<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Two sample Behrens-fisher problem for high-dimensional data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Manuscript ID</strong></td>
<td>SS-14-048R1</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://www.stat.sinica.edu.tw/statistica/">http://www.stat.sinica.edu.tw/statistica/</a></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
<td>10.5705/ss.2014.048</td>
</tr>
<tr>
<td><strong>Complete List of Authors</strong></td>
<td>Long Feng, Changliang Zou, Zhaojun Wang and Lixing Zhu</td>
</tr>
<tr>
<td><strong>Corresponding Author</strong></td>
<td>Lixing Zhu</td>
</tr>
<tr>
<td><strong>E-mail</strong></td>
<td><a href="mailto:lzhu@hkbu.edu.hk">lzhu@hkbu.edu.hk</a></td>
</tr>
</tbody>
</table>

Notice: Accepted version subject to English editing.
TWO-SAMPLE BEHRENS-FISHER PROBLEM FOR HIGH-DIMENSIONAL DATA

Long Feng¹, Changliang Zou¹, Zhaojun Wang¹, Lixing Zhu²

¹Nankai University and ²Hong Kong Baptist University

Abstract: This article is concerned with the two-sample Behrens-Fisher problem in high-dimensional settings. A novel test is proposed that is scale-invariant, asymptotically normal under certain mild conditions, and the dimensionality is allowed to grow in the rate, respectively, from square to cube of the sample size in different scenarios. We explain the necessity of bias correction for existing scale-invariant tests, otherwise they do not have well-defined limits even under the null hypothesis. We also give some examples to theoretically show the advantage of the scale-invariant test over scale-variant tests when variances of the two samples are different.

Key words and phrases: Asymptotic normality, Behrens-Fisher problem, High-dimensional data, Large-p-small-n, Two-sample test.

1. Introduction

This article is concerned with the two-sample Behrens-Fisher problem in high-dimensional settings. Assume that \(\{X_{i1}, \cdots, X_{in_i}\} \) for \(i = 1, 2\) are two independent random samples with the sizes \(n_1\) and \(n_2\), from \(p\)-variate distributions \(F(x - \mu_1)\) and \(G(x - \mu_2)\) located at \(p\)-variate centers \(\mu_1\) and \(\mu_2\). Denote \(n = n_1 + n_2\). We wish to test

\[
H_0 : \mu_1 = \mu_2 \quad \text{versus} \quad H_1 : \mu_1 \neq \mu_2,
\]

where their covariances \(\Sigma_1\) and \(\Sigma_2\) are unknown. For testing (1.1), Bai and Saranadasa (1996) proposed a test statistic based on \(M_n = \|\bar{X}_1 - \bar{X}_2\|^2\) which is developed under the equality of the two covariances, say \(\Sigma_1 = \Sigma_2 = \Sigma\). The key feature of the Bai and Saranadasa’s proposal is to use the Euclidian norm to replace the Mahalanobis norm since having the inverse of the sample covariance matrix is no longer beneficial when \(p/n \to c > 0\). Zhang and Xu (2009) extended this method to the \(k\)-sample high-dimensional Behrens-Fisher problem and derived the asymptotic distribution of the test statistic when \(p/n \to c < 1\).
To allow simultaneous testing for ultra high-dimensional data, Chen and Qin (2010) considered the following test statistic

\[ W_n = \frac{\sum_{i \neq j}^{n_1} X_{1i}^T X_{1j}}{n_1(n_1 - 1)} + \frac{\sum_{i \neq j}^{n_2} X_{2i}^T X_{2j}}{n_2(n_2 - 1)} - 2\frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} X_{1i}^T X_{2j}}{n_1 n_2} \]  

(1.2)

by removing \( \sum_{j=1}^{n_i} X_{ij}^T X_{ij} \) for \( i = 1, 2 \) from \( ||\bar{X}_1 - \bar{X}_2||^2 \) because these terms impose demands on the dimensionality. The asymptotic normality of \( W_n \) is established without imposing any explicit restriction between \( p \) and \( n \) directly. The restriction on the dimension is that the number of divergent eigenvalues of \( \Sigma_1 \) and \( \Sigma_2 \) is not too large and the divergence rate is not too fast.

However, as a test statistic, an important requirement is scale-invariance, otherwise the test may suffer from scalar transformations: the same dataset might generate different conclusions due to different scalar transformations. This is an obvious limitation of \( W_n \) and \( M_n \) as they are not invariant under scalar transformations. Intuitively speaking, both of the two tests take the sum of all \( p \) squared mean differences without using the information from the diagonal elements of the sample covariance, i.e., the variances of variables, and thus their test power would heavily depend on the underlying variance magnitudes. When all the components are (approximately) homogeneous, \( W_n \) or \( M_n \) would be useful in certain scenarios, whereas their superiority would be greatly deteriorated if the component variances differ much. In practice, different components may have completely different physical or biological readings and thus certainly their scales would not be identical. Hence, it is desirable to develop scalar-transformation-invariant tests for the Behrens-Fisher problem, which are able to integrate all the individual information in a relatively “fair” way.

For this, under the normality assumption, Srivastava and Du (2008) proposed a scalar-transformation-invariant test under the assumption of the equality of the two covariance matrices. Srivastava et al. (2013) extended their results to unequal covariance matrices. However, to derive the well-defined asymptotic null distribution, the dimension \( p \) must have a smaller order of \( n^2 \) otherwise, their test would not have a well-defined limit because of a non-negligible bias-term. Park and Ayyala (2013) also proposed a scale-invariant test from the idea of leave-out cross validation. However, their test is not shift-invariant. When the common mean vector is not zero, the expectation of their test statistic is not
zero either even under the null hypothesis. Therefore, the test is not powerful for significance level maintenance and power enhancement.

To overcome these problems, we propose another novel test via standardizing each component of the difference of two-sample means by the corresponding variance and suggest a simple but effective test statistic. Two desired features of the resulting test statistic are summarized as follows. On one hand, the proposed test is invariant under scalar transformations. The asymptotic normality of the proposed test can be derived under some very mild conditions similar to those in Chen and Qin (2010).

On the other hand, which is more importantly, when the dimension $p$ gets larger, the practically used assumption that every component of $x$ is standardized with variance 1 is not appropriate in high-dimensional scenarios as the standardization brings too many plug-in sample variances and does seriously affect, with very complicated bias-term, the asymptotic behaviors of the test. Actually, for such a scale of dimensionality, all exiting scale-invariant tests may also suffer asymptotically from this problem with bias-terms. The reason behind is mainly because the sample variance is only root-$n$ consistent and thus, the bias cannot be eliminated asymptotically. When $p$ has smaller order than $n^2$, bias correction may be no need, see, Srivastava and Du (2008) and Srivastava et al. (2013). However, when $p$ is of the order $n^2$ or higher, there are two sources to bring up the bias: the squared term in the test and too many variance estimators. As we commented above, the bias term will go to infinity as the sample size tends to infinity. As such, to make scale-invariant tests useful, a calibration or bias correction is necessary.

