Testing the Diagonality of a Large Covariance Matrix in a Regression Setting

Wei Lan, Ronghua Luo, Chih-Ling Tsai, Hansheng Wang, and Yunhong Yang

Peking University, Southwestern University of Finance and Economics,

University of California–Davis

Abstract

In multivariate analysis, the covariance matrix associated with a set of variables of interest (namely response variables) commonly contains valuable information about the dataset. When the dimension of response variables is considerably larger than the sample size, it is a non-trivial task to assess whether they are linear relationships between the variables. It is even more challenging to determine whether a set of explanatory variables can explain those relationships. To this end, we develop a bias-corrected test to examine the significance of the off-diagonal elements of the residual covariance matrix after adjusting for the contribution from explanatory variables. We show that the resulting test is asymptotically normal. Monte Carlo studies and a numerical example are presented to illustrate the performance of the proposed test.

KEY WORDS: Bias-Corrected Test; Covariance; Diagonality Test; High Dimensional Data; Multivariate Analysis

[†]Wei Lan, Hansheng Wang, and Yunhong Yang are all from the Guanghua School of Management, Peking University, Beijing, 100871, P. R. China. Their research is supported in part by a research grant from PKU-Guanghua Harvest Institute for Investment. Wei Lan and Hansheng Wang are also supported in part by a grant from the Center for Statistical Science at Peking University and a grant from the Fundamental Research Funds for the Central Universities at Renmin University of China. Ronghua Luo is from the School of Finance, South Western University of Finance and Economics. Chih-Ling Tsai is from the Graduate School of Managment, University of California–Davis.

1. INTRODUCTION

Covariance estimation is commonly used to study relationships among multivariate variables. Important applications include, for example, graphical modeling (Edwards, 2000; Drton and Perlman, 2004), longitudinal data analysis (Diggle and Verbyla, 1998; Smith and Kohn, 2002), and risk management (Ledoit and Wolf, 2004) among others. The total number of parameters needed for specifying a covariance matrix of a multivariate vector with dimension p is $p(p+1)/2$. When the sample size n is less than p, the large number of covariance parameters can significantly degrade the statistical efficiency of the usual sample covariance estimator, which makes interpretation difficult. It is therefore important to select the covariance structure so that the number of parameters needing to be estimated is reduced and an easy interpretation can be obtained.

There are a number of regularized estimation methods which have recently been developed to address this issue; current research has focused particularly on identifying various sparse structures; e.g. Huang et al. (2006), Meinshausen and Bühlmann (2006), Lam and Fan (2009), Zhang and Leng (2012) and Leng and Tang (2012). While many novel methods have been developed for covariance estimations, there has not yet been much discussion focusing on statistical tests of the covariance structure. In addition, the classical test statistics developed by John (1971), Nagao (1973), and Anderson (2003) are not applicable for high dimensional data, since the spectral analysis of the sample covariance matrix is inconsistent under a high dimensional setup (Bai and Silverstein, 2005). Efforts to address this problem have led to various tests to determine if the covariance matrix is an identity or a diagonal matrix; see, for example, Srivastava (2005) and Chen et al. (2010). It is noteworthy that Chen et al.'s (2010) test allows for $p \to \infty$ as $n \to \infty$ without imposing the normality assumption; thus it is quite useful for microarray studies (Efron, 2009; Chen et al., 2010). In addition, the aim of their test is to assess whether the covariance matrix exhibits sphericity (i.e., the matrix is proportional to the identity matrix, see Gleser (1966), Anderson (2003) and Onatski et al. (2013)) or identity without controlling for any covariates. As a result, Chen et al.'s (2010) test is not directly applicable for testing diagonality, in particular when explanatory variables are included in the model setting.

In practice, a set of variables of interest (namely, a set of response variables, $Y \in \mathbb{R}^p$) could be explained by another set of explanatory variables, $X \in \mathbb{R}^d$, in a linear form. For example, Fama and French (1993, 1996) introduced three variables to explain the response of stock returns, and the resulting three-factor asset pricing model has been widely used in the fields of finance and economics. To assess the significance of the off-diagonal elements after adjusting for the contribution of explanatory variables, one can naturally adapt the aforementioned methods to test whether the residual covariance matrix, $\text{cov}\{Y - E(Y|X)\}$, is diagonal or not. However, such an approach not only lacks theoretical justification but can also lead to inaccurate or misleading conclusions. This motivates us to develop a test for high dimensional data in a regression setting to investigate whether the residual covariance matrix is diagonal or not. The resulting test can be applied in various fields, such as financial theory (Fan et al., 2008) and microarray analysis (Chen et al., 2010).

The rest of the article is organized as follows. Section 2 introduces the model structure and proposes the bias-corrected test statistic. In addition, the theoretical properties of the resulting test are investigated. Section 3 presents simulation studies to illustrate the finite sample performance of the proposed test. An empirical example is also provided to demonstrate the usefulness of this test. Finally, we conclude the article with a brief discussion in Section 4. All the technical details are left to the

Appendix.

2. THEORETICAL ANALYSIS

2.1. Model Structure and A Test Statistic

Let $Y_i = (Y_{i1}, \cdots, Y_{ip})^\top \in \mathbb{R}^p$ be the p-dimensional response vector collected for the ith subject, where $1 \le i \le n$. For each subject i, we further assume that there exists a d-dimensional explanatory vector $X_i = (X_{i1}, \cdots, X_{id})^\top \in \mathbb{R}^d$. For the remainder of this article, d is assumed to be a fixed number, and $p \to \infty$ as $n \to \infty$ with the possibility that $p \gg n$. Consider the following relationship between Y_i and X_i,

$$Y_i = B^\top X_i + \mathcal{E}_i, \tag{2.1}$$

where $B = (\beta_1, \cdots, \beta_p) \in \mathbb{R}^{d\times p}$, $\beta_j = (\beta_{j1}, \cdots, \beta_{jd})^\top \in \mathbb{R}^d$ are unknown regression coefficients, and $\mathcal{E}_i = (\varepsilon_{i1}, \cdots, \varepsilon_{ip})^\top \in \mathbb{R}^p$ are independent and identically distributed vectors from a multivariate normal distributions with mean vector zero and $\text{cov}(\mathcal{E}_i) = \Sigma = (\sigma_{j_1 j_2})$ for $i = 1, \cdots, n$. For the given dataset $\mathbb{Y} = (Y_1, \cdots, Y_n)^\top \in \mathbb{R}^{n\times p}$ and $\mathbb{X} = (X_1, \cdots, X_n)^\top \in \mathbb{R}^{n\times d}$, we obtain the least squares estimator of the regression coefficient matrix B, $\widehat{B} = (\mathbb{X}^\top\mathbb{X})^{-1}(\mathbb{X}^\top\mathbb{Y}) \in \mathbb{R}^{d\times p}$. Subsequently, the covariance matrix Σ can be estimated by $\hat{\Sigma} = (\hat{\sigma}_{j_1 j_2})$, where $\hat{\sigma}_{j_1 j_2} = n^{-1}\sum_{i=1}^n \hat{\varepsilon}_{ij_1}\hat{\varepsilon}_{ij_2}$, and $\hat{\varepsilon}_{ij_1}$ and $\hat{\varepsilon}_{ij_2}$ are j_1-th and j_2-th components of $\widehat{\mathcal{E}}_i = Y_i - \widehat{B}^\top X_i$, respectively.

To test whether Σ is a diagonal matrix or not, we consider the following null and alternative hypotheses,

$$H_0 : \sigma^2_{j_1 j_2} = 0 \text{ for any } j_1 \neq j_2 \quad \text{vs.} \quad H_1 : \sigma^2_{j_1 j_2} > 0 \text{ for some } j_1 \neq j_2. \tag{2.2}$$

If the null hypothesis is correct, we should expect the absolute value of the off-diagonal element, $\hat{\sigma}_{j_1 j_2}$, to be small for any $j_1 \neq j_2$. Hence, we naturally consider the test statistic, $T^* = \sum_{j_1<j_2} \hat{\sigma}^2_{j_1 j_2}$. Under the null hypothesis, we can further show that $\text{var}^{1/2}(T^*) = O(n^{-3/2}p^{3/2})$ provided that $p/n \to \infty$, which motivates us to propose the following test statistic

$$T = n^{3/2}p^{-3/2} \sum_{j_1<j_2} \hat{\sigma}^2_{j_1 j_2}.$$

Clearly, one should reject the null hypothesis of diagonality if the value of T is sufficiently large. However, we need to develop some theoretical justification to determine what value of T is sufficiently large.

2.2. The Bias of Test Statistic

To understand the asymptotic behavior of the test statistics T, we first compute the expectation of T in the following theorem.

Theorem 1. *Under the null hypothesis H_0, we have*

$$E(T) = \frac{1}{2}\left(\frac{n-d}{n}\right)^{1/2}\left(\frac{n-d}{p}\right)^{1/2}\left(pM^2_{1,p} - M_{2,p}\right),$$

where $M_{\kappa,p} = p^{-1}\sum_{j=1}^{p} \sigma^{\kappa}_{jj}$ for $\kappa = 1$ and 2.

The proof is given in Appendix A. Theorem 1 indicates that $E(T)$ is not exactly zero, which yields some bias. To further investigate the property of bias, we assume that $M_{\kappa,p} \to M_\kappa$ as $p \to \infty$ for some $|M_\kappa| < \infty$. Then, $E(T) \approx M_1^2\sqrt{np} \to \infty$. As mentioned earlier, under the null hypothesis, we have $\text{var}^{1/2}(T^*) = O(n^{-3/2}p^{3/2})$ if $p/n \to \infty$, which leads to $\text{var}(T) = O(1)$. Accordingly, $E(T)/\text{var}^{1/2}(T) = O(\sqrt{np}) \to \infty$, which suggests that $T/\{\text{var}(T)\}^{1/2}$ is not asymptotically distributed as a standard normal random variable. This implies that we cannot ignore the bias due to T in asymptotic test, so we need to turn to methods of bias correction. To this end, we obtain an unbiased estimator of $E(T)$ as given below.

Theorem 2. *Under H_0, we have $E(\widehat{Bias}) = E(T)$, where*

$$\widehat{Bias} = \frac{n^{3/2}}{2(n-d)p^{3/2}}\Big[tr^2(\hat{\Sigma}) - tr(\hat{\Sigma}^{(2)})\Big] \quad and \quad \hat{\Sigma}^{(2)} = (\hat{\sigma}^2_{j_1 j_2}). \tag{2.3}$$

The proof is given in Appendix B. Theorem 2 shows that $\widehat{\text{Bias}}$ is an unbiased estimator of $E(T)$. This motivates us to consider the bias-corrected statistic, $T - \widehat{\text{Bias}}$, whose asymptotic properties will be presented in the following subsection.

2.3. The Bias-Corrected (BC) Test Statistic

After adjusting T by its bias estimator $\widehat{\text{Bias}}$, we next study its variance.

Theorem 3. *Assume that* $\min\{n, p\} \to \infty$ *and* $M_{\kappa,p} = p^{-1}\sum_{j=1}^{p}\sigma_{jj}^{\kappa} \to M_\kappa$ *for some constant* $|M_\kappa| < \infty$ *and for all* $\kappa \leq 4$. *Under H_0, then* $var(T - \widehat{Bias}) = (n/p)M_{2,p}^2 + o(n/p)$.

