The Annals of Statistics
2019, Vol. 47, No. 5, 2977–3008
https://doi.org/10.1214/18-AOS1771

TESTING FOR INDEPENDENCE OF LARGE DIMENSIONAL VECTORS

BY TARAS BODNAR, HOLGER DETTE[1] AND NESTOR PAROLYA

Stockholm University, Ruhr University Bochum and Leibniz University Hannover

In this paper, new tests for the independence of two high-dimensional vectors are investigated. We consider the case where the dimension of the vectors increases with the sample size and propose multivariate analysis of variance-type statistics for the hypothesis of a block diagonal covariance matrix. The asymptotic properties of the new test statistics are investigated under the null hypothesis and the alternative hypothesis using random matrix theory. For this purpose, we study the weak convergence of linear spectral statistics of central and (conditionally) noncentral Fisher matrices. In particular, a central limit theorem for linear spectral statistics of large dimensional (conditionally) noncentral Fisher matrices is derived which is then used to analyse the power of the tests under the alternative.

The theoretical results are illustrated by means of a simulation study where we also compare the new tests with several alternative, in particular with the commonly used corrected likelihood ratio test. It is demonstrated that the latter test does not keep its nominal level, if the dimension of one sub-vector is relatively small compared to the dimension of the other sub-vector. On the other hand, the tests proposed in this paper provide a reasonable approximation of the nominal level in such situations. Moreover, we observe that one of the proposed tests is most powerful under a variety of correlation scenarios.

1. Introduction. Estimation and testing the structure of the covariance matrix are important problems that have a number of applications in practice. For instance, the covariance matrix plays an important role in the determination of the optimal portfolio structure following the well-known mean-variance analysis of Markowitz (1952). It is also used in prediction theory where the problem of forecasting future values of the process based on its previous observations arises. In such applications, misspecification of the covariance matrix might lead to significant errors in the optimal portfolio structure and predictions. The problem becomes even more difficult if the dimension is of similar order or even larger as the sample size. A number of such situations are present in biostatistics, wireless communications and finance (see, e.g., Fan and Li (2006), Johnstone (2006) and references therein).

Received August 2017; revised May 2018.

[1]Supported by the DFG Research Unit 1735, DE 502/26-2.

MSC2010 subject classifications. Primary 60B20, 60F05, 62H15; secondary 62H20, 62F05.

Key words and phrases. Testing for independence, large dimensional covariance matrix, noncentral Fisher random matrix, linear spectral statistics, asymptotic normality.

The sample covariance matrix is the commonly used estimator in practice. However, in the case of large dimension (compared to the sample size), a number of studies demonstrate that the sample covariance does not perform well as an estimator of the population covariance matrix and numerous authors have recently addressed this problem. One approach is based on the construction of improved estimators in particular shrinkage-type estimators which reduce the variability of the sample covariance matrix at the cost of an additional bias (see, Ledoit and Wolf (2012), Wang et al. (2015) or Bodnar, Gupta and Parolya (2014), Bodnar, Gupta and Parolya (2016) among others). Alternatively, several authors impose structural assumptions on the population covariance matrix such as a block diagonal structure (e.g., Devijver and Gallopin (2018)), Toeplitz matrix (see, Cai, Ren and Zhou (2013)), band matrix (see, Bickel and Levina (2008)) or general sparsity assumptions (see Cai, Liu and Luo (2011), Cai and Shen (2011), Cai and Zhou (2012) among others) and show, that the population covariance matrix can be estimated consistently in these cases, even for large dimensions. However, these techniques may fail if the structural assumptions are not satisfied, and consequently it is desirable to validate the corresponding assumptions regarding the postulated structure of the covariance matrix.

In the present paper, we consider the problem of testing for a block diagonal structure of the covariance, which has found considerable interest in the literature. Early work in this direction has been done by Mauchly (1940), who proposed a likelihood ratio test for the hypothesis of sphericity of a normal distribution, that is, the independence of all components. This method has been extended by Gupta and Xu (2006) to the nonnormal case and by Bai et al. (2009) and Jiang and Yang (2013) to the high-dimensional case. An alternative approach is based on the empirical distance between the sample covariance matrix and the target (e.g., a multiplicity of the identity matrix) and was initially suggested by John (1971) and Nagao (1973). These tests can also be extended for testing the corresponding hypotheses in the high-dimensional setup (see Ledoit and Wolf (2002), Birke and Dette (2005), Fisher, Sun and Gallagher (2010), Chen, Zhang and Zhong (2010)). Other authors use the distributional properties of the largest eigenvalue of the sample covariance matrix to construct tests (see, e.g., Johnstone (2001, 2008)).

In the problem of testing the independence between two (or more) groups of random variables under the assumption of normality, the likelihood ratio approach has also found considerable interest in the literature. The main results for a fixed dimension can be found in the text books of Muirhead (1982) and Anderson (2003). Recently, Jiang and Yang (2013) have extended the likelihood ratio approach to the case of high-dimensional data, while Hyodo et al. (2015) and Yamada, Hyodo and Nishiyama (2017) used an empirical distance approach to test for a block diagonal covariance matrix.

In Section 2, we introduce the testing problem (in the case of two blocks) and demonstrate by means of a small simulation study that the likelihood ratio test does not yield a reliable approximation of the nominal level, if the size of one block is small compared to the other one. In Section 3, we introduce three alternative test statistics which are motivated from classical multivariate analysis of variance (MANOVA) and are defined as linear spectral statistics of a Fisher matrix. We derive their asymptotic distributions under the null hypotheses and illustrate the approximation of the nominal level by means of a simulation study. A comparison with the commonly used likelihood ratio test shows that the new tests provide a reasonable approximation of the nominal level in situations where the likelihood ratio test fails. Section 4 is devoted to the analysis of statistical properties of the new tests under the alternative hypothesis. For this purpose, we present a new central limit theorem for a (conditionally) noncentral Fisher random matrix which is of own interest and can be used to study some properties of the power of the new tests. Finally, most technical details and proofs are given in Appendix A and in the Supplementary Material (see Bodnar, Dette and Parolya (2019)).

2. Testing for independence. Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be a sample of i.i.d. observations from a p-dimensional normal distribution with zero mean vector and covariance matrix $\boldsymbol{\Sigma}$, that is, $\mathbf{x}_1 \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$. We define the $p \times n$ dimensional observation matrix $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ and denote by

$$\mathbf{S} = \frac{1}{n}\mathbf{X}\mathbf{X}^\top$$

the sample covariance matrix which is used as an estimate of $\boldsymbol{\Sigma}$. It is well known that $n\mathbf{S}$ has a p-dimensional Wishart distribution with n degrees of freedom and covariance matrix $\boldsymbol{\Sigma}$, that is, $n\mathbf{S} \sim W_p(n, \boldsymbol{\Sigma})$. In the following, we consider partitions of the population and the sample covariance matrix given by

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \quad \text{and} \quad n\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix}, \tag{2.1}$$

respectively, where $\boldsymbol{\Sigma}_{ij} \in \mathbb{R}^{p_i \times p_j}$ and $\mathbf{S}_{ij} \in \mathbb{R}^{p_i \times p_j}$ with $i, j = \{1, 2\}$ and $p_1 + p_2 = p$. We are interested in the hypothesis that the sub-vectors $\mathbf{x}_{1,1}$ and $\mathbf{x}_{1,2}$ of size p_1 and p_2 in the vector $\mathbf{x}_1 = (\mathbf{x}_{1,1}^\top, \mathbf{x}_{1,2}^\top)^\top$ are independent, or equivalently that the covariance matrix is block diagonal, that is,

$$H_0 : \boldsymbol{\Sigma}_{12} = \mathbf{O} \quad \text{versus} \quad H_1 : \boldsymbol{\Sigma}_{12} \neq \mathbf{O}. \tag{2.2}$$

Here, the symbol $\mathbf{O}$ denotes a matrix of an appropriate order with all entries equal to 0. It is worthwhile to mention that the case of nonzero mean vector can be treated exactly in the same way observing that the centered sample covariance matrix has a $\frac{1}{n-1}W_p(n-1, \boldsymbol{\Sigma})$ distribution. Thus, one needs to normalize the sample covariance matrix by $1/(n-1)$ instead of $1/n$ due to the *substitution principle* of Zheng, Bai and Yao (2015b) and the results presented in our paper will still remain valid.

Throughout this paper, we consider the case where the dimension of the blocks is increasing with the sample size, that is, $p = p(n)$, $p_i = p_i(n)$, such that

$$\lim_{n\to\infty} \frac{p_i}{n} = c_i < 1, \quad i = 1, 2$$

and define $c = c_1 + c_2$. For further reference, we also introduce the quantities

$$\gamma_{1,n} = \frac{p - p_1}{p_1}, \tag{2.3}$$

$$\gamma_{2,n} = \frac{p - p_1}{n - p_1}, \tag{2.4}$$

$$h_n = \sqrt{\gamma_{1,n} + \gamma_{2,n} - \gamma_{1,n}\gamma_{2,n}}. \tag{2.5}$$

A common approach in testing for independence is the likelihood ratio test based on the statistic

$$V_n = \frac{|\mathbf{S}|}{|\mathbf{S}_{11}||\mathbf{S}_{22}|} = \frac{|\mathbf{S}_{11}||\mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}|}{|\mathbf{S}_{11}||\mathbf{S}_{22}|} = |\mathbf{I}_{p-p_1} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}|.$$

The null hypothesis is rejected for small values of V_n. Jiang, Bai and Zheng (2013) showed that under the assumptions made in this section V_n can be written in terms of a determinant of a central Fisher matrix, that is,

$$V_n = \left|\mathbf{I}_{p-p_1} - \mathbf{F}\left(\mathbf{F} + \frac{\gamma_{1,n}}{\gamma_{2,n}}\mathbf{I}_{p-p_1}\right)^{-1}\right| = \left|\frac{\gamma_{2,n}}{\gamma_{1,n}}\mathbf{F} + \mathbf{I}_{p-p_1}\right|^{-1}, \tag{2.6}$$

where $\mathbf{F} = \frac{1}{p_1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}(\frac{1}{n-p_1}(\mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}))^{-1}$. Under the null hypothesis of independent blocks, the matrix $\mathbf{F}$ is a "ratio" of two central Wishart matrices with p_1 and $n - p_1$ degrees of freedom. Naturally, it is called a central Fisher matrix with p_1 and $n - p_1$ degrees of freedom, an analogue to its one-dimensional counterpart (see Fisher (1939)). In particular, we have the following result (see Theorem 8.2 in Yao, Zheng and Bai (2015))

PROPOSITION 1. *Under the null hypothesis, we have for* $T_{\mathrm{LR}} = \log(V_n)$

$$\frac{T_{\mathrm{LR}} - (p - p_1)s_{\mathrm{LR}} - \mu_{\mathrm{LR}}}{\sigma_{\mathrm{LR}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

where the quantities $\mu_{\mathrm{LR}}, \sigma_{\mathrm{LR}}^2$ *and* s_{LR} *are defined by*

$$\mu_{\mathrm{LR}} = 1/2\log\left[\frac{(w_n^{*2} - d_n^{*2})h_n^2}{(w_n^* h_n - \gamma_{2,n} d_n^{*2})^2}\right], \qquad \sigma_{\mathrm{LR}}^2 = 2\log\left[\frac{w_n^{*2}}{w_n^{*2} - d_n^{*2}}\right],$$

$$s_{\mathrm{LR}} = \log\left(\frac{\gamma_{1,n}}{\gamma_{2,n}}(1 - \gamma_{2,n})^2\right) + \frac{1 - \gamma_{2,n}}{\gamma_{2,n}}\log(w_n^*) - \frac{\gamma_{1,n} + \gamma_{2,n}}{\gamma_{1,n}\gamma_{2,n}}\log(w_n^* - d_n^*\gamma_2/h_n)$$

$$+\begin{cases} \frac{1-\gamma_{1,n}}{\gamma_{1,n}}\log(w_n^* - d_n^* h_n) & \gamma_{1,n}\in(0,1), \\ 0 & \gamma_{1,n}=1, \\ -\frac{1-\gamma_{1,n}}{\gamma_{1,n}}\log(w_n^* - d_n^*/h_n) & \gamma_{1,n}>1, \end{cases}$$

with $w_n^* = \frac{h_n}{\sqrt{\gamma_{2,n}}}$ *and* $d_n^* = \sqrt{\gamma_{2,n}}$.

Proposition 1 shows that the likelihood ratio test, which rejects the null hypothesis whenever

$$\frac{T_{\mathrm{LR}} - (p-p_1)s_{\mathrm{LR}} - \mu_{\mathrm{LR}}}{\sigma_{\mathrm{LR}}} < -u_{1-\alpha}, \tag{2.7}$$

is an asymptotic level α test (here and throughout this paper $u_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the standard normal distribution). In Figure 1, we illustrate the approximation of the nominal level of the test (2.7) by means of a small simulation study for the sample size $n = 100$, dimension $p = 60$ and different values of p_1 and p_2. We considered a centered p-dimensional normal distribution where the blocks $\mathbf{\Sigma}_{11}$ and $\mathbf{\Sigma}_{22}$ in the block diagonal matrix $\mathbf{\Sigma}$ are constructed as follows. For the first block $\mathbf{\Sigma}_{11}$, we took p_1 uniformly distributed eigenvalues on the interval $(0, 1]$ while the corresponding eigenvectors are simulated from the Haar distribution on the unit sphere. The p_2 eigenvalues of the second block $\mathbf{\Sigma}_{22}$ are drawn from a uniform distribution on the interval $[1, 10]$ while the corresponding eigenvectors are again Haar distributed. The matrices $\mathbf{\Sigma}_{11}$ and $\mathbf{\Sigma}_{22}$ are then fixed for the generation of multivariate normal distributed random variables ($\mathbf{\Sigma}_{12} = \mathbf{O}$). The plots show the empirical distribution of the statistic $(T_{\mathrm{LR}} - (p - p_1)s_{\mathrm{LR}} - \mu_{\mathrm{LR}})/\sigma_{\mathrm{LR}}$ using 1000 simulation runs and the density of a standard normal distribution. We observe a reasonable approximation if the dimension p_1 of the sub-vector $\mathbf{x}_{1,1}$ is large compared to the dimension p of the vector $\mathbf{x}_1$, that is $\gamma_{1,n} \le 1$ (see the upper part of Figure 1). However, if $\gamma_{1,n} >> 1$, there arises a strong bias (see the lower part of Figure 1) and the asymptotic statement in Proposition 1 cannot be used to obtain critical value for the test (2.7).

