Mathematical Statistics
Stockholm University

Testing for Independence of Large Dimensional Vectors

Taras Bodnar
Holger Dette
Nestor Parolya

Research Report 2017:15

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se

Mathematical Statistics
Stockholm University
Research Report **2017:15**,
http://www.math.su.se

Testing for Independence of Large Dimensional Vectors

August 2017

Taras Bodnar[a], Holger Dette[b] and Nestor Parolya[c]
[a] *Department of Mathematics, Stockholm University, SE-10691 Stockhom, Sweden*
e-mail: taras.bodnar@math.su.se
[b] *Department of Mathematics, Ruhr University Bochum, D-44870 Bochum, Germany*
e-mail: holger.dette@ruhr-uni-bochum.de
[c] *Institute of Statistics, Leibniz University Hannover, D-30167 Hannover, Germany*
e-mail: parolya@statistik.uni-hannover.de

Abstract

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) non-central Fisher matrices. In particular, a central limit theorem for linear spectral statistics of large dimensional (conditionally) non-central 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.

AMS 2010 subject classifications: 60B20, 60F05, 62H15, 62H20, 62F05
Keywords: testing for independence, large dimensional covariance matrix, non-central Fisher random matrix, linear spectral statistics, asymptotic normality

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).

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 et al. (2014, 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 (2016)), Toeplitz matrix (see, Cai et al. (2013)), band matrix (see, Bickel and Levina (2008)) or general sparsity assumptions (see Cai et al. (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 non-normal 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 et al. (2010), Chen et al. (2010)). Other authors use the distributional properties of the largest eigenvalue of the sample covariance matrix to construct tests (see Johnstone (2001, 2008) for example).

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 et al. (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) non-central 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 the appendix (see, Section 5).

2 Testing for independence

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

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

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

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

respectively, where $\mathbf{\Sigma}_{ij} \in \mathbb{R}^{p_i \times p_j}$ 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, i.e.

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

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 non-zero mean vector can be treated exactly in the same way observing that the centred sample covariance matrix, has a $\frac{1}{n-1}W_p(n-1, \mathbf{\Sigma})$ distribution.

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

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

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

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

The null hypothesis is rejected for small values of V_n. Jiang et al. (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

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

where $\mathbf{F} = \frac{1}{p_1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\Big(\frac{1}{n-p_1}(\mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12})\Big)^{-1}$. 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 et al. (2015))

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

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

where the quantities μ_{LR}, σ_{LR}^2 *and* s_{LR} *are defined by*

$$\begin{aligned}
\mu_{LR} &= 1/2 \log\left[\frac{({w_n^*}^2 - {d_n^*}^2)h_n^2}{(w_n^* h_n - \gamma_{2,n}{d_n^*}^2)^2}\right], \qquad \sigma_{LR}^2 = 2\log\left[\frac{{w_n^*}^2}{{w_n^*}^2 - {d_n^*}^2}\right], \\
s_{LR} &= \log\Big(\frac{\gamma_{1,n}}{\gamma_{2,n}}(1 - \gamma_{2,n})^2\Big) + \frac{1 - \gamma_{2,n}}{\gamma_{2,n}}\log(w_n^*) - \frac{\gamma_{1,n} + \gamma_{2,n}}{\gamma_{1,n}\gamma_{2,n}}\log(w_n^* - d_n^*\gamma_2/h_n) \\
&+ \begin{cases} \frac{1-\gamma_{1,n}}{\gamma_{1,n}}\log(w_n^* - d_n^* h_n), & \gamma_{1,n} \in (0, 1) \\ 0, & \gamma_{1,n} = 1 \\ -\frac{1-\gamma_{1,n}}{\gamma_{1,n}}\log(w_n^* - d_n^*/h_n), & \gamma_{1,n} > 1 \end{cases}
\end{aligned}$$

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

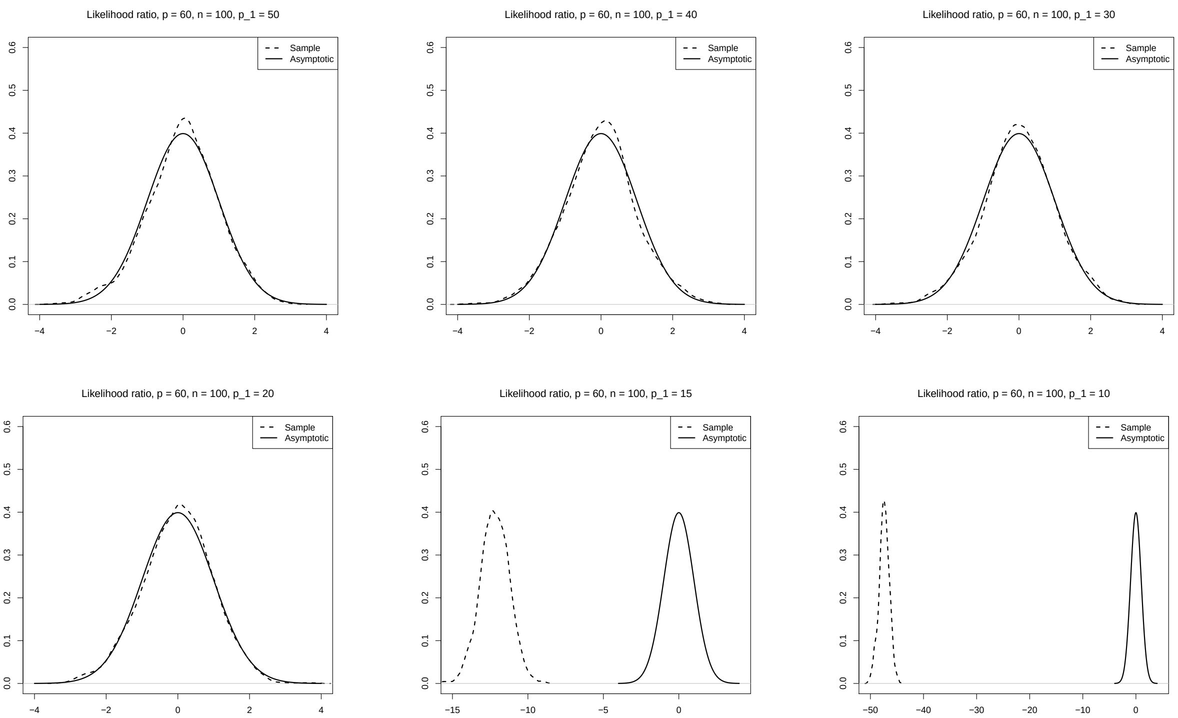

Figure 1: *Simulated distribution of the statistic* $(T_{LR} - (p - p_1)s_{LR} - \mu_{LR})/\sigma_{LR}$ *und the null hypothesis for sample size* $n = 100$*, dimension* $p = 60$ *and various values of* $p_1 = 50, 45, 40, 30, 15, 10$*. The solid curve shows the standard normal distribution.*

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

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

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 $(T_{LR} - (p - p_1)s_{LR} - \mu_{LR})/\sigma_{LR}$ 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 $\gamma_{1,n} \leq 1$ (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 non-singular 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

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

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

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 $\mathbf{\Sigma}_{22\cdot1} = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$ and $\mathbf{S}_{22\cdot1} = \mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$ the corresponding Schur complements. From Theorem 3.2.10 of Muirhead (1982), it follows that

$$\begin{aligned}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}|\mathbf{S}_{11} &\sim \mathcal{N}_{p-p_1,p_1}(\widetilde{\mathbf{\Sigma}}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{S}_{11}^{1/2}, \mathbf{\Sigma}_{22\cdot1} \otimes \mathbf{I}_{p_1}),\\ \mathbf{S}_{22\cdot1} &\sim W_{p-p_1}(n-p_1, \mathbf{\Sigma}_{22\cdot1}),\end{aligned}$$

and the Schur complement $\mathbf{S}_{22\cdot1}$ is independent of $\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}$ and $\mathbf{S}_{11}$. Hence, under the null hypothesis,

$$\begin{aligned}\widehat{\mathbf{W}} &= \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12} \sim W_{p-p_1}(p_1, \mathbf{\Sigma}_{22\cdot1}),\\ \hat{\mathbf{T}} &= \mathbf{S}_{22\cdot1} \sim W_{p-p_1}(n-p_1, \mathbf{\Sigma}_{22\cdot1}),\end{aligned}$$

and $\widehat{\mathbf{W}}$ and $\hat{\mathbf{T}}$ are independent. Under the alternative hypothesis H_1, $\widehat{\mathbf{W}}$ and $\hat{\mathbf{T}}$ are still independent as well as $\hat{\mathbf{T}} \sim W_{p-p_1}(n-p_1, \mathbf{\Sigma}_{22\cdot1})$, but $\widehat{\mathbf{W}}$ has a non-central Wishart distribution conditionally on $\mathbf{S}_{11}$, i.e.,

$$\widehat{\mathbf{W}}|\mathbf{S}_{11} \sim W_{p-p_1}(p_1, \mathbf{\Sigma}_{22\cdot1}, \mathbf{\Omega}_1(\mathbf{S}_{11}))$$

where the non-centrality parameter is given by

$$\mathbf{\Omega}_1 = \mathbf{\Omega}_1(\mathbf{S}_{11}) = \mathbf{\Sigma}_{22\cdot1}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{S}_{11}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}.$$