The proof is given in Appendix C. Theorem 3 demonstrates that $\text{var}(T - \widehat{\text{Bias}}) = O(n/p)$ and we can show that its associated term $M_{2,p}$ can be estimated by the following unbiased estimators

$$\widehat{M}_{2,p} = n^2 p^{-1}\Big[(n-d)^2 + 2(n-d)\Big]^{-1}\sum_{j=1}^{p}\hat{\sigma}_{jj}^2.$$

This drives us to consider the following bias-corrected (BC) test statistic.

$$Z = \frac{T - \widehat{\text{Bias}}}{(n/p)^{1/2}\widehat{M}_{2,p}}, \tag{2.4}$$

whose asymptotic normality is established below.

Theorem 4. *Assume that* $\min\{n, p\} \to \infty$ *and* $M_{\kappa,p} = p^{-1}\sum_{j=1}^{p}\sigma_{jj}^{\kappa} \to_p M_\kappa$ *for some constant* $|M_\kappa| < \infty$ *and for all* $\kappa \leq 4$. *Under H_0, we have* $Z \to_d N(0, 1)$.

The proof is given in Appendix D. Theorem 4 indicates that the asymptotic null distribution of Z is standard normal, as long as $\min\{n, p\} \to \infty$. Applying this theorem, we are able to test the significance of the off-diagonal elements. Specifically, for a given significance level α, we reject the null hypotheses of diagonality if $Z > z_{1-\alpha}$, where Z is the test statistics given in (2.4) and z_α stands for the αth quantile of a standard normal distribution. Simulation studies, reported in the next section, suggest that such a testing procedure can indeed control the empirical size very well. It is noteworthy that Nagao's (1973) diagonal test is valid only when p is fixed. To accommodate high dimensional data, Schott (2005) developed a testing procedure via the correlation coefficient matrix, which is useful if one is interested in testing whether $\mathrm{cov}(Y)$ is diagonal. However, it cannot be directly applied to test the diagonality of $\mathrm{cov}\{Y - E(Y|X)\}$, unless the predictor dimension d is appropriately taken into consideration; see Remarks 2 and 3 below for detailed discussions.

Remark 1. In multivariate models, researchers (e.g., Anderson, 2003, and Schott, 2005) have proposed various methods to test whether or not the covariance matrix is diagonal. It is interesting to recall that Anderson (2003) introduced the likelihood ratio test in the field of factor analysis as a method of examining the number of factors. In fact, identifying the number of common factors is similar to testing the diagonality of the covariance matrix of the specific factor. This leads us to propose our approach for testing whether the covariance matrix of the error vector is diagonal, after controlling for the effect of explanatory variables.

Remark 2. By Theorem 2, $\widehat{\mathrm{Bias}}$ is an exactly unbiased estimator of T, and it contains the quantity $(n - d)$. For the sake of simplicity, one may consider replacing $(n - d)$ in the denominator of (2.3) by n so that the multiplier of $\widehat{\mathrm{Bias}}$ becomes $n^{1/2}/(2p^{3/2})$. Under this replacement, however, $E(T - \widehat{\mathrm{Bias}}) \neq 0$ and it is of the order $O(n^{1/2}/p^{1/2})$, which has the same order as $(n/p)^{1/2}\widehat{M_{2,p}}$ given in the denominator of (2.4). Hence, the resulting test statistic is no longer distributed as a standard normal. This suggests that the predictor dimension (i.e., d) plays an important role for bias correction in our proposed BC-test statistic.

Remark 3. Although the BC-test in (2.4) shares some merits with the Schott (2005) test, there are three major differences given below. First, the BC-test considers $\min\{p, n\} \to \infty$, while the Schott test assumes that $p/n \to c$ for some finite constant $c > 0$. Second, the BC-test takes the predictors into consideration, which is not the focus of the Schott test. Third, the BC-test is obtained from the covariance matrix. In contrast, the Schott test is constructed via the correlation matrix and it is scale invariant. It is not surprising that the asymptotic theory of the Schott test is more sophisticated than that of the BC-test. According to an anonymous referee's suggestion as well as an important finding in Remark 2, we have carefully extended the Schott test to the model with predictors. We name it the Adjusted Schott (AS) test, which is

$$Z_{\text{adj}} = \sqrt{\frac{(n-d)^2(n-d+2)}{(n-d-1)p(p-1)}}\left[\sum_{j_1<j_2} \hat{r}_{j_1j_2}^2 - \frac{p(p-1)}{2(n-d)}\right], \tag{2.5}$$

where $\hat{r}_{j_1j_2} = \hat{\sigma}_{j_1j_2}/\{\hat{\sigma}_{j_1j_1}^{1/2}\hat{\sigma}_{j_2j_2}^{1/2}\}$. Following the techniques of Schott (2005) with slightly complicated calculations, we are able to demonstrate that Z_{adj} is asymptotic standard normal under the null hypothesis. However, its validity is established only when $p/n \to c$ for some finite constant $c > 0$, as assumed by Schott (2005). In high dimensional data with $p \gg n$, the asymptotic theory is far more complicated and needs further investigation.

3. NUMERICAL STUDIES

3.1. Simulation Results

To evaluate the finite sample performance of the bias-corrected test, we conduct Monte Carlo simulations. We consider model (2.1), where the predictor $X_i = (X_{ij})$ is generated from a multivariate normal distribution with $E(X_{ij}) = 0$, $\text{var}(X_{ij}) = 1$, and $\text{cov}(X_{ij_1}, X_{ij_2}) = 0.5$ for any $j_1 \neq j_2$. In addition, the regression coefficients β_{jk} are independently generated from a standard normal distribution. Moreover, the error vector $\mathcal{E}_i = (\varepsilon_{ij})$ is generated as follows: (i.) the ε_{ij} are generated from normal distributions with mean 0; (ii.) the variance of ε_{ij} (i.e., σ_{jj}) is simulated independently from a uniform distribution on [0,1]; (iii.) the correlation between ε_{ij_1} and ε_{ij_2} for any $j_1 \neq j_2$ is fixed to be a constant ρ.

We simulated 1,000 realizations with a nominal level of 5%, each consisting of two sample sizes ($n = 100$ and 200), three dimensions of multivariate responses ($p = 100$, 500, and 1,000), and four dimensions of explanatory variables ($d = 0$, 1, 3, and 5). The value of $\rho = 0$ corresponds to the null hypothesis. Schott (2005) developed a diagonal test in high dimensional data under $d = 0$. For the sake of comparison, we include the Schott (2005, Section 2) test by calculating the sample correlation with the estimated residual $\widehat{\mathcal{E}}_i$ rather than the response Y_i. In addition, we consider the Adjusted Schott test given in (2.5).

Under normal errors, Table 1 reports the sizes of the BC, Schott, and Adjusted Schott tests. When $d = 0$, all three tests perform well. However, the performance of the Schott test deteriorates with $d > 0$. This indicates that the Schott test cannot be directly applied to assess the diagonality of the residual covariance matrix after incorporating the contribution from explanatory variables. In contrast, after adjusting for the effective sample size from n to $n-d$, the performance of the Adjusted Schott test becomes satisfactory, which indicates that the predictor dimension d is indeed critical.

Figure 1: Power functions for testing $H_0 : \rho = 0$ with normal errors and $n = 100$.

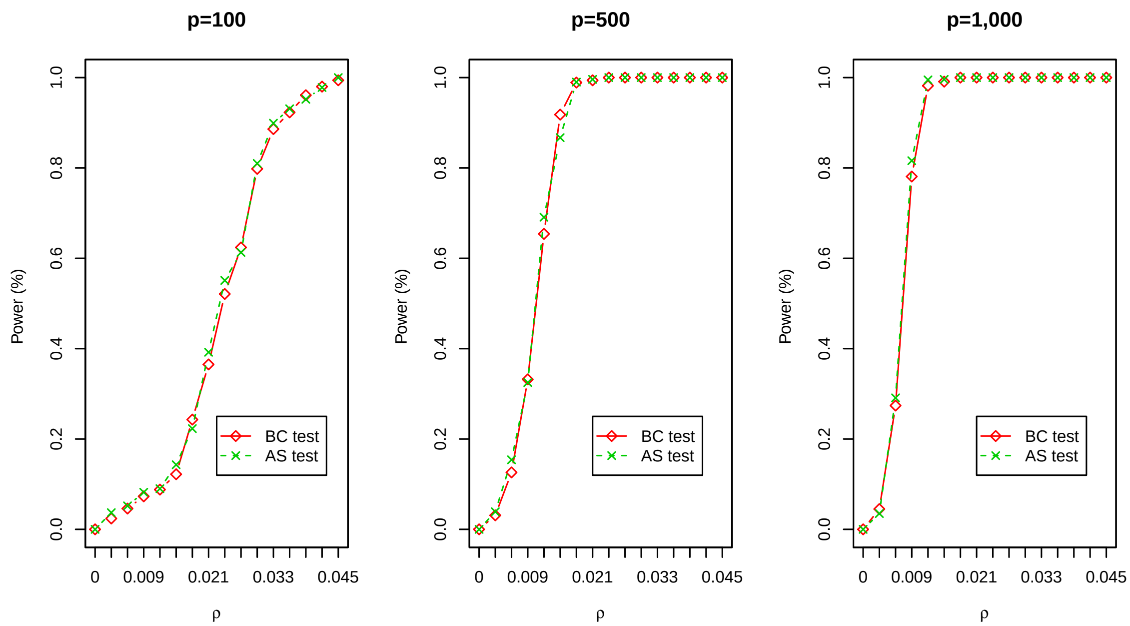

Table 1 also indicates that the BC test controls the empirical sizes well across the various sample sizes, dimensions of response variables, and dimensions of explanatory variables. To examine their robustness, we also generated errors from the Student (t_5), mixture ($0.9N(0, 1) + 0.1N(0, 3^2)$), and Gamma(4,0.5) distributions. Table 1 shows that, under these error distributions, both tests control the empirical size adequately. Finally, we investigate the power of the BC test and the AS test. For the sake of illustration, we consider only the case with normal errors, $n = 100$, and $d = 1$. Figure 1 depicts the power functions for three dimensions of response variables ($p = 100$, 500 and $1,000$) respectively. It shows that the powers of the two tests are almost identical, and the power increases as p becomes large. Since all simulation results for $n = 200$ show a similar pattern, we do not report them here. In sum, both BC and AS tests are reliable and powerful, while the theoretical justification for the AS test needs further study as mentioned in Remark 3.

Finally, upon the suggestion from the AE and an anonymous referee, we have compared our proposed test with Chen et al.'s (2010) and the tests mentioned in Onatski et al. (2013) for testing high dimensional covariances via simulation studies with $d = 0$. The results show that all these tests perform well in testing for identity under normal error, while BC test is superior to Chen et al's and Onatski et al's tests for testing diagonality. This finding is not surprising since Chen et al's and Onatski et al's tests are not developed for examining diagonality.

