

Estimation of false discovery proportion in multiple testing: From normal to chi-squared test statistics

Lilun Du

*Department of ISOM, Hong Kong University of Science and Technology
Hong Kong
e-mail: dulilun@ust.hk*

and

Chunming Zhang

*Department of Statistics, University of Wisconsin, Madison, WI 53706
USA
e-mail: cmzhang@stat.wisc.edu*

Abstract: Multiple testing based on chi-squared test statistics is common in many scientific fields such as genomics research and brain imaging studies. However, the challenges of designing a formal testing procedure when there exists a general dependence structure across the chi-squared test statistics have not been well addressed. To address this gap, we first adopt a latent factor structure ([14]) to construct a testing framework for approximating the false discovery proportion (FDP) for a large number of highly correlated chi-squared test statistics with a finite number of degrees of freedom k . The testing framework is then used to simultaneously test k linear constraints in a large dimensional linear factor model with some observable and unobservable common factors; the result is a consistent estimator of the FDP based on the associated factor-adjusted p -values. The practical utility of the method is investigated through extensive simulation studies and an analysis of batch effects in a gene expression study.

MSC 2010 subject classifications: Primary 62H15, 62H25; secondary 62G10.

Keywords and phrases: Chi-squared distribution, factor-adjusted procedure, false discovery proportion, linear factor model, multiple comparison, restricted-PCA.

Received January 2016.

Contents

1	Introduction	1049
2	Testing framework from normal to chi-squared test statistics	1051
2.1	Multiple testing under multivariate non-central chi-squared distribution	1051
2.2	Approximation of FDP	1053

3	Simultaneously testing k linear constraints in a large dimensional linear factor model	1055
3.1	Model and assumptions	1055
3.2	A factor-adjusted multiple testing procedure	1056
3.3	Consistency of the estimated FDP	1058
3.4	Departure from normality	1060
4	Restricted-PCA algorithm on an initial subset	1060
5	Simulation studies	1063
5.1	FDP estimate: From negative dependence to consistency	1065
5.2	Factor-adjusted procedure via various algorithms for estimating the latent common factors	1065
5.3	Performance of the factor-adjusted procedure beyond normality	1068
6	An application: Tackling batch effects in gene expression among ethnic groups	1068
7	Concluding remarks	1071
A	Notation index	1072
B	Proofs of Proposition 1, Proposition 2, and Theorem 1	1073
C	Consistency of the Restricted-PCA algorithm	1081
D	Negative dependence of the BH type estimator	1088
	Acknowledgements	1089
	References	1089

1. Introduction

Large-scale multiple testing is common in analysis of high dimensional data, and it has wide application in scientific fields such as biology, medicine, genomics, neuroscience, economics, and finance. For example, in a genome-wide association study, hundreds of thousands of genetic polymorphisms were scanned simultaneously; in this process, multiple testing procedures, especially false discovery rate (FDR) methods ([4]), play an important role in detecting the association between genetic variants and some complex diseases ([32]).

Benjamini and Hochberg's (BH) procedure and its variants (e.g., the step-wise multiple testing procedures in [23] and false discovery rate estimators in [27]) have been shown to be theoretically valid when the test statistics are independent or have some form of weak dependence such as positive regression dependence on a subset (PRDS) ([5]; [23]) or general weak dependence on p -values ([27]). There has been some recent research on quantifying the effect of correlation in false discovery rate analysis. For instance, [7] argued that the difficulties caused by dependence in multiple testing tend to diminish as the number of simultaneous tests increases when the dependence is weak, provided that the test statistics have light-tailed marginal distributions; [31] generalized the popular random effect model that is assumed on the p -values to a conditional dependence model that allows weak dependence among the null hypotheses and can be useful to characterize the spatial structures of the null hypotheses. When the dependence is strong, the effect of correlation on the variance of some error

measures, such as the number of false discoveries or the estimator of false discovery rate, does not diminish as the number of tests increases and should be carefully accounted for ([22]; [11]; [25]; [2]). In particular, [13] proposed that the false discovery proportion (FDP) can be consistently estimated when the test statistics follow a multivariate normal distribution with arbitrary covariance dependence. Their method is quite useful in applications where the normality approximation of the test statistics is adequate, such as the expression quantitative trait loci (eQTL) in genome-wide association studies and the classical two-sample comparison in microarray studies. In practice, chi-squared or F tests are popular in many areas, including (but not limited to), the analysis of variance (ANOVA) for testing fixed or random effects in the analysis of experimental data, the Pearson's chi-squared goodness of fit test for independence, and many variants of likelihood-ratio tests. In all of the above testing problems, it would be beneficial to extend [13]'s method to accommodate the chi-squared or F tests in multiple testing under general dependence structures. It is worth noting that [23] showed that a class of multivariate F distributions that arises in many-to-one comparisons of variances with one-sided alternatives satisfies the PRDS property, and thus the BH procedure can be legally applied. However, this class of multivariate F distributions is very restrictive in applications as it is constructed based on a series of independent chi-squared random variables.

In this article, we provide a new method for solving multiple testing problems that arise when the test statistics follow a multivariate non-central chi-squared distribution with arbitrary underlying covariance dependence. Specifically, we construct a testing framework for a large number of highly correlated chi-squared test statistics, based on a small panel of data with dimension m by k . The dataset consists of k independent multivariate normally distributed random vectors with an arbitrary common covariance dependence structure, where k is the number of degrees of freedom for each chi-squared test statistic and m is the number of tests. For each multivariate normally distributed random vector, we adopt a latent factor structure ([14]) to model the arbitrary covariance dependence structure. That is, the covariance dependence can be additively decomposed into two separate sources of variation: (i) major dependent variation shared across the tests in terms of low dimensional latent common factors; and (ii) test-specific weakly dependent variation. Under this setting, we approximate the FDP based on the associated p -values with respect to a threshold t by summing a series of non-central chi-squared distribution functions, within which are found the latent common factors and factor loadings.

We then use the testing framework to simultaneously test k linear constraints in a large dimensional linear factor model, with another latent factor structure assumed on the random errors. This model has wide application in genomics research where some unobservable batch effects (or confounders) that contribute to the gene expression differentiation can be viewed as latent common factors. This naturally leads to a factor-adjusted multiple testing procedure ([20]; [16]). The factor-adjusted procedure is more robust and powerful than the traditional unadjusted method, because the latent common factors are taken into account at an early stage before any tests are performed, resulting in a new ranking

of the hypotheses being tested. To make the testing procedure operational, the latent common factors need to be accurately estimated, both theoretically and computationally. Both goals are achieved via a proposed restricted principal component analysis (Restricted-PCA) algorithm. We further show that the estimated FDP of the associated factor-adjusted p -values converges to the true one, even when idiosyncratic errors are allowed to be weakly correlated. Note that factor-adjusted multiple testing that accounts for latent factors in a linear model has been studied in the literature; see [20], [16], [28], and [17]. Although we present latent models in a similar fashion, our model settings (see (13) and (14)) are more general in the sense that the idiosyncratic errors can accommodate various dependence structures. Moreover, we provide a simple and analytically tractable Restricted-PCA algorithm to adjust for the latent factors. This is the key ingredient of our method; it enables us to derive a consistent estimated FDP.

The rest of this article is organized as follows. The testing framework for approximating the FDP based on a large number of highly correlated chi-squared test statistics is presented in detail in Section 2. Section 3 discusses the connection between simultaneously testing k linear constraints in the large dimensional linear factor model and the testing framework, and forms a consistent estimator of FDP based on the factor-adjusted p -values. The Restricted-PCA algorithm is described in Section 4. The model’s numerical performances including both simulation studies and a real data application are extensively investigated in Sections 5 and 6, respectively. Concluding remarks in Section 7 end this article. All of the proofs are deferred to the Appendix. For notational convenience, three tables summarizing the most common recurring notations in this article are provided in Appendix A.

2. Testing framework from normal to chi-squared test statistics

In this section, we systematically construct and study a testing framework for a large number of highly correlated chi-squared test statistics based on a small panel of data with dimension m by k , where m is the number of tests and k is the number of degrees of freedom for each chi-squared test statistic.

2.1. Multiple testing under multivariate non-central chi-squared distribution

We begin with a small panel of data $\mathbf{W} = (w_{i\ell})_{m \times k} \in \mathbb{R}^{m \times k}$. As a convention, we denote by $\mathbf{W}_{\cdot\ell} = (w_{1\ell}, \dots, w_{m\ell})^\top$ the ℓ -th column of \mathbf{W} , and assume that $\{\mathbf{W}_{\cdot\ell}, \ell = 1, \dots, k\}$ are independent and $\mathbf{W}_{\cdot\ell}$ follows a multivariate normal distribution with a mean $\tilde{\boldsymbol{\mu}}_{\cdot\ell} = (\mu_{1\ell}, \dots, \mu_{m\ell})^\top$ and correlation matrix $\Sigma_{\mathbf{W}} = (\sigma_{\mathbf{W},i_1i_2})_{m \times m}$ ($\sigma_{\mathbf{W},ii} = 1$, for $i = 1, \dots, m$). For notational convenience, we denote by $\mathbf{W}_{i\cdot} = (w_{i1}, \dots, w_{ik})^\top$ and $\tilde{\boldsymbol{\mu}}_{i\cdot} = (\mu_{i1}, \dots, \mu_{ik})^\top$ the i -th row of \mathbf{W} and its mean, respectively. Under the above setup, the joint distribution of $(\|\mathbf{W}_{1\cdot}\|_2^2, \dots, \|\mathbf{W}_{m\cdot}\|_2^2)^\top$ is termed the multivariate non-central chi-squared

distribution with k degrees of freedom, the accompanying correlation matrix $\Sigma_{\mathbf{W}}$, and the non-centrality parameters $(\|\tilde{\boldsymbol{\mu}}_1\|_2^2, \dots, \|\tilde{\boldsymbol{\mu}}_m\|_2^2)^\top$.

Within \mathbf{W} , suppose we are interested in testing the following hypotheses:

$$H_{0i} : \tilde{\boldsymbol{\mu}}_{i\cdot} = \mathbf{0} \quad \text{v.s.} \quad H_{1i} : \tilde{\boldsymbol{\mu}}_{i\cdot} \neq \mathbf{0}, \quad \text{for } i = 1, \dots, m. \tag{1}$$

By construction, $T_i = \|\mathbf{W}_{i\cdot}\|_2^2$ can serve as the test statistic for the i -th hypothesis, which follows a chi-squared distribution with degrees of freedom k under the true null. Therefore, the p -value for the i -th test is calculated as $P_i = 1 - \chi_k(T_i; 0)$, where $\chi_k(t; \lambda)$ is the cumulative distribution function (CDF) of a chi-squared random variable with degrees of freedom k and non-centrality parameter λ . Denote by \mathcal{I}_0 and \mathcal{I}_1 the sets of indices corresponding to the true null and non-null, and let m_0 and m_1 be the cardinality of \mathcal{I}_0 and \mathcal{I}_1 , respectively. Hence, the asymptotic proportion of the true null is defined as $\pi_0 = \lim_{m \rightarrow \infty} m_0/m$.

As noted by [10], correlation must be accounted for in deciding which null hypotheses are significant because the accuracy of FDR controlling techniques is compromised in high correlation situations. To capture the correlation structure of $\Sigma_{\mathbf{W}}$, we assume that $\{\mathbf{W}_{i\cdot}, i = 1, \dots, m\}$ satisfies the following latent factor structure ([3]; [14]):

$$\mathbf{W}_{i\cdot} = \tilde{\boldsymbol{\mu}}_{i\cdot} + \tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_i + \tilde{\boldsymbol{\eta}}_{i\cdot}, \quad \text{for } i = 1, \dots, m, \tag{2}$$

where $\tilde{\mathbf{Z}} = (\tilde{\mathbf{Z}}_1, \dots, \tilde{\mathbf{Z}}_k)^\top \in \mathbb{R}^{k \times r}$ and $\tilde{\mathbf{Z}}_\ell = (\tilde{Z}_{\ell 1}, \dots, \tilde{Z}_{\ell r})^\top$, for $\ell = 1, \dots, k$, is a $r \times 1$ vector of latent common factors with the identification restriction $E(\tilde{\mathbf{Z}}_\ell) = \mathbf{0}$ and $\text{cov}(\tilde{\mathbf{Z}}_\ell) = \mathbf{I}_r$. The accompanied $\tilde{\boldsymbol{\gamma}}_i$ are the unknown factor loadings and $\tilde{\boldsymbol{\eta}}_{i\cdot} = (\tilde{\eta}_{i1}, \dots, \tilde{\eta}_{ik})^\top$ are the idiosyncratic errors that are independent of $\tilde{\mathbf{Z}}$. In the spirit of the model setup in [13], we further assume that $\{\tilde{\boldsymbol{\eta}}_{\cdot\ell} = (\eta_{1\ell}, \dots, \eta_{m\ell})^\top, \ell = 1, \dots, k\}$ are independent and that $\tilde{\boldsymbol{\eta}}_{\cdot\ell}$ satisfies a weak dependence assumption as

$$\begin{aligned} &\tilde{\boldsymbol{\eta}}_{\cdot\ell} \stackrel{d}{\sim} \mathbb{N}(\mathbf{0}, \Sigma_{\tilde{\boldsymbol{\eta}}}) \quad \text{with } \Sigma_{\tilde{\boldsymbol{\eta}}} = (\sigma_{\tilde{\boldsymbol{\eta}}, i_1 i_2})_{m \times m}, \\ &\{m(m-1)\}^{-1} \sum_{i_1 \neq i_2} \sum_{i_1 \neq i_2} |\rho_{\tilde{\boldsymbol{\eta}}, i_1 i_2}| = O(m^{-\delta}), \quad \text{for some } \delta > 0, \end{aligned} \tag{3}$$

where $\rho_{\tilde{\boldsymbol{\eta}}, i_1 i_2} = \sigma_{\tilde{\boldsymbol{\eta}}, i_1 i_2} / (\sigma_{\tilde{\boldsymbol{\eta}}, i_1 i_1} \sigma_{\tilde{\boldsymbol{\eta}}, i_2 i_2})^{1/2}$ is the coefficient of correlation between $\tilde{\eta}_{i_1 \ell}$ and $\tilde{\eta}_{i_2 \ell}$. Here we would like to point out that $\tilde{\mathbf{Z}}$ drives the dependence among the m tests, hence the number of factor r , depends on m . Consequently, the mean $\tilde{\boldsymbol{\mu}}_{i\cdot}$, the factor loadings $\tilde{\boldsymbol{\gamma}}_i$ and the variance of the idiosyncratic error $\sigma_{\tilde{\boldsymbol{\eta}}, i_1 i_2}$ also depend on m . For notational simplicity, we suppress the dependent m in the model assumptions (2) and (3).

Remark 1. [13] introduced a definition for weakly dependent normal variables to guarantee the validity of their approximated FDP. Their definition is essentially in the same form as in (3) except that the correlation matrix is replaced by the covariance matrix. If the diagonals of the covariance matrix are bounded away from zero and infinity, these two definitions are mutually inclusive. A

more general definition of weakly correlated standard normal variables is given in Definition 1 of [2], which is a necessary and sufficient condition for the consistency of the associated empirical cumulative distribution function in \mathcal{L}_2 norm. Another related condition was given in [25], where they assumed that the empirical moments of all of the pairwise correlations of the chi-squared test statistics shrink to zero when the number of tests diverges to infinity. Under this condition and a novel decomposition for bivariate chi-squared density function, they showed that the fraction of discoveries has asymptotically vanishing variance. We remark here that the technique for decomposing the bivariate chi-squared density function is still not applicable to our case because the condition in (3) is imposed directly on the underlying correlation matrix.

2.2. Approximation of FDP

Traditional multiple testing procedures are based on a sequence of p -values for a family of hypotheses, where the hypotheses with p -values less than or equal to some threshold t are rejected. Define the following empirical processes:

$$\begin{aligned} V(t) &= \#\{i \in \mathcal{I}_0 : P_i \leq t\}, \\ S(t) &= \#\{i \in \mathcal{I}_1 : P_i \leq t\}, \quad \text{and} \\ R(t) &= \#\{i \in \mathcal{I} : P_i \leq t\} \end{aligned} \tag{4}$$

for any $t \in [0, 1]$, where $\mathcal{I} = \mathcal{I}_0 \cup \mathcal{I}_1$. Then $V(t)$, $S(t)$, and $R(t)$ are the number of falsely rejected hypotheses, the number of correctly rejected hypotheses, and the total number of rejected hypotheses, respectively. The false discovery proportion with respect to the threshold t is defined as $\text{FDP}(t) = V(t)/\{R(t) \vee 1\}$ with $R(t) \vee 1 = \max\{R(t), 1\}$.

Note that $V(t)$ is unobserved but realized through an experiment, whereas $R(t)$ can be observed. Essentially, estimating the FDP is equivalent to approximating $V(t)/m_0$. As discussed in [27], if the p -values satisfy the weak dependence assumption, then $V(t)/m_0$ would converge to some distribution function almost surely. However, a strong dependence structure on $\Sigma_{\mathbf{W}}$ would affect or even violate the convergence of $V(t)/m_0$. From (2), after controlling for the effect of the latent common factors $\tilde{\mathbf{Z}}$, the correlation among the idiosyncratic errors $\tilde{\eta}_{\cdot\ell}$ is weak, according to assumption (3). Hence, it is possible to seek an approximated expression for $V(t)/m_0$ conditional on the latent common factors $\tilde{\mathbf{Z}}$, which in turn can be used to estimate the FDP.

Proposition 1. *Suppose assumptions (2) and (3) hold. If we further assume that (i): $\lim_m \sup_i \|\tilde{\boldsymbol{\mu}}_{i\cdot}\|_2^2 < \infty$; (ii): $\lim_m \inf_i \sigma_{\tilde{\eta},ii} > 0$; (iii): for each realization of $\tilde{\mathbf{Z}}$, $\lim_m \sup_i \|\tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_i\|_2^2 < \infty$, then for a finite number of degrees of freedom k and any $t \in [0, 1]$,*

$$\lim_{m_0 \rightarrow \infty} \left[m_0^{-1} V(t) - m_0^{-1} \sum_{i \in \mathcal{I}_0} \left\{ 1 - \chi_k \left(\frac{\chi_k^{-1}(1-t; 0)}{\sigma_{\tilde{\eta},ii}}; \lambda_{0i} \right) \right\} \right] = 0, \quad \text{a.s.} \tag{5}$$

$$\lim_{m \rightarrow \infty} \left[m^{-1} R(t) - m^{-1} \sum_{i=1}^m \left\{ 1 - \chi_k \left(\frac{\chi_k^{-1}(1-t; 0)}{\sigma_{\tilde{\boldsymbol{\eta}}, ii}}; \lambda_{1i} \right) \right\} \right] = 0, \quad a.s. \quad (6)$$

$$\lim_{m \rightarrow \infty} \left[\text{FDP}(t) - \frac{\sum_{i \in \mathcal{I}_0} \left\{ 1 - \chi_k \left(\frac{\chi_k^{-1}(1-t; 0)}{\sigma_{\tilde{\boldsymbol{\eta}}, ii}}; \lambda_{0i} \right) \right\}}{\sum_{i=1}^m \left\{ 1 - \chi_k \left(\frac{\chi_k^{-1}(1-t; 0)}{\sigma_{\tilde{\boldsymbol{\eta}}, ii}}; \lambda_{1i} \right) \right\}} \right] = 0, \quad a.s. \quad (7)$$

where $\lambda_{0i} = \|\tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_i\|_2^2/\sigma_{\tilde{\boldsymbol{\eta}}, ii}$, $\lambda_{1i} = \|\tilde{\boldsymbol{\mu}}_{i\cdot} + \tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_i\|_2^2/\sigma_{\tilde{\boldsymbol{\eta}}, ii}$, and $\chi_k^{-1}(t; \lambda)$ is the inverse function of $\chi_k(t; \lambda)$.

The proof can be found in Appendix B. The above results could be generalized to the case where the number of degrees of freedoms k is allowed to grow to infinity by assuming that the average of the coefficients of correlation grows to zero at the rate $O(m^{-\delta}k^{-1})$ in assumption (3). Essentially, this Proposition indicates that the difference between two random parts (i.e., $V(t)$ or $R(t)$ and the corresponding approximation) is negligible. This is achieved by assuming that the latent factors part $\tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_i$ is fixed, and by deriving the analysis solely on the dependence structure of $\tilde{\boldsymbol{\eta}}_i$.

Proposition 1 approximates the corresponding FDP(t) by a series of non-central chi-squared distribution functions, which contain some latent common factors and factor loadings. In practice, the approximation of $V(t)$ in (5) is intractable due to the unknown true null set \mathcal{I}_0 . If the corresponding summation of the alternative set \mathcal{I}_1 in (5) is negligible, we can conservatively estimate $V(t)$ as

$$\sum_{i=1}^m \left\{ 1 - \chi_k \left(\frac{\chi_k^{-1}(1-t; 0)}{\sigma_{\tilde{\boldsymbol{\eta}}, ii}}; \lambda_{0i} \right) \right\}. \quad (8)$$

Recall that $\text{FDP}(t) = V(t)/\{R(t) \vee 1\}$, in which $R(t)$ is observed. From (8), we propose to estimate the FDP(t) as

$$\widehat{\text{FDP}}(t) = \left[\sum_{i=1}^m \left\{ 1 - \chi_k \left(\frac{\chi_k^{-1}(1-t; 0)}{\sigma_{\tilde{\boldsymbol{\eta}}, ii}}; \lambda_{0i} \right) \right\} \right] / \{R(t) \vee 1\}. \quad (9)$$

To demonstrate the usefulness of the testing framework, in the next section we obtain a consistent estimator of the FDP when the test statistics are multivariate non-central chi-squared distributed within a large dimensional linear factor model.

Remark 2. Our approximated FDP in Proposition 1 is closely related to two existing studies in the literature as follows.

The weak dependence case given in [27].

By letting $\tilde{\mathbf{Z}} = \mathbf{0}$, $\mathbf{W}_{i\cdot}$'s are weakly correlated in the absence of the latent common factors $\tilde{\mathbf{Z}}$, in which the summation part of (5) is $m_0 t$. This essentially implies that the empirical distribution of the chi-squared type p-values under the true null converges to the standard uniform distribution, which is a generalization of Theorem 1 in [2]. Thus, we can adopt the method of [27] to estimate FDP(t) as $m\hat{\pi}_0(\lambda)t/\{R(t) \vee 1\}$ with

$\widehat{\pi}_0(\lambda) = \#\{i; P_i > \lambda\} / \{m(1 - \lambda)\}$. Here λ is selected such that the number of hypotheses under the alternative with p -values greater than λ is negligible.

The strong dependence case given in [13].

When the number of degrees of freedom k reduces to 1, the approximated FDP in (7) can also be reformulated as

$$\lim_{m \rightarrow \infty} \left[\text{FDP}(t) - \frac{\sum_{i \in \mathcal{I}_0} \left\{ \Phi\left(\frac{z_{t/2} + \tilde{\gamma}_i^\top \tilde{\mathbf{Z}}_1}{(\sigma_{\tilde{\eta}, ii})^{1/2}}\right) + \Phi\left(\frac{z_{t/2} - \tilde{\gamma}_i^\top \tilde{\mathbf{Z}}_1}{(\sigma_{\tilde{\eta}, ii})^{1/2}}\right) \right\}}{\sum_{i=1}^m \left\{ \Phi\left(\frac{z_{t/2} + \mu_{i1} + \tilde{\gamma}_i^\top \tilde{\mathbf{Z}}_1}{(\sigma_{\tilde{\eta}, ii})^{1/2}}\right) + \Phi\left(\frac{z_{t/2} - \mu_{i1} - \tilde{\gamma}_i^\top \tilde{\mathbf{Z}}_1}{(\sigma_{\tilde{\eta}, ii})^{1/2}}\right) \right\}} \right] \stackrel{\text{a.s.}}{=} 0, \tag{10}$$

where $\Phi(\cdot)$ and $z_{t/2} = \Phi^{-1}(t/2)$ are the CDF and the $t/2$ lower quantile of a standard normal distribution, respectively. The above approximation is exactly the same as in [13]. Thus, our approximation of FDP can accommodate the cases where the test statistics are normally or chi-squared distributed.