For technical reasons we will use the normalized versions of $\widehat{\mathbf{W}}$ and $\hat{\mathbf{T}}$ throughout this paper. Thus, the distributional properties of $\mathbf{W} = \frac{1}{p_1}\widehat{\mathbf{W}}$ and $\mathbf{T} = \frac{1}{n-p_1}\hat{\mathbf{T}}$ are very similar to the ones observed for the within and between covariance matrices in the multivariate analysis of variance (MANOVA) model (see Fujikoshi et al. (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 non-central 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:

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

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

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

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

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

instead of $tr(\mathbf{W}\mathbf{T}^{-1}(\mathbf{I}+\mathbf{W}\mathbf{T}^{-1})^{-1})$. 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[1] 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)). This representation is used intensively in proof of our first main result, which provides the asymptotic distribution of T_W, T_{LH}, and T_{BNP} under the null hypothesis in (2.8). The details of the proof are deferred to the Appendix in Section 5.

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

$$\begin{aligned}
\mu_W &= 1/2\log\left[\frac{(w_n^2-d_n^2)h_n^2}{(w_nh_n-\gamma_{2,n}d_n)^2}\right], & \sigma_W^2 = 2\log\left[\frac{w_n^2}{w_n^2-d_n^2}\right],\\
\mu_{LH} &= \frac{\gamma_{2,n}}{(1-\gamma_{2,n})^2}, & \sigma_{LH}^2 = \frac{2h_n^2}{(1-\gamma_{2,n})^4},\\
\mu_{BNP} &= -\frac{(1-\gamma_{2,n})^2 w_n^2(d_n^2-\gamma_{2,n})}{(w_n^2-d_n^2)^2}, & \sigma_{BNP}^2 = 2\frac{d^2(1-\gamma_{2,n})^4}{w_n^2(1+d_n)(w_n^2-d_n^2)^4}(w_n^2(w_n^2+d_n)+d_n^3(w_n^2-1))\,,
\end{aligned}$$

[1]The formal definition of the linear spectral statistic is given in Appendix by A.1.

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*

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

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

$$\frac{T_a - (p-p_1)s_a - \mu_a}{\sigma_a} > u_{1-\alpha} \tag{3.4}$$

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 Figure 2 - 4 we display the rejection probabilities of the test (3.4) 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 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 Figure 3 and 4 show a reasonable approximation of the nominal level in all considered scenarios.

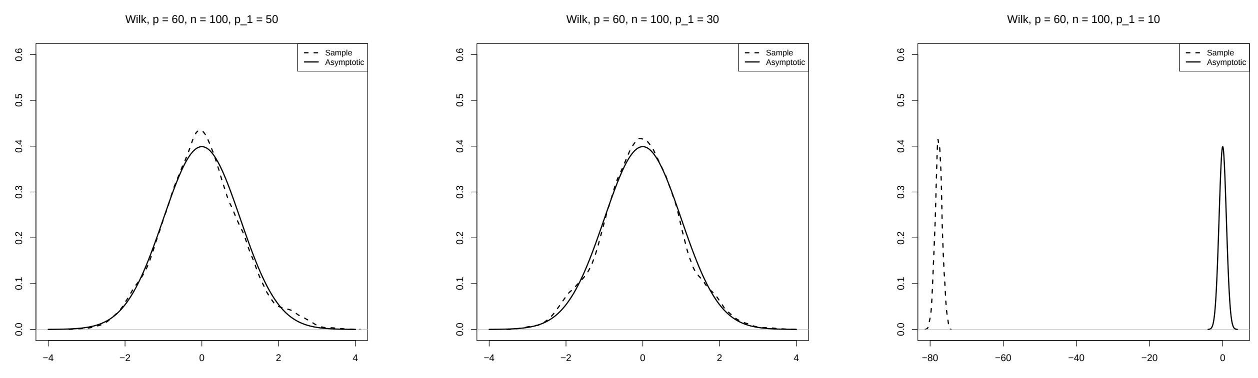

Figure 2: *Simulated distribution of the statistic* $(T_W - (p-p_1)s_W - \mu_W)/\sigma_W$ *und the null hypothesis for sample size* $n = 100$, *dimension* $p = 60$ *and various values of* $p_1 = 50, 30, 10$. *The solid curve shows the standard normal distribution.*

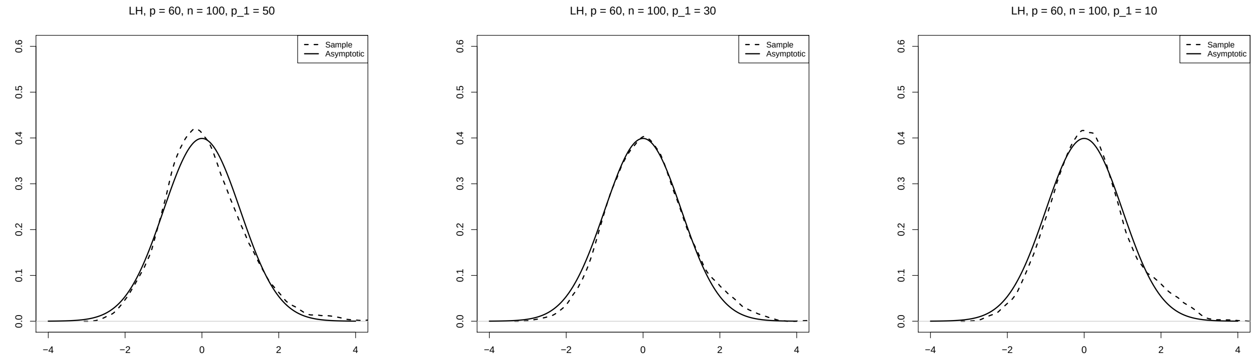

Figure 3: *Simulated distribution of the statistic* $(T_{LH} - (p - p_1)s_{LH} - \mu_{LH})/\sigma_{LH}$ *und the null hypothesis for sample size* $n = 100$*, dimension* $p = 60$ *and various values of* $p_1 = 50, 30, 10$*. The solid curve shows the standard normal distribution.*

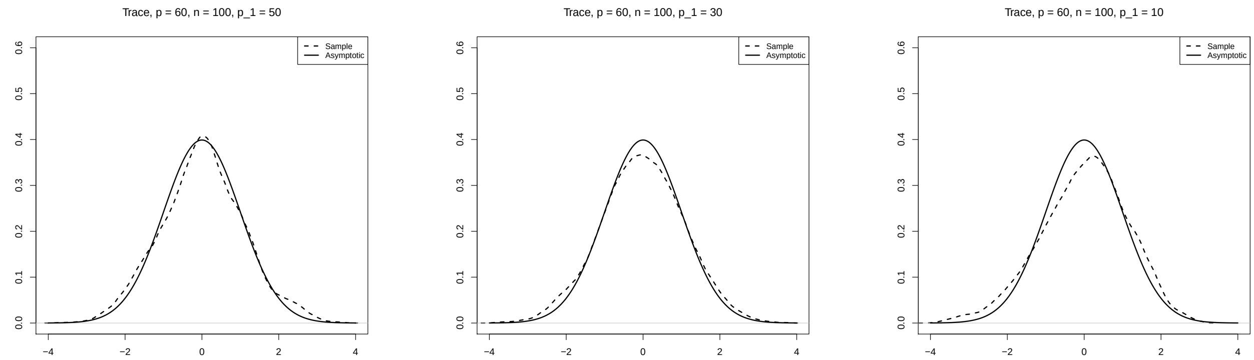

Figure 4: *Simulated distribution of the statistic* $(T_{BNP} - (p - p_1)s_{BNP} - \mu_{BNP})/\sigma_{BNP}$ *und the null hypothesis for sample size* $n = 100$*, dimension* $p = 60$ *and various values of* $p_1 = 50, 30, 10$*. The solid curve shows the standard normal distribution.*

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 $\mathbf{W}\mathbf{T}^{-1}$ has a (conditionally) non-central Fisher distribution in this case. The following two results, which are proved in the Appendix and of independent interest, specify the asymptotic distribution of the empirical spectral distribution of the matrix $\mathbf{W}\mathbf{T}^{-1}$ under H_1. Throughout the paper

$$m_G(z) = \int_{-\infty}^{+\infty} \frac{dG(t)}{t - z}$$

denotes the Stieltjes transform of a distribution function G.