3.2. A Real Example

To further demonstrate the practical usefulness of our proposed method, we consider an important finance application. Specifically, we employ our method to address an critical question: how many common factors (i.e., explanatory variables) are needed to fully describe the covariance structure of security returns? By Trzcinka (1986), Brown (1989), and Connor and Korajczky (1993), this is one of the fundamental problems in portfolio theory and asset pricing. To this end, we collect the data from a commercial database, which contains weekly returns for all of the stocks traded on the Chinese stock market during the period of 2008–2009. After eliminating those stocks with missing values, we obtain a total of $p = 1,058$ stocks with sample size $n = 103$.

We consider as our explanatory variables, several of the factors most commonly used in the finance literature to explain the covariance structure of stock returns. The first such factor is X_{i1} = market index, in this case, returns for the Shanghai Composite Index. The market index is clearly the most important factor for stock returns because it reflects the overall performance of the stock market. As a result, it serves as the foundation for the Capital Asset Pricing Model (Sharpe, 1964; Lintner, 1965; Mossin, 1966). Empirical studies, however, have suggested that the market index alone cannot fully explain the correlation structure of stock returns. Fama and French (1993, 1996) proposed the Three-Factor model to address this problem; they include the market index as well as two other factors which are denoted by $X_{i2} = \text{SMB}$ and $X_{i3} = \text{HML}$. Specifically, X_{i2} measures the size premium (i.e., the difference in returns between portfolios of small capitalization firms and large capitalization firms) and X_{i3} is the book-to-market premium (i.e., the difference in returns between portfolios of high book-to-market firms and low book-to-market firms). Finally, recent advances in behavioral finance suggest that stock returns have non-trivial momentum, which is captured by the difference in returns between portfolios of high and low prior returns. To this end, Jegadeesh and Titman (1993) and Carhart (1997) proposed the momentum factor, which is denoted by X_{i4}.

In our analysis, we consider four nested models, $\mathcal{M}_0 = \emptyset$, $\mathcal{M}_1 = \{X_{i1}\}$, $\mathcal{M}_2 = \{X_{i1}, X_{i2}, X_{i3}\}$, and $\mathcal{M}_3 = \{X_{i1}, X_{i2}, X_{i3}, X_{i4}\}$ and apply the proposed method to each candidate model; this gives test statistics of 20,560, 3,357, 228, and 215, respectively. Similar results are obtained via the AS test. We draw two conclusions from these results. The first comes from observing the differences between these values. As expected, the Fama-French model ($\mathcal{M}_2$) improves enormously on both the model with no predictors and the model with only the market index, and while the fourth factor (momentum) does improve on the Fama-French model, its contribution is clearly small. The proposed statistical method, therefore, provides additional confirmation that the Fama-French model is an extremely important finance model, even in datasets with $p > n$. Secondly, the addition of a fourth factor still does not allows us to accept the null hypothesis of a diagonal covariance matrix. This suggests that there may be factors unique to the Chinese stock market which contribute significantly to the covariance structure. To explore this idea further, we applied the principle component method of factor analysis to the residuals of $\mathcal{M}_3$ and found that the test statistic continued to decline with the inclusion of as many as 75 of the additional factors we identified. While this additional finding is interesting, it lacks of insightful financial interpretations, and so we believe that further research on risk factors in the Chinese stock market is warranted.

4. DISCUSSIONS

In this paper, We propose a bias-corrected test to assess the significance of the off-diagonal elements of a residual covariance matrix for high-dimensional data. This test takes into account the information from explanatory variables, which broadens the application of covariance analysis. Although the results are developed in the context of a linear regression, it could certainly be extended to nonparametric regressions; see for example, Fan and Gijbels (1996), Härdle et al. (2000), and Xia (2008). In the theoretical development of our test statistic, we focus principally on the normal error assumption. It could, however, be useful to obtain a test for diagonality with weaker assumptions such as sub-Gaussian errors, although simulation studies show that the proposed test performs well for non-normal errors. Moreover, it would be interesting to extend our test to the correlation matrix with $d > 0$ and $\min\{n, p\} \to \infty$. Finally, a generalization of the test statistic to the case with $d > n$ could be interest. In this context, the shrinkage regression estimates, e.g., LASSO of Tibshirani (1996), SCAD of Fan and Li (2001) and Group LASSO of Yuan and Lin (2006), could be useful in developing a test statistic. Based on our limited studies, obtaining a shrinkage estimator of B that is consistent for variable selection will be important in extending our proposed test. With a consistently selected model, we believe that the usual OLS-type estimator and the theoretical development of the resulting test statistic can be achieved. Consequently, it is essential to develop an effective shrinkage method for both $d > n$ and $n \to \infty$.

APPENDIX

Appendix A. Proof of Theorem 1

To facilitate the proof, we will refer to the following lemmas, so we present them first. The proof of Lemma 2 follows directly from the proof of Lemma 3 in Magnus (1986). Its proof is therefore omitted.

Lemma 1. *Let U_1 and U_2 be two $m \times 1$ independent random vectors with mean vector 0 and covariance matrix I_m, where I_m is an $m \times m$ identity matrix. Then for any $m \times m$ projection matrix A, we have (a) $E(U_1^\top AU_1) = tr(A)$ and (b) $E[(U_1^\top AU_2)^2] = tr(A)$. Further assume U_1 and U_2 follow multivariate normal distributions, then we have (c) $E[(U_1^\top AU_2)^2 U_1^\top AU_1 U_2^\top AU_2] = 4tr(A) + 4tr^2(A) + tr^3(A)$, (d) $E[(U_1^\top AU_1)^2] = tr^2(A) + 2tr(A)$, and (e) $E[(U_1^\top AU_2)^4] = 3tr^2(A) + 6tr(A)$.*

Proof. The proofs of (a) and (b) are straightforward, and are therefore omitted. In addition, results (d) and (e) can be directly obtained from Proposition 1 of Chen et al. (2010). As a result, we only need to show part (c). Using the fact that $U_1^\top AU_2 \in \mathbb{R}^1$, $U_1^\top AU_1 \in \mathbb{R}^1$, $U_2^\top AU_2 \in \mathbb{R}^1$, $U_1^\top AU_2 = U_2^\top AU_1$, and U_1 and U_2 are mutually independent, we have

$$
\begin{aligned}
E\Big\{(U_1^\top AU_2)^2 U_1^\top AU_1 U_2^\top AU_2\Big\} &= tr\Big\{E\Big(U_1^\top AU_1 AU_1 U_1^\top\Big) E\Big(U_2^\top AU_2 AU_2 U_2^\top\Big)\Big\} \\
&= tr\Big\{E\Big(AU_1U_1^\top\Big)^2 E\Big(AU_2U_2^\top\Big)^2\Big\}. \qquad \text{(A.1)}
\end{aligned}
$$

Next, let $A = (a_{ij})$ and $U_1 = (U_{1j})$, where $a_{ij} = a_{ji}$ since A is a projection matrix. Then we have $(AU_1U_1^\top)^2 = AU_1(U_1^\top AU_1)U_1^\top = (U_1^\top AU_1)AU_1U_1^\top = (\tilde{a}_{ij})$, where

$$
\tilde{a}_{ij} = \left(\sum_{k=1}^{m}\sum_{h=1}^{m} a_{kh}U_{1k}U_{1h}\right)\left(\sum_{l=1}^{m} a_{il}U_{1l}\right)U_{1j}.
$$

Using the fact that U_1 is a m-dimensional standard normal random vector, we obtain

$$E(\tilde{a}_{ij}) = 2 \sum_{1 \le k \le m}^{k \ne j} a_{ik} a_{kj} + a_{ij} \sum_{1 \le k \le m}^{k \ne j} a_{kk} + 3 a_{ij} a_{jj}$$

$$= 2 \sum_{1 \le k \le m} a_{ik} a_{kj} + a_{ij} \sum_{1 \le k \le m} a_{kk}.$$

As a result, we have $E(AU_1U_1^\top)^2 = 2A^2 + tr(A) \cdot A$. Because U_1 and U_2 are independent identically distributed random variables and A is a projection matrix, the right-hand side of (A.1) is equal to $tr(2A^2 + tr(A) \cdot A)^2 = 4tr(A^4) + 4tr(A^3)tr(A) + tr^2(A)tr(A^2) = 4tr(A) + 4tr^2(A) + tr^3(A)$. This completes the proof.

Lemma 2. *Let U be an $m \times 1$ normally distributed random vector with mean vector 0 and covariance matrix I_m, and let A be a $m \times m$ symmetric matrix. Then, for the fixed integer s, we have that,*

$$E(U^\top AU)^s = \sum_v \gamma_s(v) \prod_{j=1}^{s} \left\{ tr(A^j) \right\}^{n_j}, \text{ where } \gamma_s(v) = s! 2^v \prod_{j=1}^{s} \left[n_j! (2j)^{n_j} \right]^{-1},$$

with the summation over all possible $v = (n_1, n_2, \cdots, n_s)^\top \in \mathbb{R}^s$ such that $\sum_{j=1}^s jn_j = s$ and n_j $(1 \le j \le s)$ is a nonnegative integer.

Proof of Theorem 1. Let $\varepsilon_j = (\varepsilon_{1j}, \varepsilon_{2j}, \cdots, \varepsilon_{nj})^\top \in \mathbb{R}^n$ for $j = 1, \cdots, d$. From the regression model (2.1), we know that ε_j has mean 0 and covariance matrix $\sigma_{jj} I$, where $I \in \mathbb{R}^{n \times n}$ is a $n \times n$ matrix. Furthermore, the j-th residual and the (j_1, j_2)-th estimator of $\hat{\Sigma}$ are $\hat{\varepsilon}_j = (I - H)\varepsilon_j$ and $\hat{\sigma}_{j_1 j_2} = n^{-1} \hat{\varepsilon}_{j_1}^\top \hat{\varepsilon}_{j_2} = n^{-1} \varepsilon_{j_1}^\top (I - H) \varepsilon_{j_2}$, respectively, where $H = \mathbb{X}(\mathbb{X}^\top \mathbb{X})^{-1} \mathbb{X}^\top \in \mathbb{R}^{n \times n}$ is an $n \times n$ projection matrix. Under H_0, ε_{j_1} and ε_{j_2} are independent if $j_1 \ne j_2$. Applying Lemma 1(b), we have that

$E(\hat{\sigma}^2_{j_1 j_2}) = n^{-2}(n-d)\sigma_{j_1 j_1}\sigma_{j_2 j_2}$. Accordingly, we obtain that

$$\begin{aligned} E(T) &= E\bigg\{ n^{3/2} p^{-3/2} \sum_{j_1<j_2} \hat{\sigma}^2_{j_1 j_2} \bigg\} = \left(\frac{n-d}{\sqrt{n}}\right) p^{-3/2} \sum_{j_1<j_2} \sigma_{j_1 j_1}\sigma_{j_2 j_2} \\ &= \left(\frac{n-d}{2\sqrt{n}}\right) p^{-3/2} \bigg\{ \sum_{j_1=1}^{p}\sum_{j_2=1}^{p} \sigma_{j_1 j_1}\sigma_{j_2 j_2} - \sum_{j=1}^{p} \sigma^2_{jj} \bigg\} \\ &= \frac{1}{2}\left(\frac{n-d}{n}\right)^{1/2}\left(\frac{n-d}{p}\right)^{1/2}\Big(pM^2_{1,p} - M_{2,p}\Big). \end{aligned} \tag{A.2}$$

This completes the proof.