The remainder of the paper is organized as follows. In the next section, the test statistic is constructed and its asymptotic normality is established. A bias correction to the expectation of the proposed test is developed and the associated plug-in estimators are suggested. Simulation comparison is conducted in Section 3. Section 4 contains a sensor detection example to illustrate the application of the proposed test. Finally several remarks in Section 5 conclude the paper. All technical details and some additional simulation results are provided in the Appendix in a supplementary file.
2. High-dimensional two-sample location tests

To motivate our proposed test statistic, we briefly review some classical methods for Behrens-Fisher problem. For the univariate Behrens-Fisher problem, Fisher (1935; 1939) considered $\tau = (\bar{x}_1 - \bar{x}_2)/\sqrt{s_1^2/n_1 + s_2^2/n_2}$, where $\bar{x}_i$ and $s_i$ are the $i$-th sample mean and standard deviation, respectively. For the multivariate Behrens-Fisher problem, James (1954), Yao (1965) and Johansen (1980) proposed their test procedures based on the following test statistic:

$$Q_1 = (\bar{X}_1 - \bar{X}_2)^T \left( \frac{S_1}{n_1} + \frac{S_2}{n_2} \right)^{-1} (\bar{X}_1 - \bar{X}_2),$$

where $\bar{X}_i$ and $S_i$ are the $i$-th sample mean and covariance matrices, respectively. When the dimension $p > n$, the sample covariance matrices $S_1$ and $S_2$ are not positive definite and accordingly their test procedures are no longer feasible. Zhang and Xu (2009) simply removed the term $\left( \frac{S_1}{n_1} + \frac{S_2}{n_2} \right)^{-1}$ in $Q_1$. Consequently, their test statistic is essentially of $L^2$-norm and thus not scale-invariant.

Although having the inverse of the sample covariance matrix is no longer beneficial when $p/n \to c > 0$ (Bai and Saranadasa 1996), the information provided by the sample variances should still be useful. The following statistic is then motivated:

$$Q_2 = (\bar{X}_1 - \bar{X}_2)^T \left( \frac{D_1}{n_1} + \frac{D_2}{n_2} \right)^{-1} (\bar{X}_1 - \bar{X}_2),$$

where $D_i$ is the diagonal matrix of $S_i$. $Q_2$ can be written as:

$$Q_2 = \sum_{k=1}^{p} \frac{(\hat{\mu}_{1k} - \hat{\mu}_{2k})^2}{\hat{\sigma}_{1k}^2/n_1 + \hat{\sigma}_{2k}^2/n_2},$$

where $\hat{\mu}_{ik}$ and $\hat{\sigma}_{ik}$ are the sample mean and standard deviation of the $k$-th variable for the $i$-th sample, respectively. Thus, $Q_2$ is simply a sum of the $p$ squared univariate Fisher’s test statistics. Note that Srivastava et al. (2013) used standardized $Q_2$ as the test statistic: $\hat{q}_n = p^{-1/2}(Q_2 - p)$. If $n = O(p^\delta)$, $\delta > 0.5$ and under the normality assumption, this statistic asymptotically has mean zero through the asymptotic equivalence: $\hat{q}_n \to \tilde{q}_n$ in probability where

$$\tilde{q}_n = p^{-1/2} \left( \sum_{k=1}^{p} \frac{(\hat{\mu}_{1k} - \hat{\mu}_{2k})^2}{\hat{\sigma}_{1k}^2/n_1 + \hat{\sigma}_{2k}^2/n_2 - p} \right),$$
with zero mean.

Srivastava et al. (2013) claimed in their setting $Q_2$ is almost identical to the one proposed by Chen and Qin (2010). But when $p$ is of the order of $n^2$ or higher, this is not true any more and $Q_2$ will have no well-defined limit under the null. Actually, when $p$ has the order of $n^2$ or higher, the above equivalence between $\hat{q}_n$ and $\tilde{q}_n$ as shown below in a simple scenario. Consider that $n_1 = n_2$ and $\Sigma_1 = \Sigma_2 = \mathbf{I}_p$, we have the following result about $\hat{q}_n$.

**Proposition 1** Assume that Conditions (C1)-(C5) hold. When $p = n^{2+\alpha}$, $0 < \alpha < 1$, then, under $H_0$, $P\left(\frac{\hat{q}_n}{\sqrt{\text{var}(\hat{q}_n)}} > z_\alpha\right) \to 1$, where $z_\alpha$ is the upper $\alpha$ quantile of $N(0, 1)$.

The justification almost exactly follows the same arguments used to prove Theorem 1 in the appendix. We can show that $\{\hat{q}_n - E(\hat{q}_n)\}/\sqrt{\text{var}(\hat{q}_n)} \xrightarrow{d} N(0, 1)$, with $E(\hat{q}_n) = 4p^{1/2}n^{-1} + p^{-1/2}n^{-1} \sum_{k=1}^p (\kappa_{1k} - \kappa_{2k})^2 + o(1)$ and $\text{var}(\hat{q}_n) = 2 + o(1)$, where $\kappa_{ik} = E(X_{ijk}^2 - \mu_{ik})^2$. Thus, $E(\hat{q}_n)/\sqrt{\text{var}(\hat{q}_n)} = O(n^{\alpha/2})$ even in the normal case with $\sum_{k=1}^p (\kappa_{1k} - \kappa_{2k})^2 = 0$.

This result shows that, even under the simple case, the size of the Srivastava et al.’s (2013) test is totally distorted when the dimension gets higher. They also proposed a ratio consistent estimator of $\text{var}(\hat{q}_n)$. However, when $p = O(n^{2+\alpha})$, $\alpha > 0$, their estimator $\text{var}(\hat{q}_n)$ will not be ratio-consistent anymore. Further, they proposed a correction term $c_{p,n}$ to adjust the empirical size. In finite sample cases we conduct later, this size-adjusted value $c_{p,n}$ makes the variance estimator always larger than the real asymptotic variance of $\hat{q}_n$; see Figure 5.2.

