Chapter 8
Hypothesis Testing for High-Dimensional Data

Wei Biao Wu, Zhipeng Lou, and Yuefeng Han

Abstract We present a systematic theory for tests for means of high-dimensional data. Our testing procedure is based on an invariance principle which provides distributional approximations of functionals of non-Gaussian vectors by those of Gaussian ones. Differently from the widely used Bonferroni approach, our procedure is dependence-adjusted and has an asymptotically correct size and power. To obtain cutoff values of our test, we propose a half-sampling method which avoids estimating the underlying covariance matrix of the random vectors. The latter method is shown via extensive simulations to have an excellent performance.

Keywords Gaussian approximation · Goodness-of-Fit Test · Half-sampling · High-dimensional data · Hypothesis testing · Large p small n · Rademacher weighted differencing

8.1 Introduction

With the advance of modern data collection techniques, high-dimensional data appear in various fields including physics, biology, healthcare, finance, marketing, social network, and engineering among others. A common feature in such datasets is that the data dimension or the number of involved parameters can be quite large. As a fundamentally important problem in the study of such data, one would like to perform statistical inference of those parameters such as multiple testing or construction of confidence regions. With that one is able to provide an answer to the question whether there is signal in the dataset, or whether the dataset consists only of random noises. Due to the high-dimensionality, the inferential procedures developed for low-dimensional problems may no longer be valid in the high-dimensional setting. Different approaches should be designed to account for high-dimensionality.

W. B. Wu (✉) · Z. Lou · Y. Han
Department of Statistics, University of Chicago, Chicago, IL, USA
e-mail: wbwu@galton.uchicago.edu; zplou@galton.uchicago.edu; yfhan@uchicago.edu

W. K. Härdle et al. (eds.), *Handbook of Big Data Analytics*, Springer Handbooks of Computational Statistics, https://doi.org/10.1007/978-3-319-18284-1_8

There exists a huge literature on multiple testing; see, for example, Dudiot and van der Laan (2008), Efron (2010) and Dickhaus (2014).

We now introduce the setting of our testing problem. Assume that $X_1, X_2, \ldots,$ are independent and identically distributed (i.i.d.) p-dimensional random vectors, with mean vector $\mu = (\mu_1, \ldots, \mu_p)^T = E(X_i)$ and covariance matrix $\Sigma = \mathrm{cov}(X_i) = (\sigma_{jk})_{j,k \le p}$. We are testing the hypothesis of existence of a signal

$$H_0 : \mu = 0 \text{ vs } H_A : \mu \neq 0 \tag{8.1}$$

based on the sample $X_1, \ldots, X_n$. This formulation is actually very general and its solution can be applied to many other problems; see Sect. 8.2. We can estimate μ by the sample mean vector $\hat{\mu} = \bar{X}_n = n^{-1} \sum_{i=1}^n X_i$. The classical Hotelling's T-squared test has the form

$$T = \bar{X}_n \hat{\Sigma}_n^{-1} \bar{X}_n, \tag{8.2}$$

where

$$\hat{\Sigma}_n = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)(X_i - \bar{X}_n)^T \tag{8.3}$$

is the sample covariance matrix estimate of Σ. If p is small and fixed, by the Central Limit Theorem (CLT),

$$\sqrt{n}(\bar{X}_n - \mu) \Rightarrow N(0, \Sigma). \tag{8.4}$$

By the Law of Large Numbers, if Σ is non-singular,

$$\hat{\Sigma}_n^{-1} \to \Sigma^{-1} \text{ almost surely.} \tag{8.5}$$

Clearly (8.4) and (8.5) imply that under H_0, the Hotelling's T-squared statistic $nT \Rightarrow \chi_p^2$ (χ^2 distribution with degrees of freedom p). Thus we can reject H_0 at level $0 < \alpha < 1$ if $nT > \chi_{p,1-\alpha}^2$, the $(1-\alpha)$th quantile of χ_p^2.

In the high-dimensional situation in which p can be much larger than n, the CLT (8.4) is no longer valid; see Portnoy (1986). Furthermore, $\hat{\Sigma}_n$ is singular and thus T is not well-defined. Also the matrix convergence (8.5) may not hold, see Marčenko and Pastur (1967). In this chapter we shall apply a testing functional approach that does not use $\hat{\Sigma}_n^{-1}$ or the precision matrix Σ^{-1}. A function $g : \mathbb{R}^p \to [0, \infty)$ is said to be a testing functional if the following requirements are satisfied: (1) (monotonicity) for any $x = (x_1, \ldots, x_p)^T \in \mathbb{R}^p$ and $0 < c < 1$, $g(cx) \le g(x)$; (2) (identifiability) $g(x) = 0$ if and only if $x = 0$. We shall consider the test statistic

$$T_n = g(\sqrt{n}\bar{X}_n). \tag{8.6}$$

Examples of g include the L^2-based test with $g(x) = \sum_{j=1}^p x_j^2$, the L^∞-based test with $g(x) = \max_{j \le p} |x_j|$, the weighted empirical process $g(x) =$

$\sup_{u\geq 0}(\sum_{j=1}^{p}\mathbf{1}_{|x_j|\geq u}h(u))$, where $h(\cdot)$ is a nonnegative-valued non-decreasing function, among others. We reject H_0 in (8.1) if T_n is too big.

As a theoretical foundation, we base our testing procedure on the following invariance principle result

$$\sup_{t\in\mathbb{R}}|P[g(\sqrt{n}(\bar{X}_n-\mu))\leq t]-P[g(\sqrt{n}\bar{Z}_n)\leq t]|\to 0, \tag{8.7}$$

where $Z, Z_1, Z_2, \ldots$ are i.i.d. $N(0,\Sigma)$ random vectors and $\bar{Z}_n = n^{-1}\sum_{i=1}^{n}Z_i =_{\mathcal{D}} n^{-1/2}Z$. Interestingly, though the CLT (8.4) does not generally hold in the high-dimensional setting, the testing functional form (8.7) may still be valid. Chernozhukov et al. (2014) proved (8.7) with the L^∞ norm $g(x) = \max_{j\leq p}|x_j|$, while Xu et al. (2014) consider the L^2 based test with $g(x) = \sum_{j=1}^{p}x_j^2$. In Sect. 8.5 we shall provide a sufficient condition so that (8.7) holds for certain testing functionals.

In applying (8.7) for testing (8.1), one needs to know the distribution of $g(\sqrt{n}\bar{Z}_n) =_{\mathcal{D}} g(Z)$ so that a suitable cutoff value can be obtained. The latter problem is highly nontrivial since the covariance matrix Σ, which is viewed as a nuisance parameter here, is typically not known and the associated estimation issue can be quite challenging. In Sect. 8.5 we shall propose a half-sampling technique which can avoid estimating the nuisance covariance matrix Σ.

8.2 Applications

Our paradigm (8.1) is actually quite general and it can be applied to testing of high-dimensional covariance matrices, testing of independence of high-dimensional data, analysis of variances with non-normal and heteroscedastic errors.

8.2.1 Testing of Covariance Matrices

There is a huge literature on testing covariance matrices such as uncorrelatedness, sphericity, or other patterns. For Gaussian data, tests for $\Sigma = \sigma^2\mathrm{I}_p$, where I_p is the identity matrix, can be found in Ahmad (2010), Birke and Dette (2005), Chen et al. (2010), Fisher et al. (2010) and Ledoit and Wolf (2002). Tests for equality of covariance matrices are studied in Bai et al. (2009) and Jiang et al. (2012), and for sphericity is in Onatski et al. (2013). Minimax properties are considered in Cai and Ma (2013). For other contributions, see Qu and Chen (2012), Schott (2005, 2007), Srivastava (2005), Xiao and Wu (2013) and Zhang et al. (2013).

