{\rho}_{ij1}^2, \tilde{\rho}_{ij2}^2\}$ and we use $\tilde{\rho}_{ij}^2$ to replace ρ_{ij1}^2 and ρ_{ij2}^2 in (4). We propose the test statistic

$$T_{ij} = \frac{\hat{\rho}_{ij1} - \hat{\rho}_{ij2}}{\sqrt{\frac{\hat{\kappa}_1}{n_1}\left(1-\tilde{\rho}_{ij}^2\right)^2 + \frac{\hat{\kappa}_2}{n_2}\left(1-\tilde{\rho}_{ij}^2\right)^2}} \tag{5}$$

for testing the individual hypotheses $H_{0ij} : \rho_{ij1} = \rho_{ij2}$. Note that under H_{0ij}, $\tilde{\rho}_{ij}^2$ is a consistent estimator of ρ_{ij1} and ρ_{ij2}. On the other hand, under the alternative H_{1ij}, $\sqrt{\frac{\hat{\kappa}_1}{n_1}(1-\tilde{\rho}_{ij}^2)^2 + \frac{\hat{\kappa}_2}{n_2}(1-\tilde{\rho}_{ij}^2)^2} \leq \sqrt{\frac{\hat{\kappa}_1}{n_1}(1-\tilde{\rho}_{ij1}^2)^2 + \frac{\hat{\kappa}_2}{n_2}(1-\tilde{\rho}_{ij2}^2)^2}$. Hence, T_{ij} will be more powerful than the test statistic using $\tilde{\rho}_{ij1}$ and $\tilde{\rho}_{ij2}$ to estimate ρ_{ij1} and ρ_{ij2}, respectively.

Before introducing the multiple testing procedure, it is helpful to understand the basic properties of the test statistics T_{ij}, which are in general correlated. It can be proved that, under the null hypothesis H_{0ij} and certain regularity conditions,

$$\sup_{0 \leq t \leq b\sqrt{\log p}} \left|\frac{\mathrm{P}(|T_{ij}| \geq t)}{2 - 2\Phi(t)} - 1\right| \to 0 \quad \text{as } (n_1, n_2) \to \infty$$

uniformly in $1 \leq i < j \leq p$ and $p \leq n^r$ for any $b > 0$ and $r > 0$, where Φ is the cumulative distribution function of the standard normal distribution; see Proposition 1 in Section 3.

Denote the set of true null hypotheses by

$$\mathcal{H}_0 = \{(i, j) : 1 \leq i < j \leq p, \rho_{ij1} = \rho_{ij2}\}.$$

Since the asymptotic null distribution of each test statistic T_{ij} is standard normal, it is easy to see that

$$\mathrm{P}(\max_{(i,j)\in\mathcal{H}_0} |T_{ij}| \geq 2\sqrt{\log p}) \to 0 \quad \text{as } (n_1, n_2, p) \to \infty. \tag{6}$$

We now develop the multiple testing procedure. Let t be the threshold level such that the null hypotheses H_{0ij} are rejected whenever $|T_{ij}| \geq t$. Then the false discovery proportion (FDP) of the procedure is

$$\frac{\sum_{(i,j)\in\mathcal{H}_0} I\{|T_{ij}| \geq t\}}{\max\left\{\sum_{1\leq i<j\leq p} I\{|T_{ij}| \geq t\}, 1\right\}}.$$

An ideal threshold level for controlling the false discovery proportion at a prespecified level $0 < \alpha < 1$ is

$$\tilde{t}_1 = \inf\left\{0 \leq t \leq 2\sqrt{\log p} : \times \frac{\sum_{(i,j)\in\mathcal{H}_0} I\{|T_{ij}| \geq t\}}{\max\left\{\sum_{1\leq i<j\leq p} I\{|T_{ij}| \geq t\}, 1\right\}} \leq \alpha\right\},$$

where the constraint $0 \leq t \leq 2\sqrt{\log p}$ is used here due to the tail bound (6).

The ideal threshold $\tilde{t}_1$ is unknown and needs to be estimated because it depends on the knowledge of the set of the true null hypotheses $\mathcal{H}_0$. A key step in developing the FDR procedure is the estimation of $G_0(t)$ defined by

$$G_0(t) := \frac{1}{q_0}\sum_{(i,j)\in\mathcal{H}_0} I\{|T_{ij}| \geq t\}, \tag{7}$$

where $q_0 = \mathrm{Card}(\mathcal{H}_0)$. Note that $G_0(t)$ is the true proportion of the nulls falsely rejected by the procedure among all the true nulls at the threshold level t. In some applications such as the PheWAS problem in genomics, the sample sizes can be very large. In this case, it is natural to use the tail of normal distribution $G(t) = 2 - 2\Phi(t)$ to approximate $G_0(t)$. In fact, we have

$$\sup_{0\leq t\leq b_p}\left|\frac{G_0(t)}{G(t)} - 1\right| \to 0 \tag{8}$$

in probability as $(n_1, n_2, p) \to \infty$, where $b_p = \sqrt{4\log p - a_p}$ and $a_p = 2\log(\log p)$. The range $0 \leq t \leq b_p$ is nearly optimal for (8) to hold in the sense that a_p cannot be replaced by any constant in general.

Large-Scale Correlation Tests With Normal Approximation (LCT-N). Let $0 < \alpha < 1$ and define

$$\hat{t} = \inf\left\{0 \leq t \leq b_p : \frac{G(t)(p^2 - p)/2}{\max\left\{\sum_{1\leq i<j\leq p} I\{|T_{ij}| \geq t\}, 1\right\}} \leq \alpha\right\}, \tag{9}$$

Downloaded by [University of Pennsylvania] at 06:26 06 June 2016

where $G(t) = 2 - 2\Phi(t)$. If $\hat{t}$ does not exist, then set $\hat{t} = \sqrt{4\log p}$. We reject H_{0ij} whenever $|T_{ij}| \geq \hat{t}$.

Remark 1. In the above procedure, we use $G(t)$ to estimate $G_0(t)$ when $0 \leq t \leq b_p$. For $t > b_p$, $G(t)$ is not a good approximation of $G_0(t)$ because the convergence rate of $G_0(t)/G(t) \to 1$ is very slow. Furthermore, $G(t)$ is not even a consistent estimator of $G_0(t)$ when $t \geq \sqrt{4\log p - \log(\log p) + O(1)}$ since $p^2 G(t)$ is bounded. Thus, we threshold the test $|T_{ij}|$ with $\sqrt{4\log p}$ directly to control the FDP.

Note that Benjamini–Hochberg (BH) procedure with p-values $G(|T_{ij}|)$ is equivalent to rejecting H_{0ij} if $|T_{ij}| \geq \hat{t}_{\mathrm{BH}}$, where

$$\hat{t}_{\mathrm{BH}} = \inf\left\{t \geq 0 : \frac{G(t)(p^2 - p)/2}{\max\left\{\sum_{1\leq i<j\leq p} I\{|T_{ij}| \geq t\}, 1\right\}} \leq \alpha\right\}.$$

It is important to restrict the range of t to $[0, b_p]$ in (9). The BH procedure uses $G(t)$ to estimate $G_0(t)$ for all $t \geq 0$. As a result, when the number of true alternatives $|\mathcal{H}_0^c|$ is fixed as $p \to \infty$, the BH method is unable to control the FDP with some positive probability, even in the independent case. To see this, we let $H_{01}, \ldots, H_{0m}$ be m null hypotheses and m_1 be the number of true alternatives. Let $\mathrm{FDP}_{\mathrm{BH}}$ be the true FDP of the BH method with independent true p-values and the target $\mathrm{FDR} = \alpha$. If m_1 is fixed as $m \to \infty$, then Proposition 2.1 in Liu and Shao (in press) proved that, for any $0 < \beta < 1$, there exists some constant $c > 0$ such that $\underline{\lim}_{m\to\infty} \mathrm{P}(\mathrm{FDP}_{\mathrm{BH}} \geq \beta) \geq c$.

Remark 2. In the multiple testing procedure given above, we use $p(p-1)/2$ as the estimate for the number q_0 of the true nulls. In many applications, the number of the true significant alternatives is relatively small. In such "sparse" settings, one has $q_0/((p^2 - p)/2) \approx 1$ and the true FDR level of the testing procedure would be close to the nominal level α. See Section 5 for discussions on the numerical performance of the procedure.