3. Simultaneously testing k linear constraints in a large dimensional linear factor model

The testing framework in Section 2 is derived under a very special data matrix $\mathbf{W} \in \mathbb{R}^{m \times k}$, which is motivated by the structure of multivariate chi-squared distribution. To illustrate the practical utility of the testing framework, we provide a specific example, that is, simultaneously testing k linear constraints in a large dimensional linear factor model.

3.1. Model and assumptions

Assume we observe a large panel of data, that is, $\mathbf{Y} = (Y_{ij})_{m \times n} \in \mathbb{R}^{m \times n}$, with a total of m units, and each unit has n observations. Let $\mathbf{X}_j = (X_{j1}, \dots, X_{jp})^\top \in \mathbb{R}^p$ be the observable explanatory variables. To link the panel data to the explanatory variables, we consider the following multivariate linear regression:

$$\mathbf{Y}_{i\cdot} = \mathbf{X}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_{i\cdot}, \quad \text{for } i = 1, \dots, m, \tag{11}$$

where $\mathbf{Y}_{i\cdot} = (Y_{i1}, \dots, Y_{in})^\top \in \mathbb{R}^n$, $\mathbf{X}_i = (\mathbf{X}_{i1}, \dots, \mathbf{X}_{in})^\top \in \mathbb{R}^{n \times p}$ is the design matrix, $\boldsymbol{\beta}_i = (\beta_{i1}, \dots, \beta_{ip})^\top \in \mathbb{R}^p$ are the regression coefficients of unit i , and $\boldsymbol{\varepsilon}_{i\cdot} = (\varepsilon_{i1}, \dots, \varepsilon_{in})^\top \in \mathbb{R}^n$ are the random errors that are independent of \mathbf{X} . Moreover, we assume that $\{\boldsymbol{\varepsilon}_{\cdot j} = (\varepsilon_{1j}, \dots, \varepsilon_{mj})^\top, j = 1, \dots, n\}$ are independent with mean $\mathbf{0}$ and covariance matrix $\Sigma_{\mathcal{E}} = (\sigma_{\mathcal{E}, ij})_{m \times m}$, and collect the error panel as $\mathcal{E} = (\boldsymbol{\varepsilon}_{\cdot 1}, \dots, \boldsymbol{\varepsilon}_{\cdot n}) \in \mathbb{R}^{m \times n}$. To model the dependence structure of $\Sigma_{\mathcal{E}}$, we further assume that $\{\boldsymbol{\varepsilon}_{i\cdot}, i = 1, \dots, m\}$ follows a similar latent factor structure as in (2), that is,

$$\boldsymbol{\varepsilon}_{i\cdot} = \mathbf{Z} \boldsymbol{\gamma}_i + \boldsymbol{\eta}_{i\cdot}, \quad \text{for } i = 1, \dots, m, \tag{12}$$

where $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_n)^\top \in \mathbb{R}^{n \times r}$ and $\{\mathbf{Z}_j = (Z_{j1}, \dots, Z_{jr})^\top, j = 1, \dots, n\}$ are independent and each is an $r \times 1$ vector of latent common factors with the identification restriction $E(\mathbf{Z}_j) = \mathbf{0}$ and $\text{cov}(\mathbf{Z}_j) = \mathbf{I}_r$, $\boldsymbol{\gamma}_i$ are the corresponding factor loadings ($r \times 1$), and $\boldsymbol{\eta}_{i\cdot} = (\eta_{i1}, \dots, \eta_{in})^\top$ are the idiosyncratic errors, which are independent of \mathbf{Z} . Moreover, $\{\boldsymbol{\eta}_{\cdot j} = (\eta_{1j}, \dots, \eta_{mj})^\top, j = 1, \dots, n\}$ are supposed to be *i.i.d.*, and $\boldsymbol{\eta}_{\cdot j}$ satisfies a similar weak dependence assumption as in (3), that is,

$$\boldsymbol{\eta}_{\cdot j} \stackrel{d}{\sim} \mathbb{N}(\mathbf{0}, \Sigma_\eta) \quad \text{with } \Sigma_\eta = (\sigma_{\eta, i_1 i_2})_{m \times m}, \text{ and} \\ \{m(m-1)\}^{-1} \sum_{i_1 \neq i_2} |\rho_{\eta, i_1 i_2}| = O(m^{-\delta}), \text{ for some } \delta > 0, \quad (13)$$

with $\rho_{\eta, i_1 i_2} = \sigma_{\eta, i_1 i_2} / (\sigma_{\eta, i_1 i_1} \sigma_{\eta, i_2 i_2})^{1/2}$ being the coefficient of correlation between $\eta_{i_1 \ell}$ and $\eta_{i_2 \ell}$. After incorporating the latent factor structure (12), model (11) can be further written as

$$\mathbf{Y}_{i\cdot} = \mathbf{X}\boldsymbol{\beta}_i + \mathbf{Z}\boldsymbol{\gamma}_i + \boldsymbol{\eta}_{i\cdot}, \quad \text{for } i = 1, \dots, m. \quad (14)$$

(14) is called the large dimensional linear factor model with some observable and unobservable common factors.

In regression analysis, testing the regression coefficients is of great interest in practice. Most testing problems related to the regression coefficients for the i -th unit in (14) can be formulated by a set of linear constraints as follows:

$$H_{0i} : \mathbf{A}\boldsymbol{\beta}_i = \mathbf{0} \quad \text{v.s.} \quad H_{1i} : \mathbf{A}\boldsymbol{\beta}_i \neq \mathbf{0}, \quad (15)$$

where $\mathbf{A} \in \mathbb{R}^{k \times p}$ is a full low rank matrix with $\text{rank}(\mathbf{A}) = k$, and $k \leq p$. Stacking all of the m hypotheses, as in (15), multiplicity must be accounted for. In the following Sections 3.2 and 3.3, we first elaborate a factor-adjusted multiple testing procedure to address the multiplicity issue, and then resort to the theoretical framework in Section 2 to provide convincing mathematical proof to show its consistency.

3.2. A factor-adjusted multiple testing procedure

Motivated by the parametric F -test in linear regression and by ignoring the latent common factors \mathbf{Z} , we construct an unadjusted test statistic for (15) and approximate the corresponding p -value as follows:

$$\begin{aligned} T_i^u &= (\mathbf{A}\widehat{\boldsymbol{\beta}}_i)^\top \{\widehat{\sigma}_{\mathcal{E}, ii} \mathbf{A}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top\}^{-1} (\mathbf{A}\widehat{\boldsymbol{\beta}}_i) \text{ and} \\ P_i^u &= 1 - \chi_k(T_i^u; 0), \end{aligned} \quad (16)$$

where $\widehat{\boldsymbol{\beta}}_i = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}_{i\cdot}$ is the ordinary least-squares estimate of $\boldsymbol{\beta}_i$, and $\widehat{\sigma}_{\mathcal{E}, ii} = \mathbf{Y}_{i\cdot}^\top \mathbf{Q}(\mathbf{X}) \mathbf{Y}_{i\cdot} / (n - p)$ with the operator \mathbf{Q} defined as $\mathbf{Q}(\mathbf{X}) = \mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$. One problem of the parametric F -test in (16) is that the ranking of the statistical significance is completely determined by that of $\{T_i^u\}$ or $\{P_i^u\}$,

which is undesirable and inefficient in practice. Moreover, as mentioned in [8], the approximated FDP based on unadjusted p -values tends to be highly unstable due to dependence, even when the dependence is fully parametrized and used in the procedure. In other words, any downstream methods (e.g., adjusting test statistics) are naturally at a disadvantage. To provide a more robust and powerful remedy, the latent common factors \mathbf{Z} must be taken into account at an early stage before any tests are performed, resulting in a new ranking of the hypotheses being tested. Based on the above rationale, the testing problem (15) can be conducted using factor-adjusted test statistics under model setup (14). The idea was tentatively introduced in [20] and [16].

From model (14), $\mathbf{Y}_i = \mathbf{X}\beta_i + \mathbf{Z}\gamma_i + \boldsymbol{\eta}_i$, the latent common factors \mathbf{Z} should be extracted from $\widehat{\beta}_i$ before the test statistics are formed. This motivates us to introduce an oracle factor-adjusted test statistic and a p -value for the i -th unit as follows:

$$\begin{aligned} T_i^{\text{ad}} &= \frac{\{\mathbf{A}\widehat{\beta}_i - \mathbf{A}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Z}\gamma_i\}^\top}{\{\sigma_{\boldsymbol{\eta},ii} \mathbf{A}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top\}^{-1}} \\ P_i^{\text{ad}} &= 1 - \chi_k(T_i^{\text{ad}}; 0). \end{aligned} \tag{17}$$

To make the testing procedure operational, the latent factors part $\mathbf{Z}\gamma_i$ and the variance of the idiosyncratic error $\sigma_{\boldsymbol{\eta},ii}$ should be estimated by some algorithm and then used to replace the true ones. Specifically, the factor-adjusted test statistics and the associated p -values for (15) can be constructed as follows:

$$\begin{aligned} \widehat{T}_i^{\text{ad}} &= \frac{\{\mathbf{A}\widehat{\beta}_i - \mathbf{A}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \widehat{\mathbf{Z}}\widehat{\gamma}_i\}^\top}{\{\widehat{\sigma}_{\boldsymbol{\eta},ii} \mathbf{A}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top\}^{-1}} \\ \widehat{P}_i^{\text{ad}} &= 1 - \chi_k(\widehat{T}_i^{\text{ad}}; 0), \end{aligned} \tag{18}$$

where $\widehat{\mathbf{Z}}\widehat{\gamma}_i$ and $\widehat{\sigma}_{\boldsymbol{\eta},ii}$ can be obtained by the algorithms discussed in Section 4. As in (4), we denote by $\{V^{\text{ad}}(t), \widehat{V}^{\text{ad}}(t)\}$ and $\{R^{\text{ad}}(t), \widehat{R}^{\text{ad}}(t)\}$ the number of false rejections and the number of total rejections with respect to a threshold t based on $\{P_i^{\text{ad}}, \widehat{P}_i^{\text{ad}}\}$'s, respectively. Hence, the false discovery proportion of the oracle factor-adjusted p -values is defined as $\text{FDP}^{\text{ad}}(t) = V^{\text{ad}}(t)/\{R^{\text{ad}}(t) \vee 1\}$.

Intuitively, if the latent common factor and the variance of the idiosyncratic error can be accurately estimated, the behavior of $\widehat{P}_i^{\text{ad}}$ s would resemble that of P_i^{ad} s to a large extent. Moreover, the dependence between the oracle factor-adjusted test statistics or p -values only comes from the idiosyncratic error part $\boldsymbol{\eta}_i$, which is weakly dependent according to assumption (3). Hence, we can adopt the traditional multiple testing procedure to estimate the $\text{FDP}^{\text{ad}}(t)$ ([27]) as

$$\widehat{\text{FDP}}_\lambda^{\text{ad}}(t) = m\widehat{\pi}_0(\lambda)t/\{\widehat{R}^{\text{ad}}(t) \vee 1\}, \tag{19}$$

where $\widehat{\pi}_0(\lambda) = \sum_{i=1}^m \mathbf{I}(\widehat{P}_i^{\text{ad}} > \lambda)/\{m(1 - \lambda)\}$ is an estimate of the proportion of true null π_0 and λ is defined in Remark 2.

3.3. Consistency of the estimated FDP

In this subsection, we connect the chi-squared type factor-adjusted p -values to the testing framework in Section 2 to explore the theoretical property of the estimated FDP in (19). It has been shown that the BH procedure ([4]) and Storey's method ([27]) continue to control the FDR even when the test statistics or the p -values are weakly dependent. Does the weak dependence assumption hold for the factor-adjusted p -values? To answer this question, we propose a two-stage procedure. In the first stage, we connect the structure of the oracle factor-adjusted p -values (i.e., P_i^{ad}) to the theoretical framework in Section 2 to show that the weak dependence assumption of the oracle factor-adjusted p -values is satisfied; in the second stage, we supply a sufficient condition on the estimators of the latent factors and the variance of the idiosyncratic error, under which the difference between the factor-adjusted p -values (i.e., \hat{P}_i^{ad}) and the oracle factor-adjusted p -value is asymptotically negligible.

Stage I: Weak dependence of the oracle factor-adjusted p -values

To this end, we first heuristically argue that by decomposing $(T_1^{\text{ad}}, \dots, T_m^{\text{ad}})^\top$, a data matrix \mathbf{W}^{ad} as in Section 2.1 can be recovered through \mathbf{Y} and the design matrix \mathbf{X} , and \mathbf{W}^{ad} satisfies assumptions (2) and (3). This can be achieved by resorting to Proposition 1 in the testing framework. More specifically, by comparing the structure of $(T_1, \dots, T_m)^\top$ in Section 2 with that of $(T_1^{\text{ad}}, \dots, T_m^{\text{ad}})^\top$, we observe that

$$\begin{aligned} \mathbf{W}_i^{\text{ad}} &\triangleq \{\sigma_{\eta,ii} \mathbf{A}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top\}^{-1/2} \{\mathbf{A} \hat{\boldsymbol{\beta}}_i - \mathbf{A}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Z} \boldsymbol{\gamma}_i\} \\ &= \{\sigma_{\eta,ii} \mathbf{A}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top\}^{-1/2} \{\mathbf{A} \boldsymbol{\beta}_i + \mathbf{A}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\eta}_i\} \\ &= \{\sigma_{\eta,ii} \mathbf{A}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top\}^{-1/2} \mathbf{A} \boldsymbol{\beta}_i + (\sigma_{\eta,ii})^{-1/2} \mathcal{P}^\top \boldsymbol{\eta}_i, \end{aligned} \quad (20)$$

where $\mathcal{P} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top \{\mathbf{A}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top\}^{-1/2} \in \mathbb{R}^{n \times k}$. Note that the term \mathbf{W}_i^{ad} is closely related to the model assumptions (2) and (3) in the testing framework. To link the parameters in the large dimensional linear factor model to our testing framework, we directly compare (20) with (2), which yields

$$\begin{aligned} \tilde{\boldsymbol{\mu}}_i &= (\sigma_{\eta,ii})^{-1/2} \{\mathbf{A}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top\}^{-1/2} \mathbf{A} \boldsymbol{\beta}_i, \\ \tilde{\mathbf{Z}} \tilde{\boldsymbol{\gamma}}_i &= \mathbf{0}, \\ \tilde{\boldsymbol{\eta}}_i &= (\sigma_{\eta,ii})^{-1/2} \mathcal{P}^\top \boldsymbol{\eta}_i, \end{aligned} \quad (21)$$

where $\tilde{\boldsymbol{\mu}}_i$, $\tilde{\boldsymbol{\gamma}}_i$, $\tilde{\mathbf{Z}}$, and $\tilde{\boldsymbol{\eta}}_i$ are as in (2) and (3). Note that the matrix $\mathcal{P} \in \mathbb{R}^{n \times k}$ connects $\boldsymbol{\eta}_i \in \mathbb{R}^n$ to $\tilde{\boldsymbol{\eta}}_i \in \mathbb{R}^k$ as in Section 2. Accordingly, we term \mathcal{P} the connection matrix. As $\mathcal{P}^\top \mathcal{P} = \mathbf{I}_k$, one can readily verify that $\tilde{\boldsymbol{\eta}}_i \sim \mathbb{N}(\mathbf{0}, \mathbf{I}_k)$, and $\tilde{\boldsymbol{\eta}}_{i,\ell} = D_\eta \boldsymbol{\eta} \mathcal{P}_\ell \sim \mathbb{N}(\mathbf{0}, D_\eta \Sigma_\eta D_\eta)$, where \mathcal{P}_ℓ is the ℓ -th column of \mathcal{P} and $D_\eta = \text{diag}\{(\sigma_{\eta,11})^{-1/2}, \dots, (\sigma_{\eta,mm})^{-1/2}\}$. Hence, assumptions (2) and (3) in Section 2 are satisfied. In particular, as the latent common factors \mathbf{Z} have been adjusted before the test statistics are formed, there are no common factors involved in (20). Therefore, the corresponding non-centrality parameter $\lambda_{0i}^{\text{ad}} / \lambda_{1i}^{\text{ad}}$ and the variance of the idiosyncratic error $\sigma_{\tilde{\boldsymbol{\eta}},ii}^{\text{ad}}$, as in (5), are reduced to $0 / \|\tilde{\boldsymbol{\mu}}_i\|_2^2$

and 1, respectively. From Proposition 1, we obtain

$$\lim_{m \rightarrow \infty} \left[m_0^{-1} V^{\text{ad}}(t) - t \right] = 0, \quad \text{a.s. and} \quad (22)$$

$$\lim_{m \rightarrow \infty} \left[m^{-1} R^{\text{ad}}(t) - \pi_0 t - (1 - \pi_0) F_1(t) \right] = 0, \quad \text{a.s.} \quad (23)$$

where $F_1(t) = \lim_{m \rightarrow \infty} m_1^{-1} \sum_{i \in \mathcal{I}_1} \{1 - \chi_k(\chi_k^{-1}(1 - t; 0); \lambda_{1i}^{\text{ad}})\}$ and the noncentrality parameter $\lambda_{1i}^{\text{ad}} = (\mathbf{A}\beta_i)^\top \{\mathbf{A}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top\} \mathbf{A}\beta_i / \sigma_{\eta,ii}$.

Stage II: When can we ignore the difference between the oracle factor-adjusted p-values and the factor-adjusted p-values?

In the second stage, we supply a sufficient condition for the estimators of the latent common factors and the variance of the idiosyncratic error, under which the differences between $m_0^{-1} \widehat{V}^{\text{ad}}(t)$ and $m_0^{-1} V^{\text{ad}}(t)$ ($m^{-1} \widehat{R}^{\text{ad}}(t)$ and $m^{-1} R^{\text{ad}}(t)$) are negligible. The results are summarized in the following Proposition.

Proposition 2. *For any estimators $\widehat{\mathbf{Z}}, \widehat{\gamma}_i$, and $\widehat{\sigma}_{\eta,ii}$ that satisfy $\max_{i \in \mathcal{I}} |\widehat{\sigma}_{\eta,ii} - \sigma_{\eta,ii}| \rightarrow_p 0$ and $\max_{i \in \mathcal{I}} \|\mathcal{P}^\top \widehat{\mathbf{Z}} \widehat{\gamma}_i - \mathcal{P}^\top \{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\} \mathbf{Z}\gamma_i\|_2 \rightarrow_p 0$, we can obtain, with a probability tending to one,*

$$\lim_{m \rightarrow \infty} \{m_0^{-1} \widehat{V}^{\text{ad}}(t) - m_0^{-1} V^{\text{ad}}(t)\} = 0, \quad \text{a.s. and} \quad (24)$$

$$\lim_{m \rightarrow \infty} \{m^{-1} \widehat{R}^{\text{ad}}(t) - m^{-1} R^{\text{ad}}(t)\} = 0. \quad \text{a.s.} \quad (25)$$

The above conditions essentially mean that $\widehat{\sigma}_{\eta,ii} - \sigma_{\eta,ii}$ and $\mathcal{P}^\top \widehat{\mathbf{Z}} \widehat{\gamma}_i - \mathcal{P}^\top \mathbf{Z}\gamma_i$ converge to zero in probability uniformly for all i , so that the effect of estimating the latent common factors vanishes uniformly for all $i \in \mathcal{I}$, as m and n increase. For more detailed derivations, see the proof in Appendix B.

Combining the above two stages, the consistency of $\widehat{\text{FDP}}_\lambda^{\text{ad}}(t)$ can be readily obtained, which is summarized in the following theorem.

Theorem 1. *Suppose the conditions in Proposition 2 and assumption (13) hold. Then for a finite number of degrees of freedom k , we can obtain*

$$\lim_{m,n \rightarrow \infty} \left| \widehat{\text{FDP}}_\lambda^{\text{ad}}(t) - c \times \text{FDP}^{\text{ad}}(t) \right| = 0, \quad \text{a.s.} \quad (26)$$

where $c = [\pi_0 + (1 - \pi_0)\{1 - F_1(\lambda)\} / (1 - \lambda)] / \pi_0$.

The proof can be found in Appendix B. Theorem 1 indicates that the estimated FDP can consistently match the true FDP of the oracle factor-adjusted p-values when the tuning parameter λ is well chosen, such that $F_1(\lambda) = 1$.

Remark 3. *We remark here that we can use techniques similar to those in Section 3.3 to decompose the unadjusted test statistics T_i^{u} and connect the decomposed formulas to the theoretical framework in Section 2 to approximate the false discovery proportion for the unadjusted p-values P_i^{u} with respect to a*

threshold t as

$$\widehat{\text{FDP}}^u(t) = \sum_{i=1}^m \chi_k \left(\frac{\chi_k^{-1}(1-t)}{\widehat{\sigma}_{\boldsymbol{\eta},ii}/\widehat{\sigma}_{\boldsymbol{\varepsilon},ii}}; \|\mathcal{P}^\top \widehat{\mathbf{Z}} \widehat{\boldsymbol{\gamma}}_i\|_2^2 / \widehat{\sigma}_{\boldsymbol{\eta},ii} \right) / \{R^u(t) \vee 1\},$$

where $R^u(t) = \sum_{i=1}^m \mathbb{I}(P_i^u \leq t)$ is the number of total rejections. Nevertheless, this estimator overestimates the true FDP under the non-sparse assumption, and the associated power is lower than that based on the factor-adjusted procedure.

3.4. Departure from normality

In this subsection, we discuss the possible relaxation of the assumption of normality (13) on $\boldsymbol{\eta}_j$. We conjecture that as long as $\boldsymbol{\eta}_j$'s are *i.i.d.* with mean $\mathbf{0}$ and a finite covariance matrix $\Sigma_{\boldsymbol{\eta}}$, the theoretical properties in Section 3.3 are still valid. For illustration, we consider the large dimensional linear factor model (14) as a one-way ANOVA model with p treatments. As the design matrix \mathbf{X} has a special structure within this model, \mathcal{P}_ℓ (the ℓ -th column of the connection matrix) is a vector with p distinct values. For notational convenience, suppose that the g -th treatment has n_g replications with $\sum_{g=1}^p n_g = n$,

and $\mathcal{P}_\ell = \left(\overbrace{\frac{\varpi_{1\ell}}{\sqrt{n_1}}, \dots, \frac{\varpi_{1\ell}}{\sqrt{n_1}}}^{n_1}, \dots, \overbrace{\frac{\varpi_{g\ell}}{\sqrt{n_g}}, \dots, \frac{\varpi_{g\ell}}{\sqrt{n_g}}}^{n_g}, \dots, \overbrace{\frac{\varpi_{p\ell}}{\sqrt{n_p}}, \dots, \frac{\varpi_{p\ell}}{\sqrt{n_p}}}^{n_p} \right)^\top$ with $\sum_{g=1}^p \varpi_{g\ell}^2 = 1$. Consequently, by definition, $\tilde{\boldsymbol{\eta}}_{\cdot,\ell}$ can be decomposed as

$$\tilde{\boldsymbol{\eta}}_{\cdot,\ell} = D_{\boldsymbol{\eta}} \sum_{g=1}^p \varpi_{g\ell} \left(\sum_{h=1}^{n_g} \frac{1}{\sqrt{n_g}} \boldsymbol{\eta}_{\cdot,gh} \right),$$

where $\{\boldsymbol{\eta}_{\cdot,gh}, g = 1, \dots, p, h = 1, \dots, n_g\}$ are the idiosyncratic errors in (12) under the one-way ANOVA model. Then, as $\min\{n_g, g = 1, \dots, p\} \rightarrow \infty$, according to the central limit theorem (CLT), $\tilde{\boldsymbol{\eta}}_{\cdot,\ell}$ is multivariate normally distributed with mean $\mathbf{0}$ and covariance matrix $D_{\boldsymbol{\eta}} \Sigma_{\boldsymbol{\eta}} D_{\boldsymbol{\eta}}$, asymptotically. Therefore, all of the results in Section 3.3 are expected to hold beyond normality. For simplicity, we leave the rigorous proof of the above conjecture for future research. To support our conjecture, in Section 5 we evaluate the performance of the proposed factor-adjusted procedure using various idiosyncratic error distributions.