Motivated by the poor quality of the approximation of the finite sample distribution of the likelihood ratio test by a normal distribution if the dimension p_1 is small compared to the dimension p_2, we now construct alternative tests for the hypothesis (2.2), which will yield a more stable approximation of the nominal level. For this purpose, we first note that a nonsingular partitioned matrix $\mathbf{\Sigma}$ in (2.1) is block diagonal (i.e., $\mathbf{\Sigma}_{21} = \mathbf{O}$) if and only if $\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12} = \mathbf{O}$. Therefore, a test for independence can also be obtained by testing the hypotheses

$$H_0 : \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12} = \mathbf{O} \quad \text{versus} \quad H_1 : \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12} \neq \mathbf{O}. \tag{2.8}$$

In the following section, we will propose three tests for the hypothesis (2.8) as an alternative to the likelihood ratio test.

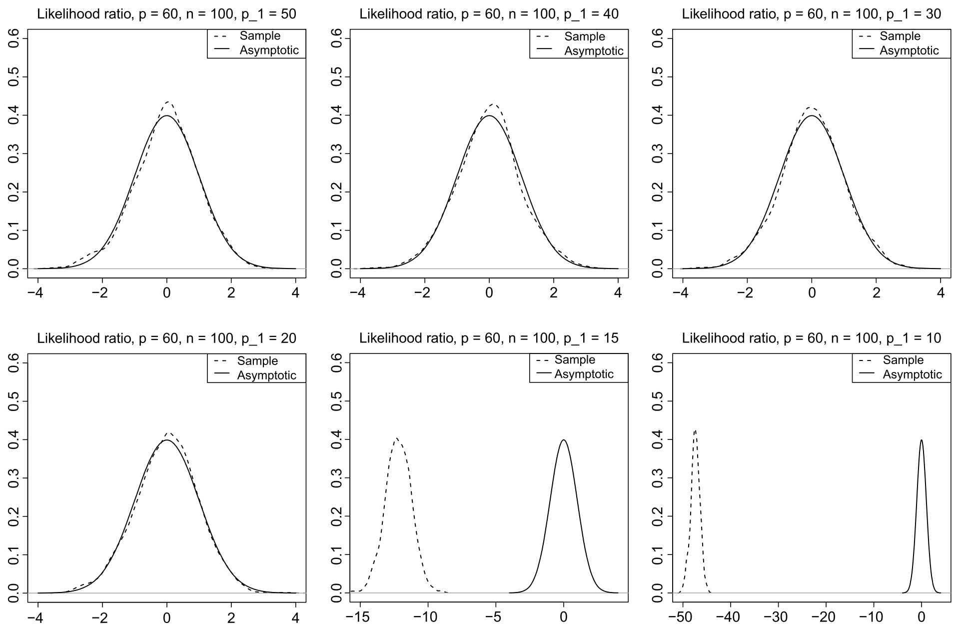

FIG. 1. *Simulated distribution of the statistic* $(T_{\mathrm{LR}} - (p - p_1)s_{\mathrm{LR}} - \mu_{\mathrm{LR}})/\sigma_{\mathrm{LR}}$ *und the null hypothesis for sample size* $n = 100$, *dimension* $p = 60$ *and various values of* $p_1 = 50, 45, 40, 30, 15, 10$. *The solid curve shows the standard normal distribution.*

3. Alternative tests for independence and their null distribution. Recall the definition of the matrices $\mathbf{\Sigma}$ and $\mathbf{S}$ in (2.1) and denote by $\mathbf{\Sigma}_{22\cdot 1} = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$ and $\mathbf{S}_{22\cdot 1} = \mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$ the corresponding Schur complements. From Theorem 3.2.10 of Muirhead (1982), it follows that

$$\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}|\mathbf{S}_{11} \sim \mathcal{N}_{p-p_1,p_1}(\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{S}_{11}^{1/2}, \mathbf{\Sigma}_{22\cdot 1} \otimes \mathbf{I}_{p_1}),$$
$$\mathbf{S}_{22\cdot 1} \sim W_{p-p_1}(n - p_1, \mathbf{\Sigma}_{22\cdot 1}),$$

and the Schur complement $\mathbf{S}_{22\cdot 1}$ is independent of $\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}$ and $\mathbf{S}_{11}$. Hence, under the null hypothesis,

$$\widehat{\mathbf{W}} = \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12} \sim W_{p-p_1}(p_1, \mathbf{\Sigma}_{22\cdot 1}),$$
$$\widehat{\mathbf{T}} = \mathbf{S}_{22\cdot 1} \sim W_{p-p_1}(n - p_1, \mathbf{\Sigma}_{22\cdot 1}),$$

and $\widehat{\mathbf{W}}$ and $\widehat{\mathbf{T}}$ are independent. Under the alternative hypothesis H_1, $\widehat{\mathbf{W}}$ and $\widehat{\mathbf{T}}$ are still independent as well as $\widehat{\mathbf{T}} \sim W_{p-p_1}(n - p_1, \mathbf{\Sigma}_{22\cdot 1})$, but $\widehat{\mathbf{W}}$ has a noncentral Wishart distribution conditionally on $\mathbf{S}_{11}$, that is,

$$\widehat{\mathbf{W}}|\mathbf{S}_{11} \sim W_{p-p_1}(p_1, \mathbf{\Sigma}_{22\cdot 1}, \mathbf{\Omega}_1(\mathbf{S}_{11}))$$

where the noncentrality parameter is given by

$$\mathbf{\Omega}_1 = \mathbf{\Omega}_1(\mathbf{S}_{11}) = \mathbf{\Sigma}_{22\cdot 1}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{S}_{11}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}.$$

For technical reasons, we will use the normalized versions of $\widehat{\mathbf{W}}$ and $\widehat{\mathbf{T}}$ throughout this paper. Thus, the distributional properties of $\mathbf{W} = \frac{1}{p_1}\widehat{\mathbf{W}}$ and $\mathbf{T} = \frac{1}{n-p_1}\widehat{\mathbf{T}}$ are very similar to the ones observed for the within and between covariance matrices in the multivariate analysis of variance (MANOVA) model (see Fujikoshi, Himeno and Wakaki (2004), Schott (2007), Kakizawa and Iwashita (2008)). More precisely, $p_1\mathbf{W}$ and $(n-p_1)\mathbf{T}$ are independent (under both hypotheses) and they possess Wishart distributions under the null hypothesis. However, under the alternative hypothesis the matrix $p_1\mathbf{W}$ has only conditionally on $\mathbf{S}_{11}$ a noncentral Wishart distribution, while the unconditional distribution appears to be a more complicated matrix-variate distribution. The similarity to MANOVA motivates the application of three tests which are usually used in this context and are given by:

(i) Wilks' Λ statistic:

$$T_W = -\log(|\mathbf{T}|/|\mathbf{T}+\mathbf{W}|) = \log(|\mathbf{I}+\mathbf{W}\mathbf{T}^{-1}|) = \sum_{i=1}^{p-p_1} \log(1+v_i); \tag{3.1}$$

(ii) Lawley–Hotelling's trace criterion:

$$T_{\mathrm{LH}} = \operatorname{tr}(\mathbf{W}\mathbf{T}^{-1}) = \sum_{i=1}^{p-p_1} v_i; \tag{3.2}$$

(iii) Bartlett–Nanda–Pillai's trace criterion:

$$T_{\mathrm{BNP}} = \operatorname{tr}(\mathbf{W}\mathbf{T}^{-1}(\mathbf{I}+\mathbf{W}\mathbf{T}^{-1})^{-1}) = \sum_{i=1}^{p-p_1} \frac{v_i}{1+v_i}, \tag{3.3}$$

where $v_1 \ge v_2 \ge \cdots \ge v_{p-p_1}$ denote the ordered eigenvalues of the matrix $\mathbf{W}\mathbf{T}^{-1}$. A statistic very similar to (3.3) was proposed by Jiang, Bai and Zheng (2013), who used

$$\operatorname{tr}\left(\mathbf{W}\mathbf{T}^{-1}\left(\frac{\gamma_1}{\gamma_2}\mathbf{I}+\mathbf{W}\mathbf{T}^{-1}\right)^{-1}\right) = \sum_{i=1}^{p-p_1} \frac{v_i}{\frac{\gamma_1}{\gamma_2}+v_i}$$

instead of $\operatorname{tr}(\mathbf{W}\mathbf{T}^{-1}(\mathbf{I}+\mathbf{W}\mathbf{T}^{-1})^{-1})$. It is remarkable that all proposed test statistics are functions of the eigenvalues of $\mathbf{W}\mathbf{T}^{-1}$ and can be presented as linear spectral statistics calculated for the random matrix $\mathbf{W}\mathbf{T}^{-1}$, which is the so-called Fisher matrix under the null hypothesis H_0 (see Zheng (2012)).

A linear spectral statistics for the matrix $\mathbf{W}\mathbf{T}^{-1}$ is generally defined by

$$\mathrm{LSS}_n = (p-p_1)\int_0^\infty f(x)\,dF_n(x) = \sum_{i=1}^{p-p_1} f(v_i), \tag{3.4}$$

where $v_1 \ge v_2 \ge \cdots \ge v_{p-p_1}$ are the ordered eigenvalues of the matrix $\mathbf{W}\mathbf{T}^{-1}$. The symbol

$$F_n(x) = \frac{1}{p-p_1}\sum_{i=1}^{p-p_1} \mathbb{1}_{(-\infty,x]}(v_i)$$

denotes the corresponding empirical spectral distribution and the symbol $\mathbb{1}_{\mathcal{A}}$ is the indicator function of the set $\mathcal{A}$. Define

$$F_n^*(dx) = q_n(x)\mathbb{1}_{[a_n,b_n]}(x)\,dx + (1-1/\gamma_{1,n})\mathbb{1}_{\gamma_{1,n}>1}\delta_0(dx) \quad \text{with}$$
$$q_n(x) = \frac{1-\gamma_{2,n}}{2\pi x(\gamma_{1,n}+\gamma_{2,n}x)}\sqrt{(b_n-x)(x-a_n)},$$
$$a_n = \frac{(1-h_n)^2}{(1-\gamma_{2,n})^2}, \qquad b_n = \frac{(1+h_n)^2}{(1-\gamma_{2,n})^2},$$

where $\gamma_{1,n}$, $\gamma_{2,n}$ and h_n are defined by (2.3), (2.4) and (2.5), respectively. Note that F_n^* is a finite sample proxy of limiting spectral distribution F of F_n, which is obtained by replacing $\gamma_{1,n}$ and $\gamma_{2,n}$ by their corresponding limits (see Bai and Silverstein (2010)), that is,

$$F(dx) = q(x)\mathbb{1}_{[a,b]}(x)\,dx + (1-1/\gamma_1)\mathbb{1}_{\gamma_1>1}\delta_0(dx) \quad \text{with} \tag{3.5}$$
$$q(x) = \frac{1-\gamma_2}{2\pi x(\gamma_1+\gamma_2 x)}\sqrt{(b-x)(x-a)}, \tag{3.6}$$
$$a = \frac{(1-h)^2}{(1-\gamma_2)^2}, \qquad b = \frac{(1+h)^2}{(1-\gamma_2)^2},$$

where

$$\gamma_1 = \lim_{n\to\infty}\gamma_{1,n} = \lim_{n\to\infty}\frac{p-p_1}{p_1}, \qquad \gamma_2 = \lim_{n\to\infty}\gamma_{2,n} = \lim_{n\to\infty}\frac{p-p_1}{n-p_1},$$
$$h = \lim_{n\to\infty}h_n = \sqrt{\gamma_1+\gamma_2-\gamma_1\gamma_2}.$$

The representations of T_W, T_{LH} and T_{BNP} in terms of the eigenvalues of the random matrix $\mathbf{W}\mathbf{T}^{-1}$ are used intensively in the proof of our first main result, which provides their asymptotic distribution under the null hypothesis in (2.8). The details of the proof are deferred to Appendix B of the Supplementary Material (see Bodnar, Dette and Parolya (2019)).