Appendix B. Proof of Theorem 2

By Lemma 1(a), we have that $E(\hat{\sigma}_{jj}) = n^{-1}(n-d)\sigma_{jj}$. Then, using the fact that $\hat{\sigma}_{j_1 j_1}$ and $\hat{\sigma}_{j_2 j_2}$ are uncorrelated, we are able to calculate $E[tr^2(\hat{\Sigma})]$ as follows.

$$\begin{aligned} E\Big[tr^2(\hat{\Sigma})\Big] &= E\bigg[\sum_{j=1}^{p} \hat{\sigma}_{jj}\bigg]^2 = E\bigg\{ \sum_{j_1\neq j_2} \hat{\sigma}_{j_1 j_1}\hat{\sigma}_{j_2 j_2} + \sum_{j=1}^{p} \hat{\sigma}^2_{jj} \bigg\} \\ &= \left(\frac{n-d}{n}\right)^2 \sum_{j_1\neq j_2} \sigma_{j_1 j_1}\sigma_{j_2 j_2} + E\bigg(\sum_{j=1}^{p} \hat{\sigma}^2_{jj}\bigg) \\ &= p\left(\frac{n-d}{n}\right)^2 \Big(pM^2_{1,p} - M_{2,p}\Big) + E\Big[tr(\hat{\Sigma}^{(2)})\Big]. \end{aligned}$$

This, together with (A.2), implies that

$$\frac{n^{3/2}}{2(n-d)p^{3/2}} \bigg\{ E\Big[tr^2(\hat{\Sigma})\Big] - E\Big[tr(\hat{\Sigma}^{(2)})\Big] \bigg\} = E(T),$$

which completes the proof.

Appendix C. Proof of Theorem 3

Note that $\mathrm{var}\{T - \widehat{\mathrm{Bias}}\} = E\{\widehat{\mathrm{Bias}}^2\} + E\{T^2\} - 2E\{T\widehat{\mathrm{Bias}}\}$, where the right-hand side of this equation contains three components. They can be evaluated separately according to the following three steps.

Step 1. We first compute $E\{\widehat{\text{Bias}}^2\}$. Recall that $\widehat{\text{Bias}} = 2^{-1}(n-d)^{-1}n^{3/2}p^{-3/2}\{tr^2(\hat{\Sigma}) - tr(\hat{\Sigma}^{(2)})\}$ defined in Theorem 2, we then obtain

$$\widehat{\text{Bias}}^2 = 4^{-1}(n-d)^{-2}n^3p^{-3}\sum_{i_1\neq j_1}\sum_{i_2\neq j_2}\hat{\sigma}_{i_1i_1}\hat{\sigma}_{j_1j_1}\hat{\sigma}_{i_2i_2}\hat{\sigma}_{j_2j_2}. \tag{A.3}$$

Because $\hat{\sigma}_{ii} = n^{-1}\hat{\varepsilon}_i^\top\hat{\varepsilon}_i = n^{-1}\varepsilon_i^\top(I-H)\varepsilon_i$, we have $\hat{\sigma}_{i_1i_1}\hat{\sigma}_{j_1j_1}\hat{\sigma}_{i_2i_2}\hat{\sigma}_{j_2j_2} = n^{-4}\varepsilon_{i_1}^\top(I-H)\varepsilon_{i_1}\varepsilon_{j_1}^\top(I-H)\varepsilon_{j_1}\varepsilon_{i_2}^\top(I-H)\varepsilon_{i_2}\varepsilon_{j_2}^\top(I-H)\varepsilon_{j_2}$. According to the values of i_1, i_2, j_1, and j_2, we subsequently calculate the expectation of $\hat{\sigma}_{i_1i_1}\hat{\sigma}_{j_1j_1}\hat{\sigma}_{i_2i_2}\hat{\sigma}_{j_2j_2}$ according to the following three cases.

Case I: i_1, i_2, j_1, and j_2 are mutually different. Then, $\hat{\sigma}_{i_1i_1}$, $\hat{\sigma}_{j_1j_1}$, $\hat{\sigma}_{i_2i_2}$, and $\hat{\sigma}_{j_2j_2}$ are mutually independent. By Lemma 1(a), we have

$$E\Big(\hat{\sigma}_{i_1i_1}\hat{\sigma}_{j_1j_1}\hat{\sigma}_{i_2i_2}\hat{\sigma}_{j_2j_2}\Big)$$

$$= n^{-4}E\Big\{\varepsilon_{i_1}^\top(I-H)\varepsilon_{i_1}\Big\}E\Big\{\varepsilon_{j_1}^\top(I-H)\varepsilon_{j_1}\Big\}E\Big\{\varepsilon_{i_2}^\top(I-H)\varepsilon_{i_2}\Big\}E\Big\{\varepsilon_{j_2}^\top(I-H)\varepsilon_{j_2}\Big\}$$
$$= n^{-4}\sigma_{i_1i_1}\sigma_{j_1j_1}\sigma_{i_2i_2}\sigma_{j_2j_2}\Big\{tr(I-H)\Big\}^4 = n^{-4}(n-d)^4\sigma_{i_1i_1}\sigma_{j_1j_1}\sigma_{i_2i_2}\sigma_{j_2j_2}. \tag{A.4}$$

Case II: $i_1 = i_2$, $j_1 = j_2$, but $i_1 \neq j_1$. Then, $\hat{\sigma}_{i_1i_1}\hat{\sigma}_{j_1j_1}\hat{\sigma}_{i_2i_2}\hat{\sigma}_{j_2j_2} = \hat{\sigma}_{i_1i_1}^2\hat{\sigma}_{j_1j_1}^2$. By Lemma 1(d), we obtain that

$$E\Big(\hat{\sigma}_{i_1i_1}\hat{\sigma}_{j_1j_1}\hat{\sigma}_{i_2i_2}\hat{\sigma}_{j_2j_2}\Big) = n^{-4}E\Big\{\varepsilon_{i_1}^\top(I-H)\varepsilon_{i_1}\Big\}^2E\Big\{\varepsilon_{j_1}^\top(I-H)\varepsilon_{j_1}\Big\}^2$$

$$= n^{-4}\sigma_{i_1i_1}^2\sigma_{j_1j_1}^2\Big\{tr^2(I-H) + 2tr(I-H)\Big\}^2$$

$$= n^{-4}\sigma_{i_1i_1}^2\sigma_{j_1j_1}^2(n-d)^2(n-d+2)^2. \tag{A.5}$$

CASE III: $i_1 = i_2$, but i_1, j_1, and j_2 are mutually different. Then, $\hat{\sigma}_{i_1i_1}\hat{\sigma}_{j_1j_1}\hat{\sigma}_{i_2i_2}\hat{\sigma}_{j_2j_2} = \hat{\sigma}^2_{i_1i_1}\hat{\sigma}_{j_1j_1}\hat{\sigma}_{j_2j_2}$. By Lemma 1(a) and 1(d), we have

$$E\Big(\hat{\sigma}_{i_1i_1}\hat{\sigma}_{j_1j_1}\hat{\sigma}_{i_2i_2}\hat{\sigma}_{j_2j_2}\Big) = n^{-4}E\Big\{\varepsilon_{i_1}^\top(I-H)\varepsilon_{i_1}\Big\}^2 E\Big\{\varepsilon_{j_1}^\top(I-H)\varepsilon_{j_1}\Big\}E\Big\{\varepsilon_{j_2}^\top(I-H)\varepsilon_{j_2}\Big\}$$

$$= n^{-4}\sigma^2_{i_1i_1}\sigma_{j_1j_1}\sigma_{j_2j_2}(n-d)^3(n-d+2). \tag{A.6}$$

This, together with (A.3), (A.4), and (A.5), leads to

$$4p^3n^{-1}E\{\widehat{\text{Bias}}^2\} = \left(\frac{n-d}{n}\right)^2\mathcal{S}_1 + 2\left(\frac{n-d+2}{n}\right)^2\mathcal{S}_2$$

$$+4\left(\frac{n-d}{n}\right)\left(\frac{n-d+2}{n}\right)\mathcal{S}_3, \tag{A.7}$$

where

$$\mathcal{S}_1 = \sum_{i_1\neq j_1\neq i_2\neq j_2}\sigma_{i_1i_1}\sigma_{j_1j_1}\sigma_{i_2i_2}\sigma_{j_2j_2},$$

$$\mathcal{S}_2 = \sum_{i\neq j}\sigma^2_{ii}\sigma^2_{jj},$$

$$\text{and}\quad \mathcal{S}_3 = \sum_{i\neq j_1}\sum_{i\neq j_2}\sigma^2_{ii}\sigma_{j_1j_1}\sigma_{j_2j_2}.$$

Step 2. We next consider $E(T^2)$. Using the fact that $T = 2^{-1}n^{3/2}p^{-3/2}\sum_{i\neq j}\hat{\sigma}^2_{ij}$ and $\hat{\sigma}_{ij} = n^{-1}\varepsilon_i^\top(I-H)\varepsilon_j$, we have

$$T^2 = 4^{-1}n^3p^{-3}\sum_{i_1\neq j_1}\sum_{i_2\neq j_2}\hat{\sigma}^2_{i_1j_1}\hat{\sigma}^2_{i_2j_2}$$

$$= 4^{-1}n^{-1}p^{-3}\sum_{i_1\neq j_1}\sum_{i_2\neq j_2}\left(\varepsilon_{i_1}^\top(I-H)\varepsilon_{j_1}\right)^2\left(\varepsilon_{i_2}^\top(I-H)\varepsilon_{j_2}\right)^2.$$

Applying the same procedure as that used in Step 1, we compute $E(T^2)$ according to the following three different cases.