4. Restricted-PCA algorithm on an initial subset

This section provides a new algorithm for estimating the latent common factors, and shows that they indeed satisfy the sufficient condition given in Proposition 2. In the literature, the method of principal component analysis (PCA) has been developed for consistently estimating the latent common factors \mathbf{Z} in model (12) if the pure error matrix \mathcal{E} is observed; see, e.g., [3] and [14]. In practice, \mathcal{E} is not directly observed; rather, we observe \mathbf{Y} , which is the sum of $\mathbf{B}\mathbf{X}^\top$ and \mathcal{E} with $\mathbf{B} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m)^\top \in \mathbb{R}^{m \times p}$ (model (11)). Under model (14), [20] introduced an

iterative re-weighted surrogate variables analysis (IRW-SVA) to estimate \mathbf{Z} . In essence, the weight vector in IRW-SVA is iteratively estimated and converges to its steady state such that big weights are assigned to the rows of \mathbf{Y} corresponding to the true null whereas small weights (close to zero) are assigned to those related to non-null. Note that under the null hypothesis that $\beta_i = 0$ (in other cases, we can transform $\mathbf{X}\beta_i$ as $\mathbf{X}(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top (\mathbf{A}\beta_i)$), \mathcal{E}_i is observable as it is equal to \mathbf{Y}_i . Hence, when the small weights vanish after several iterations, the subsequent standard factor analysis ([3]; [14]) can accurately estimate the latent factors \mathbf{Z} . Analogously, [17] suggested using negative control genes to perform factor analysis to remove unwanted variation (RUV), where the negative control genes can be regarded as a subset mimicking the true null set. In their approach, the negative control genes are genes known as a *priori* not to be differentially expressed with respect to the biological factor of interest. In summary, both methods essentially perform the standard PCA algorithm on a subset of the units, which can be treated as a subset-PCA algorithm. As an alternative to the subset algorithm, [16] proposed approximating the latent common factors using some variants of EM algorithms based on a hybrid residual matrix (on a full set) obtained through a combination of restricted and unrestricted least squares. As mentioned in [16], the estimation of the regression coefficients β_i differs depending on whether i falls in the true null. That is, β_i should be estimated by the unrestricted least-squares estimate for $i \in \mathcal{I}_1$, and by the restricted least-squares estimate for $i \in \mathcal{I}_0$. To achieve a preliminary classification, they suggested separating the true null and non-null sets based on the ranking of the unadjusted p -values. However, thus far, there is no rigorous theoretical support to explain how the effect of the estimated true null set (either by the empirical posterior probability in [20], by some prior information in [17], or by the unadjusted p -values in [16]) affects the estimate of the latent common factors.

To quantify how the preliminary estimated null set affects the subsequent factor analysis, we propose estimating the latent common factors (i.e., \mathbf{Z}) via a restricted-PCA algorithm on an initial subset; we show that the estimate of the latent common factors is consistent. Specifically, we first estimate all of the regression coefficients β_i in an selected initial subset using the restricted least-squares estimate. The PCA method can be subsequently used to estimate the latent factors based on the induced residual matrix. As long as the initial subset possesses some sparsity properties, the proposed estimation procedure can still be valid; see Proposition 3 for details. Accordingly, we term this method the restricted principal component analysis (Restricted-PCA) algorithm; it can be regarded as a subset PCA algorithm. The new method is outlined below.

- (a) [Determine the initial set]: Select an initial subset \mathcal{I}_* with its cardinality $Card(\mathcal{I}_*) = m_*$, such that the number of alternative hypotheses within \mathcal{I}_* is negligible.
- (b) [Estimate the latent factors]: We first extract the effect of the observable explanatory variables \mathbf{X} by regressing \mathbf{Y}_i on \mathbf{X} under the constraint $\mathbf{A}\beta_i = \mathbf{0}$ for $i \in \mathcal{I}_*$, and obtain the restricted residual matrix

- as $\widehat{\mathcal{E}}_{\mathcal{I}_*}^\top = \{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\}\mathbf{Y}_{\mathcal{I}_*}^\top$. We next define by $\{(\widehat{\Lambda}_e, \widehat{\boldsymbol{\rho}}_e), e = 1, \dots, n\}$ the eigenvalue-eigenvector pairs of $(nm_*)^{-1}\widehat{\mathcal{E}}_{\mathcal{I}_*}^\top\widehat{\mathcal{E}}_{\mathcal{I}_*}$, with $\widehat{\Lambda}_1 \geq \dots \geq \widehat{\Lambda}_n$. The number of factors can be selected by maximizing the eigenvalue ratios as $\widehat{r} = \arg \max_{1 \leq e \leq n-1} \widehat{\Lambda}_e / \widehat{\Lambda}_{e+1}$ ([1]). Accordingly, $\widehat{\mathbf{Z}} = n^{1/2}(\widehat{\boldsymbol{\rho}}_1, \dots, \widehat{\boldsymbol{\rho}}_{\widehat{r}})$.
- (c) [Estimate the unknown parameters]: Using least-squares estimate, for $i \in \mathcal{I}$, $\boldsymbol{\gamma}_i$ and $\sigma_{\boldsymbol{\eta}, ii}$ can be estimated by $\widehat{\boldsymbol{\gamma}}_i = (\widehat{\mathbf{Z}}^\top \widehat{\mathbf{Z}})^{-1} \widehat{\mathbf{Z}}^\top \mathcal{Q}(\mathbf{X}) \mathbf{Y}_i$ and $\widehat{\sigma}_{\boldsymbol{\eta}, ii} = n^{-1} \widehat{\boldsymbol{\eta}}_i^\top \widehat{\boldsymbol{\eta}}_i$, respectively, where $\widehat{\boldsymbol{\eta}}_i = \mathcal{Q}(\widehat{\mathbf{Z}}) \mathcal{Q}(\mathbf{X}) \mathbf{Y}_i$.

According to the theoretical results given in [1] and [29], the number of latent common factors r can be selected by the eigenvalue ratio test in part (b) with a probability tending to one. Consequently, we assume that r is known when investigating the theoretical properties of the Restricted-PCA algorithm. In addition, we impose some regularity conditions, which are not the weakest possible conditions, but those that facilitate the technical derivations.

Conditions.

- C1. Assume $p = O(n^{\zeta_1})$, $r = O(n^{\zeta_2})$, and $\log(m) = O(n^{\zeta_3})$, for some constants $0 \leq \zeta_1 < 1$, $0 \leq \zeta_2 < 1/2$, and $0 < \zeta_3 < 1/2$.
- C2. Assume $\sigma_0^2 \leq \lim_m \inf_i \sigma_{\boldsymbol{\eta}, ii} \leq \lim_m \sup_i \sigma_{\boldsymbol{\eta}, ii} \leq \sigma_1^2$ for some positive constants σ_0^2 and σ_1^2 .
- C3. Assume there exists some positive definite matrix Σ_γ of dimension r such that $m^{-1} \sum_{i=1}^m (\sqrt{r} \boldsymbol{\gamma}_i)(\sqrt{r} \boldsymbol{\gamma}_i)^\top \rightarrow \Sigma_\gamma$, with the eigenvalues of Σ_γ being uniformly bounded from zero and infinity. Moreover, there exists some constant γ_{\max} such that $\lim_{m \rightarrow \infty} \max_i \|\boldsymbol{\gamma}_i\|_2^2 = \gamma_{\max}$.
- C4. Assume $\max_{1 \leq e \leq r} \|\widehat{\boldsymbol{\rho}}_e - \mathbf{Z}_{\cdot e}\|_2 = O_p(m^{-\nu})$ for some positive constant $\nu > 0$, where $\widehat{\boldsymbol{\rho}}_e$ is the e -th eigenvectors of $m^{-1} \mathcal{E}^\top \mathcal{E}$ and $\mathbf{Z}_{\cdot e}$ is the e -th column of \mathbf{Z} .
- C5. Assume that $\text{Card}(\mathcal{I}_* \cap \mathcal{I}_1) / \text{Card}(\mathcal{I}_*) = o(n^{-1-\zeta_2/2})$.

Condition C1 allows a diverging number of factors (observable and latent) to capture all of the correlations among a large number of units, which is more realistic because the correlation structure is not truly low-dimensional under this setting. By condition C1, the number of units m could grow exponentially with the sample size n , so that m can be much larger than n . Condition C2 is commonly assumed in the literature; see, for example, [29]. Condition C3 requires that the \mathcal{L}_2 norm of the factors loadings (i.e., $\|\boldsymbol{\gamma}_i\|_2$) are bounded, so that the variance of the response would not go to infinity. Moreover, $\sqrt{r} \boldsymbol{\gamma}_i$ s can be regarded as m independent realizations from a non-degenerate population with a finite second moment. Conditions C4 is directly borrowed from [12], and can be verified by [3]. Condition C5 requires that the signals on the alternative within the initial subset \mathcal{I}_* are sparse. Under the above the conditions, we demonstrate the following result.

Proposition 3. *Suppose conditions C1–C5 hold, and further assume that η_{ij} is normally distributed. Then, for sufficiently large m and n , we obtain that*

$$\max_{i \in \mathcal{I}} \|\mathcal{P}^\top \widehat{\mathbf{Z}} \widehat{\boldsymbol{\gamma}}_i - \mathcal{P}^\top \{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\} \mathbf{Z} \boldsymbol{\gamma}_i\|_2$$

$$= O_p(\sqrt{r/n}\sqrt{r + \sqrt{r}\sqrt{\log(m)}}) + O_p(m^{-\nu}\sqrt{r}) + O_p(b_{m,n}) \text{ and} \quad (27)$$

$$\begin{aligned} & \max_{i \in \mathcal{I}} |\widehat{\sigma}_{\eta,ii} - \sigma_{\eta,ii}| \\ &= O_p(\sqrt{\log(m)/n} + p/n + r/\sqrt{n}) + O_p(b_{m,n}/\sqrt{n}), \end{aligned} \quad (28)$$

where ν is as in condition C4 and

$$\begin{aligned} b_{m,n} &= \sqrt{rn} \text{Card}(\mathcal{I}_*)^{-1} \sum_{i \in \mathcal{I}_* \cap \mathcal{I}_1} \|(\mathbf{A}\Sigma_{\mathbf{X}}^{-1}\mathbf{A}^\top)^{-1/2}(\mathbf{A}\beta_i)\|_2^2 \\ & \quad + n \text{Card}(\mathcal{I}_*)^{-1} \text{Card}(\mathcal{I}_* \cap \mathcal{I}_1)^{1/2} \sqrt{\sum_{i \in \mathcal{I}_* \cap \mathcal{I}_1} \|(\mathbf{A}\Sigma_{\mathbf{X}}^{-1}\mathbf{A}^\top)^{-1/2}(\mathbf{A}\beta_i)\|_2^2} \end{aligned}$$

with $\Sigma_{\mathbf{X}} \in \mathbb{R}^{p \times p}$ being the asymptotic covariance of the design matrix \mathbf{X} .

The proof can be found in Appendix C. The Restricted-PCA algorithm first adjusts the main effect \mathbf{X} via a restricted least-squares on an selected initial subset, followed by estimating the latent factors using the restricted residual matrix with the row indices corresponding to the initial subset \mathcal{I}_* . In practice, \mathcal{I}_* can be selected by the empirical posterior probability as in [20], the unadjusted p -values as in [16], or the prior information as in [17]. The number of alternative hypotheses in the initial subset \mathcal{I}_* affects the performance of the estimated latent factors, resulting in two additional terms (i.e., $b_{m,n}$) on the asymptotic variance in (27) and (28). Note that both terms are of the order $\text{Card}(\mathcal{I}_* \cap \mathcal{I}_1)/\text{Card}(\mathcal{I}_*) \times n\sqrt{r}$ if the signals $\mathbf{A}\beta_i$'s under the alternative are comparable and bounded, which is asymptotically ignorable under condition C5. Under the local alternative that $\mathbf{A}\beta_i$ is of order $n^{-1/2}$, the term $b_{m,n}$ would go to zero if $\sqrt{r}\text{Card}(\mathcal{I}_* \cap \mathcal{I}_1)/\text{Card}(\mathcal{I}_*)$ goes to zero. This indicates that the choice of the initial subset is flexible.

(27) implies that $\widehat{\mathbf{Z}}\widehat{\gamma}_i$ obtained from the Restricted-PCA algorithm does not converge to its true counterpart $\mathbf{Z}\gamma_i$, but in the form of $\{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\}\mathbf{Z}\gamma_i$. Thus, the Restricted-PCA algorithm produces biased estimators in the sense that it projects $\widehat{\mathbf{Z}}\widehat{\gamma}_i$ into a sub-space spanned by $\mathcal{Q}(\mathbf{X})$ and \mathcal{P} . However, the additional space spanned by the connection matrix \mathcal{P} induced by restricted least-squares is sufficient to verify the sufficient condition of Proposition 2. Compared to previous methods ([20]; [16]; [17]), the restricted PCA algorithm has a formal theoretical justification. The good theoretical properties of the Restricted-PCA algorithm are essential to derive the consistency of the proposed $\widehat{\text{FDP}}_\lambda^{\text{ad}}(t)$; see Theorem 1 for details.

5. Simulation studies

In this section, we gauge the performance of our proposed procedure on the basis of one important metric: the accuracy of the estimated FDP, via simultaneously testing k linear constraints in the large dimensional linear factor model (11). It is well known that lifestyle, geography, and biotic factors are at least as important as genetic divergence in modulating gene expression variation in humans.

This enables us to use one-way ANOVA to model the gene differentiation, where lifestyle, geography, and biotic factors serve as common factors (observed and unobserved) and the idiosyncratic error explains some genetic factors that contribute limit to the observed differentiation such as methylation. To make the simulation more realistic, the model setup and parameters are calibrated from a real data set in a biological environment, detecting differentially expressed genes in peripheral blood leukocyte samples from desert nomadic, mountain agrarian, and coastal urban Moroccan Amazigh individuals ([18]).

Model setup: one-way ANOVA model

- (1) To mimic the biological environment of detecting differentially expressed genes across three geographically defined populations ([18]), we set the large dimensional linear factor model (11) to

$$Y_{ij} = \beta_{i1}X_{j1} + \beta_{i2}X_{j2} + \beta_{i3}X_{j3} + \varepsilon_{ij}, \quad i = 1, \dots, m, \quad j = 1, \dots, n,$$

where $X_{j1} = 1$, $X_{j2} = \mathbf{I}(j \in \text{mountain agrarian})$, and $X_{j3} = \mathbf{I}(j \in \text{desert nomadic})$, with $\mathbf{I}(\cdot)$ denoting an indicator function. Each simulated replication consists of m tests and n samples. The samples are split into three equal sizes of groups parameterized by \mathbf{X} , where the first column parameterizes the intercept, and the second and the third columns denote group membership.

- (2) We are interested in testing whether these three groups show differential expression on each gene i (i.e., \mathbf{Y}_i), that is,

$$H_{0i} : \beta_{i2} = \beta_{i3} = 0 \quad \text{v.s.} \quad H_{1i} : \beta_{i2} \neq 0 \text{ or } \beta_{i3} \neq 0.$$

For each simulated replication, 5% of the tests have nonzeros β_{i2} or β_{i3} with the order C/\sqrt{n} , where C is a constant. The reference level β_{i1} is randomly sampled from a standard normal distribution.

- (3) It is well known that microarray gene expression data are confounded with some batch effects ([19]), such as individual's age and gender, resulting in strong correlation between genes. Many of these batch effects are unfortunately unknown to statisticians or unavailable in many public gene expression data repositories ([9]). Based on these observations, we assume the following latent factor structure on ε_{ij} as

$$\varepsilon_{ij} = \gamma_{i1}Z_{j1} + \gamma_{i2}Z_{j2} + \eta_{ij},$$

to account for the potential dependence among the genes induced by some unobservable batch effects. Here, Z_{j1} and Z_{j2} are discrete random variables, which are *i.i.d.* taking values -1 and 1 with equal probability 0.5 , and γ_{i1} and γ_{i2} are sampled from a standard normal distribution.

- (4) The idiosyncratic errors $\eta_{i,j}$ used to account for limited genetic variation are sampled from a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\Sigma_{\boldsymbol{\eta}} = (\sigma_{\boldsymbol{\eta},i_1i_2})_{m \times m}$ with $\sigma_{\boldsymbol{\eta},i_1i_2} = \rho^{|i_1-i_2|}$ and $\rho = 0.9$. One can easily verify that the AR(1) covariance structure satisfies

the weak dependence assumption (13). The results for other covariance structures, for example, the banded and blocked covariance structures, are numerically similar to those for the AR(1) covariance structure.

To determine the initial subset used in the Restricted-PCA algorithm, we adopt the variance filter proposed in [6] to select 95% of tests with smaller $\hat{\sigma}_{\mathcal{E},ii}$ throughout Sections 5 and 6. Unless otherwise specified, we set the tuning parameter $\lambda = 0.2$ and the threshold $t = 0.01$.

5.1. FDP estimate: From negative dependence to consistency

We first confirm that $\widehat{\text{FDP}}_{\lambda}^{\text{ad}}(t)$ in (19) can consistently estimate the true FDP of the oracle factor-adjusted p -values. For this purpose, we set $\beta_{i2} = \beta_{i3} = C/\sqrt{n}$ for $i \in \mathcal{I}_1$ with $C = 6, 12$, respectively. The number of tests m and the sample size n are calibrated from two real data sets ($m = 10177, n = 46$ in [18] and $m = 4397, n = 166$ in [26]). From Figure 1(a)-(d), when the signal is $6/\sqrt{n}$, the estimated FDP and the true FDP follow a 45 degree line, and as m and n increase, the variation decreases. Remarkably, from Figure 1(e)-(h), the estimated and the true FDPs tend to exhibit a negative dependence pattern as the signals increase. According to the discussion given in [24], the reason can be explained as follows. When the signals are large enough (i.e., $\beta_{i2} = \beta_{i3} = 12/\sqrt{n}$) such that all of the p -values under the alternative are smaller than the threshold $t = 0.01$, the slope of any two replications is approximated roughly as $-\frac{m_0}{m_1} \times t$ (-0.19 in our setting). Rigorous proofs of this negative dependence pattern can found in Appendix D. Moreover, under the strong signal setting, the true FDP is expressed as $\frac{V^{\text{ad}}(t)}{m_1 + V^{\text{ad}}(t)}$, with the mean asymptotically close to $\frac{m_0 t}{m_1 + m_0 t} = \frac{\pi_0 t}{(1 - \pi_0) + \pi_0 t} = 0.16$. This is confirmed by the x-axis of Figure 1(g)-(h). In summary, the negative dependence pattern does not violate the consistency of our FDP estimator, because the result of Theorem 1 is based on the deviation between the estimated and the true values, a quantity that is analogous to the length of the “negative” line, and as m and n increase, the length decreases.

5.2. Factor-adjusted procedure via various algorithms for estimating the latent common factors

To demonstrate the effectiveness of the Restricted-PCA algorithm in estimating the FDP, we adapt the factor-adjusted procedure by replacing the Restricted-PCA algorithm with three algorithms for estimating the latent common factors as follows.

- PCA: similar to the Restricted-PCA algorithm except that the residuals are obtained using unrestricted least-squares.
- IRW-SVA: the iterative re-weighted surrogate variable analysis algorithm given in [20]. For each iteration, it first calculates the empirical posterior probability as a weight vector, that is, $\hat{P}(\beta_i = \mathbf{0}, \gamma_i \neq \mathbf{0} \mid \mathbf{Y}, \mathbf{X}, \hat{\mathbf{Z}}^{(\text{old})})$, and

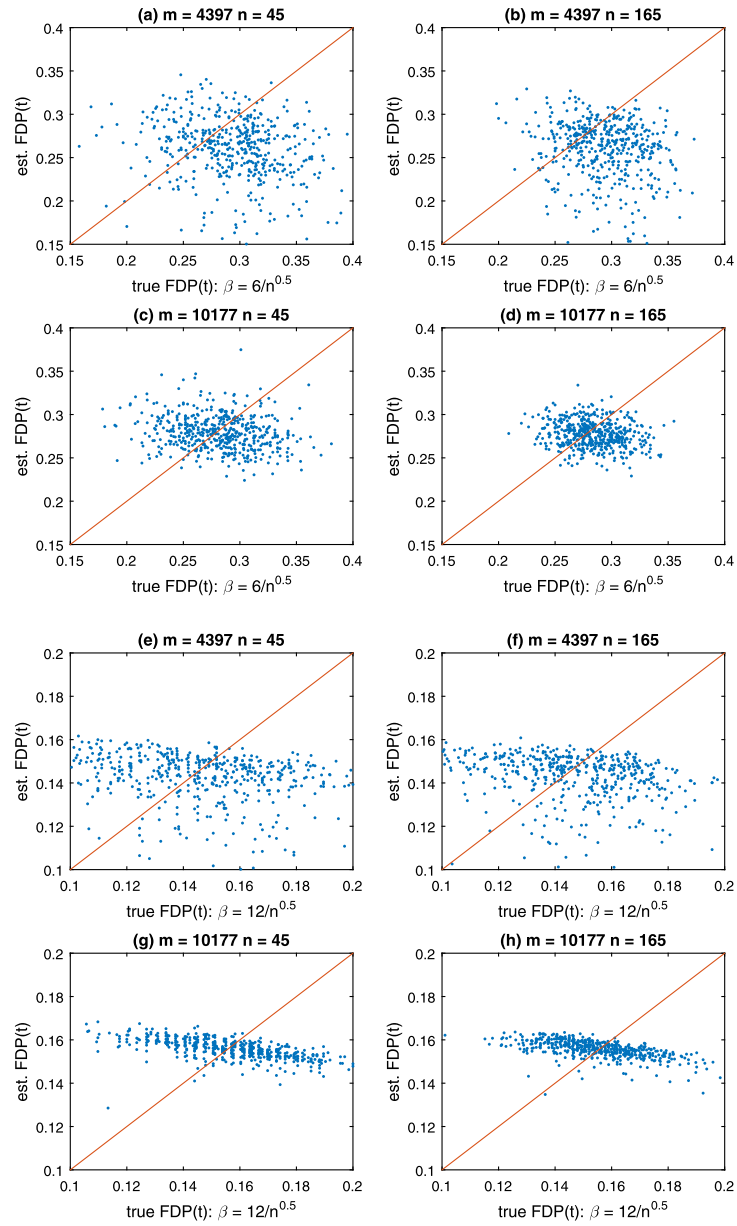


FIG 1. The $\widehat{\text{FDP}}_{\lambda}^{\text{ad}}(t)$ given by the factor-adjusted (FA) procedure with the latent factors estimated by the Restricted-PCA algorithm for various signals $\beta \equiv \beta_{i2} = \beta_{i3} = C/\sqrt{n}$, $C = 6, 12$, respectively. Here, the x-axis is the true FDP of the oracle factor-adjusted p-values with the threshold $t = 0.01$, and the results are based on 500 replications.

then updates $\widehat{\mathbf{Z}}^{(\text{new})}$ using right singular vectors of a weighted matrix obtained by multiplying the weight vector to the corresponding rows of \mathbf{Y} . Once the final estimation $\widehat{\mathbf{Z}}$ is formed, $\widehat{\boldsymbol{\gamma}}_i$ is calculated through the least-squares estimate based on the model $\mathbf{Y}_{i\cdot} = \mathbf{X}\boldsymbol{\beta}_i + \widehat{\mathbf{Z}}\boldsymbol{\gamma}_i + \boldsymbol{\eta}_{i\cdot}$, as if $\widehat{\mathbf{Z}}$ was observed.

