TECHNICAL REPORTS AND PREPRINTS

Number 13-02: April 2013

Test of Significance on High Dimensional Covariance Matrix Structures

Hengjian Cui\textsuperscript{1}, Shurong Zheng\textsuperscript{2} and Runze Li\textsuperscript{3}

\textsuperscript{1}Department of Statistics, Capital Normal University, China
\textsuperscript{2}KLAS and School of Mathematics and Statistics, Northeast Normal University, China
\textsuperscript{3}Department of Statistics and The Methodology Center, The Pennsylvania State University
Test of Significance on High Dimensional Covariance Matrix Structures

Hengjian Cui, Shurong Zheng and Runze Li
Capital Normal University, Northeast Normal University
& The Pennsylvania State University
April 28, 2013

Abstract

This paper is concerned with test of significance on high dimensional covariance structures. In the classical multivariate analysis, the likelihood ratio tests have been derived for testing various covariance structures including test of sphericity and test of independence under a multinormality assumption. In this paper, our primary interest is to study the asymptotic behaviors of such likelihood ratio tests under the setting in which the limit of the ratio of the dimension of data to the sample size lies between 0 and 1. Using the modern random matrix theory, we derive the limiting null distributions of the likelihood ratio tests without multinormality assumption. Our theoretical analysis indicates that the likelihood ratio tests diverge to infinity as the sample size increases, and therefore the traditional likelihood ratio theory becomes invalid. We further conduct Monte Carlo simulation study on the likelihood ratio tests. Our simulation results show that the limiting null distribution provides quite accurate approximation to the null distribution so that the critical values taken from the limiting null distribution keeps the Type I error rate very well. Our simulation results clearly show that using the critical values from the corresponding chi-square distributions, the traditional limiting null distribution of the likelihood ratio tests, cannot preserve Type
I error rate. Indeed the Type I error rate with the traditional critical values may be close to one under some settings. We further examine the power of the likelihood ratio tests using the newly derived null distribution. The simulation results show that the likelihood ratio tests are still very powerful with the new limiting null distribution. A real data example is used to illustrate the likelihood ratio tests.

Keywords: Covariance matrix structure; random matrix theory; test of equicorrelation; test of independence; test of sphericity.

*Hengjian Cui is Professor, Department of Statistics, Capital Normal University, China. Email: hjcui@bmu.edu.cn. His research was supported by National Natural Science Foundation of China (NNSFC) grants 11071022, 11028103 and 11231010. Shurong Zheng is Professor, KLAS and School of Mathematics and Statistics, Northeast Normal University, China. Email: zhengsr@neau.edu.cn. Her research was supported by NNSFC grant 11171058 and Program for New Century Excellent Talents in University, NECT-11-9616. Runze Li is Distinguished Professor, Department of Statistics and The Methodology Center, The Pennsylvania State University, University Park, PA 16802-2111. Email: rzli@psu.edu. His research was supported by National Institute on Drug Abuse (NIDA) grant P50-DA10075 and a NNSFC grant 11028103. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or NIDA.
1 Introduction

High dimensional data analysis has become increasingly important in various research fields. Various regularization methods have been proposed for variable selection and feature selection in regression analysis with high dimensional covariates (Tibshirani, 1996; Fan and Li, 2001; Efron, Hastie, Johnstone and Tibshirani, 2004). Fan and Li (2006) gives a brief review of regularization methods to deal with several challenges in analysis of high dimensional data analysis. This paper is concerned with test of hypotheses for high dimensional data. Tests of high-dimensional normal mean have proposed in the literature. When the dimension of data exceeds the sample size, the sample covariance matrix becomes singular, and therefore the corresponding Hotelling's $T^2$ become invalid. Dempster (1958) proposed an alternative to Hotelling $T^2$ for two sample normal mean problem. The research topic of testing high-dimensional normal mean has received more and more attentions in the recent literature. Bai and Saranadasa (1996) demonstrated the effect of dimensionality for test of two sample high-dimensional normal means. The inverse of the sample covariance matrix plays a key role in the construction of test for high-dimensional normal mean problems. Thus, spectral analysis of large covariance matrix becomes essential in test of high-dimensional normal mean problems. Bai (1999) provides a review of methodologies in spectral analysis of large dimensional random matrices. Chen and Qin (2010) proposed a new test for two sample mean problem with normality assumption. Zhong and Chen (2011) demonstrated some tests for regression coefficients with traditional critical value cannot preserve their Type I error rate.

Test of covariance structure is of great importance in multivariate data analysis. Various tests for covariance matrix have been developed in the classical multivariate analysis (Anderson, 2003). However, these tests become invalid when the dimension $p_n$ of data is large relative to the sample size $n$ (Ledoit and Wolf, 2002). Several alternatives to the classical tests of covariance structure have been developed in the literature (Srivastava, 2005; Birke and Dette, 2005; Schott, 2007). Bai, Jiang, Yao and Zheng (2009) proposed correction to the likelihood ratio tests for testing whether the covariance matrix equals to a given one such as the identity matrix when $y_n=p/n \rightarrow y \in (0,1)$. They also proposed correcting the
likelihood ratio test for two sample covariance problem by using the modern random matrix theory (Bai and Silverstein, 2004). With $y \in [0, 1)$, Wang, Yang, Miao and Cao (2012) studied the likelihood ratio test for the covariance matrix being the identity matrix. Wang, Cao and Miao (2013) further investigated its asymptotic power of the likelihood ratio test. When $p/n \to \infty$, Chen, Zhang and Zhong (2010) studied testing of sphericity for high dimensional covariance matrices, and Qiu and Chen (2012) further studied testing bandedness of covariance matrices. Li and Chen (2012) studied testing two sample covariance problems when $p/n_k \to \infty$, where $n_k, k = 1, 2$, stands for the sample size of each sample. In this paper we shall study tests of three commonly-used covariance structures under the setting $p/n \to y \in [0, 1)$.

In this paper, we focus on the asymptotic behavior of test of sphericity, test of independence and test of equicorrelation. Under a unified framework, we derive the limiting null distributions of the corresponding likelihood ratio tests in the classical multivariate analysis without multinormality assumption. Our theoretical analysis reveals a phenomenon that the limiting null distribution behaves very differently for $y = 0$, $y \in (0, 1)$ and $p/n \to 1$. We further conduct simulation study to examine the Type I error rate and the power. Our simulation results implies that the limiting null distribution approximates the null distribution quite well when the sample size is moderate. The corresponding tests with critical value obtained from the limiting null distribution are able to keep their Type I error rate very well. Our simulation results also empirically demonstrate the likelihood ratio tests with the corrected critical values are powerful under various alternatives.

The rest of this paper is organized as follows. In Section 2, we derive the limiting null distribution of the likelihood ratio test for sphericity, test of independence and test of equicorrelation for high-dimensional covariance matrices without normality assumption. In Section 3, we investigate the finite sample performance of the limiting null distributions and examine the Type I error rate and power of the tests with critical values obtained from the limiting null distribution. A real data example is also given to illustrate the proposed procedure. Technical proofs of the main theorem are given in the Appendix.
2 Some tests on covariance matrix structures

Suppose that \( \{X_1, X_2, \cdots, X_n\} \) is an independent and identically distributed random sample from a \( p \)-dimensional population \( X \) with mean \( E(X) = \mu \), and covariance matrix \( \text{Cov}(X) = \Sigma \). Of interest is to study the asymptotic behaviors of some classical likelihood ratio test statistics on covariance matrix structures when \( p/n \to y \in (0, 1) \) as \( n \to \infty \). That is, we are interested in situations in which the dimension \( p \) and the sample size \( n \) have the same magnitude. Henceforth, we suppress the subscript \( n \) in \( p_n \) for simplicity. Denote by \( \bar{X} \) and \( \hat{\Sigma}_n \) by the sample mean and the sample covariance matrix, respectively. That is,

\[
\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad \hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})',
\]

We shall focus on testing sphericity, uncorrelated structure and compound symmetry of the covariance matrix. The related techniques developed in this paper may be applicable for testing other covariance structures.

2.1 Test of Sphericity

In this section, we study test of sphericity, in which

\[
H_{10} : \Sigma = cI_p
\]

for some unknown positive constant \( c \). The alternative hypothesis \( H_{11} \) is that \( \Sigma \neq cI_p \) for any positive constant \( c \).

In classical multivariate analysis, it is assumed that \( X \sim \mathcal{N}(\mu, \Sigma) \), the multivariate normal distribution with mean \( \mu \) and covariance matrix \( \Sigma \). Under the normality assumption, the logarithm of the likelihood ratio test statistic for \( p < n \) is

\[
T_{n1} = \frac{pn}{2} \log \left\{ \frac{1}{p} \text{tr}(\hat{\Sigma}_n) \right\} - \frac{n}{2} \log |\hat{\Sigma}_n|,
\]

where \( \text{tr}(A) \) and \( |A| \) stand for the trace and the determinant of the matrix \( A \), respectively. Denote by \( \lambda_j \)'s the eigenvalues of \( \hat{\Sigma}_n \). Then

\[
T_{n1} = (n/2) \left\{ p \log \left( p^{-1} \sum_{j=1}^{p} \lambda_j \right) - \sum_{j=1}^{p} \log \lambda_j \right\}.
\]
When $p$ is fixed and finite, this test statistic has been well studied (Section 10.7, Anderson, 2003). Using the theory of the general likelihood ratio test, the limiting null distribution of $2T_{n1}$ is a chi-square distribution. For high-dimensional data, the likelihood ratio test theory may become invalid. In particular, we shall demonstrate in the next theorem that when $p/n \to y \in (0, 1)$, the limiting null distribution of $2T_{n1}$ is not a chi-square distribution any more. To broaden the applicability of our theory, we relax the normality assumption by imposing the following moment condition:

(A) Assume that $\mathbf{X}$ can be represented as $\mathbf{X} = \mu + \Sigma^{1/2} \mathbf{W}$, where $\mathbf{W} = (w_1, \ldots, w_p)'$, and $w_1, \ldots, w_p$ being independent and identically distributed and $E(w_j) = 0$, $E(w_j^2) = 1$ and $E(w_j^4) = \kappa < \infty$.