CASE I. i_1, i_2, j_1, and j_2 are mutually different. Then, $\hat{\sigma}_{i_1j_1}$ and $\hat{\sigma}_{j_2j_2}$ are mutually independent. By Lemma 1(b), we have

$$E\Big(\varepsilon_{i_1}^\top(I-H)\varepsilon_{j_1}\Big)^2\Big(\varepsilon_{i_2}^\top(I-H)\varepsilon_{j_2}\Big)^2 = E\Big(\varepsilon_{i_1}^\top(I-H)\varepsilon_{j_1}\Big)^2 E\Big(\varepsilon_{i_2}^\top(I-H)\varepsilon_{j_2}\Big)^2$$

$$= (n-d)^2\sigma_{i_1i_1}\sigma_{j_1j_1}\sigma_{i_2i_2}\sigma_{j_2j_2}. \tag{A.8}$$

CASE II. $i_1 = i_2$, $j_1 = j_2$, but $i_1 \neq j_1$. Then, $\hat{\sigma}_{i_1j_1}^2\hat{\sigma}_{i_2j_2}^2 = \hat{\sigma}_{i_1j_1}^4$. By Lemma 1(e), we obtain that

$$E\Big(\varepsilon_{i_1}^\top(I-H)\varepsilon_{j_1}\Big)^2\Big(\varepsilon_{i_2}^\top(I-H)\varepsilon_{j_2}\Big)^2 = E\Big(\varepsilon_{i_1}^\top(I-H)\varepsilon_{j_1}\Big)^4$$

$$= \sigma_{i_1i_1}^2\sigma_{j_1j_1}^2\Big\{3tr^2(I-H) + 6tr(I-H)\Big\} = 3(n-d)(n-d+2)\sigma_{i_1i_1}^2\sigma_{j_1j_1}^2. \tag{A.9}$$

CASE III. $i_1 = i_2$, but i_1, j_1, and j_2 are mutually different. Because $\{\varepsilon_{i_1}^\top(I-H)\varepsilon_{j_2}\}^2$ and $\varepsilon_{i_1}^\top(I-H)\varepsilon_{i_1}$ are scalars, we apply Lemma 1(d) and have

$$E\Big(\{\varepsilon_{i_1}^\top(I-H)\varepsilon_{j_1}\}^2\{\varepsilon_{i_1}^\top(I-H)\varepsilon_{j_2}\}^2\Big)$$

$$= E\Big[tr\Big(\{\varepsilon_{j_1}^\top(I-H)\varepsilon_{i_1}\}\{\varepsilon_{i_1}^\top(I-H)\varepsilon_{j_1}\}\{\varepsilon_{i_1}^\top(I-H)\varepsilon_{j_2}\}^2\Big)\Big]$$

$$= tr\Big\{E\Big(\varepsilon_{j_1}\varepsilon_{j_1}^\top(I-H)\varepsilon_{i_1}\varepsilon_{i_1}^\top(I-H)\{\varepsilon_{i_1}^\top(I-H)\varepsilon_{j_2}\}^2\Big)\Big\}$$

$$= tr\Big\{E\Big(\varepsilon_{j_1}\varepsilon_{j_1}^\top\Big)E\Big[(I-H)\varepsilon_{i_1}\varepsilon_{i_1}^\top(I-H)\{\varepsilon_{i_1}^\top(I-H)\varepsilon_{j_2}\}^2\Big]\Big\}$$

$$= \sigma_{j_1j_1}E\Big(\varepsilon_{i_1}^\top(I-H)\varepsilon_{i_1}\{\varepsilon_{i_1}^\top(I-H)\varepsilon_{j_2}\}\{\varepsilon_{j_2}^\top(I-H)\varepsilon_{i_1}\}\Big)$$

$$= \sigma_{j_1j_1}tr\Big[E\Big\{\varepsilon_{i_1}^\top(I-H)\varepsilon_{i_1}\varepsilon_{i_1}\varepsilon_{i_1}^\top(I-H)\Big\}E\Big\{\varepsilon_{j_2}\varepsilon_{j_2}^\top(I-H)\Big\}\Big]$$

$$= \sigma_{j_1j_1}\sigma_{j_2j_2}E\Big(\varepsilon_{i_1}^\top(I-H)\varepsilon_{i_1}\Big)^2 = (n-d)(n-d+2)\sigma_{i_1i_1}^2\sigma_{j_1j_1}\sigma_{j_2j_2}. \qquad \text{(A.10)}$$

Using the results of (A.8), (A.9), and (A.10), we obtain

$$4p^3n^{-1}E\{T^2\} = \left(\frac{n-d}{n}\right)^2\mathcal{S}_1 + 6\left(\frac{n-d}{n}\right)\left(\frac{n-d+2}{n}\right)\mathcal{S}_2$$

$$+4\left(\frac{n-d}{n}\right)\left(\frac{n-d+2}{n}\right)\mathcal{S}_3. \qquad \text{(A.11)}$$

Step 3. Finally we compute $E\{T\widehat{\text{Bias}}\}$. After algebraic simplification, we obtain the following expression

$$T\widehat{\text{Bias}} = 4^{-1}n^3(n-d)^{-1}p^{-3}\sum_{i_1\neq j_1}\sum_{i_2\neq j_2}\hat{\sigma}_{i_1j_1}^2\hat{\sigma}_{i_2i_2}\hat{\sigma}_{j_2j_2}$$

$$= 4^{-1}n^{-1}(n-d)^{-1}p^{-3}\sum_{i_1\neq j_1}\sum_{i_2\neq j_2}\{\varepsilon_{i_1}^\top(I-H)\varepsilon_{j_1}\}^2\varepsilon_{i_2}^\top(I-H)\varepsilon_{i_2}\varepsilon_{j_2}^\top(I-H)\varepsilon_{j_2}.$$

Similar to Steps 1 and 2, we consider three cases given below to calculate this quantity separately.

Case I. i_1, i_2, j_1, and j_2 are mutually different. Then, $\hat{\sigma}_{i_1j_1}^2$, $\hat{\sigma}_{i_2i_2}$ and $\hat{\sigma}_{j_2j_2}$ are mutually independent. By Lemma 1(a) and 1(b), we have

$$E\Big(\{\varepsilon_{i_1}^\top(I-H)\varepsilon_{j_1}\}^2\varepsilon_{i_2}^\top(I-H)\varepsilon_{i_2}\varepsilon_{j_2}^\top(I-H)\varepsilon_{j_2}\Big)$$

$$= E\Big(\{\varepsilon_{i_1}^\top(I-H)\varepsilon_{j_1}\}^2\Big)E\Big(\varepsilon_{i_2}^\top(I-H)\varepsilon_{i_2}\Big)E\Big(\varepsilon_{j_2}^\top(I-H)\varepsilon_{j_2}\Big)$$

$$= (n-d)^3\sigma_{i_1i_1}\sigma_{j_1j_1}\sigma_{i_2i_2}\sigma_{j_2j_2}. \qquad \text{(A.12)}$$

Case II. $i_1 = i_2$, $j_1 = j_2$, but $i_1 \neq j_1$. In this case, $\hat{\sigma}_{i_1j_1}^2\hat{\sigma}_{i_2i_2}\hat{\sigma}_{j_2j_2} = \hat{\sigma}_{i_1j_1}^2\hat{\sigma}_{i_1i_1}\hat{\sigma}_{j_1j_1}$.

By Lemma 1(c), we obtain that

$$E\Big(\{\varepsilon_{i_1}^\top(I-H)\varepsilon_{j_1}\}^2\varepsilon_{i_2}^\top(I-H)\varepsilon_{i_2}\varepsilon_{j_2}^\top(I-H)\varepsilon_{j_2}\Big)$$

$$=E\Big(\{\varepsilon_{i_1}^\top(I-H)\varepsilon_{j_1}\}^2\varepsilon_{i_1}^\top(I-H)\varepsilon_{i_1}\varepsilon_{j_1}^\top(I-H)\varepsilon_{j_1}\Big)$$

$$=(n-d+2)^2(n-d)\sigma_{i_1i_1}^2\sigma_{j_1j_1}\sigma_{j_2j_2}. \tag{A.13}$$

Case III. $i_1=i_2$, but i_1, j_1, and j_2 are all different. In this case, $\hat\sigma_{i_1j_1}^2\hat\sigma_{i_2i_2}\hat\sigma_{j_2j_2}=\hat\sigma_{i_1j_1}^2\hat\sigma_{i_1i_1}\hat\sigma_{j_2j_2}$. We apply Lemma 1(d) and have

$$E\Big(\{\varepsilon_{i_1}^\top(I-H)\varepsilon_{j_1}\}^2\varepsilon_{i_1}^\top(I-H)\varepsilon_{i_1}\Big)=tr\Big\{E\Big(\varepsilon_{i_1}^\top(I-H)\varepsilon_{j_1}\varepsilon_{j_1}^\top(I-H)\varepsilon_{i_1}\varepsilon_{i_1}^\top(I-H)\varepsilon_{i_1}\Big)\Big\}$$

$$=tr\Big\{E\Big((I-H)\varepsilon_{j_1}\varepsilon_{j_1}^\top(I-H)\varepsilon_{i_1}\varepsilon_{i_1}^\top\big[\varepsilon_{i_1}^\top(I-H)\varepsilon_{i_1}\big]\Big)\Big\}$$

$$=tr\Big\{(I-H)E(\varepsilon_{j_1}\varepsilon_{j_1}^\top)(I-H)E\Big(\varepsilon_{i_1}\varepsilon_{i_1}^\top\big[\varepsilon_{i_1}^\top(I-H)\varepsilon_{i_1}\big]\Big)\Big\}$$

$$=\sigma_{j_1j_1}tr\Big[E\Big((I-H)\varepsilon_{i_1}\varepsilon_{i_1}^\top\big[\varepsilon_{i_1}^\top(I-H)\varepsilon_{i_1}\big]\Big)\Big]=\sigma_{j_1j_1}E\Big(\varepsilon_{i_1}^\top(I-H)\varepsilon_{i_1}\Big)^2$$

$$=(n-d)(n-d+2)\sigma_{i_1i_1}^2\sigma_{j_1j_1}.$$

Accordingly, we obtain that

$$E\Big(\{\varepsilon_{i_1}(I-H)\varepsilon_{j_1}\}^2\varepsilon_{i_1}(I-H)\varepsilon_{i_1}\varepsilon_{j_2}(I-H)\varepsilon_{j_2}\Big)$$

$$=E\Big(\{\varepsilon_{i_1}(I-H)\varepsilon_{j_1}\}^2\varepsilon_{i_1}(I-H)\varepsilon_{i_1}\Big)E\Big(\varepsilon_{j_2}(I-H)\varepsilon_{j_2}\Big)$$

$$=(n-d)^2(n-d+2)\sigma_{i_1i_1}^2\sigma_{j_1j_1}\sigma_{j_2j_2}. \tag{A.14}$$

Combing the result of (A.14) with those of (A.12) and (A.13), we have

$$4p^3n^{-1}E\{T\widehat{\text{Bias}}\} = \left(\frac{n-d}{n}\right)^2 \mathcal{S}_1 + 2\left(\frac{n-d+2}{n}\right)^2 \mathcal{S}_2$$

$$+4\left(\frac{n-d}{n}\right)\left(\frac{n-d+2}{n}\right)\mathcal{S}_3. \tag{A.15}$$

Consequently, (A.7), (A.11), and (A.15) in conjunction with the fact that $\mathcal{S}_2 = p^2M_{2,p}^2 - pM_{4,p}$, imply that

$$\text{var}\{T - \widehat{\text{Bias}}\} = E\{T^2\} + E\{\widehat{\text{Bias}}^2\} - 2E\{T\widehat{\text{Bias}}\}$$

$$= \left(4p^3n^{-1}\right)^{-1} 4\left(\frac{n-d+2}{n}\right)\left(\frac{n-d-1}{n}\right)\mathcal{S}_2.$$

$$= n^{-1}p^{-3}(n-d+2)(n-d-1)\left(p^2M_{2,p}^2 - pM_{4,p}\right)$$

$$= \left(\frac{n-d+2}{p}\right)\left(\frac{n-d-1}{n}\right)M_{2,p}^2 - n^{-1}p^{-2}(n-d+2)(n-d-1)M_{4,p}. \tag{A.16}$$

In addition, by the assumption that $M_{\kappa,p} \to_p M_\kappa$ with $|M_\kappa| < \infty$ and $\kappa \le 4$, we have $|p^{-2}n^{-1}(n-d+2)(n-d-1)M_{4,p}| \le p^{-2}n^{-1}n^2|M_{4,p}| = O(n/p^2)$. As a result, the right-hand side of (A.16) is

$$\left(\frac{n-d+2}{p}\right)\left(\frac{n-d-1}{n}\right)M_{2,p}^2 + O(n/p^2)$$

$$= \frac{nM_{2,p}^2}{p} - \left(\frac{2d-1}{p}\right)M_{2,p}^2 + O(n^{-1}p^{-1}) + O(n/p^2). \tag{A.17}$$

Employing the assumption of $\min\{n,p\} \to \infty$, we know that $n \to \infty$. Hence, the second and third terms in (A.17) are negligible as compared with the first term, $nM_{2,p}^2/p = O(n/p)$. Analogously, the last term in (A.17) is also negligible since $p \to \infty$. In sum, we have $\text{var}(T - \widehat{\text{Bias}}) = nM_{2,p}^2/p + o(n/p)$. This completes the proof.