FAMT: the factor analysis and multiple testing method given in [16]. Specifically, it first obtains a hybrid residual matrix by combining subsets of restricted and unrestricted residuals, and then applies the standard factor analysis EM algorithm to learn a set of latent factors. The subsets are determined using a preliminary p -value (unadjusted p -value) cut-off of 0.05.

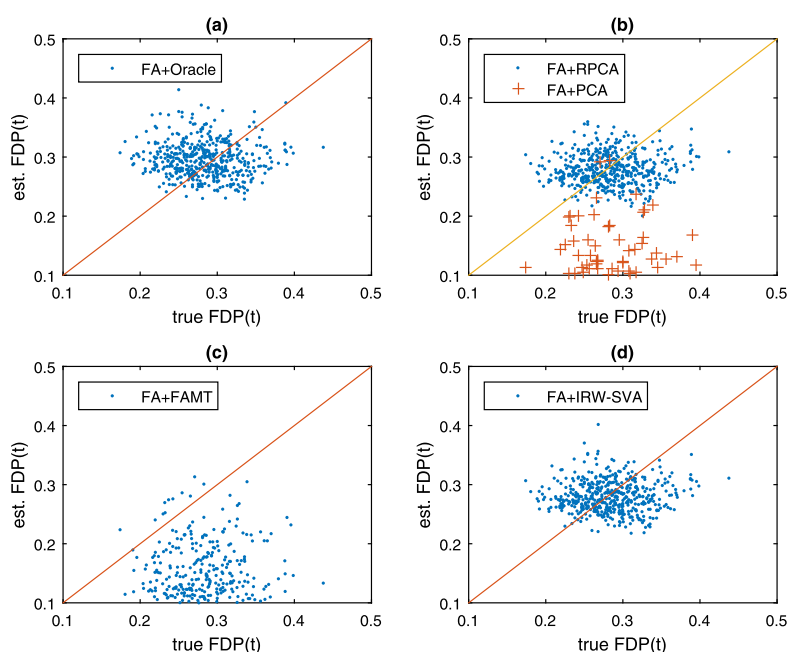


FIG 2. The $\widehat{\text{FDP}}_{\lambda}^{\text{ad}}(t)$ versus $\text{FDP}^{\text{ad}}(t)$ for the factor-adjusted (FA) procedure with the latent factors estimated by five algorithms (Oracle, Restricted-PCA, PCA, FAMT, and IRW-SVA), respectively. Here, $m = 4397, n = 165, \beta_{i2} = \beta_{i3} = 6/\sqrt{n}$ for $i \in \mathcal{I}_1$ with the threshold $t = 0.01$, and the results are based on 500 replications.

To benchmark the performance of the above algorithms, we consider an oracle procedure, where the true latent factors \mathbf{Z} and factor loadings $\boldsymbol{\gamma}_i$'s are directly incorporated into $\widehat{\text{FDP}}_{\lambda}^{\text{ad}}(t)$. This oracle procedure serves as a gold standard because the factor-adjusted procedure with one of these algorithms attempts to get as close as to the oracle one as possible. From Figure 2, the oracle procedure performs the best. Among these three algorithms, PCA estimates $\mathbf{Z}\boldsymbol{\gamma}_i$ by unrestricted least-squares. Because it ignores the fact that $\beta_{i2} = \beta_{i3} = 0$ for $i \in \mathcal{I}_0$, this estimator converges to $\mathcal{Q}(\mathbf{X})\mathbf{Z}\boldsymbol{\gamma}_i$, a quantity orthogonal to the

connection matrix \mathcal{P} . In such a circumstance, the sufficient condition given in Proposition 2 is not satisfied; the result is a lower estimated FDP as shown in Figure 2(b). Compared with PCA, the cut-off value of 0.05 for the unadjusted p -values is too subjective for FAMT to classify the true null and non-null sets in an initial step. As a result, there are certain hypotheses in the true null that are incorrectly identified as of statistical significance. Thus the hybrid residual matrix used in the subsequent factor analysis is unbalanced (with more rows erroneously estimated from unrestricted least-squares). Consequently, the FAMT-based factor-adjusted procedure fails to capture the true FDP when the true value is large; see Figure 2(c) for details. In contrast, from Figure 2(b), the Restricted-PCA algorithm borrows strength from both the restricted least-squares estimate and the sparsity assumption on the selected initial subset, thus is able to assist the factor-adjusted procedure to estimate the true FDP consistently. Note that IRW-SVA obtains as good results as Restricted-PCA, as it estimates space spanned by the latent factors \mathbf{Z} well due to the powerful weight vector, as shown in Figure 2(d).

5.3. Performance of the factor-adjusted procedure beyond normality

The results in Sections 5.1 and 5.2 have shown that the factor-adjusted procedure can consistently estimate the true FDP when the idiosyncratic error $\eta_{i,j}$ is normally distributed. Does this conclusion hold for other types of distributions? Towards this end, we modify the model setup (4) given in Section 5.1 by replacing normal distribution on $\eta_{i,j}$ with double exponential distribution, t distribution with the number of degrees of freedom 3, and chi-squared distribution with 3 degrees of freedom (a centered version so that it has mean 0), respectively. To make a fair comparison, we independently sample $\eta_{i,j}$ from each designated distribution. Figure 3 supports the conjecture made in Section 3.4. The factor-adjusted procedure in conjunction with Restricted-PCA continues to correctly track the true pattern of FDP when the idiosyncratic error is of double exponential distribution or t distribution with low degrees of freedom, both having finite second moment and being symmetric with respect to zero. Surprisingly, under the chi-squared distribution with degrees of freedom 3, the factor-adjusted procedure performs quite well with reasonable variation, revealing that our method is robust against asymmetric distributions.

6. An application: Tackling batch effects in gene expression among ethnic groups

In this section, we apply our method to a microarray data set with strong batch effects. This data set is publicly available in the gene expression omnibus (GEO): GSE5859, which contains the gene expression of 8793 genes for 166 subjects (60 European descent individuals from the Utah pedigrees of the Center d'Etude du Polymorphism Humain (CEU), 41 Han Chinese in Beijing (CHB), 41 Japanese in Tokyo (JPT), and 24 validation samples from Han Chinese of Los Angeles

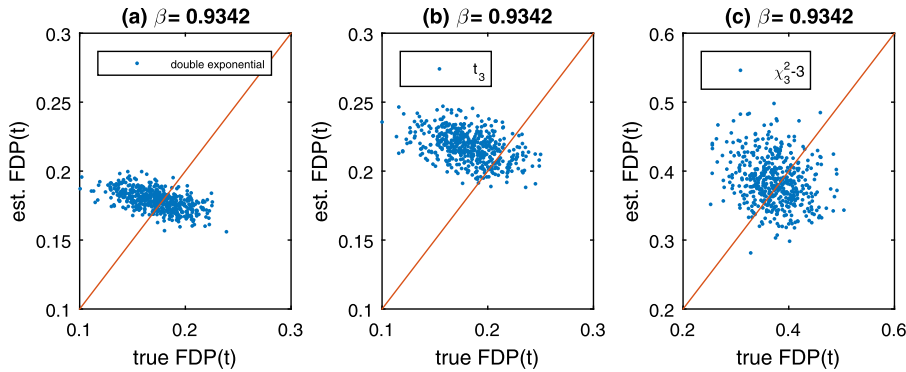


FIG 3. The $\widehat{\text{FDP}}_{\lambda}^{\text{ad}}(t)$ versus $\text{FDP}^{\text{ad}}(t)$ for the factor-adjusted (FA+Restricted-PCA) procedure when the idiosyncratic error η_{ij} is independently sampled from double exponential distribution, t distribution with the number of degrees of freedom 3, and chi-squared distribution with 3 degrees of freedom (a centered version so that it has mean 0), respectively. Here, $m = 4397, n = 165, \beta = \beta_{i2} = \beta_{i3} = 12/\sqrt{n}$ for $i \in \mathcal{I}_1$ with the threshold $t = 0.01$, and the results are based on 500 replications.

(CHLA)). While the previous work in [26] found a large proportion of differentially expressed genes among the populations or specific genetic polymorphisms, some follow-up articles (e.g., [19]) questioned the validity of these results due to strong batch effects. They argued that the pervasive signature of differential expression observed in [26] is a systematic bias introduced during the sample preparation or microarray expression measurement. Specifically, they found that arrays used to measure expression for the CEU individuals were primarily processed from year 2003 to 2004, whereas arrays used to measure expression for the Asian descent individuals were all processed in 2005-2006. Other unobserved variations may exist among the samples. We intend to further tackle the unobservable batch effects in gene expression among the four distinct groups using the factor-adjusted procedure in conjunction with the Restricted-PCA algorithm.

The model setup is identical to that in the simulation studies except that we compare four distinct groups here. Before conducting significance analysis, genes that are not well expressed in tissue must be excluded as genes with little variance in their expression are not proper candidates for differential expression. Moreover, genes with little variance tend to exacerbate the test statistics and distort the associated p -values to zero. Based on this rationale, a preliminary screening step is implemented to eliminate 50% of the genes using an independent filter called overall sample variance; see [6] for details. For the remaining 4397 genes, we calculated the unadjusted p -value, and four types of factor-adjusted p -values based on the PCA, Restricted-PCA, IRW-SVA, or FAMT algorithms, respectively; their corresponding histograms are shown in Figure 4. The unadjusted p -values display troubling behavior, with most p -values shrinking to zero, and this phenomenon further deteriorates with the factor adjustment using the PCA, IRW-SVA, and FAMT algorithms. As a

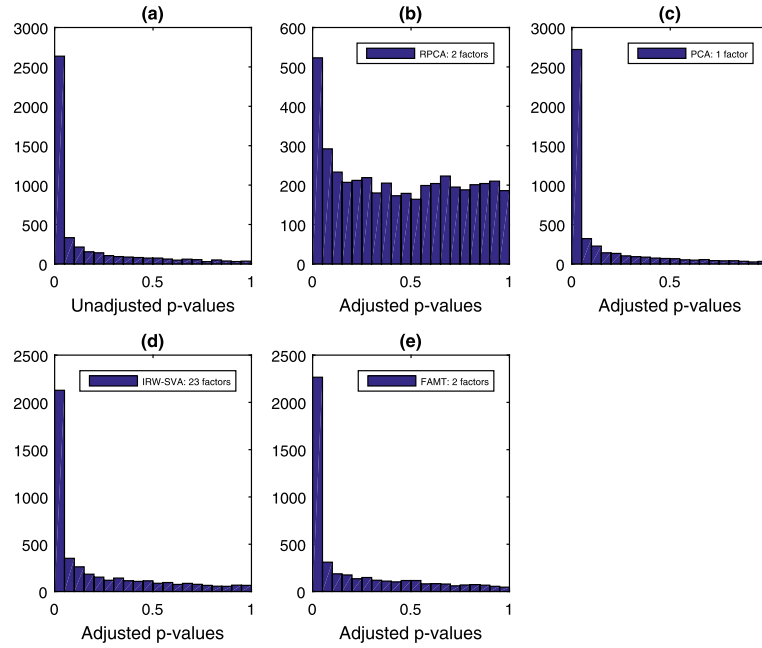


FIG 4. (a): Histogram of the unadjusted p -values resulting from tests of differential expression among four groups without a latent factor structure; (b)-(e): Histogram of the factor-adjusted p -values when the latent factors are estimated from Restricted-PCA, PCA, IRW-SVA, and FAMT algorithms, respectively.

comparison, the factor-adjusted p -values induced by the Restricted-PCA algorithm have a two-component mixture distribution with a peak component at zero for the non-null p -values and a uniform distribution for the true null p -values.

To further gauge the performance of our Restricted-PCA algorithm, we adapt the factor-adjusted procedure as proposed in Section 3.2 with latent factors estimated from Restricted-PCA, PCA, IRW-SVA, and FAMT, respectively. From Figure 5(a), we observe that the factor-adjusted procedure combined with PCA severely underestimates FDP, as it estimates only one principal component and by design this component is orthogonal to the design matrix. The same dilemma occurs for the factor-adjusted procedure based on the FAMT algorithm. In particular, due to the instability of the method for selecting the number of latent factors, IRW-SVA detects 23 principal components as the latent common factors. The redundant components naturally lead to an inaccurate FDP estimator. In contrast, Restricted-PCA estimates two principal components as surrogate variables for batch effects and incorporates them to correct the null distribution of the factor-adjusted p -values. Hence, the Restricted-PCA based factor-adjusted procedure provides more accurate estimation of FDP and finds fewer significant genes at commonly used significance levels, as shown in Figure 5(d).

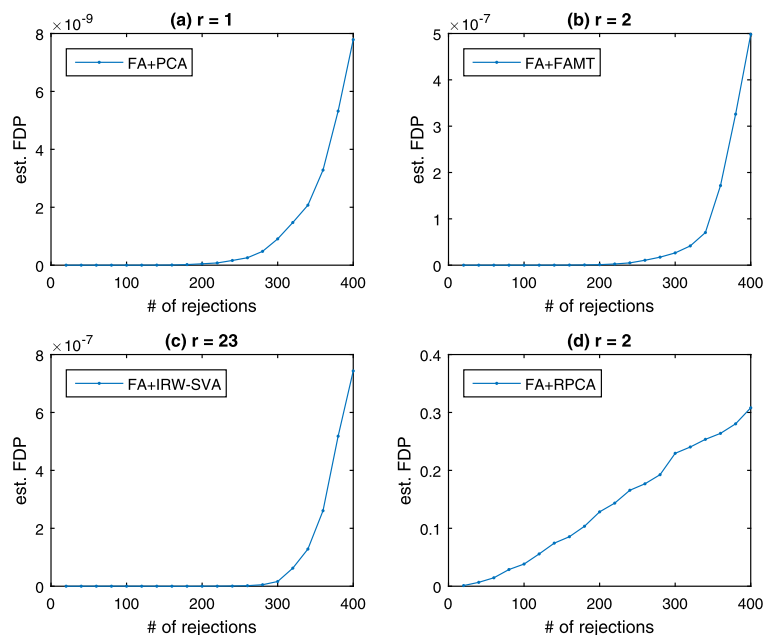


FIG 5. Estimated FDP versus the number of rejected genes from 20 to 400 in multiples of 20 for the factor-adjusted (FA) procedure combined with four various algorithms, respectively. Here, r denotes the estimated number of latent factors by each specific algorithm.

7. Concluding remarks

In this article, we provide a new method for accurately approximating the true FDP when the test statistics follow a multivariate non-central chi-squared distribution with arbitrary covariance dependence; this method shares a similar flavor with previous works based on the assumption of normality ([13]; [12]). Our testing framework is constructed based on k independent multivariate normal random vectors with an arbitrary common covariance dependence structure, where k is the number of degrees of freedom for each chi-squared test statistic. Under assumptions (2) and (3), we theoretically approximate a valid FDP expression for a large number of highly correlated chi-squared test statistics. This new type of approximation is then used to simultaneously test k linear constraints in a large dimensional linear factor model with a latent factor structure. The specific testing problem is solved via a factor-adjusted procedure, which is more robust and powerful than the traditional unadjusted method, because the latent common factors are taken into account at an early stage before any tests are performed, resulting in a new ranking of the hypotheses being tested. We further propose a Restricted-PCA algorithm to estimate the latent common factors, and incorporate them to form a FDP estimator based on the factor-adjusted p -values. Under some mild conditions, we show that this estimator can consistently estimate the true FDP.

To broaden the usefulness of the proposed procedure, we conclude the article by identifying two possible research avenues. Theoretically, it would be of interest to extend our testing framework to accommodate multivariate non-central chi-squared distribution with large degrees of freedom. This situation is particularly useful in large-scale hypotheses testing problems with a nonparametric alternative; see the generalized likelihood ratio principle in [15] for details. Empirically, multiple testing procedures based on chi-squared test statistics are popular in applications, such as the semiparametric inference for identifying significantly activated voxels in brain imaging studies ([33]) and the genome-wide SNP-set testing based on the logistic kernel machine model for assessing the association between the genotyped SNP-set and disease risk ([30]). In both scenarios, nonparametric functions are involved in addition to explanatory covariates, a setting which is more general than the large dimensional linear factor model. How to appropriately connect these extended models to our testing framework and provide valid multiple testing procedures are topics for future study.

Appendix A: Notation index

The following three tables summarize the most common recurring notations, and indicate the section where each symbol is defined. As a convention, we denote by $\|\mathbf{v}\|_2$ the \mathcal{L}_2 norm of a vector \mathbf{v} , and $\|\mathbf{M}\|_F = \{\text{tr}(\mathbf{M}^\top \mathbf{M})\}^{1/2}$ the Frobenius norm of an arbitrary matrix \mathbf{M} , respectively. Moreover, we denoted by $\mathbf{M}_{\mathcal{S}_1, \mathcal{S}_2}$ the submatrix of \mathbf{M} with the rows and columns corresponding to the index sets \mathcal{S}_1 and \mathcal{S}_2 , where \mathcal{S}_1 or \mathcal{S}_2 could be a single index i or j , or even a full set. When \mathcal{S}_1 or \mathcal{S}_2 is a full set, we use a centerdot \cdot to replace it for notational simplicity.

TABLE 1
General notations

Symbol	Description	Section
m	The total number of tests or units	2
\mathcal{I}_0	The true null set	2
\mathcal{I}_1	The alternative set	2
\mathcal{I}	The full set $\{i; i = 1, \dots, m\}$	2
m_0	The number of true null hypotheses	2
m_1	The number of alternative hypotheses	2
n	The number of repeated observations	3
k	The degrees of freedom of the chi-squared test	2
k	The number of linear constraints	3
$i = 1, \dots, m$	The indices used for the tests	3
$j = 1, \dots, n$	The indices used for the observations	3
$\ell = 1, \dots, k$	The indices used for the degrees of freedom	2
p	The number of observable factors	3
r	The number of latent factors	3
$\lim_m a_m$	The limit of the sequence of $\{a_m\}$ exists	2

TABLE 2
Notations related to the model $\mathbf{W}_{i.} = \tilde{\boldsymbol{\mu}}_i + \tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_i + \tilde{\boldsymbol{\eta}}_i$.

Symbol	Description	Section
$\mathbf{W} \in \mathbb{R}^{m \times k}$	The small panel of data matrix for test statistics	2
$\tilde{\mathbf{Z}} \in \mathbb{R}^{k \times r}$	The latent factor matrix attached to \mathbf{W}	2
$\tilde{\boldsymbol{\gamma}}_i \in \mathbb{R}^r$	The factor loadings associated with $\tilde{\mathbf{Z}}$	2
$\tilde{\boldsymbol{\eta}}_i \in \mathbb{R}^{m \times k}$	The idiosyncratic error matrix attached to \mathbf{W}	2
$T_i = \ \mathbf{W}_{i.}\ _2^2$	The test statistic for testing $\tilde{\boldsymbol{\mu}}_i = \mathbf{0}$	2
$P_i = 1 - \chi_k(T_i; 0)$	The associated p -value for testing $\tilde{\boldsymbol{\mu}}_i = \mathbf{0}$	2
$V(t)$	The numbers of false rejections for P_i s at the threshold t	2
$R(t)$	The numbers of total rejections for P_i s at the threshold t	2
$\text{FDP}(t)$	The false discovery proportion for P_i s at the threshold t	2

TABLE 3
Notations related to the model $\mathbf{Y}_{i.} = \mathbf{X}\boldsymbol{\beta}_i + \mathbf{Z}\boldsymbol{\gamma}_i + \boldsymbol{\eta}_i$.

Symbol	Description	Section
$\mathbf{Y} \in \mathbb{R}^{m \times n}$	The observed panel data matrix	3
$\mathbf{X} \in \mathbb{R}^{n \times p}$	The observed design matrix	3
$\mathbf{Z} \in \mathbb{R}^{n \times r}$	The latent factor matrix attached to \mathbf{Y}	3
$\boldsymbol{\gamma}_i \in \mathbb{R}^r$	The factor loadings associated with \mathbf{Z}	3
$\boldsymbol{\eta}_i \in \mathbb{R}^{m \times n}$	The idiosyncratic error matrix attached to \mathbf{Y}	3
T_i^u	The unadjusted test statistic for testing $\mathbf{A}\boldsymbol{\beta}_i = \mathbf{0}$	3
P_i^u	The associated unadjusted p -value	3
T_i^{ad}	The oracle factor-adjusted test statistic for testing $\mathbf{A}\boldsymbol{\beta}_i = \mathbf{0}$	3
P_i^{ad}	The oracle factor-adjusted p -value	3
$V^{\text{ad}}(t)$	The numbers of false rejections for P_i^{ad} s at the threshold t	3
$R^{\text{ad}}(t)$	The numbers of total rejections for P_i^{ad} s at the threshold t	3
$\text{FDP}^{\text{ad}}(t)$	The false discovery proportion for P_i^{ad} s at the threshold t	3
$\hat{T}_i^{\text{ad}}, \hat{P}_i^{\text{ad}}$	The realized versions of T_i^{ad} and P_i^{ad}	3
$\hat{V}^{\text{ad}}(t), \hat{R}^{\text{ad}}(t)$	The realized versions of $V^{\text{ad}}(t)$ and $R^{\text{ad}}(t)$	3
$\widehat{\text{FDP}}_{\lambda}^{\text{ad}}(t)$	The estimator of $\text{FDP}^{\text{ad}}(t)$	3
$\hat{\mathbf{Z}}$	The estimator of \mathbf{Z} by Restricted-PCA	4
\mathcal{I}_*	The initial subset used for Restricted-PCA	4
m_*	The cardinality of \mathcal{I}_*	4

Appendix B: Proofs of Proposition 1, Proposition 2, and Theorem 1

Before proving the main results in Section 2 and Section 3, we first present a strong law of large numbers for weakly dependent data below.