Assume that we have data matrix $\mathbf{Y}_n = (Y_{i,j})_{1\leq i\leq n, 1\leq j\leq p}$, where $(Y_{i,j})_{j=1}^{p}$, $i = 1,\ldots,n$, are i.i.d. p-dimensional random vectors. Let

$$\sigma_{jk} = \mathrm{cov}(Y_{1,j}, Y_{1,k}), \quad 1\leq j,k\leq p, \tag{8.8}$$

be the covariance function. Consider testing hypothesis for uncorrelatedness:

$$H_0 : \sigma_{jk} = 0 \text{ for all } j \neq k. \tag{8.9}$$

For simplicity assume that $E(Y_{i,j}) = 0$. For a pair $a = (j,k)$ write $X_{i,a} = Y_{i,j}Y_{i,k}$, and $\bar{X}_a = n^{-1}\sum_{i=1}^n X_{i,a}$ and the (p^2-p)-dimensional vector $\bar{X} = (\bar{X}_a)_{a\in\mathcal{A}}$, where $\mathcal{A} = \{(j,k) : j \neq k, j \leq p, k \leq p\}$. The hypothesis H_0 in (8.9) can be tested by using the test statistics $T = g(\sqrt{n}\bar{X})$. Xiao and Wu (2013) considered the L^∞ based test with $g(x) = \max_i |x_i|$, generalizing the result in Jiang (2004) which concerns the special case for i.i.d. vectors with independent entries. Han and Wu (2017) performed an L^2 based test for patterns of covariances with the test statistic

$$T = \sum_{a\in\mathcal{A}} \bar{X}_a^2 = \sum_{j\neq k} \hat{\sigma}_{jk}^2. \tag{8.10}$$

With slight modifications, one can also test the sphericity hypothesis

$$H_0 : \Sigma = \sigma^2 \mathrm{I}_p \text{ for some } \sigma^2 > 0, \tag{8.11}$$

where I_p is the $p \times p$ identity matrix. Let $\mathcal{A}_0 = \{(j,k) : j,k \leq p\}$ with diagonal entries added to $\mathcal{A}$. For $a = (j,j) \in \mathcal{A}_0$, let $X_{i,a} = Y_{i,j}^2 - \sigma^2$. If σ^2 is known, then H_0 in (8.11) can be rejected at level $\alpha \in (0,1)$ if $T = g(\sqrt{n}\bar{X}) > t_{1-\alpha}$, where $t_{1-\alpha}$ is the $(1-\alpha)$th quantile of $g(Z)$ and Z is a centered Gaussian vector with covariance structure $\mathrm{cov}(Z_a, Z_b) = E(X_{i,a}X_{i,b})$, $a, b \in \mathcal{A}_0$. In the case that σ^2 is not known, we shall use an estimate. For example, we can let $\hat{\sigma}^2 = n^{-1}\sum_{j=1}^n \hat{\sigma}_{jj}^2$, and consider $X_{i,a}^\circ = Y_{i,j}^2 - \hat{\sigma}^2$. Let $X_{i,a}^\circ = X_{i,a}$ if $a = (j,k)$ with $j \neq k$. The hypothesis H_0 in (8.11) can be tested by the statistic $T^\circ = g(\sqrt{n}\bar{X}^\circ)$.

8.2.2 *Testing of Independence*

Let $Y_i = (Y_{i,j})_{j=1}^p$, $i = 1, \ldots, n$, be i.i.d. p-dimensional random vectors with joint cumulative distribution function

$$F_{j_1,\ldots,j_d}(y_{j_1}, \ldots, y_{j_d}) = P(Y_{i,j_1} \leq y_{j_1}, \ldots, Y_{i,j_d} \leq y_{j_d}). \tag{8.12}$$

Consider the problem of testing whether entries of Y_i are independent. Assume that the marginal distributions are standard uniform$[0,1]$. For $\mathbf{j} = (j_1, \ldots, j_d)$, write $F_{\mathbf{j}}(y_{\mathbf{j}}) = F_{j_1,\ldots,j_d}(y_{j_1}, \ldots, y_{j_d})$. For fixed d, the hypothesis of d-wise independence is

$$H_0 : \ F_{\mathbf{j}}(y_{\mathbf{j}}) = y_{j_1} \ldots y_{j_d} \text{ holds for all } y_1, \ldots, y_d \in (0,1) \text{ and } \mathbf{j} \in \mathcal{A}_d, \tag{8.13}$$

where $\mathcal{A}_d = \{\mathbf{j} = (j_1, \ldots, j_d) : j_1 < \cdots < j_d \le p\}$. Pairwise and triple-wise independence correspond to $d = 2$ and $d = 3$, respectively. We estimate $F_{\mathbf{j}}(y_{\mathbf{j}})$ by the empirical cdf

$$\hat{F}_{\mathbf{j}}(y_{\mathbf{j}}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{Y_{i,\mathbf{j}} \le y_{\mathbf{j}}}, \tag{8.14}$$

where the notation $Y_{i,\mathbf{j}} \le y_{\mathbf{j}}$ means $Y_{i,j_h} \le y_{j_h}$ for all $h = 1, \ldots, d$. Let $y_{\mathbf{m}_1}, \ldots, y_{\mathbf{m}_N}$, $N \to \infty$, be a dense set of $[0,1]^d$. For example, we can choose them to be the lattice set $\{1/K, \ldots, (K-1)/K\}^d$ with $N = (K-1)^d$. Let X_i, $1 \le i \le n$, be the $Np!/(d!(p-d)!)$-dimensional vector with the $(\ell\mathbf{j})$th component being $\mathbf{1}_{Y_{i,\mathbf{j}} \le y_{\mathbf{m}_\ell}} - \prod_{h \in \mathbf{m}_\ell} y_h$, $1 \le \ell \le N$, $\mathbf{j} \in \mathcal{A}_d$. Then the L^2-based test for (8.13) on the dense set $(y_{\mathbf{m}_\ell})_{\ell=1}^N$ has the form $n|\bar{X}|_2^2$.

8.2.3 Analysis of Variance

Consider the following two-way ANOVA model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \varepsilon_{ijk}, i = 1, \ldots, I, j = 1, \ldots, J, k = 1, \ldots, K, \tag{8.15}$$

where μ is the grand mean, α_i and β_j are the main effects from the first and the second factors, respectively, and δ_{ij} are the interaction effect. Assume that $(Y_{ijk})_{i \le I, j \le J}$, $k = 1, \ldots, K$, are i.i.d. Consider the hypothesis of interaction:

$$H_0 : \delta_{ij} = 0 \text{ for all } i = 1, \ldots, I,\ j = 1, \ldots, J. \tag{8.16}$$

In the classical ANOVA procedure, one assumes that ε_{ijk}, $i \le I, j \le J$, are i.i.d. $N(0, \sigma^2)$ and makes use of the fact that the sum of squares

$$SS_I = \sum_{i=1}^{I}\sum_{j=1}^{J} (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot})^2 \tag{8.17}$$

is distributed as $\sigma^2 \chi^2_{(I-1)(J-1)}$. Here $\bar{Y}_{ij\cdot} = K^{-1}\sum_{k=1}^{K} Y_{ijk}$ and other sample averages $\bar{Y}_{i\cdot\cdot}$, $\bar{Y}_{\cdot j\cdot}$ and $\bar{Y}_{\cdot\cdot\cdot}$ are similarly defined. The null hypothesis H_0 is rejected at level $\alpha \in (0,1)$ if

$$\frac{SS_I}{(I-1)(J-1)} > SS_E F_{(I-1)(J-1),IJ(K-1),1-\alpha} \tag{8.18}$$

where $F_{(I-1)(J-1),IJ(K-1),1-\alpha}$ is the $(1-\alpha)$th quantile of the F-distribution $F_{(I-1)(J-1),IJ(K-1)}$ and

$$SS_E = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ijk} - \bar{Y}_{ij\cdot})^2}{IJ(K-1)} \tag{8.19}$$

is an estimate of σ^2.

The classical ANOVA procedure can be invalid when the assumption that ε_{ijk}, $i \le I, j \le J$ are i.i.d. $N(0, \sigma^2)$ is violated. In the latter case SS_I may no longer have a χ^2 distribution. However we can still approximate the distribution of SS_I in terms of (8.7). For $a = (i, j)$ let $X_{ak} = \bar{Y}_{ijk} - \bar{Y}_{i\cdot k} - \bar{Y}_{\cdot jk} + \bar{Y}_{\cdot\cdot k}$. Then $SS_I = \sum_{a\in\mathcal{A}} \bar{X}_{a\cdot}^2$, where $\bar{X}_{a\cdot} = K^{-1}\sum_{k=1}^{K} X_{ak}$.