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 $\mathbf{W}\mathbf{T}^{-1}$ *converges almost surely to some deterministic function* s*, which is the unique solution of the following system of equations*

$$\frac{s(z)}{1+\gamma_2 z s(z)} = m_H\left(z(1+\gamma_2 z s(z))\right),$$

$$\frac{m_H(z)}{1+\gamma_1 m_H(z)} = m_{\tilde{H}}((1+\gamma_1 m_H(z))[(1+\gamma_1 m_H(z))z-(1-\gamma_1)]), \tag{4.1}$$

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

subject to the condition that $\Im\{s(z)\}$ *is of the same sign as* $\Im\{z\}$*.*

We will use this result to derive a CLT for the linear spectral statistics of the matrix $\mathbf{W}\mathbf{T}^{-1}$, 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 notations as follows

$$\begin{aligned}
\delta(z) &= \gamma_1 m_H(z) && (4.3)\\
\tilde{\delta}(z) &= \delta(z) - \frac{1-\gamma_1}{z} &&\\
\eta(z) &= (1+\delta(z))(1+\tilde{\delta}(z)) && (4.4)\\
\xi(z) &= \frac{\delta'(z)}{(z\eta(z))'} && (4.5)\\
\Psi(z) &= \left(\frac{1}{1+\delta(z)} - 2\xi(z)z + \frac{1-\gamma_1}{1+\delta(z)}\xi(z)\right)^{-1}, && (4.6)\\
r &= 2\frac{(1+\sqrt{\gamma_1})^2 + \lambda_{max}(\mathbf{R})(1-\sqrt{c_1})^2}{(1-\sqrt{\gamma_2})^2} && (4.7)
\end{aligned}$$

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*

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

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

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

$$\begin{aligned}
Cov[X_f, X_g] &= -\frac{1}{2\pi^2}\oint\oint f(z_1)g(z_2)\frac{\partial^2\log(z_1b(z_1)-z_2b(z_2))}{\partial z_1\partial z_2}\,\mathrm{d}z_1\,\mathrm{d}z_2\\
&- \frac{1}{2\pi^2}\oint\oint f(z_1)g(z_2)\frac{\partial^2\log(z_1b(z_1)\eta(z_1b(z_1))-z_2b(z_2)\eta(z_2b(z_2)))}{\partial z_1\partial z_1}\,\mathrm{d}z_1\,\mathrm{d}z_2\\
&- \frac{1}{2\pi^2}\oint\oint f(z_1)g(z_2)\left[\theta_{\tilde{b},\tilde{H}}(z_1b(z_1))\theta_{\tilde{b},\tilde{H}}(z_2b(z_2))\left(\frac{\partial^2\log\left[\frac{\underline{m}_{\tilde{H}}(z_2b(z_2))-\underline{m}_{\tilde{H}}(z_1b(z_1))}{(z_2b(z_2)-z_1b(z_1))}\right]}{\partial z_1\partial z_2}\right)\right]\mathrm{d}z_1\,\mathrm{d}z_2
\end{aligned} \tag{4.9}$$

respectively, where

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

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 non-overlapping.*

It follows from the proof of Theorem 2 that

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

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 $2\left((1+\sqrt{\gamma_1})^2+\lambda_{max}(\mathbf{R})(1-\sqrt{c_1})^2\right)$ and the quantity r defined in (4.7) is an upper bound for the limiting spectrum of the matrix $\mathbf{W}\mathbf{T}^{-1}$.

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 Section 5 reveals a very interesting fact that the resulting asymptotic distributions depend neither on the eigenvectors of the non-centrality matrix $\mathbf{\Omega}_1$ nor on the eigenvectors of the matrix $\mathbf{R}=\mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot1}^{-1/2}$ 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. This result can not be deduced from previous literature (see, e.g., Yao (2013)).

Remark 2 (Generalizations and simplifications)**.** The non-central Fisher matrix in our case arises only conditionally on $\mathbf{S}_{11}$ where the non-centrality matrix $\mathbf{\Omega}_1$ is random in our framework. As a consequence Theorem 3 generalizes the result of Yao (2013), where a deterministic non-centrality 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

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

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 et al. (2017).

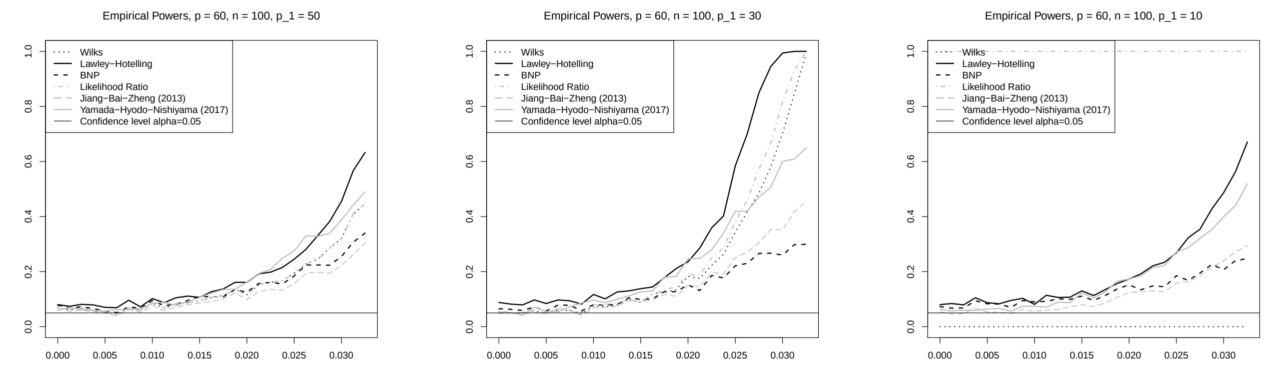

Figure 5: *Empirical power of different tests for block diagonality for sample size* $n = 100$, *dimension* $p = 60$ *and various values of* $p_1 = 50, 30, 10$ *as a function of the correlation coefficient* $\rho = \frac{\sigma_{12}}{\sigma}$ *in* $[0, 0.0325]$.

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. 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

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

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 5. Here, we also included the trace criterion recently proposed by Jiang and Yang (2013) and the test proposed by Yamada et al. (2017), which is based on an empirical distance between the full and a block diagonal covariance matrix. Figure 5 justifies our theoretical findings, i.e., 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 in absolute value, 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 et al. (2017), and Wilk's test with the corrected likelihood ratio (LR) criterion are ranked on the second and third, respectively, while the Bartlett-Nanda-Pillai's (BNP) trace criterion is on the fourth position and the trace criterion of Jiang and Yang (2013) shows the worst performance. The same ranking was observed for $p_1 = 50$ with the only difference of a decreasing power of all tests. 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 most interesting one. Indeed, here we observe that the Wilk's and the LR tests are not reliable anymore, while 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.

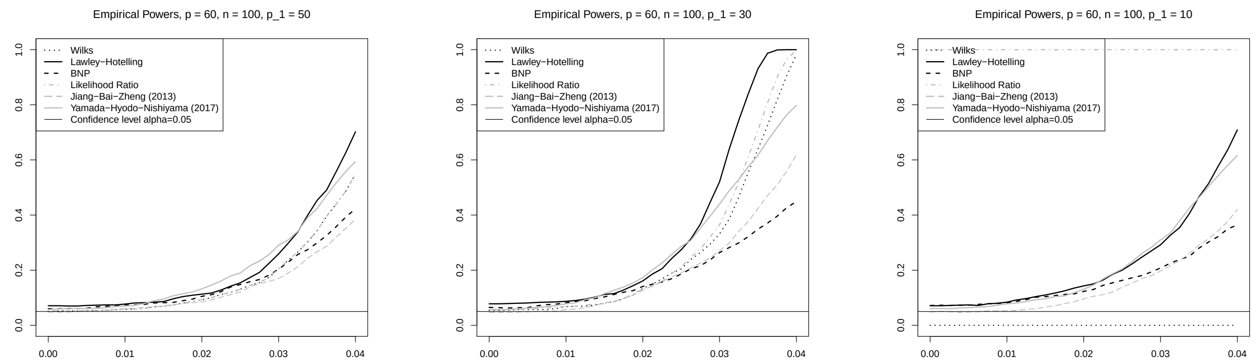

Figure 6: *Empirical power of different tests for block diagonality for sample size* $n = 100$, *dimension* $p = 60$ *and various values of* $p_1 = 50, 30, 10$ *as a function of the correlation coefficient* $\rho = \frac{\sigma_{12}}{\sigma}$ *in* $[0, 0.04]$ *and sparsity level of* 20%.

