A Behrens-Fisher problem for general factor models in high dimensions

Masashi Hyodo[a], Takahiro Nishiyama[b], Tatjana Pavlenko[c]

[a]*Faculty of Economics, Kanagawa University, 3-27-1 Rokkakubashi, Kanagawa-ku, Yokohama-shi, Kanagawa, Japan. caicmhy@gmail.com*
[b]*Department of Business Administration, Senshu University, 2-1-1, Higashimita, Tama-ku, Kawasaki-shi, Kanagawa 214-8580, Japan. nishiyama@isc.senshu-u.ac.jp*
[c]*Department of Mathematics, KTH Royal Institute of Technology, SE-100, 44, Stockholm, Sweden*

Abstract

We revisit, in an original and challenging high-dimensional framework, the well-known Behrens-Fisher problem and propose a testing procedure which accommodates a low-dimensional latent factor model. The developed inferential framework is general, as it applies to the problems where the underlying populations may be non-normal, the dimension of the population mean vectors may highly exceed the sample size, design may be unbalanced and the loading factor dimensions may be different. Under high-dimensional asymptotic regime combined with fairly weak technical conditions, we show that null limiting distributions of the test statistics follow a weighted mixture of chi-square distributions, which depends only on the spectrum of the noise covariance matrix and the number of latent factors. As these latter are usually unknown in practice, we exploit an estimation procedure which builds on recent advances in random matrix theory. The asymptotic power of the proposed test is established. The numerical study confirms good analytical properties of the new test that compares favorably to the existing procedures used in a similar context. Real data applications are demonstrated with an empirical study on a leukemia data set.

Key words: High-dimensional data, High-dimensional testing problem, Latent factor model, Two-sample test.
2000 *Mathematics Subject Classification.* 62H15, 62F03, 62F05.

1. Introduction

Let $\mathbf{x}_{gi} = (x_{gi1}, \ldots, x_{gip})^\top \sim \mathcal{F}_g$ be iid p-dimensional random vectors collected from the ith subject in the gth population with mean vector $\boldsymbol{\mu}_g$ and covariance matrix $\boldsymbol{\Sigma}_g$, where $\mathcal{F}_g$ denotes the distribution function for gth population, $i \in [\![n_g]\!]$, $g \in [\![2]\!]$ and $[\![k]\!]$ denotes the set $\{1, \ldots, k\}$ for $k \in \mathbb{N}$. Specifically, we design the test procedure for testing

$$\mathcal{H} : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2, \quad \mathcal{A} : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2 \tag{1.1}$$

in the setting where p may be large and even $p \gg n_g$, $\mathcal{F}_g$ may be non-normal and $\boldsymbol{\Sigma}_g$ may be unequal which, along with n_g also allowed to be unequal.

As the classical methods of mean comparisons for low-dimensional data, of which the best-known is Hotelling's T^2 test, do not work when $p > n_g$ and need to be modified, a number of useful two-sample tests have been proposed for high-dimensional settings. The construction of many such tests has been motivated by the work of Bai and Saranadasa (1996), who proposed to substitute an identity matrix $\mathbf{I}$ for the pooled sample covariance matrix in Hotelling's T^2 statistic under the assumption of a common covariance matrix $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$. Chen and Qin (2010) extended the L_2-norm-based construction and established asymptotic properties of the proposed test under much weaker conditions, in particular by relaxing the homoscedasticity assumption of Bai and Saranadasa (1996). Note that their tests are guaranteed under the *weak dependence*–type condition (sphericity condition) which stated as

$$\mathrm{tr}(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_\ell\boldsymbol{\Sigma}_h) = o[\mathrm{tr}^2\{(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^2\}] \text{ for } i, j, \ell, h = 1 \text{ or } 2, \text{ as } p \to \infty, \tag{1.2}$$

see e.g., assumption (3.6) in Chen and Qin (2010). Here, $\mathrm{tr}(\cdot)$ is the trace operator of a matrix. The sphericity assumption is crucial for establishing the asymptotic normality of the test statistic proposed by Chen and Qin (2010); see Theorem 1 of their paper. However, besides of being difficult to verify in practice, sphericity assumption can be easily violated in covariance models where the eigenvalues of $\boldsymbol{\Sigma}_g$ are dominated by few top ones.

Therefore, it is necessary to establish a test method in a situation where the sphericality condition is not satisfied. Recently, some of proposals has been put forth in the literature with this motivations in mind. There are two different approaches to this effort. One is the test under the covariance structure called the strongly spiked structure, and the other is test under the low-dimensional factor model. Under these two settings, the sphericity conditions (1.2) is not satisfied, and the essential difference between the two settings is the structure of the noise term. The test for (1.1) under strongly spiked structure is proposed by Ishii (2017), Aoshima and Yata (2018), and Ishii et al. (2019). On the other hands, Ma et al. (2015) proposed the test for (1.1) under the low-dimensional factor model with homoscedasticity assumption, i.e., an assumption of common covariance matrix $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$. However, this assumption is a very strong assumption which is hard to be practically verified in $p \gg n_g$ settings.

As mentioned above, while the tests under the strongly spiked structure are well developed, the test under the factor model are still evolving. Therefore, in this research, we are interested to develop a high-dimensional test for difference of mean vectors, by relaxing the standard latent factor model assumptions, e.g., large-sample setting $p < n_g$, normality and homoscedasticity. In other words, we aim to improve the Ma et al.'s method under the low-dimensional factor model.

A factor model is a convenient structural assumption on the covariance matrix which is popular in a wide spectrum of modern applied fields like genetics, microbiome and metagenomic, fMRI, economics and finance, or more generally in high-dimensional data where the dependence of measurements can be attributed to a relatively small number of latent factors. A factor model assumes that for each $g \in [\![2]\!]$, the observable vector $\mathbf{x}_{gi}$ is decomposable into a latent factor and an idiosyncratic (noise) component as follows:

$$\mathbf{x}_{gi} = \boldsymbol{\mu}_g + \mathbf{B}_g \mathbf{z}_{gi} + \boldsymbol{\Psi}_g^{1/2} \boldsymbol{\epsilon}_{gi}, \tag{1.3}$$

where $\boldsymbol{\mu}_g \in \mathbb{R}^p$ is a deterministic intercept vector, $\mathbf{z}_{gi} = (\mathrm{z}_{gi1}, \dots, \mathrm{z}_{gid_g})^\top$ is the d_g-dimensional latent (unobservable) factor vector, and $\boldsymbol{\epsilon}_{gi} = (\epsilon_{gi1}, \dots, \epsilon_{gip})^\top$ is the p-dimensional error (noise) vector which is uncorrelated with the latent factor. In what follows, we assume that $d_g \in \mathbb{N}$ is a fixed number. Further, $\mathbf{B}_g = (\mathbf{b}_{g1}, \dots, \mathbf{b}_{gp})^\top$ denotes the loading matrix where for each $j \in [\![p]\!]$, $\mathbf{b}_{gj} = (b_{gj1}, \dots, b_{gjd_g})^\top \in \mathbb{R}^{d_g}$ is a non-random vector, and $\boldsymbol{\Psi}_g = \mathrm{diag}(\psi_{g1}, \dots, \psi_{gp})$ is the non-random $p \times p$ diagonal matrix whose elements are $\psi_{g1} > 0, \dots, \psi_{gp} > 0$. For the latent vector $\mathbf{z}_{gi}$ and error vector $\boldsymbol{\epsilon}_{gi}$, we further assume that $\mathrm{z}_{gi\ell}$ are iid with $\mathrm{E}(\mathrm{z}_{gi\ell}) = 0$, $\mathrm{E}(\mathrm{z}_{gi\ell}^2) = 1$ and $\mathrm{E}(\mathrm{z}_{gi\ell}^4) = \kappa_{\mathrm{z}_g} < \infty$, and ϵ_{gij} are iid with $\mathrm{E}(\epsilon_{gij}) = 0$, $\mathrm{E}(\epsilon_{gij}^2) = 1$ and $\mathrm{E}(\epsilon_{gij}^4) = \kappa_{\epsilon_g} < \infty$ for $g \in [\![2]\!]$, $i \in [\![n_g]\!]$, $j \in [\![p]\!]$ and $\ell \in [\![d_g]\!]$. Structural assumptions of the model (1.3) imply that

$$\mathrm{E}(\mathbf{x}_{gi}) = \boldsymbol{\mu}_g, \ \mathrm{cov}(\mathbf{x}_{gi}) = \mathbf{B}_g \mathbf{B}_g^\top + \boldsymbol{\Psi}_g := \boldsymbol{\Sigma}_g, \tag{1.4}$$

where $\mathbf{\Sigma}_g \in \mathbb{R}^{p\times p}_{>0}, g \in [\![2]\!]$ and $\mathbb{R}^{p\times p}_{>0}$ denotes the space of real, symmetric, positive definite, $p \times p$ matrices.

We do not assume that $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$; out test statistics, along with their limit properties are studied under heteroscedasticity, i.e., solves a general, two-sample Behrens-Fisher problem for the latent factor model (1.3). In addition, with our approach, the test can still work well even when the two underlying covariance matrices are actually equal. Our testing procedure can accommodate the class of highly spiked high-dimensional covariance models of $\mathbf{\Sigma}_g$'s where the few leading eigenvalues may be extremely large. Furthermore, this relaxation is of crucial importance for the asymptotic theory of the proposed tests. Due to the restrictive framework of the normal asymptotic theory, we shift focus to more flexible approximation types, specifically, to a chi-square mixture-type of asymptotic approximation, which leads to the totally different testing procedures compared to those of Bai and Saranadasa (1996) or Chen and Qin (2010). Our asymptotic results are valid with no specific distributional or moment assumptions on the data. Numerical studies demonstrate that the chi-square mixture asymptotic approximation allows for better size control under a variety of practical scenarios of the factor model in high dimensions.

The rest of this paper is organized as follows. Section 2 lays out a high-dimensional asymptotic framework, presents the new test statistics along with their limiting properties, and provides the data-driven test procedures. Section 3 provides evaluation of a finite sample performance of the proposed tests where the simulation study is followed by the real-data applications. Discussion and concluding remarks are provided in Section 4. All proofs and auxiliary technical results are delegated to Appendix.

1.1. Notational convention

Let $\boldsymbol{v} \in \mathbb{R}^p$ and $\mathbf{M} \in \mathbb{R}^{p\times p}_{>0}$. Throughout the paper, we let $\|\boldsymbol{v}\|$ and $\mathrm{tr}(\mathbf{M})$ denote the L_2-norm of a vector $\boldsymbol{v}$ and the trace operator of a matrix $\mathbf{M}$, respectively. $\lambda_1(\mathbf{M}) \geqslant \lambda_2(\mathbf{M}) \geqslant \cdots \geqslant \lambda_p(\mathbf{M})$ denote the eigenvalues of a matrix $\mathbf{M}$. The symbols $\rightsquigarrow$ and $\xrightarrow{\mathbb{P}}$ denote convergence in distribution and convergence in probability, respectively. The symbol $\chi^2(\nu)$ denotes a central chi-square distribution with ν degrees of freedom. We recall that $[\![m]\!]$ denotes the set $\{1, \ldots, m\}$ for $m \in \mathbb{N}$, and denote by $[\![\ell, m]\!]$ the set $\{\ell, \ell+1, \ldots, m\}$ for $1 < \ell < m \in \mathbb{N}$. For two sequences of real numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n = o(b_n)$ if $\limsup_{n\to\infty} |a_n|/|b_n| = 0$, $a_n = O(b_n)$ if $\limsup_{n\to\infty} |a_n|/|b_n| < \infty$ and $a_n \asymp b_n$ if $a_n = O(b_n)$ and $b_n = O(a_n)$, these asymptotic notations extend naturally to probabilistic setups, denoted by $o_{\mathbb{P}}$, $O_{\mathbb{P}}$ and $\overset{\mathbb{P}}{\asymp}$ where limits are in the sense of convergence in probability.