The normal approximation is suitable when the sample sizes are large. On the other hand, when the sample sizes are small, the following bootstrap procedure can be used to improve the accuracy of the approximation. Let $\mathcal{X}^* = \{\boldsymbol{X}_k^*, 1 \leq k \leq n_1\}$ and $\mathcal{Y}^* = \{\boldsymbol{Y}_k^*, 1 \leq k \leq n_2\}$ be resamples drawn randomly with replacement from $\{\boldsymbol{X}_k, 1 \leq k \leq n_1\}$ and $\{\boldsymbol{Y}_k, 1 \leq k \leq n_2\}$, respectively. Set $\boldsymbol{X}_k^* = (X_{k,1}^*, \ldots, X_{k,p}^*)'$, $1 \leq k \leq n_1$ and $\boldsymbol{Y}_k^* = (Y_{k,1}^*, \ldots, Y_{k,p}^*)'$, $1 \leq k \leq n_2$. Let

$$\hat{\rho}_{ij1}^* = \frac{\sum_{k=1}^{n_1}\left(X_{k,i}^* - \bar{X}_i^*\right)\left(X_{k,j}^* - \bar{X}_j^*\right)}{\sqrt{\sum_{k=1}^{n_1}\left(X_{k,i}^* - \bar{X}_i^*\right)^2 \sum_{k=1}^{n_1}\left(X_{k,j}^* - \bar{X}_j^*\right)^2}},$$

where $\bar{X}_i^* = \frac{1}{n_1}\sum_{k=1}^{n_1} X_{k,i}^*$ and $\bar{Y}_j^* = \frac{1}{n_2}\sum_{k=1}^{n_2} Y_{k,j}^*$. We define $\hat{\rho}_{ij2}^*$ in a similar way. Let

$$T_{ij}^* = \frac{\hat{\rho}_{ij1}^* - \hat{\rho}_{ij2}^* - (\hat{\rho}_{ij1} - \hat{\rho}_{ij2})}{\sqrt{\frac{\hat{\kappa}_1}{n_1}\left(1 - \hat{\rho}_{ij1}^{*2}\right)^2 + \frac{\hat{\kappa}_2}{n_2}\left(1 - \hat{\rho}_{ij2}^{*2}\right)^2}}. \tag{10}$$

For some given positive integer N, we replicate the above procedure N times independently and obtain $T_{ij,1}^*, \ldots, T_{ij,N}^*$. Let

$$G_{N,n}^*(t) = \frac{2}{N(p^2 - p)}\sum_{k=1}^{N}\sum_{1\leq i<j\leq p} I\left\{\left|T_{ij,k}^*\right| \geq t\right\}.$$

In the bootstrap procedure, we use the conditional (given the data) distribution of $\hat{\rho}_{ij1}^* - \hat{\rho}_{ij2}^* - (\hat{\rho}_{ij1} - \hat{\rho}_{ij2})$ to approximate the null distribution. The signal is not present because the conditional mean of $(\hat{\rho}_{ij1}^* - \hat{\rho}_{ij2}^*) - (\hat{\rho}_{ij1} - \hat{\rho}_{ij2})$ is zero. Proposition 1 in Section 3 shows that, under some regularity conditions,

$$\sup_{0\leq t\leq b_p}\left|\frac{G_{N,n}^*(t)}{G_0(t)} - 1\right| \to 0 \tag{11}$$

in probability. Equation (11) leads us to propose the following multiple testing procedure for correlations.

Large-Scale Correlation Tests With Bootstrap (LCT-B). Let $0 < \alpha < 1$ and define

$$\hat{t} = \inf\left\{0 \leq t \leq b_p : \times \frac{G_{n,N}^*(t)(p^2 - p)/2}{\max\left\{\sum_{1\leq i<j\leq p} I\{|T_{ij}| \geq t\}, 1\right\}} \leq \alpha\right\}. \tag{12}$$

If $\hat{t}$ does not exist, then let $\hat{t} = \sqrt{4\log p}$. We reject H_{0ij} whenever $|T_{ij}| \geq \hat{t}$.

The procedure requires to choose the bootstrap replication time N. The theoretical analysis in Section 3 shows that it can be taken to be any positive integer. The simulation shows that the performance of the procedure is quite insensitive to the choice of N.

3. THEORETICAL PROPERTIES

We now investigate the properties of the multiple testing procedure for correlations introduced in Section 2. It will be shown that, under mild regularity conditions, the procedure controls the FDR asymptotically at any prespecified level $0 < \alpha < 1$. In addition, it also controls the FDP accurately.

Let $\mathrm{FDP}(\hat{t})$ and $\mathrm{FDR}(\hat{t})$ be, respectively, the false discovery proportion and the false discovery rate of the multiple testing procedure defined in (9) and (12),

$$\mathrm{FDP}(\hat{t}) = \frac{\sum_{(i,j)\in\mathcal{H}_0} I\{|T_{ij}| \geq \hat{t}\}}{\max\left(\sum_{1\leq i<j\leq p} I\{|T_{ij}| \geq \hat{t}\}, 1\right)}$$

and

$$\mathrm{FDR}(\hat{t}) = \mathrm{E}(\mathrm{FDP}(\hat{t})).$$

For given positive numbers k_p and s_p, define the collection of symmetric matrices $\mathcal{A}(k_p, s_p)$ by

$$\mathcal{A}(k_p, s_p) = \left\{(a_{ij})_{p\times p} : a_{ij} = a_{ji},\ \mathrm{Card}\{1 \leq i \leq p : \times |a_{ij}| \geq k_p\} \leq s_p,\ \forall\, 1 \leq j \leq p\right\}. \tag{13}$$

We introduce some conditions on the dependence structure of $\boldsymbol{X}$ and $\boldsymbol{Y}$.

(C1) Suppose that, for some $0 < \theta < 1$, $\gamma > 0$, and $0 < \xi < \min\{(1-\theta)/(1+\theta), 1/3\}$, we have $\max_{1\leq i<j\leq p}|\rho_{ijh}| \leq \theta$, $h = 1, 2$, and $\boldsymbol{R}_h \in \mathcal{A}(k_p, s_p)$, $h = 1, 2$, for some $k_p = (\log p)^{-2-\gamma}$ and $s_p = O(p^{\xi})$.
The assumption $\max_{1\leq i<j\leq p}|\rho_{ijh}| \leq \theta$, $h = 1, 2$, is natural as the correlation matrix would be singular if $\max_{1\leq i<j\leq p}|\rho_{ijh}| = 1$. The assumption $\boldsymbol{R}_h \in \mathcal{A}(k_p, s_p)$

means that every variable can be highly correlated (i.e., $\rho_{ijl} \geq k_p$) with at most s_p other variables. The conditions on the correlations in (C1) are quite weak.

Besides the above dependence conditions, we also need an assumption on the covariance structures of $\boldsymbol{X}$ and $\boldsymbol{Y}$. Let $(\sigma_{ij1})_{p\times p}$ and $(\sigma_{ij2})_{p\times p}$ be the covariance matrices of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively.

(C2) Suppose that there exist constants $\kappa_1 \geq \frac{1}{3}$ and $\kappa_2 \geq \frac{1}{3}$ such that for any $i, j, k, l \in \{1, 2, \ldots, p\}$,

$$\begin{aligned} &\mathrm{E}(X_i - \mu_{i1})(X_j - \mu_{j1})(X_k - \mu_{k1})(X_l - \mu_{l1}) \\ &\quad = \kappa_1(\sigma_{ij1}\sigma_{kl1} + \sigma_{ik1}\sigma_{jl1} + \sigma_{il1}\sigma_{jk1}), \\ &\mathrm{E}(Y_i - \mu_{i2})(Y_j - \mu_{j2})(Y_k - \mu_{k2})(Y_l - \mu_{l2}) \\ &\quad = \kappa_2(\sigma_{ij2}\sigma_{kl2} + \sigma_{ik2}\sigma_{jl2} + \sigma_{il2}\sigma_{jk2}). \end{aligned} \tag{14}$$

It is easy to see that $\kappa_1 \equiv \frac{1}{3}\mathrm{E}(X_i - \mu_{i1})^4/[\mathrm{E}(X_i - \mu_{i1})^2]^2$ and $\kappa_2 \equiv \frac{1}{3}\mathrm{E}(Y_i - \mu_{i2})^4/[\mathrm{E}(Y_i - \mu_{i2})^2]^2$. Condition (C2) holds, for example, for all the elliptically contoured distributions (Anderson 2003). Note that the asymptotically normality result (4) holds under Condition (C2) and the null $H_{0ij} : \rho_{ij1} = \rho_{ij2}$.

We also impose exponential type tail conditions on $\boldsymbol{X}$ and $\boldsymbol{Y}$.

(C3) *Exponential Tails:* There exist some constants $\eta > 0$ and $K > 0$ such that

$$\begin{aligned} &\mathrm{E}\exp\left(\eta\left|X_i - \mu_{i1}\right|/\sigma_{ii1}^{1/2}\right) \leq K \quad \text{and} \\ &\mathrm{E}\exp\left(\eta\left|Y_i - \mu_{i2}\right|/\sigma_{ii2}^{1/2}\right) \leq K \quad \text{for all } i. \end{aligned}$$

Let $n = n_1 + n_2$. We first show that under $p \leq n^r$ for some $r > 0$, (C2) and (C3), the distributions of T_{ij} and $G^*_{n,N}(t)$ are asymptotic normally distributed and $G_0(t)$ is well approximated by $G^*_{N,n}(t)$.