There are two resources to create this bias-term. By dealing with them, we can define a test that can be asymptotically unbiased. The details are as follows. First, similarly as Bai and Saranadasa (1996), the term $\sum_{j=1}^{n_1} X_{ij}^2$ in $(\hat{\mu}_{1k} - \hat{\mu}_{2k})^2$ may impose demands on the dimensionality, where $\mathbf{X}_{ij} = (X_{ij1}, \ldots, X_{ijp})^T$. In other words, $Q_2$ is not possible to handle larger dimension even when a bias correction is made. To attack this challenge, motivated by Chen and Qin (2010),
after removing the terms like $\sum_{j=1}^{n_i} X_{ijk}^2$ in $(\hat{\mu}_{1k} - \hat{\mu}_{2k})^2$, we define a statistic as

$$Q_3 = \sum_{k=1}^{p} \frac{1}{\sigma_{1k}^2/n_1 + \sigma_{2k}^2/n_2} \left\{ \frac{1}{n_1(n_1-1)} \sum_{i \neq j} X_{1ik}X_{1jk} + \frac{1}{n_2(n_2-1)} \sum_{i \neq j} X_{2ik}X_{2jk} \right\} - \frac{2}{n_1n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} X_{1ik}X_{2jk} \equiv \sum_{k=1}^{p} A_k \frac{A_k}{\sigma_{1k}^2/n_1 + \gamma \sigma_{2k}^2/n_2}.$$  

Note that $\sum_{k=1}^{p} A_k = W_n$ is the test statistic proposed by Chen and Qin (2010), which is with no scale-invariance property, while $Q_3$ scales each $A_k$ in $W_n$ by its corresponding variance estimator. The following modification of $Q_3$ can be as an initial test statistic:

$$T_n = \frac{1}{n_1} Q_3 = \sum_{k=1}^{p} A_k \frac{A_k}{\sigma_{1k}^2 + \gamma \sigma_{2k}^2}.$$  

where $\gamma = n_1/n_2$. It is called an initial test statistic because we will see that $T_n$ cannot directly be used as the test for the hypotheses we want to check, but a standardized version should be used. Although $E(A_k) = 0$, the expectation of $T_n$ is not exactly zero since $A_k$ is not independent of $\sigma_{1k}^2 + \gamma \sigma_{2k}^2$. To this end, we make an analysis about $T_n$ such that we can deal with the second resource of the bias-term.

Both the tests statistics proposed by Bai and Saranadasa (1996) and Chen and Qin (2010) are invariant under the orthogonal transformation, say, $X_{ij} \rightarrow PX_{ij}$ where $P$ is an orthogonal matrix. In contrast, $T_n$ is not invariant under this orthogonal transformation but it is invariant under location shifts and the group of scalar transformations, say, $X_{ij} \rightarrow DX_{ij} + c$ for $i = 1, 2, j = 1, \ldots, n_i$, where $c$ is a constant vector, $D = \text{diag}(d_1, \ldots, d_p)$ and $d_1, \ldots, d_p$ are non-zero constants.

To establish the asymptotic normality of $T_n$ under $H_0$, we need the following conditions. Assume, like Bai and Saranadasa (1996) and Chen and Qin (2010) did, $X_{ij}$’s come from the following multivariate model:

$$X_{ij} = \Gamma_i z_{ij} + \mu_i \quad \text{for} \quad j = 1, \ldots, n_i, \ i = 1, 2, \quad (2.1)$$

where each $\Gamma_i$ is a $p \times m$ matrix for some $m \geq p$ such that $\Gamma_i^T \Gamma_i = \Sigma_i$, and $\{z_{ij}\}_{j=1}^{n_i}$ are $m$-variate independent and identically distributed (i.i.d.) random
vectors such that

\[ E(z_i) = 0, \quad \text{var}(z_i) = I_n, \quad E(z_j^2) = 3 + \Delta_i, \quad E(z_j^2) = m_{ij} \in (0, \infty), \]

\[ E(z_{ik1}^{\alpha_1}z_{ik2}^{\alpha_2} \cdots z_{ikq}^{\alpha_q}) = E(z_{ik1}^{\alpha_1})E(z_{ik2}^{\alpha_2}) \cdots E(z_{ikq}^{\alpha_q}), \]

for a positive integer \( q \) such that \( \sum_{k=1}^q \alpha_k \leq 8 \) and \( k_1 \neq k_2 \cdots \neq k_q \). The data structure (2.2) generates a rich collection of \( X_i \) from \( z_i \) with a given covariance. Additionally, we need the following conditions: as \( n, p \to \infty \)

(C1) \( n_1/(n_1 + n_2) \to \lambda \in (0, 1) \).

(C2) \( \text{tr} (\Lambda \Sigma_1 \Lambda^2 \Sigma_2 \Lambda^2 \Sigma_h \Lambda) = o(\text{tr}^2(\{ (\Lambda \Sigma_1 \Lambda + \Lambda \Sigma_2 \Lambda)^2 \})) \) for \( i, j, l, h = 1 \) or \( 2 \) where \( \Lambda = \text{diag}\{ (\sigma_{11}^2 + \gamma \sigma_{21}^2)^{-1/2}, \cdots, (\sigma_{ip}^2 + \gamma \sigma_{2p}^2)^{-1/2} \} \).

(C3) \( \frac{\text{var}(T_n)}{n} \to 0; \)

(C4) Define \( \Pi_{1i} = E(\Lambda(X_{ij} - \mu_i) (\Lambda(X_{ij} - \mu_i))^T), \Pi_{2i} = E(\Lambda(X_{ij} - \mu_i) (X_{ij} - \mu_i)^T \Lambda)^3, i = 1, 2. \) We assume that \( n_i^{-4} \text{tr}(\Pi_{1i}^2) = o(\text{var}(T_n)) \) and \( n_i^{-4} \text{tr}(\Lambda \Sigma_i \Lambda \Pi_{2i}) = o(\text{var}(T_n)) \) for \( i = 1, 2. \)

(C5) \( (\mu_1 - \mu_2)^T \Lambda^2 \Sigma_1 \Lambda^2 (\mu_1 - \mu_2) = o(n^{-1} \text{tr}[((\Lambda \Sigma_1 \Lambda + \Lambda \Sigma_2 \Lambda)^2)]), \) for \( i = 1, 2. \)

And \( (\mu_1 - \mu_2)^T \Lambda^2 (\mu_1 - \mu_2)^2 = o(n^{-1} \text{tr}[((\Lambda \Sigma_1 \Lambda + \Lambda \Sigma_2 \Lambda)^2)]). \)