THEOREM 1. *Under the assumptions stated in Section 2, we have*

$$\frac{T_a-(p-p_1)s_\alpha-\mu_a}{\sigma_a} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

where the index $a\in\{W,\mathrm{LH},\mathrm{BNP}\}$ *represents the statistic under consideration defined in* (3.1), (3.2) *and* (3.3), *respectively. The asymptotic means and variances are given by*

$$\mu_W = 1/2\log\left[\frac{(w_n^2-d_n^2)h_n^2}{(w_nh_n-\gamma_{2,n}d_n)^2}\right], \qquad \sigma_W^2 = 2\log\left[\frac{w_n^2}{w_n^2-d_n^2}\right],$$

$$\mu_{\mathrm{LH}} = \frac{\gamma_{2,n}}{(1-\gamma_{2,n})^2}, \qquad \sigma_{\mathrm{LH}}^2 = \frac{2h_n^2}{(1-\gamma_{2,n})^4},$$

$$\mu_{\mathrm{BNP}} = -\frac{(1-\gamma_{2,n})^2 w_n^2 (d_n^2 - \gamma_{2,n})}{(w_n^2 - d_n^2)^2},$$

$$\sigma_{\mathrm{BNP}}^2 = 2\frac{d^2(1-\gamma_{2,n})^4 (w_n^2(w_n^2 + d_n) + d_n^3(w_n^2 - 1))}{w_n^2(1+d_n)(w_n^2 - d_n^2)^4},$$

where $w_n > d_n > 0$ *satisfy* $w_n^2 + d_n^2 = (1-\gamma_{2,n})^2 + 1 + h_n^2$, $w_n d_n = h_n$ *and the quantities* $\gamma_{1,n}$, $\gamma_{2,n}$ *and* h_n *are defined by* (2.3), (2.4) *and* (2.5), *respectively. The centering parameters are given by*

$$\begin{aligned} s_W = {} & -\log\bigl((1-\gamma_{2,n})^2\bigr) - \frac{1-\gamma_{2,n}}{\gamma_{2,n}} \log(w_n) \\ & + \frac{\gamma_{1,n} + \gamma_{2,n}}{\gamma_{1,n}\gamma_{2,n}} \log(w_n - d_n \gamma_{2,n}/h_n) \\ & - \begin{cases} \dfrac{1-\gamma_{1,n}}{\gamma_{1,n}} \log(w_n - d_n h_n) & \gamma_{1,n} \in (0,1), \\ 0 & \gamma_{1,n} = 1, \\ -\dfrac{1-\gamma_{1,n}}{\gamma_{1,n}} \log(w_n - d_n/h_n) & \gamma_{1,n} > 1, \end{cases} \\ s_{\mathrm{LH}} = {} & \frac{1}{1-\gamma_{2,n}}, \\ s_{\mathrm{BNP}} = {} & \frac{1-\gamma_{2,n}}{w_n^2 - \gamma_{2,n}}. \end{aligned}$$

Theorem 1 provides a simple asymptotic level α test by rejecting the null hypothesis H_0 if

$$\frac{T_a - (p - p_1)s_a - \mu_a}{\sigma_a} > u_{1-\alpha}. \tag{3.7}$$

We illustrate the quality of the approximation in Theorem 1 by means of a small simulation study. For the sake of comparison with the likelihood ratio test, we use the same scenario as in Section 2, that is, $n = 100$, $p = 60$ and different values for p_1. In Figures 2–4, we display the rejection probabilities of the test (3.7) under the null hypothesis in the case of the Wilk test, the Lawley–Hotelling's and the Bartlett–Nanda–Pillai's trace criterion. From the results depicted in Figure 2, we observe that the statistic T_W exhibits similar problems as the statistic of the likelihood ratio test. If the dimension p_1 is too small,

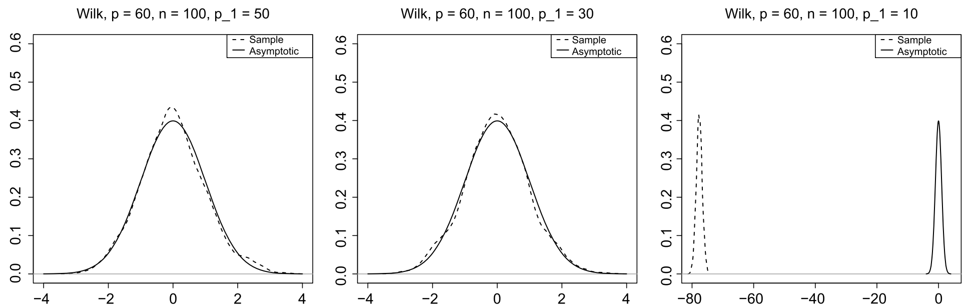

FIG. 2. *Simulated distribution of the statistic* $(T_W - (p - p_1)s_W - \mu_W)/\sigma_W$ *und the null hypothesis for sample size* $n = 100$, *dimension* $p = 60$ *and various values of* $p_1 = 50, 30, 10$. *The solid curve shows the standard normal distribution.*

the approximation provided by Theorem 1 is not reliable. This fact seems to be related to the use of the log determinant criterion. On the other hand, the Lawley–Hotelling's and the Bartlett–Nanda–Pillai's trace criterion yield test statistics which do not possess these drawbacks. The results in Figures 3 and 4 show a reasonable approximation of the nominal level in all considered scenarios.

In order to investigate the properties of two adjusted tests T_{BNP} and T_{LH} for small dimensions and small sample sizes, we provide additional results for $p = 16$, $n = 25$ and different values of $p_1 = 13, 8, 3$. The results are depicted in Figures 5 and 6 and indicate a good approximation of the nominal level although a small-sample effect is present. Note that this effect is more pronounced for the LH test as for the BNP. Thus the results are still reliable and there is again no large bias as in case of LR and Wilk's statistics when the dimension p_1 is much smaller than $p - p_1$.

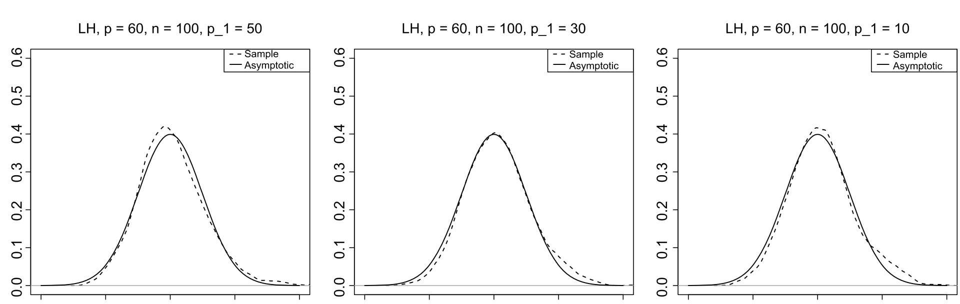

FIG. 3. *Simulated distribution of the statistic* $(T_{\mathrm{LH}} - (p - p_1)s_{\mathrm{LH}} - \mu_{\mathrm{LH}})/\sigma_{\mathrm{LH}}$ *und the null hypothesis for sample size* $n = 100$, *dimension* $p = 60$ *and various values of* $p_1 = 50, 30, 10$. *The solid curve shows the standard normal distribution.*

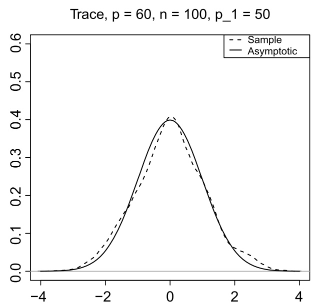

FIG. 4. *Simulated distribution of the statistic* $(T_{\mathrm{BNP}} - (p - p_1)s_{\mathrm{BNP}} - \mu_{\mathrm{BNP}})/\sigma_{\mathrm{BNP}}$ *und the null hypothesis for sample size* $n = 100$, *dimension* $p = 60$ *and various values of* $p_1 = 50, 30, 10$. *The solid curve shows the standard normal distribution*.

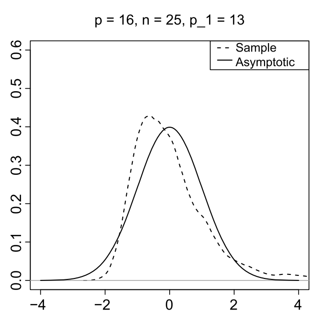

FIG. 5. *Simulated distribution of the statistic* $(T_{\mathrm{LH}} - (p - p_1)s_{\mathrm{LH}} - \mu_{\mathrm{LH}})/\sigma_{\mathrm{LH}}$ *und the null hypothesis for sample size* $n = 25$, *dimension* $p = 16$ *and various values of* $p_1 = 13, 8, 3$. *The solid curve shows the standard normal distribution.*

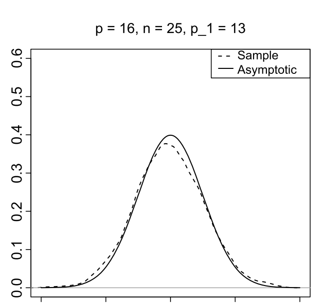

FIG. 6. *Simulated distribution of the statistic* $(T_{\mathrm{BNP}} - (p - p_1)s_{\mathrm{BNP}} - \mu_{\mathrm{BNP}})/\sigma_{\mathrm{BNP}}$ *und the null hypothesis for sample size* $n = 25$, *dimension* $p = 16$ *and various values of* $p_1 = 13, 8, 3$. *The solid curve shows the standard normal distribution.*

4. Distributional properties under alternative hypothesis. In this section, we derive the distribution of the considered linear spectral statistics under the alternative hypothesis. The main difficulty consists in the fact that under the alternative the random matrix $\mathbf{WT}^{-1}$ has a (conditionally) noncentral Fisher distribution in this case.

The following two results, which are proved in Appendix A and of independent interest, specify the asymptotic distribution of the empirical spectral distribution of the matrix $\mathbf{WT}^{-1}$ under H_1. Throughout the paper,

$$m_Q(z) = \int_{-\infty}^{+\infty} \frac{dQ(t)}{t-z}$$

denotes the Stieltjes transform of a distribution function Q.

THEOREM 2. *Consider the alternative hypothesis H_1 in (2.2) and assume that the assumptions of Section 2 are satisfied. If the the matrix $\mathbf{R} = \mathbf{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot 1}^{-1/2}$ is bounded in spectral norm and its spectral distribution converges weakly to some function G, then for any $z \in \mathbb{C}\setminus\mathbb{R}$ the Stieltjes transform of the empirical spectral distribution of the matrix $\mathbf{WT}^{-1}$ converges almost surely to some deterministic function s, which is the unique solution of the following system of equations*:

$$\begin{aligned} &\frac{s(z)}{1+\gamma_2 z s(z)} = m_H\big(z(1+\gamma_2 z s(z))\big), \\ &\frac{m_H(z)}{1+\gamma_1 m_H(z)} = m_{\tilde{H}}\big((1+\gamma_1 m_H(z))\big[(1+\gamma_1 m_H(z))z - (1-\gamma_1)\big]\big), \end{aligned} \tag{4.1}$$

$$\begin{aligned} &m_{\tilde{H}}(z)\big(1-(c-c_1)-(c-c_1)zm_{\tilde{H}}(z)\big)c_1^{-1} \\ &\qquad = m_G\left(\frac{c_1 z}{1-(c-c_1)-(c-c_1)zm_{\tilde{H}}(z)}\right), \end{aligned} \tag{4.2}$$

subject to the condition that $\Im\{s(z)\}$ is of the same sign as $\Im\{z\}$. The functions H and $\tilde{H}$ denote the limiting spectral distributions of the matrices $\mathbf{W}$ and $\tilde{\mathbf{R}} = 1/p_1 \mathbf{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{S}_{11}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot 1}^{-1/2}$, respectively.

Note that the matrix $\tilde{\mathbf{R}}$ from Theorem 2 can be interpreted as the sample covariance matrix generated from a population with the covariance matrix equal to $\frac{p_1}{n}\mathbf{R}$.

We will use this result to derive a CLT for the linear spectral statistics of the matrix $\mathbf{WT}^{-1}$, which can be used for the analysis of the test proposed in Section 3 under the alternative hypothesis. For this purpose, we introduce some useful notation as follows:

$$\delta(z) = \gamma_1 m_H(z), \tag{4.3}$$

$$\tilde{\delta}(z) = \delta(z) - \frac{1-\gamma_1}{z},$$

$$\eta(z) = \bigl(1+\delta(z)\bigr)\bigl(1+\tilde{\delta}(z)\bigr), \tag{4.4}$$

$$\xi(z) = \frac{\delta'(z)}{(z\eta(z))'}, \tag{4.5}$$

$$\Psi(z) = \left(\frac{1}{1+\delta(z)} - 2\xi(z)z + \frac{1-\gamma_1}{1+\delta(z)}\xi(z)\right)^{-1}, \tag{4.6}$$

$$r = 2\frac{(1+\sqrt{\gamma_1})^2 + \lambda_{\max}(\mathbf{R})(1+\sqrt{c_1})^2}{(1-\sqrt{\gamma_2})^2}. \tag{4.7}$$