2. Methodology and main results

This section considers the testing problem (1.1) in the high-dimensional asymptotic setting specified in Section 2.1. We shall first present our testing procedure in the oracle setting in Section 2.2 where the number of factors d_g and the noise covariance matrices $\mathbf{\Psi}_g$ in the model (1.3) are assumed to be known. Clearly, in practice, the assumption of known model parameters imposes limitations of the use of the results. A data-driven testing procedure is given in Section 2.3 for a general case of unknown d_g and $\mathbf{\Psi}_g$, as a natural next logical step in this line of research.

2.1. Asymptotic set up and some preliminary studies

We shall now formalize the asymptotic viewpoint which we adopt for the latent factor model of our interest. Our analysis takes place in an asymptotic setting where both the number of features p as well as the sample sizes n_g go to infinity. In the model (1.3), we choose the loading matrix $\mathbf{B}_g$ and the noise covariance matrix $\mathbf{\Psi}_g$ according to n_g and p with the following assumptions.

(A1) $p = p(n) \to \infty$ as a function of $n = n_1 + n_2$ such that p tents to infinity along with $n \to \infty$ and $n/p \to \eta \in (0, \infty)$. $n_1, n_2 \to \infty$ in such a way that $n_1/n_2 \to \gamma \in (0, \infty)$.

(A2) Let $\psi_{g\max} = \max\{\psi_{g1}, \ldots, \psi_{gp}\}$ for $g \in [\![2]\!]$. Then, $\psi_{g\max}/p \to 0$ and $(1/p)\mathbf{B}_g^\top \mathbf{B}_h \to \mathbf{A}_{gh}$ as $p \to \infty$, where $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are positive definite matrices.

Assumption (A1) specifies the general regime of p growth relative to $n = n_1 + n_2$ in the high-dimensional asymptotic setting to provide valid inference when both indices go to infinity. This general regime, in what follows will be denoted by $(n, p) \to \infty$. The second part of (A1) is a natural regularity condition for the two-sample problem which ensures that the sample sizes n_1 and n_2 grow proportionally. It is needed to keep the limit non-degenerate when $n_g \to \infty$. Technical assumption (A2) is set to take care of the limiting behavior of the loading matrices $\mathbf{B}_g$, $g \in [\![2]\!]$.

To appreciate these assumptions we note that, unlike common practice in the literature on high-dimensional testing, our approach is less restrictive as it does not impose any structural conditions on $\mathbf{\Sigma}_g$. The boundary conditions on the maximum eigenvalue of $\mathbf{\Sigma}_g$, e.g., $\lambda_{\max}(\mathbf{\Sigma}) = o(p^{1/2})$ required in Bai and Saranadasa (1996), or sphericity condition (1.2) required in Chen and Qin (2010) for the normal convergence of the proposed test statistics under non-normality and heteroscedasticity (see eq. (3.6)-(3.8) in Chen and Qin (2010)), may collapse for many covariance structures useful for high-dimensional mean testing. One specific example of such structure is the latent factor model at hand. It is for such cases that our test is designed.

Let

$$\mathbf{B} = \begin{pmatrix} \sqrt{\frac{n}{n_1}}\mathbf{B}_1 & -\sqrt{\frac{n}{n_2}}\mathbf{B}_2 \end{pmatrix}.$$

Under (A1) and (A2), $(1/p)\mathbf{B}^\top\mathbf{B} \to \mathbf{A}$, where

$$\mathbf{A} = \begin{pmatrix} (1+\gamma^{-1})\mathbf{A}_{11} & -\sqrt{(1+\gamma)(1+\gamma^{-1})}\mathbf{A}_{12} \\ -\sqrt{(1+\gamma)(1+\gamma^{-1})}\mathbf{A}_{21} & (1+\gamma)\mathbf{A}_{22} \end{pmatrix}$$

is a semi positive definite matrix. The following theorem is the result of the spectral properties of the covariance matrices $\mathrm{cov}(\sqrt{n/n_1}\mathbf{x}_{11} - \sqrt{n/n_2}\mathbf{x}_{21})$, $\mathbf{\Sigma}_1$, and $\mathbf{\Sigma}_2$ under the factor model (1.3).

Theorem 1. *Let $\mathbf{\Omega} = \mathrm{cov}(\sqrt{n/n_1}\mathbf{x}_{11} - \sqrt{n/n_2}\mathbf{x}_{21})$. For the factor model (1.3) satisfying condition (A2), the eigenvalues of $\mathbf{\Omega}$ and $\mathbf{\Sigma}_g$ obey the following properties*

$$\text{(i)} \quad \frac{\lambda_\ell(\mathbf{\Omega})}{p} = \begin{cases} \lambda_\ell(\mathbf{A}) + o(1), & \ell \in [\![d_1 + d_2]\!] \\ o(1), & \ell \in [\![d_1 + d_2 + 1, p]\!] \end{cases}$$

$$\text{(ii)} \quad \frac{\lambda_\ell(\mathbf{\Sigma}_g)}{p} = \begin{cases} \lambda_\ell(\mathbf{A}_{gg}) + o(1), & \ell \in [\![d_g]\!] \\ o(1), & \ell \in [\![d_g + 1, p]\!]. \end{cases}$$

Proof. See, Appendix B. □

As the discussion given just before Theorem 1 states, we note that by the result (ii) the sphericity condition (1.2) does not hold for the model (1.3) as $p \to \infty$. Indeed, to illustrate this in a simplified case, consider the homoscedastic model of Ma et al. (2015) and set $\mathbf{B}_1 = \mathbf{B}_2$ and $\mathbf{\Psi}_1 = \mathbf{\Psi}_2$, that is, $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}$. Then it follows from (ii) that

$$\{\mathrm{tr}(\mathbf{\Sigma}^2)\}^2 = p^4 \left[\sum_{\ell=1}^{d_0} \{\lambda_\ell(\mathbf{A}_0)\}^2\right]^2 + o(p^4),\ \mathrm{tr}(\mathbf{\Sigma}^4) = p^4 \sum_{\ell=1}^{d_0} \{\lambda_\ell(\mathbf{A}_0)\}^4 + o(p^4),$$

where $d_0 = d_1 = d_2$ and $(1/p)\mathbf{B}_1^\top \mathbf{B}_1 = (1/p)\mathbf{B}_2^\top \mathbf{B}_2 \to \mathbf{A}_0$ as $p \to \infty$. Hence, we note that $\operatorname{tr}(\mathbf{\Sigma}^4)/\{\operatorname{tr}(\mathbf{\Sigma}^2)\}^2 \nrightarrow 0$ as $p \to \infty$, implying that the spericity condition does not hold.

2.2. Asymptotic distribution theory for the oracle procedure

For the latent factor model data set up specified by (1.3) and (1.4), let $\overline{\mathbf{x}}_g$ and $\mathbf{S}_g$ be the gth group sample mean vector and sample covariance matrix, respectively, defined as

$$\overline{\mathbf{x}}_g = \frac{1}{n_g}\sum_{i=1}^{n_g} \mathbf{x}_{gi}, \quad \mathbf{S}_g = \frac{1}{n_g - 1}\sum_{i=1}^{n_g} (\mathbf{x}_{gi} - \overline{\mathbf{x}}_g)(\mathbf{x}_{gi} - \overline{\mathbf{x}}_g)^\top, \quad g \in [\![2]\!].$$

To obtain some initial theoretical results in testing (1.1), we temporarily assume that the noise covariance matrices $\mathbf{\Psi}_g$ as well as the number of latent factors d_g in (1.3) are known (to an oracle) and introduce the following oracle statistic

$$T = \frac{n}{p}\left\{\|\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2\|^2 - \frac{1}{n_1}\operatorname{tr}(\mathbf{\Psi}_1) - \frac{1}{n_2}\operatorname{tr}(\mathbf{\Psi}_2)\right\}. \tag{2.1}$$

The following theorem states the null asymptotic distribution of the oracle test statistic T.

Theorem 2. *Suppose that the null hypothesis $\mathcal{H}$ from (1.1) is true. For a latent factor model (1.3) satisfying conditions (A1)-(A2), the statistic T is asymptotically distributed as $\sum_{\ell=1}^{d_1+d_2} \lambda_\ell(\mathbf{A})\chi_\ell^2(1)$, where $\chi_1^2(1), \ldots, \chi_{d_1+d_2}^2(1)$ are mutually independent, chi-square distributed random variables with 1 degree of freedom and $\lambda_\ell(\mathbf{A})$ is the ℓth largest eigenvalue of matrix $\mathbf{A}$.*

Proof. See, Appendix C. □

Remark 1. *The convergence result stated in Theorem 2 holds in a more general asymptotic regime when $(p, n) \to \infty$ without imposing any explicit restrictions between n and p.*

Some comments are in order to explain the structure of the statistic T defined in (2.1). Under the null hypothesis $\mathcal{H}$ from (1.1) and a latent factor model (1.3) satisfying conditions (A1)-(A2), a careful study of the proof of Theorem 2 reveals that $\mathrm{E}\{(n/p)\|\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2\|^2\} = \operatorname{tr}(\mathbf{A}) + (n/pn_1)\operatorname{tr}(\mathbf{\Psi}_1) + (n/pn_2)\operatorname{tr}(\mathbf{\Psi}_2) + o(1)$. It is well known that the expectation of $\sum_{\ell=1}^{d_1+d_2} \lambda_\ell(\mathbf{A})\chi_\ell^2(1)$

is $\mathrm{tr}(\mathbf{A})$, (see, e.g., eq. (11) in Zhang et al. (2021)). Hence, we need to remove $(n/pn_1)\mathrm{tr}(\mathbf{\Psi}_1)+(n/pn_2)\mathrm{tr}(\mathbf{\Psi}_2)$ from $(n/p)\|\overline{\mathbf{x}}_1-\overline{\mathbf{x}}_2\|^2$, which results in the test statistic proposed in (2.1). It is important to note that the suggested bias-adjustment yields almost surely nonnegative statistic T; indeed, by noting that $T=\mathbf{y}^\top\mathbf{A}\mathbf{y}+o_{\mathbb{P}}(1)$ we get $\mathbb{P}\{T\geqslant 0\}\to 1$, where $\mathbf{y}=(\sqrt{n_1}\overline{\mathbf{z}}_1^\top,\sqrt{n_2}\overline{\mathbf{z}}_2^\top)^\top$, $\overline{\mathbf{z}}_g=n_g^{-1}\sum_{i=1}^{n_g}\mathbf{z}_{gi}$ for $g\in[\![2]\!]$.

On the basis of the limiting null distribution, the asymptotically α-level oracle test can be defined as follows:

$$\boldsymbol{\phi}_\alpha(d,\mathbf{\Psi})=\mathbb{1}(T\geqslant t_\alpha),\tag{2.2}$$

where $d=\{d_1,d_2\}$, $\mathbb{1}(\cdot)$ is the indicator function, $\mathbf{\Psi}=\{\mathbf{\Psi}_1,\mathbf{\Psi}_2\}$, t_α is the $(1-\alpha)$-quantile of the cumulative distribution function of the random variable $\sum_{\ell=1}^{d_1+d_2}\lambda_\ell(\mathbf{A})\chi^2_\ell(1)$. The null hypothesis is rejected if and only if $\boldsymbol{\phi}_\alpha(\cdot)=1$.

The message is that the chi-square mixture approximation approach of Theorem 2 is of obvious appeal for the development of L_2-norm-based mean testing procedures for (1.1) under factor models. Ma et al. (2015) exploited this approximation and derived the asymptotic null distribution of their proposed statistic under homoscedastisity assumption. As we will see in Section 2.3, to apply Theorem 2 in studying the statistics considered in the present paper, we simply need to properly determine estimators of d_g and $\mathrm{tr}(\mathbf{\Psi}_g)$.