In the representation in Condition (A), it is natural to assume that $w_j$ is standardized so that $E(w_j) = 0$ and $E(w_j^2) = 1$. If $w_j$ is symmetric with finite kurtosis, then Condition (A) is satisfied. Of course, multivariate normal distribution satisfies Condition (A). Many other distributions may also satisfy Condition (A). Condition (A) implies that multinormality assumption indeed is unnecessary. This widens the application of theories developed in this work. Define $y_{n-1} = p/(n-1)$, $a_1(y) = [(y-1)/y] \log(1-y) - 1$, $a_2(y) = -0.5 \{ \log(1-y) + (\kappa - 3)y \}$ and $a_3(y) = -2y - 2 \log(1-y)$. The following theorem presents the asymptotic null distribution of $T_{n1}$.

**Theorem 2.1** Suppose that $\lim_{n \to \infty} y_n = y \in [0, 1)$. Under Assumption (A) and under $H_{10}$, it follows that

\[
(2/n)T_{n1} + p a_1(y_{n-1}) \xrightarrow{d} N(a_2(y), a_3(y)). \tag{2.2}
\]

where the symbol $\xrightarrow{d}$ stands for the weak convergence.

The proof of Theorem 2.1 is given in the Appendix. For example, $p = (n-1)/2$, and $y = 0.5$, $a_1(y_{n-1}) \approx \log(2) - 1$, $a_2(y) = \log(2)/2$ and $a_3(y) = 2 \log(2) - 1$, the Wilks phenomenon becomes invalid. Figure 1 depicts the plots of $a_j(y)$ with $\kappa = 3$. This figure clearly implies that the limiting null distribution of $2T_{n1}$ does not follow a chi-square distribution any
more, and the general theory of likelihood ratio test becomes invalid. When $y_n \uparrow 1$, then $a_1(y_{n-1}) \downarrow -1, a_2(y_m) \uparrow \infty$ and $a_3(y_n) \uparrow \infty$. Thus, the limiting distribution of $T_{n1}$ with $y = 1$ is quite different from that with $y \in [0, 1)$.

When $p$ is fixed and finite, it follows by the classical Wilks theorem that $2T_{n1} \overset{D}{\rightarrow} \chi^2(q)$ with $q = p(p+1)/2 - 1$. Denote $Z = (\chi^2(q) - q)/\sqrt{2q}$, which tends to $N(0, 1)$ in distribution as $p \to \infty$. Thus, when $p$ is large, $\frac{2}{n} T_{n1} - y_n \frac{p+1}{2} \approx y_n Z$. This is consistent with Theorem 2.1 since $y \to 0+$, and then all three $a_j(y) \to 0$.

We next study the limiting distribution under alternative hypothesis. When $c = 1$ or a known constant, some authors have studied the asymptotic power under different alternatives. Under the assumption of multinormality on the observed data, Wang, Gao and Miao (2013) studied the asymptotic power of likelihood ratio test with $c = 1$ under general alternative $H_1 : \Sigma \neq I_p$. The authors derived an explicit expression of the power when its corresponding likelihood ratio test tends to a constant. Onatski, Moreira and Hallin (2013) studied asymptotic behavior of the likelihood ratio test under the setting in which the observations follows $N(0, \Sigma)$ with covariance structure $\Sigma = c(I_p + hvv^T)$, where $v$ is a $p$-dimensional nonzero vector and $h$ is a scalar. The authors studied the asymptotic power of the following testing hypothesis

$$H_0 : h = 0 \quad \text{versus} \quad H_1 : h \neq 0.$$  

In this paper, we will study the asymptotic power for more general alternatives. Under $H_{10}$, all eigenvalues of $\Sigma$ equal $c$, and therefore $G_p(t) = p^{-1} \sum_{j=1}^p I(\lambda_j \leq t)$ takes value 0 for $t < c$ and 1 for $t \geq c$. This motivates us to consider a specific alternative $H_{11} : G_p(t) \to G(t)$, in which $G(t)$ is not degenerated to a single point distribution. We have the following limiting distribution under this alternative.

**Theorem 2.2** Suppose that $\lim_{n \to \infty} y_n = y \in [0, 1)$. Under Assumption (A) and under $H_1 : G_p(t) \to G(t)$, a non-degenerated distribution, and $\Sigma$ satisfies (A.3) of Lemma A.3 in Appendix, it follows that

$$(2/n) T_{n1} - p[\log(b_1) - b_2] \overset{d}{\rightarrow} N\left( (b_1^{-1}, -1) \mu_1^{\text{CLT}}, (b_1^{-1}, -1) \Sigma_1^{\text{CLT}}(b_1^{-1}, -1)' \right),$$
where \( b_1 > 0, b_2 \in \mathbb{R}^1, \mu_1^{\text{CLT}} \in \mathbb{R}^2 \) and \( \Sigma_1^{\text{CLT}} \) is some \( 2 \times 2 \) positive definite matrix, which are given in the proof of Theorem 2.2.

2.2 Test of independence

Under a normality assumption, test of independence is equivalent to test whether \( \Sigma \) is diagonal or not. With slightly abuse of terminology, we still use test of independence for the following null hypothesis without normality assumption:

\[
H_{20} : \Sigma = \text{diag}(c_1, \ldots, c_p),
\]

(2.3)

where \( c_i, i = 1, \ldots, p \) are unknown positive constant versus \( H_{21} : \Sigma \) is not diagonal. Write \( X_i = (x_{i1}, \ldots, x_{ip})' \). Thus, under \( H_{20} \) and under normality assumption on \( X \), the maximum likelihood estimator for \( c_j \) is

\[
\hat{c}_j = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2,
\]

where \( \bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \). Thus, for \( p < n \), the logarithm of the corresponding likelihood ratio statistic under normality assumption is

\[
T_{n2} = \frac{n}{2} \sum_{j=1}^p \log(\hat{c}_j) - \frac{n}{2} \log |\hat{\Sigma}_n|.
\]

Using the random matrix theory, we can derive the limiting null distribution of \( T_{n2} \), which is given the following theorem.

**Theorem 2.3** Suppose that \( \lim_{n \to \infty} y_n = y \in [0, 1) \). Under Assumption (A) and under \( H_{20} \), it follows that

\[
(2/n)T_{n2} + y_{n-1} + p a_1(y_{n-1}) \xrightarrow{d} N(a_2(y), a_3(y)).
\]

The proof of Theorem 2.3 is given in the Appendix.

We have also studied the asymptotic distribution under an alternative hypothesis. It can be shown that the limiting distribution under the alternative hypothesis can be expressed by the limiting spectral distribution of the alternative covariance matrix and its Stieltjes transform. However, it is numerical intractable since there is no existing algorithm to evaluate the limiting distribution. Thus, we opt not to present the limiting distribution here.
2.3 Test of equicorrelation

In this section, we study the test of equicorrelation. Let \( \mathbf{1} \) be a \( p \)-dimensional vector with all elements being 1. Test of equicorrelation corresponds to the following null hypothesis:

\[
H_{30} : \Sigma = c(1 - \rho)\mathbf{1}_p + c\rho \mathbf{1}\mathbf{1}'
\]

where \( c > 0 \) and \(-1/(p-1) < \rho < 1\) are unknown constants. Under normality assumption \( \mathbf{X} \sim N(\mu, \Sigma) \), the maximum likelihood estimators of \( c \) and \( \rho \) under \( H_{30} \) are as

\[
\hat{c} = \frac{\text{tr}(\hat{\Sigma}_n)}{p} \quad \text{and} \quad \hat{\rho} = \frac{1'\hat{\Sigma}_n \mathbf{1} - \text{tr}(\hat{\Sigma}_n)}{(p-1)\text{tr}(\hat{\Sigma}_n)}.
\]

Let \( \hat{\Sigma}_n = (1 - \hat{\rho})\mathbf{1}_p + \hat{\rho} \mathbf{1}\mathbf{1}' \), then we have

\[
|\hat{\Sigma}_n| = (1 - \hat{\rho})^{p-1} + (p - 1)\hat{\rho}(1 - \hat{\rho})^{p-2}
\]

\[
\hat{\Sigma}_n^{-1} = \frac{1}{1 - \hat{\rho}} \cdot \mathbf{1}_p - \frac{\hat{\rho}}{(1 - \hat{\rho})[1 + (p - 1)\hat{\rho}]} \cdot \mathbf{1}\mathbf{1}'
\]

\[
\frac{1}{n} \sum_{i=1}^{n} (\mathbf{X}_i - \bar{\mathbf{X}})' \hat{\Sigma}_n^{-1} (\mathbf{X}_i - \bar{\mathbf{X}}) = \frac{\text{tr}(\hat{\Sigma}_n)}{1 - \hat{\rho}} - \frac{\hat{\rho}}{(1 - \hat{\rho})[1 + (p - 1)\hat{\rho}]} \cdot 1'\hat{\Sigma}_n \mathbf{1}.
\]

Thus, the logarithm of the likelihood ratio statistic is

\[
T_{n3} = \frac{n}{2} \left\{- (p - 1)\log(p - 1) + (p - 1)\log \left( \frac{\text{tr}(\hat{\Sigma}_n)}{p} \right) + \log \left( \frac{1'\hat{\Sigma}_n \mathbf{1}}{p} \right) - \log |\hat{\Sigma}_n| \right\}.
\]

The following theorem presents the limiting null distribution of \( T_{n3} \).