THEOREM 3. *If the assumptions of Theorem 2 are satisfied, then for any pair f, g of analytic functions in an open region of the complex plane containing the interval $[0, r]$ the random vector*

$$\left((p-p_1)\int_0^\infty f(x)\,d\bigl(F_n(x) - F_n^*(x)\bigr), (p-p_1)\int_0^\infty g(x)\,d\bigl(F_n(x) - F_n^*(x)\bigr)\right)^\top$$

converges weakly to a Gaussian vector $(X_f, X_g)^\top$ with mean and covariances given by

$$\begin{aligned} \mathrm{E}[X_f] &= \frac{1}{4\pi i}\oint f(z)\,d\log\bigl(q(z)\bigr) + \frac{1}{2\pi i}\oint f(z)B\bigl(zb(z)\bigr)\,d\bigl(zb(z)\bigr) \\ &\quad + \frac{1}{2\pi i}\oint f(z)\theta_{b,H}(z)\Bigl(\theta_{\tilde{b},\tilde{H}}\bigl(zb(z)\bigr) \\ &\quad \times \frac{c_1^2\int \underline{m}_{\tilde{H}}^3(zb(z))t^2(c_1 + t\underline{m}_{\tilde{H}}(zb(z)))^{-3}\,dG(t)}{(1-c_1\int \underline{m}_{\tilde{H}}^2(zb(z))t^2(c_1+t\underline{m}_{\tilde{H}}(zb(z)))^{-2}\,dG(t))^2}\Bigr)\,dz, \end{aligned} \tag{4.8}$$

$$\begin{aligned} &\mathrm{Cov}[X_f, X_g] \\ &\quad = -\frac{1}{2\pi^2}\oint\oint f(z_1)g(z_2)\frac{\partial^2 \log(z_1 b(z_1) - z_2 b(z_2))}{\partial z_1\,\partial z_2}\,dz_1\,dz_2 \\ &\qquad - \frac{1}{2\pi^2}\oint\oint f(z_1)g(z_2) \\ &\qquad \times \frac{\partial^2 \log(z_1 b(z_1)\eta(z_1 b(z_1)) - z_2 b(z_2)\eta(z_2 b(z_2)))}{\partial z_1\,\partial z_1}\,dz_1\,\mathrm{d}z_2 \\ &\qquad - \frac{1}{2\pi^2}\oint\oint f(z_1)g(z_2)\Bigl[\theta_{\tilde{b},\tilde{H}}\bigl(z_1 b(z_1)\bigr)\theta_{\tilde{b},\tilde{H}}\bigl(z_2 b(z_2)\bigr) \\ &\qquad \times \left(\frac{\partial^2 \log[\frac{\underline{m}_{\tilde{H}}(z_2 b(z_2)) - \underline{m}_{\tilde{H}}(z_1 b(z_1))}{(z_2 b(z_2) - z_1 b(z_1))}]}{\partial z_1\,\partial z_2}\right)\Bigr]\,\mathrm{d}z_1\,\mathrm{d}z_2 \end{aligned} \tag{4.9}$$

respectively, where

$$
\begin{aligned}
& b(z) = 1 + \gamma_2 z s(z), \\
(4.10) \quad & \tilde{b}(z) = 1 + \gamma_1 m_H(z), \\
& q(z) = 1 - \gamma_2 \int \frac{b^2(z)\, dH(t)}{(t/z - b(z))^2}, \\
& \theta_{\tilde{b},\tilde{H}}(z) \\
& \quad = \tilde{b}(z) \Big/ \Big(1 - \gamma_1 m_{\tilde{H}}\big(\tilde{b}(z)\big(\tilde{b}(z)z - (1-\gamma_1)\big)\big) \\
& \qquad - \tilde{b}(z)\gamma_1\big(2z\tilde{b}(z) - (1-\gamma_1)\big) \int \frac{d\tilde{H}(t)}{[t - (\tilde{b}(z)(\tilde{b}(z)z - (1-\gamma_1)))]^2} \Big) \\
& \underline{m}_{\tilde{H}}(z) = -\frac{1-c_1}{z} + c_1 m_{\tilde{H}}(z) \\
& B(z) = \Psi^2(z)\Big[-\widetilde{\omega}(z) N(z)\big(1-\delta(z)\big) + \frac{1}{1+\delta(z)} N(z) + \xi(z)\Psi^{-1}(z) \\
(4.11) \quad & \qquad + z\xi^2(z) + z^2\tilde{\delta}^2(z)\Big(\xi^2(z) - \delta(z)N(z)\Big(z - \frac{1-\gamma_1}{1+\delta(z)} + 1\Big)\Big)\Big]
\end{aligned}
$$

with

$$
N(z) = \frac{\xi'(z)\Psi^{-1}(z)}{2} - \xi^2(z) \quad \textit{and} \quad \widetilde{\omega}(z) = z^2\xi(z) + \frac{1-\gamma_1}{1+\delta(z)}\Psi^{-1}(z).
$$

Here, the integrals are taken over an arbitrary positively oriented contour which contains the interval $[0, r]$, *moreover the contours in* (4.9) *are nonoverlapping*.

There are substantial differences between the CLT derived here and the recent results in Zheng, Bai and Yao (2017). In particular, the matrix $\mathbf{W}$ does not possess the usual properties of the covariance matrix under normality anymore. Indeed, the conditional distribution of $\mathbf{W}$ given $\mathbf{S}_{11}$ is a noncentral Wishart distribution, while the unconditional distribution is defined by a very complicated integral expression. As a consequence $\mathbf{W}\mathbf{T}^{-1}$ can be interpreted as a conditionally noncentral Fisher matrix, while Zheng, Bai and Yao (2017) considered a rescaled Fisher matrix. In general, the CLT presented in Zheng, Bai and Yao (2017) is constructed for studying the asymptotic power of the test for the equality of two population covariance matrices. In contrast, the CLT derived in Theorem 3 is used to investigate the power of the test for block-diagonality, that is, $H_0 : \mathbf{\Sigma}_{12} = \mathbf{O}$.

It follows from the proof of Theorem 2 that

$$
\mathbf{W} \overset{d}{\leq} 2\Big(\frac{1}{p_1}\mathbf{X}\mathbf{X}^\top + \mathbf{M}\mathbf{M}^\top\Big), \tag{4.12}
$$

where $n\mathbf{M}\mathbf{M}^\top \sim \mathcal{W}_{p-p_1}(n, \mathbf{R})$ and all entries of $\mathbf{X}$ are independent and standard normally distributed. Consequently, the largest eigenvalue of the matrix $\mathbf{W}$ will asymptotically be smaller than

$$2\big((1+\sqrt{\gamma_1})^2 + \lambda_{\max}(\mathbf{R})(1+\sqrt{c_1})^2\big)$$

and the quantity r defined in (4.7) is an upper bound for the limiting spectrum of the matrix $\mathbf{W}\mathbf{T}^{-1}$.

This observation is quite important for controlling the tail estimates of the extreme eigenvalues of the matrix $\mathbf{W}\mathbf{T}^{-1}$, which play a vital role for the application of the Cauchy's integral formula (A.19) at the end of the proof of Theorem 3. The proof of the following result is given in Appendix A.

PROPOSITION 2. *Let $l_r > r$, where r is given in* (4.7), *then*

$$\forall k \in \mathbb{N}: \quad \mathbb{P}\big(\lambda_{\max}(\mathbf{W}\mathbf{T}^{-1}) > l_r\big) = o\big(n^{-k}\big).$$

Although the limiting mean and variance presented in Theorem 3 are very difficult to calculate in a closed form even for simple cases, there are several important implications of Theorem 3.

REMARK 1 (Eigenvectors). Going through the proof of Theorem 3 one can see that Lemma 1 in Appendix A reveals an interesting though quite expected fact that the resulting asymptotic distributions depend neither on the eigenvectors of the noncentrality matrix $\mathbf{\Omega}_1$ nor on the eigenvectors of the matrix $\mathbf{R} = \mathbf{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot 1}^{-1/2}$ for the normally distributed data. Loosely speaking, without loss of generality (w.l.g.), we can restrict ourselves to the case when $\mathbf{\Omega}_1$ and $\mathbf{R}$ are diagonal matrices, which simplifies the simulations in a remarkable way.

REMARK 2 (Generalizations and simplifications). The noncentral Fisher matrix in our case arises only conditionally on $\mathbf{S}_{11}$ where the noncentrality matrix $\mathbf{\Omega}_1$ is random in our framework. As a consequence, Theorem 3 generalizes the result of Yao (2013), where a deterministic noncentrality matrix was considered. Moreover, all the asymptotic quantities including $\delta(z)$ are expressed in a more convenient form, like $\delta(z) = \gamma_1 m_H(z)$. Finally, the expression of the bias term $B(z)$ is significantly simplified which makes it possible to do numerical computations more efficiently and to investigate the results of Theorem 3 deeper in the future.

REMARK 3 (Finite rank alternatives). Combining Theorem 2 and Theorem 3, one observes that finite rank alternatives with a bounded spectrum have no influence on the asymptotic power of the tests, because the asymptotic means and variances under the null hypothesis and alternative hypothesis coincide. Indeed, assuming that the matrix $\mathbf{R}$ has a finite rank, say k, and a bounded spectrum we get

$$m_{F^{\mathbf{R}}}(z) = \int \frac{dF^{\mathbf{R}}(t)}{t-z} = \frac{1}{p-p_1} \sum_{i=1}^{p-p_1} \frac{1}{\lambda_i(\mathbf{R}) - z}$$
$$= \frac{1}{p-p_1} \sum_{i=1}^{k} \frac{1}{\lambda_i(\mathbf{R}) - z} - \frac{p-p_1-k}{p-p_1}\frac{1}{z} \to -\frac{1}{z}.$$

Thus, it follows that $m_G(z) = -\frac{1}{z}$ and, therefore, G is the distribution function of the Dirac measure concentrated at the point 0. Consequently, we obtain $\underline{m}_{\tilde{H}}(z) = -1/z$ and the third summands in (4.8) and in (4.9) vanish, that is,

$$\int \frac{t^2}{(c_1 + t\underline{m}_{\tilde{H}}(z))^3} \, dG(t) = \int \frac{t^2}{(c_1 + t\underline{m}_{\tilde{H}}(z))^3} \delta_0(t) d(t) = 0,$$
$$\frac{\partial^2 \log[\frac{\underline{m}_{\tilde{H}}(z_2) - \underline{m}_{\tilde{H}}(z_1)}{z_2 - z_1}]}{\partial z_1 \, \partial z_2} = \frac{\underline{m}'_{\tilde{H}}(z_1)\underline{m}'_{\tilde{H}}(z_2)}{(\underline{m}_{\tilde{H}}(z_1) - \underline{m}_{\tilde{H}}(z_2))^2} - \frac{1}{(z_1 - z_2)^2} = 0,$$

for any $z, z_1, z_2 \in \mathbb{C}^+$. The other summands in (4.8) and in (4.9) do not depend on the eigenvalues of matrix $\mathbf{R}$, which reflects the alternative hypothesis H_1 via $\mathbf{\Sigma}_{12}$, thus they are expected to be equal to the corresponding quantities under the null hypothesis H_0 given in Theorem 1. Consequently, all tests based on a linear spectral statistic cannot detect the alternative hypothesis H_1 if the matrix $\mathbf{R}$ has no large eigenvalues.

On the other hand, if $\lambda_{\max}(\mathbf{R})$ is an increasing function of the dimension $p - p_1$ the spectrum of $\lambda_{\max}(\mathbf{R})$ is not bounded and Theorem 3 is not applicable. Although we have no theoretical result in this case, we expect that the power of the tests will be an increasing function of $\lambda_{\max}(\mathbf{R})$. These properties have been verified numerically by means of a simulation study.

REMARK 4 (Full rank alternatives). As we have already mentioned, the formulas in Theorem 2 and Theorem 3 are very complex, which makes it difficult to calculate the power functions of the considered tests in an analytic form. For instance, we need to solve the system of three equations in Theorem 2 which leads to the cubic equation already for $m_H(z)$ even in the simple case $\mathbf{R} = \rho^2\mathbf{I}$. On the other hand, the whole system in Theorem 2 simplifies to a quadratic equation under the null hypothesis H_0. Nevertheless, we believe that these results may be useful for future investigations of the power of the considered tests on the block diagonality of the covariance matrix. For example, one may consider the numerical approximations discussed in Zheng, Bai and Yao (2017).

To illustrate these remarks and comments, we present a comparison of the power of the different tests under consideration by means of a small simulation study.

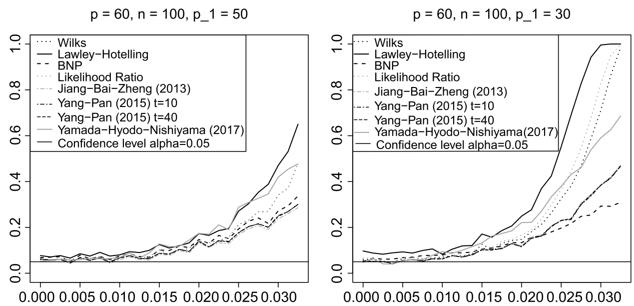

FIG. 7. *Empirical power of different tests for block diagonality for sample size* $n = 100$, *dimension* $p = 60$ *and various values of* $p_1 = 50, 30, 10$ *as a function of the correlation coefficient* $\rho = \frac{\sigma_{12}}{\sigma}$ *in* $[0, 0.0325]$.

In order to demonstrate the results in a clear way, we assume for simplicity that $\mathbf{\Sigma}_{11} = \mathbf{\Sigma}_{22} = \sigma\mathbf{I}$ which yields

$$\mathbf{R} = \frac{1}{\sigma}\left(\sigma\mathbf{I} - \frac{1}{\sigma}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{12}\right)^{-1/2}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{12}\left(\sigma\mathbf{I} - \frac{1}{\sigma}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{12}\right)^{-1/2}.$$

Note that the spectrum of matrix $\mathbf{R}$ is the same as that of the matrix $\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{12}(\sigma^2\mathbf{I} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{12})^{-1}$. First, we take $\mathbf{\Sigma}_{12}$ as a rank 1 matrix with all components equal to $\sigma_{12} \in [0, 1.3]$ (equicorrelation) and in order to assure positive definiteness of $\mathbf{\Sigma}$ in that range we choose $\sigma = 40$. Note that if σ_{12} varies in the interval $[0, 1.3]$ the correlation coefficient $\rho = \sigma_{12}/\sigma$ will change in the interval $[0, 0.0325]$. Further, we increase the rank of $\mathbf{\Sigma}_{12}$ by setting some of its elements to zero (sparsifying). The empirical rejection probailities of the proposed tests in the case of rank 1 alternatives are given in Figure 7.

For the sake of comparison, we also included the trace criterion recently proposed by Jiang, Bai and Zheng (2013), the test introduced by Yamada, Hyodo and Nishiyama (2017), which is based on an empirical distance between the full and a block diagonal covariance matrix; and the test suggested by Yang and Pan (2015) built on the sum of the canonical correlations coefficients. Note that there exists a regularized and a nonregularized version of the latter test. In general, the statistic of Yang and Pan (2015) is defined by the sum of eigenvalues of the matrix $(\mathbf{S}_{22} + t\mathbf{I}_{p-p_1})^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$ for some $t \geq 0$. Thus, in case $t = 0$ this test is equivalent to the sum of canonical correlation coefficients. Moreover, the test of Jiang, Bai and Zheng (2013) and Yang and Pan (2015) coincide for $t = 0$ because the matrix $\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$ can be written in form $\mathbf{W}\mathbf{T}^{-1}(\frac{\gamma_1}{\gamma_2}\mathbf{I} + \mathbf{W}\mathbf{T}^{-1})^{-1}$ under H_0 (see, e.g., Yao, Zheng and Bai (2015), Section 8.5.1). Thus, in order to visualize difference between them we take $t = 10$ and $t = 40$ for the test proposed by Yang and Pan (2015). Further, the simulations showed that taking larger t will lead only to a slight increase of power in the case where p_1 is not equal to $p - p_1$. Nevertheless, it must be mentioned that the regularized test is applicable even in the case $p - p_1 > n$ while all of other considered tests need both $p_1 < n$ and $p - p_1 < n$.