2.3. Data-driven test procedure

Estimation of the number of factors, d_g, is another important issue of the practical use of high-dimensional factor models. There is a rich source of literature where the number of factors is determined by leveraging random matrix theory and the present study is aligned with this perspective. Specifically, we exploit the covariance matrix eigenvalue's ratio-based estimation technique presented in Onatski (2010), and later in Wang (2012) and Ahn and Horenstein (2013), and propose a simple, tuning parameter-free estimator of d_g which is consistent for high-dimensional factor models. This in turn makes it possible to construct a fully data-driven test procedure for testing (1.1).

We have so far focused on the oracle case in which d_g and $\mathrm{tr}(\mathbf{\Psi}_g)$ are known. Obviously, in practice, the consistency is expected to hold with unknown parameters replaced by their estimators. In what follows, we show with a bit of technical work, that the statement of Theorem 2 remains valid for the "plug-in" version of T, where the unknown d_g and $\mathrm{tr}(\mathbf{\Psi}_g)$ are replaced with any consistent estimators $\widehat{d}_g$ and $\widehat{\mathrm{tr}(\boldsymbol{\Psi}_g)}$, respectively, specified under the general asymptotic regime of (A1). This latter fact will be used to develop the asymptotic theory of the proposed test. In fact, since only consistency is required for estimators $\widehat{d}_g$ and $\widehat{\mathrm{tr}(\boldsymbol{\Psi}_g)}$, our asymptotic theory developed in this section is valid for a general class of tests for mean difference in factor models with consistently estimated parameters d_g and $\mathrm{tr}(\boldsymbol{\Psi}_g)$.

To estimate the number of factors d_g, we explore the idea of Wang (2012) and Ahn and Horenstein (2013). We focus on the criteria function which is proposed by Ahn and Horenstein (2013)

$$ER_g(k) = \frac{\lambda_k(\mathbf{S}_g)}{\lambda_{k+1}(\mathbf{S}_g)},$$

where ER refers to eigenvalue ratio. The estimator of d_g is k which satisfies

$$\widehat{d}_g = \underset{1 \leqslant k \leqslant k_{g,\max}}{\arg\max}\ ER_g(k), \tag{2.3}$$

where $k_{g,\max}$ is the prespecified upper bound of k. By using Ahn and Horenstein (2013), we obtain the following lemma which provides consistency of the estimator $\widehat{d}_g$ in high-dimensional setting.

Lemma 1. *Suppose that the factor model (1.3) satisfies conditions (A1)-(A2). Then, there exists* $c_g \in (0, 1]$ *such that* $\mathbb{P}(\widehat{d}_g = d_g) \to 1$ *as* $\min\{p, n_g\} \to \infty$*, for any* $k_{g,\max} \in (d_g, \lfloor c_g \min\{p, n_g\}\rfloor - d_g - 1]$*, where* $g \in [\![2]\!]$ *and* $\lfloor\cdot\rfloor$ *denotes the floor function.*

Proof. See, Appendix D. □

The following lemma also derived asymptotic properties for the first d_g eigenvalues of $\mathbf{S}_g$ by using Ahn and Horenstein (2013).

Lemma 2. *Suppose that the factor model (1.3) satisfies conditions (A1) and (A2). Then, for* $\ell \in [\![d_g]\!]$, $\lambda_\ell(\mathbf{S}_g)/p = \lambda_\ell(\mathbf{A}_{gg})/p + o_{\mathbb{P}}(1)$ *as* $\min\{p, n_g\} \to \infty$ *with* $g \in [\![2]\!]$.

Proof. See, Appendix E. □

Based on Lemma 1 and 2, we define the consistent plug-in estimators of $\mathrm{tr}(\boldsymbol{\Psi}_g)$ as

$$\widehat{\mathrm{tr}(\boldsymbol{\Psi}_g)} = \mathrm{tr}(\mathbf{S}_g) - \sum_{\ell=1}^{\hat{d}_g} \lambda_\ell(\mathbf{S}_g). \tag{2.4}$$

Being equipped with consistent estimator (2.4), we define the data-driven test statistic as

$$T_{FA} = \frac{n}{p}\left\{\|\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2\|^2 - \frac{1}{n_1}\widehat{\mathrm{tr}(\boldsymbol{\Psi}_1)} - \frac{1}{n_2}\widehat{\mathrm{tr}(\boldsymbol{\Psi}_2)}\right\}. \tag{2.5}$$

Below we will show that for all large enough p and n the limiting null distribution of T_{FA} remains the same as that of the oracle statistic T.

Theorem 3. *Suppose that the null hypothesis $\mathcal{H}$ of (1.1), is true. For a factor model (1.3) satisfying conditions (A1)-(A2), the statistic T_{FA} is asymptotically distributed as $\sum_{\ell=1}^{d_1+d_2}\lambda_\ell(\mathbf{A})\chi^2_\ell(1)$, where $\lambda_\ell(\mathbf{A})$ and $\chi^2_\ell(1)$ are specified in Theorem 2.*

Proof. See, Appendix F. □

Theorem 4. *For a factor model (1.3) satisfying conditions (A1)-(A2), we have*

$$\sum_{\ell=1}^{\hat{d}_1+\hat{d}_2}\widehat{\lambda_\ell(\mathbf{A})}\chi^2_\ell(1) = \sum_{\ell=1}^{d_1+d_2}\lambda_\ell(\mathbf{A})\chi^2_\ell(1) + o_{\mathbb{P}}(1).$$

Proof. See, Appendix G. □

Now, on the basis the results Lemmas 1-2 and Theorems 3 and 4 we furnish a fully data-driven test of significance of $\mathcal{H}$. Given a test statistic T_{FA}, we write $\boldsymbol{\phi}_\alpha(\widehat{d}, \widehat{\boldsymbol{\Psi}})$ for the test that uses estimators of d_g and $\boldsymbol{\Psi}$ specified respectively by (2.3) and (2.4), and has asymptotic size α. The following are six steps of the test procedure.

1. For each $g \in [\![2]\!]$, draw n_g observations from $\mathcal{F}_g$ and calculate $\overline{\mathbf{x}}_g$, $\widehat{d}_g$, $\widehat{\mathrm{tr}(\boldsymbol{\Psi}_g)}$ and T_{FA}.
2. Let $n_0 = \min\{n_1, n_2\}$. For each $i \in [\![n_0]\!]$ obtain $\mathbf{x}_{0i} = \sqrt{n/n_1}(\mathbf{x}_{1i} - \overline{\mathbf{x}}_1) - \sqrt{n/n_2}(\mathbf{x}_{2i} - \overline{\mathbf{x}}_2)$ and calculate the empirical covariance matrix as
$$\mathbf{S}_0 = \sum_{i=1}^{n_0}\mathbf{x}_{0i}\mathbf{x}_{0i}^\top/(n_0 - 1).$$
3. Using $\mathbf{S}_0$, estimate $\lambda_\ell(\mathbf{A})$ by $\widehat{\lambda_\ell(\mathbf{A})} = \lambda_\ell(\mathbf{S}_0)/p$ for $\ell \in [\![\widehat{d}_1 + \widehat{d}_2]\!]$.
4. Draw a sample of $\widehat{d}_1 + \widehat{d}_2$ independent, $\chi^2(1)$-distributed random variables and obtain $\sum_{\ell=1}^{\hat{d}_1+\hat{d}_2}\widehat{\lambda_\ell(\mathbf{A})}\chi^2_\ell(1)$.

5. Repeat step 4 sufficiently many times (10^4), obtain a Monte-Carlo estimate of the distribution of the random variable $\sum_{\ell=1}^{\hat{d}_1+\hat{d}_2} \widehat{\lambda_\ell(\mathbf{A})}\chi^2_\ell(1)$ and find its $(1-\alpha)$-quantile, $\hat{t}_\alpha$.
6. Define the corresponding test as

$$\boldsymbol{\phi}_\alpha(\hat{d}, \widehat{\boldsymbol{\Psi}}) = \mathbb{1}(T_{FA} \geqslant \hat{t}_\alpha)$$

and reject the null hypothesis if and only if $\boldsymbol{\phi}_\alpha(\hat{d}, \widehat{\boldsymbol{\Psi}}) = 1$.

We note that, by consistency of $\widehat{d}_g$ and $\widehat{\lambda_\ell(\mathbf{A})}$, $\ell \in [\![\widehat{d}_1 + \widehat{d}_2]\!]$ stated in Lemmas 1-2, the Monte-Carlo estimate of the distribution of $\sum_{\ell=1}^{\hat{d}_1+\hat{d}_2} \widehat{\lambda_\ell(\mathbf{A})}\chi^2_\ell(1)$ specified by steps 4-5 is obtained by resampling of the $\chi^2(1)$ components only.

2.4. Aspects of power

The asymptotic power of $\boldsymbol{\phi}_\alpha(\widehat{d}, \widehat{\boldsymbol{\Psi}})$ is analyzed under certain conditions on the separation between $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$. While our null distribution results in Section 2 are valid under the more general asymptotic regime of (A1), we treat here its particular case where $(p, n) \to \infty$ in such a way that $n/p \to \eta \in (0, \infty)$. Recall the definition in (2.5) and let $S_n = \{(\overline{\mathbf{x}}_1 - \boldsymbol{\mu}_1) - (\overline{\mathbf{x}}_2 - \boldsymbol{\mu}_2)\}^\top(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. In our exploration of power, it is at times convenient to write

$$\|\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2\|^2 = \|(\overline{\mathbf{x}}_1 - \boldsymbol{\mu}_1) - (\overline{\mathbf{x}}_2 - \boldsymbol{\mu}_2)\|^2 + 2S_n + \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2. \tag{2.6}$$

Theorem 5. *Suppose that conditions (A1) and (A2) hold. Under the alternative hypothesis $\mathcal{A}$, for any $0 < \alpha < 1$*

$$\mathbb{P}\left(T_{FA} \geqslant t_\alpha\right) = \begin{cases} 1 - F\left(t_\alpha - \eta\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2\right) + o(1) & \text{if } \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\| = O(1) \\ 1 & \text{if } \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\| \to \infty, \end{cases} \tag{2.7}$$

where t_α is the $(1-\alpha)$-quantile of the distribution of $\sum_{\ell=1}^{d_1+d_2} \lambda_\ell(\mathbf{A})\chi^2_\ell(1)$, $F(\cdot)$ is cumulative distribution function of the random variable

$$\mathbf{y}_0^\top \mathbf{A}\mathbf{y}_0 + \frac{2\sqrt{n}}{p}\mathbf{y}_0^\top \mathbf{B}^\top(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \tag{2.8}$$

for $\mathbf{y}_0 \sim \mathcal{N}_{d_1+d_2}(\mathbf{0}, \mathbf{I}_{d_1+d_2})$.

Proof. See, Appendix H. □

In words, Theorem 5 says that for the test of asymptotic size α based on T_{FA}, under the regime $n/p \to \eta$, the maximum achievable power against a specific alternative with $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\| = O(1)$ is $1 - F\left(t_\alpha - \eta\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2\right) + o(1)$. The test has full power when $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\| \to \infty$. Both these scenarios for the magnitude of deviation from $\mathcal{H}$ will be numerically demonstrated in our simulations in the next section.

3. Numerical studies

3.1. Simulation experiments

In this section we compare, through simulations, the performance of the proposed test $\boldsymbol{\phi}_\alpha(\hat{d}, \hat{\boldsymbol{\Psi}})$ and existing procedures suitable for a two-sample, Behrens-Fisher problem in high-dimensional data with latent factor structure. Focusing on the tests constructed via L_2-norm, we examine numerical performance of procedures by Chen and Qin (2010) and Zang et al. (2021) in terms of their size control and power. These tests are denoted respectively by Ch-Q and L2D and our proposed test is denoted by FA, in the rest of this section. Before turning to the simulations, however, we discuss the implementation of the test procedures.