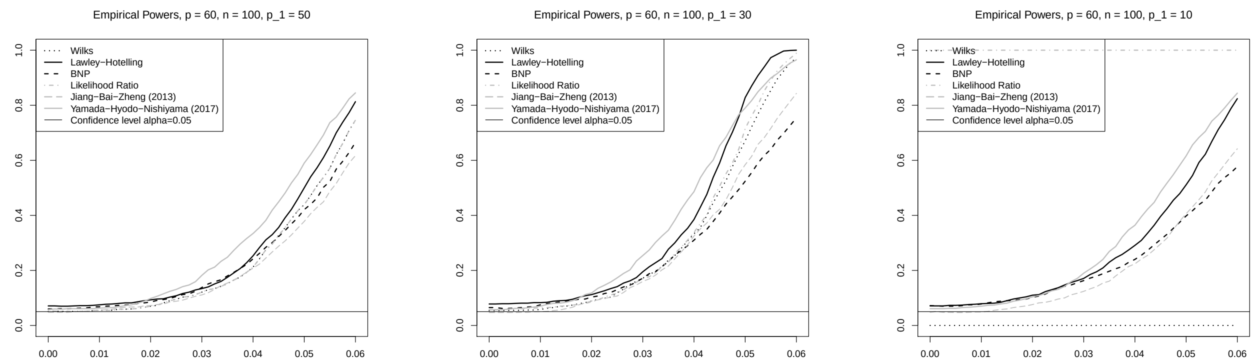

Figure 7: *Empirical power of different tests for block diagonality for sample size* $n = 100$, *dimension* $p = 60$ *and various values of* $p_1 = 50, 30, 10$ *as a function of the correlation coefficient* $\rho = \frac{\sigma_{12}}{\sigma}$ *in* $[0, 0.06]$ *and sparsity level of* 50%.

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 6 and 7.

We observe a similar behaviour as in the non-sparse case (see Figure 5). The LH test and the test proposed in Yamada et al. (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 non-sensitivity 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. 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 can not be recommended, if the size of the first block is much smaller than the size of the second one.

Acknowledgements. The authors would like to thank M. Stein who typed parts of this manuscript with considerable technical expertise. The work of H. Dette was supported by the DFG Research Unit 1735, DE 502/26-2

References

Anderson, T. W. (2003). Multivariate Statistical Analysis. John Wiley & Sons, New York.

Bai, Z., Jiang, D., Yao, J.-F., and Zheng, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. Annals of Statistics, 37:3822–3840.

Bai, Z. D. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large dimensional sample covariance matrices. Annals of Probability, 32:553–605.

Bai, Z. D. and Silverstein, J. W. (2010). Spectral Analysis of Large Dimensional Random Matrices. Springer, New York.

Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. Annals of Statistics, 36:199–227.

Birke, M. and Dette, H. (2005). A note on testing the covariance matrix for large dimension. Statistics & Probability Letters, 74:281–289.

Bodnar, T., Gupta, A. K., and Parolya, N. (2014). On the strong convergence of the optimal linear shrinkage estimator for large dimensional covariance matrix. Journal of Multivariate Analysis, 132:215–228.

Bodnar, T., Gupta, A. K., and Parolya, N. (2016). Direct shrinkage estimation of large dimensional precision matrix. Journal of Multivariate Analysis, 146:223–236.

Cai, T., Lui, W., and Luo, X. (2011). A constrained l_1 minimization approach to sparse precision matrix estimation. Journal of the American Statistical Association, 106:594–607.

Cai, T., Ren, Z., and Zhou, H. H. (2013). Optimal rates of convergence for estimating Toeplitz covariance matrices. Probability Theory and Related Fields, 156(1):101–143.

Cai, T. and Shen, X. (2011). High-Dimensional Data Analysis, Volume 2, Frontiers of Statistics. World Scientific, Singapore.

Cai, T. and Zhou, H. (2012). Minimax estimation of large covariance matrices under l_1 norm. Statistica Sinica, 22:1319–1378.

Chen, S. X., Zhang, L.-X., and Zhong, P.-S. (2010). Testing high dimensional covariance matrices. Journal of the American Statistical Association, 105:810–819.

Devijver, E. and Gallopin, M. (2016). Block-diagonal covariance selection for high-dimensional Gaussian graphical models. Journal of the American Statistical Association, to appear.

Dozier, R. B. and Silverstein, J. W. (2007). On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices. Journal of Multivariate Analysis, 98:678–694.

Fan, J. and Li, R. (2006). Statistical challenges with high dimensionality: feature selection in knowledge discovery.

Fisher, R. A. (1939). The sampling distribution of some statistics obtained from non-linear equations. Annals of Eugenics, 9(3):238–249.

Fisher, T. J., Sun, X., and Gallagher, C. M. (2010). A new test for sphericity of the covariance matrix for high dimensional data. Journal of Multivariate Analysis, 101:2554–2570.

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. Journal of the Japan Statistical Society, 34(1):19–26.

Gupta, A. and Nagar, D. (2000). Matrix Variate Distributions. Chapman and Hall/CRC, Boca Raton.

Gupta, A. K. and Xu, J. (2006). On some tests of the covariance matrix under general conditions. Annals of the Institute of Statistical Mathematics, 58:101–114.

Hachem, W., Kharouf, J., Najim, J., and Silverstein, J. W. (2012). A CLT for information-theoretic statistics of non-centered Gram random matrices. Random Matrices: Theory and Applications, 1(2):1150010.

Hyodo, M., Shutoh, N., Nishiyama, T., and Pavlenko, T. (2015). Testing block-diagonal covariance structure for high-dimensional data. Statistica Neerlandica, 69(4):460–482.

Jiang, D., Bai, Z., and Zheng, S. . (2013). Sci. China Math., 56(1):135–147.

Jiang, T. and Yang, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. Annals of Statistics, 41:2029–2074.

John, S. (1971). Some optimal multivariate tests. Biometrika, 58:123–127.

Johnstone, I. (2006). High dimensional statistical inference and random matrices.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. Annals of Statistics, 29:295–327.

Johnstone, I. M. (2008). Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy-Widom limits and rates of convergence. Annals of Statistics, 36:2638–2716.

Kakizawa, Y. and Iwashita, T. (2008). A comparison of higher-order local powers of a class of one-way MANOVA tests under general distributions. Journal of Multivariate Analysis, 99(6):1128–1153.

Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. Annals of Statistics, 30:1081–1102.

Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. Annals of Statistics, 40:1024–1060.

Markowitz, H. (1952). Portfolio selection. Journal of Finance, 7(1):77–91.

Mauchly, J. W. (1940). Significance test for sphericity of a normal N-variate distribution. Annals of Mathematical Statistics, 11:204–209.

Muirhead, R. J. (1982). Aspects of Multivariate Statistical Theory. John Wiley & Sons, New York.

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

Schott, J. R. (2007). Some high-dimensional tests for a one-way MANOVA. Journal of Multivariate Analysis, 98(9):1825–1839.

Wang, C., Pan, G. M., Tong, T. J., and Zhu, L. (2015). Shrinkage estimator of large dimensional precision matrix using random matrix theory. Statistica Sinica, 25:993–1008.

Yamada, Y., Hyodo, M., Shutoh, N., and Nishiyama, T. (2017). Testing block-diagonal covariance structure for high-dimensional data under non-normality. Journal of Multivariate Analysis, 155:305–316.

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

Yao, J., Bai, Z., and Zheng, S. (2015). Large Sample Covariance Matrices and High-Dimensional Data Analysis (No. 39). Cambridge University Press, New York.

Zheng, S. (2012). Central limit theorems for linear spectral statistics of large dimensional F-matrices. Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, 48(2):444–476.

Zheng, S., Bai, Z., and Yao, J. (2017). Clt for eigenvalue statistics of large-dimensional general fisher matrices with applications. Bernoulli, 23(2):1130–1178.

Zheng, Z., Bai, Z. D., and Yao, J. (2015). CLT for linear spectral statistics of a rescaled sample precision matrix. Random Matrices: Theory and Applications, 04(04):1550014.

5 Appendix

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

$$LSS_n = (p - p_1) \int_0^\infty f(x)\,\mathrm{d}F_n(x) = \sum_{i=1}^{p-p_1} f(v_i)\,, \tag{A.1}$$

where $v_1 \geq v_2 \geq ... \geq v_{p-p_1}$ are the ordered eigenvalues of the matrix $\mathbf{W}\mathbf{T}^{-1}$,

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

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

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

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

$$F(dx) = q(x)\mathbb{1}_{[a,b]}(x)dx + (1-1/\gamma_1)\mathbb{1}_{\gamma_1>1}\delta_0(dx) \quad \text{with} \tag{A.2}$$

$$q(x) = \frac{1-\gamma_2}{2\pi x(\gamma_1+\gamma_2 x)}\sqrt{(b-x)(x-a)}, \quad a = \frac{(1-h)^2}{(1-\gamma_2)^2}, \quad b = \frac{(1+h)^2}{(1-\gamma_2)^2}. \tag{A.3}$$

where

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

The asymptotic properties of a centred version of (A.2) have been determined by Zheng (2012), who showed that for any functions f, g, which are analytic in an open region of the complex plane containing the interval $[a, b]$, the random vector