**Remark 1** Conditions (C1) and (C5) are similar to conditions (3.3) and (3.4) in Chen and Qin (2010). To appreciate Condition (C2), consider the simple case \( \sigma_{ik} = \sigma_{ji} \), \( i, j = 1, 2, k, l = 1, \cdots, p. \) Condition (C2) then becomes \( \text{tr}(\Sigma_1 \Sigma_2 \Sigma_1 \Sigma_2) = o(\text{tr}[((\Sigma_1 + \Sigma_2)^2)]) \) for \( i, j, l, h = 1 \) or \( 2 \), which is the same as condition (3.6) in Chen and Qin (2010). Unlike Chen and Qin (2010), note that we restrict the relationship between the dimension \( p \) and sample size \( n \) in Condition (C3) due to the use of the plug-in variances. Such a “step backward” is just the price we need to pay for the scalar-transformation-invariance. Consider the simple case \( \Sigma_1 = \Sigma_2 \) with bounded eigenvalues which leads to \( \text{var}(T_n) = O(pn^{-2}) \). So, we can allow \( p = o(n^3) \), which is much more relaxed than that required in Srivastava and Du (2008). Actually, under assumptions (i)-(iii) in Srivastava and Du (2008), the test statistic \( T_n \) is not biased, i.e. \( \mu_n = o(\sqrt{\text{var}(T_n)}) \) under \( H_0. \) However, when the dimension \( p \) gets larger, such as \( p = O(n^2) \), \( \Sigma_1 = \Sigma_2 \) with bounded eigenvalues, we can obtain that \( \mu_n \sim \sqrt{\text{var}(T_n)} \). Condition (C4) is a technical condition and is verified in some specific cases in
Intuitively speaking, if all the variables are positively correlated, \( E((X_{ijk} - \mu_k)^3(X_{ijl} - \mu_l)) \) will be dominated by \( \sigma^2_{2k} E((X_{ijk} - \mu_k)(X_{ijl} - \mu_l)) \) and then \( \text{tr}(\Pi^2_i) \) is dominated by \( \text{tr}((A\Sigma_i A)^2) \). Similarly, \( \text{tr}(A\Sigma_i A\Pi_{2i}) \) is also dominated by \( \text{tr}((A\Sigma_i A)^2) \). Obviously, Condition (C4) will hold in this special case.

The following theorem establishes the asymptotic null distribution of \( T_n \).

**Theorem 1** Under Conditions (C1)–(C5), \( \{T_n - E(T_n)\}/\sqrt{\text{var}(T_n)} \xrightarrow{d} \mathcal{N}(0, 1) \), as \( p, n \to \infty \).

In order to formulate a testing procedure based on Theorem 1, both \( E(T_n) \) and \( \text{var}(T_n) \) under \( H_0 \) need to be estimated. On the surface, it is difficult to obtain a close-form of \( E(T_n) \) and \( \text{var}(T_n) \). Fortunately, as presented in the Appendix, under Conditions (C1)-(C5) stated above, we can show that

\[
E(T_n) = ||A(\mu_1 - \mu_2)||^2 + \sum_{k=1}^{p} \left\{ \frac{2\sigma_{1k}^4}{n_1(n_1 - 1)(\sigma_{1k}^2 + \gamma \sigma_{2k}^2)^2} + \frac{2\gamma \sigma_{2k}^4}{n_2(n_2 - 1)(\sigma_{1k}^2 + \gamma \sigma_{2k}^2)^2} + \frac{2}{n_1^2} \left( \frac{\sigma_{1k}^2 + \gamma \sigma_{2k}^2}{\sigma_{1k}^2 + \gamma \sigma_{2k}^2} \right)^2 \frac{\kappa_{1k}^2}{\sigma_{1k}^2 + \gamma \sigma_{2k}^2} \right\} + \sum_{k=1}^{p} \left\{ \frac{2\gamma^2 \kappa_{2k}^2}{n_2^2} \left( \frac{\mu_{2k} - \mu_{1k}}{\sigma_{1k}^2 + \gamma \sigma_{2k}^2} \right)^2 \right\} + o(\sqrt{\text{var}(T_n)})
\]

\[
\text{var}(T_n) = \left\{ \frac{2}{n_1(n_1 - 1)} \text{tr}((A\Sigma_1 A)^2) + \frac{2}{n_2(n_2 - 1)} \text{tr}((A\Sigma_2 A)^2) \right\} + \frac{4}{n_1 n_2} \text{tr}(A\Sigma_1 A^2 \Sigma_2 A) \right\} (1 + o(1)),
\]

and \( \kappa_{ik} = E(X_{ijk} - \mu_{ik})^3 \), \( \nu_{ik} = E(X_{ijk} - \mu_{ik})^4 \), \( i = 1, 2 \). Although the presentation is very complicated, the bias-term is estimable and then correctable.

It is required to obtain the sample variance \( \hat{\sigma}_{2k}^2 \) and skewness \( \hat{\kappa}_{ik} = n_i^{-1} \sum_{j=1}^{n_i} (X_{ijk} - \hat{\mu}_{ik})^3 \). The corresponding estimator of \( E(T_n) \) under \( H_0 \) is obtained by plugging-in
relevant estimators, i.e.,
\[
\hat{\mu}_n = E(T_n) = \sum_{k=1}^{p} \left\{ \frac{2\hat{\sigma}_{1k}^4}{n_1(n_1-1)(\hat{\sigma}_{1k}^2 + \gamma \hat{\sigma}_{2k}^2)^2} + \frac{2\gamma \hat{\sigma}_{2k}^4}{n_2(n_2-1)(\hat{\sigma}_{1k}^2 + \gamma \hat{\sigma}_{2k}^2)^2} + \frac{2}{n_1^2 (\hat{\sigma}_{1k}^2 + \gamma \hat{\sigma}_{2k}^2)^3} + \frac{2\gamma^2}{n_2^2 (\hat{\sigma}_{1k}^2 + \gamma \hat{\sigma}_{2k}^2)^3} - \frac{4\gamma}{n_1n_2 (\hat{\sigma}_{1k}^2 + \gamma \hat{\sigma}_{2k}^2)^3} \right\}.
\]