Appendix D. Proof of Theorem 4

To prove this theorem, we need to demonstrate (I.) $(T - \widehat{\text{Bias}})/\text{var}^{1/2}(T - \widehat{\text{Bias}})$ is asymptotically normal; and (II.) $\widehat{M}_{2,p} \to_p M_{2,p}$. Because (II) can be obtained by applying the same techniques as in the proof of Lemma 2.1 of Srivastava (2005), we only focus on (I). To show the asymptotic normality of $(T - \widehat{\text{Bias}})/\text{var}^{1/2}(T - \widehat{\text{Bias}})$, we need to employ the martingale central limit theorem; see Hall and Heyde (1980). To this end, we define $\mathcal{F}_r = \sigma\{\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_r\}$, which represents the σ-field generated by $\{\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_r\}$ for $r = 1, 2, \cdots, p$. We further define

$$a_{p,r} = n^{-1/2}p^{-3/2} \sum_{i<j\le r} \Big[\{\varepsilon_i^\top (I-H)\varepsilon_j\}^2 - (n-d)^{-1}\{\varepsilon_i^\top (I-H)\varepsilon_i\}\{\varepsilon_j^\top (I-H)\varepsilon_j\}\Big]. \quad \text{(A.18)}$$

Obviously, $a_{p,r} \in \mathcal{F}_r$, and one can easily verify that $T - \widehat{\text{Bias}} = a_{p,p}$. Then, set $\Delta_{p,r} = a_{p,r} - a_{p,r-1}$ with $a_{p,0} = 0$. Furthermore, we can show that $E(a_{p,r}|\mathcal{F}_q) = a_{p,q}$ for any $q < r$. This implies that, for an arbitrarily fixed p, $\{\Delta_{p,r}, 1 \le r \le p\}$ is a martingale difference sequence with respect to $\{\mathcal{F}_q, 1 \le q \le p\}$. Moreover, define $\sigma_{p,r}^2 = E(\Delta_{p,r}^2|\mathcal{F}_{r-1})$. Accordingly, by the martingale central limit theorem (Hall and Heyde, 1980), it suffices to show that

$$\frac{\sum_{r=1}^p \sigma_{p,r}^2}{\text{var}(T - \widehat{\text{Bias}})} \to_p 1 \quad \text{and} \quad \frac{\sum_{r=1}^p E(\Delta_{p,r}^4)}{\text{var}^2(T - \widehat{\text{Bias}})} \to_p 0. \quad \text{(A.19)}$$

This can be done in three steps given below. In the first step, we obtain an analytical expression of $\sigma_{p,r}^2$, which facilitates subsequent technical proofs. The second step demonstrates the first part of (A.19), while the last step verifies the second part of (A.19).

Step 1. Using the fact that $n^{1/2}p^{3/2}\Delta_{p,r} = \sum_{i=1}^{r-1} \big[\{\varepsilon_i^\top (I-H)\varepsilon_r\}^2 - (n-d)^{-1}\{\varepsilon_i^\top (I-$

$H)\varepsilon_i\}\{\varepsilon_r^\top(I-H)\varepsilon_r\}\big]$, we have

$$np^3\sigma_{p,r}^2 = np^3E(\Delta_{p,r}^2|\mathcal{F}_{r-1}) = E\Bigg(\sum_{i=1}^{r-1}\sum_{j=1}^{r-1}\{\varepsilon_i^\top(I-H)\varepsilon_r\}^2\{\varepsilon_j^\top(I-H)\varepsilon_r\}^2$$

$$+(n-d)^{-2}\sum_{i=1}^{r-1}\sum_{j=1}^{r-1}\{\varepsilon_r^\top(I-H)\varepsilon_r\}^2\{\varepsilon_i^\top(I-H)\varepsilon_i\}\{\varepsilon_j^\top(I-H)\varepsilon_j\}$$

$$-2(n-d)^{-1}\sum_{i=1}^{r-1}\sum_{j=1}^{r-1}\{\varepsilon_i^\top(I-H)\varepsilon_r\}^2\{\varepsilon_j^\top(I-H)\varepsilon_j\}\{\varepsilon_r^\top(I-H)\varepsilon_r\}\Big|\mathcal{F}_{r-1}\Bigg). \qquad \text{(A.20)}$$

To obtain the explicit expression of $\sigma_{p.r}^2$, we next calculate the three terms on the right-hand side of (A.20) separately.

The 1st Term in (A.20). It is noteworthy that

$$E\Big(\{\varepsilon_i^\top(I-H)\varepsilon_r\}^2\{\varepsilon_j^\top(I-H)\varepsilon_r\}^2\Big|\mathcal{F}_{r-1}\Big)$$

$$= tr\Big\{E\Big(\{\varepsilon_i^\top(I-H)\varepsilon_r\}^2\varepsilon_r\varepsilon_r^\top\Big|\mathcal{F}_{r-1}\Big)(I-H)\varepsilon_j\varepsilon_j^\top(I-H)\Big\}. \qquad \text{(A.21)}$$

This allows us to focus on the computation of $E\big(\{\varepsilon_i^\top(I-H)\varepsilon_r\}^2\varepsilon_r\varepsilon_r^\top|\mathcal{F}_{r-1}\big)$ in the first term's calculation. For the sake of simplicity, let $\{\varepsilon_i^\top(I-H)\varepsilon_r\}^2\varepsilon_r\varepsilon_r^\top = C = (c_{gh})$ and $(I-H) = B = (b_{ij})$, where g, h, i, and j range from 1 to n. Accordingly, we have

$$c_{gh} = \sum_{k_1=1}^{n}\sum_{k_2=1}^{n}\sum_{l_1=1}^{n}\sum_{l_2=1}^{n}\Big(\varepsilon_{rg}\varepsilon_{rh}\Big)\Big(\varepsilon_{il_1}b_{l_1k_1}\varepsilon_{rk_1}\Big)\Big(\varepsilon_{il_2}b_{l_2k_2}\varepsilon_{rk_2}\Big).$$

Because ε_r is a n-dimensional normal vector with mean 0 and variance $\sigma_{rr}I$, this leads to $E(c_{gh}|\mathcal{F}_{r-1}) = \sigma_{rr}^2\sum_{l_1,l_2}(b_{l_1g}b_{l_2h}+b_{l_1h}b_{l_2g})\varepsilon_{il_1}\varepsilon_{il_2} = 2\sigma_{rr}^2\sum_{l_1,l_2}b_{l_1g}b_{l_2h}\varepsilon_{il_1}\varepsilon_{il_2}$ for $g \neq h$. In the case of $g = h$, we have

$$E(c_{gh}|\mathcal{F}_{r-1}) = \sigma_{rr}^2\sum_{k,l_1,l_2}^{k\neq g}b_{l_1k}b_{l_2k}\varepsilon_{il_1}\varepsilon_{il_2} + 3\sigma_{rr}^2\sum_{l_1,l_2}b_{l_1g}b_{l_2g}\varepsilon_{il_1}\varepsilon_{il_2}$$

$$= 2\sigma_{rr}^2 \sum_{l_1,l_2} \varepsilon_{il_1} b_{l_1 g} b_{l_2 g} \varepsilon_{il_2} + \sigma_{rr}^2 \sum_{l_1,l_2} \varepsilon_{il_1} \Big(\sum_{k=1}^{n} b_{l_1 k} b_{l_2 k} \Big) \varepsilon_{il_2}$$

$$= 2\sigma_{rr}^2 \sum_{l_1,l_2} \varepsilon_{il_1} b_{l_1 g} b_{l_2 g} \varepsilon_{il_2} + \sigma_{rr}^2 \varepsilon_i^\top (I-H)^2 \varepsilon_i$$

Since the (g,h)-th element of $(I-H)\varepsilon_i\varepsilon_i^\top(I-H)$ is $\sum_{l_1,l_2} \varepsilon_{il_1} b_{l_1 g} b_{l_2 h} \varepsilon_{il_2}$, we obtain

$$E\Big(\{\varepsilon_i^\top (I-H)\varepsilon_r\}^2 \varepsilon_r \varepsilon_r^\top \Big) = 2\sigma_{rr}^2 (I-H)\varepsilon_i\varepsilon_i^\top (I-H) + \sigma_{rr}^2 \varepsilon_i^\top (I-H)^2 \varepsilon_i I$$

$$= 2\sigma_{rr}^2 (I-H)\varepsilon_i\varepsilon_i^\top (I-H) + \sigma_{rr}^2 \varepsilon_i^\top (I-H) \varepsilon_i I,$$

because $I-H$ is a projection matrix. This, together with (A.21), leads to

$$E\Big(\{\varepsilon_i^\top (I-H)\varepsilon_r\}^2 \{\varepsilon_j^\top (I-H)\varepsilon_r\}^2 \Big| \mathcal{F}_{r-1} \Big)$$

$$= tr\Bigg[\Big\{ 2\sigma_{rr}^2 (I-H)\varepsilon_i\varepsilon_i^\top (I-H) \Big\} \Big\{ (I-H)\varepsilon_j\varepsilon_j^\top (I-H) \Big\} \Bigg]$$

$$+ tr\Bigg[\Big\{ \sigma_{rr}^2 \varepsilon_i^\top (I-H)\varepsilon_i I \Big\} \Big\{ (I-H)\varepsilon_j\varepsilon_j^\top (I-H) \Big\} \Bigg]$$

$$= 2\sigma_{rr}^2 \{\varepsilon_i^\top (I-H)\varepsilon_j\}^2 + \sigma_{rr}^2 \{\varepsilon_i^\top (I-H)\varepsilon_i\}\{\varepsilon_j^\top (I-H)\varepsilon_j\}. \qquad \text{(A.22)}$$

This completes the calculation of the major component of the first term in (A.20).