8.3 Tests Based on L^∞ Norms

Fan et al. (2007) considered the L^∞ norm based test of (8.1) with the form

$$M_n = \max_{j\le p} \frac{\sqrt{n}|\hat{\mu}_j - \mu_j|}{\hat{\sigma}_j}, \text{ where } \hat{\sigma}_j^2 = \frac{1}{n}\sum_{i=1}^{n}(X_{ij} - \hat{\mu}_j)^2. \tag{8.20}$$

Assume that the dimension p satisfies

$$\log p = o(n^{1/3}) \tag{8.21}$$

and the uniform bounded third moment condition

$$\max_{j\le p} E|X_{ij} - \mu_j|^3 = O(1). \tag{8.22}$$

Let Φ be the standard normal cumulative distribution function and $z_\alpha = \Phi^{-1}(\alpha)$. Then

$$P(M_n \ge z_{1-\alpha/(2p)}) \le \alpha + o(1). \tag{8.23}$$

Namely, if we perform the test by rejecting H_0 of (8.1) whenever $M_n \ge z_{1-\alpha/(2p)}$, the familywise type I error of the latter test is asymptotically bounded by α. As a finite sample correction, the cutoff value $z_{1-\alpha/(2p)}$ in (8.23) can be replaced by the t-distribution quantile $t_{n-1,1-\alpha/(2p)}$ with degree of freedom $n-1$, noting that $(n-1)^{1/2}\hat{\mu}_j/\hat{\sigma}_j \sim t_{n-1}$ if X_{ij} are Gaussian. Due to the Bonferroni correction, the test by Fan et al. (2007) can be quite conservative if the dependence among entries of X_i is strong. For example, if $X_{i1} = X_{i2} = \cdots = X_{ip}$, then instead of using the cutoff value $z_{1-\alpha/(2p)}$, one should use $z_{1-\alpha/2}$, since the cutoff value $z_{1-\alpha/(2p)}$ leads to

the extremely conservative type I error $\alpha/(2p)$. If entries of X_i are independent and X_i is Gaussian, then the type I error is $1-(1-\alpha/p)^p \to 1-e^{-\alpha}$ and it is slightly conservative. For example, when $\alpha = 0.05$, $1-e^{-\alpha} = 0.04877058$.

Liu and Shao (2013) obtained Gumbel convergence of M_n under the following conditions: (1) for some $r > 3$, the uniform bounded rth moment conditions $\max_{j\le p} E|X_{ij}-\mu_j|^r = O(1)$ holds, which is slightly stronger than (8.22) and (2) weak dependence among entries of X_i. For $\Sigma = (\sigma_{jk})_{j,k\le p}$, assume the correlation matrix $R = (r_{jk})_{j,k\le p}$ with $r_{jk} = \sigma_{jk}/(\sigma_{jj}^{1/2}\sigma_{kk}^{1/2})$ has the property: for some $\gamma > 0$,

$$\max \#\{j \le p : |r_{jk}| \ge (\log p)^{-1-\gamma}\} = O(p^{\rho}) \tag{8.24}$$

holds for all $\rho > 0$. Then under (8.21), Theorem 3.1 in Liu and Shao (2013) asserts the Gumbel convergence

$$M_n - 2\log p + \log\log p \Rightarrow \mathcal{G}, \tag{8.25}$$

where $\mathcal{G}$ follows the Gumbel distribution $P(\mathcal{G} \le y) = \exp(-e^{-y/2}/\pi^{1/2})$. By (8.25), one can reject H_0 in (8.1) at level $\alpha \in (0,1)$ based on the L^∞ norm test

$$\max_{j\le p} \frac{\sqrt{n}|\hat{\mu}_j|}{\hat{\sigma}_j} > 2\log p - \log\log p + g_{1-\alpha}, \tag{8.26}$$

where $g_{1-\alpha}$ is chosen such that $P(\mathcal{G} \le g_{1-\alpha}) = 1-\alpha$. Clearly the latter test has an asymptotically correct size.

Applying Theorem 2.2 in Chernozhukov et al. (2014), we can have the following Gaussian approximation result. Assume that there exist constants $c_1, c_2 > 0$ such that $c_1 \le E(X_{ij}-\mu_j)^2 \le c_2$ holds for all $j \le p$ and assume that $u = u_{n,p}$ satisfies

$$P\left[\max_{j\le p} |X_{1j}-\mu_j| \ge u\right] = o(n^{-1}) \tag{8.27}$$

Let $m_k = \max_{j\le p}(E|X_{1j}-\mu_j|^k)^{1/k}$ and further assume that

$$n^{-1/8}(m_3^{3/4} + m_4^{1/2})(\log(pn))^{7/8} + n^{-1/2}(\log(pn))^{3/2}u \to 0. \tag{8.28}$$

Let $Z \sim N(0,R)$. Then we have the Gaussian approximation result: as $n \to \infty$

$$\sup_t |P(M_n \ge t) - P(|Z|_\infty \ge t)| \to 0. \tag{8.29}$$

Let $t_{1-\alpha}$ be the $(1-\alpha)$th quantile of $|Z|_\infty$. The Gaussian approximation (8.29) leads to L^∞ norm based test: H_0 is rejected at level α if $\max_{j\le p}\sqrt{n}|\hat{\mu}_j|/\hat{\sigma}_j \ge t_{1-\alpha}$. In comparison with the result in Fan et al. (2007), the latter test has an asymptotically correct size and it is dependence adjusted. To obtain an estimate for the cutoff value $t_{1-\alpha}$, Chernozhukov et al. (2014) proposed a Gaussian Multiplier Bootstrap (GMB)

method. Given $X_1, \ldots, X_n$, let $\hat{t}_{1-\alpha}$ be such that

$$P\left(\max_{j\le p} n^{-1/2}|\sum_{i=1}^{n} X_{ij}e_i| \ge \hat{t}_{1-\alpha}|X_1, \ldots, X_n\right) = \alpha, \tag{8.30}$$

where e_i are i.i.d. $N(0,1)$ random variables independent of $(X_{ij})_{i\ge 1, j\ge 1}$. Note that $\hat{t}_{1-\alpha}$ can be numerically calculated by extensive Monte Carlo simulations. In Sect. 8.5 we shall propose a Hadamard matrix and a Rademacher weighted approaches. The simulation study in Sect. 8.6 shows that, for finite-sample performance, the latter approach gives a more accurate size than the method based on Gaussian Multiplier Bootstrap (8.30).

Chen et al. (2016) generalized Fan, Hall and Yao's L^∞ norm to high-dimensional dependent vectors. Assume that $(X_i)_{i\in\mathbb{Z}}$ is a p-dimensional stationary process of the form

$$X_t = G(\mathcal{F}_t) = (G_1(\mathcal{F}_t), \ldots, G_p(\mathcal{F}_t))^T, \tag{8.31}$$

where ε_t, $t \in \mathbb{Z}$, are i.i.d. random variables, $\mathcal{F}_t = (\ldots, \varepsilon_{t-1}, \varepsilon_t)$ and $G(\cdot)$ is a measurable function such that X_t is well-defined. Assume that the long-run covariance matrix

$$\Sigma_\infty = \sum_{i=-\infty}^{\infty} \mathrm{cov}(X_0, X_i) = (\omega_{jl})_{j,l\le p} \tag{8.32}$$

exists. Let $\varepsilon_i^*, \varepsilon_j, i, j \in \mathbb{Z}$, be i.i.d. random variables. Assume that X_t has finite rth moment, $r > 2$. Define the functional dependence measures (see, Wu 2005, 2011) as

$$\theta_r(m) = \max_{j\le p} \|X_{ij} - G_j(\ldots, \varepsilon_{i-m-2}, \varepsilon_{i-m-1}, \varepsilon_{i-m}^*, \varepsilon_{i-m+1}, \ldots, \varepsilon_i)\|_r. \tag{8.33}$$

If X_i are i.i.d., then $\Sigma_\infty = \Sigma$ and $\theta_r(m) = 0$ if $m \ge 1$. We say that (X_t) is *geometric moment contraction* (GMC; see Wu and Shao 2004) if there exist $\rho \in (0,1)$ and $a_1 > 0$ such that

$$\theta_r(m) \le a_1\rho^m = a_1 e^{-a_2 m} \text{ with } a_2 = -\log\rho. \tag{8.34}$$

Let $\mu = EX_t$. To test the hypothesis H_0 in (8.1), Chen et al. (2016) introduced the following dependence-adjusted versions of Fan, Hall, and Yao's M_n. Let $n = mk$, where $m \asymp n^{1/4}$ and blocks $B_l = \{i : m(l-1)+1 \le i \le ml\}$. Let $Y_{lj} = \sum_{i\in B_l} X_{ij}$, $1 \le j \le p$, $1 \le l \le k$, be the block sums. Define the block-normalized sum

$$M_n^\circ = \max_{j\le p} \frac{\sqrt{n}|\hat{\mu}_j - \mu_j|}{\hat{\sigma}_j^\circ}, \text{ where } (\hat{\sigma}_j^\circ)^2 = \frac{1}{mk}\sum_{l=1}^{k}(Y_{lj} - m\hat{\mu}_j)^2, \tag{8.35}$$

and the interlacing normalized sum: let $k^* = k/2$, $\mu_j^\dagger = (mk^*)^{-1}\sum_{l=1}^{k^*} Y_{2lj}$ and

$$M_n^\dagger = \max_{j\le p} \frac{\sqrt{n/2}|\mu_j^\dagger - \mu_j|}{\hat{\sigma}_j^\dagger}, \text{ where } (\hat{\sigma}_j^\dagger)^2 = \frac{1}{mk^*}\sum_{l=1}^{k^*}(Y_{2lj} - m\mu_j^\dagger)^2. \quad (8.36)$$