**Theorem 2.4** Suppose that \( \lim_{n \to \infty} y_n = y \in [0, 1) \). Under Assumption (A) and under \( H_{30} \), it follows that

\[
(2/n)T_{n3} + \rho \omega_3(y_{n-1}) - \left[ (p - 1)\log \left( \frac{n(p - n + 1)}{np - n} \right) + \log \left( \frac{n - 1}{n} \right) \right] \xrightarrow{d} N(a_2(y), a_3(y)).
\]

The proof of Theorem 2.4 is given in the Appendix.

3 Numerical studies

In this section, we conduct Monte Carlo simulation study to compare Type I error rates of \( T_{nk} \), \( k = 1, 2 \) and 3 using critical values obtained from the new limiting null distribution.
with the corresponding traditional likelihood ratio tests which use critical values obtained from their limiting chi-square distributions. We further examine powers of \( T_{nk} \), \( k = 1, 2 \) and 3 via Monte Carlo simulation. We illustrate the proposed methodology by an empirical analysis of a Chinese stock market data set.

**Example 1.** In this example, data are generated from \( N(0, \Sigma) \). To examine the Type I error rate of \( T_{n1} \), we set \( \Sigma = \mathbf{I}_p \) and significant level 0.05. We first set \( y_n = 0.05 \) and 0.10, and consider various combinations of \((n, p)\). For each combination, we conduct 5000 replications. Thus, the Monte Carlo error is \( 1.96\sqrt{0.05 \times 0.95/5000} = 0.006 \). The simulation results are summarized in the top two panels of Table 1, in which GLRT stands for the likelihood ratio test using critical value from the new limiting distribution, and LRT for the traditional likelihood ratio test using critical value from the limiting chi-square distribution. Table 1 clearly show that the GLRT keeps the Type I error rate very well, while the LRT keeps Type I error rate poorly when \( n \) and \( p \) are relatively large. We further fix \( n = 200 \), and let \( p \) vary from 5 to 100, and conduct 5000 simulation for each \( p \). The simulation results are reported in the bottom panel of Table 1, from which it can be seen that the GLRT keeps Type I error rate very well, but the LRT fails quickly as \( p \) increases.

<table>
<thead>
<tr>
<th>((n, p))</th>
<th>(100,5)</th>
<th>(200,10)</th>
<th>(600,30)</th>
<th>(1000,50)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLRT</td>
<td>0.048</td>
<td>0.053</td>
<td>0.049</td>
<td>0.054</td>
</tr>
<tr>
<td>LRT</td>
<td>0.055</td>
<td>0.064</td>
<td>0.091</td>
<td>0.117</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>((n, p))</th>
<th>(50,5)</th>
<th>(100,10)</th>
<th>(300,30)</th>
<th>(500,50)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLRT</td>
<td>0.050</td>
<td>0.049</td>
<td>0.052</td>
<td>0.054</td>
</tr>
<tr>
<td>LRT</td>
<td>0.069</td>
<td>0.076</td>
<td>0.151</td>
<td>0.278</td>
</tr>
</tbody>
</table>

\(n = 200\)

<table>
<thead>
<tr>
<th>((n, p))</th>
<th>(200,5)</th>
<th>(200,10)</th>
<th>(200,50)</th>
<th>(200,100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLRT</td>
<td>0.054</td>
<td>0.054</td>
<td>0.045</td>
<td>0.054</td>
</tr>
<tr>
<td>LRT</td>
<td>0.059</td>
<td>0.062</td>
<td>0.814</td>
<td>1.000</td>
</tr>
</tbody>
</table>

We next examine the power of \( T_{n1} \). To this end, we first set \( \Sigma = \text{diag}(a, \mathbf{I}_{p-1}) \) and let
$\alpha$ vary. The power curves are depicted in the left panel of Figure 2, from which it can be seen that (a) the GLRT keeps Type I error rate well when $\alpha = 1$, (b) the GLRT becomes more powerful when the sample size increases, while the value of $\eta$ (0.05 or 0.10) keep unchanged, and (c) when the sample size $n = 200$ is fixed, the GLRT loses its power as the dimension $p$ increases. We further examine the power of $T_{n1}$ when $\Sigma = (1 - r)I_p + r11'$ with different values of $r$. Note that this covariance matrix has one eigenvalue $(p - 1)r + 1$ and all other eigenvalues $1 - r$. In particular, when $r = 0$, it corresponds to the null hypothesis.

The power curves are depicted in the right panel of Figure 2. The GRLT keeps its Type I error rate very well in this case. The patterns of (b) and (d) are similar to those of (a) and (c). However, the pattern of (f) shows that the power increases as the dimension $p$ increases. This is because the difference between the largest eigenvalue and the other eigenvalues is $pr$, and become larger as $p$ increases.

**Example 2.** In this example, we examine the performance of $T_{n2}$. Data are generated from $N(0, \Sigma)$. To examine Type I error rate of $T_{n2}$, we set $\Sigma = I_p$ and consider various combinations of $n$ and $p$. For each combination $(n, p)$, we conduct 5000 simulations. The Type I error rates are reported in Table 2, which has the same pattern as that in Table 1. Overall, the GLRT keeps Type I error rate very well, while LRT fails to maintain its Type I error rate.

<table>
<thead>
<tr>
<th></th>
<th>$y_n = 0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(n, p)$</td>
<td>(100,5)</td>
</tr>
<tr>
<td>GLRT</td>
<td>0.046</td>
</tr>
<tr>
<td>LRT</td>
<td>0.051</td>
</tr>
</tbody>
</table>

For $n = 200$

<table>
<thead>
<tr>
<th>$(n, p)$</th>
<th>(200,5)</th>
<th>(200,10)</th>
<th>(200,50)</th>
<th>(200,100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLRT</td>
<td>0.047</td>
<td>0.045</td>
<td>0.048</td>
<td>0.055</td>
</tr>
<tr>
<td>LRT</td>
<td>0.063</td>
<td>0.067</td>
<td>0.782</td>
<td>1.000</td>
</tr>
</tbody>
</table>
To investigate the empirical power of $T_{n2}$, we consider $\Sigma = \text{diag}(R_1, I_{p-2})$, where

$$R_1 = \begin{pmatrix} 2 & r \\ r & 0.5 \end{pmatrix}$$

with $r$ ranges from 0 to 1. The power curves are depicted in the left panel of Figure 3. When $r = 0$, the corresponding $\Sigma$ is diagonal. The GLRT keeps its Type I error rate very well. When $y_n$ is fixed at 0.05, the performance of the GLRT becomes better as the sample size increases. When the sample size $n$ is fixed at 200, the performance of the GLRT is getting worse as the dimension increases. We further examine the empirical power of $T_{n2}$ when $\Sigma = (1 - r)I_p + r11'$ for $r \in [0, 1)$. The power curves are depicted in the right panel of Figure 3. The power is near 0.05 when $r = 0$, which implies that the GLRT keeps its Type I error rate well. The power quickly increases to one as $r$ increase.

**Example 3.** We assess the performance of the GLRT $T_{n3}$ in this example. Data are generated from $N(0, \Sigma)$. We first examine the Type I error rate of $T_{n3}$. To this end, we set $\Sigma = 0.5I_p + 0.511'$, and conduct 5000 simulations for each combination of $(n, p)$ listed in Table 3, from which we can see that the GLRT keeps Type I error rate well, but the LRT does not.

<table>
<thead>
<tr>
<th>$y_n = 0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(n, p)$</td>
</tr>
<tr>
<td>GLRT</td>
</tr>
<tr>
<td>LRT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(n, p)$</td>
</tr>
<tr>
<td>GLRT</td>
</tr>
<tr>
<td>LRT</td>
</tr>
</tbody>
</table>

We next investigate the power of $T_{n3}$. We consider two alternatives: (a) $\Sigma = 0.5I_p +$
$0.511' + rR_2$ with $r \in (-0.5, 0.5)$ and

$$R_2 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and (b) $\Sigma' = 0.5(1+r)I_p + 0.5(1-r)11' + 0.5rR_3$, where $r \in (0, 1)$ and $R_3$ has $(j,j-1)$- and $(j-1,j)$-elements being one, and others being zero. The power curves for each combination of $(n, p)$ listed in Table 3 based on 5000 simulations are depicted in Figure 4, from which it can be seen that $T_{n,3}$ keeps its Type I error rate very well, and the power increases as the sample size increases when $n_a$ is fixed.

**Example 4.** In this example, we illustrate the proposed tests by an empirical analysis of a real data set, which was extracted from a commercial database containing weekly returns for all of the stocks traded on the Chinese stock market during the period of 2008-2009. The original data set contains many missing observations. By excluding the missing observations, we have a data subset containing 132 weekly returns for 97 stocks. It is known that the stock returns depend on other factors. Here we consider three factors: $z_1$ stands for returns of the Shanghai Composite Index (i.e., the market index), $z_2$ for the difference in returns between portfolios of small capitalization firms and large capitalization firms and $z_3$ for the difference in returns between portfolios of high book-to-market ratio firms and low book-to-market ratio firms. Two popular regression models have been used for asset pricing. The first one is the capital asset pricing model (CAPM; Markowitz (1952), Sharpe (1964), Litmer (1965)) and the second one is the Fama and French’s three factor model (TFM; Fama and French, 1993, 1996). In the CAPM, it includes only the factor $z_1$. Let $X_i$ be a 97-dimensional vector for the weekly returns of 97 stocks at week $i$. The CAPM is the following regression model

$$X_i = B_{10}B_{11}z_{1i} + E_{1i},$$

(3.1)

where $B_{10}$ and below are regression coefficient vector or matrix and $E_{1i}$ and below are the random errors. Let $Z_i = (z_{1i}, z_{2i}, z_{3i})'$. The TFM considers

$$X_i = B_{20} + B_{21} Z_i + E_{2i}.$$  

(3.2)
It is of interest to test whether the CAPM is sufficient for describing the stock returns. This can be formulated as test whether $\text{Cov}(E_t)$ in (3.1) is diagonal. If the CAPM is not sufficient, it may be of great interest to test whether the TFM is sufficient for describing the stock returns and whether the TFM offers some improvement. Both CAPM and TFM are linear models. Fortunately, the proposed tests can be extended for linear models provided that the number of predictors in the linear models is finite and fixed. Specifically, suppose that $\{X_1, \cdots, X_n\}$ is a random sample from a linear model

$$X = B_0 + BZ + E,$$

where $B_0$ and $B$ consist of regression coefficients and $Z$ is a covariate vector. One may easily derive the corresponding likelihood ratio test for the covariance matrix of $E$. Assume that Assumption (A) holds for the random error vector $E$. Under conditions in Theorems 2.1 to 2.4, it can be shown by applying techniques used in the proofs of these theorems that the limiting distributions for testing $\text{Cov}(E)$ in these theorems are still valid provided that the dimension of $Z$ is finite and fixed.