Without loss of generality, we shall always take $\boldsymbol{\mu}_1 = \mathbf{0}$ in the simulations. Under the null hypothesis, $\mathcal{H} : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \mathbf{0}$, whereas under the alternative hypothesis we consider two different scenarios

$$\mathcal{A}_1 : \boldsymbol{\mu}_1 = \mathbf{0},\ \boldsymbol{\mu}_2 = \frac{\sqrt{6}\delta_1}{\sqrt{(2p+1)(p+1)p}}(1, \cdots, p)^\top,$$
$$\mathcal{A}_2 : \boldsymbol{\mu}_1 = \mathbf{0},\ \boldsymbol{\mu}_2 = \sqrt{p/n}(\mathbf{1}^\top_{\lceil p^{\delta_2}\rceil},\ \mathbf{0}^\top_{p-\lceil p^{\delta_2}\rceil})^\top,$$

where $\delta_1 \in \{1.6, 3.2, 6.4\}$ and $\delta_2 \in \{0.4, 0.6, 0.8\}$ are tuning parameters which control the magnitude of departure from the null hypothesis. $\mathcal{A}_1$ represents the alternative when $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\| = O(1)$, whereas $\mathcal{A}_2$ illustrates the case when alternative depends on p in such a way that $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\| = O(p^{\delta_2})$.

To examine the test performance for spiked covariance structure of $\mathcal{F}_g$, we consider the following two models.

M1: Under heteroscedasticity, i.e., assuming that $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$, we set $\mathbf{B}_g$ and $\boldsymbol{\Psi}_g$ as

$$\mathbf{B}_1 = (\mathbf{b}_{11}, \cdots, \mathbf{b}_{1p}),\ \mathbf{B}_2 = (\mathbf{b}_{21}, \cdots, \mathbf{b}_{2p}),\ \boldsymbol{\Psi}_1 = \mathbf{I}_p,\ \boldsymbol{\Psi}_2 = 0.5\mathbf{I}_p,$$

with $\mathbf{b}_{gj}$ for $g \in [\![2]\!]$ and $j \in [\![p]\!]$ generated as d_g-dimensional, normally distributed vectors

$$\mathbf{b}_{11}, \cdots, \mathbf{b}_{1p} \overset{iid}{\sim} \mathcal{N}_{d_1}(\mathbf{0}, \mathbf{\Omega}_1), \ \mathbf{b}_{21}, \cdots, \mathbf{b}_{2p} \overset{iid}{\sim} \mathcal{N}_{d_2}(\mathbf{0}, \mathbf{\Omega}_2),$$

where $\mathbf{\Omega}_g$ are set to be a first-order autoregressive covariance structure parameterized respectively as $\mathbf{\Omega}_1 = (0.3^{|i-j|})$ and $\mathbf{\Omega}_2 = (0.1^{|i-j|})$. For the loading factor dimensions, we take $(d_1, d_2) \in \{(2,1), (3,2)\}$.

M2: Under homoscedasticity, i.e., assuming that $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$, we set

$$\mathbf{B}_1 = \mathbf{B}_2 = (\mathbf{b}_{11}, \cdots, \mathbf{b}_{1p}), \ \mathbf{\Psi}_1 = \mathbf{\Psi}_2 = \mathbf{I}_p,$$

where $\mathbf{b}_{gj}$ are generated as for the **M1** and the loading factor dimension is set to $(d_1, d_2) \in \{(2,2), (3,3)\}$.

Observe that the specific choice of the loading matrix $\mathbf{B}_g$ in **M1** and **M2** is motivated by the technical assumption (A2) and a spiky nature of the covariance structure representing the latent factor model at hand. Indeed, one can see that

$$\frac{1}{p}\mathbf{B}_1^\top \mathbf{B}_1 \xrightarrow{\mathbb{P}} \mathbf{\Omega}_1, \quad \frac{1}{p}\mathbf{B}_2^\top \mathbf{B}_2 \xrightarrow{\mathbb{P}} \mathbf{\Omega}_2$$

as $p \to \infty$, which implies that both $\mathbf{B}_1$ and $\mathbf{B}_2$ satisfy (A2).

To include effects such as asymmetry and heavy-tailed nature of the factor vector and error term of (1.3), we consider the following three distribution scenarios for the latter.

D1: z_{gij} and ϵ_{gij} are iid $\mathcal{N}(0,1)$.

D2: Let $\tilde{z}_{gij}, \ \tilde{\epsilon}_{gij} \overset{iid}{\sim} \chi^2_{10}$, then

$$z_{gij} = \frac{\tilde{z}_{gij} - 10}{\sqrt{20}}, \ \epsilon_{gij} = \frac{\tilde{\epsilon}_{gij} - 10}{\sqrt{20}}.$$

D3: Let $\tilde{z}_{gij}, \ \tilde{\epsilon}_{gij} \overset{iid}{\sim} t_{10}$, then

$$z_{gij} = \frac{\tilde{z}_{gij}}{\sqrt{5/4}}, \ \epsilon_{gij} = \frac{\tilde{\epsilon}_{gij}}{\sqrt{5/4}}.$$

To mimic the "large p, small n" situation in a finite-sample setting and to represent an unbalanced design, we examine the test performance for the following dimensionality and sample sizes

$$(p, n_1, n_2) \in \{(100, 40, 50), (300, 80, 100), (500, 160, 200)\}.$$

We set the nominal significance levels to $\alpha \in \{0.1, 0.05, 0.01\}$.

Monte-Carlo estimates of finite-sample test sizes and powers are obtained as follows. The test statistics under consideration are then computed and empirical size of each test is obtained from 10^4 simulation runs. To assess the power of the proposed test, we fix the nominal significance level $\alpha = 0.05$. The empirical power of each test statistic is calculated under $\mathcal{A}_1$ and $\mathcal{A}_2$ with varying range of tuning parameters δ_1 and δ_2, for each combination of **M1**-**M2**, **D1**-**D3** and (p, n_1, n_2), and averaged over 10^4 simulation runs.

Monte-Carlo estimates of the test size of $\boldsymbol{\phi}_\alpha(\hat{d}, \widehat{\boldsymbol{\Psi}})$ along with Ch-Q's and L2D procedures are summarized in Tables 1-6. Tables 1 and 2 compare the empirical sizes of the tests under Models 1 and 2, respectively, whereas Tables 3-6 report the estimated power of the same.

From Tables 1 and 2, we observe that our proposed test FA provides valid asymptotic test with an accurate size control for most of the simulation settings. The accuracy of FA for **D2** (chi-square distributed factor vectors and error terms) as a seriously non-normal case is particularly noticeable. Likewise, is the case for the dependence structures **M1**-**M2** involving compound symmetry, both being highly spiked covariance matrices with only two distinct eigenvalues. The test sizes for FA get very close to the nominal levels even for $(p, n_1, n_2) = (100, 40, 50)$, indicating that the asymptotic described by Theorem 3 kicks in even with relatively small sample sizes. Furthermore, it is seen that FA systematically outperforms both Ch-Q and L2D tests in terms of their size control for both heteroscedastic and homoscedastic settings, and across different distributions generating factor vectors and error terms. The stability of the accurate size control for FA with increasing p a lot greater than n, and in a seriously unbalanced setting, is also evident. Moreover, while being constructed for a heteroscedastic setting, our proposed test demonstrates stably accurate size control when being applied to the homoscedastic case of **M2**. As expected, both Ch-Q and L2D tests exhibited inflated size in almost all the simulation settings, and in contrast to FA, sizes do not converge to the nominal level as both p and n_g increase together. This could be due to blindly using normal approximation for Ch-Q and due to slow convergence of the L2D test statistic to its asymptotic distribution. However, we can admit that the Welch-Satterthwaite approximation approach explored by Zang et al. (2021) for the chi-square distribution (L2D) maintains the test sizes better compared to that of the normal approximation used by Chen and Qin (2010).

More accurate results on the size control indicate that the new test FA is adapted to the presence of low-dimensional latent factor structure in high-dimensional data. This supports the main analytical findings of the present study summarized in Theorems 2 and 3, according to which the use of chi-square mixture approximation of the limit distribution of an L_2-based test statistic is preferable for high-dimensional data in presence of a latent factor structure.

Next, we consider the power of the tests, as studied in Section 2.4. As expected, FA is well adapted for detecting the alternatives we generated. As the results in Tables 3-6 show, the power of FA increases steadily with increasing the magnitude of δ_1. Under $\mathcal{A}_2$, the power grows monotonically and quickly approaches 1 not only for increasing sample size but also for increasing dimension p.

These numerical results confirm the theoretical analysis that was given in the last sections, specifically high-dimensional asymptotics for the power of T_{FA} derived in Theorem 5. As a numerical illustration of Theorem 5, powers under $\mathcal{A}_2$, as reported in Tables 5-6, increase much faster than those under $\mathcal{A}_1$ reported in Tables 3-4, as the sample size and dimension are increased. When δ_2 has been increased from 0.4 to 0.8 under $\mathcal{A}_2$, many entries of the empirical powers of the tests are observer to be 1. This could be viewed as an empirical indication of the proposed test being consistent in high dimensions. We also note in comparison, that the Ch-Q's and L2D procedures show slightly better power under lower values of δ_1 and δ_2, but this is due to the fact that the size of both these tests far exceeds the nominal significance level α.

Also, for the heteroscedastic model **M1**, the test FA under $\mathcal{A}_2$ has full power for sample sizes as small as 40 or 50, even for χ^2-distribution of error term. These results depict strong robustness of the proposed test against violation of classical large-sample factor model assumption such as normality.

3.2. A real-data application

For illustration, we apply the proposed test procedure to the leukemia dataset which has expression levels for $3,571$ human genes, for 47 patients with acute limphoblastic leukemia (ALL-group, $n_1 = 47$) and 25 patients

Table 1: Empirical sizes based on 10,000 simulation runs for **M1**: $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$

			$(d_1, d_2) = (2, 1)$			$(d_1, d_2) = (3, 2)$		
$(z_{gij}, \epsilon_{gij})$	(p, n_1, n_2)	α	FA	Ch-Q	L2D	FA	Ch-Q	L2D
(i) $\mathcal{N}(0,1)$	$(100, 40, 50)$	1	1.06	3.96	2.17	0.87	3.62	1.75
		5	4.73	7.43	6.04	4.33	7.44	5.99
		10	9.43	10.39	10.10	9.26	11.01	10.53
	$(300, 80, 100)$	1	0.84	3.78	1.64	0.86	3.26	1.58
		5	5.04	7.59	5.92	4.37	6.72	5.41
		10	10.15	10.88	10.35	9.13	10.37	9.79
	$(500, 160, 200)$	1	0.98	4.06	1.73	0.78	2.93	1.35
		5	5.36	7.60	6.22	4.64	6.75	5.28
		10	10.19	10.88	10.42	9.68	10.45	9.84
(ii) χ^2_{10}	$(100, 40, 50)$	1	1.16	4.02	2.02	0.86	3.53	1.92
		5	4.92	7.63	6.21	4.25	7.28	5.76
		10	9.88	10.92	10.54	9.05	10.79	10.46
	$(300, 80, 100)$	1	0.97	3.61	1.74	0.92	3.75	1.65
		5	4.81	7.10	5.64	4.98	7.93	6.07
		10	10.00	10.72	10.13	10.29	11.45	10.85
	$(500, 160, 200)$	1	1.06	3.64	1.90	1.04	3.18	1.53
		5	4.86	7.39	5.87	4.70	6.82	5.29
		10	10.22	10.71	10.23	9.63	10.43	9.72
(iii) t_{10}	$(100, 40, 50)$	1	0.85	3.96	2.17	0.87	3.56	1.87
		5	4.73	7.43	6.04	4.39	7.15	5.76
		10	9.43	10.39	10.10	8.85	10.65	10.30
	$(300, 80, 100)$	1	1.06	3.61	1.74	0.83	3.32	1.56
		5	4.81	7.10	5.64	4.60	6.94	5.41
		10	10.00	10.72	10.13	9.28	10.60	10.08
	$(500, 160, 200)$	1	0.96	3.59	1.75	1.01	3.37	1.48
		5	4.79	7.18	5.76	5.08	7.20	5.63
		10	9.70	10.49	10.05	9.94	10.67	10.18

with acute myeloid leukemia (AML-group, $n_2 = 25$). We are interested in testing the null hypothesis $\mathcal{H}$ on whether these two groups have the same mean expression levels. It is thus an unbalanced, high-dimensional, two-sample Behrens-Fisher problem.