Proposition 1. Suppose $p \leq n^r$ for some constant $r > 0$. Under Conditions (C2) and (C3), we have for any $r > 0$ and $b > 0$, as $(n, p) \to \infty$,

$$\sup_{(i,j)\in\mathcal{H}_0} \sup_{0\leq t\leq b\sqrt{\log p}} \left|\frac{\mathrm{P}(|T_{ij}| \geq t)}{2 - 2\Phi(t)} - 1\right| \to 0, \tag{15}$$

$$\sup_{0\leq t\leq b_p} \left|\frac{G^*_{N,n}(t)}{2 - 2\Phi(t)} - 1\right| \to 0, \tag{16}$$

and

$$\sup_{0\leq t\leq b_p} \left|\frac{G^*_{N,n}(t)}{G_0(t)} - 1\right| \to 0 \tag{17}$$

in probability, where Φ is the cumulative distribution function of the standard normal distribution.

We are now ready to state our main results. For ease of notation, we use FDP and FDR to denote FDP($\hat{t}$) and FDR($\hat{t}$), respectively. Recall that $\mathcal{H}_0 = \{(i, j) : 1 \leq i < j \leq p, \rho_{ij1} = \rho_{ij2}\}$ and $q_0 = \mathrm{Card}(\mathcal{H}_0)$. Let $\mathcal{H}_1 = \{(i, j) : 1 \leq i < j \leq p, \rho_{ij1} \neq \rho_{ij2}\}$, $q_1 = \mathrm{Card}(\mathcal{H}_1)$, and $q = (p^2 - p)/2$.

Theorem 1. Assume that $p \leq n^r$ for some $r > 0$ and $q_1 \leq cq$ for some $0 < c < 1$. Under (C1)–(C3),

$$\overline{\lim_{(n,p)\to\infty}} \mathrm{FDR} \leq \alpha, \qquad \lim_{(n,p)\to\infty} \mathrm{P}(\mathrm{FDP} \leq \alpha + \varepsilon) = 1 \tag{18}$$

for any $\varepsilon > 0$.

Theorem 1 shows that the procedures proposed in Section 2 control the FDR and FDP at the desired level asymptotically. It is quite natural to assume $q_1 \leq cq$. For example, if $q_1/q \to 1$, then the number of the zero entries of $\boldsymbol{R}_1 - \boldsymbol{R}_2$ is negligible compared with the number of the nonzero entries and the trivial procedure of rejecting all of the null hypotheses controls FDR at level 0 asymptotically. Note that r in Theorem 1 can be arbitrarily large so that p can be much larger than n ($p \gg n$).

A weak condition to ensure $\hat{t}$ in (9) and (12) exists is Equation (19), which imposes the condition on the number of significant true alternatives. The next theorem shows that, when $\hat{t}$ in (9) and (12) exists, the FDR and FDP tend to $\alpha q_0/q$, where $q = (p^2 - p)/2$.

Theorem 2. Suppose that for some $\delta > 0$,

$$\begin{aligned} &\mathrm{Card}\left\{(i, j) : \frac{|\rho_{ij1} - \rho_{ij2}|}{\sqrt{\kappa_1 + \kappa_2}} \geq 4\sqrt{\frac{(n_1 + n_2)\log p}{n_1 n_2}}\right\} \\ &\quad \geq \left(\frac{1}{\sqrt{8\pi}\alpha} + \delta\right)\sqrt{\log(\log p)}. \end{aligned} \tag{19}$$

Then, under the conditions of Theorem 1, we have

$$\begin{aligned} &\lim_{(n,p)\to\infty} \frac{\mathrm{FDR}}{\alpha q_0/q} = 1 \quad \text{and} \quad \frac{\mathrm{FDP}}{\alpha q_0/q} \to 1 \\ &\quad \text{in probability as} (n, p) \to \infty. \end{aligned} \tag{20}$$

From Theorem 2, we see that if $\boldsymbol{R}_1 - \boldsymbol{R}_2$ is sparse such that the number of nonzero entries is of order $o(p^2)$, then $q_0/q \to 1$. So the FDR tends to α asymptotically. The sparsity assumption is commonly imposed in the literature on estimation of high-dimensional covariance matrix. See, for example, Bickel and Levina (2008) and Cai and Liu (2011).

The multiple testing procedure in this article is related to that in Storey, Taylor, and Siegmund (2004). Let $p_1, \ldots, p_q$ be the p-values. Storey, Taylor, and Siegmund (2004) estimated the number of true null hypotheses q_0 by $\hat{q}_0 = \sum_{k=1}^q I\{p_i \geq \lambda\}/(1 - \lambda)$ with some well-chosen λ and then incorporate $\hat{q}_0$ into the BH method for FDR control. It is possible to use similar idea to estimate q_0 and improve the power in our problem. However, the theoretical results in Storey, Taylor, and Siegmund (2004) are not applicable in our setting. In their Theorem 4, to control FDR, they required $\widehat{\mathrm{FDR}}^{\infty}_{\lambda}(t) < \alpha$, which implies the number of true alternative hypotheses $q_1/q \to \pi_1$ with some positive $\pi_1 > 0$. This excludes the sparse setting $q_1 = o(q)$, which is of particular interest in this article. They assumed that the true p-values are known. This is a very strong condition and will not be satisfied in our setting. Moreover, their dependence condition is imposed on the p-values by assuming the law of large numbers (7) in Storey, Taylor, and Siegmund (2004). Note that we only have the asymptotic distributions $G^*_{n,N}(t)$ and $N(0, 1)$ for the test statistic. Our dependence condition is imposed on the correlation matrix, which is more natural.

Downloaded by [University of Pennsylvania] at 06:26 06 June 2016

Downloaded by [University of Pennsylvania] at 06:26 06 June 2016

4. ONE-SAMPLE CASE

As mentioned in the introduction, multiple testing of correlations in the one-sample case also has important applications. In this section, we consider the one-sample testing problem where we observe a random sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ from a p-dimensional distribution with mean $\boldsymbol{\mu}$ and correlation matrix $\boldsymbol{R} = (\rho_{ij})_{p\times p}$, and wish to simultaneously test the hypotheses

$$H_{0ij}: \quad \rho_{ij} = 0 \quad \text{versus} \quad H_{1ij}: \quad \rho_{ij} \neq 0, \quad \text{for} \quad 1 \le i < j \le p. \tag{21}$$

As mentioned in the introduction, Fisher's z-transformation does not work well for non-Gaussian data in general. Using the same argument as in the two-sample case, we may use the following test statistic for testing each $H_{0ij}: \rho_{ij} = 0$,

$$\frac{|\hat{\rho}_{ij}|}{\sqrt{\frac{\hat{\kappa}}{n}}\left(1-\hat{\rho}_{ij}^2\right)},$$

where $\hat{\kappa} = \frac{1}{3p}\sum_{i=1}^{p} \frac{n\sum_{k=1}^{n}(X_{k,i}-\bar{X}_i)^4}{(\sum_{k=1}^{n}(X_{k,i}-\bar{X}_i)^2)^2}$ is an estimate of $\kappa \equiv \frac{1}{3}\frac{\mathrm{E}(X_i-\mu_i)^4}{[\mathrm{E}(X_i-\mu_i)^2]^2}$. The false discovery rate can be controlled in a similar way as in Section 2 and all the theoretical results in Section 3 also hold in the one-sample case.

There is in fact a different test statistic that requires weaker conditions for the asymptotic normality for the one-sample testing problem (21). Note that (21) is equivalent to

$$H_{0ij}: \quad \sigma_{ij} = 0 \quad \text{versus} \quad H_{1ij}: \quad \sigma_{ij} \neq 0, \quad \text{for} \quad 1 \le i < j \le p. \tag{22}$$

Hence, we propose to use the following normalized sample covariance as the test statistic

$$T_{ij} = \frac{\sum_{k=1}^{n}(X_{ki}-\bar{X}_i)(X_{kj}-\bar{X}_j)}{\sqrt{n\hat{\theta}_{ij}}}, \tag{23}$$

where

$$\hat{\theta}_{ij} = \frac{1}{n}\sum_{k=1}^{n}\left[(X_{ki}-\bar{X}_i)(X_{kj}-\bar{X}_j) - \hat{\sigma}_{ij}\right]^2$$

is a consistent estimator of the variance $\theta_{ij} = \mathrm{var}((X_i - \mu_i)(X_j - \mu_j))$. Note that Cai and Liu (2011) used a similar idea to construct an adaptive thresholding procedure for estimation of sparse covariance matrix. By the central limit theorem and the law of large numbers, we have T_{ij} converging in law to $N(0, 1)$ under the null H_{0ij} and the finite fourth moment condition, $\mathrm{E}(X_i - \mu_i)^4/\sigma_{ii}^2 < \infty$.