We now conduct test of independence for the random error vector in model (3.1). In this case, $y_{n-2} = 37/(132 - 2)$ and the estimated $\hat{k} = 5.2625$ for the residuals of model (3.1). $[(2/n)T_{n2} + y_{n-2} + \rho \alpha_1(y_{n-2}) - \alpha_2(y_{n-2})]/\sqrt{\alpha_2(y_{n-2})} = 21.68$. Thus, $T_{n2}$ has $P$-value 0.0000. This implies that the CAPM is not sufficient to describe the stock returns. We next perform test of independence for the random error vector in model (3.2). $y_{n-4} = 97/(132 - 4)$ and the estimated $\hat{k} = 5.1554$ for the residuals of this model. The corresponding $[(2/n)T_{n2} + y_{n-4} + \rho \alpha_1(y_{n-4}) - \alpha_2(y_{n-4})]/\sqrt{\alpha_2(y_{n-4})} = 13.67$, which has significantly decrease from the one for the CAMP. This implies that the TFM offers improvement over the CAPM. However, $T_{n2}$ also has $P$-value 0.0000. This implies that the TFM is not sufficient to describe the stocks returns either. Our conclusion is consistent with Lan, et al. (2013), in which a different data subset was analyzed. This empirical analysis implies that more factors are needed for a better description of the stock returns. As a result, further financial research along this direction is still needed.
Appendix

In this section, we first derive some results on random matrix theory that will be used to derive the limiting distributions in Section 2. The proofs of theorems in Section 2 will be given in Section A.2.

A.1 Some results on random matrix theory

Let \( \{x_{ki} \in \mathbb{R}, k, i = 1, 2, \cdots \} \) be a double array of independent and identically distributed random variables with mean 0 and variance 1. Let \( \mathbf{X}_i = (x_{1i}, x_{2i}, \cdots, x_{pi})' \), and consider \( \mathbf{X}_1, \cdots, \mathbf{X}_n \) as an independent and identically distributed random sample from a \( p \)-dimensional distribution with mean 0 and covariance matrix \( \mathbf{I}_p \). Therefore the unbiased sample covariance matrix is

\[
\mathbf{S}_n = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'.
\]

To derive the limiting distribution of \( T_{nk}, k=1, 2 \) and 3, we need the limiting distributions of functionals of eigenvalues of \( \mathbf{S}_n \).

We first define the Marchenko-Pastur law. For \( 0 < \theta \leq 1 \), let \( a(\theta) = (1 - \sqrt{\theta})^2 \) and \( b(\theta) = (1 + \sqrt{\theta})^2 \). The Marchenko-Pastur distribution of index \( \theta \), denoted as \( F^\theta \), is the distribution on \([a(\theta), b(\theta)]\) with the following density function

\[
f^\theta(x) = \frac{1}{2\pi x} \sqrt{[b(\theta) - x][x - a(\theta)]}, \quad a(\theta) \leq x \leq b(\theta).
\]  \hspace{1cm} (A.1)

Let \( y_n = \frac{\lambda}{n} \to y \in [0, 1) \) and \( F^y, F^{y_n} \) be the Marchenko-Pastur law of index \( y \) and \( y_n \), respectively. Let \( \mathcal{U} \) be an open set of the complex plane, including \([a(y), b(y)]\) and \( \mathcal{A} \) be the set of analytic functions \( f : \mathcal{U} \to \mathbb{C} \). We consider the empirical process \( G_n := \{ G_n(f) \} \) indexed by \( \mathcal{A} \),

\[
G_n(f) = \sum_{j=1}^{n} f(\lambda_j) I(\lambda_j > 0) - p \cdot \int_{0}^{+\infty} f(x) F^{y_n-1}(dx), \quad f \in \mathcal{A},
\]  \hspace{1cm} (A.2)

where \( \lambda_j \) are eigenvalues of \( \mathbf{S}_n \) and \( I(\cdot) \) is an indicator function. Lemmas A.1, A.2 and A.3 below play an important role in the proofs of the theorems. The proofs of Lemmas A.1,
A.2 and A.3 are quite technical and therefore are given in the supplemental material of this paper.

**Lemma A.1** Assume that \( f_1, \ldots, f_k \in A \), and \( \{X_{ij}\} \) are independent and identically distributed random variables with \( E X_{11} = 0 \), \( E X_{11}^2 = 1 \), and \( \kappa = E X_{11}^4 < \infty \). Moreover, \( \frac{y}{n} \to y \in [0, 1] \) as \( n,p \to \infty \). Then the random vector \((G_n(f_1), \ldots, G_n(f_k))\) weakly converges to a \( k \)-dimensional Gaussian vector with mean vector,

\[
m(f_j) = \frac{1}{4} \int_{a(y)}^{b(y)} f_j(x) \, dx - \frac{1}{2\pi} \int_{a(y)}^{b(y)} \sqrt{4y - (x - 1 - y)^2} \, dx \]  

\[-y(\kappa - 3) \cdot \frac{1}{2\pi i} \int f_j(z) \frac{y m^2(z)(1 + m(z))^{-3}}{1 - ym^2(z)(1 + m(z))} \, dz\]

and covariance function

\[
u(f_j, f_i) = -\frac{1}{2\pi^2} \int \int \frac{f_j(z_1)f_i(z_2)}{(m(z_1) - m(z_2))^2} \, dm(z_1) \, dm(z_2) + y(\kappa - 3) \cdot \frac{1}{2\pi i} \int \frac{f_j(z_1)}{(1 + m(z_1))^2} \, dm(z_1) \cdot \frac{1}{2\pi i} \int \frac{f_i(z_2)}{(1 + m(z_2))^2} \, dm(z_2)
\]

where \( m(z) \equiv m_{F^y}(z) \) is the Stieltjes Transform of \( F^y \equiv (1 - y)I_{[0,\infty]} + yF^y \), the contours are non-overlapping and both contain the support set \( [(1 - \sqrt{y})^2, (1 + \sqrt{y})^2] \) of \( F^y \).

To apply Lemma A.1 for our settings, we need the following properties of several integrations.

**Lemma A.2** Under conditions of Lemma A.1, we have

\[
\int_{a(y_{n-1})}^{b(y_{n-1})} x f_{n-1}(x) \, dx = 1
\]

\[
\int_{a(y_{n-1})}^{b(y_{n-1})} \log(x) f_{n-1}(x) \, dx = \frac{y_{n-1} - 1}{y_{n-1}} \log(1 - y_{n-1}) - 1
\]

\[
\frac{a(y) + b(y)}{4} - \frac{1}{2\pi} \int_{a(y)}^{b(y)} \frac{x}{\sqrt{4y - (x - 1 - y)^2}} \, dx = 0
\]

\[
\frac{\log(a(y)) + \log(b(y))}{4} - \frac{1}{2\pi} \int_{a(y)}^{b(y)} \frac{\log(x)}{\sqrt{4y - (x - 1 - y)^2}} \, dx = \frac{\log(1 - y)}{2}
\]

\[
\frac{1}{2\pi^2} \int \int \frac{z_1 z_2}{(m(z_1) - m(z_2))^2} \, dm(z_1) \, dm(z_2) = 2y
\]

\[
\frac{1}{2\pi^2} \int \int \frac{z_1 \log(z_2)}{(m(z_1) - m(z_2))^2} \, dm(z_1) \, dm(z_2) = 2y
\]

\[
\frac{1}{2\pi^2} \int \int \frac{\log(z_1) \log(z_2)}{(m(z_1) - m(z_2))^2} \, dm(z_1) \, dm(z_2) = -2 \log(1 - y)
\]
\[
\frac{1}{2\pi i} \oint z \frac{ym^3(z)(1 + m)^{-3}}{1 - ym^2(z)(1 + m)^{-2}} dz = 0
\]
\[
\frac{1}{2\pi i} \oint \log(z) - \frac{ym^3(z)(1 + m)^{-3}}{1 - ym^2(z)(1 + m)^{-2}} dz = \frac{y}{2}
\]
\[
\frac{1}{2\pi i} \oint \frac{z}{(1 + m(z))^2} dm(z) = 1 \quad \text{and} \quad \frac{1}{2\pi i} \oint \frac{\log(z)}{(1 + m(z))^2} dm(z) = 1.
\]

Let \( e_i \) be a \( p \)-dimensional column vector whose \( i \)-th element equals 1 and the others equal 0. Denote \( B_{1i}(z) = e_i^T \Sigma^{1/2}(m(z)\Sigma + I)^{-1} \Sigma^{1/2} e_i \) and \( B_{2i}(z) = e_i^T \Sigma^{1/2}(m(z)\Sigma + I)^{-2} \Sigma^{1/2} e_i \).