The data were obtained from affymetrix oligonucleotide microarrays and are publicly available at

`https://web.stanford.edu/~hastie/CASI_files/DATA/leukemia.html`.

The description of the datasets and preprocessing protocols are due to Dudoit et al. (2002). Following the protocol, we preprocessed the dataset which resulted in $p = 3,571$ variables representing genes with highest minimal intensity across the samples, with n_1 and n_2 given above.

Applying the six steps of the data-driven test procedure from Section 2.3 to the data, we obtain the estimators of the number of loading factors

Table 2: Empirical sizes based on 10,000 simulation runs for **M2**: $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$

			$d_1 = d_2 = 2$			$d_1 = d_2 = 3$		
$(z_{gij}, \epsilon_{gij})$	(p, n_1, n_2)	α	FA	Ch-Q	L2D	FA	Ch-Q	L2D
(i) $\mathcal{N}(0,1)$	$(100, 40, 50)$	1	1.00	5.27	1.69	0.90	4.86	1.42
		5	5.05	9.39	5.76	4.67	9.15	5.77
		10	10.01	12.66	10.42	9.65	12.65	10.20
	$(300, 80, 100)$	1	1.04	5.00	1.54	0.87	4.81	1.38
		5	4.86	8.54	5.41	4.85	9.08	5.43
		10	9.60	11.65	9.53	9.75	12.22	9.96
	$(500, 160, 200)$	1	1.05	4.73	1.59	1.02	4.60	1.44
		5	4.83	8.46	5.25	4.69	8.84	5.20
		10	9.62	11.59	9.35	9.92	12.24	9.69
(ii) χ^2_{10}	$(100, 40, 50)$	1	1.26	5.15	1.80	0.91	5.02	1.52
		5	4.92	8.96	5.81	4.81	9.13	5.79
		10	9.70	12.25	9.89	9.43	12.41	10.16
	$(300, 80, 100)$	1	0.97	4.79	1.52	0.90	4.63	1.29
		5	4.83	8.26	5.20	4.77	8.62	5.16
		10	9.53	11.36	9.18	9.32	12.09	9.51
	$(500, 160, 200)$	1	0.91	4.85	1.43	1.14	4.97	1.73
		5	5.00	8.56	5.44	5.16	8.96	5.43
		10	9.56	11.37	9.48	9.80	12.06	9.80
(iii) t_{10}	$(100, 40, 50)$	1	1.28	5.12	1.95	0.96	4.98	1.57
		5	4.96	9.07	5.79	4.51	8.69	5.58
		10	9.97	12.58	10.26	9.28	12.21	9.74
	$(300, 80, 100)$	1	1.05	4.74	1.72	0.97	4.86	1.50
		5	4.66	8.34	5.24	4.90	9.05	5.38
		10	9.48	11.44	9.14	10.11	12.65	10.04
	$(500, 160, 200)$	1	0.94	4.66	1.52	1.03	4.65	1.57
		5	4.80	8.46	5.12	4.76	8.5	5.04
		10	9.62	11.54	9.32	9.69	11.91	9.63

and eigenvalues of $\mathbf{A}$, respectively as $\widehat{d}_1 = 3$, $\widehat{d}_2 = 2$, and $\widehat{\lambda_1(\mathbf{A})} \approx 0.14$, $\widehat{\lambda_2(\mathbf{A})} \approx 0.11$, $\widehat{\lambda_3(\mathbf{A})} \approx 0.11$, $\widehat{\lambda_4(\mathbf{A})} \approx 0.09$, and $\widehat{\lambda_5(\mathbf{A})} \approx 0.07$. The observed value of the test statistic is $T_{FA} \approx 9.45$. We further obtain an approximate critical value of the test as an upper 5% quantile of the empirical distribution of the test statistic $\sum_{\ell=1}^{\widehat{d}_1+\widehat{d}_2} \widehat{\lambda_\ell(\mathbf{A})}\chi^2_\ell(1)$, which is derived by 10^6 simulation runs. This gives $\widehat{t}_{0.05} \approx 1.16$. The results indicate that the null hypothesis $\mathcal{H}$ is rejected, i.e., the ALL- and AML-group means, statistically, discernibly different from each other at the level 0.05 and at any reasonable nominal level.

4. Concluding remarks

A test statistic for testing the equality of mean vectors is presented for high-dimensional, non-normal, unbalanced data under two-sample Behrens-

Table 3: Empirical powers under $\mathcal{A}_1$ based on 10,000 simulation runs for **M1**: $\mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$

			$(d_1, d_2) = (2, 1)$			$(d_1, d_2) = (3, 2)$		
$(z_{gij}, \epsilon_{gij})$	(p, n_1, n_2)	δ_1	FA	Ch-Q	L2D	FA	Ch-Q	L2D
(i) $\mathcal{N}(0,1)$	$(100, 40, 50)$	1.6	7.43	12.12	9.78	6.66	10.41	8.36
		3.2	25.46	39.77	32.27	17.35	27.71	22.52
		6.4	100.00	100.00	100.00	99.93	100.00	99.99
	$(300, 80, 100)$	1.6	6.72	10.01	8.01	7.83	12.06	9.44
		3.2	17.54	26.06	20.83	33.81	47.05	38.15
		6.4	100.00	100.00	100.00	99.99	100.00	99.97
	$(500, 160, 200)$	1.6	6.99	11.05	8.68	6.51	10.19	7.92
		3.2	24.51	34.82	28.04	16.7	24.99	19.46
		6.4	100.00	100.00	100.00	100.00	100.00	100.00
(ii) χ^2_{10}	$(100, 40, 50)$	1.6	7.33	11.67	9.30	5.98	9.98	7.96
		3.2	25.13	38.97	31.69	17.52	27.54	22.74
		6.4	100.00	100.00	100.00	99.80	99.98	99.96
	$(300, 80, 100)$	1.6	7.06	10.20	8.31	7.78	11.53	8.93
		3.2	17.64	26.66	21.16	34.81	47.86	38.80
		6.4	100.00	100.00	100.00	99.97	99.99	99.98
	$(500, 160, 200)$	1.6	7.84	11.19	8.69	6.61	9.85	7.55
		3.2	24.37	35.18	27.60	17.30	24.50	19.40
		6.4	100.00	100.00	100.00	100.00	100.00	100.00
(iii) t_{10}	$(100, 40, 50)$	1.6	7.17	11.86	9.19	6.50	10.41	8.36
		3.2	25.33	39.41	32.02	17.45	27.71	22.56
		6.4	99.99	100.00	100.00	99.86	100.00	99.93
	$(300, 80, 100)$	1.6	6.34	9.87	7.62	7.88	11.81	9.04
		3.2	17.56	26.04	20.86	34.33	47.65	37.85
		6.4	100.00	100.00	100.00	100.00	100.00	100.00
	$(500, 160, 200)$	1.6	7.03	10.36	7.99	6.69	9.46	7.42
		3.2	24.51	34.82	28.04	16.69	24.02	18.85
		6.4	100.00	100.00	100.00	100.00	100.00	100.00

Fisher problem for the low-dimensional factor model. The test statistic is formulated as a bias-adjusted, squared L_2-norm of the sample mean difference, where the adjustment terms are adapted to the correlation structure of the noise component of the latent factor model. The corresponding asymptotic theory is then used to derive the null- and non-null limiting distributions of the proposed test when both sample size and dimension go to infinity. The asymptotic theory of the test is developed under few mild assumptions and accommodates a wide class of highly spiked, high-dimensional covariance models for $\mathcal{F}_g$, which usually represent a factor structure.

We propose the chi-square mixture type of asymptotic approximations of the test statistic along with a Monte Carlo simulation scheme for computation of the test's critical values. This is in contrast with the approach in e.g., Chen and Qin (2010) where the asymptotic theory of the proposed tests centers around the normal limits of the null distribution. Both our theo-

Table 4: Empirical powers under $\mathcal{A}_1$ based on 10,000 simulation runs for **M2**: $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$

			$d_1 = d_2 = 2$			$d_1 = d_2 = 3$		
$(z_{gij}, \epsilon_{gij})$	(p, n_1, n_2)	δ_1	FA	Ch-Q	L2D	FA	Ch-Q	L2D
(i) $\mathcal{N}(0,1)$	$(100, 40, 50)$	1.6	6.99	12.92	8.04	5.89	11.25	6.82
		3.2	18.03	33.93	21.82	10.46	19.14	11.98
		6.4	99.98	100.00	100.00	80.96	99.15	90.67
	$(300, 80, 100)$	1.6	5.62	9.69	6.13	5.62	10.72	6.59
		3.2	14.08	21.11	15.14	15.11	22.32	16.88
		6.4	100.00	100.00	100.00	100.00	100.00	100.00
	$(500, 160, 200)$	1.6	5.33	9.40	5.78	5.45	10.34	6.70
		3.2	17.47	23.73	18.17	16.96	22.88	17.94
		6.4	100.00	100.00	100.00	100.00	100.00	100.00
(ii) χ^2_{10}	$(100, 40, 50)$	1.6	6.79	12.51	8.03	5.78	10.71	6.60
		3.2	17.93	33.59	21.51	10.42	19.54	11.99
		6.4	99.84	100.00	100.00	80.46	98.71	89.34
	$(300, 80, 100)$	1.6	6.11	10.28	6.55	5.67	10.37	6.25
		3.2	14.18	21.32	15.25	13.31	21.21	14.74
		6.4	100.00	100.00	100.00	100.00	100.00	100.00
	$(500, 160, 200)$	1.6	5.67	9.88	6.36	5.39	9.92	5.87
		3.2	18.31	25.91	18.11	16.74	22.17	17.19
		6.4	100.00	100.00	100.00	100.00	100.00	100.00
(iii) t_{10}	$(100, 40, 50)$	1.6	7.05	12.32	8.21	6.09	10.90	7.11
		3.2	18.18	34.22	22.00	10.92	19.95	12.44
		6.4	99.91	100.00	99.99	81.19	98.94	90.00
	$(300, 80, 100)$	1.6	5.82	10.15	6.15	5.61	10.35	6.21
		3.2	13.31	20.01	15.41	14.22	23.20	16.43
		6.4	100.00	100.00	100.00	100.00	100.00	100.00
	$(500, 160, 200)$	1.6	5.28	9.28	5.79	5.51	9.73	6.06
		3.2	17.13	23.32	18.73	16.21	22.13	17.91
		6.4	100.00	100.00	100.00	100.00	100.00	100.00

retical findings and numerical studies justify that the proposed construction of the bias-adjusted L_2-based test statistic as well as its chi-square mixture approximation allows for better test size control and are therefore more suitable for the high-dimensional models with a latent factor structure than the tests proposed by e.g., Chen and Qin (2010) or Zang et al. (2021) where the structural aspects of the underlying population's distributions are not taken into account.