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

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

$$\begin{aligned} E[X_f] &= \frac{1}{2\pi i}\oint f(z)\,\mathrm{dlog}\left(\frac{\frac{1-c}{1-c_1}m_0^2(z)+2m_0(z)+2-c/c_1}{\frac{1-c}{1-c_1}m_0^2(z)+2m_0(z)+1}\right) \\ &+ \frac{1}{2\pi i}\oint f(z)\,\mathrm{dlog}\left(\frac{1-\frac{c-c_1}{1-c_1}m_0^2(z)}{(1+m_0^2(z))^2}\right) \\ \mathrm{Cov}[X_f, X_g] &= -\frac{1}{2\pi^2}\oint\oint \frac{f(z_1)g(z_2)}{(m_0(z_1)-m_0(z_2))^2}\,\mathrm{d}m_0(z_1)\,\mathrm{d}m_0(z_2) \end{aligned}$$

respectively. Here $m_0(z) = \underline{m}_{\gamma_2}(-\underline{m}(z))$ with $\underline{m}_{\gamma_2}(z) = -\frac{1-\gamma_2}{z} + \gamma_2 m_{\gamma_2}(z)$ and $\underline{m}(z) = -\frac{1-\gamma_1}{z} + \gamma_1 m(z)$, where $m(z)$ denotes the Stieltjes transform of the function (A.2) and $m_{\gamma_2}(z)$ is the Stieltjes transformation of the matrix $\mathbf{W}$ under H_0. The integrals are taken over arbitrary positively oriented countur which contains the interval $[a, b]$. Note that this result is only applicable under the null hypothesis H_0, because under H_1 the unconditional distribution of the random matrix $\mathbf{W}$ is no longer a central Wishart distribution. Therefore further investigation are needed in this situation (see the proof of Theorem 2)

The distributions of the test statistics T_W, T_{LH}, and T_{BNP} are obtained as special cases using the functions $f_W(x) = \log(1+x)$, $f_{LH}(x) = x$ and $f_{BNP}(x) = \frac{x}{1+x}$, respectively. Thus, we need to calculate the asymptotic means, variances and the terms $\int f(x)dF(x)$ in these cases. The asymptotic means and variances for f_W and f_{LH} can be deduced from Examples 4.1 and 4.2 in Zheng (2012) and we only need to find the corresponding quantities for f_{BNP}.

Let w and d be the positive solutions of the equation

$$|1+hz|^2 + (1-\gamma_2)^2 = |w+dz|^2 \tag{A.4}$$

for any $z \in \mathbb{C}$ with $|z| = 1$ which also satisfy

$$w^2 + d^2 = h^2 + 1 + (1-\gamma_2)^2 \quad \text{and} \quad wd = h$$

and, consequently, it holds that

$$(1-\gamma_2)^2 = (1-d^2)(w^2-1)\ , \quad (1+h)^2 = (w+d)^2 - (1-\gamma_2)^2\ , \ (1-h)^2 = (w-d)^2 - (1-\gamma_2)^2\,.$$

Further, without loss of generality[2] we assume that $w > d$.

In using that $|1+hz|^2 = (1+hz)(h+z)/z$, $|w+dz|^2 = (w+dz)(d+wz)$ and due to Corollary 3.2 of Zheng (2012), we get

$$\begin{aligned}
\mathrm{E}[X_{f_{BNP}}] &= \lim_{r\downarrow 1} \frac{1}{4\pi i} \oint\limits_{|z|=1} \frac{|1+hz|^2/(1-\gamma_2)^2}{|1+hz|^2/(1-\gamma_2)^2+1}\left[\frac{1}{z-r^{-1}} + \frac{1}{z+r^{-1}} - \frac{2}{z+\gamma_2/h}\right]\mathrm{d}z \\
&= \lim_{r\downarrow 1} \frac{1}{4\pi i} \oint\limits_{|z|=1} \frac{|1+hz|^2}{|w+dz|^2}\left[\frac{1}{z-r^{-1}} + \frac{1}{z+r^{-1}} - \frac{2}{z+\gamma_2/h}\right]\mathrm{d}z \\
&= \lim_{r\downarrow 1} \frac{1}{4\pi i} \oint\limits_{|z|=1} \frac{(1+hz)(h+z)}{(w+dz)(d+wz)}\left[\frac{1}{z-r^{-1}} + \frac{1}{z+r^{-1}} - \frac{2}{z+\gamma_2/h}\right]\mathrm{d}z \\
&= \lim_{r\downarrow 1} \frac{1}{2}\left[\left.\frac{(1+hz)(h+z)}{(w+dz)(d+wz)}\right|_{z=r^{-1}} + \left.\frac{(1+hz)(h+z)}{(w+dz)(d+wz)}\right|_{z=-r^{-1}} - 2\left.\frac{(1+hz)(h+z)}{(w+dz)(d+wz)}\right|_{z=-\frac{\gamma_2}{h}}\right] \\
&+ \lim_{r\downarrow 1} \frac{1}{2w}\left.\frac{(1+hz)(h+z)}{(w+dz)}\left[\frac{1}{z-r^{-1}} + \frac{1}{z+r^{-1}} - \frac{2}{z+\gamma_2/h}\right]\right|_{z=-d/w} \\
&= \lim_{r\downarrow 1} \frac{1}{2}\left[\frac{(1+hr^{-1})(h+r^{-1})}{(w+dr^{-1})(d+wr^{-1})} + \frac{(1-hr^{-1})(h-r^{-1})}{(w-dr^{-1})(d-wr^{-1})}\right. \\
&- \left. 2\,\frac{(1-\gamma_2)(h-\gamma_2/h)}{(w-\gamma_2 d/h)(d-\gamma_2 w/h)} + 2\frac{(1-dh/w)(h-d/w)}{(w^2-d^2)}\left(-\frac{d/w}{(d/w)^2-r^{-2}} - \frac{1}{\gamma_2/h-d/w}\right)\right].
\end{aligned}$$

[2] It holds that $|w+dz|^2 = |d+wz|^2$ for $|z| = 1$.

Taking now the limit $r \downarrow 1$ in (A.5) we obtain the following result for the mean

$$\begin{aligned}
\mathrm{E}[X_{f_{BNP}}] &= \frac{1}{2}\left[\frac{(1+h)^2}{(w+d)^2}+\frac{(1-h)^2}{(w-d)^2}\right]-\frac{(1-\gamma_2)(h^2-\gamma_2)}{(w^2-\gamma_2)(d^2-\gamma_2)}+\frac{(1-d^2)(w^2-1)d^2(w^2-\gamma_2)}{(w^2-d^2)^2(d^2-\gamma_2)}\\
&= \left(1-\frac{(1-\gamma_2)^2(w^2+d^2)}{(w^2-d^2)^2}\right)-\left(1+\frac{(1-\gamma_2)^2\gamma_2}{(w^2-\gamma_2)(d^2-\gamma_2)}\right)+\frac{(1-\gamma_2)^2d^2(w^2-\gamma_2)}{(w^2-d^2)^2(d^2-\gamma_2)}\\
&= -\frac{(1-\gamma_2)^2(w^2+d^2)}{(w^2-d^2)^2}-\frac{(1-\gamma_2)^2\gamma_2}{(w^2-\gamma_2)(d^2-\gamma_2)}+\left(\frac{(1-\gamma_2)^2\gamma_2(w^2-\gamma_2)}{(w^2-d^2)^2(d^2-\gamma_2)}+\frac{(1-\gamma_2)^2(w^2-\gamma_2)}{(w^2-d^2)^2}\right)\\
&= \frac{(1-\gamma_2)^2}{(w^2-d^2)^2}\left(-(d^2+\gamma_2)-\gamma_2\left[\frac{(w^2-d^2)^2-(w^2-\gamma_2)^2}{(w^2-\gamma_2)(d^2-\gamma_2)}\right]\right)\\
&= -\frac{(1-\gamma_2)^2w^2(d^2-\gamma_2)}{(w^2-d^2)^2}\,.
\end{aligned}$$

Similarly, we have for the variance

$$\begin{aligned}
\mathrm{Var}[X_{f_{BNP}}] &= -\lim_{r\downarrow 1}\frac{1}{2\pi^2}\oint\limits_{|z_2|=1}\frac{(1+hz_2)(h+z_2)}{(w+dz_2)(d+wz_2)}\left(\oint\limits_{|z_1|=1}\frac{(1+hz_1)(h+z_1)}{(w+dz_1)(d+wz_1)(z_1-rz_2)^2}\,\mathrm{d}z_1\right)\mathrm{d}z_2\\
&= -\lim_{r\downarrow 1}\frac{i}{\pi}\oint\limits_{|z_2|=1}\frac{(1+hz_2)(h+z_2)}{(w+dz_2)(d+wz_2)}\left(\left.\frac{(1+hz_1)(h+z_1)}{w(w+dz_1)(z_1-rz_2)^2}\right|_{z_1=-\frac{d}{w}}\right)\mathrm{d}z_2\\
&= -\frac{i}{\pi}\oint\limits_{|z_2|=1}\frac{(1+hz_2)(h+z_2)}{(w+dz_2)(d+wz_2)}\left(-\frac{dw(1-d^2)(w^2-1)}{(w^2-d^2)(d+wz_2)^2}\right)\mathrm{d}z_2\\
&= -\frac{h(1-\gamma_2)^2}{w^3(w^2-d^2)}\left[\left.\frac{\partial^2}{\partial z_2^2}\frac{(1+hz_2)(h+z_2)}{(w+dz_2)}\right|_{z_2=-d/w}\right]\\
&= -2\frac{hd(w^2-1)(1-\gamma_2)^2}{w^3(w^2-d^2)^2}\left[1-\frac{(w^2-1)d}{w^2-d^2}+\frac{d^2w^2(1-d^2)}{(w^2-d^2)^2}\right]\\
&= 2\frac{d^2(1-\gamma_2)^4}{w^2(1+d)(w^2-d^2)^4}\left(w^2(w^2+d)+d^3(w^2-1)\right).
\end{aligned}$$