**Lemma A.3.** Suppose that the population covariance matrix \( \Sigma \) has the limiting spectral distribution \( G(t) \), a non-degenerated distribution, and

\[
M(z) = \lim_{p \to \infty} \frac{1}{p} \sum_{i=1}^p [B_{1i}(z)B_{2i}(z)], \quad C(z_1, z_2) = \lim_{p \to \infty} \frac{1}{p} \sum_{i=1}^p B_{1i}(z_1)B_{1i}(z_2)
\]

exist and \( C(z_1, z_2) \) is differentiable with respect to \( z_1 \) and \( z_2 \), and \( m(z) \) satisfies

\[
z = -\frac{1}{tm(z)} + y \int \frac{t}{1 + tm(z)} dG(t).
\]

Denote \( \lambda_j \) are eigenvalues of \( S_n \Sigma^{-1} S_n^T \) is the limiting spectral distribution of \( S_n \) when \( p/n \to y \in [0, 1) \), and \( F_{y^{n-1}G} \) is obtained by replacing \( y \) by \( y_{n-1} \) in \( F_{yG} \). Under Assumption (A), we have that

\[
G_n(f) = \left\{ \sum_{j=1}^p f(\lambda_j)I_{\lambda_j>0} - p \cdot \int_0^{+\infty} f(x) F_{y_{n-1}G}(dx) \right\}
\]

converges to a Gaussian vector with mean \( EX_f \) and covariance function \( \text{Cov}(X_f, X_g) \) as follows.

\[
EX_f = -\frac{1}{2\pi i} \oint f(z) \frac{ym^3(z)t^2(1 + tm(z))^{-3}dG(t)}{(1 - y \int m^2(z)t^2(1 + tm(z))^{-2}dG(t))^2} dz
\]
\[
-\frac{y(\kappa - 3)}{2\pi i} \cdot \oint f(z) \frac{m^3(z)M(z)}{1 - y \int m^2(z)M(z)G(t)} dz
\]
\]

and

\[
\text{Cov}(X_f, X_g) = -\frac{1}{2\pi^2} \oint \oint \frac{f(z_1)g(z_2)}{(m(z_1) - m(z_2))^2} \frac{d}{dz_1}m(z_1) \frac{d}{dz_2}m(z_2) dz_1 dz_2
\]
\[
-\frac{y(\kappa - 3)}{4\pi^2} \oint \oint f_1(z_1)f_2(z_2) - \frac{d^2}{dz_1 dz_2} [m(z_1)m(z_2)C(z_1, z_2)] dz_1 dz_2
\]
A.2 Proofs of Theorems

Proof of Theorem 2.1. Note that $\lambda_j$’s are eigenvalues of the sample covariance matrix $S_n$. Thus, $\lambda_j/c$, $j = 1, \ldots, p$ are the eigenvalues of $c^{-1}S_n$. Thus, under $H_0 : \Sigma = \sigma I_p$ and under Assumption (A), $c^{-1}S_n$ is the sample covariance matrix of $(w_{ik}, \ldots, w_{ip})'$ with $\{w_{ik}, i, k = 1, 2, \ldots, \}$ being a double array of independent random variables with mean 0 and variance 1. Therefore we can apply Lemma A.1 for $c^{-1}S_n$.

For ease of notation, define $L_{n1}^* = 2T_{n1}/n$. Then it follows that

$$
L_{n1}^* = p \log \left\{ \frac{1}{p} \text{tr}(S_n) \right\} - \log |S_n| = p \log \left\{ \frac{1}{p} \text{tr}(S_n) \right\} - \log |S_n| = p \log \left\{ \frac{1}{p} \text{tr}(S_n/c) \right\} - \log |S_n/c| = p \cdot \log \left( \frac{1}{p} \sum_{j=1}^{p} \frac{\lambda_j}{c} \right) - \sum_{j=1}^{p} \log \frac{\lambda_j}{c}.
$$

To derive the limiting distribution of $L_{n1}^*$, we first derive the joint asymptotic distribution of $(p^{-1} \sum_{j=1}^{p} \lambda_j/c, p^{-1} \sum_{j=1}^{p} \log(\lambda_j/c))$. By Lemma A.1, it follows by taking $g_1(x) = x$, $g_2(x) = \log(x)$ that

$$
\left( \begin{array}{c}
p \left\{ \frac{1}{p} \sum_{j=1}^{p} \frac{\lambda_j}{c} - \int g_1(x) f^{y_{n-1}}(x) dx \right\} \\
p \left\{ \frac{1}{p} \sum_{j=1}^{p} \log \frac{\lambda_j}{c} - \int g_2(x) f^{y_{n-1}}(x) dx \right\}
\end{array} \right) \overset{d}{\rightarrow} N(\mu_0, \Sigma_0),
$$

(A.7)

where $f^{y_{n-1}}(x)$ is the density of Parzenko-Pastur law with index $y_{n-1}$, and

$$
\mu_0 = (m(g_1), m(g_2))', \quad \Sigma_0 = \left( \begin{array}{cc}
\nu(g_1, g_1) & \nu(g_1, g_2) \\
\nu(g_1, g_2) & \nu(g_2, g_2)
\end{array} \right).
$$

Let $b = \int g_1(x) f^{y_{n-1}}(x) dx$. It follows by the delta-method that

$$
p \cdot \left\{ \log \left( \frac{1}{p} \sum_{j=1}^{p} \frac{\lambda_j}{c} \right) - \log(b) - \frac{1}{p} \sum_{j=1}^{p} \log \frac{\lambda_j}{c} + \int g_2(x) f^{y_{n-1}}(x) dx \right\} \overset{d}{\rightarrow} N \left( \left( \frac{1}{b}, -1 \right)', \mu_0, \left( \frac{1}{b}, -1 \right) \Sigma_0 \left( \frac{1}{b}, -1 \right) \right).
$$

Hence it follows that

$$
\left[ p \log \left( \frac{1}{p} \sum_{j=1}^{p} \lambda_j \right) - \sum_{j=1}^{p} \log \lambda_j \right] \sim \left[ p \log(b) - p \int g_2(x) f^{y_{n-1}}(x) dx \right]
$$

$$
\overset{d}{\rightarrow} N \left( \left( \frac{1}{b}, -1 \right)', \mu_0, \left( \frac{1}{b}, -1 \right) \Sigma_0 \left( \frac{1}{b}, -1 \right) \right).
$$
This implies that

$$L_{a_1} = \left[ p \log(b) - p \int g_2(x) f_{y_{n-1}}(x) dx \right] \xrightarrow{d} N \left( \left( \frac{1}{b}, -1 \right), \left( \frac{1}{b}, -1 \right) \right) \Sigma_n \left( \frac{1}{b}, -1 \right).$$

We next derive \( \mu_0, \Sigma_0 \), \( b = \int g_1(x) f_{y_{n-1}}(x) dx \) and \( \int g_2(x) f_{y_{n-1}}(x) dx \). By Lemma A.2, we have

$$\int g_1(x) f_{y_{n-1}}(x) dx = 1$$
$$\int g_2(x) f_{y_{n-1}}(x) dx = \frac{y_{n-1} - 1}{y_{n-1}} \log(1 - y_{n-1}) - 1.$$

As a result,

$$b = \int g_1(x) f_{y_{n-1}}(x) dx = 1 \quad \text{and} \quad \int g_2(x) f_{y_{n-1}}(x) dx = \frac{y_{n-1} - 1}{y_{n-1}} \log(1 - y_{n-1}) - 1. \quad (A.8)$$

By using Lemmas A.1 and A.2, we have

$$m(g_1) = \frac{a(y) + b(y)}{4} - \frac{1}{2\pi} \int_{a(y)}^{b(y)} \frac{x}{\sqrt{4y - (x - 1 - y)^2}} dx$$
$$- (\kappa - 3) \cdot \frac{1}{2\pi i} \oint \frac{ym^3(z)(1 + m)^{-3}}{1 - ym^2(z)(1 + m)^{-2}} dz$$
$$= 0$$

and

$$m(g_2) = \frac{\log(a(y)) + \log(b(y))}{4} - \frac{1}{2\pi} \int_{a(y)}^{b(y)} \frac{\log(x)}{\sqrt{4y - (x - 1 - y)^2}} dx$$
$$- (\kappa - 3) \cdot \frac{1}{2\pi i} \oint \frac{\log(z)}{ym^2(z)(1 + m)^{-3}} dz$$
$$= \frac{\log(1 - y)}{2} - (\kappa - 3) \cdot \frac{y}{2}.$$

Therefore, we have a closed form \( \mu_0 \):

$$\mu_0 = (m(g_1), m(g_2))' = \left( 0, \frac{\log(1 - y)}{2} - (\kappa - 3) \cdot \frac{y}{2} \right)'.$$

(A.9)