Lemma 1. *Let $\{x_{i,m}, 1 \leq i \leq m, m \geq 1\}$ be a triangular array of real-valued random variables. Suppose the following three conditions hold, that is (1): $|x_{i,m}| \leq 1$ a.s.; (2): $\sum_{m \geq 1} a_m/m < \infty$ with $a_m = E\{(\frac{1}{m} \sum_{i \leq m} x_{i,m})^2\}$; (3): $\lim_{q \rightarrow \infty} \sup_{i \leq m_q} \sup_{s < m_{q+1} - m_q} |x_{i,m_q} - x_{i,m_q+s}| = 0$ a.s., for any subsequence $\{m_q\}_{q=1}^{\infty}$ that $m_q \rightarrow \infty$ and $m_{q+1}/m_q \rightarrow 1$. Then,*

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i \leq m} x_{i,m} = 0 \quad \text{a.s.}$$

Proof of Lemma 1 The result can be obtained using similar techniques as those in Theorem 1 of [21]. By condition (2) that $\sum_{m \geq 1} a_m/m < \infty$ and Lemma 2 of [21], there exists an increasing sequence of integers $\{m_q\}$ such that $\sum_{q \geq 1} a_{m_q} < \infty$ and $m_{q+1}/m_q \rightarrow 1$. By Lemma 3 of [21], we can obtain that $m_q^{-1} \sum_{i \leq m_q} x_{i,m_q} \rightarrow 0$ a.s. Lemma 1 can be proved by applying a suitable maximal inequality to interpolate between $\{m_q, q \geq 1\}$. Specifically, for $m_q \leq m < m_{q+1}$,

$$\begin{aligned} & \left| \frac{1}{m} \sum_{i \leq m} x_{i,m} \right| \\ & \leq \left| \frac{1}{m_q} \sum_{i \leq m_q} x_{i,m} \right| + \sup_{1 \leq s \leq m_{q+1} - m_q} \left| \frac{1}{m_q} \sum_{i=m_q+1}^{m_q+s} x_{i,m} \right| \\ & \leq \left| \frac{1}{m_q} \sum_{i \leq m_q} x_{i,m} - \frac{1}{m_q} \sum_{i \leq m_q} x_{i,m_q} \right| + \left| \frac{1}{m_q} \sum_{i \leq m_q} x_{i,m_q} \right| \\ & \quad + \sup_{s < m_{q+1} - m_q} \left| \frac{1}{m_q} \sum_{i=m_q+1}^{m_q+s} x_{i,m} \right| \\ & \leq \frac{1}{m_q} \sum_{i \leq m_q} |x_{i,m} - x_{i,m_q}| + \left| \frac{1}{m_q} \sum_{i \leq m_q} x_{i,m_q} \right| + \sup_{s < m_{q+1} - m_q} \left| \frac{1}{m_q} \sum_{i=m_q+1}^{m_q+s} x_{i,m} \right| \\ & \leq \sup_{i \leq m_q} \sup_{s < m_{q+1} - m_q} |x_{i,m_q+s} - x_{i,m_q}| + \left| \frac{1}{m_q} \sum_{i \leq m_q} x_{i,m_q} \right| \\ & \quad + \sup_{s < m_{q+1} - m_q} \left| \frac{1}{m_q} \sum_{i=m_q+1}^{m_q+s} x_{i,m} \right|, \end{aligned}$$

where the second term converges to zero almost surely as $m_q \rightarrow \infty$. By condition (1) that $|x_{i,m_q+s}| \leq 1$, the third term is bounded by $\frac{m_{q+1} - m_q}{m_q} \rightarrow 0$. The first term is asymptotically negligible by condition (3) that $x_{i,m_q} - x_{i,m_q+s}$ converges to zero uniformly. Combining the above results together, the proof is completed. ■

Proof of Proposition 1 Whenever necessary, we use superscript $\{ \}^m$ to emphasize that the vector or matrix therein depends on m . We first prove (6). The expectation of $I(P_i \leq t \mid \tilde{\mathbf{Z}})$ is derived as

$$\begin{aligned} & P(P_i \leq t \mid \tilde{\mathbf{Z}}) \\ & = P(T_i \geq \chi_k^{-1}(1-t; 0) \mid \tilde{\mathbf{Z}}) \\ & = P(\|\mathbf{W}_i\|_2^2 \geq \chi_k^{-1}(1-t; 0) \mid \tilde{\mathbf{Z}}) \\ & = P(\|\tilde{\boldsymbol{\mu}}_i + \tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_i + \tilde{\boldsymbol{\eta}}_i\|_2^2 \geq \chi_k^{-1}(1-t; 0) \mid \tilde{\mathbf{Z}}) \\ & = P\left(\left\| \frac{\tilde{\boldsymbol{\mu}}_i + \tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_i}{(\sigma_{\tilde{\boldsymbol{\eta}},ii})^{1/2}} + \frac{\tilde{\boldsymbol{\eta}}_i}{(\sigma_{\tilde{\boldsymbol{\eta}},ii})^{1/2}} \right\|_2^2 \geq \frac{\chi_k^{-1}(1-t; 0)}{\sigma_{\tilde{\boldsymbol{\eta}},ii}} \mid \tilde{\mathbf{Z}}\right) \\ & = 1 - \chi_k\left(\frac{\chi_k^{-1}(1-t; 0)}{\sigma_{\tilde{\boldsymbol{\eta}},ii}}; \lambda_{1i}\right), \end{aligned} \tag{29}$$

where $\lambda_{1i} = \|\tilde{\boldsymbol{\mu}}_{i\cdot} + \tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_i\|_2^2/\sigma_{\tilde{\boldsymbol{\eta}},ii}$. From (29), the goal is to show that

$$R(t)/m - \sum_{i=1}^m \mathbb{P}(P_i \leq t \mid \tilde{\mathbf{Z}})/m \xrightarrow{a.s.} 0. \tag{30}$$

We first show a weaker version of (30), that is,

$$\sum_{i=1}^m \mathbb{I}(P_i \leq t \mid \tilde{\mathbf{Z}})/m - \sum_{i=1}^m \mathbb{P}(P_i \leq t \mid \tilde{\mathbf{Z}})/m \xrightarrow{a.s.} 0. \tag{31}$$

As discussed in Section 2.1, the conditional distribution $\mathbb{P}(P_i \leq t \mid \tilde{\mathbf{Z}})$ depends on m . Hence, we write $x_{i,m} = \mathbb{I}(P_i \leq t \mid \tilde{\mathbf{Z}}) - \mathbb{P}(P_i \leq t \mid \tilde{\mathbf{Z}})$. By Lemma 1, the conclusion (31) holds if we can verify the three conditions of Lemma 1. The first condition is trivial as $x_{i,m} \leq 1$ a.s. To verify the second condition, it suffices to show that

$$a_m = E\left\{\frac{1}{m} \sum_{i \leq m} x_{i,m}\right\}^2 = O(m^{-\delta}), \text{ for some } \delta > 0. \tag{32}$$

Starting from (32), we note that

$$\begin{aligned} & E\left\{\frac{1}{m} \sum_{i \leq m} x_{i,m}\right\}^2 \\ &= \text{var}\left(m^{-1} \sum_{i=1}^m \mathbb{I}(P_i \leq t \mid \tilde{\mathbf{Z}})\right) \\ &= m^{-2} \sum_{i=1}^m \text{var}\left(\mathbb{I}(P_i \leq t \mid \tilde{\mathbf{Z}})\right) \\ &\quad + m^{-2} \sum_{i \neq i^*} \sum \text{cov}\left(\mathbb{I}(P_i \leq t \mid \tilde{\mathbf{Z}}), \mathbb{I}(P_{i^*} \leq t \mid \tilde{\mathbf{Z}})\right). \end{aligned} \tag{33}$$

The first term in (33) is $O_p(m^{-1})$ due to the fact that $\text{var}\left(\mathbb{I}(P_i \leq t \mid \tilde{\mathbf{Z}})\right) \leq \frac{1}{4}$. For the second term, the covariance is given by

$$\begin{aligned} & \mathbb{P}(P_i \leq t, P_{i^*} \leq t \mid \tilde{\mathbf{Z}}) - \mathbb{P}(P_i \leq t \mid \tilde{\mathbf{Z}})\mathbb{P}(P_{i^*} \leq t \mid \tilde{\mathbf{Z}}) \\ &= \mathbb{P}\left(\|\tilde{\boldsymbol{\mu}}_{i\cdot} + \tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_i + \tilde{\boldsymbol{\eta}}_{i\cdot}\|_2^2 \geq \chi_k^{-1}(1-t; 0), \right. \\ &\quad \left. \|\tilde{\boldsymbol{\mu}}_{i^*\cdot} + \tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_{i^*} + \tilde{\boldsymbol{\eta}}_{i^*\cdot}\|_2^2 \geq \chi_k^{-1}(1-t; 0) \mid \tilde{\mathbf{Z}}\right) \\ &\quad - \left\{1 - \chi_k\left(\frac{\chi_k^{-1}(1-t; 0)}{\sigma_{\tilde{\boldsymbol{\eta}},ii}}; \lambda_{1i}\right)\right\} \left\{1 - \chi_k\left(\frac{\chi_k^{-1}(1-t; 0)}{\sigma_{\tilde{\boldsymbol{\eta}},i^*i^*}}; \lambda_{1i^*}\right)\right\}. \end{aligned}$$

Without loss of generality, we assume $\rho_{\tilde{\boldsymbol{\eta}},i^*i^*} > 0$, while the calculation is similar for $\rho_{\tilde{\boldsymbol{\eta}},i^*i^*} \leq 0$. By definition, $\{(\tilde{\eta}_{i\ell}/(\sigma_{\tilde{\boldsymbol{\eta}},ii})^{1/2}, \tilde{\eta}_{i^*\ell}/(\sigma_{\tilde{\boldsymbol{\eta}},i^*i^*})^{1/2})^\top, \ell = 1, \dots, k\}$ are independent and each follows a bivariate normal distribution with mean $\mathbf{0}$

and covariance matrix $\Sigma_{ii^*} = (\sigma_{i',j'})_{2 \times 2}$ with $\sigma_{11} = \sigma_{22} = 1$, $\sigma_{12} = \sigma_{21} = \rho_{\tilde{\eta},ii^*}$. Consequently, we decompose $\tilde{\eta}_{i\cdot}$ and $\tilde{\eta}_{i^*\cdot}$ as

$$\begin{aligned} \tilde{\eta}_{i\cdot}/(\sigma_{\tilde{\eta},ii})^{1/2} &= (\rho_{\tilde{\eta},ii^*})^{1/2}\mathbf{c} + (1 - \rho_{\tilde{\eta},ii^*})^{1/2}\mathbf{u} \quad \text{and} \\ \tilde{\eta}_{i^*\cdot}/(\sigma_{\tilde{\eta},i^*i^*})^{1/2} &= (\rho_{\tilde{\eta},ii^*})^{1/2}\mathbf{c} + (1 - \rho_{\tilde{\eta},ii^*})^{1/2}\mathbf{v}, \end{aligned}$$

where $\mathbf{u} = (u_1, \dots, u_k)^\top$, $\mathbf{v} = (v_1, \dots, v_k)^\top$, and $\mathbf{c} = (c_1, \dots, c_k)^\top$. In particular, $\{u_\ell, v_\ell, c_\ell\}_{\ell=1}^k$ are mutually independent standard normal random variables. Then, the first term of the (i, i^*) -th covariance can be expressed in terms of $\{\mathbf{u}, \mathbf{v}, \mathbf{c}\}$ as,

$$\begin{aligned} &P\left(\|\tilde{\boldsymbol{\mu}}_{i\cdot} + \tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_i + \tilde{\boldsymbol{\eta}}_{i\cdot}\|_2^2 \geq \chi_k^{-1}(1-t; 0), \right. \\ &\quad \left. \|\tilde{\boldsymbol{\mu}}_{i^*\cdot} + \tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_{i^*} + \tilde{\boldsymbol{\eta}}_{i^*\cdot}\|_2^2 \geq \chi_k^{-1}(1-t; 0) \mid \tilde{\mathbf{Z}}\right) \\ &= P\left(\left\|\frac{\tilde{\boldsymbol{\mu}}_{i\cdot} + \tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_i + (\sigma_{\tilde{\eta},ii})^{1/2}(\rho_{\tilde{\eta},ii^*})^{1/2}\mathbf{c}}{(\sigma_{\tilde{\eta},ii})^{1/2}(1 - \rho_{\tilde{\eta},ii^*})^{1/2}} + \mathbf{u}\right\|_2^2 \geq \frac{\chi_k^{-1}(1-t; 0)}{\sigma_{\tilde{\eta},ii}(1 - \rho_{\tilde{\eta},ii^*})}, \right. \\ &\quad \left. \left\|\frac{\tilde{\boldsymbol{\mu}}_{i^*\cdot} + \tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_{i^*} + (\sigma_{\tilde{\eta},i^*i^*})^{1/2}(\rho_{\tilde{\eta},ii^*})^{1/2}\mathbf{c}}{(\sigma_{\tilde{\eta},i^*i^*})^{1/2}(1 - \rho_{\tilde{\eta},ii^*})^{1/2}} + \mathbf{v}\right\|_2^2 \geq \frac{\chi_k^{-1}(1-t; 0)}{\sigma_{\tilde{\eta},i^*i^*}(1 - \rho_{\tilde{\eta},ii^*})} \mid \tilde{\mathbf{Z}}\right) \\ &= \int P\left(\left\|\frac{\tilde{\boldsymbol{\mu}}_{i\cdot} + \tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_i + (\sigma_{\tilde{\eta},ii})^{1/2}(\rho_{\tilde{\eta},ii^*})^{1/2}\mathbf{c}}{(\sigma_{\tilde{\eta},ii})^{1/2}(1 - \rho_{\tilde{\eta},ii^*})^{1/2}} + \mathbf{u}\right\|_2^2 \geq \frac{\chi_k^{-1}(1-t; 0)}{\sigma_{\tilde{\eta},ii}(1 - \rho_{\tilde{\eta},ii^*})}, \right. \\ &\quad \left. \left\|\frac{\tilde{\boldsymbol{\mu}}_{i^*\cdot} + \tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_{i^*} + (\sigma_{\tilde{\eta},i^*i^*})^{1/2}(\rho_{\tilde{\eta},ii^*})^{1/2}\mathbf{c}}{(\sigma_{\tilde{\eta},i^*i^*})^{1/2}(1 - \rho_{\tilde{\eta},ii^*})^{1/2}} + \mathbf{v}\right\|_2^2 \geq \frac{\chi_k^{-1}(1-t; 0)}{\sigma_{\tilde{\eta},i^*i^*}(1 - \rho_{\tilde{\eta},ii^*})} \mid \tilde{\mathbf{Z}}, \mathbf{c}\right) \\ &\quad \phi(\mathbf{c})d\mathbf{c} \\ &= \int \left\{1 - \chi_k\left(\frac{\chi_k^{-1}(1-t; 0)}{\sigma_{\tilde{\eta},ii}(1 - \rho_{\tilde{\eta},ii^*})}; \lambda'_{1i}\right)\right\} \left\{1 - \chi_k\left(\frac{\chi_k^{-1}(1-t; 0)}{\sigma_{\tilde{\eta},i^*i^*}(1 - \rho_{\tilde{\eta},ii^*})}; \lambda'_{1i^*}\right)\right\} \phi(\mathbf{c})d\mathbf{c}, \end{aligned} \tag{34}$$

where

$$\begin{aligned} \phi(\mathbf{c})d\mathbf{c} &= \prod_{\ell=1}^k \phi(c_\ell)dc_1 \dots dc_k, \\ \lambda'_{1i} &= \|(\tilde{\boldsymbol{\mu}}_{i\cdot} + \tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_i)/(\sigma_{\tilde{\eta},ii})^{1/2} + (\rho_{\tilde{\eta},ii^*})^{1/2}\mathbf{c}\|_2^2/(1 - \rho_{\tilde{\eta},ii^*}), \\ \lambda'_{1i^*} &= \|(\tilde{\boldsymbol{\mu}}_{i^*\cdot} + \tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_{i^*})/(\sigma_{\tilde{\eta},i^*i^*})^{1/2} + (\rho_{\tilde{\eta},ii^*})^{1/2}\mathbf{c}\|_2^2/(1 - \rho_{\tilde{\eta},ii^*}). \end{aligned}$$

For notational simplicity, we set $q = (\rho_{\tilde{\eta},ii^*})^{1/2}$, $G_i(q) = \frac{\chi_k^{-1}(1-t; 0)}{\sigma_{\tilde{\eta},ii}(1 - \rho_{\tilde{\eta},ii^*})}$, $\lambda_i(q) = \lambda'_{1i}$, and denote by $f_k(t; \lambda)$ the density function of $\chi_k(t; \lambda)$. By assumption (3) and Cauchy-Schwarz inequality,

$$\frac{\sum_{i \neq i^*} |\rho_{\tilde{\eta},ii^*}|^{1/2}}{m(m-1)} \leq \frac{\{\sum_{i \neq i^*} |\rho_{\tilde{\eta},ii^*}|\}^{1/2} \{\sum_{i \neq i^*} 1\}^{1/2}}{m(m-1)} = O(m^{-\delta/2}) \rightarrow 0. \tag{35}$$

Denote by \mathcal{A} and \mathcal{A}^c the set of indices of (i, i^*) in $\{(i, i^*); i \neq i^*\}$ such that $\rho_{\tilde{\eta},ii^*} \rightarrow 0$ and its complement, respectively. By (35), $\#\{\mathcal{A}^c\}/\{m(m-1)\} =$

$O(m^{-\delta/2})$ with $\#$ being the counting operator. For $(i, i^*) \in \mathcal{A}$, $G_i(q) \rightarrow G_i(0) = \frac{\chi_k^{-1}(1-t;0)}{\sigma_{\tilde{\eta},ii}}$ and $\lambda_i(q) \rightarrow \lambda_{1i}$. For each non-central chi-squared distribution function $\chi_k(\cdot; \cdot)$ appeared in (34) and $(i, i^*) \in \mathcal{A}$, we apply Taylor expansion with respect to q , that is,

$$\begin{aligned} & \chi_k\left(\frac{\chi_k^{-1}(1-t;0)}{\sigma_{\tilde{\eta},ii}(1-\rho_{\tilde{\eta},ii^*})}; \lambda'_{1i}\right) \\ \triangleq & \chi_k(G_i((\rho_{\tilde{\eta},ii^*})^{1/2}); \lambda_i((\rho_{\tilde{\eta},ii^*})^{1/2})) = \chi_k(G_i(q); \lambda_i(q)) \\ = & \chi_k(G_i(0); \lambda_i(0)) \\ & + \left\{ \frac{\partial \chi_k(G_i(q); \lambda_i(q))}{\partial G_i} \frac{\partial G_i(q)}{\partial q} + \frac{\partial \chi_k(G_i(q); \lambda_i(q))}{\partial \lambda_i} \frac{\partial \lambda_i(q)}{\partial q} \right\} \Big|_{q=0} q \\ & + \frac{1}{2} \left[\frac{\partial^2 \chi_k(G_i(q); \lambda_i(q))}{\partial G_i^2} \left\{ \frac{\partial G_i(q)}{\partial q} \right\}^2 + 2 \frac{\partial^2 \chi_k(G_i(q); \lambda_i(q))}{\partial G_i \partial \lambda_i} \frac{\partial G_i(q)}{\partial q} \frac{\partial \lambda_i(q)}{\partial q} \right. \\ & + \left. \frac{\partial^2 \chi_k(G_i(q); \lambda_i(q))}{\partial \lambda_i^2} \left\{ \frac{\partial \lambda_i(q)}{\partial q} \right\}^2 + \frac{\partial \chi_k(G_i(q); \lambda_i(q))}{\partial G_i} \frac{\partial^2 G_i(q)}{\partial q^2} \right. \\ & \left. + \frac{\partial \chi_k(G_i(q); \lambda_i(q))}{\partial \lambda_i} \frac{\partial^2 \lambda_i(q)}{\partial q^2} \right] \Big|_{q=0} q^2 + R(|q|^3), \end{aligned}$$

where

$$\begin{aligned} \frac{\partial \chi_k(G_i(q); \lambda_i(q))}{\partial G_i} &= f_k(G_i(q); \lambda_i(q)), \\ \frac{\partial \chi_k(G_i(q); \lambda_i(q))}{\partial \lambda_i} &= \frac{1}{2} \chi_{k+2}(G_i(q); \lambda_i(q)) - \frac{1}{2} \chi_k(G_i(q); \lambda_i(q)), \\ \frac{\partial^2 \chi_k(G_i(q); \lambda_i(q))}{\partial G_i \partial \lambda_i} &= \frac{1}{2} f_{k+2}(G_i(q); \lambda_i(q)) - \frac{1}{2} f_k(G_i(q); \lambda_i(q)), \\ \frac{\partial^2 \chi_k(G_i(q); \lambda_i(q))}{\partial G_i^2} &= \frac{1}{2} f_{k-2}(G_i(q); \lambda_i(q)) - \frac{1}{2} f_k(G_i(q); \lambda_i(q)), \\ \frac{\partial^2 \chi_k(G_i(q); \lambda_i(q))}{\partial \lambda_i^2} &= \frac{1}{4} \chi_{k+4}(G_i(q); \lambda_i(q)) - \frac{1}{2} \chi_{k+2}(G_i(q); \lambda_i(q)) \\ & \quad + \frac{1}{4} \chi_k(G_i(q); \lambda_i(q)), \\ \frac{\partial G_i(q)}{\partial q} &= \frac{\chi_k^{-1}(1-t;0)}{\sigma_{\tilde{\eta},ii}} \times \frac{2q}{(1-q^2)^2}, \\ \frac{\partial^2 G_i(q)}{\partial q^2} &= \frac{\chi_k^{-1}(1-t;0)}{\sigma_{\tilde{\eta},ii}} \times \left\{ \frac{2}{(1-q^2)^2} + \frac{8q^2}{(1-q^2)^3} \right\}, \\ \frac{\partial \lambda_i(q)}{\partial q} &= \frac{2}{1-q^2} \times \mathbf{c}^\top \{(\tilde{\boldsymbol{\mu}}_{i\cdot} + \tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_i)/(\sigma_{\tilde{\eta},ii})^{1/2} + \mathbf{c}q\} \\ & \quad + \frac{2q}{(1-q^2)^2} \times \|(\tilde{\boldsymbol{\mu}}_{i\cdot} + \tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_i)/(\sigma_{\tilde{\eta},ii})^{1/2} + \mathbf{c}q\|_2^2, \\ \frac{\partial^2 \lambda_i(q)}{\partial q^2} &= \frac{2}{1-q^2} \times \|\mathbf{c}\|_2^2 \\ & \quad + \frac{8q}{(1-q^2)^2} \times \mathbf{c}^\top \{(\tilde{\boldsymbol{\mu}}_{i\cdot} + \tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_i)/(\sigma_{\tilde{\eta},ii})^{1/2} + \mathbf{c}q\} \\ & \quad + \left\{ \frac{2}{(1-q^2)^2} + \frac{8q^2}{(1-q^2)^3} \right\} \end{aligned}$$

$$\times \|(\tilde{\boldsymbol{\mu}}_{i\cdot} + \tilde{\mathbf{Z}}\tilde{\boldsymbol{\gamma}}_i)/(\sigma_{\tilde{\boldsymbol{\eta}},ii})^{1/2} + \mathbf{c}q\|_2^2.$$