By Chen et al. (2016), we have the following result: Assume exists a constant $\zeta > 0$ such that the long-run variance $\omega_{jj} \ge \zeta$ for $j \le p$, (8.34) holds with $r = 3$, and

$$\log p = o(n^{1/4}). \quad (8.37)$$

Then (8.23) holds for both the block-normalized sum M_n° and the interlacing normalized sum $M_n^\dagger$. Note that, while (8.37) still allows ultra high dimensions, due to dependence, the allowed dimension p in condition (8.37) is smaller than the one in (8.21). Additionally, if the GMC (8.34) holds with some $r > 3$, (8.24) holds with the long-run correlation matrix $R = D^{-1/2}\Sigma_\infty D^{-1/2}$, where $D = \mathrm{diag}(\Sigma_\infty)$, and for some $0 < \tau < 1/4$,

$$\log p = o(n^\tau), \quad (8.38)$$

then we have the Gumbel convergence for the interlacing normalized sum:

$$M_n^\dagger - 2\log p + \log\log p \Rightarrow \mathcal{G}, \quad (8.39)$$

where $\mathcal{G}$ is given in (8.25). Similarly as (8.26), one can perform the following test which has an asymptotically correct size: we reject H_0 in (8.1) at level $\alpha \in (0, 1)$ if

$$\max_{j\le p} \frac{\sqrt{n/2}|\mu_j^\dagger|}{\hat{\sigma}_j^\dagger} > 2\log p - \log\log p + g_{1-\alpha}. \quad (8.40)$$

8.4 Tests Based on L^2 Norms

In this section we shall consider the test which is based on the L^2 functional with $g(x) = \sum_{j=1}^p x_j^2$. Let $\lambda_1 \ge \cdots \ge \lambda_p \ge 0$ be the eigenvalues of Σ. For $Z \sim N(0, \Sigma)$, we have the distributional equality $g(Z) = Z^T Z =_{\mathcal{D}} \sum_{j=1}^p \lambda_j \eta_j^2$, where η_j are i.i.d. standard $N(0, 1)$ random variables. Let $f_k = (\sum_{j=1}^p \lambda_j^k)^{1/k}$, $k > 0$, and $f = f_2$. Then $Eg(Z) = f_1 = \mathrm{tr}(\Sigma)$ and $\mathrm{var}(g(Z)) = 2f^2$. Xu et al. (2014) provide a sufficient condition for the invariance principle (8.7) with the quadratic functional g. For some $0 < \delta \le 1$ let $q = 2 + \delta$.

Condition 1 *Let $\delta > 0$. Assume $EX_1 = 0$, $E|X_1|^{2q} < \infty$ and let*

$$K_\delta(X)^q := E\left|\frac{|X_1|_2^2 - f_1}{f}\right|^q < \infty \tag{8.41}$$

$$D_\delta(X)^q := E\left|\frac{X_1^T X_2}{f}\right|^q < \infty. \tag{8.42}$$

Observe that Condition 1, (8.41) and (8.42) are Lyapunov-type conditions. Assume that

$$\frac{K_0(X)^2}{n} + \frac{K_\delta(X)^q}{n^{q-1}} + \frac{E(X_1^T \Sigma X_1)^{q/2}}{n^{\delta/2} f^q} + \frac{D_\delta(X)^q}{n^\delta} \to 0 \text{ as } n \to \infty. \tag{8.43}$$

Then (8.7) holds (cf Xu et al. 2014). Consequently we have

$$\sup_{t\in\mathbb{R}} |P((n|\bar{X}_n|_2^2 - f_1)/f \le t) - P(V \le t)| \to 0, \text{ where } V = \sum_{j=1}^{p} f^{-1}\lambda_j(\eta_j^2 - 1). \tag{8.44}$$

In the literature, researchers primarily focus on developing the central limit theorem

$$R_n := \frac{n|\bar{X}_n|_2^2 - f_1}{f} = \frac{n\bar{X}_n^T\bar{X}_n - f_1}{f} \Rightarrow N(0, 2) \tag{8.45}$$

or its modified version; see, for example, Bai and Saranadasa (1996), Chen and Qin (2010) and Srivastava (2009). Xu et al. (2014) clarified an important issue on the CLT of R_n. By the Lindeberg–Feller central limit theorem, $V \Rightarrow N(0, 2)$ as $p \to \infty$ holds if and only if $\lambda_1/f \to 0$. The distributional approximation (8.44) indicates that, if λ_1/f does not go to 0, then the central limit theorem cannot hold for R_n.

Let $t_{1-\alpha}$ be the $(1-\alpha)$th quantile of $g(Z) = |Z|^2 = Z^T Z$. By (8.7) we can reject (8.1) at level $\alpha \in (0, 1)$ if

$$n|\bar{X}_n|^2 > t_{1-\alpha} \tag{8.46}$$

To calculate $t_{1-\alpha}$, one needs to know the eigenvalues $\lambda_1, \ldots, \lambda_p$. However, estimation of those eigenvalues is a very challenging problem, in particular if one does not impose certain structural assumptions on Σ. In Sect. 8.5.2 we shall propose a half-sampling based approach which does not need estimation of the covariance matrix Σ.

The L^∞ based tests discussed in Sect. 8.3 have a good power when the alternative consists of few large signals. If the signals are small and have a similar magnitude, then the L^2 test is more powerful. To this end, assume that there exists a constant $c > 0$ and a small $\delta > 0$ such that $c\delta \le \mu_j \le \delta/c$ holds for all $j = 1, \ldots, p$. We can interpret δ as the departure parameter (from the null H_0 with $\mu = 0$). For the L^∞-based test to have power approaching to 1, one necessarily requires that $\sqrt{n}\delta \to \infty$.

Elementary calculation shows that, under the much weaker condition $np^{1/2}\delta^2 \to \infty$, then the power of the L^2 based test, or the probability that event (8.46) occurs going to one. In the latter condition, larger dimension p is actually a blessing as it requires a smaller departure δ.

8.5 Asymptotic Theory

In Sects. 8.3 and 8.4, we discussed the classical L^∞ and L^2 functionals, respectively. For a general testing functional, we have the following invariance principle (cf Theorem 1), which asserts that functionals of sample means of non-Gaussian random vectors $X_1, X_2, \ldots$ can be approximated by those of Gaussian vectors $Z_1, Z_2, \ldots$ with same covariance structure. Assume $g \in \mathbb{C}^3(\mathbb{R}^p)$. For $\mathbf{x} = (x_1, \ldots, x_p)^T$ write $g_j = g_j(\mathbf{x}) = \partial g(\mathbf{x})/\partial x_j$. Similarly we define the partial derivatives g_{jk} and g_{jkl}. For all $j, k, l = 1, \ldots, p$, assume that

$$\kappa_{jkl} := \sup_{\mathbf{x}\in\mathbb{R}^p} (|g_j g_k g_l| + |g_{jk} g_l| + |g_{jl} g_k| + |g_{kl} g_j| + |g_{jkl}|) < \infty. \tag{8.47}$$

For $Z_1 \sim N(0, \Sigma)$ write $Z_1 = (Z_{11}, \ldots, Z_{1p})^T$. Define

$$\mathcal{K}_p = \sum_{j,k,l=1}^{p} \kappa_{jkl}(E|X_{1j}X_{1k}X_{1l}| + E|Z_{1j}Z_{1k}Z_{1l}|). \tag{8.48}$$

For $g(Z_1) =_{\mathcal{D}} g(\sqrt{n}\bar{Z}_n)$, we assume that its c.d.f. $F(t) = P[g(Z) \le t]$ is Hölder continuous: there exists $\ell_p > 0$, index $\alpha > 0$, such that for all $\psi > 0$, the concentration function

$$\sup_{t\in\mathbb{R}} P(t \le g(Z_1) \le t + \psi) \le \ell_p \psi^\alpha. \tag{8.49}$$

Theorem 1 (Lou and Wu (2018)) *Assume (8.47), (8.49) and* $\mathcal{K}_p \ell_p^{3/\alpha} = o(\sqrt{n})$. *Then*

$$\sup_{t\in\mathbb{R}} |P[g(\sqrt{n}(\bar{X}_n - \mu)) \le t] - P[g(\sqrt{n}\bar{Z}_n) \le t]| = O(\ell_p^3 \mathcal{K}_p^\alpha n^{-\alpha/2}) \to 0. \tag{8.50}$$