Figure 7 justifies our theoretical findings, that is, none of the tests can detect the alternatives for $\rho \in [0, 0.01]$ (the power function in this region is basically flat and close to the nominal level 0.05). On the other hand, if the correlation is greater than 0.01, then all of the tests gain power. For $p_1 = 30$ (case of equal blocks) all test are powerful enough to reject H_0 if the correlation is greater than 0.03. These results are in accordance with the discussion in Remark 3, because in the considered scenario the largest eigenvalue of the matrix $\mathbf{R}$ is given by

$$\frac{p_1(p - p_1)\rho^2}{1 - p_1(p - p_1)\rho^2}.$$

Thus, if the correlation coefficient ρ is close to $1/\sqrt{p_1(p - p_1)}$ we will get a spike (note that $1/\sqrt{p_1(p - p_1)} \approx 0.0333$ if $p_1 = 30$, $p = 60$). Moreover, here we have a clear winner—the Lawley–Hotelling's (LH) trace criterion. The test of Yamada, Hyodo and Nishiyama (2017) and Wilk's test with the corrected likelihood ratio (LR) criterion are ranked on the second and third, respectively. The regularized test of Yang and Pan (2015) together with the trace criterion of Jiang, Bai and Zheng (2013) are on the fourth position, while the Bartlett–Nanda–Pillai's (BNP) trace criterion shows the worst performance. Interestingly, the tests Jiang, Bai and Zheng (2013) and Yang and Pan (2015) cannot be visually distinguished neither for $t = 10$ nor for $t = 40$.

A similar ranking was observed for $p_1 = 50$ with the difference of a decreasing power of all tests and a slight increase of power for Yang and Pan (2015) with respect to its benchmark for $t = 0$, that is, Jiang, Bai and Zheng (2013). Note that Wilk's test and the LR test have the same power for $p_1 = 50$. In light of the previous findings obtained under the null hypothesis H_0, the case $p_1 = 10$ is the most interesting one. Indeed, here we observe that Wilk's and the LR test are not reliable anymore (they either always (Wilk's) or never (LR) reject H_0). On the other hand, the other tests show a similar behaviour as in the case $p_1 = 50$. As before, the LH test is the most powerful in all three situations.

In order to investigate the robustness of the tests, we increase the sparsity of the matrix $\mathbf{\Sigma}_{12}$, where 20% and 50% of the elements are set randomly to zero, while all other elements are still equal to σ_{12}. By this procedure, we increase the probability that $\mathbf{\Sigma}_{12}$ has full rank. The results are summarized in Figures 8 and 9.

We observe a similar behaviour as in the nonsparse case (see Figure 7). The LH test and the test proposed in Yamada, Hyodo and Nishiyama (2017) show the best performance. The latter is slightly better than the LH test for the sparsity level 50%, while a superiority of the LH test could be observed for a sparsity level of 20%. Of course, by increasing the sparsity level we make the alternative hypothesis harder to detect. For this reason the nonsensitivity interval $[0, 0.01]$ (the interval where the test is not sensitive to the alternative H_1) is increased to $[0, 0.02]$ and $[0, 0.03]$ in case of 20% and 50% sparsity levels, respectively.

Moreover, in the Supplementary Material (see Bodnar, Dette and Parolya (2019)) we have also investigated the performance of the different tests for $p = 10$,

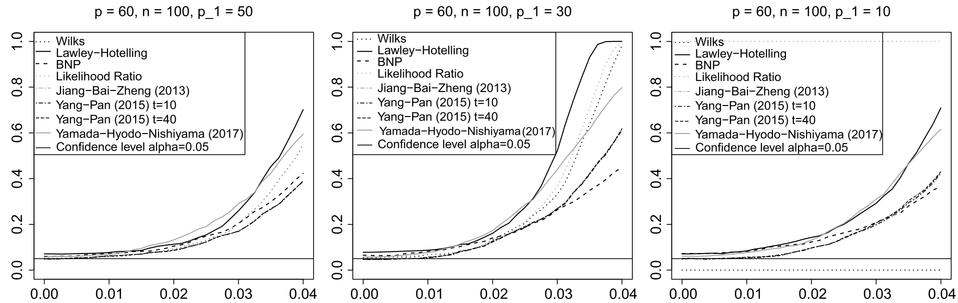

FIG. 8. *Empirical power of different tests for block diagonality for sample size* $n = 100$, *dimension* $p = 60$ *and various values of* $p_1 = 50, 30, 10$ *as a function of the correlation coefficient* $\rho = \frac{\sigma_{12}}{\sigma}$ *in* $[0, 0.04]$ *and sparsity level of* 20%.

100, 200, 250, 290 and $n = 300$ (see Figures 10–14 in Appendix B of the Supplementary Material). Our findings still remain unchanged except for the case $p = 290$: here, all tests have a substantial loss in power and the LH test does not keep the nominal level. Indeed, this is an expected result, because in the case $p = 290$ and $n = 300$ the ratio p/n is close to one, and the sample covariance matrix is a very unstable estimator.

As a conclusion, although the LH trace criterion is the most simple one among the linear spectral statistics of the matrix $\mathbf{WT}^{-1}$ ($f = id$), it seems to be the most robust and powerful test on the block diagonality of the large-dimensional covariance matrix. On the other hand, the corrected LR and Wilk's criteria cannot be recommended, if the size of the first block is much smaller than the size of the second one.

REMARK 5. A possible reason for the superior performance of the LH test are the specific alternatives considered in our numerical experiments. In particular, results coincide with the findings in Pillai and Jayachandran (1967), where the

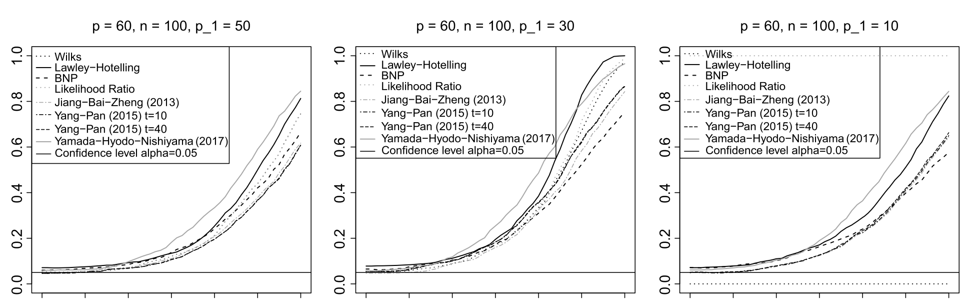

FIG. 9. *Empirical power of different tests for block diagonality for sample size* $n = 100$, *dimension* $p = 60$ *and various values of* $p_1 = 50, 30, 10$ *as a function of the correlation coefficient* $\rho = \frac{\sigma_{12}}{\sigma}$ *in* $[0, 0.06]$ *and sparsity level of* 50%.

LH test showed the best performance in the case, where the eigenvalues of the matrix $\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ are far apart. Thus, taking very sparse or other full rank alternatives could have an considerable impact on the dominance property of the LH test.

APPENDIX A: PROOFS

PROOF OF PROPOSITION 2. Because $\mathbf{W}$ and $\mathbf{T}$ are positive semi-definite we have

$$\begin{aligned}\mathbb{P}(\lambda_{\max}(\mathbf{W}\mathbf{T}^{-1}) > l_r) &\le \mathbb{P}(\lambda_{\max}(\mathbf{W})\lambda_{\max}(\mathbf{T}^{-1}) > l_r) \\ &\le \mathbb{P}(\lambda_{\max}(\mathbf{W}) > l_r(1-\sqrt{\gamma_2})^2) + o(n^{-k}),\end{aligned} \tag{A.1}$$

where the last inequality follows from inequality (1.9b) in Bai and Silverstein (2004). Furthermore, using this inequality again and (4.12) we get

$$\begin{aligned}&\mathbb{P}(\lambda_{\max}(\mathbf{W}) > l_r(1-\sqrt{\gamma_2})^2) \\ &\quad\le \mathbb{P}(2\lambda_{\max}(1/p_1\mathbf{X}\mathbf{X}^\top) + 2\lambda_{\max}(\mathbf{M}\mathbf{M}^\top) > l_r(1-\sqrt{\gamma_2})^2) \\ &\quad\le \mathbb{P}\left(\lambda_{\max}(1/p_1\mathbf{X}\mathbf{X}^\top) > \frac{l_r}{2}(1-\sqrt{\gamma_2})^2 - \lambda_{\max}(\mathbf{R})(1+\sqrt{c_1})^2\right) \\ &\quad\quad + o(n^{-k}).\end{aligned} \tag{A.2}$$

Finally, combining (A.1), (A.2) and using (4.7) with $l_r > r$, we arrive at

$$\mathbb{P}(\lambda_{\max}(\mathbf{W}\mathbf{T}^{-1}) > l_r) \le \mathbb{P}(\lambda_{\max}(1/p_1\mathbf{X}\mathbf{X}^\top) > (1+\sqrt{\gamma_1})^2) + o(n^{-k}) = o(n^{-k}),$$

where the last equality follows again from (1.9a) of Bai and Silverstein (2004). $\square$

PROOF OF THEOREM 2. Since $(n-p_1)\mathbf{T} \sim W_{p-p_1}(n-p_1, \boldsymbol{\Sigma}_{22\cdot 1})$, $p_1\mathbf{W}|\mathbf{S}_{11} \sim W_{p-p_1}(p_1, \boldsymbol{\Sigma}_{22\cdot 1}, \boldsymbol{\Omega}_1)$ with $\boldsymbol{\Omega}_1 = \boldsymbol{\Omega}_1(\mathbf{S}_{11}) = \boldsymbol{\Sigma}_{22\cdot 1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{S}_{11}\boldsymbol{\Sigma}_{11}^{-1} \times \boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{22\cdot 1}^{-1}\mathbf{M}\mathbf{M}^\top$, $\mathbf{S}_{11} \sim W_{p_1}(n, \boldsymbol{\Sigma}_{11})$, and $\mathbf{T}$ is independent of $\mathbf{W}$ and $\mathbf{S}_{11}$ we get the following stochastic representations for $\mathbf{T}$ and $\mathbf{W}$ expressed as[2]

$$\begin{aligned}\mathbf{W} &\overset{d}{=} \frac{1}{p_1}\boldsymbol{\Sigma}_{22\cdot 1}^{1/2}(\mathbf{X} + \boldsymbol{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{M})(\mathbf{X} + \boldsymbol{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{M})^\top\boldsymbol{\Sigma}_{22\cdot 1}^{1/2}, \\ \mathbf{T} &\overset{d}{=} \frac{1}{n-p_1}\boldsymbol{\Sigma}_{22\cdot 1}^{1/2}\mathbf{Y}\mathbf{Y}^\top\boldsymbol{\Sigma}_{22\cdot 1}^{1/2},\end{aligned}$$

[2]Here, we use the definition of the noncentral Wishart distribution given by Muirhead (1982), Definition 10.3.1.

where $\mathbf{X} \sim \mathcal{N}_{p-p_1,p_1}(\mathbf{O}, \mathbf{I} \otimes \mathbf{I})$, $\mathbf{Y} \sim \mathcal{N}_{p-p_1,n-p_1}(\mathbf{O}, \mathbf{I} \otimes \mathbf{I})$, and $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{S}_{11}$ are mutually independent. Then the stochastic representation of $\mathbf{W}\mathbf{T}^{-1}$ is given by

$$\mathbf{W}\mathbf{T}^{-1} \stackrel{d}{=} \frac{1}{p_1}\boldsymbol{\Sigma}_{22\cdot 1}^{1/2}(\mathbf{X} + \boldsymbol{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{M})(\mathbf{X} + \boldsymbol{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{M})^{\top}\boldsymbol{\Sigma}_{22\cdot 1}^{1/2} \times \left(\frac{1}{n-p_1}\boldsymbol{\Sigma}_{22\cdot 1}^{1/2}\mathbf{Y}\mathbf{Y}^{\top}\boldsymbol{\Sigma}_{22\cdot 1}^{1/2}\right)^{-1}.$$

The last equality in distribution implies that the spectral distribution of $\mathbf{W}\mathbf{T}^{-1}$ is the same as the spectral distribution of $\widetilde{\mathbf{W}}\widetilde{\mathbf{T}}^{-1}$ with

$$\widetilde{\mathbf{W}} = \frac{1}{p_1}(\mathbf{X} + \boldsymbol{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{M})(\mathbf{X} + \boldsymbol{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{M})^{\top} \quad \text{and} \quad \widetilde{\mathbf{T}} = \frac{1}{n-p_1}\mathbf{Y}\mathbf{Y}^{\top}.$$