When the sample size is large, the normal approximation can be used as in (9). On the other hand, if the sample size is small, then we can use a similar bootstrap method to estimate the proportion of the nulls falsely rejected among all the true nulls,

$$\frac{1}{q_0}\sum_{(i,j)\in\mathcal{H}_0} I\{|T_{ij}| \ge t\},$$

where $\mathcal{H}_0 = \{(i, j) : 1 \le i < j \le p, \rho_{ij} = 0\}$ and $q_0 = \mathrm{Card}(\mathcal{H}_0)$. Let $\mathcal{X}_j^* = \{X_{kj}^*, 1 \le k \le n\}$ be a resample drawn randomly with replacement from $\{X_{kj}, 1 \le k \le n\}$. Let the resamples $\mathcal{X}_j^*$, $1 \le j \le p$, be independent given $\{X_{kj}, 1 \le k \le n, 1 \le j \le p\}$ and set $\boldsymbol{X}_k^* = (X_{k1}^* - \bar{X}_1, \ldots, X_{kp}^* - \bar{X}_p)'$, $1 \le k \le n$. We construct the bootstrap test statistics T_{ij}^* from $\boldsymbol{X}_1^*, \ldots, \boldsymbol{X}_n^*$ as in (23). The above procedure is replicated N times independently, which yields $T_{ij,1}^*, \ldots, T_{ij,N}^*$. Let

$$G_{n,N}^*(t) = \frac{2}{N(p^2-p)}\sum_{k=1}^{N}\sum_{1\le i<j\le p} I\{|T_{ij,k}^*| \ge t\}. \tag{24}$$

Finally, we use the same FDR control procedure as defined in (12).

In the one-sample case, the dependence condition (C1) can be weakened significantly.

(C1*) Suppose that for some $\gamma > 0$ and $\xi > 0$, we have

$$\mathrm{Card}\{(i, j) : 1 \le i < j \le p, \quad |\rho_{ij}| \ge (\log p)^{-2-\gamma}\} \le Cp^2/(\log p)^{1+\xi}.$$

In (C1*), the number of pairs of strong correlated variables can be as large as $p^2/(\log p)^{1+\xi}$. Similar to Theorems 1 and 2 in the two-sample case, we have the following results for the one-sample case. Let $\mathcal{H}_1 = \{(i, j) : 1 \le i < j \le p, \rho_{ij} \neq 0\}$, $q_1 = \mathrm{Card}(\mathcal{H}_1)$, and $q = (p^2 - p)/2$.

Theorem 3. Assume that $p \le n^r$ for some $r > 0$ and $q_1 \le cq$ for some $0 < c < 1$. Suppose the distribution of $\boldsymbol{X}$ satisfies Condition (C1*), (C2), and (C3), then

$$\overline{\lim_{(n,p)\to\infty}} \mathrm{FDR} \le \alpha \quad \text{and} \quad \lim_{(n,p)\to\infty} \mathrm{P}(\mathrm{FDP} \le \alpha + \varepsilon) = 1 \tag{25}$$

for any $\varepsilon > 0$.

Theorem 3 shows that for simultaneous testing of the correlations in the one-sample case, the dependence condition (C1) can be substantially weakened to (C1*). As in Theorem 2, if the number of significant true alternatives is at least of order $\sqrt{\log(\log p)}$, then Theorem 4 shows that the FDR and FDP will converge to $\alpha q_0/q$.

Theorem 4. Suppose that for some $\delta > 0$,

$$\mathrm{Card}\left\{(i, j) : \frac{|\sigma_{ij}|}{\sqrt{\theta_{ij}}} \ge 4\sqrt{\frac{\log p}{n}}\right\} \ge \left(\frac{1}{\sqrt{8\pi}\alpha} + \delta\right)\sqrt{\log(\log p)}.$$

Then, under conditions of Theorem 3,

$$\lim_{(n,p)\to\infty}\frac{\mathrm{FDR}}{\alpha q_0/q} = 1 \quad \text{and} \quad \frac{\mathrm{FDP}}{\alpha q_0/q} \to 1 \quad \text{in probability as } (n, p) \to \infty. \tag{26}$$

5. NUMERICAL STUDY

In this section, we study the numerical properties of the multiple testing procedure defined in Section 2 through the analysis of both simulated and real data. Section 5.1 examines the performance of the multiple testing procedure by simulations. A real data analysis is discussed in Section 5.2.

5.1 Simulation

We study in this section the performance of the testing procedure by a simulation study. In particular, the numerical performance of the proposed procedure is compared with that

Downloaded by [University of Pennsylvania] at 06:26 06 June 2016

of the procedures based on Fisher's z transformation (3) together with the Benjamini–Hochberg method (Benjamini and Hochberg 1995) and Benjamini–Yekutieli method (Benjamini and Yekutieli 2001). We denote these two procedures by F_z-B-H and F_z-B-Y, respectively.

5.1.1 Two-Sample Case: Comparison With F_z-B-H and F_z-B-Y. The sample correlation matrix is invariant to the variances. Hence, we only consider the simulation for $\sigma_{ii1} = \sigma_{ii2} = 1, i = 1, \ldots, p$. Two covariance matrix models are considered.

- *Model 1.* $\boldsymbol{R}_1 = \boldsymbol{\Sigma}_1 = \mathrm{diag}(\mathbf{D}_1, \mathbf{D}_2 \ldots, \mathbf{D}_{p/5})$, where $\mathbf{D}_k$ is a 5×5 matrix with 1 on the diagonal and ρ for all the off-diagonal entries. $\boldsymbol{R}_2 = \boldsymbol{\Sigma}_2 = \mathrm{diag}(\boldsymbol{I}, \boldsymbol{A})$, where $\boldsymbol{I}$ is a $(p/4) \times (p/4)$ identity matrix and $\boldsymbol{A} = \mathrm{diag}(\mathbf{D}_{p/20+1}, \ldots, \mathbf{D}_{p/5})$.
- *Model 2.* $\boldsymbol{R}_1 = \boldsymbol{\Sigma}_1 = \mathrm{diag}(\hat{\mathbf{D}}_1, \hat{\mathbf{D}}_2 \ldots, \hat{\mathbf{D}}_{[p/m_2]}, \hat{\boldsymbol{I}})$, where $\hat{\mathbf{D}}_k$ is an $m_1 \times m_1$ matrix with 1 on the diagonal and ρ for all the off-diagonal entries. $\hat{\boldsymbol{I}}$ is a $(p - m_1[p/m_1]) \times (p - m_1[p/m_1])$ identity matrix. $\boldsymbol{R}_2 = \boldsymbol{\Sigma}_2 = \mathrm{diag}(\tilde{\mathbf{D}}_1, \tilde{\mathbf{D}}_2 \ldots, \tilde{\mathbf{D}}_{[p/m_2]}, \tilde{\boldsymbol{I}})$, where $\tilde{\mathbf{D}}_k$ is an $m_2 \times m_2$ matrix with 1 on the diagonal and ρ for all the off-diagonal entries. $\tilde{\boldsymbol{I}}$ is a $(p - m_2[p/m_2]) \times (p - m_2[p/m_2])$ identity matrix.

The value of ρ will be specified in different distributions for the population. We will take $(m_1, m_2) = (80, 40)$ in Model 2 to consider the strong correlation case. The following four distributions are considered.

- *Normal mixture distribution.* $\boldsymbol{X} = U_1 \boldsymbol{Z}_1$ and $\boldsymbol{Y} = U_2 \boldsymbol{Z}_2$, where U_1 and U_2 are independent uniform random variables on $(0, 1)$ and $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are independent random vectors with distributions $N(0, \boldsymbol{\Sigma}_1)$ and $N(0, \boldsymbol{\Sigma}_2)$, respectively. Let $\rho = 0.8$.
- *Normal distribution.* $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent random vectors with distributions $N(0, \boldsymbol{\Sigma}_1)$ and $N(0, \boldsymbol{\Sigma}_2)$, respectively. Let $\rho = 0.6$.
- *t distribution.* $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are independent random vectors with iid components having t_6 distributions. Let $\boldsymbol{X} = \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{Z}_1$ and $\boldsymbol{Y} = \boldsymbol{\Sigma}_2^{1/2} \boldsymbol{Z}_2$ with $\rho = 0.6$.
- *Exponential distribution.* $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are independent random vectors with iid components having exponential distributions with parameter 1. Let $\boldsymbol{X} = \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{Z}_1$ and $\boldsymbol{Y} = \boldsymbol{\Sigma}_2^{1/2} \boldsymbol{Z}_2$ with $\rho = 0.6$.

The normal mixture distribution ($\kappa_1 \neq 1$ and $\kappa_2 \neq 1$) allows us to check the influence of nonnormality of the data on the procedures based on Fisher's z transformation. We also give the comparison between our procedure and the one based on Fisher's z transformation when the distribution is truly multivariate normal distributed.

Note that the normal mixture distribution and the normal distribution satisfy the elliptically contoured distributions condition. On the other hand, the t distribution and exponential distribution generated by the above way do not satisfy (C2) and the t distribution does not satisfy (C3) either. So it allows us to check the influence of conditions (C2) and (C3) on our method.