To apply Theorem 1 for hypothesis testing, we need to know the c.d.f. $F(t) = P[g(Z) \le t]$. Note that $F(\cdot)$ depends on g and the covariance matrix Σ. Thus we can also write $F(\cdot) = F_{g,\Sigma}(\cdot)$. If Σ is known, the distribution of $g(Z)$ is completely known and its cdf $F(t) = P[g(Z) \le t]$ can be calculated either analytically or by extensive Monte Carlo simulations. Let $t_{1-\alpha}$, $0 < \alpha < 1$, be the $(1-\alpha)$th quantile of $g(Z)$. Namely

$$P[g(Z) > t_{1-\alpha}] = \alpha. \tag{8.51}$$

Then the null hypothesis H_0 in (8.1) is rejected at level α if the test statistic $T_n = g(\sqrt{n}\bar{X}_n) > t_{1-\alpha}$. This test has asymptotically correct size α. Additionally, the $(1-\alpha)$ confidence region for μ can be constructed as

$$\{\mu \in \mathbb{R}^p : g(\sqrt{n}(\bar{X}_n - \mu)) \le t_{1-\alpha}\} = \{\bar{X}_n + \nu \in \mathbb{R}^p : g(\sqrt{n}\nu) \le t_{1-\alpha}\}. \quad (8.52)$$

If Σ is not known, as a straightforward way to approximate $F(t) = F_{g,\Sigma}(t)$, one may use an estimate $\tilde{\Sigma}$ so that $F_{g,\Sigma}(t)$ can be approximated by $F_{g,\tilde{\Sigma}}(t)$. Here we do not adopt this approach for the following two reasons. First, it can be quite difficult to consistently estimate Σ without assuming sparseness or other structural conditions. The latter assumptions are widely used in the literature; see, for example, Bickel and Levina (2008a), Bickel and Levina (2008b), Cai et al. (2011) and Fan et al. (2013). Second, it is difficult to quantify the difference $F_{g,\tilde{\Sigma}}(\cdot) - F(\cdot)$ based on operator norm or other type of matrix convergence of the estimate $\tilde{\Sigma}$. Xu et al. (2014) argued that, for the L^2 test with $g(x) = \sum_{j=1}^p x_j^2$, one needs to use the normalized consistency of $\tilde{\Sigma}$, instead of the widely used operator norm consistency. We propose using half-sampling and balanced Rademacher schemes.

8.5.1 *Preamble: i.i.d. Gaussian Data*

In practice, however, the covariance matrix Σ is typical unknown. Assume at the outset that $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \Sigma)$ vectors. Assume that $n = 4m$, where m is a positive integer. Then we can estimate the cumulative distribution function $F(t) = P[g(Z) \le t]$ by using Hadamard matrices (see, Georgiou et al. 2003; Hedayat and Wallis 1978; Yarlagadda and Hershey 1997). We say that H is an $n \times n$ Hadamard matrix if its first row consisting all 1s, and all its entries taking values 1 or -1 such that

$$HH^T = nI_n, \quad (8.53)$$

where I_n is the $n \times n$ identity matrix. Let

$$Y_j = \frac{1}{\sqrt{n}} \sum_{i=1}^n H_{ji} X_i, \ j = 1, \ldots, n. \quad (8.54)$$

By (8.53), we have $\sum_{i=1}^n H_{ji} = 0$ for $2 \le j \le n$ and $\sum_{i=1}^n H_{ji} H_{j'i} = 0$ if $j \ne j'$. Since $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \Sigma)$, it is clear that $Y_2, \ldots, Y_n$ are also i.i.d. $N(0, \Sigma)$ vectors. Hence the random variables $g(Y_2), \ldots, g(Y_n)$ are independent and identically distributed as $g(Z)$. Therefore we can construct the empirical cumulative distribution function

$$\hat{F}_n(t) = \frac{1}{n-1} \sum_{j=2}^n \mathbf{1}_{g(Y_j) \le t}, \quad (8.55)$$

which converges uniformly to $F(t)$ as $n \to \infty$, and $t_{1-\alpha}$ can be estimated by $\hat{t}_{1-\alpha} = \hat{F}_n^{-1}(1-\alpha)$, the $(1-\alpha)$th empirical quantile of $\hat{F}_n(\cdot)$. As an important feature of the latter method, one does not need to estimate the covariance matrix Σ, the nuisance parameter. In combinatorial experiment design, however, it is highly nontrivial to construct Hadamard matrices. If n is a power of 2, then one can simply apply Sylvester's construction. The Hadamard conjecture states that a Hadamard matrix of order n exists when $4|n$. The latter problem is still open. For example, it is unclear whether a Hadamard matrix exists when $n = 668$ (see Brent et al. 2015).

8.5.2 *Rademacher Weighted Differencing*

To circumvent the existence problem of Hadamard matrices in Sect. 8.5.1, we shall construct asymptotically independent realizations by using Rademacher random variables. Let $\varepsilon_{jk}, j, k \in \mathbb{Z}$, independent of $(X_i)_{i\geq 1}$, be i.i.d. Bernoulli random variables with $P(\varepsilon_{jk} = 1) = P(\varepsilon_{jk} = -1) = 1/2$. Define the Rademacher weighted differences

$$Y_j = D(A_j), \text{ where } D(A) = \frac{|A|^{1/2}(n-|A|)^{1/2}}{n^{1/2}} \left(\frac{\sum_{i\in A} X_i}{|A|} - \frac{\sum_{i\in\{1,\ldots,n\}-A} X_i}{n-|A|} \right), \tag{8.56}$$

where the random set

$$A_j = \{1 \leq i \leq n : \varepsilon_{ji} = 1\}. \tag{8.57}$$

When defining Y_j, we require that A_j satisfies $|A_j| \neq 0$ and $|A_j| \neq n$. By the Hoeffding inequality, $|A_j|$ concentrates around $n/2$ in the sense that, for $u \geq 0$, $P(||A_j| - n/2| \geq u) \leq 2\exp(-2u^2/n)$. Alternatively, we consider the balanced Rademacher weighted differencing: let $A_1^\circ, A_2^\circ, \ldots$ be simple random sample drawn equally likely from $\mathcal{A}_m = \{A \subset \{1, \ldots, n\} : |A| = m\}$, where $m = \lfloor n/2 \rfloor$. Similarly as Y_j in (8.56), we define

$$Y_j^\circ = D(A_j^\circ). \tag{8.58}$$

Clearly, given A_j (resp. A_j°), Y_j (resp. Y_j°) has mean 0 and covariance matrix Σ. Based on Y_j in (8.56) (resp. Y_j° in (8.58)), define the empirical distribution functions

$$\hat{F}_N(t) = \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}_{g(Y_j)\leq t}, \tag{8.59}$$

where $N \to \infty$ and

$$\hat{F}_N^\circ(t) = \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}_{g(Y_j^\circ)\le t}. \tag{8.60}$$

For sets $A, B \subset \{1, \ldots, n\}$, let $A^c = \{1, \ldots, n\} - A$, $B^c = \{1, \ldots, n\} - B$ and

$$d(A,B) = \max\left\{||A\cap B| - \frac{n}{4}|,\ ||A^c \cap B| - \frac{n}{4}|,\ ||A\cap B^c| - \frac{n}{4}|,\ ||A^c \cap B^c| - \frac{n}{4}|\right\}.$$

If A, B are chosen according to a Hadamard matrix, then $d(A, B) = 0$. Assume that

$$d(A,B) \le 0.1n. \tag{8.61}$$

Then there exists an absolute constant $c > 0$ such that

$$\mathrm{cov}(D(A), D(B)) = \delta\Sigma, \text{ where } |\delta| \le c\frac{d(A,B)}{n}. \tag{8.62}$$

Again by the Hoeffding inequality, if we choose A_1, A_2 according to (8.57), there exists absolute constants $c_1, c_2 > 0$ such that $P(d(A_1, A_2) \ge u) \le c_1 \exp(-c_2u^2/n)$, indicating that (8.61) holds with probability close to 1, $d(A_1, A_2) = O_P(n^{1/2})$ and hence the weak orthogonality with $\delta(A_1, A_2) = O_P(n^{-1/2})$.

Theorem 2 (Lou and Wu (2018)) *Under conditions of Theorem 1, we have* $\sup_t |\hat{F}_N^\circ(t) - F(t)| \to 0$ *in probability as* $N \to \infty$.