From Theorem 2.1 of Zheng, Bai and Yao (2015a), it holds that the Stieltjes transform of $\widetilde{\mathbf{W}}\widetilde{\mathbf{T}}^{-1}$ given $\widetilde{\mathbf{W}}$ $m_{F^{\widetilde{\mathbf{W}}\widetilde{\mathbf{T}}^{-1}}|\widetilde{\mathbf{W}}}(z)$ converges to $s_{\widetilde{\mathbf{W}}}(z)$ which satisfies the following equation:

$$zs_{\widetilde{\mathbf{W}}}(z) = -1 + \int \frac{t\,dH(t)}{t - z(1 + \gamma_2 z s_{\widetilde{\mathbf{W}}}(z))}, \tag{A.3}$$

where $H(t) = H_{\widetilde{\mathbf{W}}}(t)$ is the limiting spectral distribution of the matrix $\widetilde{\mathbf{W}}$, which is a deterministic function following Theorem 1.1 of Dozier and Silverstein (2007). Noting that the right-hand side of (A.3) does not depend on the condition $\widetilde{\mathbf{W}}$ and rewriting (A.3), we get the limiting spectral distribution of $\mathbf{W}\mathbf{T}^{-1}$, which is equal to $\widetilde{\mathbf{W}}\widetilde{\mathbf{T}}^{-1}$, is given by $s(z) = s_{\widetilde{\mathbf{W}}}(z)$ expressed as

$$zs(z) = \int \frac{z\gamma_2(zs(z)+1)\,dH(t)}{t - z(1+\gamma_2 zs(z))} = z(\gamma_2 zs(z)+1)m_H(z(\gamma_2 zs(z)+1)),$$

where (see Theorem 1.1 of Dozier and Silverstein (2007))

$$\begin{aligned} m_H(z) &= \int \frac{(1+\gamma_1 m_H(z))\,d\tilde{H}(t)}{t - (1+\gamma_1 m_H(z))[(1+\gamma_1 m_H(z))z - (1-\gamma_1)]} \\ &= (1+\gamma_1 m_H(z)) m_{\tilde{H}}\big((1+\gamma_1 m_H(z))\big[(1+\gamma_1 m_H(z))z - (1-\gamma_1)\big]\big) \end{aligned}$$

with $\tilde{H}$ the limiting spectral distribution of

$$\begin{aligned} \widetilde{\mathbf{R}} &= 1/p_1 \boldsymbol{\Sigma}_{22\cdot 1}^{-1/2}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{S}_{11}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22\cdot 1}^{-1/2} \\ &= c_{1,n}^{-1} 1/n \boldsymbol{\Sigma}_{22\cdot 1}^{-1/2}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{S}_{11}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22\cdot 1}^{-1/2}, \end{aligned}$$

satisfying the following equation:

$$m_{\tilde{H}}(z) = \int \frac{(1-(c-c_1)-(c-c_1)zm_{\tilde{H}}(z))^{-1}\,dG(t)}{c_1^{-1}t - \frac{z}{1-(c-c_1)-(c-c_1)zm_{\tilde{H}}(z)}}$$

$$= c_1^{-1}\big(1-(c-c_1)-(c-c_1)zm_{\tilde{H}}(z)\big)^{-1}$$
$$\times m_G\left(\frac{c_1 z}{1-(c-c_1)-(c-c_1)zm_{\tilde{H}}(z)}\right)$$

where $G(t)$ is the limiting spectral distribution of the matrix $\mathbf{R} = \mathbf{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1} \times \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot 1}^{-1/2}$ which is deterministic as well. □

In the proof of Theorem 3, we make use of the following lemma which simplifies the conditions used in Theorem 2.2 of Zheng, Bai and Yao (2015a) and is proved in Appendix B of the Supplementary Material (see Bodnar, Dette and Parolya (2019)).

LEMMA 1. *Conditionally on* $\mathbf{S}_{11}$, *the distribution of the matrix* $\mathbf{WT}^{-1}$ *solely depends on the eigenvalues of the noncentrality matrix* $\mathbf{\Omega}_1(\mathbf{S}_{11})$ *and does not depend on the corresponding eigenvectors. Moreover, the unconditional distribution of the eigenvalues of matrix* $\mathbf{WT}^{-1}$ *depends only on the eigenvalues of the matrix* $\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot 1}^{-1}\mathbf{\Sigma}_{21}$.

The results of Lemma 1 shows that both the unconditional distribution of the eigenvalues of $\mathbf{WT}^{-1}$ and its conditional distribution given $\mathbf{S}_{11}$ depend only on the eigenvalues of $\mathbf{\Omega}_1(\mathbf{S}_{11})$ and of $\widetilde{\mathbf{R}} = 1/p_1\mathbf{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{S}_{11}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot 1}^{-1/2}$, respectively, for any fixed dimension p and sample size n. Consequently, without loss of generality both matrices $\mathbf{\Omega}_1(\mathbf{S}_{11})$ and of $\widetilde{\mathbf{R}}$ can be taken as diagonal. These simplify the validation of the conditions present in Theorem 2.2.1 and Theorem 2.2.2 of Yao (2013).

PROOF OF THEOREM 3. Throughout the proof of Theorem 3, we assume that the complex number z belongs to the arbitrary positively oriented contour $\mathcal{C}$, which contains the limiting support $[0, r]$. We consider

$$(p-p_1)\big(m_{F^{\mathbf{WT}^{-1}}}(z)-s_n(z)\big)$$
$$= (p-p_1)\big(m_{F^{\mathbf{WT}^{-1}}}(z)-s_n^*(z)\big)+(p-p_1)\big(s_n^*(z)-s_n(z)\big), \tag{A.4}$$

where $s_n(z)$ and $s_n^*(z)$ are unique roots of the following equations:

$$zs_n(z) = -1+\int\frac{t\,dH_n(t)}{t-z(1+\gamma_{2,n}zs_n(z))}, \tag{A.5}$$

$$zs_n^*(z) = -1+\int\frac{t\,dF_n^{\mathbf{W}}(t)}{t-z(1+\gamma_{2,n}zs_n^*(z))} \tag{A.6}$$

with $\gamma_{2,n} = \frac{p-p_1}{n-p_1}$. The symbol H_n denotes the discretized limiting distribution of $\mathbf{W}$ with γ_2 replaced by $\gamma_{2,n}$ and $F_n^{\mathbf{W}}$ stands for the empirical spectral distribution of $\mathbf{W}$.

Following the proof of Theorem 2.2 by Zheng, Bai and Yao (2015a), we get that the first summand $(p-p_1)(m_{F^{\mathbf{WT}^{-1}}}(z)-s_n^*(z))$ in (A.4) conditionally on the matrix $\mathbf{W}$ converges to a Gaussian process $M_1(z)$ with the mean function

$$\mathrm{E}\big(M_1(z)\big)=\frac{\gamma_2 b^3(z)}{z^2q^2(z)}\int\frac{t\,dH(t)}{(t/z-b(z))^3}=\frac{1}{2}\big(\log\big(q(z)\big)\big)' \tag{A.7}$$

and the covariance function

$$\begin{aligned}\operatorname{Cov}\big(M_1(z_1),M_1(z_2)\big)&=2\frac{(z_1b(z_1))'(z_2b(z_2))'}{(z_1b(z_1)-z_2b(z_2))^2}\\&=2\frac{\partial\log((z_1b(z_1)-z_2b(z_2)))}{\partial z_1\,\partial z_2},\end{aligned} \tag{A.8}$$

where

$$\begin{aligned}b(z)&=1+\gamma_2 zs(z),\\q(z)&=1-\gamma_2\int\frac{b^2(z)\,dH(t)}{(t/z-b(z))^2}\end{aligned} \tag{A.9}$$

for z_1 and z_2 from $\mathcal{C}$. Since all quantities in (A.7)–(A.9) do not depend on the condition $\mathbf{W}$, we get that this is also the unconditional distribution and both summands in (A.4) are independent.

Next, we derive the asymptotic distribution of the second summand $(p-p_1)(s_n^*(z)-s_n(z))$ in (A.4). Let

$$b_n^*(z)=1+\gamma_{2,n}zs_n^*(z)\quad\text{and}\quad b_n(z)=1+\gamma_{2,n}zs_n(z).$$

Then, by using the definition of the Stieltjes transform, (A.5), and (A.6) we get

$$\begin{aligned}&(p-p_1)\big(s_n^*(z)-s_n(z)\big)\\&\quad=(p-p_1)\big(b_n^*(z)m_{F_n^{\mathbf{W}}}\big(zb_n^*(z)\big)-b_n(z)m_{H_n}\big(zb_n(z)\big)\big)\\&\quad=(p-p_1)\big(b_n^*(z)-b_n(z)\big)m_{F_n^{\mathbf{W}}}\big(zb_n^*(z)\big)\\&\qquad+(p-p_1)b_n(z)\big(m_{F_n^{\mathbf{W}}}\big(zb_n^*(z)\big)-m_{F_n^{\mathbf{W}}}\big(zb_n(z)\big)\big)\\&\qquad+(p-p_1)b_n(z)\big(m_{F_n^{\mathbf{W}}}\big(zb_n(z)\big)-m_{H_n}\big(zb_n(z)\big)\big)\\&\quad=(p-p_1)\gamma_{2,n}z\big(s_n^*(z)-s_n(z)\big)m_{F_n^{\mathbf{W}}}\big(zb_n^*(z)\big)\\&\qquad+(p-p_1)b_n(z)\gamma_{2,n}z^2\big(s_n^*(z)-s_n(z)\big)\int\frac{dF_n^W(t)}{(t-zb_n^*(z))(t-zb_n(z))}\\&\qquad+(p-p_1)b_n(z)\big(m_{F_n^{\mathbf{W}}}\big(zb_n(z)\big)-m_{H_n}\big(zb_n(z)\big)\big).\end{aligned}$$

Hence,

$$
\begin{aligned}
&(p-p_1)\big(s_n^*(z)-s_n(z)\big)\\
&\quad=(p-p_1)\big(m_{F_n^{\mathbf{W}}}\big(zb_n(z)\big)-m_{H_n}\big(zb_n(z)\big)\big)\\
&\qquad\times\frac{b_n(z)}{1-\gamma_{2,n}zm_{F_n^{\mathbf{W}}}(zb_n^*(z))-b_n(z)\gamma_{2,n}z^2\int\frac{dF_n^W(t)}{(t-zb_n^*(z))(t-zb_n(z))}},
\end{aligned}
$$

where

$$
\begin{aligned}
&\frac{b_n(z)}{1-\gamma_{2,n}zm_{F_n^{\mathbf{W}}}(zb_n^*(z))-b_n(z)\gamma_{2,n}z^2\int\frac{dF_n^W(t)}{(t-zb_n^*(z))(t-zb_n(z))}}\\
&\quad\stackrel{a.s.}{\to}\theta_{b,H}(z)=\frac{b(z)}{1-\gamma_2zm_H(zb(z))-b(z)\gamma_2z^2\int\frac{dH(t)}{(t-zb(z))^2}}=\frac{b^2(z)}{q(z)},
\end{aligned}
$$

where the last equality follows from (A.9) and

$$\gamma_2zb(z)m_H\big(zb(z)\big)=b(z)-1. \tag{A.10}$$

Next, we derive the asymptotic distribution of $(p-p_1)(m_{F_n^{\mathbf{W}}}(zb_n(z))-m_{H_n}(zb_n(z)))$. It holds that

$$
\begin{aligned}
&(p-p_1)\big(m_{F_n^{\mathbf{W}}}\big(zb_n(z)\big)-m_{H_n}\big(zb_n(z)\big)\big)\\
&\quad=(p-p_1)\big(m_{F_n^{\mathbf{W}}}\big(zb_n(z)\big)-m_{H_n^{\mathbf{S}_{11}}}\big(zb_n(z)\big)\big) &\text{(A.11)}\\
&\qquad+(p-p_1)\big(m_{H_n^{\mathbf{S}_{11}}}\big(zb_n(z)\big)-m_{H_n}\big(zb_n(z)\big)\big) &\text{(A.12)}
\end{aligned}
$$

where $m_{H_n^{\mathbf{S}_{11}}}(z)$ and $m_{H_n}(z)$ are the unique solutions of the equations

$$
\begin{aligned}
&\frac{m_{H_n^{\mathbf{S}_{11}}}(z)}{(1+\gamma_{1,n}m_{H_n^{\mathbf{S}_{11}}}(z))}\\
&\quad=\int\frac{dF_n^{\widetilde{\mathbf{R}}}(t)}{t-(1+\gamma_{1,n}m_{H_n^{\mathbf{S}_{11}}}(z))[(1+\gamma_{1,n}m_{H_n^{\mathbf{S}_{11}}}(z))z-(1-\gamma_{1,n})]}, &\text{(A.13)}\\
&\frac{m_{H_n}(z)}{(1+\gamma_{1,n}m_{H_n}(z))}\\
&\quad=\int\frac{d\tilde{H}_n(t)}{t-(1+\gamma_{1,n}m_{H_n}(z))[(1+\gamma_{1,n}m_{H_n}(z))z-(1-\gamma_{1,n})]}, &\text{(A.14)}
\end{aligned}
$$

where $\widetilde{\mathbf{R}}=1/p_1\boldsymbol{\Sigma}_{22\cdot1}^{-1/2}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{S}_{11}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22\cdot1}^{-1/2}$, $\tilde{H}_n(t)$ stands for its discretized limiting spectral distribution, and $F_n^{\widetilde{\mathbf{R}}}(t)$ is the empirical spectral distribution of $\widetilde{\mathbf{R}}$.