To conclude, the proposed testing methodology can be generally considered for high-dimensional, two-sample Behrens-Fisher problem when the data poses an unknown degree of heteroscedasticity with underlying low-dimensional latent factor structure. A very good performance of the test is observed for a number of practically used distributions of the factor vector and error term of (1.1) and spiked, heteroscedastic covariance structures of $\mathcal{F}_g$, where the dimension may far exceed the sample size, and for a moderate

Table 5: Empirical powers under $\mathcal{A}_2$ based on 10,000 simulation runs for **M1**: $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$

			$(d_1, d_2) = (2, 1)$			$(d_1, d_2) = (3, 2)$		
$(z_{gij}, \epsilon_{gij})$	(p, n_1, n_2)	δ_2	FA	Ch-Q	L2D	FA	Ch-Q	L2D
(i) $\mathcal{N}(0,1)$	$(100, 40, 50)$	0.4	16.20	24.41	19.98	12.48	20.32	16.03
		0.6	62.74	82.25	72.84	42.81	63.51	54.17
		0.8	100.00	100.00	100.00	99.99	100.00	100.00
	$(300, 80, 100)$	0.4	30.86	45.52	36.28	23.33	33.27	27.06
		0.6	100.00	100.00	100.00	99.96	100.00	99.99
		0.8	100.00	100.00	100.00	100.00	100.00	100.00
	$(500, 160, 200)$	0.4	53.18	72.39	58.37	38.03	51.68	42.11
		0.6	100.00	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00	100.00
(ii) χ^2_{10}	$(100, 40, 50)$	0.4	16.45	45.09	36.32	12.37	21.08	16.52
		0.6	62.11	80.42	71.51	43.82	64.53	54.60
		0.8	100.00	100.00	100.00	99.99	100.00	100.00
	$(300, 80, 100)$	0.4	31.57	45.09	36.32	21.88	31.81	25.68
		0.6	100.00	100.00	100.00	99.89	99.98	99.96
		0.8	100.00	100.00	100.00	100.00	100.00	100.00
	$(500, 160, 200)$	0.4	52.53	71.91	57.85	39.41	52.58	42.86
		0.6	100.00	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00	100.00
(iii) t_{10}	$(100, 40, 50)$	0.4	16.21	25.45	20.49	12.27	20.56	16.25
		0.6	63.99	81.39	73.26	43.28	63.85	54.05
		0.8	99.99	100.00	100.00	99.98	100.00	100.00
	$(300, 80, 100)$	0.4	31.99	45.82	36.86	22.93	33.42	27.02
		0.6	100.00	100.00	100.00	99.92	100.00	99.97
		0.8	100.00	100.00	100.00	100.00	100.00	100.00
	$(500, 160, 200)$	0.4	53.88	72.93	58.92	37.86	51.64	41.81
		0.6	100.00	100.00	100.00	100.00	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00	100.00

number of independent samples.

Acknowledgments

The research of the first author was supported by JSPS KAKENHI Grant Number 20K11712. The research of the second author was supported by JSPS KAKENHI Grant Number 20K11714 and general support of Senshu Research Abroad Program (FY 2020).

Appendix: Proof of main results

A. Technical lemma

The proofs of some main results rely on the following technical lemma.

Table 6: Empirical powers under $\mathcal{A}_2$ based on 10,000 simulation runs for **M2**: $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$

			$d_1 = d_2 = 2$			$d_1 = d_2 = 3$		
$(z_{gij}, \epsilon_{gij})$	(p, n_1, n_2)	δ_2	FA	Ch-Q	L2D	FA	Ch-Q	L2D
(i) $\mathcal{N}(0,1)$	$(100, 40, 50)$	0.4	10.27	19.33	12.43	8.88	15.88	9.96
		0.6	28.01	51.47	31.80	19.35	35.77	22.06
		0.8	99.24	100.00	99.96	89.81	99.85	97.03
	$(300, 80, 100)$	0.4	15.82	26.52	16.34	13.22	23.22	14.11
		0.6	97.30	100.00	99.55	75.43	98.92	84.29
		0.8	100.00	100.00	100.00	100.00	100.00	100.00
	$(500, 160, 200)$	0.4	20.84	37.40	22.73	16.91	28.77	16.86
		0.6	100.00	100.00	100.00	99.94	100.00	100.00
		0.8	100.00	100.00	100.00	100.00	100.00	100.00
(ii) χ^2_{10}	$(100, 40, 50)$	0.4	10.20	18.22	11.64	9.02	16.55	10.32
		0.6	28.62	51.92	32.62	19.08	36.19	22.11
		0.8	98.86	100.00	99.82	89.26	99.67	96.02
	$(300, 80, 100)$	0.4	15.65	27.25	16.75	12.54	22.71	13.76
		0.6	96.44	99.99	99.10	76.12	98.61	83.13
		0.8	100.00	100.00	100.00	100.00	100.00	100.00
	$(500, 160, 200)$	0.4	21.06	37.65	22.39	16.67	29.23	17.64
		0.6	100.00	100.00	100.00	99.84	100.00	99.96
		0.8	100.00	100.00	100.00	100.00	100.00	100.00
(iii) t_{10}	$(100, 40, 50)$	0.4	9.88	18.39	11.38	9.01	16.78	10.35
		0.6	28.81	50.96	31.91	19.61	36.90	23.42
		0.8	98.72	100.00	99.90	89.19	99.72	95.99
	$(300, 80, 100)$	0.4	15.83	27.53	17.08	12.73	23.22	14.14
		0.6	96.29	99.99	99.15	75.97	98.58	83.04
		0.8	100.00	100.00	100.00	100.00	100.00	100.00
	$(500, 160, 200)$	0.4	21.12	38.00	22.37	17.05	29.32	18.16
		0.6	100.00	100.00	100.00	99.86	100.00	99.98
		0.8	100.00	100.00	100.00	100.00	100.00	100.00

Lemma A 1. *Let* $\mathbf{z}_i = (\mathrm{z}_{i1}, \ldots, \mathrm{z}_{id})^\top$ *be a* d*-dimensional random vector such that* $\mathrm{E}(\mathrm{z}_{i\ell}) = 0$*,* $\mathrm{E}(\mathrm{z}_{i\ell}^2) = 1$*,* $\mathrm{E}(\mathrm{z}_{i\ell}^4) = \kappa_{\mathrm{z}} < \infty$*, and* $\mathrm{z}_{i\ell}$ *are iid for* $i \in [\![n_g]\!]$ *and* $\ell \in [\![d]\!]$*, Then for any* $\mathbf{A} \in \mathbb{R}^{d\times d}_{>0}$ *it holds that*

$$\mathrm{E}\{(\mathbf{z}_i^\top \mathbf{A}\mathbf{z}_i)\} = \mathrm{tr}(\mathbf{A}),\ \ \mathrm{E}\{(\mathbf{z}_i^\top \mathbf{A}\mathbf{z}_i)^2\} = \kappa_{\mathrm{z}}\mathrm{tr}(\mathbf{A} \odot \mathbf{A}) + \{\mathrm{tr}(\mathbf{A})\}^2 + 2\mathrm{tr}(\mathbf{A}^2),$$

where $\odot$ *denotes the Hadamard product defined as* $(\mathbf{A} \odot \mathbf{A})_{ij} = (\mathbf{A})_{ij} \times (\mathbf{A})_{ij}$.

Proof. See, Watanabe et al. (2019). □

B. Proof of Theorem 1

At first, we show (i). By using Lemma 1 in Ma et al. (2015), for $\ell \in [\![d_1 + d_2]\!]$,

$$\begin{aligned}\lambda_\ell(\mathbf{\Omega}) &= \lambda_\ell(\mathbf{B}\mathbf{B}^\top + \mathbf{\Psi}_0)\\ &\leqslant \lambda_\ell(\mathbf{B}\mathbf{B}^\top) + \lambda_1(\mathbf{\Psi}_0) = \lambda_\ell(\mathbf{B}^\top\mathbf{B}) + \frac{n}{n_1}\psi_{1\max} + \frac{n}{n_2}\psi_{2\max},\\ \lambda_\ell(\mathbf{\Omega}) &= \lambda_\ell(\mathbf{B}\mathbf{B}^\top + \mathbf{\Psi}_0) \geqslant \lambda_\ell(\mathbf{B}\mathbf{B}^\top) = \lambda_\ell(\mathbf{B}^\top\mathbf{B}),\end{aligned}$$

where $\mathbf{\Psi}_0 = (n/n_1)\mathbf{\Psi}_1 + (n/n_2)\mathbf{\Psi}_2$. From this result, under (A2), $\lambda_\ell(\mathbf{\Omega})/p \to \lambda_\ell(\mathbf{A})$ for $\ell \in [\![d_1 + d_2]\!]$. On the other hand, by Lemma 1 in Ma et al. (2015), for $\ell \in [\![d_1 + d_2 + 1, p]\!]$

$$\lambda_\ell(\mathbf{\Omega}) = \lambda_\ell(\mathbf{B}\mathbf{B}^\top + \mathbf{\Psi}_0) \leqslant \lambda_\ell(\mathbf{B}\mathbf{B}^\top) + \lambda_1(\mathbf{\Psi}_0) = \frac{n}{n_1}\psi_{1\max} + \frac{n}{n_2}\psi_{2\max}.$$

From this result, under (A2), $\lambda_\ell(\mathbf{\Omega})/p \to 0$ for $\ell \in [\![d_1 + d_2 + 1, p]\!]$.

Next, we show (ii). Under (A2), $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are positive definite matrices since $\mathbf{A}$ is positive definite matrix and $1 + \gamma^{-1}, 1 + \gamma > 0$. From this fact, we can also prove (ii) using arguments similar to those used in proof of the statement (i). The proof of Theorem 1 is then complete.

C. Proof of Theorem 2

We define $\overline{\mathbf{z}}_g = n_g^{-1}\sum_{i=1}^{n_g}\mathbf{z}_{gi}$ and $\overline{\boldsymbol{\epsilon}}_g = n_g^{-1}\sum_{i=1}^{n_g}\boldsymbol{\epsilon}_{gi}$ for $g \in [\![2]\!]$. Then under the null hypothesis $\mathcal{H}$ and the factor model (1.3), T can be rewritten as

$$\begin{aligned}T =& \frac{n}{p}(\mathbf{B}_1\overline{\mathbf{z}}_1 - \mathbf{B}_2\overline{\mathbf{z}}_2)^\top(\mathbf{B}_1\overline{\mathbf{z}}_1 - \mathbf{B}_2\overline{\mathbf{z}}_2) + \frac{2n}{p}(\mathbf{B}_1\overline{\mathbf{z}}_1 - \mathbf{B}_2\overline{\mathbf{z}}_2)^\top(\mathbf{\Psi}_1^{1/2}\overline{\boldsymbol{\epsilon}}_1 - \mathbf{\Psi}_2^{1/2}\overline{\boldsymbol{\epsilon}}_2)\\ &+ \frac{n}{p}(\mathbf{\Psi}_1^{1/2}\overline{\boldsymbol{\epsilon}}_1 - \mathbf{\Psi}_2^{1/2}\overline{\boldsymbol{\epsilon}}_2)^\top(\mathbf{\Psi}_1^{1/2}\overline{\boldsymbol{\epsilon}}_1 - \mathbf{\Psi}_2^{1/2}\overline{\boldsymbol{\epsilon}}_2) - \frac{n}{pn_1}\mathrm{tr}(\mathbf{\Psi}_1) - \frac{n}{pn_2}\mathrm{tr}(\mathbf{\Psi}_2)\\ =& T_1 + 2T_2 + T_3, \end{aligned} \tag{C.1}$$

where

$$\begin{aligned}T_1 =& \frac{1}{p}\mathbf{y}^\top\mathbf{B}^\top\mathbf{B}\mathbf{y}, \quad T_2 = \frac{1}{p}\mathbf{y}^\top\mathbf{B}^\top(\mathbf{\Psi}_1^{1/2}\overline{\boldsymbol{\epsilon}}_1 - \mathbf{\Psi}_2^{1/2}\overline{\boldsymbol{\epsilon}}_2),\\ T_3 =& \frac{n}{p}\left\{(\mathbf{\Psi}_1^{1/2}\overline{\boldsymbol{\epsilon}}_1 - \mathbf{\Psi}_2^{1/2}\overline{\boldsymbol{\epsilon}}_2)^\top(\mathbf{\Psi}_1^{1/2}\overline{\boldsymbol{\epsilon}}_1 - \mathbf{\Psi}_2^{1/2}\overline{\boldsymbol{\epsilon}}_2) - \frac{1}{n_1}\mathrm{tr}(\mathbf{\Psi}_1) - \frac{1}{n_2}\mathrm{tr}(\mathbf{\Psi}_2)\right\},\end{aligned}$$

and $\mathbf{y} = (\sqrt{n_1}\bar{\mathbf{z}}_1^\top, \sqrt{n_2}\bar{\mathbf{z}}_2^\top)^\top$. Then under (1.3) and (A1)-(A2), the following results hold true

$$T_1 = O_{\mathbb{P}}(1),\ T_2 = o_{\mathbb{P}}(1),\ T_3 = o_{\mathbb{P}}(1). \tag{C.2}$$