By Lemma A.1 again, it follows that

$$v(y, g_1) = - \frac{1}{2\pi^2} \oint \oint \frac{g_1(z_1)g_1(z_2)}{\left( a(z_1) - m(z_2) \right)^2} dm(z_1) dm(z_2)$$
$$+ y(\kappa - 3) \cdot \frac{1}{2\pi i} \oint \frac{g_1(z_1)}{(1 + m(z_1))^2} dm(z_1) \cdot \frac{1}{2\pi i} \oint \frac{g_2(z_2)}{(1 + m(z_2))^2} dm(z_2).$$
So by Lemma A.2, we have

\[
v(g_1, g_1) = -\frac{1}{2\pi^2} \int \int \frac{g_1(z_1)g_1(z_2)}{(m(z_1) - m(z_2))^2} dm(z_1) dm(z_2) \\
+ y(\kappa - 3) \cdot \frac{1}{2\pi i} \int \frac{g_1(z_1)}{(1 + m(z))^2} dm(z_1) \cdot \frac{1}{2\pi i} \int \frac{g_1(z_2)}{(1 + m(z))^2} dm(z_2)
= 2y + y(\kappa - 3)
\]

\[
v(g_1, g_2) = -\frac{1}{2\pi^2} \int \int \frac{g_1(z_1)g_2(z_2)}{(m(z_1) - m(z_2))^2} dm(z_1) dm(z_2) \\
+ y(\kappa - 3) \cdot \frac{1}{2\pi i} \int \frac{g_1(z_1)}{(1 + m(z))^2} dm(z_1) \cdot \frac{1}{2\pi i} \int \frac{g_2(z_2)}{(1 + m(z))^2} dm(z_2)
= 2y + y(\kappa - 3)
\]

\[
v(g_2, g_2) = -\frac{1}{2\pi^2} \int \int \frac{g_2(z_1)g_2(z_2)}{(m(z_1) - m(z_2))^2} dm(z_1) dm(z_2) \\
+ y(\kappa - 3) \cdot \frac{1}{2\pi i} \int \frac{g_2(z_1)}{(1 + m(z))^2} dm(z_1) \cdot \frac{1}{2\pi i} \int \frac{g_2(z_2)}{(1 + m(z))^2} dm(z_2)
= -2\log(1 - y) + y(\kappa - 3)
\]

As a result,

\[
\Sigma_0 = \begin{pmatrix}
v(g_1, g_1) & v(g_1, g_2) \\
v(g_1, g_2) & v(g_2, g_2)
\end{pmatrix} = \begin{pmatrix}
2y + y(\kappa - 3) & 2y + y(\kappa - 3) \\
2y + y(\kappa - 3) & -2\log(1 - y) + y(\kappa - 3)
\end{pmatrix}
\]

(A.10)

By (A.8)-(A.10), we obtain

\[
\frac{L_{n1}^*}{p} \left\{ \frac{(y_{n-1} - 1)}{y_{n-1}} \log(1 - y_{n-1}) - 1 \right\}
\xrightarrow{d} N \left( -\frac{\log(1 - y)}{2} + (\kappa - 3), \frac{y}{2} - 2y - 2\log(1 - y) \right).
\]

That is,

\[
\frac{(2/n)T_{n1} + p\alpha_1(y_{n-1})}{\alpha} \xrightarrow{d} N \left( \alpha_2(y), \alpha_3(y) \right).
\]

The proof of Theorem 2.1 is completed.

**Proof of Theorem 2.2.** Let \( g_1(x) = x \) and \( g_2(x) = \log x \). Then by the expressions \( E_Xf \) and \( \text{Cov}(X, X) \) in Lemma A.3, we have \( \mu_{i1}^{\text{CLT}} = (\mu_{i1}^{\text{CLT}}, \mu_{i2}^{\text{CLT}})^T \) and \( \Sigma_{i1}^{\text{CLT}} = (\Sigma_{i1}^{\text{CLT}}) \) for \( i, j = 1, 2, \) where \( \mu_{i1}^{\text{CLT}} = E_X g_1, \mu_{i2}^{\text{CLT}} = E_X g_2, \Sigma_{i11}^{\text{CLT}} = \text{Cov}(X, g_1), \Sigma_{i12}^{\text{CLT}} = \Sigma_{i21}^{\text{CLT}} = \text{Cov}(X, g_2), \) and \( \Sigma_{i22}^{\text{CLT}} = \text{Cov}(g_2, g_2) \). Moreover, we have

\[
b_1 = \int_a^b x f^{\nu_{n-1}G}(x) dx, \quad b_2 = \int_a^b \log x \cdot f^{\nu_{n-1}G}(x) dx
\]
where $f^{y_{n-1}}(x) = \frac{1}{y_{n-1}} \lim_{z \to x} \Im(m(z))$ and $z = -\frac{1}{\im\theta_{n}} + y \int_{1+\im\theta_{n}(x)}^{t} dG(t)$. In fact, the integrals in $b_{1}, b_{2}, \mu_{1}^{\text{CLT}}$, and $\Sigma_{1}^{\text{CLT}}$ can be approximated by Riemann sums. This completes the proof of Theorem 2.2.

**Proof of Theorem 2.3.** Denote $L_{\text{m2}}^{*} = 2T_{\text{m2}}/n$. Thus,

$$L_{\text{m2}}^{*} = \sum_{j=1}^{p} \log \left( \frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \bar{x}_{j})^{2} \right) - \log | \Sigma | = \sum_{j=1}^{p} \log \left( \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_{j})^{2} \right) - \log | S_{n} |$$

where $\bar{x}_{j} = \frac{1}{n} \sum_{i=1}^{n} x_{ij}$. By the Taylor expansion of $\log(1 + x)$, we have

$$= \sum_{j=1}^{p} \log \left( \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_{j})^{2} \right)$$

$$= \sum_{j=1}^{p} \log \left( \frac{1}{(n-1) \cdot c_{j}} \sum_{i=1}^{n} (x_{ij} - \bar{x}_{j})^{2} \right) + \sum_{j=1}^{p} \log(c_{j})$$

$$= \sum_{j=1}^{p} \left( \frac{1}{(n-1) \cdot c_{j}} \sum_{i=1}^{n} (x_{ij} - \bar{x}_{j})^{2} - 1 \right) + \sum_{j=1}^{p} \log(c_{j})$$

$$\frac{1}{2} \sum_{j=1}^{p} w_{j} \left( \frac{1}{(n-1) \cdot c_{j}} \sum_{i=1}^{n} (x_{ij} - \bar{x}_{j})^{2} - 1 \right)^{2} + o_{p}(1)$$

$$= \text{tr}(S_{n}^{*}) - p + \sum_{j=1}^{p} \log(c_{j}) - \frac{1}{2} \sum_{j=1}^{p} \left( \frac{1}{(n-1) \cdot c_{j}} \sum_{i=1}^{n} (x_{ij} - \bar{x}_{j})^{2} - 1 \right)^{2} + o_{p}(1)$$

$$= \text{tr}(S_{n}^{*}) - p + \sum_{j=1}^{p} \log(c_{j}) - \frac{p}{n-1} + o_{p}(1),$$

where $S_{n}^{*} = \Sigma^{1/2} S_{n} \Sigma^{1/2}$ with $\Sigma^{1/2} = \text{diag}(\sqrt{c_{1}}, \sqrt{c_{2}}, \cdots, \sqrt{c_{p}})$. Under $H_{21}$, $S_{n}^{*}$ is the sample covariance matrix of $(w_{i1}, \cdots, w_{ip})^{t}$ with $\{w_{ik}, i = 1, 2, \cdots, \}$ being a double array of independent and identically distributed random variables with mean 0 and variance 1.

Applying Lemma A.1 for $S_{n}^{*}$, we have

$$L_{\text{m2}}^{*} + y_{n-1} + p \cdot \int \log(x) f^{y_{n-1}}(x) dx$$

$$= (\text{tr} S_{n}^{*} - p) - (\log | S_{n}^{*} | - p \cdot \int \log(x) f^{y_{n-1}}(x) dx) + o_{p}(1), \quad (A.11)$$

where $f^{y_{n-1}}(x)$ is the density of Marchenko-Pastur distribution of index $y_{n-1}$. Thus, it is sufficient to derive the limiting distribution of

$$(\text{tr} S_{n}^{*} - p) - \left( \log | S_{n}^{*} | - p \cdot \int \log(x) f^{y_{n-1}}(x) dx \right).$$
Using (A.7), under conditions of Theorem 2.3, we have

\[
(\text{tr} \mathbf{S}_n - p) - \left( \log |\mathbf{S}_n^*| - p \cdot \int \log(x) f^{\text{un}\cdot}(x) dx \right)
\]

\[
\xrightarrow{d} N \left( -\frac{\log(1 - y)}{2} + (\kappa - 3) \cdot \frac{y}{2}, -2 \log(1 - y) - 2y \right)
\]

That is, under \(H_0\)

\[
\frac{2}{n} T_{n2} + y_{n-1} + p \cdot \left( \frac{y_{n-1} - 1}{y_{n-1}} \log(1 - y_{n-1}) - 1 \right)
\]

\[
\xrightarrow{d} N \left( -\frac{\log(1 - y)}{2} + (\kappa - 3) \cdot \frac{y}{2}, -2 \log(1 - y) - 2y \right)
\]

This completes the proof of Theorem 2.3.