Next, we estimate the trace terms \(\text{tr}(\Lambda \Sigma_1 \Lambda^2 \Sigma_2 \Lambda)\), \(i = 1, 2\) and \(\text{tr}(\Lambda \Sigma_1 \Lambda^2 \Sigma_2 \Lambda)\) in \(\text{var}(T_n)\). Chen and Qin (2010) proposed effective estimators for these terms by applying the “leave-one-out” idea. In a similar spirit, we may use
\[
\text{tr}(\Lambda \Sigma_1 \Lambda^2) = \frac{1}{n_1(n_1-1)} \sum_{i \neq j}^{n_1} \sum_{l=1}^{p} \left( \frac{X_{iil} - \hat{\mu}_{il(i,j)}}{\hat{\sigma}_{1l}^2 + \gamma \hat{\sigma}_{2l}^2} \right)^2,
\]
where \(\hat{\mu}_{il(i,j)}\) is the \(i\)-the sample mean after excluding \(X_{ijl}\) and \(X_{ikl}\). However, similar to the expectation of \(T_n\), since \(X_{ijk}\) is not independent of \(\hat{\sigma}_{2k}\), using such an estimator would yield bias-terms which are not negligible when \(n = O(p^{1/2})\). It seems more complex to calculate those terms numerically than the expectation of \(T_n\) in a high-dimensional setting. As such, we suggest a remedy which is motivated by Chen et al. (2010) as follows:
\[
\text{tr}(\Lambda \Sigma_1 \Lambda^2) = \frac{1}{2P^4 n_s} \sum_{s=1}^{s} (X_{s_{i1}} - X_{s_{i2}})^T D_{s_{i1},i2,i3,i4}^{-1} (X_{s_{i1}} - X_{s_{i4}}) \times (X_{s_{i3}} - X_{s_{i4}})^T D_{s_{i1},i2,i3,i4}^{-1} (X_{s_{i3}} - X_{s_{i4}}), s = 1, 2, \]
\[
\text{tr}(\Lambda \Sigma_1 \Lambda^2 \Sigma_2 \Lambda) = \frac{1}{4P^2 n_s^2} \sum_{i \neq j}^{n_1} \sum_{l=1}^{n_2} \left( \frac{X_{i1l} - X_{i2l}}{\hat{\sigma}_{1l}^2 + \gamma \hat{\sigma}_{2l}^2} \right)^2 D_{s_{i1},i2,i3,i4}^{-1} (X_{s_{i3}} - X_{s_{i4}})^T (X_{s_{i1}} - X_{s_{i4}})^2,
\]
where
\[
D_{1(i_1,i_2,i_3,i_4)} = \text{diag}(\hat{\sigma}_{11}^2 + \gamma \hat{\sigma}_{21}^2, \ldots, \hat{\sigma}_{1p(i_1,i_2,i_3,i_4)} + \gamma \hat{\sigma}_{2p(i_1,i_2,i_3,i_4)}),
\]
\[
D_{2(i_1,i_2,i_3,i_4)} = \text{diag}(\hat{\sigma}_{11}^2 + \gamma \hat{\sigma}_{21}^2, \ldots, \hat{\sigma}_{1p(i_1,i_2,i_3,i_4)} + \gamma \hat{\sigma}_{2p(i_1,i_2,i_3,i_4)}),
\]
\[
D_{s(i_1,i_2,i_3,i_4)} = \text{diag}(\hat{\sigma}_{11}^2 + \gamma \hat{\sigma}_{21}^2, \ldots, \hat{\sigma}_{1p(i_1,i_2,i_3,i_4)} + \gamma \hat{\sigma}_{2p(i_1,i_2,i_3,i_4)}),
\]
and \(\hat{\sigma}_{sk(i_1,\ldots,i_l)}^2\) is the \(s\)-th sample variance after excluding \(X_{s_{i1}}, j = 1, \ldots, l, s = 1, 2, l = 2, 4, k = 1, \ldots, p\). Through this article, we use \(\sum\) to denote summations over distinct indexes. For example, in \(\text{tr}(\Lambda \Sigma_1 \Lambda^2)\), the summation is over the set \(\{i_1 \neq i_2 \neq i_3 \neq i_4\}\), for all \(i_1, i_2, i_3, i_4 \in \{1, \ldots, n_1\}\) and \(P^m_n = n!/(n-m)!\).
Remark 2. The estimators of $\text{tr}(\Sigma_i^2)$ and $\text{tr}(\Sigma_1\Sigma_2)$ proposed by Chen and Qin (2010) are not translation-invariant. According to the proof of Theorem 2 in Chen and Qin (2010), $E(\text{tr}(\Sigma_i^2)_{CQ}) = \text{tr}(\Sigma_i^2) + \mu_\Sigma \mu/(n-2)$. Thus, it is easy to verify that the estimator $\text{tr}(\Sigma_i^2)_{CQ}$ will be different when $X_{ij}$ is transformed to $X_{ij} + \theta$. From an asymptotic viewpoint, the ratio consistency of $\text{tr}(\Sigma_i^2)_{CQ}$ relies on the condition that $\mu_\Sigma \mu/(n-2) = o(\text{tr}(\Sigma_i^2))$. This assumption is fairly restrictive because the expectation of $X_{ij}$ could be in any scale in two-sample location testing applications. As such, the estimator of $\text{tr}(R^2)$ proposed by Park and Ayyala (2013) is not translation-invariant either.

Remark 3. Both the estimators proposed by Srivastava and Du (2008) and Srivastava et al. (2013) are for $\text{tr}(R^2)$ by the plug-in sample variances. Thus, there would be also the corresponding bias-terms in their variance estimators when $p$ is large. Moreover, in large $p$ cases, the variance estimator will be inevitably larger than the real variance of their proposed test statistic. See more information in Section 3.

The next proposition establishes the ratio-consistency of the two estimators of $E(T_n)$ and $\text{var}(T_n)$.

Proposition 2 Suppose conditions (C1)–(C4) hold. Then, we have $\hat{\mu}_n = E(T_n) + o_p(\sqrt{\text{var}(T_n)})$, and

$$\frac{\text{tr}(\Sigma_i^2)}{\text{tr}(\Sigma_i^2)} \xrightarrow{p} 1, \quad i = 1, 2 \quad \text{and} \quad \frac{\text{tr}(\Sigma_1\Sigma_2)}{\text{tr}(\Sigma_1\Sigma_2)} \xrightarrow{p} 1.$$ 

As a consequence, a ratio-consistent estimator of $\text{var}(T_n)$ under $H_0$ is

$$\hat{\sigma}_n^2 = \text{var}(T_n) = \left\{ \frac{2}{n_1(n_1-1)} \text{tr}(\Sigma_1\Sigma_2) + \frac{2}{n_2(n_2-1)} \text{tr}(\Sigma_2\Sigma_2) + \frac{4}{n_1n_2} \text{tr}(\Sigma_1\Sigma_2) \right\}.$$ 

Now we are in the position to define the final test statistic: $(T_n - \hat{\mu}_n)/\hat{\sigma}_n$. Its asymptotic normality under the null hypothesis is stated as follows.