Due to Theorem 2.23 in Yao et al. (2015), the terms $\int_b^b f(x)\,\mathrm{d}F(x)$ can be calculated in the following way

$$\int\limits_b^b f(x)\,\mathrm{d}F(x) = -\frac{h^2(1-\gamma_2)}{4\pi i}\oint\limits_{|z|=1} f\left(\frac{|1+hz|^2}{(1-\gamma_2)^2}\right)\frac{(1-z^2)^2}{z(1+hz)(z+h)(\gamma_2 z+h)(\gamma_2+hz)}\,\mathrm{d}z\,, \tag{A.5}$$

where the interval $[a,b]$ is the support of limiting spectral distribution F of the Fisher matrix $\mathbf{WT}^{-1}$ defined in (A.3). The function f_{LH} has already been considered in Yao et al. (2015), Example 2.25, that is $s_{LH}=\int_b^b x\,\mathrm{d}F(x)=\frac{1}{1-\gamma_2}$. Next we determine he corresponding terms for f_W and f_{BNP} noting that

$$s_W = -\frac{h^2(1-\gamma_2)}{4\pi i}\oint\limits_{|z|=1}\frac{\log\left((1-\gamma_2)^{-2}|w+dz|^2\right)(1-z^2)^2}{z(1+hz)(z+h)(\gamma_2 z+h)(\gamma_2+hz)}\,\mathrm{d}z = -\log\left((1-\gamma_2)^2\right)+I_1+I_2\,,$$

where we used again (A.4) and the terms I_1 and I_2 are defined by

$$\begin{aligned} I_1 &= -\frac{h^2(1-\gamma_2)}{4\pi i} \oint\limits_{|z|=1} \frac{\log(w+dz)(1-z^2)^2}{z(1+hz)(z+h)(\gamma_2 z+h)(\gamma_2+hz)} \mathrm{d}z \\ I_2 &= -\frac{h^2(1-\gamma_2)}{4\pi i} \oint\limits_{|z|=1} \frac{\log(w+d\bar{z})(1-z^2)^2}{z(1+hz)(z+h)(\gamma_2 z+h)(\gamma_2+hz)} \mathrm{d}z\,. \end{aligned}$$

A change of variables yields $I_1 = I_2$ and we obtain (see Yao et al. (2015) for detailed calculation)

$$2I_1 = -\frac{h^2(1-\gamma_2)}{2\pi i} \oint\limits_{|z|=1} \frac{\log(w+dz)(1-z^2)^2}{z(1+hz)(z+h)(\gamma_2 z+h)(\gamma_2+hz)} \mathrm{d}z \begin{cases} \frac{1-\gamma_1}{\gamma_1}\log(w-dh), & \gamma_1 \in (0,1) \\ 0, & \gamma_1 = 1 \\ -\frac{1-\gamma_1}{\gamma_1}\log(w-d/h), & \gamma_1 > 1\,, \end{cases}$$

which yields the desired representation of s_w. Similarly, we obtain

$$\begin{aligned} s_{BNP} &= -\frac{h^2(1-\gamma_2)}{4\pi i} \oint\limits_{|z|=1} \frac{(1+hz)(h+z)}{(w+dz)(d+wz)} \frac{(1-z^2)^2}{z(1+hz)(z+h)(\gamma_2 z+h)(\gamma_2+hz)} \mathrm{d}z \\ &= -\frac{h^2(1-\gamma_2)}{4\pi i} \oint\limits_{|z|=1} \frac{(1-z^2)^2}{z(w+dz)(d+wz)(\gamma_2 z+h)(\gamma_2+hz)} \mathrm{d}z \\ &= -\frac{h^2(1-\gamma_2)}{2} \left(\left.\frac{(1-z^2)^2}{hz(w+dz)(d+wz)(\gamma_2 z+h)}\right|_{z=-\frac{\gamma_2}{h}} + \left.\frac{(1-z^2)^2}{zw(w+dz)(\gamma_2 z+h)(\gamma_2+hz)}\right|_{z=-\frac{d}{w}} \right. \\ &+ \left. \left.\frac{(1-z^2)^2}{(w+dz)(d+wz)(\gamma_2 z+h)(\gamma_2+hz)}\right|_{z=0} \right) \\ &= \frac{1-\gamma_2}{w^2-\gamma_2}\,, \end{aligned}$$

which completes the proof of Theorem 1. □

Proof of Theorem 2: Since $(n-p_1)\mathbf{T} \sim W_{p-p_1}(n-p_1, \mathbf{\Sigma}_{22\cdot1})$, $p_1\mathbf{W}|\mathbf{S}_{11} \sim W_{p-p_1}(p_1, \mathbf{\Sigma}_{22\cdot1}, \mathbf{\Omega}_1)$ with $\mathbf{\Omega}_1 = \mathbf{\Omega}_1(\mathbf{S}_{11}) = \mathbf{\Sigma}_{22\cdot1}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{S}_{11}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12} = \mathbf{\Sigma}_{22\cdot1}^{-1}\mathbf{M}\mathbf{M}^\top$, $\mathbf{S}_{11} \sim W_{p-p_1}(p_1, \mathbf{\Sigma}_{22\cdot1})$, 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

$$\begin{aligned} \mathbf{W} &\overset{d}{=} \frac{1}{p_1}\mathbf{\Sigma}_{22\cdot1}^{1/2}(\mathbf{X}+\mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{M})(\mathbf{X}+\mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{M})^\top\mathbf{\Sigma}_{22\cdot1}^{1/2} \\ \mathbf{T} &\overset{d}{=} \frac{1}{n-p_1}\mathbf{\Sigma}_{22\cdot1}^{1/2}\mathbf{Y}\mathbf{Y}^\top\mathbf{\Sigma}_{22\cdot1}^{1/2}\,, \end{aligned}$$

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

$$\mathbf{W}\mathbf{T}^{-1} \overset{d}{=} \frac{1}{p_1}\mathbf{\Sigma}_{22\cdot1}^{1/2}(\mathbf{X}+\mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{M})(\mathbf{X}+\mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{M})^\top\mathbf{\Sigma}_{22\cdot1}^{1/2}\left(\frac{1}{n-p_1}\mathbf{\Sigma}_{22\cdot1}^{1/2}\mathbf{Y}\mathbf{Y}^\top\mathbf{\Sigma}_{22\cdot1}^{1/2}\right)^{-1}.$$

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

$$\widetilde{\mathbf{W}} = \frac{1}{p_1}(\mathbf{X}+\mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{M})(\mathbf{X}+\mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{M})^\top \quad\text{and}\quad \widetilde{\mathbf{T}} = \frac{1}{n-p_1}\mathbf{Y}\mathbf{Y}^\top.$$

From Theorem 2.1 of Zheng et al. (2015) 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

$$zs_{\widetilde{\mathbf{W}}}(z) = -1 + \int \frac{tdH(t)}{t - z(1+\gamma_2 z s_{\widetilde{\mathbf{W}}}(z))}\,, \tag{A.6}$$

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.6) does not depend on the condition $\widetilde{\mathbf{W}}$ and rewriting (A.6), 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))

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

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

$$\widetilde{\mathbf{R}} = 1/p_1 \mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{S}_{11}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot1}^{-1/2} = c_{1,n}^{-1}1/n\mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{S}_{11}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot1}^{-1/2}.$$

satisfying the following equation

$$\begin{aligned} m_{\tilde{H}}(z) &= \int \frac{(1-(c-c_1)-(c-c_1)zm_{\tilde{H}}(z))^{-1}dG(t)}{c_1^{-1}t - \frac{z}{1-(c-c_1)-(c-c_1)zm_{\tilde{H}}(z)}} \\ &= c_1^{-1}(1-(c-c_1)-(c-c_1)zm_{\tilde{H}}(z))^{-1}m_G\left(\frac{c_1 z}{1-(c-c_1)-(c-c_1)zm_{\tilde{H}}(z)}\right) \end{aligned}$$

where $G(t)$ is the limiting spectral distribution of the matrix $\mathbf{R} = \mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot1}^{-1/2}$ 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 et al. (2015).