In the simulation, we generate two groups of independent samples from $\boldsymbol{X}$ and $\boldsymbol{Y}$. Let the sample sizes $n_1 = n_2 = 50$ and $n_1 = n_2 = 100$ and let the dimension $p = 250$, 500, and 1000. The number of the bootstrap resamples is taken to be $N = 50$ and the nominal false discovery rate $\alpha = 0.2$. Based on 100 replications, we calculate the average empirical false discovery rates

$$\text{Average}\left\{\frac{\sum_{(i,j)\in\mathcal{H}_0} I\{|T_{ij}| \geq \hat{t}\}}{\max\left\{\sum_{1\leq i<j\leq p} I\{|T_{ij}| \geq \hat{t}\}, 1\right\}}\right\}$$

and the average empirical powers

$$\text{Average}\left\{\frac{\sum_{(i,j)\in\mathcal{H}_1} I\{|T_{ij}| \geq \hat{t}\}}{\sum_{1\leq i<j\leq p} I\{\rho_{ij1} \neq \rho_{ij2}\}}\right\},$$

where $\mathcal{H}_1 = \{(i, j) : 1 \leq i < j \leq p, \rho_{ij1} \neq \rho_{ij2}\}$.

The simulation results for Model 1 in terms of the empirical FDR are summarized in Table 1 and the results on the empirical powers are given in Table 2. It can be seen from the two tables that, for the normal mixture distribution, the proposed procedure with bootstrap approximation (LCT-B) has significant advantages on controlling the FDR. It performs much better than the proposed procedure with normal approximation (LCT-N) when the sample size is small. Note that the performance of LCT-N becomes better as n increases. Both procedures in (9) and (12) outperform the one based on Fisher's z transformation (3) on FDR control. For the multivariate normal distribution, our methods have more power than F_z-B-H and F_z-B-Y. The latter method is also quite conservative. For the other two distributions that do not satisfy (C2), the empirical FDRs of F_z-B-H are larger

Table 1. Empirical false discovery rates ($\alpha = 0.2$), Model 1

		Normal mixture		N(0,1)	
$p \backslash n_1 = n_2$		50	100	50	100
250	F_z-B-H	0.9519	0.9479	0.3084	0.2511
	F_z-B-Y	0.6400	0.6136	0.0411	0.0256
	LCT-B	0.2267	0.1096	0.1068	0.1045
	LCT-N	0.4897	0.3065	0.3270	0.2450
500	F_z-B-H	0.9750	0.9721	0.3253	0.2511
	F_z-B-Y	0.7293	0.6714	0.0341	0.0249
	LCT-B	0.2368	0.0935	0.1039	0.0834
	LCT-N	0.5137	0.2977	0.3204	0.2334
1000	F_z-B-H	0.9871	0.9861	0.3669	0.2594
	F_z-B-Y	0.8052	0.7629	0.0428	0.0226
	LCT-B	0.2420	0.0620	0.1012	0.0567
	LCT-N	0.5479	0.2804	0.3304	0.2227
		t_6		Exp(1)	
250	F_z-B-H	0.3204	0.2473	0.3738	0.2846
	F_z-B-Y	0.0430	0.0278	0.0693	0.0351
	LCT-B	0.0703	0.0890	0.0943	0.0817
	LCT-N	0.0903	0.0323	0.0721	0.0097
500	F_z-B-H	0.3487	0.2530	0.4328	0.3040
	F_z-B-Y	0.0384	0.0255	0.0768	0.0345
	LCT-B	0.0612	0.0639	0.0915	0.0568
	LCT-N	0.0868	0.0228	0.0845	0.0065
1000	F_z-B-H	0.3870	0.2711	0.4975	0.3309
	F_z-B-Y	0.0523	0.0261	0.0958	0.0396
	LCT-B	0.0565	0.0434	0.1050	0.0355
	LCT-N	0.0907	0.0165	0.1018	0.0046

Table 2. Empirical powers ($\alpha = 0.2$), Model 1

$p \backslash n_1 = n_2$		Normal mixture 50	Normal mixture 100	N(0,1) 50	N(0,1) 100
250	F_z-B-H	0.9889	1.0000	0.5375	0.9405
	F_z-B-Y	0.9113	0.8782	0.2125	0.8072
	LCT-B	0.9245	0.9968	0.6445	0.9712
	LCT-N	0.9729	0.9995	0.7798	0.9792
500	F_z-B-H	0.9906	1.0000	0.4433	0.9247
	F_z-B-Y	0.8945	0.9985	0.1521	0.7576
	LCT-B	0.9074	0.9944	0.5741	0.9572
	LCT-N	0.9671	0.9996	0.7268	0.9751
1000	F_z-B-H	0.9894	1.0000	0.3593	0.8866
	F_z-B-Y	0.8768	0.9977	0.1027	0.6876
	LCT-B	0.8920	0.9979	0.5048	0.9381
	LCT-N	0.9583	0.9992	0.6784	0.9646
		t_6		Exp(1)	
250	F_z-B-H	0.5465	0.9477	0.5981	0.9565
	F_z-B-Y	0.2329	0.8252	0.2762	0.8432
	LCT-B	0.6397	0.9647	0.5593	0.9525
	LCT-N	0.6562	0.9462	0.5357	0.8806
500	F_z-B-H	0.4679	0.9228	0.5104	0.9273
	F_z-B-Y	0.1684	0.7645	0.1884	0.7763
	LCT-B	0.5536	0.9490	0.4781	0.9206
	LCT-N	0.6047	0.9244	0.4656	0.8300
1000	F_z-B-H	0.4717	0.8925	0.4405	0.9049
	F_z-B-Y	0.1134	0.6965	0.1334	0.7208
	LCT-B	0.4699	0.9273	0.4118	0.8754
	LCT-N	0.5373	0.8984	0.4067	0.7873

Table 3. Empirical false discovery rates ($\alpha = 0.2$), Model 2

$p \backslash n_1 = n_2$		Normal mixture 50	Normal mixture 100	N(0,1) 50	N(0,1) 100
250	F_z-B-H	0.4582	0.4476	0.1944	0.1797
	F_z-B-Y	0.1406	0.1356	0.0212	0.0189
	LCT-B	0.2095	0.2063	0.1845	0.1824
	LCT-N	0.2934	0.2433	0.2454	0.2163
500	F_z-B-H	0.6264	0.5993	0.2226	0.1968
	F_z-B-Y	0.2187	0.1924	0.0239	0.0174
	LCT-B	0.1722	0.1951	0.1612	0.1836
	LCT-N	0.3309	0.2694	0.2607	0.2214
1000	F_z-B-H	0.7275	0.7174	0.2436	0.2131
	F_z-B-Y	0.2700	0.2456	0.0245	0.0177
	LCT-B	0.1349	0.1632	0.1222	0.1600
	LCT-N	0.3297	0.2698	0.2626	0.2278
		t_6		Exp(1)	
250	F_z-B-H	0.1976	0.1753	0.2058	0.2051
	F_z-B-Y	0.0242	0.0171	0.0257	0.0253
	LCT-B	0.1928	0.1924	0.2111	0.2039
	LCT-N	0.1924	0.1398	0.1497	0.1100
500	F_z-B-H	0.2340	0.2067	0.2372	0.2163
	F_z-B-Y	0.0253	0.0201	0.0282	0.0215
	LCT-B	0.1694	0.1745	0.1699	0.1945
	LCT-N	0.1883	0.1377	0.1313	0.0810
1000	F_z-B-H	0.2425	0.2171	0.2597	0.2255
	F_z-B-Y	0.0234	0.0181	0.0275	0.0201
	LCT-B	0.1235	0.1675	0.1343	0.1667
	LCT-N	0.1644	0.1211	0.1101	0.0640

than α while the empirical FDRs of our method are smaller than α. However, the powers of our method are quite close to those of F_z-B-H. Note that F_z-B-Y has the lowest powers although it is able to control the FDR.

The correlation in Model 2 is much stronger than that in Model 1 and the number of true alternatives is also larger. As we can see from Tables 3 and 4, our method can still control the FDR efficiently and the powers are comparable to those of F_z-B-H and much higher than those of F_z-B-Y. As the numerical results for Model 1, the empirical FDRs of F_z-B-H are much larger than α for the normal mixture distribution. The performance of F_z-B-H is improved on the other three distributions although its empirical FDRs are somewhat higher than α when $p = 1000$ and $n = 50$.

5.1.2 One Sample Case. To examine the performance of our method in the one-sample case, we consider the following model.

- *Model 3.* $\boldsymbol{R} = \boldsymbol{\Sigma} = \mathrm{diag}(\mathbf{D}_1, \mathbf{D}_2 \ldots, \mathbf{D}_{p/5})$, where $\mathbf{D}_k$ is a 5×5 matrix with 1 on the diagonal and ρ for all the off-diagonal entries.

We consider four types of distributions and ρ is taken to be the same values as in the two-sample case. In the simulation, we let $n = 50$ and $p = 500$. The number of the bootstrap resamples is taken to be $N = 50$ and the nominal false discovery rate $\alpha = 0.2$. The empirical FDRs of three methods based on 100 replications are summarized in Table 5. As we can see from Table 5, the empirical FDRs of F_z-B-H are higher than α, especially for the normal mixture distribution. F_z-B-Y is also unable to control the FDR for the normal mixture distribution. Our method controls FDR quite well for all four distributions. Even when (C2) is not satisfied, our method can still control FDR efficiently.

We now carry out a simulation study to verify that the FDP control in the one-sample case can get benefit from the correlation. Consider the following matrix model.