Corollary 1 Under Conditions (C1)–(C4) and $H_0$, $(T_n - \hat{\mu}_n)/\hat{\sigma}_n \xrightarrow{L} N(0, 1)$. From this result, we successfully remove the bias-term. This result suggests rejecting $H_0$ with $\alpha$ level of significance if $(T_n - \hat{\mu}_n)/\hat{\sigma}_n > z_\alpha$. The ratio-consistent
estimator of \( \text{var}(T_n) \) seems complex but computes fast. For example, it takes 5s per iteration in FORTRAN using Inter Core 2.2 MHz CPU for a \( n_1 = n_2 = 30, p = 1000 \) case for each \( \text{tr}((\Lambda \Sigma_1 \Lambda)^2), \text{tr}(\Lambda \Sigma_1 \Lambda^2 \Sigma_2 \Lambda) \) and the entire procedure is generally completed in less than 20s.

Next, we discuss the asymptotic power properties of the proposed test. According to Theorem 1, the power under the local alternative (C5) is

\[
\beta_{BF}(||\Lambda(\mu_1 - \mu_2)||) = \Phi \left( -z_\alpha + \frac{\tilde{\mu}_n}{\tilde{\sigma}_n} \right),
\]

where

\[
\tilde{\mu}_n = ||\Lambda(\mu_1 - \mu_2)||^2 + \sum_{k=1}^{p} \frac{2 n_1 \kappa_{1k}(\mu_{2k} - \mu_{1k}) + 2 n_2 \kappa_{2k}(\mu_{1k} - \mu_{2k})}{(\sigma_{1k}^2 + \gamma \sigma_{2k}^2)^2}
+ \sum_{k=1}^{p} \frac{(\mu_{1k} - \mu_{2k})^2}{(\sigma_{1k}^2 + \gamma \sigma_{2k}^2)^3} \left( \frac{1}{n_1} \nu_{1k} + \frac{4}{n_1(n_1 - 1)} \sigma_{1k}^4 + \frac{\gamma}{n_2} \nu_{2k} + \frac{4 \gamma^2}{n_2(n_2 - 1)} \sigma_{2k}^4 \right),
\]

\[
\tilde{\sigma}_n^2 = \frac{2}{n_1(n_1 - 1)} \text{tr}((\Lambda \Sigma_1 \Lambda)^2) + \frac{2}{n_2(n_2 - 1)} \text{tr}((\Lambda \Sigma_2 \Lambda)^2) + \frac{4}{n_1 n_2} \text{tr}(\Lambda \Sigma_1 \Lambda^2 \Sigma_2 \Lambda),
\]

and \( \Phi(\cdot) \) is the standard normal distribution function. In contrast, Chen and Qin (2010) showed that the power of their proposed test is

\[
\beta_{CQ}(||\mu_1 - \mu_2||) = \Phi \left( -z_\alpha + \frac{n \lambda(1 - \lambda)||\mu_1 - \mu_2||^2}{\sqrt{2\text{tr}(\tilde{\Sigma}^2)}} \right),
\]

where \( \tilde{\Sigma} = (1 - \lambda) \Sigma_1 + \lambda \Sigma_2. \)

It is worth pointing out that theoretically comparing the proposed test with Chen and Qin’s (2010) test under general settings turns out to be difficult. In order to get a rough picture of the asymptotic power comparison between these two tests, we simply assume that \( \kappa_{ik} = 0, k = 1, \cdots, p, i = 1, 2. \) It is then easy to show that

\[
\sum_{k=1}^{p} \frac{(\mu_{1k} - \mu_{2k})^2}{(\sigma_{1k}^2 + \gamma \sigma_{2k}^2)^3} \left( \frac{1}{n_1} \nu_{1k} + \frac{4}{n_1(n_1 - 1)} \sigma_{1k}^4 + \frac{\gamma}{n_2} \nu_{2k} + \frac{4 \gamma^2}{n_2(n_2 - 1)} \sigma_{2k}^4 \right) = o(||\Lambda(\mu_1 - \mu_2)||^2).
\]

Now, the power of the proposed test becomes

\[
\beta_{BF}(||\Lambda(\mu_1 - \mu_2)||) = \Phi \left( -z_\alpha + \frac{n \lambda(1 - \lambda)||\Lambda(\mu_1 - \mu_2)||^2}{\sqrt{2\text{tr}(\Lambda \Sigma \Lambda)}} \right) + o(1).
\]
We consider the following representative cases:

(i) $\mu_{1k} - \mu_{2k} = \delta$, $k = 1, \cdots, p$. In this case,

$$\beta_{BF}(||\mathbf{A}(\mathbf{\mu}_1 - \mathbf{\mu}_2)||) = \Phi \left( -z_{\alpha} + \frac{n\lambda(1 - \lambda)\delta^2\text{tr}(\mathbf{A}^2)}{\sqrt{2\text{tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{A})^2}} \right) + o(1),$$

$$\beta_{CQ}(||\mathbf{\mu}_1 - \mathbf{\mu}_2||) = \Phi \left( -z_{\alpha} + \frac{np\lambda(1 - \lambda)\delta^2}{\sqrt{2\text{tr}(\mathbf{\Sigma}^2)}} \right).$$

By the Cauchy inequality, we can obtain that

$$p^2\text{tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{A})^2 \leq \text{tr}(\mathbf{\Sigma}^2)\text{tr}(\mathbf{A}^2).$$

As a consequence,

$$\beta_{CQ}(||\mathbf{\mu}_1 - \mathbf{\mu}_2||) \leq \Phi \left( -z_{\alpha} + \frac{n\lambda(1 - \lambda)\delta^2\text{tr}(\mathbf{A}^2)}{\sqrt{2\text{tr}(\mathbf{A}\mathbf{\Sigma}\mathbf{A})^2}} \right) \leq \beta_{BF}(||\mathbf{A}(\mathbf{\mu}_1 - \mathbf{\mu}_2)||).$$

When the variances of all the components are equal, the two tests are equivalently powerful from the asymptotic viewpoint. Otherwise, the proposed test would be preferable in terms of asymptotic power under local alternatives.