8.5.3 Calculating the Power

The asymptotic power expression is

$$B(\mu) = P[g(Z + \sqrt{n}\mu) \ge t_{1-\alpha}]. \tag{8.63}$$

Given the sample $X_1, \ldots, X_n$ whose mean vector μ may not necessarily be 0, based on the estimated $\hat{t}_{1-\alpha}$ from the empirical cumulative distribution functions (8.59) and (8.60), we can actually estimate the power function by the following:

$$\begin{aligned}\hat{B}(v) &= \hat{P}(g(D(A_j^\circ) + \sqrt{n}v) \ge \hat{t}_{1-\alpha}|X_1, \ldots, X_n)\\ &= \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}_{g(D(A_j^\circ)+\sqrt{n}v)\ge\hat{t}_{1-\alpha}}.\end{aligned} \tag{8.64}$$

8.5.4 An Algorithm with General Testing Functionals

For ease of application, we shall in this section provide details of testing the hypothesis H_0 in (8.1) using the Rademacher weighting scheme described in Sect. 8.5.2.

To construct a confidence region for μ, one can use (8.52) with $t_{1-\alpha}$ therein replaced by the empirical quantile $\hat{t}^{\circ}_{1-\alpha}$.

8.6 Numerical Experiments

In this section, we shall perform a simulation study and evaluate the finite-sample performance of our Algorithm 1 with $\hat{F}^{\circ}_N(t)$ defined in (8.60). Tests for mean vectors and covariance matrices are considered in Sects. 8.6.1 and 8.6.2, respectively. Section 8.6.3 contains a real data application on testing correlations between different pathways of a pancreatic ductal adenocarcinoma dataset.

8.6.1 Test of Mean Vectors

We consider three different testing functionals: for $x = (x_1, \ldots, x_p)^\top \in \mathbb{R}^p$, let

$$g_1(x) = \max_{j \le p} |x_j|, \quad g_2(x) = \sum_{j=1}^{p} |x_j|^2, \quad g_3(x) = \sup_{c \ge 0} \left\{ c^2 \sum_{j=1}^{p} |x_j|^2 \mathbf{1}_{|x_j| \ge c} \right\}.$$

For the L^∞ form $g_1(x)$, four different testing procedures are compared: the procedure using our Algorithm 1 with $\hat{F}^{\circ}_N(\cdot)$ replaced by $\hat{F}_N(\cdot)$; cf (8.59); or by

$$\hat{F}^{\dagger}_N(t) = \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}_{g(Y_j^{\dagger}) \le t}, \text{ where } Y_j^{\dagger} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varepsilon_{ji}(X_i - \bar{X}) \tag{8.65}$$

Algorithm 1: Rademacher weighted testing procedure

1. Input $X_1, \ldots, X_n$;
2. Compute the average $\bar{X}_n$ and the test statistic $T = g(\sqrt{n}\bar{X}_n)$;
3. Choose a large N in (8.60) and obtain the empirical quantile $\hat{t}^{\circ}_{1-\alpha}$;
4. Reject H_0 at level α if $T > \hat{t}^{\circ}_{1-\alpha}$;
5. Report the p-value as $\hat{F}^{\circ}_N(T)$.

and ε_{ji} are i.i.d. Bernoulli(1/2) independent of (X_{ij}); the test of Fan et al. (2007) (FHY, see (8.20) and (8.23)) and the Gaussian Multiplier Bootstrap method in Chernozhukov et al. (2014) (CCK, see (8.30)).

For $g_2(x)$, we compare the performance of our Algorithm 1 with $\hat{F}_N^{\circ}(\cdot)$, $\hat{F}_N(\cdot)$ and $\hat{F}_N^{\dagger}(\cdot)$, and also the CLT-based procedure of Chen and Qin (2010) (CQ), which is a variant of (8.45) with the numerator $n\bar{X}_n^T\bar{X}_n - f_1$ therein replaced by $n^{-1}\sum_{i\neq j} X_i^{\top}X_j$.

The portmanteau testing functional $g_3(x)$ is a marked weighted empirical process.

For our Algorithm 1 and the Gaussian Multiplier Bootstrap method, we calculate the empirical cutoff values with $N = 4000$. For each functional, we consider two models and use $n = 40, 80$ and $p = 500, 1000$. The empirical sizes for each case are calculated based on 1000 simulations.

Example 1 (Factor Model) Let Z_{ij} be i.i.d. $N(0, 1)$ and consider

$$X_i = (Z_{i1}, \ldots, Z_{ip})^{\top} + p^{\delta}(Z_{i0}, \ldots, Z_{i0})^{\top}, \quad i = 1, \ldots, n, \tag{8.66}$$

Then X_i are i.i.d. $N(0, \Sigma)$ with $\Sigma = \mathrm{I}_p + p^{2\delta}\mathbf{1}\mathbf{1}^{\top}$, where $\mathbf{1} = (1, \ldots, 1)^{\top}$. Larger δ implies stronger correlation among the entries $X_{i1}, \ldots, X_{ip}$.

Table 8.1 reports empirical sizes for the factor model with $g_1(\cdot)$ at the 5% significance level. For each choice of p, n, and δ, our Algorithm 1 with $\hat{F}_N^{\circ}(\cdot)$ and $\hat{F}_N(\cdot)$ perform reasonably well, while the empirical sizes using $\hat{F}_N^{\dagger}(\cdot)$ are generally slightly larger than 5%. The empirical sizes using Chernozhukov et al.'s (8.30) or Fan et al.'s (8.23) are substantially different from the nominal level 5%. For large δ, as expected, the procedure of Fan, Hall, and Yao can be very conservative.

The empirical sizes for the factor model using $g_2(\cdot)$ are summarized in Table 8.2. Our Algorithm 1 with $\hat{F}_N^{\circ}(\cdot)$ and $\hat{F}_N(\cdot)$ perform quite well. The empirical sizes for Chen and Qin's procedure deviate significantly from 5%. This can be explained by the fact that CLT of type (8.45) is no longer valid for model (8.66); see the discussion following (8.45) and Theorem 2.2 in Xu et al. (2014).

When using functional $g_3(x)$, our Algorithm 1 with $\hat{F}_N^{\circ}(\cdot)$ and $\hat{F}_N(\cdot)$ perform slightly better than $\hat{F}_N^{\dagger}(\cdot)$ and approximate the nominal 5% level well (Table 8.3).

Table 8.1 Empirical sizes for the factor model (8.66) with $g_1(\cdot)$

		$n = 40$					$n = 80$				
p	δ	$\hat{F}_N^{\circ}$	CCK	$\hat{F}_N$	FHY	$\hat{F}_N^{\dagger}$	$\hat{F}_N^{\circ}$	CCK	$\hat{F}_N$	FHY	$\hat{F}_N^{\dagger}$
500	−0.05	0.053	0.028	0.052	0.028	0.059	0.053	0.037	0.052	0.031	0.055
1000		0.052	0.023	0.052	0.035	0.057	0.051	0.036	0.051	0.034	0.053
500	0.05	0.051	0.034	0.054	0.014	0.064	0.047	0.030	0.044	0.018	0.047
1000		0.057	0.035	0.058	0.011	0.063	0.053	0.044	0.055	0.015	0.056
500	0.1	0.046	0.026	0.048	0.009	0.055	0.053	0.042	0.054	0.007	0.056
1000		0.059	0.041	0.059	0.007	0.063	0.052	0.045	0.054	0.008	0.056

Table 8.2 Empirical sizes for the factor model (8.66) using functional $g_2(x)$

		$n = 40$				$n = 80$			
p	δ	CQ	$\hat{F}_N^{\circ}$	$\hat{F}_N$	$\hat{F}_N^{\dagger}$	CQ	$\hat{F}_N^{\circ}$	$\hat{F}_N$	$\hat{F}_N^{\dagger}$
500	−0.05	0.078	0.055	0.061	0.066	0.063	0.048	0.047	0.048
1000		0.081	0.063	0.066	0.072	0.066	0.050	0.049	0.053
500	0.05	0.074	0.054	0.054	0.059	0.067	0.052	0.053	0.054
1000		0.075	0.054	0.052	0.056	0.076	0.058	0.057	0.059
500	0.1	0.067	0.049	0.051	0.052	0.068	0.055	0.052	0.056
1000		0.083	0.064	0.064	0.067	0.068	0.048	0.051	0.051

Table 8.3 Empirical sizes for the factor model (8.66) using functional $g_3(x)$

		$n = 40$			$n = 80$		
p	δ	$\hat{F}_N^{\circ}$	$\hat{F}_N$	$\hat{F}_N$	$\hat{F}_N^{\circ}$	$\hat{F}_N$	$\hat{F}_N^{\dagger}$
500	−0.05	0.061	0.059	0.066	0.049	0.048	0.049
1000		0.062	0.064	0.073	0.058	0.059	0.063
500	0.05	0.054	0.058	0.060	0.053	0.053	0.055
1000		0.053	0.054	0.057	0.059	0.059	0.060
500	0.1	0.049	0.049	0.051	0.053	0.053	0.055
1000		0.053	0.054	0.057	0.059	0.059	0.060

Example 2 (Multivariate t-Distribution) Consider the multivariate t_ν vector

$$X_i = (X_{i1}, \ldots, X_{ip})^\top = Y_i \sqrt{\nu / W_i} \sim t_\nu(0, \Sigma), \quad i = 1, \ldots, n \tag{8.67}$$

where the degrees of freedom $\nu = 4$, $\Sigma = (\sigma_{jk})_{j,k=1}^p$, $\sigma_{jj} = 1$ for $j = 1, \ldots, p$ and

$$\sigma_{jk} = c|j - k|^{-d}, \quad 1 \le j \ne k \le p,$$

and $Y_i \sim N(0, \Sigma)$, $W_i \sim \chi_\nu^2$ are independent. The above covariance structure allows long-range dependence among $X_{i1}, \ldots, X_{ip}$; see Veillette and Taqqu (2013).