- *Model 4.* $\boldsymbol{\Sigma} = \mathrm{diag}(\mathbf{D}_1, \mathbf{D}_2 \ldots, \mathbf{D}_k, \boldsymbol{I})$, where $\mathbf{D}_k$ is a 5×5 matrix with 1 on the diagonal and 0.6 for all the off-diagonal entries.

We take $k = 1, 5, 10, 20, 40, 80$ such that the correlation increases as k grows. Let $\boldsymbol{X} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{Z}$, where $\boldsymbol{Z}$ is the standard normal random vector. We take $n = 50$ and $p = 500$. The procedure in Section 4 with the bootstrap approximation is used in the simulation. To evaluate the performance of the FDP control, we use the l_2 distance $\mathrm{SD} := \sqrt{\sum_{i=1}^{100}(\mathrm{FDP}_i - \alpha q_0/q)^2/100}$, where FDP_i is the FDP in the ith replication. As we can see from Table 6, the distance between FDP and $\alpha q_0/q$ becomes small as k increases.

5.2 Real Data Analysis

Kostka and Spang (2004), Carter et al. (2004), and Lai et al. (2004) studied gene–gene coexpression patterns based on cancer gene expression datasets. Their analyses showed that several

Downloaded by [University of Pennsylvania] at 06:26 06 June 2016

Downloaded by [University of Pennsylvania] at 06:26 06 June 2016

Table 4. Empirical powers ($\alpha = 0.2$), Model 2

		Normal mixture		N(0,1)	
$p \backslash n_1 = n_2$		50	100	50	100
250	F_z-B-H	0.9970	1.0000	0.9208	0.9963
	F_z-B-Y	0.9730	0.9997	0.6640	0.9658
	LCT-B	0.9879	1.0000	0.9096	0.9977
	LCT-N	0.9932	0.9999	0.9381	0.9978
500	F_z-B-H	0.9955	1.0000	0.8637	0.9941
	F_z-B-Y	0.9658	0.9996	0.5482	0.9506
	LCT-B	0.9819	0.9999	0.8482	0.9943
	LCT-N	0.9901	0.9999	0.8954	0.9967
1000	F_z-B-H	0.9936	1.0000	0.8037	0.9900
	F_z-B-Y	0.9498	0.9996	0.4479	0.9257
	LCT-B	0.9753	0.9999	0.7920	0.9926
	LCT-N	0.9836	0.9999	0.8492	0.9947
		t_6		Exp(1)	
250	F_z-B-H	0.9136	0.9965	0.9165	0.9971
	F_z-B-Y	0.6548	0.9678	0.6861	0.9704
	LCT-B	0.9047	0.9972	0.8710	0.9957
	LCT-N	0.9013	0.9957	0.8607	0.9920
500	F_z-B-H	0.8576	0.9924	0.8641	0.9929
	F_z-B-Y	0.5498	0.9430	0.5771	0.9467
	LCT-B	0.8441	0.9946	0.8000	0.9912
	LCT-N	0.8394	0.9907	0.7774	0.9813
1000	F_z-B-H	0.8015	0.9881	0.8105	0.9875
	F_z-B-Y	0.4639	0.9232	0.4890	0.9196
	LCT-B	0.7655	0.9886	0.7254	0.9827
	LCT-N	0.7857	0.9856	0.7110	0.9679

transcriptional regulators, which are known to be involved in cancer, had no significant changes in their mean expression levels but were highly differentially coexpressed. As pointed out by de la Fuente (2010), these results strongly indicated that, besides differential mean expressions, coexpression changes are also highly relevant when comparing gene expression datasets.

In this section, we illustrate the proposed multiple testing procedure with an application to the detection of the changes in coexpression patterns between gene expression levels using a prostate cancer dataset (Singh et al. 2002). The dataset is available at *http://www.broad.mit.edu/cgi-bin/cancer/datasets.cgi*.

This dataset consists of two classes of gene expression data that came from 52 prostate tumor patients and 50 prostate normal patients. There are a total of 12,600 genes. We first choose 500 genes with the smallest absolute values of the two-sample t test statistics for the comparison of the means

$$t_i = |\bar{X}_i - \bar{Y}_i| \Big/ \sqrt{\hat{s}_{1i}^2/n_1 + \hat{s}_{2i}^2/n_2},$$

Table 5. Empirical FDRs for one-sample tests ($\alpha = 0.2$), Model 3

	U*N(0,1)	N(0,1)	$t(6)$	Exp(1)
F_z-B-H	0.9093	0.2923	0.3019	0.3601
F_z-B-Y	0.5304	0.0339	0.0361	0.0714
LCT-B	0.1733	0.1895	0.1859	0.1769

Table 6. Empirical distance between FDP and $\alpha q_0/q$ ($\alpha = 0.2$)

k	1	5	10	20	40	80
SD	0.3426	0.1784	0.0836	0.0433	0.0281	0.0221

where $\hat{s}_{1i}^2$ and $\hat{s}_{2i}^2$ are the sample variances of the ith gene. All of the p-values $\mathrm{P}(|N(0,1)| \geq |t_i|)$ of 500 genes are greater than 0.87; see Figure 1(a). Hence, it is very likely that all of the 500 genes are not differentially expressed in the means. The proposed multiple testing procedure is applied to investigate whether there are differentially coexpressed gene pairs between these 500 genes. As in Kostka and Spang (2004), Carter et al. (2004), and Lai et al. (2004), the aim of this analysis is to verify the phenomenon that additional information can be gained from the coexpressions even when the genes are not differentially expressed in the means.

Let r_{ij}^N (r_{ij}^T) denote the Pearson correlation coefficient between the expression levels of gene i and gene j of the prostate normal (tumor) patients. We wish to test the hypotheses H_{0ij} : $r_{ij}^N = r_{ij}^T, 1 \leq i < j \leq 500$. The pair of genes i and j is identified to be differentially coexpressed if the hypothesis H_{0ij} is rejected. See de la Fuente (2010). We compare the performance between our procedure (the number of the bootstrap resamples $N = 50$) and those based on Fisher's z transformation with the nominal FDR level $\alpha = 0.05$. Our procedure (Figure 1(b)) identifies 1341 pairs of coexpression genes and 1.07% nonzero entries of the coexpression matrix (estimation of support of $\boldsymbol{R}_1 - \boldsymbol{R}_2$). As noted by Yeung and Tegne (2002), gene regulatory networks in most biological systems are expected to be sparse. Our method thus leads to a clear and easily interpretable coexpression network. In comparison, F_z-B-H and F_z-B-Y identify, respectively, 26,373 (21.14%) and 13,794 (11.06%) pairs of coexpression genes and the estimates of the support of $\boldsymbol{R}_1 - \boldsymbol{R}_2$ are very dense and difficult to interpret (Figure 1(c) and 1(d)). This is likely due to the nonnormality of the dataset so that (3) fails to hold. As a result, the true FDR level of F_z-B-H and F_z-B-Y may be much larger than the nominal level, which leads to the large number of rejections.

6. DISCUSSION

In this article, we introduced a large-scale multiple testing procedure for correlations and showed that the procedure performs well both theoretically and numerically under certain regularity conditions. The method can also be used for testing the cross-correlations, and some of the conditions can be further weakened. We discuss in the section some of the extensions and the connections to other work.

6.1 Multiple Testing of Cross-Correlations

In some applications, it is of interest to carry out multiple testing of cross-correlations between two high-dimensional random vectors, which is closely related to the one-sample case considered in this article. Let $\boldsymbol{X} = (X_1, \ldots, X_{p_1})'$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_{p_2})'$ be two random vectors with dimension p_1 and p_2, respectively. We consider multiple correlation tests between X_i and Y_j

$$H_{0ij} : \mathrm{cov}(X_i, Y_j) = 0 \quad \text{versus} \quad H_{1ij} : \mathrm{cov}(X_i, Y_j) \neq 0$$