(ii) $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ and they are both diagonal matrices. The variances of the first half and the other half of components are $\zeta_1^2$ and $\zeta_2^2$, respectively. Assume $\mu_{1k} - \mu_{2k} = \delta$, $k = 1, \cdots, \lfloor \frac{p}{2} \rfloor$. In this setting,

$$\beta_{BF}(||\mathbf{A}(\mathbf{\mu}_1 - \mathbf{\mu}_2)||) = \Phi \left( -z_{\alpha} + \frac{n\sqrt{p}\lambda(1 - \lambda)\delta^2}{2\sqrt{2}\zeta_1^2} \right) + o(1),$$

$$\beta_{CQ}(||\mathbf{\mu}_1 - \mathbf{\mu}_2||) = \Phi \left( -z_{\alpha} + \frac{n\sqrt{p}\lambda(1 - \lambda)\delta^2}{2\sqrt{\zeta_1^2 + \zeta_2^2}} \right).$$

Thus, the asymptotic relative efficiency (ARE) of the proposed test with respect to the CQ test would be $\sqrt{\zeta_1^4 + \zeta_2^4}/(\sqrt{2}\zeta_1^2)$. It is clear that the proposed test is more powerful than CQ if $\zeta_1^2 < \zeta_2^2$ and vice versa. This ARE has a positive lower bound of $1/\sqrt{2}$ when $\zeta_1^2 > > \zeta_2^2$, whereas it can be arbitrarily large if $\zeta_1^2/\zeta_2^2$ is close to zero. This property shows the necessity for a test with the scale-invariance property.
3. Numerical Studies

Throughout this section, for each experiment we run 1,000 replications and then the standard error of size or power entries is bounded by 0.016.

3.1. The bias-term

Here we report a simulation study designed to evaluate the bias-term of $\hat{q}_n$ proposed by Srivastava et al. (2013) and $T_n - \hat{\mu}_n$ proposed by us and the quality of the corresponding variance estimator under the null hypothesis. Here we only consider the equal covariance matrices assumption $\Sigma_1 = \Sigma_2 = \Sigma = (a_{ij})$, $a_{ij} = 0.5|\!|i-j|\!|$ and $X_{ij}$ are all independent $p$-dimensional multivariate normal random vectors. We summarize simulation results by using the mean-standard deviation-ratio $E(T)/\sqrt{\text{var}(T)}$ and the variance estimator ratio $\text{var}(\hat{T})/\text{var}(T)$, where $T$ denotes either $\hat{q}_n$ or $T_n - \hat{\mu}_n$. Since the explicit form of $E(T)$ and $\text{var}(T)$ is difficult to calculate, we estimate them by simulation. Here we consider two sample sizes $n_1 = n_2 = 15, 30$ and eight dimensions $p = 30, 60, 100, 200, 300, 400, 800, 1000$.

Figure 5.1 reports the mean-standard deviation-ratio of the test statistics proposed by Srivastava et al. (2013) and us. From Figure 5.1, we observe that the bias-term in $\hat{q}_n$ apparently exists, especially when the dimension is high. There is also a little bias for our BF test when $n_1 = n_2 = 15, p = 1000$. It is not strange because the dimension is comparable to the cubic of the sample size and the condition (C3) will not hold. However, in the other cases, the mean-standard deviation-ratio of our test statistic $T_n - \hat{\mu}_n$ is approximately zero, which shows that our bias correction procedure is very effective. Figure 5.2 reports the simulation results of the variance estimator ratio. We also observe that the variance estimator proposed by Srivastava et al. (2013) is apparently larger than the real variance of their proposed test statistic. There are two reasons for this. Firstly, there is also a bias-term in the estimator $\text{var}(\hat{q}_n)$ when the dimension is high. Secondly, the correction term $c_{p,n}$ is always larger then one. In contrast, the variance estimator ratio of our test statistic is approximately one, which shows that our variance estimator is also effective, even when the dimension is very high. Because both the two ratios are higher than the acceptable level, the empirical sizes of the test statistics proposed by Srivastava et al. (2013) will also deviate from the significance level. See the next subsections for more information.

[Insert Figures 5.1-5.2 around here]
3.2. Empirical sizes and power comparison

Here we report a simulation study designed to evaluate the performance of our proposed test (abbreviated as BF). To allow a meaningful comparison with the methods proposed by Bai and Saranadasa (1996) (abbreviated as BS), Srivastava and Du (2008) (abbreviated as SD), Chen and Qin (2010) (abbreviated as CQ), and Srivastava et al. (2013) (abbreviated as SKK), we firstly consider the unequal covariance matrices assumption. In this case, the assumption of common covariances in Bai and Saranadasa (1996) and Srivastava and Du (2008) does not hold. We consider the following moving average model as Chen and Qin (2010):

\[ X_{ijk} = \rho_{i1} Z_{ij} + \rho_{i2} Z_{i(j+1)} + \cdots + \rho_{iL_i} Z_{i(j+L_i-1)} + \mu_{ij} \]

for \( i = 1, 2, j = 1, \cdots, n_i \) and \( k = 1, \cdots, p \) where \( \{Z_{ijk}\} \) are, respectively, i.i.d. random variables. Consider two scenarios for the innovation \( \{Z_{ijk}\} \): (Scenario I) all the \( \{Z_{ijk}\} \) are from \( N(0,1) \); (Scenario II) the first half components of \( \{Z_{ijk}\}_{k=1}^{p} \) are from centralized Gamma(4,1) so that it has zero mean, and the rest half components are from \( N(0,1) \). The coefficients \( \{\rho_{il}\}_{l=1}^{L_i} \) are generated independently from \( U(2,3) \) and are kept fixed once generated through our simulations. The correlations among \( X_{ijk} \) and \( X_{ijl} \) are determined by \( |k-l| \) and \( L_i \). We choose \( L_1 = 1, \) and \( L_2 = 3 \) to generate different covariances of \( X_i \). For the alternative hypothesis, we fix \( \mu_1 = 0 \) and again choose \( \mu_2 \) in two scenarios: (Case A) one allocates all of the components of equal magnitude to be nonzero; (Case B) the other allocates randomly half of components of equal magnitude to be nonzero. To make the power comparable among the configurations of \( H_1 \), we set \( \eta := ||\mu_1 - \mu_2||^2 / \sqrt{\text{tr}(\Sigma_1^2) + \text{tr}(\Sigma_2^2)} = 0.1 \) throughout the simulation.