We summarize the simulated sizes for model (8.67) in Tables 8.4, 8.5, and 8.6. As in Example 1, similar conclusions apply here. Due to long-range dependence, the procedure of Fan, Hall, and Yao appears conservative. The Gaussian Multiplier Bootstrap (8.30) yields empirical sizes that are quite different from 5%. The CLT-based procedure of Chen and Qin is severely affected by the dependence. In practice we suggest using Algorithm 1 with $\hat{F}_N^{\circ}(\cdot)$ which has a good size accuracy.

Table 8.4 Empirical sizes for multivariate t-distribution using functional $g_1(x)$

t_4		$n = 40$					$n = 80$				
c	d	$\hat{F}_N^{\dagger}$	CCK	$\hat{F}_N$	FHY	$\hat{F}_N^{\dagger}$	$\hat{F}_N^{\circ}$	CCK	$\hat{F}_N$	FHY	$\hat{F}_N^{\dagger}$
0.5	1/8	0.047	0.011	0.044	0.016	0.053	0.051	0.017	0.045	0.013	0.049
		0.059	0.015	0.056	0.014	0.061	0.055	0.017	0.055	0.022	0.059
0.5	1/4	0.057	0.010	0.055	0.023	0.061	0.050	0.016	0.050	0.022	0.053
		0.051	0.005	0.048	0.018	0.058	0.054	0.014	0.055	0.022	0.060
0.8	1/8	0.054	0.020	0.050	0.017	0.061	0.052	0.030	0.051	0.016	0.053
		0.049	0.019	0.044	0.012	0.049	0.049	0.022	0.048	0.017	0.051
0.8	1/4	0.048	0.013	0.050	0.022	0.053	0.046	0.019	0.042	0.036	0.044
		0.054	0.008	0.053	0.017	0.057	0.051	0.018	0.050	0.018	0.052

For each choice of c and d, the upper line corresponding to $p = 500$ and the second for $p = 1000$

Table 8.5 Empirical sizes for multivariate t-distribution using functional $g_2(x)$

t_4		$n = 40$				$n = 80$			
c	d	CQ	$\hat{F}_N^{\circ}$	$\hat{F}_N$	$\hat{F}_N^{\dagger}$	CQ	$\hat{F}_N^{\circ}$	$\hat{F}_N$	$\hat{F}_N^{\dagger}$
0.5	1/8	0.074	0.053	0.053	0.056	0.076	0.060	0.052	0.058
		0.073	0.055	0.050	0.054	0.077	0.062	0.061	0.064
0.5	1/4	0.067	0.052	0.044	0.051	0.073	0.055	0.054	0.057
		0.072	0.057	0.054	0.060	0.070	0.056	0.055	0.060
0.8	1/8	0.074	0.059	0.062	0.066	0.070	0.047	0.051	0.052
		0.064	0.052	0.053	0.057	0.075	0.052	0.054	0.055
0.8	1/4	0.081	0.063	0.058	0.063	0.080	0.055	0.056	0.061
		0.067	0.052	0.051	0.059	0.068	0.053	0.052	0.056

For each choice of c and d, the upper line corresponding to $p = 500$ and the second for $p = 1000$

Table 8.6 Empirical sizes for multivariate t-distribution using functional $g_3(x)$

t_4		$n = 40$			$n = 80$		
c	d	$\hat{F}_N^{\circ}$	$\hat{F}_N$	$\hat{F}_N^{\dagger}$	$\hat{F}_N^{\circ}$	$\hat{F}_N$	$\hat{F}_N^{\dagger}$
0.5	1/8	0.053	0.050	0.056	0.055	0.051	0.054
		0.050	0.049	0.056	0.059	0.055	0.060
0.5	1/4	0.052	0.048	0.053	0.056	0.056	0.060
		0.056	0.048	0.060	0.055	0.056	0.061
0.8	1/8	0.059	0.059	0.066	0.049	0.049	0.052
		0.048	0.048	0.054	0.051	0.051	0.058
0.8	1/4	0.067	0.063	0.069	0.053	0.056	0.061
		0.049	0.048	0.052	0.048	0.048	0.051

For each choice of c and d, the upper line corresponding to $p = 500$ and the second for $p = 1000$

8.6.2 Test of Covariance Matrices

8.6.2.1 Sizes Accuracy

We first consider testing for $H_{0a} : \Sigma = I$ for the following model:

$$X_{ij} = \varepsilon_{i,j}\varepsilon_{i,j+1},\ 1 \le i \le n,\ 1 \le j \le p, \tag{8.68}$$

where ε_{ij} are i.i.d. (1) standard normal; (2) centralized Gamma(4,1); and (3) the student t_5. We then study the second test $H_{0b} : \Sigma_{1,2} = 0$, by partitioning equally the entire random vector $X_i = (X_{i1}, \ldots, X_{ip})^T$ into two subvectors of $p_1 = p/2$ and $p_2 = p - p_1$. In the simulation, we generate samples of two subvectors independently according to model (8.68). We shall use Algorithm 1 with L^2 functional. Tables 8.7 and 8.8 report the simulated sizes based on 1000 replications with $N = 1000$ half-sampling implementations, and they are reasonably closed to the nominal level 5%.

8.6.2.2 Power Curve

To access the power for testing $H_0 : \Sigma = \mathrm{I}_p$ using the L^2 test, we consider the model

$$X_{ij} = \varepsilon_{i,j}\varepsilon_{i,j+1} + \rho\zeta_i,\ 1 \le i \le n,\ 1 \le j \le p, \tag{8.69}$$

where ε_{ij} and ζ_i are i.i.d. Student t_5 and ρ is chosen to be $0, 0.02, 0.04, \ldots, 0.7$. The power curve is shown in Fig. 8.1. As expected, the power increases with n.

Table 8.7 Simulated sizes of the L^2 test for H_{0a}

	$N(0,1)$		$\Gamma(4,1)$		t_5	
	p					
n	64	128	64	128	64	128
20	0.045	0.054	0.046	0.047	0.053	0.048
50	0.044	0.045	0.055	0.045	0.046	0.050
100	0.050	0.054	0.047	0.053	0.051	0.049

Table 8.8 Simulated sizes of the L^2 test for H_{0b}

	$N(0,1)$		$\Gamma(4,1)$		t_5	
	p					
n	64	128	64	128	64	128
20	0.044	0.050	0.043	0.055	0.045	0.043
50	0.045	0.043	0.049	0.044	0.053	0.045
100	0.053	0.053	0.053	0.045	0.050	0.050

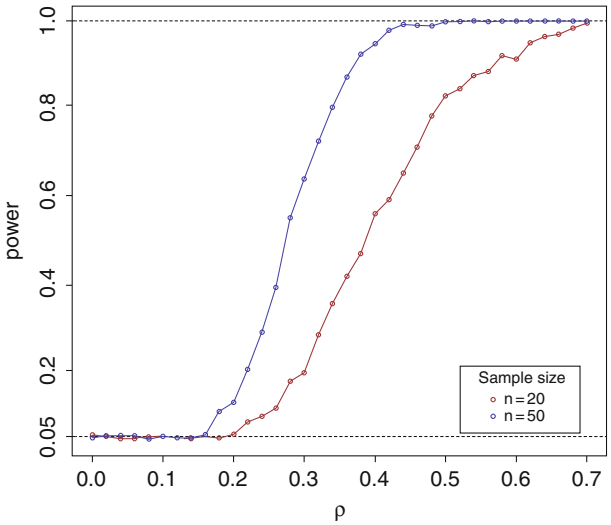

Fig. 8.1 Power curve for testing $H_0 : \Sigma = \mathrm{I}_p$ with model (8.69), and $n = 20, 50$, using the L^2 test

8.6.3 A Real Data Application

We now apply our testing procedures to a pancreatic ductal adenocarcinoma (PDAC) dataset, preprocessed from NCBI's Gene Expression Omnibus, accessible through GEO Series accession number GSE28735 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE28735). The dataset consists of two classes of gene expression levels that came from 45 pancreatic tumor patients and 45 pancreatic normal patients. There are a total of 28,869 genes. We shall test existence of correlations between two subvectors, which can be useful for identifying sets of genes which are significantly correlated.