Downloaded by [University of Pennsylvania] at 06:26 06 June 2016

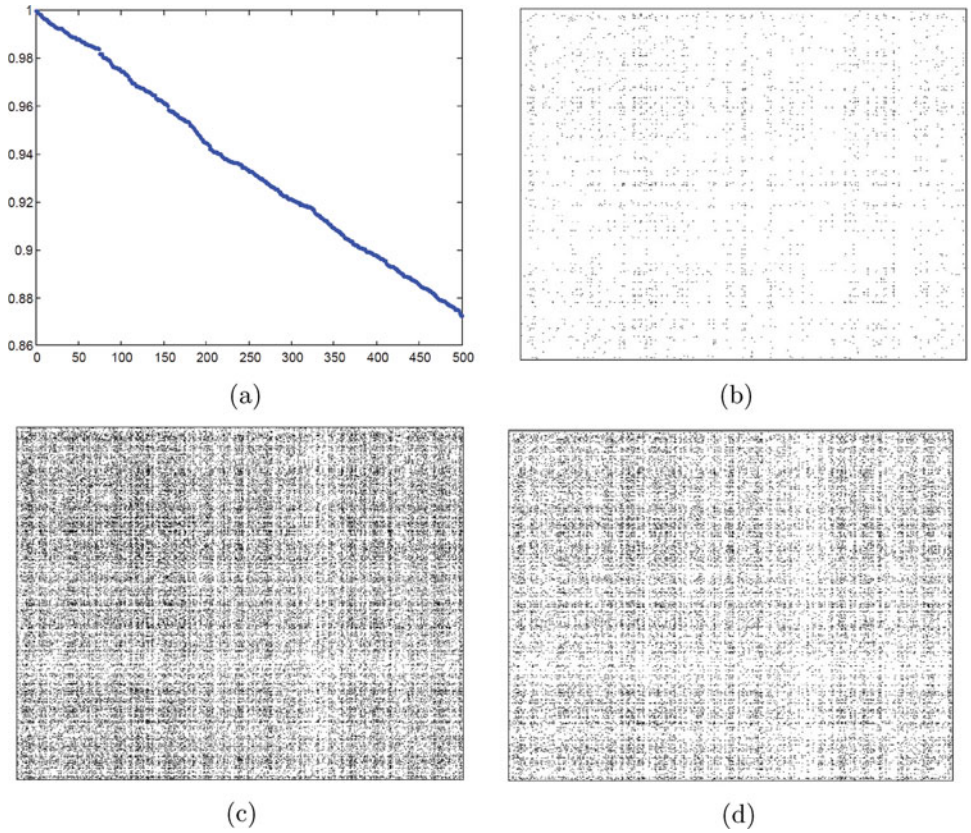

Figure 1. (a) p-Values of 500 genes. (b) Coexpression matrix (C-L). (c) Coexpression matrix (F_z-B-H). (d) Coexpression matrix (F_z-B-Y).

for $1 \le i \le p_1$ and $1 \le j \le p_2$. We can construct similar test statistics

$$T_{ij} = \frac{\sum_{k=1}^{n}(X_{ki} - \bar{X}_i)(Y_{kj} - \bar{Y}_j)}{\sqrt{n\hat{\theta}_{ij}}},$$

where

$$\hat{\theta}_{ij} = \frac{1}{n}\sum_{k=1}^{n}[(X_{ki} - \bar{X}_i)(Y_{kj} - \bar{Y}_j) - \hat{\sigma}_{ijXY}]^2,$$
$$\hat{\sigma}_{ijXY} = \frac{1}{n}\sum_{k=1}^{n}(X_{ki} - \bar{X}_i)(Y_{kj} - \bar{Y}_j).$$

The normal distribution can be used to approximate the null distribution of T_{ij} when the sample size is large. If the sample size is small, we can use $G^*_{n,N}(t)$ to approximate the null distribution of T_{ij}, where

$$G^*_{n,N}(t) = \frac{1}{Np_1p_2}\sum_{k=1}^{N}\sum_{i=1}^{p_1}\sum_{j=1}^{p_2} I\{|T^*_{ij,k}| \ge t\}.$$

Here $T^*_{ij,k}$ are constructed by the bootstrap method as in (24). The multiple testing procedure is as follows.

FDR control procedure. Let $0 < \alpha < 1$ and define

$$\hat{t} = \inf\left\{0 \le t \le b_p : \frac{G^*_{n,N}(t)p_1p_2}{\max\left\{\sum_{i=1}^{p_1}\sum_{j=1}^{p_2} I\{|T_{ij}| \ge t\}, 1\right\}} \le \alpha\right\}.$$

If $\hat{t}$ does not exist, then let $\hat{t} = \sqrt{2\log(p_1p_2)}$. We reject H_{0ij} whenever $|T_{ij}| \ge \hat{t}$.

Let $\mathcal{H}_0 = \{(i,j) : \mathrm{cov}(X_i, Y_j) = 0\}$ and $\mathcal{H}_1 = \{(i,j) : \mathrm{cov}(X_i, Y_j) \ne 0\}$. We assume the following condition holds for $\boldsymbol{X}$ and $\boldsymbol{Y}$.

(C4). For any $\mathcal{A} = \{i, j, k, l\}$, if $(i,j) \in \mathcal{H}_0$ and $(k,l) \in \mathcal{H}_0$, then

$$\begin{aligned}&\mathrm{E}[(X_i - \mathrm{E}X_i)(Y_j - \mathrm{E}Y_j)(X_k - \mathrm{E}X_k)(Y_l - \mathrm{E}Y_l)]\\ &\quad = \tau_{\mathcal{A}}\mathrm{E}[(X_i - \mathrm{E}X_i)(X_k - \mathrm{E}X_k)]\mathrm{E}[(Y_j - \mathrm{E}Y_j)(Y_l - \mathrm{E}Y_l)]\end{aligned}$$

for some positive constant $\tau_{\mathcal{A}}$.

Let $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ be the correlation matrices of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. Denote $p = p_1 + p_2$, $q = p_1p_2$, $q_0 = \mathrm{Card}(\mathcal{H}_0)$, and $q_1 = \mathrm{Card}(\mathcal{H}_1)$. Suppose that $p_1 \asymp p_2$. Then the following theorem holds.

Theorem 5. Assume that $p \le n^r$ for some $r > 0$ and $q_1 \le cq$ for some $0 < c < 1$. Under (C1), (C3), and (C4),

$$\overline{\lim_{(n,p)\to\infty}}\ \mathrm{FDR} \le \alpha, \qquad \lim_{(n,p)\to\infty}\mathrm{P}(\mathrm{FDP} \le \alpha + \varepsilon) = 1$$

for any $\varepsilon > 0$. Furthermore, if

$$\begin{aligned}&\mathrm{Card}\left\{(i,j) : \frac{|\mathrm{cov}(X_i, Y_j)|}{\sqrt{\theta_{ij,XY}}} \ge 4\sqrt{\frac{\log p}{n}}\right\}\\ &\quad \ge \left(\frac{1}{\sqrt{8\pi\alpha}} + \delta\right)\sqrt{\log(\log p)},\end{aligned}$$

Downloaded by [University of Pennsylvania] at 06:26 06 June 2016

then

$$\lim_{(n,p)\to\infty} \frac{\text{FDR}}{\alpha q_0/q} = 1 \quad \text{and}$$
$$\frac{\text{FDP}}{\alpha q_0/q} \to 1 \qquad \text{in probability as } (n, p) \to \infty,$$

where $\theta_{ij,XY} = \text{var}[(X_i - \text{E}X_i)(Y_j - \text{E}Y_j)]$.

6.2 Relations to Owen (2005)

A related work to the one-sample correlation test is Owen (2005), which studied the variance of the number of false discoveries in the tests on the correlations between a single response and p covariates. It was shown that the correlation would greatly affect the variance of the number of false discoveries. The goal in our article is different from that in Owen (2005). Here, we study the FDR control on the correlation tests between all pairs of variables. In our problem, the impact of correlation is much less serious and is even beneficial to the FDP control under the sparse setting (C1*). To see this, set $\mathcal{N} = \{(i, j) : 1 \le i < j \le p, |\rho_{ij}| \ge (\log p)^{-2-\gamma}\}$ for some $\gamma > 0$. In other words, $\mathcal{N}$ denotes the pairs with strong correlations. Suppose that $\text{Card}(\mathcal{N}) = p^\tau$ for some $0 < \tau < 2$. The larger τ indicates the stronger correlations among the variables. It follows from the proof of Theorem 2 that $\text{P}(0 \le \hat{t} \le \sqrt{(4-2\tau)\log p}) \to 1$. By the proof in the supplementary materials, we can see that the difference $\text{FDP} - \alpha q_0/q$ depends on the accuracy of the approximation

$$\sup_{0\le t\le\sqrt{(4-2\tau)\log p}} \left| \frac{\sum_{(i,j)\in\mathcal{H}_0} I\{|T_{ij}| \ge t\}}{|\mathcal{H}_0| G^*_{n,N}(t)} - 1 \right|.$$

Generally, a larger τ provides a better approximation because the range $0 \le t \le \sqrt{(4-2\tau)\log p}$ becomes smaller and $|\mathcal{H}_0| G^*_{n,N}(\sqrt{(4-2\tau)\log p})$ becomes larger. Hence, as τ increases, the FDP is better controlled. Simulation results in Section 5.1.2 also support this observation.

6.3 Relax the Conditions

In Sections 2 and 3, we require the distributions to satisfy the moment condition (C2), which is essential for the validity of the testing procedure. An important example is the class of the elliptically contoured distributions. This is clearly a much larger class than the class of multivariate normal distributions. However, in real applications, (C2) can still be violated. It is desirable to develop test statistics that can be used for more general distributions. To this end, we introduce the following test statistics that do not need the condition (C2).