**Proof of Theorem 2.4.** Under \(H_{30}\), we have \(\Sigma = c(1 - \rho)I_p + cpI_1\). Note that

\[
\Sigma^{-1} = \frac{1}{c} \left[ \frac{1}{1 - \rho} \cdot I_p - \frac{\rho}{1 - \rho(1 + (p - 1)\rho)} \cdot I_1 \right]
\]

\[
|\Sigma| = c^p (1 - \rho)^{p - 1} \cdot [1 + (p - 1) \cdot \rho].
\]

Let \(\mathbf{F} = \Sigma^{-\frac{1}{2}} \mathbf{S}_n \Sigma^{-\frac{1}{2}}\), which is the sample covariance matrix of \((w_{11}, \cdots, w_{n})'\) with \(\{w_{ik}, i, k = 1, 2, \cdots \}\) being a double array of independent and identically distributed random variables with mean 0 and variance 1. Under \(H_{30}\), we have

\[
\begin{align*}
\text{tr} \mathbf{S}_n &= c(1 - \rho) \text{tr} \mathbf{F} + cp \cdot \text{tr} 1' \mathbf{F} = c(1 - \rho) \text{tr} \mathbf{F} + cp \cdot 1' \mathbf{F} \\
\frac{1}{p} \mathbf{S}_n &= c[1 - \rho] \frac{1' \mathbf{F}}{p} + cp \mathbf{F} \cdot 1' \mathbf{F} = c[1 + (p - 1) \rho] \cdot \frac{1' \mathbf{F}}{p} \\
|\mathbf{S}_n| &= |\Sigma| \cdot |\mathbf{F}| = c^p (1 - \rho)^{p - 1} \cdot [1 + (p - 1) \cdot \rho] \cdot |\mathbf{F}|.
\end{align*}
\]

By the definition of \(T_{n3}\) and \((p - 1) \log \left( \frac{\text{tr} \Sigma_n - \frac{1}{p} \mathbf{S}_n}{p} \right) + \log \left( \frac{1}{p} \mathbf{S}_n \right) - \log |\Sigma_n| = (p - 1) \log \left( \frac{\text{tr} \mathbf{S}_n - \frac{1}{p} \mathbf{S}_n}{p} \right) + \log \left( \frac{1}{p} \mathbf{S}_n \right) - \log |\mathbf{S}_n|\), it follows that

\[
2T_{n3}/n = (p - 1) \log \left( \frac{1}{p - 1} \text{tr} \mathbf{F} - \frac{1' \mathbf{F}}{p} \right) + \log \left( \frac{1' \mathbf{F}}{p} \right) - \log |\mathbf{F}|.
\]

which can be re-expressed as

\[
(p - 1) \log \frac{p}{p - 1} + \frac{p - 1}{p} \cdot p \log \left( \frac{1}{p} \text{tr} \mathbf{F} - \frac{1' \mathbf{F}}{p^2} \right) + \log \left( \frac{1' \mathbf{F}}{p} \right) - \log |\mathbf{F}|.
\]
Under Assumption (A), we can show that $1' F^1 / p = 1 + O_p(1/\sqrt{n})$. Thus,

$$2T_{n3}/n = (p-1) \log \frac{p-1}{p-1} + \frac{p-1}{p} \cdot p \log \left( \frac{1}{p} tr F - \frac{1' F^1}{p^2} \right) - \log |F| + o_P(1).$$

By the delta method and the proof of Theorem 2.1, it follows that

$$\frac{(2/n)T_{n3} - \left[ (\mu - 1) \log \frac{p-1}{p-1} + (p-1) \log (b - \frac{1}{p}(1 - \frac{1}{n})) + \log (1 - \frac{1}{n}) - p \int g_2(x)f_{y_n-1}(x)dx \right]}{\mu_3} \xrightarrow{d} N \left( \left( \frac{p-1}{p}, \frac{1}{b}, -1 \right) \right)^\prime \mu_3, \left( \frac{p-1}{p}, \frac{1}{b}, -1 \right) \right)^\prime \Sigma_3 \left( \frac{p-1}{p}, \frac{1}{b}, -1 \right) \right)^\prime \right).$$

where $b = 1$ and $\int g_2(x)f_{y_n-1}(x)dx = \frac{y_n-1}{y_n-1} \log (1 - y_{n-1}) - 1$. Then we obtain

$$\frac{(2/n)T_{n3} - \left[ (p-1) \log \left( \frac{np-n+1}{n} \right) + \log \left( \frac{n-1}{n} \right) - p \frac{y_n-1}{y_n-1} \log (1 - y_{n-1}) + p \right]}{\mu_3} \xrightarrow{d} N \left( \frac{-\log(1 - y)}{2} + \frac{\kappa - 3}{2} \cdot \frac{y}{2} - 2y - 2 \log(1 - y) \right).$$

This completes the proof of Theorem 2.4.

References

John Wiley & Sons.


Figure 1: Plots of $a_j(y)$, $j = 1, 2$ and 3.
Figure 2: Plot of Power Curves at Level 0.05 for Example 1. Left panel is for $\Sigma = \text{diag}(\alpha, I_{p-1})$ and right panel is for $\Sigma = (1 - \tau)I_p + \tau 11'$. 

28
Figure 3: Plot of Power Curves at Level 0.05 for Example 2. Left panel is for $\Sigma = \text{diag}(|R|, I_{p-2})$ and right panel is for $\Sigma = (1 - r)I_p + r11'$. 
Figure 4: Plot of Power Curves at Level 0.05 for Example 3. Left panel is for $\Sigma = 0.5I_p + 0.511' + rR_2$ and right panel is for $\Sigma = 0.5(1+r)I_p + 0.5(1-r)11' + 0.5rR_3$. 
Supplementary material

This supplementary material consists of technical proofs of Lemmas 1, 2 and 3.

**Proof of Lemma A.1.** When \( \kappa = 3 \), Bai and Silverstein (2004) has shown that the random vector \((G_1(f_1), \ldots, G_n(f_k))\) weakly converges to a \( k \)-dimensional Gaussian vector with mean vector,

\[
-\frac{1}{2\pi i} \oint f_j(z) \frac{ym^3(z)(1 + m(z))^{-3}}{(1 - ym^2(z)(1 + m(z))^{-2})^2} \, dz \tag{A.12}
\]

and covariance function

\[
-\frac{1}{2\pi^2} \oint \oint \frac{f_j(z_1)f_l(z_2)}{(m(z_1) - m(z_2))^2} \, dm(z_1) dm(z_2), \quad j, l \in \{1, \ldots, k\} \tag{A.13}
\]

Now we want to prove results in Lemma A.1 still valid when \( \kappa \neq 3 \). By (6.40) and (6.41) of Zheng (2012) replacing \( S^{-1}_2 \) by \( I_p \) and therefore \( F_{yz}(x) = F_{[1, \infty)}(x) \), it follows that when \( X_{11} \) has \( \kappa \neq 3 \), the mean and covariance terms of \( G_n(f_j) \) and \( G_n(f_l) \) have the following additional terms than (A.12) and (A.13), respectively,

\[
(\kappa - 3) \cdot -\frac{1}{2\pi i} \oint f_j(z) \frac{ym^3(z)(1 + m(z))^{-3}}{1 - ym^2(z)(1 + m(z))^{-2}} \, dz,
\]

and

\[
(\kappa - 3) \cdot -\frac{1}{4\pi^2} \oint f_j(z_1)f_l(z_2) \frac{m'(z_1)}{(1 + m(z_1))^2} \frac{m'(z_2)}{(1 + m(z_2))^2} \, dz_1 dz_2.
\]

Thus, we obtain that the random vector \((G_n(f_1), \ldots, G_n(f_k))\) weakly converges to a \( k \)-dimensional Gaussian vector with mean vector,

\[
m(f_j) = -\frac{1}{2\pi i} \oint f_j(z) \frac{ym^3(z)(1 + m(z))^{-3}}{(1 - ym^2(z)(1 + m(z))^{-2})^2} \, dz
+ (\kappa - 3) \cdot -\frac{1}{2\pi i} \oint f_j(z) \frac{ym^3(z)(1 + m(z))^{-3}}{1 - ym^2(z)(1 + m(z))^{-2}} \, dz
= \frac{f_j(a(y)) + f_j(b(y))}{4} \frac{1}{2\pi \int \frac{b(y)}{f_j(x)}} \, dx
- y(\kappa - 3) \cdot \frac{1}{2\pi i} \oint f_j(z) \frac{ym^3(z)(1 + m(z))^{-3}}{1 - ym^2(z)(1 + m(z))^{-2}} \, dz
\]

1
and covariance function

\[
\nu(f_1, f_2) = \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{f_1(z_1)f_2(z_2)}{(m(z_1) - m(z_2))^2} dm(z_1) dm(z_2)
+ (\kappa - 3) \cdot \frac{1}{4\pi^2} \int_{\mathbb{R}} \int_{\mathbb{R}} f_1(z_1)f_2(z_2) \frac{m'(z_1)}{1 + m(z_1)} \frac{m'(z_2)}{1 + m(z_2)} \, dz_1 dz_2
\]

\[
= \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{f_1(z_1)f_2(z_2)}{(m(z_1) - m(z_2))^2} dm(z_1) dm(z_2)
+ \frac{1}{2\pi i} \int_{\mathbb{R}} \frac{f_1(z_1)}{(1 + m(z_1))^2} \, dz_1 \cdot \frac{1}{2\pi i} \int_{\mathbb{R}} \frac{f_2(z_2)}{(1 + m(z_2))^2} \, dz_2
\]

**Proof of Lemma A.2.** Since \( F_{y_{n-1}} \) is the Marčenko-Pastur law of index \( y_{n-1} \), then it follows by setting \( \varepsilon = 1 + y_{n-1} - 2\sqrt{y_{n-1}} \cos \theta \), \( 0 \leq \theta \leq \pi \) that

\[
\int x f_{y_{n-1}}(x) \, dx = \int_{a(y_{n-1})}^{b(y_{n-1})} \frac{x}{2\pi x y_{n-1}} \sqrt{(b(y_{n-1}) - x)(x - a(y_{n-1}))} \, dx
= \frac{1}{2\pi y_{n-1}} \int_{0}^{\pi} 4y_{n-1} \sin^2 \theta \, d\theta = 1
\]

and from P597 of Bai and Silverstein (2004), it follows that

\[
\frac{1}{2\pi} \int_{0}^{2\pi} \frac{2\sin^2 \theta}{1 + y_{n-1} - 2\sqrt{y_{n-1}} \cos \theta} \log |1 - \sqrt{y_{n-1}} e^{i\theta}|^2 \, d\theta = \frac{y_{n-1} - 1}{y_{n-1}} \log(1 - y_{n-1}) - 1.
\]