First, we consider the second summand in (A.12). Let

$$\tilde{b}_n^*(z) = 1 + \gamma_{1,n} m_{H_n^{\mathbf{S}_{11}}}(z) \quad \text{and} \quad \tilde{b}_n(z) = 1 + \gamma_{1,n} m_{H_n}(z).$$

Similarly, using the definition of Stieltjes transform, (A.13) and (A.14) one can write

$$\begin{aligned}
&(p - p_1)\big(m_{H_n^{\mathbf{S}_{11}}}(z) - m_{H_n}(z)\big) \\
&\quad = (p - p_1)\tilde{b}_n^*(z) m_{F_n^{\tilde{\mathbf{R}}}}\big(\tilde{b}_n^*(z)\big(\tilde{b}_n^*(z)z - (1 - \gamma_{1,n})\big)\big) \\
&\qquad - (p - p_1)\tilde{b}_n(z) m_{\tilde{H}_n}\big(\tilde{b}_n(z)\big(\tilde{b}_n(z)z - (1 - \gamma_{1,n})\big)\big) \\
&\quad = (p - p_1)\big(\tilde{b}_n^*(z) - \tilde{b}_n(z)\big) m_{F_n^{\tilde{\mathbf{R}}}}\big(\tilde{b}_n^*(z)\big(\tilde{b}_n^*(z)z - (1 - \gamma_{1,n})\big)\big) \\
&\qquad + (p - p_1)\tilde{b}_n(z)\big[m_{F_n^{\tilde{\mathbf{R}}}}\big(\tilde{b}_n^*(z)\big(\tilde{b}_n^*(z)z - (1 - \gamma_{1,n})\big)\big) \\
&\qquad - m_{F_n^{\tilde{\mathbf{R}}}}\big(\tilde{b}_n(z)\big(\tilde{b}_n(z)z - (1 - \gamma_{1,n})\big)\big)\big] \\
&\qquad + (p - p_1)\tilde{b}_n(z)\big[m_{F_n^{\tilde{\mathbf{R}}}}\big(\tilde{b}_n(z)\big(\tilde{b}_n(z)z - (1 - \gamma_{1,n})\big)\big) \\
&\qquad - m_{\tilde{H}_n}\big(\tilde{b}_n(z)\big(\tilde{b}_n(z)z - (1 - \gamma_{1,n})\big)\big)\big] \\
&\qquad = (p - p_1)\gamma_{1,n}\big(m_{H_n^{\mathbf{S}_{11}}}(z) - m_{H_n}(z)\big) m_{F_n^{\tilde{\mathbf{R}}}}\big(\tilde{b}_n^*(z)\big(\tilde{b}_n^*(z)z - (1 - \gamma_{1,n})\big)\big) \\
&\qquad + (p - p_1)\tilde{b}_n(z)\gamma_{1,n}\big(m_{H_n^{\mathbf{S}_{11}}}(z) - m_{H_n}(z)\big)\big(z\big(\tilde{b}_n^* + \tilde{b}_n\big) - (1 - \gamma_{1,n})\big) \\
&\qquad \times \int \frac{dF_n^{\tilde{\mathbf{R}}}(t)}{[t - (\tilde{b}_n^*(z)(\tilde{b}_n^*(z)z - (1 - \gamma_{1,n})))][t - (\tilde{b}_n(z)(\tilde{b}_n(z)z - (1 - \gamma_{1,n})))]} \\
&\qquad + (p - p_1)\tilde{b}_n(z)\big[m_{F_n^{\tilde{\mathbf{R}}}}\big(\tilde{b}_n(z)\big(\tilde{b}_n(z)z - (1 - \gamma_{1,n})\big)\big) \\
&\qquad - m_{\tilde{H}_n}\big(\tilde{b}_n(z)\big(\tilde{b}_n(z)z - (1 - \gamma_{1,n})\big)\big)\big].
\end{aligned}$$

Rearranging terms, we get

$$\begin{aligned}
&(p - p_1)\big(m_{H_n^{\mathbf{S}_{11}}}(z) - m_{H_n}(z)\big) \\
&\quad = (p - p_1)\big[m_{F_n^{\tilde{\mathbf{R}}}}\big(\tilde{b}_n(z)\big(\tilde{b}_n(z)z - (1 - \gamma_{1,n})\big)\big) \\
&\qquad - m_{\tilde{H}_n}\big(\tilde{b}_n(z)\big(\tilde{b}_n(z)z - (1 - \gamma_{1,n})\big)\big)\big] \\
&\qquad \times \tilde{b}_n(z)\Big(1 - \gamma_{1,n} m_{F_n^{\tilde{\mathbf{R}}}}\big(\tilde{b}_n^*(z)\big(\tilde{b}_n^*(z)z - (1 - \gamma_{1,n})\big)\big) \\
&\qquad - \tilde{b}_n(z)\gamma_{1,n}\big(z\big(\tilde{b}_n^* + \tilde{b}_n\big) - (1 - \gamma_{1,n})\big)
\end{aligned}$$

$$\times \int dF_n^{\tilde{\mathbf{R}}}(t)/\big(\big[t-\big(\tilde{b}_n^*(z)\big(\tilde{b}_n^*(z)z-(1-\gamma_{1,n})\big)\big)\big]$$
$$\times \big[t-\big(\tilde{b}_n(z)\big(\tilde{b}_n(z)z-(1-\gamma_{1,n})\big)\big)\big]\big)\Big)^{-1},$$

where

$$\begin{aligned}
\tilde{b}_n(z)\Big(&1-\gamma_{1,n}m_{F_n^{\tilde{\mathbf{R}}}}\big(\tilde{b}_n^*(z)\big(\tilde{b}_n^*(z)z-(1-\gamma_{1,n})\big)\big)\\
&-\tilde{b}_n(z)\gamma_{1,n}\big(z\big(\tilde{b}_n^*+\tilde{b}_n\big)-(1-\gamma_{1,n})\big)\\
&\times \int dF_n^{\tilde{\mathbf{R}}}(t)/\big(\big[t-\big(\tilde{b}_n^*(z)\big(\tilde{b}_n^*(z)z-(1-\gamma_{1,n})\big)\big)\big]\\
&\times \big[t-\big(\tilde{b}_n(z)\big(\tilde{b}_n(z)z-(1-\gamma_{1,n})\big)\big)\big]\big)\Big)^{-1}\\
&\xrightarrow{a.s.} \theta_{\tilde{b},\tilde{H}}(z)\\
&=\tilde{b}(z)\Big/\Big(1-\gamma_1 m_{\tilde{H}}\big(\tilde{b}(z)\big[\tilde{b}(z)z-(1-\gamma_1)\big]\big)\\
&\qquad -\tilde{b}(z)\gamma_1\big(2z\tilde{b}(z)-(1-\gamma_1)\big)\int \frac{d\tilde{H}(t)}{[t-(\tilde{b}(z)(\tilde{b}(z)z-(1-\gamma_1)))]^2}\Big),
\end{aligned}$$

where $\tilde{b}(z)$ is given in (4.10).

The application of Lemma 1.1 in Bai and Silverstein (2004) proves that $(p-p_1)(m_{H_n^{\mathbf{S}_{11}}}(zb_n(z))-m_{H_n}(zb_n(z)))$ converges to a Gaussian process $M_3(z)$ with the mean function

$$\mathrm{E}\big(M_3(z)\big)=\theta_{\tilde{b},\tilde{H}}\big(zb(z)\big)\frac{c_1^2\int \underline{m}_{\tilde{H}}^3(zb(z))t^2(c_1+t\underline{m}_{\tilde{H}}(zb(z)))^{-3}\,dG(t)}{(1-c_1\int \underline{m}_{\tilde{H}}^2(zb(z))t^2(c_1+t\underline{m}_{\tilde{H}}(zb(z)))^{-2}\,dG(t))^2}$$

and the covariance function

$$\begin{aligned}
\operatorname{Cov}\big(M_3(z_1),M_3(z_2)\big)=2\theta_{\tilde{b},\tilde{H}}&\big(z_1b(z_1)\big)\theta_{\tilde{b},\tilde{H}}\big(z_2b(z_2)\big)\\
&\times\bigg(\frac{\partial}{\partial(z_1b(z_1))}\frac{\underline{m}_{\tilde{H}}(z_1b(z_1))\frac{\partial}{\partial(z_2b(z_2))}\underline{m}_{\tilde{H}}(z_2b(z_2))}{(\underline{m}_{\tilde{H}}(z_1b(z_1))-\underline{m}_{\tilde{H}}(z_2b(z_2)))^2}\\
&\quad-\frac{1}{(z_1b(z_1)-z_2b(z_2))^2}\bigg),
\end{aligned}$$

where $\underline{m}_{\tilde{H}}(z)=-\frac{1-c_1}{z}+c_1m_{\tilde{H}}(z)$ and $G(t)$ is the limiting spectral distribution of the matrix $\mathbf{R}=\boldsymbol{\Sigma}_{22\cdot1}^{-1/2}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22\cdot1}^{-1/2}$.

In order to derive the asymptotic distribution of the first summand in (A.11), we use the results in Yao (2013) to the conditional distribution of $(p-p_1)\times(m_{F_n^{\mathbf{W}}}(zb_n(z))-m_{F_n^{\mathbf{S}_{11}}}(zb_n(z)))$ given $\mathbf{S}_{11}$.

From the proof of Theorem 2, we know that the empirical spectral distribution of $\mathbf{W}$ is the same as of $\widetilde{\mathbf{W}}$ given by

$$\widetilde{\mathbf{W}} = \left(\frac{1}{\sqrt{p_1}}\mathbf{X} + \frac{1}{\sqrt{p_1}}\boldsymbol{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{M}\right)\left(\frac{1}{\sqrt{p_1}}\mathbf{X} + \frac{1}{\sqrt{p_1}}\boldsymbol{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{M}\right)^{\top},$$

with $\mathbf{M} = \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{S}_{11}^{1/2}$. Furthermore, following Lemma 1 it is enough to consider the case where $\boldsymbol{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{M}\mathbf{M}^T\boldsymbol{\Sigma}_{22\cdot 1}^{-1/2}$ is diagonal and, consequently, $\boldsymbol{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{M}$ is pseudo-diagonal.

Finally, in using that $\mathbf{X}$ consists of i.i.d. entries which are normally distributed and applying the results of Section 2.2.2 in Yao (2013), we get that $(p - p_1)(m_{F_n^{\mathbf{W}}}(zb_n(z)) - m_{F_n^{\mathbf{S}_{11}}}(zb_n(z)))$ converges to a Gaussian process $M_2(z)$ with the mean function $\mathrm{E}(M_2(z))$ and $\mathrm{Cov}(M_2(z_1), M_2(z_2))$ given in the following lemma which is proved below the proof of the theorem.

LEMMA 2. *The random process* $(p - p_1)(m_{F_n^{\mathbf{W}}}(zb_n(z)) - m_{F_n^{\mathbf{S}_{11}}}(zb_n(z)))$ *converges to a Gaussian process* $M_2(z)$ *with the mean function* $\mathrm{E}(M_2(z))$ *and* $\mathrm{Cov}(M_2(z_1), M_2(z_2))$ *given by*

$$\mathrm{E}\big(M_2(z)\big) = B\big(zb(z)\big)$$

and the covariance function

$$\mathrm{Cov}\big(M_2(z_1), M_2(z_2)\big) = 2\frac{\partial^2 \log(z_1 b(z_1)\eta(z_1 b(z_1)) - z_2 b(z_2)\eta(z_2 b(z_2)))}{\partial(z_1 b(z_1))\,\partial(z_2 b(z_2))}$$

which are independent of $\mathbf{S}_{11}$. *The functions* $B(z)$, $\delta(z)$, $\Psi(z)$, $\xi(z)$ *and* $\eta(z)$ *are given by* (4.11), (4.3), (4.6), (4.5) *and* (4.4), *respectively.*

The proof of Lemma 2 can be found in Appendix B of the Supplementary Material (see Bodnar, Dette and Parolya (2019)). Thus, merging the results for the independent asymptotic processes $M_2(z)$ and $M_3(z)$, we get

$$(p - p_1)\big(s_n^*(z) - s_n(z)\big) \to \theta_{b,H}(z)\big(M_2(z) + M_3(z)\big),$$

that is, converges to a Gaussian process with mean and covariance functions given by

$$\begin{aligned}&\theta_{b,H}(z)\Bigg(B\big(zb(z)\big) + \theta_{\tilde{b},\tilde{H}}\big(zb(z)\big)\\ &\quad\times\frac{c_1^2\int \underline{m}_{\tilde{H}}^3(zb(z))t^2(c_1 + t\underline{m}_{\tilde{H}}(zb(z)))^{-3}\,dG(t)}{(1 - c_1\int \underline{m}_{\tilde{H}}^2(zb(z))t^2(c_1 + t\underline{m}_{\tilde{H}}(zb(z)))^{-2}\,dG(t))^2}\Bigg)\end{aligned} \tag{A.15}$$

and

$$
\begin{aligned}
&2\theta_{b,H}(z_1)\theta_{b,H}(z_2)\Big[\frac{\partial^2 \log(z_1 b(z_1)\eta(z_1 b(z_1)) - z_2 b(z_2)\eta(z_2 b(z_2)))}{\partial(z_1 b(z_1))\,\partial(z_2 b(z_2))} \\
&\quad + \theta_{\tilde{b},\tilde{H}}\big(z_1 b(z_1)\big)\theta_{\tilde{b},\tilde{H}} \\
&\quad \times \big(z_2 b(z_2)\big)\Big(\frac{\frac{\partial}{\partial(z_1 b(z_1))}\underline{m}_{\tilde{H}}(z_1 b(z_1))\frac{\partial}{\partial(z_1 b(z_1))}\underline{m}_{\tilde{H}}(z_2 b(z_2))}{(\underline{m}_{\tilde{H}}(z_1 b(z_1)) - \underline{m}_{\tilde{H}}(z_2 b(z_2)))^2} \\
&\quad - \frac{1}{(z_1 b(z_1) - z_2 b(z_2))^2}\Big)\Big], \qquad \text{(A.16)}
\end{aligned}
$$

respectively. Remind that H is the asymptotic spectral distribution of the matrix $\mathbf{W}$ (and, thus, of $\widetilde{\mathbf{W}}$). Furthermore, it holds

$$
\begin{aligned}
&\theta_{b,H}(z_1)\theta_{b,H}(z_2)\frac{\partial^2 \log(z_1 b(z_1)\eta(z_1 b(z_1)) - z_2 b(z_2)\eta(z_2 b(z_2)))}{\partial(z_1 b(z_1))\,\partial(z_2 b(z_2))} \\
&\quad = \frac{b^2(z_1)}{q(z_1)(z_1 b(z_1))'}\frac{b^2(z_2)}{q(z_2)(z_2 b(z_2))'} \\
&\quad\quad \times \frac{\partial^2 \log(z_1 b(z_1)\eta(z_1 b(z_1)) - z_2 b(z_2)\eta(z_2 b(z_2)))}{\partial z_1\,\partial z_2} \\
&\quad = \frac{\partial^2 \log(z_1 b(z_1)\eta(z_1 b(z_1)) - z_2 b(z_2)\eta(z_2 b(z_2)))}{\partial z_1\,\partial z_2}, \qquad \text{(A.17)}
\end{aligned}
$$

where the last equality in (A.17) follows from (A.10) and

$$
\begin{aligned}
q(z)\big(zb(z)\big)' &= \Big(1 - \gamma_2\big(b(z)z\big)^2\frac{m'_H(zb(z))}{(zb(z))'}\Big)\big(zb(z)\big)' \\
&= \big(zb(z)\big)' - \big(zb(z)\big)^2\Big(-\frac{1}{z^2} + \frac{(zb(z))'}{(zb(z))^2}\Big) = b^2(z).
\end{aligned}
$$