We firstly examine the empirical sizes of tests. For simplicity, the sample sizes \( n_1 = n_2 \) are respectively chosen to be 15, 20 and 30. In the supplemental file, we also present some simulation results with unbalanced designs, say \( n_1 \neq n_2 \), for which the comparison conclusion revealed below still holds (c.f., Figure S1.1). We choose three dimensions for each sample size \( p = 225, 400, 900 \). Figure 5.3 below reports the empirical sizes of those five tests. Clearly, when \( \Sigma_1 \neq \Sigma_2 \), the sizes of BS and SD are much smaller than the significance level, 0.05, especially when \( p \) is ultra-high. In comparison, both CQ and BF have reasonable sizes in most of the cases. The performance of SKK is not very encouraging. In many cases, the sizes are larger than the significance level, whereas in some other cases
where \( n_1 = n_2 = 15, p = 900 \), the sizes of SKK seem somewhat conservative. We note that there are considerable biases in the estimation of the sample correlation matrix when \( n = O(p^{1/2}) \) and the tests should be very difficult to maintain the significance level when a bias-correction has not yet been made. From the above observations, we can see that for SKK, it is difficult to make a size-adjusted version to maintain the significance level. The consequence is that its high power would not be very meaningful.

We present the powers of the tests in Figure 5.4 below for a further comparison. Since the sizes of BS and SD are extremely distortional, they are not efficient in most of cases as we can expect. Under scenario (I), the variances of components are all equal and thus the powers of CQ test are slightly larger than the BF test. However, under scenario (II), BF outperforms CQ uniformly in all the cases by a quite large margin of power, which again concurs with the asymptotic comparison in Section 2. All these results together suggest that the newly proposed BF test is scale-invariant and quite efficient and robust in testing the equality of locations, and particularly useful when the variances of components are not equal or \( \Sigma_1 \neq \Sigma_2 \). Furthermore, SKK does not perform well. Although in some cases, the power is higher than the BF test its sizes are also larger when we look at Figure 5.3, and when its sizes are smaller than the significance level, the power is also lower. Thus, all the above results suggest that SKK in high-dimensional cases would need a bias correction.

4. A real-data example

In this section, we demonstrate the proposed methodology by applying it to a real dataset from a semi-conductor manufacturing process which is under consistent surveillance via collecting variables from sensors at many measurement points. The data set contains a total of 1567 vector observations. For each observation, there are originally 591 continuous measurements. A categorical label, indicating a single production is whether a conforming yield through house-line testing, is also provided in this dataset. The goal of the data analysis is mainly to model and monitor production quality based on those sensor measurements.
In the quality control of such a process, a critical step is comparing a new sample with some reference sample to check if the new observations are in-control. Then we would update the reference sample with the sample without significant difference. Thus, it requires carefully examination and powerful testing approach because once an inferior sample is regarded as a good one, the reference sample will be potentially (highly) affected. It is interesting to consider some high-dimensional tests by incorporating the information from all the sensors. Note that in many applications, it may not be always feasible to wait for the accumulation of sufficiently large calibration samples because users often want to test the process at the start-up stages in which only a few of reference samples are available. Our proposed method is applicable for the cases with rather small sample sizes as shown in Section 3, and thus appears to be particularly useful for this example.

This dataset contains null values varying in intensity depending on the individuals features. Since the fraction of missing values is trivial in this dataset, we simply use mean imputation. In addition, 117 constant features are removed from the analysis and totally 474 variables are involved. For illustration, we artificially assume that we have \( n \) observations categorized as the nonconforming and conforming respectively and apply BF, BS, CQ and SD tests. To get a broad picture of performance comparison, we consider a bootstrap type testing procedure. Two random samples of size \( n \) are drawn from 104 nonconforming observations and 1463 conforming observations without replacement. Then all four considered tests are applied to these two samples and the corresponding test results with a significance level 0.05 are recorded. Repeat this procedure 1,000 times and then the resulting powers are given in Table 5.1. The BF and SD tests deliver satisfactory testing results and their powers increase fast with larger sample sizes. Also, it seems that the proposed BF test performs slightly better than SD and SKK when the sample size is small \((n = 15)\) which is again consistent with our preceding theoretical and numerical analysis. In contrast, BS and CQ are ineffective for this example because they are not scale-invariant. In this dataset, different components have different physical readings and their scales differ much, yielding the extremely poor performances of BS and CQ.
5. Discussion

A natural concern is whether the newly proposed test in this paper can handle the ultra-high dimensional scenarios with larger $p$, say, at an exponential rate of $n$. Unfortunately, it is very difficult, if not impossible, when there is no sparse structure for all existing scale-invariant tests to correct bias-terms. Thus in general, it is still an open problem for whether we can define a test statistic that is (at least) asymptotically unbiased without sparsity assumption on the data structure. This paper offers the insight on the reasons why all existing scale-invariant tests cannot work on ultra-high dimensional paradigms, and the results in this paper also serve as a reminder for us to carefully take this challenge into consideration. In view of this, the standardized version with shrinkage estimation under sparse structure and other conditions may help; see Cai et al. (2014).

Acknowledgment The authors thank the editor, associate editor, and two anonymous referees for their many helpful comments that have resulted in significant improvements in the article.

References


James, G. S. (1954). Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown, Biometrika, 41, 19–43.


Institute of Statistics, Nankai University, Tianjin, 300071, China
E-mail: flnankai@126.com
Institute of Statistics, Nankai University, Tianjin, 300071, China
E-mail: nk.chlzou@gmail.com
Institute of Statistics, Nankai University, Tianjin, 300071, China
E-mail: zjwang@nankai.edu.cn
Department of Mathematics, Hong Kong Baptist University, Hong Kong, China
E-mail: lzhu@hkbu.edu.hk
Figure 5.1: The mean-standard deviation-ratio $E(T)/\sqrt{\text{var}(T)}$ of the test statistics proposed by Srivastava et al. (2013) (SKK) and us (BF).

Table 5.1: Empirical power comparisons at 5% significance for the sensor dataset

<table>
<thead>
<tr>
<th>n</th>
<th>BF</th>
<th>BS</th>
<th>CQ</th>
<th>SD</th>
<th>SKK</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>0.355</td>
<td>0.067</td>
<td>0.076</td>
<td>0.288</td>
<td>0.297</td>
</tr>
<tr>
<td>20</td>
<td>0.611</td>
<td>0.048</td>
<td>0.051</td>
<td>0.616</td>
<td>0.621</td>
</tr>
<tr>
<td>30</td>
<td>0.941</td>
<td>0.058</td>
<td>0.059</td>
<td>0.971</td>
<td>0.982</td>
</tr>
</tbody>
</table>
Figure 5.2: The variance estimator ratio \( \text{var}(\hat{T})/\text{var}(T) \) of the test statistics proposed by Srivastava et al. (2013) (SKK) and us (BF).
Figure 5.3: Empirical size comparisons at 5% significance when $\Sigma_1 \neq \Sigma_2$. 
Figure 5.4: Empirical power comparisons at 5% significance when $\Sigma_1 \neq \Sigma_2$. 