Particularly, $R(|q|^3) = f_R(w_1, \dots, w_k)O(|q|^3)$ with $f_R(w_1, \dots, w_k)$ being a polynomial function of w_1, \dots, w_k with the highest order as 3. Using the above formulas and the fact that $\int c_\ell \phi(c_\ell) dc_\ell = 0$, $\int c_\ell^2 \phi(c_\ell) dc_\ell = 1$, (34) can be further simplified as

$$\begin{aligned} & \left\{1 - \chi_k\left(\frac{\chi_k^{-1}(1-t;0)}{\sigma_{\tilde{\boldsymbol{\eta}},ii}}; \lambda_{1i}\right)\right\} \left\{1 - \chi_k\left(\frac{\chi_k^{-1}(1-t;0)}{\sigma_{\tilde{\boldsymbol{\eta}},i^*i^*}}; \lambda_{1i^*}\right)\right\} \\ & + \left[\left\{\frac{1}{2}\chi_{k+2}(G_i(0); \lambda_{1i}) - \frac{1}{2}\chi_k(G_i(0); \lambda_{1i})\right\} \right. \\ & \quad \times \left. \left\{\frac{1}{2}\chi_{k+2}(G_{i^*}(0); \lambda_{1i^*}) - \frac{1}{2}\chi_k(G_{i^*}(0); \lambda_{1i^*})\right\} \left\{4k(\lambda_{1i}\lambda_{1i^*})^{1/2}\right\} \right] q^2 \\ & - \left\{1 - \chi_k(G_{i^*}(0); \lambda_{1i^*})\right\} \times \\ & \frac{1}{2} \left[\left\{\frac{1}{4}\chi_{k+4}(G_i(0); \lambda_{1i}) - \frac{1}{2}\chi_{k+2}(G_i(0); \lambda_{1i}) + \frac{1}{4}\chi_k(G_i(0); \lambda_{1i})\right\} \{4k\lambda_{1i}\} \right. \\ & \quad + f_k(G_i(0); \lambda_{1i})\{2G_i(0)\} \\ & \quad \left. + \left\{\frac{1}{2}\chi_{k+2}(G_i(0); \lambda_{1i}) - \frac{1}{2}\chi_k(G_i(0); \lambda_{1i})\right\} \{2k + 2\lambda_{1i}\} \right] q^2 \\ & - \left\{1 - \chi_k(G_i(0); \lambda_{1i})\right\} \times \\ & \frac{1}{2} \left[\left\{\frac{1}{4}\chi_{k+4}(G_{i^*}(0); \lambda_{1i^*}) - \frac{1}{2}\chi_{k+2}(G_{i^*}(0); \lambda_{1i^*}) + \frac{1}{4}\chi_k(G_{i^*}(0); \lambda_{1i^*})\right\} \right. \\ & \quad \times \{4k\lambda_{1i^*}\} + f_k(G_{i^*}(0); \lambda_{1i^*})\{2G_{i^*}(0)\} \\ & \quad \left. + \left\{\frac{1}{2}\chi_{k+2}(G_{i^*}(0); \lambda_{1i^*}) - \frac{1}{2}\chi_k(G_{i^*}(0); \lambda_{1i^*})\right\} \{2k + 2\lambda_{1i^*}\} \right] q^2 + R(|q|^3). \end{aligned}$$

By conditions (i), (ii), and (iii), all of the terms associated with q^2 are upper bounded and the bound does not depend on m , implying that the (i, i^*) -th covariance is $O(|\rho_{\tilde{\boldsymbol{\eta}},ii^*}|)$. As a result, (33) is bounded by

$$O\left(m^{-1} + m^{-2} \sum_{(i,i^*) \in \mathcal{A}} \sum |\rho_{\tilde{\boldsymbol{\eta}},ii^*}| + \#\{\mathcal{A}^c\}/m^2\right) = O(m^{-\delta/2}).$$

This completes the proof of (32), and the condition (2) is satisfied. Now we verify that the condition (3) holds. According to Lemma 2 of [21], there exists a sequence of integers $\{m_p\}$ such that $\sum_p a_{m_p} < \infty$ and $m_{p+1}/m_p \rightarrow 1$. By the continuity of the noncentral chi-squared distributions, $|x_{i,m_p+s} - x_{i,m_p}|$ can be bounded by

$$O\left(\left\|\tilde{\mathbf{Z}}^{m_p+s}\tilde{\boldsymbol{\gamma}}_i^{m_p+s}\right\|_2^2 - \left\|\tilde{\mathbf{Z}}^{m_p}\tilde{\boldsymbol{\gamma}}_i^{m_p}\right\|_2^2 + \left\|\sigma_{\tilde{\boldsymbol{\eta}},ii}^{m_p+s}\right\|_2^2 - \left\|\sigma_{\tilde{\boldsymbol{\eta}},ii}^{m_p}\right\|_2^2\right).$$

Condition (ii) implies that

$$\lim_{p \rightarrow \infty} \sup_{i \leq m_p} \sup_{s < m_{p+1} - m_p} \left| \left\|\sigma_{\tilde{\boldsymbol{\eta}},ii}^{m_p+s}\right\|_2^2 - \left\|\sigma_{\tilde{\boldsymbol{\eta}},ii}^{m_p}\right\|_2^2 \right| = 0 \quad a.s.$$

Similarly, by condition (iii), we can obtain that

$$\lim_{p \rightarrow \infty} \sup_{i \leq m_p} \sup_{s < m_{p+1} - m_p} \left| \|\tilde{\mathbf{Z}}^{m_p+s} \tilde{\boldsymbol{\gamma}}_i^{m_p+s}\|_2^2 - \|\tilde{\mathbf{Z}}^{m_p} \tilde{\boldsymbol{\gamma}}_i^{m_p}\|_2^2 \right| = 0 \quad a.s.$$

Combining the above two results, the condition (3) holds. As a result, (31) is verified. By a similar argument in [13], (30) also holds; this directly completes the proof of (6). The proof of (5) is analogous, hence we omit the details here.

Now we turn to prove (7). For ease of presentation, we denote by $\bar{R}(t)$ and $\bar{V}(t)$ as $\sum_{i=1}^m \left\{ 1 - \chi_k \left(\frac{\chi_k^{-1}(1-t;0)}{\sigma_{\eta,ii}}; \lambda_{1i} \right) \right\}$ and $\sum_{i \in \mathcal{I}_0} \left\{ 1 - \chi_k \left(\frac{\chi_k^{-1}(1-t;0)}{\sigma_{\eta,ii}}; \lambda_{0i} \right) \right\}$, respectively. Then,

$$\begin{aligned} & \lim_{m \rightarrow \infty} \left[\text{FDP}(t) - \frac{\sum_{i \in \mathcal{I}_0} \left\{ 1 - \chi_k \left(\frac{\chi_k^{-1}(1-t;0)}{\sigma_{\eta,ii}}; \lambda_{0i} \right) \right\}}{\sum_{i=1}^m \left\{ 1 - \chi_k \left(\frac{\chi_k^{-1}(1-t;0)}{\sigma_{\eta,ii}}; \lambda_{1i} \right) \right\}} \right] \\ &= \lim_{m \rightarrow \infty} \left\{ \frac{V(t)}{R(t)} - \frac{\bar{V}(t)}{\bar{R}(t)} \right\} \\ &= \lim_{m \rightarrow \infty} \frac{\{V(t)/m_0\} [\{\bar{R}(t) - R(t)\}/m] + \{R(t)/m\} [\{V(t) - \bar{V}(t)\}/m_0]}{R(t)\bar{R}(t)/(mm_0)} \\ &= 0, \quad a.s. \end{aligned}$$

where the last equation holds by the results (5) and (6), $\bar{R}(t)/m \geq t/2$, and the fact that $R(t)/m$ and $V(t)/m_0$ are bounded by 1. The proof of Proposition 1 is completed. \blacksquare

Proof of Proposition 2 We first verify the conclusion in (24). Motivated by the structure of \hat{P}_i^{ad} , we write $\hat{P}_i^{\text{ad}} = 1 - \chi_k(\|\widehat{\mathbf{W}}_{i\cdot}^{\text{ad}}\|_2^2; 0)$ with $\widehat{\mathbf{W}}_{i\cdot}^{\text{ad}} \triangleq \{\widehat{\sigma}_{\eta,ii} \mathbf{A}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top\}^{-1/2} \{\mathbf{A} \hat{\boldsymbol{\beta}}_i - \mathbf{A}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \widehat{\mathbf{Z}} \hat{\boldsymbol{\gamma}}_i\}$. By more derivations, $\widehat{\mathbf{W}}_{i\cdot}^{\text{ad}}$ is expressed as

$$\begin{aligned} \widehat{\mathbf{W}}_{i\cdot}^{\text{ad}} &= \sigma_{\eta,ii}^{1/2} / \widehat{\sigma}_{\eta,ii}^{1/2} \{\sigma_{\eta,ii} \mathbf{A}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top\}^{-1/2} \{\mathbf{A} \hat{\boldsymbol{\beta}}_i - \mathbf{A}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Z} \boldsymbol{\gamma}_i\} \\ &\quad - \widehat{\sigma}_{\eta,ii}^{-1/2} \mathcal{P}^\top (\widehat{\mathbf{Z}} \hat{\boldsymbol{\gamma}}_i - \mathbf{Z} \boldsymbol{\gamma}_i) \\ &= \sigma_{\eta,ii}^{1/2} / \widehat{\sigma}_{\eta,ii}^{1/2} \mathbf{W}_{i\cdot}^{\text{ad}} - \widehat{\sigma}_{\eta,ii}^{-1/2} \mathcal{P}^\top (\widehat{\mathbf{Z}} \hat{\boldsymbol{\gamma}}_i - \mathbf{Z} \boldsymbol{\gamma}_i). \end{aligned} \quad (36)$$

For ease of presentation, let $a = \sigma_{\eta,ii}^{-1/2} / \widehat{\sigma}_{\eta,ii}^{-1/2}$ and $b = \|\widehat{\sigma}_{\eta,ii}^{-1/2} \mathcal{P}^\top (\widehat{\mathbf{Z}} \hat{\boldsymbol{\gamma}}_i - \mathbf{Z} \boldsymbol{\gamma}_i)\|_2$. By the sufficient condition, we have $a \rightarrow_p 1$ and $b \rightarrow_p 0$ uniformly for $i \in \mathcal{I}$. By definition, for any $\varepsilon > 0$, there exist a $\delta > 0$ and sufficiently large \bar{m} and \bar{n} such that when $m > \bar{m}$ and $n > \bar{n}$, we have $\text{P}(\Omega_1) \geq 1 - \varepsilon$ and $\text{P}(\Omega_2) \geq 1 - \varepsilon$ with $\Omega_1 = \{\omega; |a(\omega) - 1| < \delta\}$ and $\Omega_2 = \{\omega; |b(\omega)| < \delta\}$. Then, when $m > \bar{m}$, $n > \bar{n}$, and $\omega \in \Omega_1 \cap \Omega_2$,

$$\begin{aligned} & m_0^{-1} \sum_{i \in \mathcal{I}_0} \text{I} \left(\|\widehat{\mathbf{W}}_{i\cdot}^{\text{ad}}\|_2 \geq \sqrt{\chi_k(1-t;0)} \right) \\ & \leq m_0^{-1} \sum_{i \in \mathcal{I}_0} \text{I} \left(a \|\mathbf{W}_{i\cdot}^{\text{ad}}\|_2 + b \geq \sqrt{\chi_k(1-t;0)} \right) \end{aligned}$$

$$\begin{aligned}
&\leq m_0^{-1} \sum_{i \in \mathcal{I}_0} \mathbb{I}\left((1 + \delta)\|\mathbf{W}_{i \cdot}^{\text{ad}}\|_2 + \delta \geq \sqrt{\chi_k(1 - t; 0)}\right) \\
&\leq m_0^{-1} \sum_{i \in \mathcal{I}_0} \mathbb{I}\left(\|\mathbf{W}_{i \cdot}^{\text{ad}}\|_2 \geq \sqrt{\chi_k(1 - t; 0)}/(1 + \delta) - \delta/(1 + \delta)\right). \quad (37)
\end{aligned}$$

Similarly, when $m > \bar{m}$, $n > \bar{n}$, and $\omega \in \Omega_1 \cap \Omega_2$, we have

$$\begin{aligned}
&m_0^{-1} \sum_{i \in \mathcal{I}_0} \mathbb{I}\left(\|\widehat{\mathbf{W}}_{i \cdot}^{\text{ad}}\|_2 \geq \sqrt{\chi_k(1 - t; 0)}\right) \\
&\geq m_0^{-1} \sum_{i \in \mathcal{I}_0} \mathbb{I}\left(\|\mathbf{W}_{i \cdot}^{\text{ad}}\|_2 \geq \sqrt{\chi_k(1 - t; 0)}/(1 - \delta) + \delta/(1 - \delta)\right). \quad (38)
\end{aligned}$$

By letting $\varepsilon \rightarrow 0$ and $\delta \rightarrow 0$, (37) and (38) imply that, with a probability tending to one,

$$\begin{aligned}
&\lim_{m, n \rightarrow \infty} m_0^{-1} \sum_{i \in \mathcal{I}_0} \mathbb{I}\left(\|\widehat{\mathbf{W}}_{i \cdot}^{\text{ad}}\|_2 \geq \sqrt{\chi_k(1 - t; 0)}\right) \\
&= \lim_{m, n \rightarrow \infty} m_0^{-1} \sum_{i \in \mathcal{I}_0} \mathbb{I}\left(\|\mathbf{W}_{i \cdot}^{\text{ad}}\|_2 \geq \sqrt{\chi_k(1 - t; 0)}\right). \quad (39)
\end{aligned}$$

Thus $\widehat{V}^{\text{ad}}(t)/m_0 - V^{\text{ad}}(t)/m_0$ converges to zero almost surely. (25) can be derived in a similar way as (24), hence the proof is omitted. \blacksquare

Proof of Theorem 1. By decomposition, for any fixed $t > 0$,

$$\begin{aligned}
&|\widehat{\text{FDP}}_{\lambda}^{\text{ad}}(t) - c\text{FDP}^{\text{ad}}(t)| \\
&= \frac{|\widehat{\text{FDP}}_{\lambda}^{\text{ad}}(t) - cV^{\text{ad}}(t)/m|}{\{cV^{\text{ad}}(t)/m\}} \frac{|\{R^{\text{ad}}(t) - \widehat{R}^{\text{ad}}(t)\}/m|}{\{R^{\text{ad}}(t)/m\} \{\widehat{R}^{\text{ad}}(t)/m\}} \\
&\quad + \frac{|\{R^{\text{ad}}(t)/m\} \{m\widehat{\pi}_0(\lambda)t - cV^{\text{ad}}(t)\}/m|}{\{R^{\text{ad}}(t)/m\} \{\widehat{R}^{\text{ad}}(t)/m\}} \\
&\leq \frac{c|\{R^{\text{ad}}(t) - \widehat{R}^{\text{ad}}(t)\}/m| + |\{m\widehat{\pi}_0(\lambda)t - cV^{\text{ad}}(t)\}/m|}{\{R^{\text{ad}}(t)/m\} \{\widehat{R}^{\text{ad}}(t)/m\}}. \quad (40)
\end{aligned}$$

From (23), $R^{\text{ad}}(t)/m \geq t/2$ almost surely when m is sufficiently large. Moreover, $|\{R^{\text{ad}}(t) - \widehat{R}^{\text{ad}}(t)\}/m|$ is negligible by the sufficient condition and Proposition 2. Thus, to bound (40), it suffices to consider $|\{m\widehat{\pi}_0(\lambda)t - cV^{\text{ad}}(t)\}/m|$ and $\widehat{R}^{\text{ad}}(t)/m$. By (24), (25), and Proposition 2, it is straightforward to show that

$$\begin{aligned}
\lim_{m \rightarrow \infty} \widehat{V}^{\text{ad}}(t)/m &= t, \quad a.s. \\
\lim_{m \rightarrow \infty} \widehat{R}^{\text{ad}}(t)/m &= \{\pi_0 \times t + (1 - \pi_0) \times F_1(t)\}. \quad a.s.
\end{aligned}$$

As a result, $\widehat{R}^{\text{ad}}(t)/m \geq t/2$ as m is sufficiently large. Similarly,

$$\widehat{\pi}_0(\lambda)t = \{m - \widehat{R}^{\text{ad}}(\lambda)\}t/\{m(1 - \lambda)\} = [1 - \widehat{R}^{\text{ad}}(\lambda)/m]t/(1 - \lambda)$$

converges to $[\pi_0 + (1 - \pi_0)\{1 - F_1(\lambda)\}/(1 - \lambda)]t = c\pi_0 t$ almost surely. This together with (22) implies that $|\{m\hat{\pi}_0(\lambda)t - cV^{\text{ad}}(t)\}/m|$ converges to zero almost surely, which completes the proof of Theorem 1. \blacksquare

Appendix C: Consistency of the Restricted-PCA algorithm

For notational simplicity, we denote by $\check{\mathbf{Z}}_j, \check{\Lambda}_e, \check{\gamma}_i, \check{\sigma}_{\eta,ii}$ and $\bar{\mathbf{Z}}_j, \bar{\Lambda}_e, \bar{\gamma}_i, \bar{\sigma}_{\eta,ii}$ the associated estimators of $\mathbf{Z}_j, \Lambda_e, \gamma_i, \sigma_{\eta,ii}$ based on the error matrices $\{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\}\mathcal{E}_{\mathcal{I}_*}^\top$ and $\mathcal{E}_{\mathcal{I}_*}^\top$ in the Restricted-PCA algorithm, respectively.

Before proving Proposition 3, we present four useful lemmas below. Lemma 2 can be obtained directly through Bonferroni inequality. Lemma 3 consists of the so-called *Weyl’s Theorem* and *sin(θ) Theorem* used in [14]. Hence, the proofs of both lemmas are omitted here. The proofs of the last two lemmas involve more tedious derivations than that in [3], [14], and [12], because we allow p and r to grow to infinity at some polynomial rates. For completeness, we supply detailed proofs for Lemma 4 and Lemma 5 in this Appendix.

Lemma 2. *Let $\{\chi_{i,n}\}_{i=1}^m$ be chi-squared distributed random variables with degrees of freedom n , then we should have*

$$\max_{i=1, \dots, m} |n^{-1}\chi_{i,n} - 1| = O_p(\sqrt{\log(m)/n}).$$

Lemma 3. *For any positive-semidefinite matrices \mathbf{M}_1 and \mathbf{M}_2 , we have*

$$\begin{aligned} |\Lambda_{i,\mathbf{M}_1} - \Lambda_{i,\mathbf{M}_2}| &\leq \|\mathbf{M}_1 - \mathbf{M}_2\|_F \quad \text{and} \\ \|\mathbf{q}_{i,\mathbf{M}_1} - \mathbf{q}_{i,\mathbf{M}_2}\|_2 &\leq \frac{\sqrt{2}\|\mathbf{M}_1 - \mathbf{M}_2\|_F}{\min\{|\Lambda_{i-1,\mathbf{M}_1} - \Lambda_{i,\mathbf{M}_2}|, |\Lambda_{i+1,\mathbf{M}_1} - \Lambda_{i,\mathbf{M}_2}|\}}, \end{aligned}$$

where $\Lambda_{i,\mathbf{M}}$ ’s and $\mathbf{q}_{i,\mathbf{M}}$ ’s are the eigenvalues and eigenvectors of any positive-semidefinite matrix \mathbf{M} .

Lemma 4. *Under conditions C1–C3, and assuming that η_{ij} is normally distributed, we can obtain that*

$$\max_{i \in \mathcal{I}} |\check{\sigma}_{\eta,ii} - \sigma_{\eta,ii}| = O_p\left(\sqrt{\log(m)/n} + p/n + r/\sqrt{n}\right).$$

Lemma 5. *Suppose conditions C1–C4 hold, and further assume that η_{ij} is normally distributed. Then, we can obtain that*

$$\begin{aligned} &\max_{i \in \mathcal{I}} \|\mathcal{P}^\top \check{\mathbf{Z}}\check{\gamma}_i - \mathcal{P}^\top \{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\}\mathbf{Z}\gamma_i\|_2 \\ &= O_p\left(\sqrt{r/n}\sqrt{r} + \sqrt{r}\sqrt{\log(m)}\right) + O_p(m^{-\nu}\sqrt{r}). \end{aligned}$$

Proof of Lemma 4 We prove the lemma in the following two steps. In the first step, we demonstrate that $\max_{i \in \mathcal{I}} |\bar{\sigma}_{\eta,ii} - \sigma_{\eta,ii}| = O_p\{\sqrt{\log(m)/n} + (r + p)/n + r/\sqrt{n} + r^2/n\}$, followed by extending the result to $\check{\sigma}_{\eta,ii}$ in the second step.

We first show that $\max_{i \in \mathcal{I}} |\bar{\sigma}_{\boldsymbol{\eta}, ii} - \sigma_{\boldsymbol{\eta}, ii}| = O_p\{\sqrt{\log(m)/n} + (r+p)/n + r/\sqrt{n} + r^2/n\}$. Note that $\bar{\sigma}_{\boldsymbol{\eta}, ii} = n^{-1} \bar{\boldsymbol{\eta}}_i^\top \bar{\boldsymbol{\eta}}_i$, with $\bar{\boldsymbol{\eta}}^\top = \mathcal{Q}(\bar{\mathbf{Z}}) \mathcal{Q}(\mathbf{X}) \mathcal{E}^\top$, where $\bar{\mathbf{Z}}$ and $\mathcal{Q}(\bar{\mathbf{Z}})$ are analogously defined in Section 4. As a result, we have $\bar{\sigma}_{\boldsymbol{\eta}, ii} = n^{-1} \mathcal{E}_i^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\bar{\mathbf{Z}}) \mathcal{Q}(\mathbf{X}) \mathcal{E}_i$, which leads to

$$\begin{aligned} & \left| \bar{\sigma}_{\boldsymbol{\eta}, ii} - \sigma_{\boldsymbol{\eta}, ii} \right| \\ &= \left| n^{-1} \mathcal{E}_i^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\bar{\mathbf{Z}}) \mathcal{Q}(\mathbf{X}) \mathcal{E}_i - n^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\mathbf{Z}) \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i \right. \\ & \quad \left. + n^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\mathbf{Z}) \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i - \sigma_{\boldsymbol{\eta}, ii} \right| \\ &\leq \left| n^{-1} \mathcal{E}_i^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\bar{\mathbf{Z}}) \mathcal{Q}(\mathbf{X}) \mathcal{E}_i - n^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\mathbf{Z}) \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i \right| \\ & \quad + \left| n^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\mathbf{Z}) \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i - \sigma_{\boldsymbol{\eta}, ii} \right|. \end{aligned} \quad (41)$$

We then consider the above two parts separately. We first consider the second part. Note that $\boldsymbol{\eta}_i$ follows a multivariate normal distribution with mean zero and covariance matrix $\sigma_{\boldsymbol{\eta}, ii} \mathbf{I}_n$. As a result, $\boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}, \mathbf{Z}) \boldsymbol{\eta}_i / \sigma_{\boldsymbol{\eta}, ii}$ follows a chi-square distribution of degree $n - r - p$. Thus, according to Lemma 2, we can obtain that

$$\begin{aligned} & \max_{i \in \mathcal{I}} \left| n^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\mathbf{Z}) \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i - \sigma_{\boldsymbol{\eta}, ii} \right| \\ &\leq \max_{i \in \mathcal{I}} \left| n^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}, \mathbf{Z}) \boldsymbol{\eta}_i - \sigma_{\boldsymbol{\eta}, ii} \right| \\ &\leq \max_{i \in \mathcal{I}} \left| n^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}, \mathbf{Z}) \boldsymbol{\eta}_i - (n - r - p) \sigma_{\boldsymbol{\eta}, ii} / n \right| + (r + p) / n \max_{i \in \mathcal{I}} \sigma_{\boldsymbol{\eta}, ii} \\ &= \max_{i \in \mathcal{I}} \sigma_{\boldsymbol{\eta}, ii} \left| n^{-1} \chi_{n-r-p} - (n - r - p) / n \right| + (r + p) / n \max_i \sigma_{\boldsymbol{\eta}, ii} \\ &= O_p(\sqrt{\log(m)/n}) + O_p((r + p)/n). \end{aligned}$$

We next consider the first part of (41). Note that

$$\begin{aligned} & n^{-1} \mathcal{E}_i^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\bar{\mathbf{Z}}) \mathcal{Q}(\mathbf{X}) \mathcal{E}_i \\ &= n^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\bar{\mathbf{Z}}) \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i + 2n^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\bar{\mathbf{Z}}) \mathcal{Q}(\mathbf{X}) \mathbf{Z} \boldsymbol{\gamma}_i \\ & \quad + n^{-1} \boldsymbol{\gamma}_i^\top \mathbf{Z}^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\bar{\mathbf{Z}}) \mathcal{Q}(\mathbf{X}) \mathbf{Z} \boldsymbol{\gamma}_i. \end{aligned} \quad (42)$$

We then consider the three parts in (42) in the following two steps.