Let $X'_{ki} = (X_{ki} - \mu_i)/\sigma_{ii1}^{1/2}$. It can be proved that, under the finite 4th moment condition $\text{E}(X_{ki} - \mu_{i1})^4/\sigma_{ii1}^2 < \infty$,

$$2\sqrt{\frac{n_1}{\theta_{ij1}}}(\hat{\rho}_{ij1} - \rho_{ij1}) \to N(0, 1), \tag{27}$$

where $i \ne j$ and

$$\theta_{ij1} = \frac{1}{n_1}\sum_{k=1}^{n_1}\left(2X'_{ki}X'_{kj} - \hat{\rho}_{ij1}X'^2_{ki} - \hat{\rho}_{ij1}X'^2_{kj}\right)^2.$$

We can estimate μ_i and σ_{ii1} in θ_{ij1} by their sample versions. Let $\hat{X}_{ki} = (X_{ki} - \bar{X}_i)/\hat{\sigma}_{ii1}^{1/2}$, where $\hat{\sigma}_{ii1} = \frac{1}{n_1}\sum_{k=1}^{n_1}(X_{ki} - \bar{X}_i)^2$, and let

$$\hat{\theta}_{ij1} = \frac{1}{n_1}\sum_{k=1}^{n_1}(2\hat{X}_{ki}\hat{X}_{kj} - \hat{\rho}_{ij1}\hat{X}^2_{ki} - \hat{\rho}_{ij1}\hat{X}^2_{kj})^2.$$

$\hat{\theta}_{ij2}$ is defined in the same way by replacing X with Y. So the test statistic

$$T'_{ij} = \frac{2(\hat{\rho}_{ij1} - \hat{\rho}_{ij2})}{\sqrt{\hat{\theta}_{ij1}/n_1 + \hat{\theta}_{ij2}/n_2}} \tag{28}$$

can be used to test the individual hypothesis $H_{0ij} : \rho_{ij1} = \rho_{ij2}$. We have the following proposition.

Proposition 2.

1. Suppose that $\text{E}(X_{ki} - \mu_{i1})^4/\sigma_{ii1}^2 < \infty$ and $\text{E}(Y_{ki} - \mu_{i2})^4/\sigma_{ii2}^2 < \infty$. Under the null hypothesis $H_{0ij} : \rho_{ij1} = \rho_{ij2}$, we have $T'_{ij} \Rightarrow N(0, 1)$.
2. Suppose that $p \le n^r$ for some $r > 0$ and (C3) holds. For any $b > 0$, we have

$$\sup_{(i,j)\in\mathcal{H}_0} \sup_{0\le t\le b\sqrt{\log p}} \left| \frac{\text{P}(|T'_{ij}| \ge t)}{2 - 2\Phi(t)} - 1 \right| \to 0,$$

Proposition 2 can be used to establish the FDR control result for multiple tests (2) by assuming some dependence condition between the test statistics T'_{ij}. However, we should point out that, although T'_{ij} does not require (C2), numerical results show that it is less powerful than the test statistic T_{ij} in Section 2.

SUPPLEMENTARY MATERIALS

In the supplement, we prove the main theorems and the key technical result, Proposition 1.

[Received February 2013. Revised November 2014.]

REFERENCES

Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis* (3rd ed.), Hoboken, NJ: Wiley. [230,233]

Benjamini, Y., and Hochberg, Y. (1995), "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society*, Series B, 57, 289–300. [229,235]

Benjamini, Y., and Yekutieli, D. (2001), "The Control of the False Discovery Rate in Multiple Testing Under Dependency," *Annals of Statistics*, 29, 1165–1188. [229,235]

Bickel, P., and Levina, E. (2008), "Covariance Regularization by Thresholding," *Annals of Statistics*, 36, 2577–2604. [233]

Cai, T. T., and Liu, W. D. (2011), "Adaptive Thresholding for Sparse Covariance Matrix Estimation," *Journal of the American Statistical Association*, 106, 672–684. [233,234]

Carter, S. L., Brechbühler, C. M., Griffin, M., and Bond, A. T. (2004), "Gene Co-Expression Network Topology Provides a Framework for Molecular Characterization of Cellular State," *Bioinformatics*, 20, 2242–2250. [229,236,237]

de la Fuente, A. (2010), "From 'Differential Expression' to 'Differential Networking'—Identification of Dysfunctional Regulatory Networks in Diseases," *Trends in Genetics*, 26, 326–333. [229,237]

Dubois, P. C. A., et al. (2010), "Multiple Common Variants for Celiac Disease Influencing Immune Gene Expression," *Nature Genetics*, 42, 295–302. [229]

Efron, B. (2004), "Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis," *Journal of the American Statistical Association*, 99, 96–104. [229]

——— (2007), "Correlation and Large-Scale Simultaneous Significance Testing," *Journal of the American Statistical Association*, 102, 93–103. [229]

Farcomeni, A. (2007), "Some Results on the Control of the False Discovery Rate Under Dependence," *Scandinavian Journal of Statistics*, 34, 275–297. [229]

Genovese, C., and Wasserman, L. (2004), "A Stochastic Process Approach to False Discovery Control," *Annals of Statistics*, 32, 1035–1061. [229]

Hawkins, D. L. (1989), "Using *U* Statistics to Derive the Asymptotic Distribution of Fisher's *Z* Statistic," *Journal of the American Statistical Association*, 43, 235–237. [230]

Hirai, M. Y., Sugiyama, K., Sawada, Y., Tohge, T., Obayashi, T., Suzuki, A., Araki, R., Sakurai, N., Suzuki, H., Aoki, K., Goda, H., Nishizawa, O. I., Shibata,D., and Saito, K. (2007), "Omics-Based Identification of Arabidopsis Myb Transcription Factors Regulating Aliphatic Glucosinolate Biosynthesis," *Proceedings of the National Academy of Sciences*, 104, 6478–6483. [229]

Kostka, D., and Spang, R. (2004), "Finding Disease Specific Alterations in the Co-Expression of Genes," *Bioinformatics*, 20, 194–199. [229,236,237]

Lai, Y., Wu, B., Chen, L., and Zhao, H. (2004), "A Statistical Method for Identifying Differential Gene-Gene Co-Expression Patterns," *Bioinformatics* , 20, 3146–3155. [229,236,237]

Lee, H. K., Hsu, A. K., and Sajdak, J. (2004), "Coexpression Analysis of Human Genes Across Many Microarray Data Sets," *Genome Research*, 14, 1085–1094. [229]

Liu, W. D., and Shao, Q. M. (in press), "Phase Transition and Regularized Bootstrap In Large-Scale t-Tests With False Discovery Rate Control," *Annals of Statistics*, 42, 2003–2025. [232]

Owen, A. B. (2005), "Variance of the Number of False Discoveries," *Journal of the Royal Statistical Society,* Series B, 67, 411–426. [239]

Qiu, X., Klebanov, L., and Yakovlev, A. (2005), "Correlation Between Gene Expression Levels and Limitations of the Empirical Bayes Methodology for Finding Differentially Expressed Genes," *Statistical Applications in Genetics and Molecular Biology*, 4. [229]

Raizada, R. D. S., Richards, T. L., Meltzoff, A., and Kuhl, P. K. (2008), "Socioeconomic Status Predicts Hemispheric Specialisation of the Left Inferior Frontal Gyrus in Young Children," *NeuroImage*, 40, 1392–1401. [229]

Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., Evans, A., Rapoport, J., and Giedd, J. (2006), "Intellectual Ability and Cortical Development in Children And Adolescents," *Nature*, 440, 676–679. [229,230]

Singh, D., Febbo, P., Ross, K., Jackson, D., Manola, J., Ladd, C., Tamayo, P., Renshaw, A., D′Amico, A., and Richie, J. (2002), "Gene Expression Correlates of Clinical Prostate Cancer Behavior," *Cancer Cell*, 1, 203–209. [237]

Storey, J. D. (2002), "A Direct Approach to False Discovery Rates," *Journal of the Royal Statistical Society*, Series B, 64, 479–498. [229]

Storey, J. D., Taylor, J., and Siegmund, D. (2004), "Strong Control, Conservative Point Estimation and Simultaneous Conservative Consistency of False Discovery Rates: A Unified Approach," *Journal of the Royal Statistical Society*, Series B, 66, 187–205. [229,233]

Sun, W., and Cai, T. T. (2007), "Oracle and Adaptive Compound Decision Rules for False Discovery Rate Control," *Journal of the American Statistical Association*, 102, 901–912. [229]

——— (2009), "Large-Scale Multiple Testing Under Dependence," *Journal of the Royal Statistical Society*, Series B, 71, 393–424. [229]

Wu, W. (2008), "On False Discovery Control Under Dependence," *Annals of Statistics*, 36, 364–380. [229]

Yeung, M. K. S., and Tegne, J. (2002), "Reverse Engineering Gene Networks using Singular Value Decomposition and Robust Regression," *Proceedings of the National Academy of Sciences*, 99, 6163–6168. [237]

Zhang, J., Li, J., and Deng, H. (2008), "Class-Specific Correlations of Gene Expressions: Identification and Their Effects on Clustering Analyses," *The American Journal of Human Genetics*, 83, 269–277. [229,230]

Zhu, D., Hero, A. O., Qin, Z. S., and Swaroop, A. (2005), "High Throughput Screening of co-Expressed Gene Pairs With Controlled False Discovery Rate (FDR) and Minimum Acceptable Strength (MAS)," *Journal of Computational Biology*, 12, 1029–1045. [229]

Downloaded by [University of Pennsylvania] at 06:26 06 June 2016