We consider genetic pathways of the PDAC dataset. Pathways are found to be highly significantly associated with the disease even if they harbor a very small amount of individually significant genes. According to the KEGG database, the pathway "hsa05212" is relevant to pancreatic cancer. Among the 28,869 genes, 66 are mapped to this pathway. We are interested in testing whether the pathway to pancreatic cancer is correlated with some common pathways, "hsa04950" (21 genes, with name "Maturity onset diabetes of the young"), "hsa04940" (59 genes, with name "Type I diabetes mellitus"), "hsa04972" (87 genes, with name "Pancreatic secretion"). Let W_i, X_i, Y_i, and Z_i be the expression levels of individual i from the tumor group for pathways "hsa05212," "hsa04950," "hsa04940," and "hsa04972," respectively. The null hypotheses are $H_{01}^T : \mathrm{cov}(W_i, X_i) = 0_{66\times 21}$, $H_{02}^T : \mathrm{cov}(W_i, Y_i) = 0_{66\times 59}$ and $H_{03}^T : \mathrm{cov}(W_i, Z_i) = 0_{66\times 87}$. Similar null hypothesis $H_{01}^N, H_{01}^N, H_{01}^N$ can be formulated for the normal group. Our L^2 test of Algorithm 1 is compared with the Gaussian multiplier bootstrap (8.30). The results are summarized

Table 8.9 Estimated p-values of tests for covariances between pathway "pancreatic cancer" and other different pathways, based on $N = 10^6$ half-sampling implementations

Pathway	Name	Tumor patients		Normal patients	
		CCK	L^2 test	CCK	L^2 test
hsa04950	Maturity onset diabetes of the young	0.013116	0.000000	0.006618	0.000000
hsa04940	Type I diabetes mellitus	0.066270	0.000000	0.074014	0.002327
hsa04972	Pancreatic secretion	0.063291	0.000003	0.095358	0.001189

in Table 8.9. The CCK test is not able to reject the null hypothesis H_{03} at 5% level since it gives a p-value of 0.063291. However using the L^2 test, H_{03} is rejected, suggesting that there is a substantial correlation between pathways "hsa05212" and "hsa04972." Similar claims can be made for other cases. The L^2 test also suggests that, at 0.1% level, for the tumor group, the hypotheses H_{02}^T and H_{03}^T are rejected, while for the normal group, the hypotheses H_{02}^N and H_{03}^N are not rejected.

References

Ahmad MR (2010) Tests for covariance matrices, particularly for high dimensional data. Technical Reports, Department of Statistics, University of Munich. http://epub.ub.uni-muenchen.de/11840/1/tr091.pdf. Accessed 3 Apr 2018

Bai ZD, Saranadasa H (1996) Effect of high dimension: by an example of a two sample problem. Stat Sin 6:311–329

Bai ZD, Jiang DD, Yao JF, Zheng SR (2009) Corrections to LRT on large-dimensional covariance matrix by RMT. Ann Stat 37:3822–3840

Bickel PJ, Levina E (2008a) Regularized estimation of large covariance matrices. Ann Stat 36:199–227

Bickel PJ, Levina E (2008b) Covariance regularization by thresholding. Ann Stat 36:2577–2604

Birke M, Dette H (2005) A note on testing the covariance matrix for large dimension. Stat Probab Lett 74:281–289

Brent RP, Osborn JH, Smith WD (2015) Probabilistic lower bounds on maxima determinants of binary matrices. Available at arxiv.org/pdf/1501.06235. Accessed 3 Apr 2018

Cai Y, Ma ZM (2013) Optimal hypothesis testing for high dimensional covariance matrices. Bernoulli 19:2359–2388

Cai T, Liu WD, Luo X (2011) A constrained l_1 minimization approach to sparse precision matrix estimation. J Am Stat Assoc 106:594–607

Chen SX, Qin Y-L (2010) A two-sample test for high-dimensional data with applications to gene-set testing. Ann Stat 38:808–835

Chen SX, Zhang L-X, Zhong P-S (2010) Tests for high-dimensional covariance matrices. J Am Stat Assoc 105:810–819

Chen XH, Shao QM, Wu WB, Xu LH (2016) Self-normalized Cramér type moderate deviations under dependence. Ann Stat 44:1593–1617

Chernozhukov V, Chetverikov D, Kato K (2014) Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. Ann Stat 41:2786–2819

Dickhaus T (2014) Simultaneous statistical inference: with applications in the life sciences. Springer, Heidelberg

Dudiot S, van der Laan M (2008) Multiple testing procedures with applications to genomics. Springer, New York

Efron B (2010) Large-scale inference: empirical Bayes methods for estimation, testing, and prediction. Cambridge University Press, Cambridge
Fan J, Hall P, Yao Q (2007) To how many simultaneous hypothesis tests can normal, Student's t or bootstrap calibration be applied. J Am Stat Assoc 102:1282–1288
Fan J, Liao Y, Mincheva M (2013) Large covariance estimation by thresholding principal orthogonal complements. J R Stat Soc Ser B Stat Methodol 75:603–680
Fisher TJ, Sun XQ, Gallagher CM (2010) A new test for sphericity of the covariance matrix for high dimensional data. J Multivar Anal 101:2554–2570
Georgiou S, Koukouvinos C, Seberry J (2003) Hadamard matrices, orthogonal designs and construction algorithms. In: Designs 2002: further computational and constructive design theory, vols 133–205. Kluwer, Boston
Han YF, Wu WB (2017) Test for high dimensional covariance matrices. Submitted to Ann Stat
Hedayat A, Wallis WD (1978) Hadamard matrices and their applications. Ann Stat 6:1184–1238
Jiang TF (2004) The asymptotic distributions of the largest entries of sample correlation matrices. Ann Appl Probab 14:865–880
Jiang DD, Jiang TF, Yang F (2012) Likelihood ratio tests for covariance matrices of high-dimensional normal distributions. J Stat Plann Inference 142:2241–2256
Ledoit O, Wolf M (2002) Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. Ann Stat 30:1081–1102
Liu WD, Shao QM (2013) A Cramér moderate deviation theorem for Hotelling's T^2-statistic with applications to global tests. Ann Stat 41:296–322
Lou ZP, Wu WB (2018) Construction of confidence regions in high dimension (Paper in preparation)
Marčenko VA, Pastur LA (1967) Distribution of eigenvalues for some sets of random matrices. Math U S S R Sbornik 1:457–483
Onatski A, Moreira MJ, Hallin M (2013) Asymptotic power of sphericity tests for high-dimensional data. Ann Stat 41:1204–1231
Portnoy S (1986) On the central limit theorem in $\mathbb{R}^p$ when $p \to \infty$. Probab Theory Related Fields 73:571–583
Qu YM, Chen SX (2012) Test for bandedness of high-dimensional covariance matrices and bandwidth estimation. Ann Stat 40:1285–1314
Schott JR (2005) Testing for complete independence in high dimensions. Biometrika 92:951–956
Schott JR (2007) A test for the equality of covariance matrices when the dimension is large relative to the sample size. Comput Stat Data Anal 51:6535–6542
Srivastava MS (2005) Some tests concerning the covariance matrix in high-dimensional data. J Jpn Stat Soc 35:251–272
Srivastava MS (2009) A test for the mean vector with fewer observations than the dimension under non-normality. J Multivar Anal 100:518–532
Veillette MS, Taqqu MS (2013) Properties and numerical evaluation of the Rosenblatt distribution. Bernoulli 19:982–1005
Wu WB (2005) Nonlinear system theory: another look at dependence. Proc Natl Acad Sci USA 102:14150–14154 (electronic)
Wu WB (2011) Asymptotic theory for stationary processes. Stat Interface 4:207–226
Wu WB, Shao XF (2004) Limit theorems for iterated random functions. J Appl Probab 41:425–436
Xiao H, Wu WB (2013) Asymptotic theory for maximum deviations of sample covariance matrix estimates. Stoch Process Appl 123:2899–2920
Xu M, Zhang DN, Wu WB (2014) L^2 asymptotics for high-dimensional data. Available at arxiv.org/pdf/1405.7244v3. Accessed 3 Apr 2018
Yarlagadda RK, Hershey JE (1997) Hadamard matrix analysis and synthesis. Kluwer, Boston
Zhang RM, Peng L, Wang RD (2013) Tests for covariance matrix with fixed or divergent dimension. Ann Stat 41:2075–2096