The detailed evaluation of the claims (C.2) is presented in what follows. At first, we verify that $T_1 = O_{\mathbb{P}}(1)$. Direct calculations give the first two moments of T_1 as

$$\begin{aligned}
\mathrm{E}(T_1) =& \frac{n}{pn_1}\mathrm{tr}(\mathbf{B}_1^\top\mathbf{B}_1) + \frac{n}{pn_2}\mathrm{tr}(\mathbf{B}_2^\top\mathbf{B}_2) = \mathrm{tr}(\mathbf{A}) + o(1),\\
\mathrm{E}(T_1^2) =& \frac{\kappa_{z_1}n^2}{n_1^3}\mathrm{tr}(\mathbf{A}_{11}\odot\mathbf{A}_{11}) + \frac{n^2}{n_1^2}\{\mathrm{tr}(\mathbf{A}_{11})\}^2 + \frac{2n^2}{n_1^2}\mathrm{tr}(\mathbf{A}_{11}^2)\\
&+ \frac{2n^2}{n_1n_2}\mathrm{tr}(\mathbf{A}_{11})\mathrm{tr}(\mathbf{A}_{22}) + \frac{4n^2}{n_1n_2}\mathrm{tr}(\mathbf{A}_{12}\mathbf{A}_{21}) + \frac{\kappa_{z_2}n^2}{n_2^3}\mathrm{tr}(\mathbf{A}_{22}\odot\mathbf{A}_{22})\\
&+ \frac{n^2}{n_2^2}\{\mathrm{tr}(\mathbf{A}_{22})\}^2 + \frac{2n^2}{n_1^2}\mathrm{tr}(\mathbf{A}_{22}^2) + o(1)\\
=&\{\mathrm{tr}(\mathbf{A})\}^2 + 2\mathrm{tr}(\mathbf{A}^2) + o(1).
\end{aligned}$$

It is now straightforward to see that $\mathrm{var}(T_1) = O(1)$ which gives the first claim of (C.2). Following analogous steps to prove the two remaining claims, we obtain the first two moments of T_2 and T_3 as

$$\begin{aligned}
\mathrm{E}(T_2) =&0,\\
\mathrm{E}(T_2^2) =& \sum_{g=1}^{2}\frac{n^2}{pn_g^2}\mathrm{tr}(\mathbf{A}_{gg}\boldsymbol{\Psi}_g) + \frac{n^2}{pn_1n_2}\mathrm{tr}(\mathbf{A}_{11}\boldsymbol{\Psi}_2) + \frac{n^2}{pn_1n_2}\mathrm{tr}(\mathbf{A}_{22}\boldsymbol{\Psi}_1) + o(p^{-1})\\
=&O(p^{-1}),
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}(T_3) =&0,\\
\mathrm{E}(T_3^2) =& \frac{2}{p^2}\mathrm{tr}\left\{\left(\frac{n}{n_1}\boldsymbol{\Psi}_1 + \frac{n}{n_2}\boldsymbol{\Psi}_2\right)^2\right\} + \frac{\kappa_\epsilon}{np^2}\mathrm{tr}\left(\frac{n^3}{n_1^3}\boldsymbol{\Psi}_1^2 + \frac{n^3}{n_2^3}\boldsymbol{\Psi}_1^2\right) = O\left(p^{-1}\right),
\end{aligned}$$

respectively. These yield $T_2 = o_{\mathbb{P}}(1)$ and $T_3 = o_{\mathbb{P}}(1)$. Consequently, in a view of (C.2), the terms T_2 and T_3 are stochastically negligible in the representation (C.1) and the asymptotics of T is dominated by T_1, which we analyze below.

To derive the asymptotic distribution of T_1, we first observe that by the multivariate central limit theorem, $\sqrt{n_1}\overline{\mathbf{z}}_1 \rightsquigarrow \mathcal{N}_{d_1}(\mathbf{0}, \mathbf{I}_{d_1})$ and $\sqrt{n_2}\overline{\mathbf{z}}_2 \rightsquigarrow \mathcal{N}_{d_2}(\mathbf{0}, \mathbf{I}_{d_2})$ as $\min\{n_1, n_2\} \to \infty$. Also, by independence of $\sqrt{n_1}\overline{\mathbf{z}}_1$ and $\sqrt{n_1}\overline{\mathbf{z}}_2$ we get $\mathbf{y} = (\sqrt{n_1}\overline{\mathbf{z}}_1^\top, \sqrt{n_2}\overline{\mathbf{z}}_2^\top)^\top \rightsquigarrow \mathcal{N}_{d_1+d_2}(\mathbf{0}, \mathbf{I}_{d_1+d_2})$ as $\min\{n_1, n_2\} \to \infty$. Further, for the matrix $\mathbf{A}$ under (A2), there exist an orthogonal matrix $\mathbf{P}$ such that $\mathbf{P}\mathbf{P}^\top = \mathbf{P}^\top\mathbf{P} = \mathbf{I}_{d_1+d_2}$ and $\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^\top$, where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1(\mathbf{A}), \ldots, \lambda_{d_1+d_2}(\mathbf{A}))$. Since $\mathbf{w} = \mathbf{P}^\top\mathbf{y} \rightsquigarrow \mathcal{N}_{d_1+d_2}(\mathbf{0}, \mathbf{I}_{d_1+d_2})$ as $\min\{n_1, n_2\} \to \infty$, we get

$$T_1 \rightsquigarrow \sum_{\ell=1}^{d_1+d_2} \lambda_\ell(\mathbf{A})\chi_\ell^2(1). \tag{C.3}$$

Combining (C.2), (C.3), and Slutsky's theorem, we have Theorem 1.

D. Proof of Lemma 1

We define $\tilde{\mathbf{x}}_{gi} = \mathbf{x}_{gi} - \boldsymbol{\mu}_g$ and $\widetilde{\mathbf{X}}_g^\top = (\tilde{\mathbf{x}}_{g1}, \ldots, \tilde{\mathbf{x}}_{gn_g})$. Then the result when changing $\lambda_\ell(\mathbf{S}_g/p)$ to $\lambda_\ell\{\widetilde{\mathbf{X}}_g\widetilde{\mathbf{X}}_g^\top/(n_g p)\}$ in $ER_g(k)$ is directly proved by Ahn and Horenstein (2013).

For the concrete proof, denote

$$\mathbf{Z}_g^\top = (\mathbf{z}_{g1}, \ldots, \mathbf{z}_{gn_g}), \quad \mathbf{E}^\top = (\boldsymbol{\Psi}_g\boldsymbol{\epsilon}_{g1}, \ldots, \boldsymbol{\Psi}_g\boldsymbol{\epsilon}_{gn_g})$$

for each $g \in [\![2]\!]$. In these notations, the factor model (1.3) becomes $\widetilde{\mathbf{X}}_g = \mathbf{Z}_g\mathbf{B}_g^\top + \mathbf{E}_g$. Further, it is easy to see that

$$\frac{1}{n_g - 1}\widetilde{\mathbf{X}}_g\widetilde{\mathbf{X}}_g^\top = \mathbf{S}_g + \frac{n_g}{n_g - 1}(\overline{\mathbf{x}}_g - \boldsymbol{\mu}_g)(\overline{\mathbf{x}}_g - \boldsymbol{\mu}_g)^\top.$$

Now, by Lemma A.5 and Lemma A.6 in Ahn and Horenstein (2013), for all $\ell \in [\![p]\!]$

$$\lambda_\ell\left(\frac{1}{n_g - 1}\widetilde{\mathbf{X}}_g\widetilde{\mathbf{X}}_g^\top\right) - \frac{n_g\|\overline{\mathbf{x}}_g - \boldsymbol{\mu}_g\|^2}{n_g - 1} \leqslant \lambda_\ell\left(\mathbf{S}_g\right) \leqslant \lambda_\ell\left(\frac{1}{n_g - 1}\widetilde{\mathbf{X}}_g\widetilde{\mathbf{X}}_g^\top\right). \tag{D.1}$$

Since $\|\overline{\mathbf{x}}_g - \boldsymbol{\mu}_g\|^2 = O_p\{\mathrm{tr}(\boldsymbol{\Sigma}_g)/n_g\} = O_p(p/n_g)$,

$$\lambda_\ell\left(\frac{1}{n_g - 1}\widetilde{\mathbf{X}}_g\widetilde{\mathbf{X}}_g^\top\right) - \frac{n_g\|\overline{\mathbf{x}}_g - \boldsymbol{\mu}_g\|^2}{n_g - 1} = \lambda_\ell\left(\frac{1}{n_g - 1}\widetilde{\mathbf{X}}_g\widetilde{\mathbf{X}}_g^\top\right) + O_p\left(\frac{p}{n_g}\right). \tag{D.2}$$

Furthermore, the inspection of Section 2 in Ahn and Horenstein (2013) reviles that the factor model (1.3) at hand, along with (A1) and (A2), is a special case of the factor model (1) under assumptions (A)–(D) in Ahn and Horenstein's asymptotic setting. Therefore, we can apply Lemma A.9 in Ahn and Horenstein (2013), by which for all $\ell \in [\![d_g + 1, \lfloor c_g \min\{n_g, p\}\rfloor - d_g]\!]$ it holds that

$$\lambda_\ell \left(\frac{1}{n_g p} \widetilde{\mathbf{X}} \widetilde{\mathbf{X}}^\top \right) \overset{\mathbb{P}}{\asymp} \frac{1}{\min\{n_g, p\}}. \tag{D.3}$$

By Lemma A.11 of Ahn and Horenstein (2013), for all $\ell \in [\![d_g]\!]$,

$$\lambda_\ell \left(\frac{1}{n_g p} \widetilde{\mathbf{X}}_g \widetilde{\mathbf{X}}_g^\top \right) = \lambda_\ell \left(\mathbf{A}_{gg} \right) + o_{\mathbb{P}}(1). \tag{D.4}$$

Combining (D.1)–(D.4) yields

$$\lambda_\ell \left(\frac{1}{p} \mathbf{S}_g \right) = \lambda_\ell \left(\mathbf{A}_{gg} \right) + o_{\mathbb{P}}(1), \quad \ell \in [\![d_g]\!], \tag{D.5}$$

$$\lambda_\ell \left(\frac{1}{p} \mathbf{S}_g \right) \overset{\mathbb{P}}{\asymp} \frac{1}{\min\{n_g, p\}}, \quad \ell \in [\![d_g + 1, \lfloor c_g \min\{n_g, p\}\rfloor - d_g]\!]. \tag{D.6}$$

For $\ell \in [\![\lfloor c_g \min\{n_g, p\}\rfloor - d_g - 1]\!]$,

$$ER_g(\ell) \overset{\mathbb{P}}{\asymp} \begin{cases} \min\{n_g, p\} & \ell = d_g \\ 1 & \ell \neq d_g \end{cases},$$

which gives consistency of $\hat{d}_g$ in high dimensions.

E. Proof of Lemma 2

Let $\widetilde{\mathbf{S}}_g = \sum_{i=1}^{n_g} (\mathbf{x}_{gi} - \boldsymbol{\mu}_g)(\mathbf{x}_{gi} - \boldsymbol{\mu}_g)^\top / (n_g - 1) = \widetilde{\mathbf{X}}_g \widetilde{\mathbf{X}}_g^\top / (n_g - 1)$. If assuming (A1) and (A2) with the factor model (1.3), then the factor model (1) and assumptions (A)–(D) in Ahn and Horenstein (2013) are satisfied. Therefore, Lemma 2 when changing $\lambda_\ell(\mathbf{S}_g/p)$ to $\lambda_\ell(\widetilde{\mathbf{S}}_g/p)$ is directly proved by Ahn and Horenstein (2013). Since $\mathbf{S}_g$ and $\widetilde{\mathbf{S}}_g$ are asymptotically equivalent, after replacing $\widetilde{\mathbf{S}}_g$ with $\mathbf{S}_g$, the result is still justified.