Similarly, we get

$$
\begin{aligned}
&\theta_{b,H}(z_1)\theta_{b,H}(z_2)\Big(\frac{\frac{\partial}{\partial z_1 b(z_1)}\underline{m}_{\tilde{H}}(z_1 b(z_1))\frac{\partial}{\partial z_2 b(z_2)}\underline{m}_{\tilde{H}}(z_2 b(z_2))}{(\underline{m}_{\tilde{H}}(z_1 b(z_1)) - \underline{m}_{\tilde{H}}(z_2 b(z_2)))^2} \\
&\quad - \frac{1}{(z_1 b(z_1) - z_2 b(z_2))^2}\Big) \\
&\quad = \frac{b^2(z_1)}{q(z_1)(z_1 b(z_1))'}\frac{b^2(z_2)}{q(z_2)(z_2 b(z_2))'}\frac{\partial^2 \log(\frac{\underline{m}_{\tilde{H}}(z_1 b(z_1)) - \underline{m}_{\tilde{H}}(z_2 b(z_2))}{z_1 b(z_1) - z_2 b(z_2)})}{\partial z_1\,\partial z_2} \\
&\quad = \frac{\partial^2 \log(\frac{\underline{m}_{\tilde{H}}(z_1 b(z_1)) - \underline{m}_{\tilde{H}}(z_2 b(z_2))}{z_1 b(z_1) - z_2 b(z_2)})}{\partial z_1\,\partial z_2}. \qquad \text{(A.18)}
\end{aligned}
$$

At last, combining the results (A.7), (A.8), (A.15), (A.16) together with (A.17) and (A.18) we get that the process $(p-p_1)(m_{F^{\mathbf{WT}^{-1}}}(z)-s_n(z))$ is asymptotically Gaussian with mean and covariance functions given by

$$\frac{1}{2}\,\mathrm{d}\log\bigl(q(z)\bigr)+\theta_{b,H}(z)\biggl(B\bigl(zb(z)\bigr)$$
$$+\theta_{\tilde{b},\tilde{H}}\bigl(zb(z)\bigr)\frac{c_1^2\int \underline{m}_{\tilde{H}}^3(zb(z))t^2(c_1+t\underline{m}_{\tilde{H}}(zb(z)))^{-3}\,dG(t)}{(1-c_1\int \underline{m}_{\tilde{H}}^2(zb(z))t^2(c_1+t\underline{m}_{\tilde{H}}(zb(z)))^{-2}\,dG(t))^2}\biggr)$$

and

$$2\biggl[\frac{\partial^2\log(z_1b(z_1)\eta(z_1b(z_1))-z_2b(z_2)\eta(z_2b(z_2)))}{\partial z_1\,\partial z_2}$$
$$+\theta_{\tilde{b},\tilde{H}}\bigl(z_1b(z_1)\bigr)\theta_{\tilde{b},\tilde{H}}\bigl(z_2b(z_2)\bigr)\biggl(\frac{\partial^2\log(\frac{\underline{m}_{\tilde{H}}(z_1b(z_1))-\underline{m}_{\tilde{H}}(z_2b(z_2))}{z_1b(z_1)-z_2b(z_2)})}{\partial z_1\,\partial z_2}\biggr)\biggr].$$

Since the process of interest $(p-p_1)(m_{F^{\mathbf{WT}^{-1}}}(z)-s_n(z))=M_{1,n}+M_{2,n}+M_{3,n}$ forms a tight sequence (see Bai and Silverstein (2004), Yao (2013) and Zheng, Bai and Yao (2015a)), the Cauchy integral formula leads to

$$\sum_{i=1}^{p-p_1} f(\lambda_i)-(p-p_1)\int f(x)F_n(dx)$$
$$=-\frac{1}{2\pi i}\oint f(z)(p-p_1)\bigl(m_{F^{\mathbf{WT}^{-1}}}(z)-s_n(z)\bigr)\,dz, \tag{A.19}$$

where λ_i is the ith eigenvalue of the matrix $\mathbf{WT}^{-1}$ and f is an arbitrary analytic function with support containing the interval $[0,r]$, which itself contains the asymptotic spectrum of the matrix $\mathbf{WT}^{-1}$. The application of (A.19) to our process together with some elementary calculus lead to the result of the theorem. □

Acknowledgements. The authors would like to thank Professor Tailen Hsing and two anonymous referees for the valuable comments and remarks which have significantly improved the previous version of this paper. Additionally, we want to thank M. Stein who typed parts of this manuscript with considerable technical expertise.

SUPPLEMENTARY MATERIAL

Supplement to "Testing for independence of large dimensional vectors" (DOI: 10.1214/18-AOS1771SUPP; .pdf). The supplementary material contains the proofs of Theorem 1, Lemma 1–2 and additional simulations provided in Figures 10–14.

REFERENCES

ANDERSON, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley Interscience, Hoboken, NJ. MR1990662

BAI, Z. D. and SILVERSTEIN, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32** 553–605. MR2040792

BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2567175

BAI, Z., JIANG, D., YAO, J.-F. and ZHENG, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** 3822–3840. MR2572444

BICKEL, P. J. and LEVINA, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. MR2387969

BIRKE, M. and DETTE, H. (2005). A note on testing the covariance matrix for large dimension. *Statist. Probab. Lett.* **74** 281–289. MR2189467

BODNAR, T., DETTE, H. and PAROLYA, N. (2019). Supplement to "Testing for independence of large dimensional vectors." DOI:10.1214/18-AOS1771SUPP.

BODNAR, T., GUPTA, A. K. and PAROLYA, N. (2014). On the strong convergence of the optimal linear shrinkage estimator for large dimensional covariance matrix. *J. Multivariate Anal.* **132** 215–228. MR3266272

BODNAR, T., GUPTA, A. K. and PAROLYA, N. (2016). Direct shrinkage estimation of large dimensional precision matrix. *J. Multivariate Anal.* **146** 223–236. MR3477661

CAI, T., LIU, W. and LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Amer. Statist. Assoc.* **106** 594–607. MR2847973

CAI, T. T., REN, Z. and ZHOU, H. H. (2013). Optimal rates of convergence for estimating Toeplitz covariance matrices. *Probab. Theory Related Fields* **156** 101–143. MR3055254

CAI, T. T. and SHEN, X., eds. (2011). *High-Dimensional Data Analysis*. *Frontiers of Statistics* **2**. World Scientific Co. Pte. Ltd., Singapore; Higher Education Press, Beijing. MR2866981

CAI, T. T. and ZHOU, H. H. (2012). Minimax estimation of large covariance matrices under ℓ_1-norm. *Statist. Sinica* **22** 1319–1349. MR3027084

CHEN, S. X., ZHANG, L.-X. and ZHONG, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. MR2724863

DEVIJVER, E. and GALLOPIN, M. (2018). Block-diagonal covariance selection for high-dimensional Gaussian graphical models. *J. Amer. Statist. Assoc.* **113** 306–314. MR3803466

DOZIER, R. B. and SILVERSTEIN, J. W. (2007). On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices. *J. Multivariate Anal.* **98** 678–694. MR2322123

FAN, J. and LI, R. (2006). Statistical challenges with high dimensionality: Feature selection in knowledge discovery. In *International Congress of Mathematicians. Vol. III* 595–622. Eur. Math. Soc., Zürich. MR2275698

FISHER, R. A. (1939). The sampling distribution of some statistics obtained from non-linear equations. *Ann. Eugenics* **9** 238–249. MR0001499

FISHER, T. J., SUN, X. and GALLAGHER, C. M. (2010). A new test for sphericity of the covariance matrix for high dimensional data. *J. Multivariate Anal.* **101** 2554–2570. MR2719881

FUJIKOSHI, Y., HIMENO, T. and WAKAKI, H. (2004). Asymptotic results of a high dimensional Manova test and power comparison when the dimension is large compared to the sample size. *J. Japan Statist. Soc.* **34** 19–26. MR2084057

GUPTA, A. K. and XU, J. (2006). On some tests of the covariance matrix under general conditions. *Ann. Inst. Statist. Math.* **58** 101–114. MR2281208

HYODO, M., SHUTOH, N., NISHIYAMA, T. and PAVLENKO, T. (2015). Testing block-diagonal covariance structure for high-dimensional data. *Stat. Neerl.* **69** 460–482. MR3414707

JIANG, D., BAI, Z. and ZHENG, S. (2013). Testing the independence of sets of large-dimensional variables. *Sci. China Math.* **56** 135–147. MR3016588

JIANG, T. and YANG, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist.* **41** 2029–2074. MR3127857

JOHN, S. (1971). Some optimal multivariate tests. *Biometrika* **58** 123–127. MR0275568

JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

JOHNSTONE, I. (2006). High dimensional statistical inference and random matrices. Instituto de Ciencias Matemáticas (ICMAT). Available at: http://www.icm2006.org/proceedings/Vol_I/17.pdf.

JOHNSTONE, I. M. (2008). Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy–Widom limits and rates of convergence. *Ann. Statist.* **36** 2638–2716. MR2485010

KAKIZAWA, Y. and IWASHITA, T. (2008). A comparison of higher-order local powers of a class of one-way MANOVA tests under general distributions. *J. Multivariate Anal.* **99** 1128–1153. MR2432484

LEDOIT, O. and WOLF, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** 1081–1102. MR1926169

LEDOIT, O. and WOLF, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Ann. Statist.* **40** 1024–1060. MR2985942

MARKOWITZ, H. (1952). Portfolio selection. *J. Finance* **7** 77–91.

MAUCHLY, J. W. (1940). Significance test for sphericity of a normal n-variate distribution. *Ann. Math. Stat.* **11** 204–209. MR0002084

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory. Wiley Series in Probability and Mathematical Statistics*. Wiley, New York. MR0652932

NAGAO, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.* **1** 700–709. MR0339405

PILLAI, K. C. S. and JAYACHANDRAN, K. (1967). Power comparisons of tests of two multivariate hypotheses based on four criteria. *Biometrika* **54** 195–210. MR0216641

SCHOTT, J. R. (2007). Some high-dimensional tests for a one-way MANOVA. *J. Multivariate Anal.* **98** 1825–1839. MR2392435

WANG, C., PAN, G., TONG, T. and ZHU, L. (2015). Shrinkage estimation of large dimensional precision matrix using random matrix theory. *Statist. Sinica* **25** 993–1008. MR3409734

YAMADA, Y., HYODO, M. and NISHIYAMA, T. (2017). Testing block-diagonal covariance structure for high-dimensional data under non-normality. *J. Multivariate Anal.* **155** 305–316. MR3607897

YANG, Y. and PAN, G. (2015). Independence test for high dimensional data based on regularized canonical correlation coefficients. *Ann. Statist.* **43** 467–500. MR3316187

YAO, J. (2013). Estimation and fluctuations of functionals of large random matrices. Telecom ParisTech, tel-00909521v1.

YAO, J., ZHENG, S. and BAI, Z. (2015). *Large Sample Covariance Matrices and High-Dimensional Data Analysis. Cambridge Series in Statistical and Probabilistic Mathematics* **39**. Cambridge Univ. Press, New York. MR3468554

ZHENG, S. (2012). Central limit theorems for linear spectral statistics of large dimensional F-matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **48** 444–476. MR2954263

ZHENG, S., BAI, Z. and YAO, J. (2015a). CLT for linear spectral statistics of a rescaled sample precision matrix. *Random Matrices Theory Appl.* **4** 1550014, 43. MR3418843

ZHENG, S., BAI, Z. and YAO, J. (2015b). Substitution principle for CLT of linear spectral statistics of high-dimensional sample covariance matrices with applications to hypothesis testing. *Ann. Statist.* **43** 546–591. MR3316190

ZHENG, S., BAI, Z. and YAO, J. (2017). CLT for eigenvalue statistics of large-dimensional general Fisher matrices with applications. *Bernoulli* **23** 1130–1178. MR3606762

T. BODNAR
DEPARTMENT OF MATHEMATICS
STOCKHOLM UNIVERSITY
SE-10691 STOCKHOM
SWEDEN
E-MAIL: taras.bodnar@math.su.se

H. DETTE
DEPARTMENT OF MATHEMATICS
RUHR UNIVERSITY BOCHUM
D-44870 BOCHUM
GERMANY
E-MAIL: holger.dette@ruhr-uni-bochum.de

N. PAROLYA
INSTITUTE OF STATISTICS
LEIBNIZ UNIVERSITY HANNOVER
D-30167 HANNOVER
GERMANY
E-MAIL: parolya@statistik.uni-hannover.de