The 2nd Term in (A.20). Employing Lemma 1(d), we obtain that

$$(n-d)^{-2} E\Big(\{\varepsilon_r^\top (I-H)\varepsilon_r\}^2 \{\varepsilon_i^\top (I-H)\varepsilon_i\}\{\varepsilon_j^\top (I-H)\varepsilon_j\} | \mathcal{F}_{r-1} \Big)$$

$$= (n-d)^{-1}(n-d+2)\sigma_{rr}^2 \{\varepsilon_i^\top (I-H)\varepsilon_i\}\{\varepsilon_j^\top (I-H)\varepsilon_j\}. \qquad \text{(A.23)}$$

The 3rd Term in (A.20). Lastly, we evaluate the third term of (A.20) by computing its major component,

$$E\Big(\{\varepsilon_i^\top(I-H)\varepsilon_r\}^2\{\varepsilon_r^\top(I-H)\varepsilon_r\}\Big|\mathcal{F}_{r-1}\Big)$$

$$= E\Big[\{\varepsilon_r^\top(I-H)\varepsilon_r\}\{\varepsilon_r^\top(I-H)\varepsilon_i\}\{\varepsilon_i^\top(I-H)\varepsilon_r\}\Big|\mathcal{F}_{r-1}\Big]$$

$$= tr\Bigg\{\Big[E\{\varepsilon_r^\top(I-H)\varepsilon_r\}\varepsilon_r\varepsilon_r^\top\Big](I-H)\varepsilon_i\varepsilon_i^\top(I-H)\Bigg\}. \tag{A.24}$$

For the sake of simplicity, let $\{\varepsilon_r^\top(I-H)\varepsilon_r\}\varepsilon_r\varepsilon_r^\top = \tilde{C} = (\tilde{c}_{gh})$, where

$$\tilde{c}_{gh} = \sum_{k=1}^{n}\sum_{l=1}^{n}\varepsilon_{rg}\varepsilon_{rh}(\varepsilon_{rk}b_{kl}\varepsilon_{rl}).$$

When $g \neq h$, we have $E(\tilde{c}_{gh}) = 2\sigma_{rr}^2 b_{gh}$; otherwise, we obtain that

$$E(\tilde{c}_{gg}) = b_{gg}E(\varepsilon_{rg}^4) + \sum_{k\neq g} b_{kk}E(\varepsilon_{rg}^2)E(\varepsilon_{rk}^2) = 3b_{gg}\sigma_{rr}^2 + \sum_{k\neq g} b_{kk}\sigma_{rr}^2$$

$$= 2\sigma_{rr}^2 b_{gg} + \sigma_{rr}^2\sum_{i=1}^{n} b_{kk} = 2\sigma_{rr}^2 b_{gg} + \sigma_{rr}^2 tr(I-H) = 2\sigma_{rr}^2 b_{gg} + (n-d)\sigma_{rr}^2.$$

Accordingly, we have $E(\varepsilon_r^\top(I-H)\varepsilon_r\varepsilon_r\varepsilon_r^\top) = 2\sigma_{rr}^2(I-H) + (n-d)\sigma_{rr}^2 I$. The above results lead to

$$E\Big(\{\varepsilon_i^\top(I-H)\varepsilon_r\}^2\{\varepsilon_r^\top(I-H)\varepsilon_r\}\{\varepsilon_j^\top(I-H)\varepsilon_j\}\Big|\mathcal{F}_{r-1}\Big)$$

$$= tr\Bigg[\Big\{2\sigma_{rr}^2(I-H) + (n-d)\sigma_{rr}^2 I)\Big\}\Big\{(I-H)\varepsilon_i\varepsilon_i^\top(I-H)\Big\}\Bigg]\{\varepsilon_j^\top(I-H)\varepsilon_j\}$$

$$= (n-d+2)\sigma_{rr}^2\{\varepsilon_i^\top(I-H)\varepsilon_i\}\{\varepsilon_j^\top(I-H)\varepsilon_j\}. \tag{A.25}$$

which completes the calculation of the major component of the third term. This, together with (A.20), (A.22) and (A.23), yields,

$$\sigma_{p,r}^2 = 2n^{-1}p^{-3}\sigma_{rr}^2 \sum_{i=1}^{r-1}\sum_{j=1}^{r-1} \Big[\{\varepsilon_i^\top (I-H)\varepsilon_j\}^2 - (n-d)^{-1}\{\varepsilon_i(I-H)\varepsilon_i\}\{\varepsilon_j(I-H)\varepsilon_j\}\Big]$$

$$= 4n^{-1/2}p^{-3/2}\sigma_{rr}^2 a_{p,r-1} + 2n^{-1}p^{-3}D_r, \tag{A.26}$$

where $D_r = \sigma_{rr}^2 \sum_{i=1}^{r-1}\{1-(n-d)^{-1}\}\{\varepsilon_i^\top(I-H)\varepsilon_i\}^2$ and $a_{p,r}$ is defined in (A.18).

Step 2. We verify the first part of (A.19). Because $\Delta_{p,r}$ is a martingale sequence, one can verify that $E(\sigma_{p,1}^2+\sigma_{p,2}^2+\cdots+\sigma_{p,p}^2) = \mathrm{var}\big\{T-\widehat{\mathrm{Bias}}\big\}$. Accordingly, we only need to show that $\mathrm{var}(\sum_{r=1}^p \sigma_{p,r}^2)/\mathrm{var}^2(T-\widehat{\mathrm{Bias}}) \to 0$. To this end, we focus on calculating $\mathrm{var}(\sum_{r=1}^p \sigma_{p,r}^2)$. Because Theorem 3 implies that $\mathrm{var}(T-\widehat{\mathrm{Bias}}) = O(np^{-1})$, (A.26) suggests that we can prove the first part of (A.19) by demonstrating the following results.

$$(i.)\ \mathrm{var}\Big(\sum_{r=1}^p D_r\Big) = o\big(n^4p^4\big) \ \text{ and } \ (ii.)\ \mathrm{var}\Big(\sum_{r=1}^p \sigma_{rr}^2 a_{p,r-1}\Big) = o\big(n^3p\big). \tag{A.27}$$

To prove equation (i), we first note that

$$\sum_{r=1}^p D_r = \Big\{1-(n-d)^{-1}\Big\}\sum_{i=1}^{p-1}\Bigg\{\varepsilon_i^\top(I-H)\varepsilon_i \sum_{r=i+1}^p \sigma_{rr}^2\Bigg\},$$

where $\{\varepsilon_i^\top(I-H)\varepsilon_i : 1 \le i \le p-1\}$ are mutually independent. After algebraic simplification, we have

$$\mathrm{var}\Big(\sum_{r=1}^p D_r\Big) = \Big\{1-(n-d)^{-1}\Big\}^2 \sum_{i=1}^{p-1}\Big(\sum_{r=i+1}^p \sigma_{rr}^2\Big)^2 \mathrm{var}\Big\{\varepsilon_i^\top(I-H)\varepsilon_i\Big\}.$$

$$= \Big\{1-(n-d)^{-1}\Big\}^2 \sum_{i=1}^{p-1}\Big(\sum_{r=i+1}^p \sigma_{rr}^2\Big)^2 \Big\{2\sigma_{ii}^2(n-d)\Big\}$$

$$\leq 2n \sum_{i=1}^{p-1} \Big(\sum_{r=i+1}^{p} \sigma_{rr}^2 \Big)^2 \sigma_{ii}^2, \tag{A.28}$$

where the last inequality is using the fact that $\{1-(n-d)^{-1}\}^2(n-d) = (n-d-1)\{1-(n-d)^{-1}\} \leq n$. Since $\sum_{r=i+1}^{p} \sigma_{rr}^2 \leq \sum_{r=1}^{p} \sigma_{rr}^2 = pM_{2,p}$ and $M_{2,p} \to_p M_2$ with $|M_2| < \infty$, the right-hand side of (A.28) can be further bounded from infinity by

$$2n \sum_{i=1}^{p-1} \Big(pM_{2,p} \Big)^2 \sigma_{ii}^2 \leq 2np^2 M_{2,p}^2 \sum_{i=1}^{p-1} \sigma_{ii}^2 \leq 2np^3 M_{2,p}^3 = o(n^4 p^4).$$

This verifies the equation (i) in (A.27).

We next show equation (ii). Because $a_{p,r} = \sum_{s=1}^{r} \Delta_{p,s}$, we obtain that

$$\sum_{r=1}^{p} \sigma_{rr}^2 a_{p,r-1} = \sum_{r=1}^{p} \sigma_{rr}^2 \sum_{s=1}^{r-1} \Delta_{p,s} = \sum_{s=1}^{p-1} \Delta_{p,s} \left(\sum_{r=s+1}^{p} \sigma_{rr}^2 \right).$$

Furthermore, using the fact that $\{\Delta_{p,s} : 1 \leq s \leq p-1\}$ is a martingale sequence, we have

$$\operatorname{var} \left(\sum_{r=1}^{p} \sigma_{rr}^2 a_{p,r} \right) = \sum_{s=1}^{p-1} E\Big(\Delta_{p,s}^2\Big) \left(\sum_{r=s+1}^{p} \sigma_{rr}^2 \right)^2 \leq p^2 M_{2,p}^2 \sum_{s=1}^{p} E\Big(\Delta_{p,s}^2\Big).$$

By Cauchy's inequality, the right-hand side of the above inequality can be further bounded from infinity by

$$p^2 M_{2,p}^2 p \left\{ p^{-1} \sum_{s=1}^{p} E(\Delta_{p,s}^4) \right\}^{1/2}. \tag{A.29}$$

Moreover, using the result that will be demonstrated in (A.30), we have $\sum_{s=1}^{p} E(\Delta_{p,s}^4) = O(n^2 p^{-3})$. This, together with the assumption, $M_{2,p} \to_p M_2$ with $|M_2| < \infty$, implies that the right-hand side of (A.29) is the order of $O(np) = o(n^3 p)$, which completes the proof of equation (ii) in (A.27).

Step 3. We finally show the second part of (A.19). It is noteworthy that

$$\Delta_{p,r} = n^{-1/2}p^{-3/2}\sum_{i=1}^{r-1}\Big[\{\varepsilon_i^\top(I-H)\varepsilon_r\}^2 - (n-d)^{-1}\{\varepsilon_i^\top(I-H)\varepsilon_i\}\{\varepsilon_r^\top(I-H)\varepsilon_r\}\Big]$$

$$= n^{-1/2}p^{-3/2}\sum_{i=1}^{r-1}\{\varepsilon_i^\top A_r\varepsilon_i\},$$

where $A_r = (I-H)\varepsilon_r\varepsilon_r^\top(I-H) - (n-d)^{-1}(I-H)\{\varepsilon_r^\top(I-H)\varepsilon_r\}$ that is a symmetric matrix, and it is related to the current observation ε_r only. When $i \neq r$, one can show that $tr(A_r) = 0$ and $E(\varepsilon_i^\top A_r\varepsilon_i|\varepsilon_r) = 0$. Using these results, we have

$$n^2p^6E\Big(\Delta_{p,r}^4\Big|\varepsilon_r\Big) = E\Bigg(\sum_{i_1=1}^{r-1}\sum_{i_2=1}^{r-1}\sum_{i_3=1}^{r-1}\sum_{i_4=1}^{r-1}\varepsilon_{i_1}^\top A_r\varepsilon_{i_1}\varepsilon_{i_2}^\top A_r\varepsilon_{i_2}\varepsilon_{i_3}^\top A_r\varepsilon_{i_3}\varepsilon_{i_4}^\top A_r\varepsilon_{i_4}\Big|\varepsilon_r\Bigg)$$

$$= E\Bigg(3\sum_{i=1}^{r-1}\sum_{j\neq i}\{\varepsilon_i^\top A_r\varepsilon_i\}^2\{\varepsilon_j^\top A_r\varepsilon_j\}^2 + \sum_{i=1}^{r-1}\{\varepsilon_i^\top A_r\varepsilon_i\}^4\Big|\varepsilon_r\Bigg).$$