STEP I. We first consider $n^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\bar{\mathbf{Z}}) \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i - n^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\mathbf{Z}) \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i$. According to the results of Theorem 2 in [29] that $\text{tr}\{\mathcal{Q}(\mathbf{Z}) - \mathcal{Q}(\bar{\mathbf{Z}})\}^2 = O_p(r^2/n)$, we can obtain that

$$\begin{aligned} & \max_{i \in \mathcal{I}} \left| n^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\bar{\mathbf{Z}}) \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i - n^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\mathbf{Z}) \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i \right| \\ &= \max_{i \in \mathcal{I}} \left| n^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \{ \mathcal{Q}(\bar{\mathbf{Z}}) - \mathcal{Q}(\mathbf{Z}) \} \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i \right| \\ &\leq \left[\text{tr}\{ \mathcal{Q}(\bar{\mathbf{Z}}) - \mathcal{Q}(\mathbf{Z}) \}^2 \right]^{1/2} \max_{i \in \mathcal{I}} n^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i = O_p(r/n^{1/2}), \end{aligned}$$

where the last equality is due to the fact that $\max_i \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i = O_p(n - p)$ by Lemma 2.

STEP II. We next consider $2n^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\bar{\mathbf{Z}}) \mathcal{Q}(\mathbf{X}) \mathbf{Z} \boldsymbol{\gamma}_i$, which is less than $|2n^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\bar{\mathbf{Z}}) \mathbf{Z} \boldsymbol{\gamma}_i|$. As $\mathcal{Q}(\mathbf{Z})$ is orthogonal to \mathbf{Z} , the order of $2n^{-1} \boldsymbol{\eta}_i^\top \mathcal{Q}(\bar{\mathbf{Z}}) \mathbf{Z} \boldsymbol{\gamma}_i$ is equivalent to that of $2n^{-1} \boldsymbol{\eta}_i^\top \{ \mathcal{Q}(\bar{\mathbf{Z}}) - \mathcal{Q}(\mathbf{Z}) \} \mathbf{Z} \boldsymbol{\gamma}_i$. By the results of Theorem

2 in [29] that $\text{tr}\{\mathcal{Q}(\mathbf{Z}) - \mathcal{Q}(\overline{\mathbf{Z}})\}^2 = O_p(r^2/n)$, the above term is further bounded as

$$\begin{aligned} & \max_{i \in \mathcal{I}} 2n^{-1} \boldsymbol{\eta}_i^\top \{\mathcal{Q}(\overline{\mathbf{Z}}) - \mathcal{Q}(\mathbf{Z})\} \mathbf{Z} \boldsymbol{\gamma}_i \\ & \leq [\text{tr}\{\mathcal{Q}(\overline{\mathbf{Z}}) - \mathcal{Q}(\mathbf{Z})\}^2]^{1/2} \times \max_{i \in \mathcal{I}} |2n^{-1} \boldsymbol{\eta}_i^\top \mathbf{Z} \boldsymbol{\gamma}_i| \\ & = O_p(r/\sqrt{n}) \times O_p(\sqrt{\log(m)}/\sqrt{n}) = O_p(r/n). \end{aligned}$$

STEP III. We last consider $n^{-1} \boldsymbol{\gamma}_i^\top \mathbf{Z}^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\overline{\mathbf{Z}}) \mathcal{Q}(\mathbf{X}) \mathbf{Z} \boldsymbol{\gamma}_i$. By condition C3 and the fact that $\text{tr}(\mathbf{Z}^\top \mathcal{Q}(\overline{\mathbf{Z}}) \mathbf{Z}) = O_p(r^2)$ as demonstrated in [29], we obtain

$$\max_{i \in \mathcal{I}} n^{-1} \boldsymbol{\gamma}_i^\top \mathbf{Z}^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\overline{\mathbf{Z}}) \mathcal{Q}(\mathbf{X}) \mathbf{Z} \boldsymbol{\gamma}_i \leq n^{-1} \max_{i \in \mathcal{I}} \|\boldsymbol{\gamma}_i\|^2 \text{tr}(\mathbf{Z}^\top \mathcal{Q}(\overline{\mathbf{Z}}) \mathbf{Z}) = O_p(r^2/n).$$

Combining all of these results above, we have completed the first step of the proof for Lemma 4.

We next consider the second step of Lemma 4. Note that $\{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\} \mathcal{E}^\top$ can be written as $\{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\} \mathbf{Z} \boldsymbol{\gamma}^\top + \{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\} \boldsymbol{\eta}^\top$. As a result, $\{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\} \mathcal{E}^\top$ also follows a latent factor structure of dimension r . Thus, by the results in the first step of the proof, we can obtain that

$$\max_{i \in \mathcal{I}} |\check{\sigma}_{\boldsymbol{\eta}, ii} - \sigma_{\boldsymbol{\eta}, ii}| = O_p \left\{ \sqrt{\log(m)/n} + (r+p)/n + r/\sqrt{n} + r^2/n \right\}.$$

This completes the proof of Lemma 4. \blacksquare

Proof of Lemma 5 Similar to the proof of Lemma 4, we only need to prove that

$$\max_{i \in \mathcal{I}} \|\mathcal{P}^\top \overline{\mathbf{Z}} \overline{\boldsymbol{\gamma}}_i - \mathcal{P}^\top \mathbf{Z} \boldsymbol{\gamma}_i\|_2 = O_p \left(\sqrt{r/n} \sqrt{r + \sqrt{r} \sqrt{\log(m)}} \right) + O_p(m^{-\nu} \sqrt{r}).$$

By Cauchy–Schwarz inequality, we can obtain

$$\begin{aligned} & \|\mathcal{P}^\top \overline{\mathbf{Z}} \overline{\boldsymbol{\gamma}}_i - \mathcal{P}^\top \mathbf{Z} \boldsymbol{\gamma}_i\|_2 \\ & = \|\mathcal{P}^\top \overline{\mathbf{Z}} \overline{\boldsymbol{\gamma}}_i - \mathcal{P}^\top \mathbf{Z} \overline{\boldsymbol{\gamma}}_i + \mathcal{P}^\top \mathbf{Z} \overline{\boldsymbol{\gamma}}_i - \mathcal{P}^\top \mathbf{Z} \boldsymbol{\gamma}_i\|_2 \\ & \leq \sqrt{2} \|\mathcal{P}^\top \mathbf{Z} \overline{\boldsymbol{\gamma}}_i - \mathcal{P}^\top \mathbf{Z} \boldsymbol{\gamma}_i\|_2 + \sqrt{2} \|\mathcal{P}^\top \overline{\mathbf{Z}} \overline{\boldsymbol{\gamma}}_i - \mathcal{P}^\top \mathbf{Z} \overline{\boldsymbol{\gamma}}_i\|_2. \end{aligned} \quad (43)$$

We then consider the above two terms separately. The first term of (43) can be bounded by

$$\begin{aligned} & \|\mathcal{P}^\top \mathbf{Z} \overline{\boldsymbol{\gamma}}_i - \mathcal{P}^\top \mathbf{Z} \boldsymbol{\gamma}_i\|_2^2 \\ & = (\overline{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i)^\top \left(\mathbf{Z}^\top \mathcal{P} \mathcal{P}^\top \mathbf{Z} \right) (\overline{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i) \\ & \leq (\overline{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i)^\top \left(\sum_{\ell=1}^k \mathbf{Z}^\top \mathcal{P}_\ell \mathcal{P}_\ell^\top \mathbf{Z} \right) (\overline{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i). \end{aligned}$$

Note that $\mathbf{Z}^\top \mathcal{P}_\ell$ are *i.i.d.* with mean zero and identity matrix of dimension r , due to the fact that $\mathcal{P}^\top \mathcal{P} = \mathbf{I}_k$. Hence, the trace of $\sum_{\ell=1}^k \mathbf{Z}^\top \mathcal{P}_\ell \mathcal{P}_\ell^\top \mathbf{Z}$ is $O_p(r)$. As a result,

$$\max_{i \in \mathcal{I}} \|\mathcal{P}^\top \mathbf{Z} \overline{\boldsymbol{\gamma}}_i - \mathcal{P}^\top \mathbf{Z} \boldsymbol{\gamma}_i\|_2$$

$$\begin{aligned}
&\leq \sqrt{r} \max_{i \in \mathcal{I}} \|\bar{\gamma}_i - \gamma_i\|_2 \\
&= \sqrt{r} \max_{i \in \mathcal{I}} \|n^{-1} \bar{\mathbf{Z}}^\top \mathcal{Q}(\mathbf{X}) \mathbf{Z} \gamma_i + n^{-1} \bar{\mathbf{Z}}^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i - \gamma_i\|_2 \\
&\leq \sqrt{r}/n \max_{i \in \mathcal{I}} \|\bar{\mathbf{Z}}^\top \mathcal{Q}(\mathbf{X}) \mathbf{Z} \gamma_i - \mathbf{Z}^\top \mathcal{Q}(\mathbf{X}) \mathbf{Z} \gamma_i\|_2 \\
&\quad + \sqrt{r}/n \max_{i \in \mathcal{I}} \|\bar{\mathbf{Z}}^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i\|_2. \tag{44}
\end{aligned}$$

We consider the above two parts separately. First note that

$$\begin{aligned}
&\|\bar{\mathbf{Z}}^\top \mathcal{Q}(\mathbf{X}) \mathbf{Z} \gamma_i - \mathbf{Z}^\top \mathcal{Q}(\mathbf{X}) \mathbf{Z} \gamma_i\|_2^2 \\
&= \gamma_i^\top \mathbf{Z}^\top \mathcal{Q}(\mathbf{X}) (\bar{\mathbf{Z}} - \mathbf{Z}) (\bar{\mathbf{Z}} - \mathbf{Z})^\top \mathcal{Q}(\mathbf{X}) \mathbf{Z} \gamma_i \\
&\leq \|\gamma_i\|_2^2 \times \text{tr}\{\mathbf{Z}^\top (\bar{\mathbf{Z}} - \mathbf{Z}) (\bar{\mathbf{Z}} - \mathbf{Z})^\top \mathbf{Z}\} \\
&\leq n \|\gamma_i\|_2^2 \times \text{tr}\{(\bar{\mathbf{Z}} - \mathbf{Z}) (\bar{\mathbf{Z}} - \mathbf{Z})^\top\}.
\end{aligned}$$

Moreover, according to condition C3, we have $\max_i \|\gamma_i\|_2^2 = O(1)$. Consequently, we only need to consider $\text{tr}\{(\mathbf{Z} - \bar{\mathbf{Z}})^\top (\mathbf{Z} - \bar{\mathbf{Z}})\}$. According to condition C4, one can easily verify that

$$\text{tr}\{(\mathbf{Z} - \bar{\mathbf{Z}})^\top (\mathbf{Z} - \bar{\mathbf{Z}})\} = \sum_{e \leq r} \|\bar{\boldsymbol{\rho}}_e - \mathbf{Z}_{\cdot e}\|_2^2 = O_p(rm^{-2\nu}).$$

Here, $(\mathbf{Z}_{\cdot 1}, \dots, \mathbf{Z}_{\cdot r}) = \mathbf{Z}$. Consequently, we have

$$\sqrt{r}/n \max_{i \in \mathcal{I}} \|\bar{\mathbf{Z}}^\top \mathcal{Q}(\mathbf{X}) \mathbf{Z} \gamma_i - \mathbf{Z}^\top \mathcal{Q}(\mathbf{X}) \mathbf{Z} \gamma_i\|_2 = O_p(m^{-\nu} r / \sqrt{n}). \tag{45}$$

We next consider the second part of (44). By Cauchy–Schwarz inequality, we have

$$\|\bar{\mathbf{Z}}^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i\|_2^2 \leq 2\|(\bar{\mathbf{Z}} - \mathbf{Z})^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i\|_2^2 + 2\|\mathbf{Z}^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i\|_2^2.$$

The first part can be further bounded as

$$\begin{aligned}
&\max_{i \in \mathcal{I}} \|(\bar{\mathbf{Z}} - \mathbf{Z})^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i\|_2^2 \\
&= \max_{i \in \mathcal{I}} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) (\bar{\mathbf{Z}} - \mathbf{Z}) (\bar{\mathbf{Z}} - \mathbf{Z})^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i \\
&\leq \text{tr}\{(\bar{\mathbf{Z}} - \mathbf{Z}) (\bar{\mathbf{Z}} - \mathbf{Z})^\top\} \times \max_{i \in \mathcal{I}} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i \\
&= O_p(m^{-2\nu} rn).
\end{aligned}$$

For the second part, similar derivations yield that

$$\begin{aligned}
&\max_{i \in \mathcal{I}} \|\mathbf{Z}^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i\|_2^2 \\
&= \max_{i \in \mathcal{I}} \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \mathbf{Z} \mathbf{Z}^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i \\
&\leq \max_{i \in \mathcal{I}} \boldsymbol{\eta}_i^\top \mathbf{Z} \mathbf{Z}^\top \boldsymbol{\eta}_i \\
&= n \times \max_{i \in \mathcal{I}} (\mathbf{Z}^\top \boldsymbol{\eta}_i / \sqrt{n})^\top (\mathbf{Z}^\top \boldsymbol{\eta}_i / \sqrt{n}) \\
&= n \times \max_{i \in \mathcal{I}} \chi_{i,r} = O_p\left(n(r + \sqrt{r} \sqrt{\log(m)})\right).
\end{aligned}$$

Consequently, combining the above two results, we have

$$\sqrt{r}/n \max_{i \in \mathcal{I}} \|\bar{\mathbf{Z}}^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i\|_2 = O_p(\sqrt{r/n} \sqrt{r + \sqrt{r} \sqrt{\log(m)}}). \quad (46)$$

Combining the results in (45) and (46) yields that

$$\max_{i \in \mathcal{I}} \|\mathcal{P}^\top \mathbf{Z} \bar{\boldsymbol{\gamma}}_i - \mathcal{P}^\top \mathbf{Z} \boldsymbol{\gamma}_i\|_2 = O_p(\sqrt{r/n} \sqrt{r + \sqrt{r} \sqrt{\log(m)}}) + O_p(m^{-\nu} r / \sqrt{n}). \quad (47)$$

We next consider the second term of (43). Note that

$$\begin{aligned} & \|\mathcal{P}^\top \bar{\mathbf{Z}} \bar{\boldsymbol{\gamma}}_i - \mathcal{P}^\top \mathbf{Z} \boldsymbol{\gamma}_i\|_2^2 \\ &= \bar{\boldsymbol{\gamma}}_i^\top (\mathbf{Z} - \bar{\mathbf{Z}})^\top \mathcal{P} \mathcal{P}^\top (\mathbf{Z} - \bar{\mathbf{Z}}) \bar{\boldsymbol{\gamma}}_i \\ &\leq \|\bar{\boldsymbol{\gamma}}_i\|_2^2 \times \lambda_{\max}(\mathcal{P} \mathcal{P}^\top) \times \text{tr}\{(\mathbf{Z} - \bar{\mathbf{Z}})^\top (\mathbf{Z} - \bar{\mathbf{Z}})\} \\ &= \|\bar{\boldsymbol{\gamma}}_i\|_2^2 \times \lambda_{\max}(\mathcal{P}^\top \mathcal{P}) \times \text{tr}\{(\mathbf{Z} - \bar{\mathbf{Z}})^\top (\mathbf{Z} - \bar{\mathbf{Z}})\} \\ &= \|\bar{\boldsymbol{\gamma}}_i\|_2^2 \times \text{tr}\{(\mathbf{Z} - \bar{\mathbf{Z}})^\top (\mathbf{Z} - \bar{\mathbf{Z}})\}. \end{aligned}$$

As a result, we can obtain that

$$\begin{aligned} & \max_{i \in \mathcal{I}} \|\mathcal{P}^\top \bar{\mathbf{Z}} \bar{\boldsymbol{\gamma}}_i - \mathcal{P}^\top \mathbf{Z} \boldsymbol{\gamma}_i\|_2 \\ &\leq O_p(\sqrt{r} m^{-\nu}) \times \max_{i \in \mathcal{I}} \|\bar{\boldsymbol{\gamma}}_i\|_2 \\ &\leq O_p(\sqrt{r} m^{-\nu}) \times \max_{i \in \mathcal{I}} \{\sqrt{2} \|\boldsymbol{\gamma}_i\|_2 + \sqrt{2} \|\bar{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|_2\} \\ &= O_p(\sqrt{r} m^{-\nu}). \end{aligned} \quad (48)$$

Combining these results in (47) and (48), we thus have

$$\max_{i \in \mathcal{I}} \|\mathcal{P}^\top \bar{\mathbf{Z}} \bar{\boldsymbol{\gamma}}_i - \mathcal{P}^\top \mathbf{Z} \boldsymbol{\gamma}_i\|_2 = O_p(\sqrt{r/n} \sqrt{r + \sqrt{r} \sqrt{\log(m)}}) + O_p(\sqrt{r} m^{-\nu}), \quad (49)$$

which completes the proof of Lemma 5. \blacksquare

Proof of Proposition 3. To verify (27), based on the result of Lemma 5, it suffices to prove that

$$\max_{i \in \mathcal{I}} \|\mathcal{P}^\top \hat{\mathbf{Z}} \hat{\boldsymbol{\gamma}}_i - \mathcal{P}^\top \check{\mathbf{Z}} \check{\boldsymbol{\gamma}}_i\|_2 = O_p(b_{m,n}).$$

Note that $\hat{\boldsymbol{\gamma}}_i = (\hat{\mathbf{Z}}^\top \hat{\mathbf{Z}})^{-1} \hat{\mathbf{Z}}^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\varepsilon}_i$ and $\check{\boldsymbol{\gamma}}_i = (\check{\mathbf{Z}}^\top \check{\mathbf{Z}})^{-1} \check{\mathbf{Z}}^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\varepsilon}_i$. Then, for any $i \in \mathcal{I}$, we can have

$$\begin{aligned} & \|\mathcal{P}^\top \hat{\mathbf{Z}} \hat{\boldsymbol{\gamma}}_i - \mathcal{P}^\top \check{\mathbf{Z}} \check{\boldsymbol{\gamma}}_i\|_2^2 \\ &= \|\mathcal{P}^\top \hat{\mathbf{Z}} (\hat{\mathbf{Z}}^\top \hat{\mathbf{Z}})^{-1} \hat{\mathbf{Z}}^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\varepsilon}_i - \mathcal{P}^\top \check{\mathbf{Z}} (\check{\mathbf{Z}}^\top \check{\mathbf{Z}})^{-1} \check{\mathbf{Z}}^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\varepsilon}_i\|_2^2 \\ &= \|\mathcal{P}^\top \{\hat{\mathbf{Z}} (\hat{\mathbf{Z}}^\top \hat{\mathbf{Z}})^{-1} \hat{\mathbf{Z}}^\top - \check{\mathbf{Z}} (\check{\mathbf{Z}}^\top \check{\mathbf{Z}})^{-1} \check{\mathbf{Z}}^\top\} \mathcal{Q}(\mathbf{X}) \boldsymbol{\varepsilon}_i\|_2^2 \\ &= \lambda_{\max}(\mathcal{P} \mathcal{P}^\top) \times \boldsymbol{\varepsilon}_i^\top \mathcal{Q}(\mathbf{X}) \{\hat{\mathbf{Z}} (\hat{\mathbf{Z}}^\top \hat{\mathbf{Z}})^{-1} \hat{\mathbf{Z}}^\top - \check{\mathbf{Z}} (\check{\mathbf{Z}}^\top \check{\mathbf{Z}})^{-1} \check{\mathbf{Z}}^\top\}^2 \mathcal{Q}(\mathbf{X}) \boldsymbol{\varepsilon}_i \\ &\leq \lambda_{\max}(\mathcal{P}^\top \mathcal{P}) \times \|\hat{\mathbf{Z}} (\hat{\mathbf{Z}}^\top \hat{\mathbf{Z}})^{-1} \hat{\mathbf{Z}}^\top - \check{\mathbf{Z}} (\check{\mathbf{Z}}^\top \check{\mathbf{Z}})^{-1} \check{\mathbf{Z}}^\top\|_F^2 \times \boldsymbol{\varepsilon}_i^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\varepsilon}_i, \end{aligned} \quad (50)$$

where $\lambda_{\max}(\cdot)$ represents the largest eigenvalue of any matrix therein. We first consider $\mathcal{E}_i^\top \mathcal{Q}(\mathbf{X}) \mathcal{E}_i$. By $\mathcal{E}_i = \mathbf{Z}\gamma_i + \boldsymbol{\eta}_i$, we have $\mathcal{E}_i^\top \mathcal{Q}(\mathbf{X}) \mathcal{E}_i = \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i + 2\boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \mathbf{Z}\gamma_i + \gamma_i^\top \mathbf{Z}^\top \mathcal{Q}(\mathbf{X}) \mathbf{Z}\gamma_i$. Note that $\boldsymbol{\eta}_i$ follows a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\sigma_{\boldsymbol{\eta},ii} \mathbf{I}_n$. Thus, $\boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i / \sigma_{\boldsymbol{\eta},ii}$ follows a chi-squared distribution with degrees of freedom $n - p$. By Lemma 2 and condition C2, we have $\max_i \boldsymbol{\eta}_i^\top \mathcal{Q}(\mathbf{X}) \boldsymbol{\eta}_i = O_p(n)$. In addition, by condition C3 and the identification condition $\text{cov}(\mathbf{Z}_j) = \mathbf{I}_r$, we can obtain that $\max_i \gamma_i^\top \mathbf{Z}^\top \mathcal{Q}(\mathbf{X}) \mathbf{Z}\gamma_i \leq n \max_i \gamma_i^\top \gamma_i \leq n\gamma_{\max}$, with probability approaching one. Combing these results, we have $\max_i \mathcal{E}_i^\top \mathcal{Q}(\mathbf{X}) \mathcal{E}_i = O_p(n)$. We next consider the first part of (50). By Cauchy-Schwarz inequality,

$$\begin{aligned} & \|\widehat{\mathbf{Z}}(\widehat{\mathbf{Z}}^\top \widehat{\mathbf{Z}})^{-1} \widehat{\mathbf{Z}}^\top - \check{\mathbf{Z}}(\check{\mathbf{Z}}^\top \check{\mathbf{Z}})^{-1} \check{\mathbf{Z}}^\top\|_F^2 \\ &= \|\widehat{\mathbf{Z}}\widehat{\mathbf{Z}}^\top - \check{\mathbf{Z}}\check{\mathbf{Z}}^\top\|_F^2/n^2 \\ &= \|\widehat{\mathbf{Z}}\widehat{\mathbf{Z}}^\top - \check{\mathbf{Z}}\check{\mathbf{Z}}^\top + \widehat{\mathbf{Z}}\check{\mathbf{Z}}^\top - \check{\mathbf{Z}}\widehat{\mathbf{Z}}^\top\|_F^2/n^2 \\ &\leq 2\|\widehat{\mathbf{Z}}\widehat{\mathbf{Z}}^\top - \check{\mathbf{Z}}\check{\mathbf{Z}}^\top\|_F^2/n^2 + 2\|\widehat{\mathbf{Z}}\check{\mathbf{Z}}^\top - \check{\mathbf{Z}}\widehat{\mathbf{Z}}^\top\|_F^2/n^2 \\ &= 4\|\widehat{\mathbf{Z}} - \check{\mathbf{Z}}\|_F^2/n. \end{aligned}$$