Lemma 1. *Conditionally on $\mathbf{S}_{11}$ the distribution of the matrix $\hat{\mathbf{W}}\hat{\mathbf{T}}^{-1}$ solely depends on the eigenvalues of the non-centrality matrix $\mathbf{\Omega}_1(\mathbf{S}_{11})$ and does not depend on the corresponding eigenvectors. Moreover, the unconditional distribution of the eigenvalues of matrix $\hat{\mathbf{W}}\hat{\mathbf{T}}^{-1}$ depends only on the eigenvalues of the matrix $\widetilde{\mathbf{R}} = \mathbf{\Sigma}_{11}^{-1/2}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot1}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1/2}$. Precisely, it holds for $p_1 \geq p - p_1$*

$$\begin{aligned} g_{v_1,v_2,\dots,v_{p-p_1}}(l_1,l_2,\dots,l_{p-p_1}) &= \frac{\pi^{(p-p_1)^2/2}\Gamma_{p-p_1}\left(\frac{n}{2}\right)}{\Gamma_{p-p_1}\left(\frac{p_1}{2}\right)\Gamma_{p-p_1}\left(\frac{n-p_1}{2}\right)\Gamma_{p-p_1}\left(\frac{p-p_1}{2}\right)}|\mathbf{I}_{p_1}+\widetilde{\mathbf{R}}|^{-n/2}\frac{|\mathbf{L}|^{p_1-(p+1)/2}}{|\mathbf{I}_{p-p_1}+\mathbf{L}|^{n/2}}\prod_{i<j}^{p-p_1}(l_i-l_j) \\ &\times\ {}_2F_1^{(p-p_1)}\left(\frac{n}{2},\frac{n}{2};\frac{p_1}{2};(\mathbf{I}_{p_1}+\widetilde{\mathbf{R}})^{-1}\widetilde{\mathbf{R}},\mathbf{L}(\mathbf{I}_{p-p_1}+\mathbf{L})^{-1}\right), \end{aligned}$$

whereas for $p_1 < p - p_1$ *we have*

$$g_{v_1,v_2,...,v_{p-p_1}}(l_1, l_2, ..., l_{p-p_1}) = \frac{\Gamma_{p_1}\left(\frac{n}{2}\right)}{\Gamma_{p_1}\left(\frac{p-p_1}{2}\right)\Gamma_{p_1}\left(\frac{n-p+p_1}{2}\right)} \frac{\pi^{p_1^2/2}}{\Gamma_{p_1}\left(\frac{p_1}{2}\right)} |\mathbf{I}_{p_1} + \widetilde{\mathbf{R}}|^{-n/2} \frac{|\mathbf{L}|^{(p-1)/2-p_1}}{|\mathbf{I}_{p_1} + \mathbf{L}|^{n/2}} \prod_{i<j}^{p_1} (l_i - l_j)$$
$$\times \; {}_2F_1^{(p_1)}\left(\frac{n}{2}, \frac{n}{2}; \frac{p-p_1}{2}; (\mathbf{I}_{p_1} + \widetilde{\mathbf{R}})^{-1}\widetilde{\mathbf{R}}, \mathbf{L}(\mathbf{I}_{p-p_1} + \mathbf{L})^{-1}\right)$$

with $\mathbf{L} = diag(l_1, \ldots, l_{p-p_1})$.

Proof. First, we note that this distribution is independent of $\mathbf{\Sigma}_{22\cdot 1}$ since the eigenvalues of $\widehat{\mathbf{W}}\hat{\mathbf{T}}^{-1}$ coincide with the eigenvalues of $\widetilde{\mathbf{W}}\widetilde{\mathbf{T}}^{-1}$ with

$$\widetilde{\mathbf{W}} = \mathbf{\Sigma}_{22\cdot 1}^{-1/2}\hat{\mathbf{W}}\mathbf{\Sigma}_{22\cdot 1}^{-1/2} \quad \text{and} \quad \widetilde{\mathbf{T}} = \mathbf{\Sigma}_{22\cdot 1}^{-1/2}\hat{\mathbf{T}}\mathbf{\Sigma}_{22\cdot 1}^{-1/2},$$

where $\mathbf{\Sigma}_{22\cdot 1}^{1/2}$ denotes the symmetric square root of $\mathbf{\Sigma}_{22\cdot 1}$; $\widetilde{\mathbf{T}} \sim W_{p-p_1}(n - p_1, \mathbf{I})$; $\widetilde{\mathbf{W}}|\mathbf{S}_{11} \sim W_{p-p_1}(p_1, \mathbf{I}, \mathbf{\Omega}_1(\mathbf{S}_{11}))$; $\widetilde{\mathbf{T}}$ and $(\widetilde{\mathbf{W}}, \mathbf{S}_{11})$ are independent.

We distinguish between the following two cases: (a) $p_1 \geq p - p_1$ and (b) $p_1 < p - p_1$.

Case (a): $p_1 \geq p - p_1$

Here, we first note, that the eigenvalues of $\widetilde{\mathbf{W}}^{1/2}\widetilde{\mathbf{T}}^{-1}\widetilde{\mathbf{W}}^{1/2}$, where $\widetilde{\mathbf{W}}^{1/2}$ is the symmetric square root of $\widetilde{\mathbf{W}}$, coincides with $\widetilde{\mathbf{W}}\widetilde{\mathbf{T}}^{-1}$ and, consequently, with $\hat{\mathbf{W}}\hat{\mathbf{T}}^{-1}$. Furthermore, the application of (Muirhead, 1982, Theorem 10.4.1) leads to the conditional density of $\widetilde{\mathbf{W}}^{1/2}\widetilde{\mathbf{T}}^{-1}\widetilde{\mathbf{W}}^{1/2}$ given $\mathbf{S}_{11}$, expressed as

$$g_{\widetilde{\mathbf{W}}^{1/2}\widetilde{\mathbf{T}}^{-1}\widetilde{\mathbf{W}}^{1/2}|\mathbf{S}_{11}}(\mathbf{F}|\mathbf{S}_{11}) = etr\left(-\frac{1}{2}\mathbf{\Omega}_1(\mathbf{S}_{11})\right) {}_1F_1\left(\frac{n}{2}; \frac{p_1}{2}; \frac{1}{2}\mathbf{\Omega}_1(\mathbf{S}_{11})\mathbf{F}(\mathbf{I}_{p-p_1} + \mathbf{F})^{-1}\right)$$
$$\times \; \frac{\Gamma_{p-p_1}\left(\frac{n}{2}\right)}{\Gamma_{p-p_1}\left(\frac{p_1}{2}\right)\Gamma_{p-p_1}\left(\frac{n-p_1}{2}\right)} \frac{|\mathbf{F}|^{p_1-(p+1)/2}}{|\mathbf{I}_{p-p_1} + \mathbf{F}|^{n/2}} \quad \text{for} \quad \mathbf{F} > \mathbf{O},$$

where $etr(\mathbf{A}) = \exp(tr(\mathbf{A}))$ for a symmetric matrix $\mathbf{A}$, $\Gamma_m(q) = \pi^{m(m-1)/4}\prod_{i=1}^m \Gamma\left(q - \frac{1}{2}(i-1)\right)$, $q > (m-1)/2$ is the multivariate m-dimensional gamma function, the statement $\mathbf{F} > \mathbf{O}$ means that $\mathbf{F}$ is positive definite, and ${}_1F_1\left(\frac{n}{2}; \frac{p_1}{2}; \frac{1}{2}\mathbf{\Omega}_1(\mathbf{S}_{11})\mathbf{F}(1 + \mathbf{F})^{-1}\right)$ is the hypergeometric function of matrix argument defined by (see, e.g., (Gupta and Nagar, 2000, Section 1.6))

$${}_qF_k(a_1, ..., a_q; b_1, ..., b_k; \mathbf{A}) = \sum_{i=0}^{\infty}\sum_{\kappa} \frac{(a_1)_\kappa ...(a_q)_\kappa}{(b_1)_\kappa ...(b_k)_\kappa} \frac{C_\kappa(\mathbf{A})}{i!}, \tag{A.7}$$

where $\mathbf{A} : m \times m$ is a symmetric matrix, $\sum_\kappa$ denotes the summation over all partitions κ, $(c)_\kappa = \frac{\Gamma_m(c,\kappa)}{\Gamma_m(c)}$ with $\Gamma_m(c, \kappa) = \pi^{m(m-1)/4}\prod_{j=1}^m \Gamma\left(c + i_j - \frac{1}{2}(j-1)\right)$, $c > (m-1)/2 - i_m$, for $\kappa = (i_1, ..., i_m)$ with $i_1 \geq ...i_m \geq 0$ and $\sum_{j=1}^m i_j = i$, $C_\kappa(\mathbf{A})$ stands for the zonal polynomial (c.f., (Gupta and Nagar, 2000, Section 1.5)).