In addition, one can verity by Lemma 2 that there exists constants C_1, C_2, and C_3 such that

$$n^2p^6E\Big(\Delta_{p,r}^4\Big|\varepsilon_r\Big) \leq C_1\sum_{i=1}^{r-1}\sum_{j\neq i}\sigma_{ii}^2\sigma_{jj}^2 tr^2(A_r^2) + C_2\sum_{i=1}^{r-1}\sigma_{ii}^4\Big[tr^2(A_r^2) + tr(A_r^4)\Big]$$

$$\leq C_3p^2\Big[tr^2(A_r^2) + tr(A_r^4)\Big].$$

After algebraic simplification, we obtain that

$$tr(A_r^2) = tr\Big((I-H)\varepsilon_r\varepsilon_r^\top(I-H)\varepsilon_r\varepsilon_r^\top(I-H) + (n-d)^{-2}(I-H)\{\varepsilon_r^\top(I-H)\varepsilon_r\}^2$$

$$-2(n-d)^{-1}(I-H)\varepsilon_r\varepsilon_r^\top(I-H)\{\varepsilon_r^\top(I-H)\varepsilon_r\}\Big)$$

$$= \{1-(n-d)^{-1}\}\{\varepsilon_r^\top(I-H)\varepsilon_r\}^2$$

and $tr(A_r^4) = \{1-4(n-d)^{-1}+6(n-d)^{-2}-3(n-d)^{-3}\}\{\varepsilon_r^\top(I-H)\varepsilon_r\}^4$. Consequently, we have

$$n^2p^6E(\Delta_{p,r}^4) \leq E\Big\{C_3p^2tr^2(A_r^2)+C_3p^2tr(A_r^4)\Big\} \leq C_4p^2E\Big(\{\varepsilon_r^\top(I-H)\varepsilon_r\}^4\Big)$$

$$\leq C_4p^2E(\varepsilon_r^\top\varepsilon_r)^4 = C_4p^2n^4E\Big(\varepsilon_r^\top\varepsilon_r/n\Big)^4 \leq 2C_4p^2n^4\sigma_{rr}^4,$$

where C_4 is a positive constant and the last inequality is due to the fact that $\varepsilon_r^\top\varepsilon_r/n \to_p \sigma_{rr}$. This implies that

$$E\Big(\sum_{r=1}^{p}\Delta_{p,r}^4\Big) \leq 2C_4n^2p^{-3}\sigma_{rr}^4 = o\Big(n^2p^{-2}\Big), \tag{A.30}$$

which proves the second part of (A.19) and thus completes the proof of Theorem 4.

REFERENCES

Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis,* 3rd Ed., John Wiley: New York.

Bai, Z. D. and Silverstein, J. W. (2005), *Spectral Analysis of Large Dimensional Random Matrices*, Scientific Press: Beijing.

Brown, S. J. (1989), “The number of factors in security returns,” *The Journal of Finance*, 44, 1247–1262.

Carhart, M. M. (1997), “On persistency in mutual fund performance,” *Journal of Finance*, 52, 57–82.

Chen, S. X., Zhang, L. X., and Zhong, P. S. (2010), “Tests for high dimensional covariance matrices,” *Journal of the American Statistical Association*, 105, 810–819.

Connor, G. and Korajczky, R. (1993), "A test for the number of factors in an approximate factor model," *The Journal of Finance*, 48, 1263–1291.

Diggle, P. and Verbyla, A. (1998), "Nonparametric estimation of covariance structure in longitudinal data," *Biometrics*, 54, 401–415.

Drton, M. and Perlman, M. (2004), "Model selection for gaussian concentration graphs," *Biometrika*, 91, 591–602.

Edwards, D. M. (2000), *Introduction to Graphical Modeling*, Springer, New York.

Efron, B. (2009), "Are a set of microarrays independent of each other?" *The Annals of Applied Statistics*, 3, 992–942.

Fama, E. and French, K. (1993), "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33, 3–56.

— (1996), "Multifactor explanations of asset pricing anomalies," *The Journal of Finance*, 51, 55–84.

Fan, J., Fan, Y., and Lv, J. (2008), "High dimensional covariance matrix estimation using a factor model," *Journal of Econometrics*, 147, 186–197.

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modeling and Its Applications*, Chapman and Hall: New York.

Fan, J. and Li, R. (2001), "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, 96, 1348–1360.

Gleser, L. J. (1966), "A note on the sphericity test," *The Annals of Mathematical Statistics*, 37, 464–467.

Hall, P. and Heyde, C. C. (1980), *Martingale Limit Theory and its Application*, New York: Academic Press.

Härdle, W., Liang, H., and Gao, J. (2000), *Partially Linear Models*, Heidelberg: Springer.

Huang, J. Z., Liu, N., Pourahmadi, M., and Liu, L. (2006), "Covariance selection and estimation via penalized normal likelihood," *Biometrika*, 93, 85–98.

Jegadeesh, N. and Titman, S. (1993), "Returns to buying winners and selling lowers: implications for stock market efficiency," *Journal of Finance*, 48, 65–91.

John, S. (1971), "Some optimal multivariate test," *Biometrika*, 59, 123–127.

Lam, C. and Fan, J. (2009), "Sparsistency and rates of convergence in large covariance matrices estimation," *The Annals of Statistics*, 37, 4254–4278.

Ledoit, O. and Wolf, M. (2004), "Honey, I shrunk the sample covariance matrix," *Journal of Portfolio Management*, 4, 110–119.

Leng, C. and Tang, C. (2012), "Sparse matrix graphical models," *Journal of the American Statistical Association*, 4, 1187–1200.

Lintner, J. (1965), "The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets," *Review of Economics and Statistics*, 47, 13–37.

Magnus, J. R. (1986), "The exact moments of a ratio of quadratic forms in normal variables," *Annals of Economics and Statistics*, 4, 95–109.

Markowitz, H. M. (1952), "Portfolio Selection," *Journal of Finance*, 7, 77–91.

Meinshausen, N. and Bühlmann, P. (2006), "High-dimensional graphs and variable selection with the lasso," *The Annals of Statistics*, 34, 1436–1462.

Mossin, J. (1966), "Equilibrium in a capital asset market," *Econometrika*, 34, 768–783.

Nagao, H. (1973), "On some test criteria for covariance matrix," *The Annals of Statistics*, 1, 700–709.

Onatski, A., Moreira, M. J., and Hallin, M. (2013), "Asymptotic power of sphericity tests for high dimensional data," *The Annals of Statistics*, 41, 1055–1692.

Schott, J. R. (2005), "Testing for complete independence in high dimensions," *Biometrika*, 92, 951–956.

Sharpe, W. (1964), "Capital asset prices: A theory of market equilibrium under conditions of risks," *Journal of Finance*, 19, 425–442.

Smith, M. and Kohn, R. (2002), "Parsimonious covariance matrix estimation for longitudinal data," *Journal of the American Statistical Association*, 97, 1141–1153.

Srivastava, M. S. (2005), "Some tests concerning the covariance matrix in high dimensional data," *Journal of Japan Statistical Society*, 35, 251–272.

Tibshirani, R. J. (1996), "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society, Series B*, 58, 267–288.

Trzcinka, C. (1986), "On the number of factors in the arbitrage pricing model," *The Journal of Finance*, 41, 347–368.

Xia, Y. (2008), “A multiple-index model and dimension reduction,” *Journal of the American Statistical Association*, 103, 1631–1640.

Yuan, M. and Lin, Y. (2006). “Model selection and estimation in regression with grouped variables,” *Journal of the Royal Statistical Society: Series B*, 68, 49–67.

Zhang, W. and Leng, C. (2012), “A moving average Cholesky factor model in covariance modeling for longitudinal data,” *Biometrika*, 99, 141–150.

Table 1: The sizes of bias-corrected (BC) and Schott tests for testing $H_0 : \rho = 0$ with a nominal level of 5%

		Bias-Corrected (BC) Test				Schott Test				Adjusted Schott Test			
	Error	d				d				d			
p	Distribution	0	1	3	5	0	1	3	5	0	1	3	5
							$n = 100$						
100	Normal	0.045	0.049	0.037	0.038	0.042	0.125	0.472	0.831	0.042	0.042	0.055	0.046
	Student	0.061	0.046	0.054	0.058	0.064	0.124	0.450	0.844	0.064	0.046	0.059	0.058
	Mixture	0.069	0.053	0.049	0.051	0.063	0.132	0.462	0.821	0.063	0.056	0.050	0.053
	Gamma	0.046	0.056	0.050	0.045	0.057	0.124	0.444	0.836	0.057	0.055	0.064	0.062
500	Normal	0.048	0.046	0.031	0.046	0.051	0.836	1.000	1.000	0.051	0.052	0.046	0.050
	Student	0.056	0.065	0.055	0.046	0.043	0.800	1.000	1.000	0.043	0.057	0.061	0.053
	Mixture	0.056	0.058	0.059	0.053	0.050	0.789	1.000	1.000	0.050	0.068	0.069	0.066
	Gamma	0.062	0.048	0.055	0.044	0.052	0.812	1.000	1.000	0.052	0.060	0.042	0.064
1000	Normal	0.057	0.054	0.048	0.042	0.049	1.000	1.000	1.000	0.049	0.054	0.056	0.045
	Student	0.061	0.059	0.061	0.053	0.055	1.000	1.000	1.000	0.055	0.066	0.061	0.054
	Mixture	0.058	0.060	0.047	0.047	0.060	1.000	1.000	1.000	0.060	0.062	0.050	0.057
	Gamma	0.041	0.070	0.044	0.049	0.046	1.000	1.000	1.000	0.046	0.077	0.064	0.061
							$n = 200$						
100	Normal	0.052	0.043	0.054	0.041	0.054	0.075	0.190	0.323	0.054	0.040	0.053	0.050
	Student	0.055	0.046	0.055	0.043	0.062	0.090	0.185	0.331	0.062	0.061	0.053	0.040
	Mixture	0.055	0.057	0.068	0.037	0.062	0.088	0.220	0.364	0.062	0.055	0.070	0.053
	Gamma	0.054	0.050	0.041	0.049	0.050	0.079	0.191	0.354	0.050	0.047	0.060	0.057
500	Normal	0.041	0.047	0.065	0.053	0.033	0.345	0.977	1.000	0.033	0.054	0.053	0.047
	Student	0.066	0.065	0.058	0.046	0.043	0.335	0.984	1.000	0.043	0.061	0.060	0.050
	Mixture	0.057	0.051	0.048	0.059	0.045	0.379	0.985	1.000	0.045	0.053	0.046	0.057
	Gamma	0.051	0.059	0.040	0.048	0.055	0.349	0.984	1.000	0.055	0.050	0.049	0.040
1000	Normal	0.042	0.043	0.041	0.045	0.041	0.810	1.000	1.000	0.041	0.048	0.051	0.052
	Student	0.056	0.050	0.057	0.054	0.050	0.794	1.000	1.000	0.050	0.048	0.063	0.058
	Mixture	0.036	0.048	0.052	0.070	0.042	0.792	1.000	1.000	0.042	0.053	0.057	0.069
	Gamma	0.055	0.047	0.049	0.050	0.055	0.811	1.000	1.000	0.055	0.049	0.049	0.047