Thus, it follows that

\[
\int \log(x) f_{y_{n-1}}(x) \, dx = \int_{a(y_{n-1})}^{b(y_{n-1})} \frac{\log x}{2\pi x y_{n-1}} \sqrt{(b(y_{n-1}) - x)(x - a(y_{n-1}))} \, dx
= \frac{1}{2\pi y_{n-1}} \int_{0}^{\pi} \left[ \frac{\log(1 + y_{n-1} - 2\sqrt{y_{n-1}} \cos \theta)}{1 + y_{n-1} - 2\sqrt{y_{n-1}} \cos \theta} \right] 4y_{n-1} \sin^2 \theta \, d\theta
= \frac{1}{2\pi} \int_{0}^{2\pi} \frac{2\sin^2 \theta}{1 + y_{n-1} - 2\sqrt{y_{n-1}} \cos \theta} \log |1 - \sqrt{y_{n-1}} e^{i\theta}|^2 \, d\theta
= \frac{y_{n-1} - 1}{y_{n-1}} \log(1 - y_{n-1}) - 1.
\]

Moreover, we have

\[
\frac{a(y) + b(y)}{4} = \frac{1}{2\pi} \int_{a(y)}^{b(y)} \frac{x}{\sqrt{2y - (x - 1 - y)^2}} \, dx
= \frac{1 + y}{2} - \frac{1}{2\pi} \int_{0}^{\pi} \left[ 1 + y - 2\sqrt{y} \cos \theta \right] \, d\theta
= \frac{1 + y}{2} - \frac{1 + y}{2} = 0
\]
and

\[
\frac{\log(a(y)) + \log(b(y))}{4} = \frac{1}{2\pi} \int_{a(y)}^{b(y)} \frac{\log(x)}{\sqrt{y - (x - 1 - y)^2}} \, dx
\]

\[
= \frac{\log(1 - y)}{2} - \frac{1}{2\pi} \int_{0}^{2\pi} \log(1 + y - 2\sqrt{y} \cos \theta) \, d\theta
\]

\[
= \frac{\log(1 - y)}{2} - \frac{1}{4\pi} \int_{0}^{2\pi} \log|1 - \sqrt{ye^{i\theta}}|^2 \, d\theta
\]

where \( \int_{0}^{2\pi} \log|1 - \sqrt{ye^{i\theta}}|^2 \, d\theta = 0 \) (Bai and Silverstein, 2004).

We know that \( m(z) \) satisfies

\[
z = -\frac{1}{m(z)} + \frac{y}{1 + m(z)}
\]

(see Page 556 of Bai and Silverstein (2004)). Let \( m_i = m(z_i), \quad i = 1, 2 \). For fixed \( m_2 \), we have on a contour enclosing 1, \( (y - 1)^{-1} \) and \( -1 \), but not 0,

\[
\oint \frac{\log(z(m_1))}{(m_1 - m_2)^2} \, dm_1 = \oint \frac{\frac{y}{m_2} - \frac{y}{1 + m_2}}{(m_1 - m_2)^2} \, dm_1
\]

\[
= \oint \frac{(1 + m_1)^2 - ym_1^2}{ym_1(m_1 - m_2)} \left( \frac{-1}{m_1 + 1} + \frac{1}{m_1 - \frac{1}{y - 1}} \right) \, dm_1
\]

\[
= 2\pi i \cdot \left( \frac{1}{m_2 + 1} - \frac{1}{m_2 - \frac{1}{y - 1}} \right)
\]

and

\[
\oint \frac{-\frac{1}{m_1} + \frac{y}{1 + m_1}}{(m_1 - m_2)^2} \, dm_1 = 2\pi i \cdot \frac{y}{(m_2 + 1)^2}
\]

Then

\[
-\frac{1}{2\pi^2} \oint \oint \frac{z_1 z_2}{(m(z_1) - m(z_2))^2} \, dm(z_1) dm(z_2)
\]

\[
= \frac{y^2}{\pi i} \oint \frac{1}{(m_2 + 1)^2} \left( \frac{1}{1 + m_2} + \frac{1 - y}{y} \right) \sum_{j=0}^{\infty} (1 + m_2)^j \, dm_2 = 2y,
\]
and

$$-\frac{1}{2\pi^2} \oint \oint \frac{z_1 \log(z_2)}{(\mu(z_1) - \mu(z_2))^2} \, dm(z_1) \, dm(z_2)$$

$$= \frac{y}{\pi i} \oint \left( \frac{1}{m + 1} - \frac{1}{1 + m} \right) \left( \frac{1}{1 + m} + \frac{1 - y}{y} \right) \cdot \frac{1}{1 - (1 + m)^{-1}} \, dm_2$$

$$= \frac{y}{\pi i} \oint \left( \frac{1}{m + 1} - \frac{1}{1 + m} \right) \left( \frac{1}{1 + m} + \frac{1 - y}{y} \right) \sum_{j=0}^{\infty} \frac{1}{(1 + m)^2} \, dm_2$$

$$= 2y.$$

Moreover, Bai and Silverstein (2004) proved that

$$-\frac{1}{2\pi^2} \oint \oint \frac{\log(z_1) \log(z_2)}{(\mu(z_1) - \mu(z_2))^2} \, dm(z_1) \, dm(z_2) = -2 \log(1 - y).$$

Furthermore, we have

$$\frac{1}{2\pi i} \oint \frac{y m^3(z)(1 + m)^{-3}}{1 - y m^2(z)(1 + m)^{-2}} \, dz$$

$$= \frac{y}{2\pi i} \oint \left( -\frac{1}{m} + \frac{y}{1 + m} \right) \frac{m}{(1 + m)^3} \, dm$$

$$= -\frac{y}{2\pi i} \oint \left( \frac{1}{1 + m} \right)^3 \frac{d m}{1 + m} + \frac{y^2}{2\pi i} \oint \frac{m}{(1 + m)^4} \, dm = 0$$

where $z = -\frac{1}{m} + \frac{y}{1 + m}$ and $\frac{d m(z)}{d z} = \frac{m^2}{1 - y m^2(z)(1 + m)^{-2}}$ (see P596 of Bai and Silverstein (2004)).

$$\frac{1}{2\pi i} \oint \log(z) \frac{y m^3(z)(1 + m)^{-3}}{1 - y m^2(z)(1 + m)^{-2}} \, dz$$

$$= \frac{y}{2\pi i} \cdot \frac{1}{2} \oint \log \left( -\frac{1}{m} + \frac{y}{1 + m} \right) \frac{m^2}{(1 + m)^2} \, dm$$

$$= -\frac{y}{2\pi i} \cdot \frac{1}{2} \oint \frac{m^2}{(1 + m)^2} \, dm$$

$$= -\frac{1}{2\pi i} \cdot \frac{1}{2} \oint \frac{m^2(1 + m)^2 - y m^2}{(1 + m)^2} \, dm$$

$$= \frac{1}{2\pi i} \cdot \frac{1}{2} \oint \frac{m^2(1 + m)^2 - y m^2}{(1 + m)^3} \, dm - \frac{1}{2\pi i} \cdot \frac{1}{2} \oint \frac{m^2(1 + m)^2 - y m^2}{(1 + m)^2} \, dm$$

$$= \frac{3y - 1}{2} - \frac{1}{2y} \cdot \frac{(2y + 1)(y - 1)}{2y} = \frac{y}{2}.$$

$$\frac{1}{2\pi i} \oint \frac{z}{(1 - m(z))^2} \, dm(z) = \frac{1}{2\pi i} \oint \left( -\frac{1}{m} + \frac{y}{1 + m} \right) \frac{1}{(1 + m(z))^2} \, dm(z)$$

$$= \frac{1}{2\pi i} \oint \left( -\frac{1}{m} + \frac{y}{1 + m} \right) \frac{1}{(1 + m(z))^2} \, dm(z) + \frac{1}{2\pi i} \oint \frac{y}{(1 + m(z))^2} \, dm(z)$$

$$= 1.$$
\[
\frac{1}{2\pi i} \oint \frac{\log(z)}{(1+m(z))^2} dm(z) = -\oint \log \left( -\frac{1}{m} + \frac{\nu}{1+m} \right) d\frac{1}{1+m(z)} \\
= \frac{1}{2\pi i} \oint \frac{m^2 - (1+m)^2}{-\frac{1}{m} + \frac{\nu}{1+m}} \frac{1}{1+m(z)} dm(z) \\
= \frac{1}{2\pi i} \oint \frac{(1+m)^2 - \nu m^2}{y m(1+m)} \left( -\frac{1}{1+m} + \frac{1}{m - \frac{1}{\nu-1}} \right) dm(z) \\
= \frac{1}{2\pi i} \oint \frac{(1+m)^2 - \nu m^2}{y m(1+m)} dm(z) + \frac{1}{2\pi i} \oint \frac{(1+m)^2 - \nu m^2}{y m(1+m)} \frac{1}{m - \frac{1}{\nu-1}} dm(z) \\
= 1.
\]

**Proof of Lemma A.3.** Let \( D_j(z) = S_n - \frac{1}{n-1} (X_j - \mu)(X_j - \mu)' - z \cdot I \). Then by (2.9) of Bai and Silverstein (2004), \( D_j^{-1} \) can be replaced by \(-z^{-1}(m(z) \Sigma + I_p)^{-1} \). By (2.17) of Bai and Silverstein (2004), \( b_p(z) \) can replaced by \(-z m(z) \). Then replacing \( S_2^{-1/2} \) by \( \Sigma^{1/2} \), \( b_p(z) \) by \(-z m(z) \) and \( D_{n1}^{-1} \) by \(-z^{-1}(m(z) \Sigma + I_p)^{-1} \) in (6.40) and (6.41) of Zheng (2012), we can show that \( G_n(f) \) converges to a Gaussian vector with mean \( EX_f \) and covariance function \( \text{Cov}(X_f, X_g) \) as in (A.5) and (A.6). This completes the proof of Lemma A.3.