F. Proof of Theorem 3

Using Lemma A1,

$$\begin{aligned} \mathrm{E}\{\mathrm{tr}(\mathbf{S}_g)\} =& \mathrm{tr}(\mathbf{\Sigma}_g), \\ \mathrm{E}\left[\{\mathrm{tr}(\mathbf{S}_g)\}^2\right] =& \frac{\kappa_{z_g}\mathrm{tr}(\mathbf{A}_{gg}\odot\mathbf{A}_{gg}) + \kappa_{\epsilon_g}\mathrm{tr}(\mathbf{\Psi}_g^2)}{n_g} + \frac{2}{n_g-1}\mathrm{tr}(\mathbf{\Sigma}_g^2) + \{\mathrm{tr}(\mathbf{\Sigma}_g)\}^2. \end{aligned}$$

From these results, under the model 1.3 and assumptions (A1)-(A2), $\mathrm{var}\{\mathrm{tr}(\mathbf{\Sigma}_g)/p\} = O(1/n_g)$. Thus,

$$\frac{\mathrm{tr}(\mathbf{S}_g)}{p} = \frac{\mathrm{tr}(\mathbf{\Sigma}_g)}{p} + o_{\mathbb{P}}(1). \tag{F.1}$$

From Lemma 1 and Lemma 2,

$$\frac{\sum_{\ell=1}^{\hat{d}_g}\lambda_\ell(\mathbf{S}_g)}{p} = \frac{\sum_{\ell=1}^{d_g}\lambda_\ell(\mathbf{A}_{gg})}{p} + o_{\mathbb{P}}(1) = \frac{\mathrm{tr}(\mathbf{A}_{gg})}{p} + o_{\mathbb{P}}(1). \tag{F.2}$$

Combining (F.1) and (F.2),

$$\frac{\widehat{\mathrm{tr}(\mathbf{\Psi}_g)}}{p} = \frac{\mathrm{tr}(\mathbf{\Sigma}_g)}{p} - \frac{\mathrm{tr}(\mathbf{A}_{gg})}{p} + o_{\mathbb{P}}(1) = \frac{\mathrm{tr}(\mathbf{\Psi}_g)}{p} + o_{\mathbb{P}}(1).$$

From this result, under the model (1.3) and assumptions (A1) and (A2), $T_{FA} = T + o_{\mathbb{P}}(1)$. Thus, combining Theorem 2 and Slutsky's theorem, we have Theorem 3.

G. Proof of Theorem 4

From Ahn and Horenstein (2013), under (A1) and (A2),

$$\widehat{\lambda_\ell(\mathbf{A})} = \begin{cases} \lambda_\ell(\mathbf{A}) + o_{\mathbb{P}}(1) & \text{for } \ell \in [\![d_0]\!] \\ o_{\mathbb{P}}(1) & \text{for } \ell \in [\![d_0+1, d_1+d_2]\!] \end{cases}.$$

We write

$$\begin{aligned} &\sum_{\ell=1}^{\hat{d}_1+\hat{d}_2}\hat{\lambda}_\ell(\mathbf{A})\chi_\ell^2(1) - \sum_{\ell=1}^{d_1+d_2}\lambda_\ell(\mathbf{A})\chi_\ell^2(1), \\ &= \sum_{\ell=1}^{\hat{d}_1+\hat{d}_2}\left\{\hat{\lambda}_\ell(\mathbf{A}) - \lambda_\ell(\mathbf{A})\right\}\chi_\ell^2(1) + \sum_{\ell=1}^{\hat{d}_1+\hat{d}_2}\lambda_\ell(\mathbf{A})\chi_\ell^2(1) - \sum_{\ell=1}^{d_1+d_2}\lambda_\ell(\mathbf{A})\chi_\ell^2(1). \end{aligned}$$

Thus, it is sufficient to show the following (i) and (ii):

$$\text{(i)} \quad \sum_{\ell=1}^{\widehat{d}_1+\widehat{d}_2} \left\{ \widehat{\lambda}_\ell(\mathbf{A}) - \lambda_\ell(\mathbf{A}) \right\} \chi_\ell^2(1) = o_{\mathbb{P}}(1)$$

$$\text{(ii)} \quad \sum_{\ell=1}^{\widehat{d}_1+\widehat{d}_2} \lambda_\ell(\mathbf{A}) \chi_\ell^2(1) - \sum_{\ell=1}^{d_1+d_2} \lambda_\ell(\mathbf{A}) \chi_\ell^2(1) = o_{\mathbb{P}}(1).$$

At first, we show (i). For any $\epsilon > 0$, we have

$$\begin{aligned} &\Pr\left(\left| \sum_{\ell=1}^{\widehat{d}_1+\widehat{d}_2} \left\{ \widehat{\lambda}_\ell(\mathbf{A}) - \lambda_\ell(\mathbf{A}) \right\} \chi_\ell^2(1) \right| > \epsilon \right) \\ &\leqslant \Pr\left(\left| \sum_{\ell=1}^{d_1+d_2} \left\{ \widehat{\lambda}_\ell(\mathbf{A}) - \lambda_\ell(\mathbf{A}) \right\} \chi_\ell^2(1) \right| > \epsilon \right) + \Pr\left\{ (\widehat{d}_1, \widehat{d}_2) \neq (d_1, d_2) \right\}. \end{aligned} \tag{G.1}$$

Since $\widehat{\lambda}_\ell(\mathbf{A}) = \lambda_\ell(\mathbf{A}) + o_{\mathbb{P}}(1)$, we get

$$\Pr\left(\left| \sum_{\ell=1}^{d_1+d_2} \left\{ \widehat{\lambda}_\ell(\mathbf{A}) - \lambda_\ell(\mathbf{A}) \right\} \chi_\ell^2(1) \right| > \epsilon \right) = o(1). \tag{G.2}$$

Also, by Lemma 1, we note that

$$\Pr\left\{ (\widehat{d}_1, \widehat{d}_2) \neq (d_1, d_2) \right\} \leqslant \Pr\left(\widehat{d}_1 \neq d_1 \right) + \Pr\left(\widehat{d}_2 \neq d_2 \right) = o(1). \tag{G.3}$$

Plugging (G.2) and (G.3) into (G.1) yields

$$\Pr\left(\left| \sum_{\ell=1}^{\widehat{d}_1+\widehat{d}_2} \left\{ \widehat{\lambda}_\ell(\mathbf{A}) - \lambda_\ell(\mathbf{A}) \right\} \chi_\ell^2(1) \right| > \epsilon \right) = o(1),$$

which proves (i).

Next, we show (ii). For any $\epsilon > 0$, we have

$$
\begin{aligned}
&\Pr\left(\left|\sum_{\ell=1}^{\widehat{d}_1+\widehat{d}_2} \lambda_\ell(\mathbf{A})\chi_\ell^2(1) - \sum_{\ell=1}^{d_1+d_2} \lambda_\ell(\mathbf{A})\chi_\ell^2(1)\right| > \epsilon\right) \\
&= \Pr\left\{\left|\sum_{\ell=1}^{\widehat{d}_1+\widehat{d}_2} \lambda_\ell(\mathbf{A})\chi_\ell^2(1) - \sum_{\ell=1}^{d_1+d_2} \lambda_\ell(\mathbf{A})\chi_\ell^2(1)\right| > \epsilon,\ (\widehat{d}_1, \widehat{d}_2) \neq (d_1, d_2)\right\} \\
&\leqslant \Pr\left\{(\widehat{d}_1, \widehat{d}_2) \neq (d_1, d_2)\right\} \\
&\leqslant \Pr\left(\widehat{d}_1 \neq d_1\right) + \Pr\left(\widehat{d}_2 \neq d_2\right) = o(1).
\end{aligned}
$$

This proves (ii) and finishes the proof of Theorem 4.

H. Proof of Theorem 5

From (2.6), we note that

$$
T_{FA} = T_0 + \frac{2n}{p}S_n + \frac{n}{p}\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2, \tag{H.1}
$$

where

$$
\begin{aligned}
T_0 =& \frac{n}{p}\left\{\|(\overline{\mathbf{x}}_1 - \boldsymbol{\mu}_1) - (\overline{\mathbf{x}}_2 - \boldsymbol{\mu}_2)\|^2 - \frac{1}{n_1}\mathrm{tr}(\boldsymbol{\Psi}_1) - \frac{1}{n_2}\mathrm{tr}(\boldsymbol{\Psi}_2)\right\}, \\
S_n =& \{(\overline{\mathbf{x}}_1 - \boldsymbol{\mu}_1) - (\overline{\mathbf{x}}_2 - \boldsymbol{\mu}_2)\}^\top(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2).
\end{aligned}
$$

If $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2 = O(1)$, then under (A1) and (A2),

$$
T_0 \rightsquigarrow \mathbf{y}_0^\top \mathbf{A}\mathbf{y}_0,\ \frac{n}{p}S_n \rightsquigarrow \frac{\sqrt{n}}{p}\mathbf{y}_0^\top \mathbf{B}^\top(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2), \tag{H.2}
$$

where $\mathbf{y}_0 \sim \mathcal{N}_{d_1+d_2}(\mathbf{0}, \mathbf{I}_{d_1+d_2})$. Combining (H.1) and (H.2), we obtain the asymptotic power under $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2 = O(1)$.

If $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2 \to \infty$ under (A1) and (A2),

$$
T_0 = O_{\mathbb{P}}(1),\ \frac{n}{p}S_n = O_{\mathbb{P}}(\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|). \tag{H.3}
$$

Combining (H.1) and (H.3), $T_{FA} = \eta\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2 + o_{\mathbb{P}}(\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2)$ under (A1), (A2) and $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2 \to \infty$. In other words, the power converges to 1 under (A1), (A2) and $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2 \to \infty$.

References

[1] S. C. Ahn, A. R. Horenstein, Eigenvalue ratio test for the number of factors. Econometrica 81 (2013) 1203–1227.

[2] M. Aoshima, K. Yata, Two-Sample Tests for High-Dimension, Strongly Spiked Eigenvalue Models. Statist. Sinica 28 (2018) 43–62.

[3] Z. D. Bai, H. Saranadasa, Effect of high dimension: by an example of a two sample problem. Statist. Sinica 6 (1996) 311–329.

[4] S. X. Chen, Y. L. Qin, A two-sample test for high-dimensional data with applications to gene-set testing. Ann. Statist. 38 (2010) 808–835.

[5] S. Dudoit, J. Fridlyand, P. T. Speed, Comparison of discrimination methods for the classification of tumors using gene expression data. J. Amer. Statist. Assoc. 97 (2002) 77–87.

[6] A. Ishii, A high-dimensional two-sample test for non-Gaussian data under a strongly spiked eigenvalue model. J. Japan Statist. Soc. 47 (2017) 273–291.

[7] A. Ishii, K. Yata, M. Aoshima, Inference on high-dimensional mean vectors under the strongly spiked eigenvalue model. Jpn. J. Stat. Data Sci. 2 (2019) 105–128.

[8] Y. Ma, W. Lan, H. Wang, A high dimensional two-sample test under a low dimensional factor structure. J. Multivariate Anal. 140 (2015) 162–170.

[9] A. Onatski, Determining the number of factors from empirical distribution of eigenvalues. Rev. Econom. Statist. 92 (2010) 1004–1016.

[10] H. Wang, Factor profiled sure independence screening. Biometrika 99 (2012) 15–28.

[11] H. Watanabe, M. Hyodo, Y. Yamada, T. Seo, Estimation of misclassification probability for a distance-based classifier in high-dimensional data. Hiroshima Math. J. 49 (2019) 175–193.

[12] J. T. Zhang, B. Zhou, J. Guo, T. Zhu, Two-sample Behrens-Fisher problems for high-dimensional data: A normal reference approach. J. Statist. Plann. Inference 213 (2021) 142–161.