Furthermore, by Lemma 3, we can obtain

$$\begin{aligned} & \|\widehat{\mathbf{Z}} - \check{\mathbf{Z}}\|_F \\ &\leq \sqrt{2r} \|m_*^{-1} \widehat{\mathcal{E}}_{\mathcal{L}_*}^\top \widehat{\mathcal{E}}_{\mathcal{L}_*} - m_*^{-1} \{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\} \mathcal{E}_{\mathcal{L}_*}^\top \mathcal{E}_{\mathcal{L}_*} \{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\}\|_F / K_{\min}, \end{aligned}$$

where $K_{\min} = \min_{2 \leq e \leq r-1} \min\{|\widehat{\Lambda}_{e-1} - \check{\Lambda}_e|, |\widehat{\Lambda}_{e+1} - \check{\Lambda}_e|\}$. Using Cauchy-Schwarz inequality again,

$$\begin{aligned} & \|m_*^{-1} \widehat{\mathcal{E}}_{\mathcal{L}_*}^\top \widehat{\mathcal{E}}_{\mathcal{L}_*} - m_*^{-1} \{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\} \mathcal{E}_{\mathcal{L}_*}^\top \mathcal{E}_{\mathcal{L}_*} \{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\}\|_F \\ &= \|m_*^{-1} \{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\} \mathbf{X} \mathbf{B}_{\mathcal{L}_*}^\top \mathbf{B}_{\mathcal{L}_*} \mathbf{X}^\top \{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\} \\ &\quad + 2m_*^{-1} \{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\} \mathbf{X} \mathbf{B}_{\mathcal{L}_*}^\top \mathcal{E}_{\mathcal{L}_*} \{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\}\|_F \\ &\leq \sqrt{2} \|m_*^{-1} \{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\} \mathbf{X} \mathbf{B}_{\mathcal{L}_*}^\top \mathbf{B}_{\mathcal{L}_*} \mathbf{X}^\top \{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\}\|_F \\ &\quad + \sqrt{2} \|2m_*^{-1} \{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\} \mathbf{X} \mathbf{B}_{\mathcal{L}_*}^\top \mathcal{E}_{\mathcal{L}_*} \{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\}\|_F \\ &= \sqrt{2} \|m_*^{-1} \{\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top \mathbf{V}^{-1} \mathbf{A} \mathbf{B}_{\mathcal{L}_* \cap \mathcal{I}_1}^\top\} \{\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top \mathbf{V}^{-1} \mathbf{A} \mathbf{B}_{\mathcal{L}_* \cap \mathcal{I}_1}^\top\}^\top\|_F \\ &\quad + \sqrt{2} \|2m_*^{-1} \{\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top \mathbf{V}^{-1} \mathbf{A} \mathbf{B}_{\mathcal{L}_* \cap \mathcal{I}_1}^\top\} \mathcal{E}_{\mathcal{L}_* \cap \mathcal{I}_1} \{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\}\|_F, \end{aligned} \tag{51}$$

where $\mathbf{V} = \mathbf{A}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top$. We next consider the above two parts separately. We first consider the first part,

$$\begin{aligned} & \|m_*^{-1} \{\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top \mathbf{V}^{-1} \mathbf{A} \mathbf{B}_{\mathcal{L}_* \cap \mathcal{I}_1}^\top\} \{\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top \mathbf{V}^{-1} \mathbf{A} \mathbf{B}_{\mathcal{L}_* \cap \mathcal{I}_1}^\top\}^\top\|_F \\ &= m_*^{-1} \|(\mathbf{A} \mathbf{B}_{\mathcal{L}_* \cap \mathcal{I}_1}^\top)^\top \mathbf{V}^{-1} (\mathbf{A} \mathbf{B}_{\mathcal{L}_* \cap \mathcal{I}_1}^\top)\|_F \\ &= m_*^{-1} n \|(\mathbf{A} \mathbf{B}_{\mathcal{L}_* \cap \mathcal{I}_1}^\top)^\top (\mathbf{A} \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \mathbf{A}^\top)^{-1} (\mathbf{A} \mathbf{B}_{\mathcal{L}_* \cap \mathcal{I}_1}^\top)\|_F \\ &\leq m_*^{-1} n \|(\mathbf{A} \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \mathbf{A}^\top)^{-1/2} \mathbf{A} \mathbf{B}_{\mathcal{L}_* \cap \mathcal{I}_1}^\top\|_F^2. \end{aligned}$$

The second part in (51) can be derived in a similar way as

$$\|m_*^{-1} \{\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top \mathbf{V}^{-1} \mathbf{A} \mathbf{B}_{\mathcal{L}_* \cap \mathcal{I}_1}^\top\} \mathcal{E}_{\mathcal{L}_* \cap \mathcal{I}_1} \{\mathcal{Q}(\mathbf{X}) + \mathcal{P}\mathcal{P}^\top\}\|_F^2$$

$$\begin{aligned}
 &= m_*^{-2} \text{tr} \left[\{ \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top \mathbf{V}^{-1} \mathbf{A} \mathbf{B}_{\mathcal{I}_* \cap \mathcal{I}_1}^\top \} \mathcal{E}_{\mathcal{I}_* \cap \mathcal{I}_1} \{ \mathcal{Q}(\mathbf{X}) + \mathcal{P} \mathcal{P}^\top \} \mathcal{E}_{\mathcal{I}_* \cap \mathcal{I}_1}^\top \right. \\
 &\quad \left. \{ \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{A}^\top \mathbf{V}^{-1} \mathbf{A} \mathbf{B}_{\mathcal{I}_* \cap \mathcal{I}_1}^\top \}^\top \right] \\
 &= m_*^{-2} n \text{tr} \left[\mathcal{E}_{\mathcal{I}_* \cap \mathcal{I}_1} \{ \mathcal{Q}(\mathbf{X}) + \mathcal{P} \mathcal{P}^\top \} \mathcal{E}_{\mathcal{I}_* \cap \mathcal{I}_1}^\top (\mathbf{A} \mathbf{B}_{\mathcal{I}_* \cap \mathcal{I}_1}^\top)^\top (\mathbf{A} \Sigma_{\mathbf{X}}^{-1} \mathbf{A}^\top)^{-1} \mathbf{A} \mathbf{B}_{\mathcal{I}_* \cap \mathcal{I}_1}^\top \right] \\
 &\leq m_*^{-2} n \text{tr} \{ \mathcal{E}_{\mathcal{I}_* \cap \mathcal{I}_1} \mathcal{E}_{\mathcal{I}_* \cap \mathcal{I}_1}^\top (\mathbf{A} \mathbf{B}_{\mathcal{I}_* \cap \mathcal{I}_1}^\top)^\top (\mathbf{A} \Sigma_{\mathbf{X}}^{-1} \mathbf{A}^\top)^{-1} \mathbf{A} \mathbf{B}_{\mathcal{I}_* \cap \mathcal{I}_1}^\top \} \\
 &\leq m_*^{-2} n \lambda_{\max}(\mathcal{E}_{\mathcal{I}_* \cap \mathcal{I}_1} \mathcal{E}_{\mathcal{I}_* \cap \mathcal{I}_1}^\top) \times \text{tr} \{ (\mathbf{A} \mathbf{B}_{\mathcal{I}_* \cap \mathcal{I}_1}^\top)^\top (\mathbf{A} \Sigma_{\mathbf{X}}^{-1} \mathbf{A}^\top)^{-1} \mathbf{A} \mathbf{B}_{\mathcal{I}_* \cap \mathcal{I}_1}^\top \} \\
 &= m_*^{-2} n \lambda_{\max}(\mathcal{E}_{\mathcal{I}_* \cap \mathcal{I}_1} \mathcal{E}_{\mathcal{I}_* \cap \mathcal{I}_1}^\top) \times \| (\mathbf{A} \Sigma_{\mathbf{X}}^{-1} \mathbf{A}^\top)^{-1/2} (\mathbf{A} \mathbf{B}_{\mathcal{I}_* \cap \mathcal{I}_1}^\top) \|_F^2. \tag{52}
 \end{aligned}$$

One can easily verify that $\lambda_{\max}(m^{-1} n^{-1} \mathcal{E} \mathcal{E}^\top) = O_p(r^{-1})$ by condition C3 and the identification condition $\text{cov}(\mathbf{Z}_j) = \mathbf{I}_r$. Thus, (52) is further bounded by

$$r^{-1} m_*^{-2} n^2 \times \text{Card}(\mathcal{I}_* \cap \mathcal{I}_1) \| (\mathbf{A} \Sigma_{\mathbf{X}}^{-1} \mathbf{A}^\top)^{-1/2} (\mathbf{A} \mathbf{B}_{\mathcal{I}_* \cap \mathcal{I}_1}^\top) \|_F^2.$$

As a result, (51) can be bounded by

$$\begin{aligned}
 b_{m,n} &= \sqrt{r} m_*^{-1} n \| (\mathbf{A} \Sigma_{\mathbf{X}}^{-1} \mathbf{A}^\top)^{-1/2} (\mathbf{A} \mathbf{B}_{\mathcal{I}_* \cap \mathcal{I}_1}^\top) \|_F^2 \\
 &\quad + m_*^{-1} n \times \text{Card}(\mathcal{I}_* \cap \mathcal{I}_1)^{1/2} \| (\mathbf{A} \Sigma_{\mathbf{X}}^{-1} \mathbf{A}^\top)^{-1/2} (\mathbf{A} \mathbf{B}_{\mathcal{I}_* \cap \mathcal{I}_1}^\top) \|_F.
 \end{aligned}$$

We next consider K_{\min} . Note that $|\widehat{\Lambda}_{e-1} - \check{\Lambda}_e| \geq |\check{\Lambda}_{e-1} - \check{\Lambda}_e| - |\widehat{\Lambda}_{e-1} - \check{\Lambda}_{e-1}|$. By Lemma 3 again,

$$\begin{aligned}
 &|\widehat{\Lambda}_{e-1} - \check{\Lambda}_{e-1}| \\
 &\leq \| m_*^{-1} \widehat{\mathcal{E}}_{\mathcal{I}_*}^\top \widehat{\mathcal{E}}_{\mathcal{I}_*} - m_*^{-1} \{ \mathcal{Q}(\mathbf{X}) + \mathcal{P} \mathcal{P}^\top \} \mathcal{E}_{\mathcal{I}_*}^\top \mathcal{E}_{\mathcal{I}_*} \{ \mathcal{Q}(\mathbf{X}) + \mathcal{P} \mathcal{P}^\top \} \|_F \\
 &= O_p(b_{m,n}).
 \end{aligned}$$

If we further assume that $|\check{\Lambda}_{e-1} - \check{\Lambda}_e| \geq d_{m,n}$ for some positive constant $d_{m,n}$ and $e = 2, \dots, r$ (see Theorem 2 in [12] for details), then $|\check{\Lambda}_{e-1} - \check{\Lambda}_e| \geq d_{m,n}$, for any $e = 2, \dots, r$. Consequently, with a probability approaching one, $2K_{\min} \geq \min_{2 \leq e \leq r} |\widehat{\Lambda}_{e-1} - \check{\Lambda}_e| \geq d_{m,n} > 0$. Combing these two results, we can obtain that

$$\max_{i \in \mathcal{I}} \| \mathcal{P}^\top \widehat{\mathbf{Z}} \widehat{\boldsymbol{\gamma}}_i - \mathcal{P}^\top \check{\mathbf{Z}} \check{\boldsymbol{\gamma}}_i \|_2 = O_p(b_{m,n}),$$

which completes the proof of (27).

Now let us turn to prove (28). By Lemma 4, it suffices to show that

$$\max_{i \in \mathcal{I}} |\widehat{\sigma}_{\boldsymbol{\eta}, ii} - \check{\sigma}_{\boldsymbol{\eta}, ii}| = O_p(b_{m,n}/n^{1/2}).$$

To this end, for any $i \in \mathcal{I}$,

$$\begin{aligned}
 &|\widehat{\sigma}_{\boldsymbol{\eta}, ii} - \check{\sigma}_{\boldsymbol{\eta}, ii}| \\
 &= |\mathcal{E}_{i\cdot}^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\widehat{\mathbf{Z}}) \mathcal{Q}(\mathbf{X}) \mathcal{E}_{i\cdot} - \mathcal{E}_{i\cdot}^\top \mathcal{Q}(\mathbf{X}) \mathcal{Q}(\check{\mathbf{Z}}) \mathcal{Q}(\mathbf{X}) \mathcal{E}_{i\cdot}| / n \\
 &= |\mathcal{E}_{i\cdot}^\top \mathcal{Q}(\mathbf{X}) \{ \mathcal{Q}(\widehat{\mathbf{Z}}) - \mathcal{Q}(\check{\mathbf{Z}}) \} \mathcal{Q}(\mathbf{X}) \mathcal{E}_{i\cdot}| / n \\
 &\leq \| \widehat{\mathbf{Z}} (\widehat{\mathbf{Z}}^\top \widehat{\mathbf{Z}})^{-1} \widehat{\mathbf{Z}}^\top - \check{\mathbf{Z}} (\check{\mathbf{Z}}^\top \check{\mathbf{Z}})^{-1} \check{\mathbf{Z}}^\top \|_F / n \times \mathcal{E}_{i\cdot}^\top \mathcal{Q}(\mathbf{X}) \mathcal{E}_{i\cdot}.
 \end{aligned}$$

The above term is essentially the same as (50), which is of the order $O_p(b_{m,n}/n^{1/2})$. \blacksquare

Appendix D: Negative dependence of the BH type estimator

For notational simplicity, we use a similar set of notations as those defined in Section 2 to define the p -values $(P_i, i = 1, \dots, m)$, the p -values threshold (t) , the true null and non-null sets $(\mathcal{I}_0$ and $\mathcal{I}_1)$, their corresponding cardinality $(m_0$ and $m_1)$, the number of false rejections $(V(t))$, the total number of total rejections $(R(t))$, the true FDP $(\text{FDP}(t))$, and so forth. According to [4] and [27], we define the BH type estimator for the true FDP as

$$\widehat{\text{FDP}}_\lambda(t) = \frac{m\widehat{\pi}_0(\lambda)t}{R(t) \vee 1},$$

where $\widehat{\pi}_0(\lambda) = \{m - R(\lambda)\}/\{m(1 - \lambda)\}$ is an estimate of π_0 and λ is a tuning parameter that $\lambda \geq t$.

Following [24], this part shows that the BH type FDP estimator (i.e., $\widehat{\text{FDP}}_\lambda(t)$) shares a negative dependence pattern when all of the p -values under the alternative are smaller than the fixed threshold t . Under this condition, $R(t) = m_1 + V(t)$. Then for any two replications of experiments (use subscript $l = 1, 2$), the slope of two pairs of points $[\{\widehat{\text{FDP}}_\lambda(t)_l, \text{FDP}(t)_l\}, l = 1, 2]$ can be derived as

$$\begin{aligned} \text{slope} &= \frac{\widehat{\text{FDP}}_\lambda(t)_1 - \widehat{\text{FDP}}_\lambda(t)_2}{\text{FDP}(t)_1 - \text{FDP}(t)_2} \\ &= \frac{m\widehat{\pi}_0(\lambda)_1 t / R(t)_1 - m\widehat{\pi}_0(\lambda)_2 t / R(t)_2}{V(t)_1 / R(t)_1 - V(t)_2 / R(t)_2} \\ &= \frac{1}{1 - \lambda} \left\{ \frac{mt / R(t)_1 - mt / R(t)_2}{V(t)_1 / R(t)_1 - V(t)_2 / R(t)_2} \right. \\ &\quad \left. - \frac{R(\lambda)_1 / R(t)_1 - R(\lambda)_2 / R(t)_2}{V(t)_1 / R(t)_1 - V(t)_2 / R(t)_2} t \right\}. \end{aligned} \quad (53)$$

We first analyze the first term of (53). Using the fact that $R_l(t) = m_1 + V_l(t)$, for $l = 1, 2$, it can be simplified as

$$\begin{aligned} &\frac{1}{1 - \lambda} \times \frac{mt / R(t)_1 - mt / R(t)_2}{V(t)_1 / R(t)_1 - V(t)_2 / R(t)_2} \\ &= \frac{mt}{1 - \lambda} \times \frac{R(t)_2 - R(t)_1}{V(t)_1 R(t)_2 - V(t)_2 R(t)_1} \\ &= \frac{mt}{1 - \lambda} \times \frac{V(t)_2 - V(t)_1}{V(t)_1 [m_1 + V(t)_2] - V(t)_2 [m_1 + V(t)_1]} \\ &= -\frac{mt}{(1 - \lambda)} \times \frac{1}{m_1}. \end{aligned} \quad (54)$$

(54) implies that the slope of two replications for the BH procedure in [4] is $-mt/m_1$. By similar technique and the fact that $t \leq \lambda$, the second term can be decomposed as

$$-\frac{t}{1 - \lambda} \times \frac{R(\lambda)_1 / R(t)_1 - R(\lambda)_2 / R(t)_2}{V(t)_1 / R(t)_1 - V(t)_2 / R(t)_2}$$

$$\begin{aligned}
&= -\frac{t}{1-\lambda} \times \frac{R(\lambda)_1/R(t)_1 - R(\lambda)_1/R(t)_2}{V(t)_1/R(t)_1 - V(t)_2/R(t)_2} \\
&\quad - \frac{t}{1-\lambda} \times \frac{[R(\lambda)_1 - R(\lambda)_2]/R(t)_2}{V(t)_1/R(t)_1 - V(t)_2/R(t)_2} \\
&= \frac{t}{1-\lambda} \times \frac{R_1(\lambda)}{m_1} - \frac{t}{1-\lambda} \times \frac{R_1(t)}{m_1} \times \frac{V(\lambda)_1 - V(\lambda)_2}{V(t)_1 - V(t)_2}. \tag{55}
\end{aligned}$$

Combing (54) and (55), we have slope = $-\frac{mt}{m_1(1-\lambda)} \times [1 - R_1(\lambda)/m + R_1(t)/m \times \Delta]$ with $\Delta = \frac{V(\lambda)_1 - V(\lambda)_2}{V(t)_1 - V(t)_2}$. We conclude that the slope is negative if $\Delta > -\frac{m - R_1(\lambda)}{R_1(t)}$. ■

Acknowledgements

The authors are grateful to the Editor, an Associate Editor, and two anonymous referees for their insightful comments and constructive suggestions. C. Zhang's research is supported by the NSF grants DMS-1308872, DMS-1521761, Wisconsin Alumni Research Foundation, and National Natural Science Foundation of China, grants 11690014. L. Du's research is partially supported by IGN15BM04, SBI16BM01, and Hong Kong RGC ECS26301216.

References

- [1] Ahn, S. C., and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, **81**, 1203–1227. [MR3064065](#)
- [2] Azriel, D., and Schwartzman, A. (2015). The Empirical distribution of a large number of correlated normal variables. *J. Am. Statist. Ass.*, **110**, 1217–1228. [MR3420696](#)
- [3] Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, **71**, 135–171. [MR1956857](#)
- [4] Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B*, **57**, 289–300. [MR1325392](#)
- [5] Benjamini, Y., and Yekutieli, D. (2001). The control of false discovery rate in multiple testing under dependency. *Ann. Statist.*, **29**, 1165–1188. [MR1869245](#)
- [6] Bourgon, R., Gentleman, R., and Huber, W. (2010). Independent filtering increases detection power for high-throughput experiments. *Proc. Natl. Acad. Sci.*, **107**, 9546–9551.
- [7] Clarke, S., and Hall, P. (2009). Robustness of multiple testing procedures against dependence. *Ann. Statist.*, **37**, 332–358. [MR2488354](#)
- [8] Desai, K. H., and Storey, J. D. (2012). Cross-dimensional inference of dependent high-dimensional data. *J. Am. Statist. Ass.*, **107**, 135–151. [MR2949347](#)
- [9] Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, **30**, 207–210.

- [10] Efron, B. (2007). Correlation and large-scale simultaneous significance testing. *J. Am. Statist. Ass.*, **102**, 93–103. [MR2293302](#)
- [11] Efron, B. (2010). Correlated z-values and the accuracy of large-scale statistical estimates. *J. Am. Statist. Ass.*, **105**, 1042–1055. [MR2752597](#)
- [12] Fan, J., and Han, X. (2017). Estimation of false discovery proportion with unknown dependence. *J. R. Statist. Soc. B*, to appear.
- [13] Fan, J., Han, X., and Gu, W. (2012). Estimating false discovery proportion under arbitrary covariance dependence. *J. Am. Statist. Ass.*, **107**, 1019–1035. [MR3010887](#)
- [14] Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *J. R. Statist. Soc. B*, **75**, 603–680. [MR3091653](#)
- [15] Fan, J., Zhang, C. M., and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.*, **29**, 153–193. [MR1833962](#)
- [16] Friguet, C., Kloareg, M., and Causeur, D. (2009). A factor model approach to multiple testing under dependence. *J. Am. Statist. Ass.*, **104**, 1406–1415. [MR2750571](#)
- [17] Gagnon-Bartsch, J. A., and Speed, T. P. (2012). Using control genes to correct for unwanted variation in microarray data. *Biostatistics*, **13**, 539–552.
- [18] Idaghdour, Y., Storey, J. D., Jadallah, S., and Gibson, G. (2008). A genome-wide gene expression signature of environmental geography in leukocytes of Moroccan Amazighs. *PLoS genetics*, **4**, e1000052.
- [19] Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., Geman, D., Baggerly, K., and Irizarry, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.*, **11**, 733–739.
- [20] Leek, J. T., and Storey, J. D. (2008). A general framework for multiple testing dependence. *Proc. Natl. Acad. Sci.*, **105**, 18718–18723.
- [21] Lyons, R. (1988). Strong laws of large numbers for weakly correlated random variables. *Michigan Mathematical Journal*, **35**, 353–359. [MR0978305](#)
- [22] Owen, A. B. (2005). Variance of the number of false discoveries. *J. R. Statist. Soc. B*, **67**, 411–426. [MR2155346](#)
- [23] Sarkar, S. K. (2002). Some results on false discovery rate in stepwise multiple testing procedures. *Ann. Statist.*, **30**, 239–257. [MR1892663](#)
- [24] Schwartzman, A. (2012). Comment: FDP vs FDR and the effect of conditioning. *J. Am. Statist. Ass.*, **107**, 1039–1041. [MR3010890](#)
- [25] Schwartzman, A., and Lin, X. (2011). The effect of correlation in false discovery rate estimation. *Biometrika*, **98**, 199–214. [MR2804220](#)
- [26] Spielman, R. S., Bastone, L. A., Burdick, J. T., Morley, M., Ewens, W. J., and Cheung, V. G. (2007). Common genetic variants account for differences in gene expression among ethnic groups. *Nature Genetics*, **39**, 226–231.
- [27] Storey, J. D., Taylor, J. E., and Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Statist. Soc. B*, **66**, 187–205. [MR2035766](#)

- [28] Sun, Y., Zhang, N. R., and Owen, A. B. (2012). Multiple hypothesis testing adjusted for latent variables, with an application to the AGEMAP gene expression data. *Ann. Appl. Statist.*, **6**, 1664–1688. [MR3058679](#)
- [29] Wang, H. (2012). Factor profiled independence screening. *Biometrika*, **99**, 15–28. [MR2899660](#)
- [30] Wu, M. C., Kraft, P., Epstein, M. P., Taylor, D. M., Chanock, S. J., Hunter, D. J., and Lin, X. (2010). Powerful SNP-set analysis for case-control genome-wide association studies. *Am. J. Hum. Genet.*, **86**, 929–942.
- [31] Wu, W. (2008). On false discovery control under dependence. *Ann. Statist.*, **36**, 364–380. [MR2387975](#)
- [32] Zeggini, E., Weedon, M. N., Lindgren, C. M., Frayling, T. M. *et al.* (2007). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*, **316**, 1336–1341.
- [33] Zhang, C. M., and Yu, T. (2008). Semiparametric detection of significant activation for brain fMRI. *Ann. Statist.*, **36**, 1693–1725. [MR2435453](#)