Let $\mathcal{O}(p - p_1)$ denote the space of $(p - p_1) \times (p - p_1)$ orthogonal matrices. Then, the density of the eigenvalues of $\widetilde{\mathbf{W}}^{1/2}\widetilde{\mathbf{T}}^{-1}\widetilde{\mathbf{W}}^{1/2}$, $v_1 \geq v_2 \geq ... \geq v_{p-p_1} > 0$, given $\mathbf{S}_{11}$ is expressed as (see, e.g., (Muirhead, 1982, Theorem 3.2.17))

$$g_{v_1,v_2,...,v_{p-p_1}|\mathbf{S}_{11}}(l_1,l_2,...,l_{p-p_1}|\mathbf{S}_{11}) = \frac{\Gamma_{p-p_1}\left(\frac{n}{2}\right)}{\Gamma_{p-p_1}\left(\frac{p_1}{2}\right)\Gamma_{p-p_1}\left(\frac{n-p_1}{2}\right)} \frac{\pi^{(p-p_1)^2/2}}{\Gamma_{p-p_1}\left(\frac{p-p_1}{2}\right)} \prod_{i<j}^{p-p_1}(l_i-l_j)$$

$$\times \; etr\left(-\frac{1}{2}\mathbf{\Omega}_1(\mathbf{S}_{11})\right) \int\limits_{\mathbf{H}\in\mathcal{O}(p-p_1)} \frac{|\mathbf{HLH}'|^{p_1-(p+1)/2}}{|\mathbf{I}_{p-p_1}+\mathbf{HLH}'|^{n/2}} {}_1F_1\left(\frac{n}{2};\frac{p_1}{2};\frac{1}{2}\mathbf{\Omega}_1(\mathbf{S}_{11})\mathbf{HL}(\mathbf{I}_{p-p_1}+\mathbf{L})^{-1}\mathbf{H}'\right)\mathrm{d}\mathbf{H}$$

$$= \frac{\Gamma_{p-p_1}\left(\frac{n}{2}\right)}{\Gamma_{p-p_1}\left(\frac{p_1}{2}\right)\Gamma_{p-p_1}\left(\frac{n-p_1}{2}\right)} \frac{\pi^{(p-p_1)^2/2}}{\Gamma_{p-p_1}\left(\frac{p-p_1}{2}\right)} \frac{|\mathbf{L}|^{p_1-(p+1)/2}}{|\mathbf{I}_{p-p_1}+\mathbf{L}|^{n/2}} \prod_{i<j}^{p-p_1}(l_i-l_j)$$

$$\times \; etr\left(-\frac{1}{2}\mathbf{\Omega}_1(\mathbf{S}_{11})\right) {}_1F_1^{(p-p_1)}\left(\frac{n}{2};\frac{p_1}{2};\frac{1}{2}\mathbf{\Omega}_1(\mathbf{S}_{11}),\mathbf{L}(\mathbf{I}_{p-p_1}+\mathbf{L})^{-1}\right)$$

where the last equality follows from (Gupta and Nagar, 2000, Theorem 1.6.1) and ${}_qF_k^{(m)}$ denotes the hypergeometric function of two matrices defined by

$${}_qF_k^{(m)}(a_1,...,a_q;b_1,...,b_k;\mathbf{A},\mathbf{B}) \;=\; \sum_{i=0}^{\infty}\sum_{\kappa}\frac{(a_1)_\kappa...(a_q)_\kappa}{(b_1)_\kappa...(b_k)_\kappa}\frac{C_\kappa(\mathbf{A})C_\kappa(\mathbf{B})}{k!C_\kappa(\mathbf{I}_m)}, \tag{A.8}$$

Using (A.8), integrating over $\mathbf{S}_{11}$, applying (Gupta and Nagar, 2000, Lemma 1.5.2) and (A.7), we get the unconditional distribution of $v_1 \geq v_2 \geq ... \geq v_{p-p_1} > 0$ given by

$$g_{v_1,v_2,...,v_{p-p_1}}(l_1,l_2,...,l_{p-p_1}) = \int\limits_{\mathbf{S}_{11}>\mathbf{O}} g_{v_1,v_2,...,v_{p-p_1}|\mathbf{S}_{11}}(l_1,l_2,...,l_{p-p_1}|\mathbf{S}_{11})g(\mathbf{S}_{11})\mathrm{d}\mathbf{S}_{11}$$

$$= \frac{\Gamma_{p-p_1}\left(\frac{n}{2}\right)}{\Gamma_{p-p_1}\left(\frac{p_1}{2}\right)\Gamma_{p-p_1}\left(\frac{n-p_1}{2}\right)} \frac{\pi^{(p-p_1)^2/2}}{\Gamma_{p-p_1}\left(\frac{p-p_1}{2}\right)} \frac{|\mathbf{L}|^{p_1-(p+1)/2}}{|\mathbf{I}_{p-p_1}+\mathbf{L}|^{n/2}} \prod_{i<j}^{p-p_1}(l_i-l_j)\frac{1}{2^{p_1 n/2}\Gamma_{p_1}\left(\frac{n}{2}\right)}|\mathbf{\Sigma}_{11}|^{-n/2}$$

$$\times \int\limits_{\mathbf{S}_{11}>\mathbf{O}} |\mathbf{S}_{11}|^{(n-p_1-1)/2} etr\left(-\frac{1}{2}\mathbf{S}_{11}\mathbf{\Sigma}_{11}^{-1}\right) etr\left(-\frac{1}{2}\mathbf{S}_{11}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot1}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\right)$$

$$\times \; {}_1F_1^{(p-p_1)}\left(\frac{n}{2};\frac{p_1}{2};\frac{1}{2}\mathbf{S}_{11}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot1}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1},\mathbf{L}(\mathbf{I}_{p-p_1}+\mathbf{L})^{-1}\right)\mathrm{d}\mathbf{S}_{11}$$

$$= \frac{\pi^{(p-p_1)^2/2}\Gamma_{p-p_1}\left(\frac{n}{2}\right)}{\Gamma_{p-p_1}\left(\frac{p_1}{2}\right)\Gamma_{p-p_1}\left(\frac{n-p_1}{2}\right)\Gamma_{p-p_1}\left(\frac{p-p_1}{2}\right)} |\mathbf{I}_{p_1}+\mathbf{\Sigma}_{11}^{-1/2}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot1}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1/2}|^{-n/2} \frac{|\mathbf{L}|^{p_1-(p+1)/2}}{|\mathbf{I}_{p-p_1}+\mathbf{L}|^{n/2}} \prod_{i<j}^{p-p_1}(l_i-l_j)$$

$$\times \; {}_2F_1^{(p-p_1)}\left(\frac{n}{2},\frac{n}{2};\frac{p_1}{2};(\mathbf{I}_{p_1}+\mathbf{\Sigma}_{11}^{-1/2}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot1}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1/2})^{-1}\mathbf{\Sigma}_{11}^{-1/2}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot1}^{-1}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1/2},\mathbf{L}(\mathbf{I}_{p-p_1}+\mathbf{L})^{-1}\right).$$

Now the statement of the lemma follows from the definition of the hypergeometric function.

Case (b): $p_1 < p - p_1$

In this case the matrix $\hat{\mathbf{W}}\hat{\mathbf{T}}^{-1}$ is not longer a positive definite matrix. In order to derive the joint distribution of the non-zero eigenvalues of $\hat{\mathbf{W}}\hat{\mathbf{T}}^{-1}$, we use that $\hat{\mathbf{W}} = \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$ and $\hat{\mathbf{T}} = \mathbf{ZZ}'$, where $\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}|\mathbf{S}_{11} \sim \mathcal{N}_{p-p_1,p_1}\left(\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{S}_{11}^{1/2},\mathbf{\Sigma}_{22\cdot1}\otimes\mathbf{I}_{p_1}\right)$ and $\mathbf{Z} \sim \mathcal{N}_{p-p_1,n-p_1}(\mathbf{O},\mathbf{\Sigma}_{22\cdot1}\otimes\mathbf{I}_{n-p_1})$. Then the application of (Muirhead, 1982, Theorem 10.4.4) leads to the conditional density of $\mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\hat{\mathbf{T}}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}$ given $\mathbf{S}_{11}$, which is expressed as

$$g_{\mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\hat{\mathbf{T}}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}|\mathbf{S}_{11}}(\mathbf{F}|\mathbf{S}_{11}) \;=\; etr\left(-\frac{1}{2}\widetilde{\mathbf{\Omega}_1}(\mathbf{S}_{11})\right) {}_1F_1\left(\frac{n}{2};\frac{p-p_1}{2};\frac{1}{2}\widetilde{\mathbf{\Omega}_1}(\mathbf{S}_{11})\mathbf{F}(\mathbf{I}_{p_1}+\mathbf{F})^{-1}\right)$$

$$\times \; \frac{\Gamma_{p_1}\left(\frac{n}{2}\right)}{\Gamma_{p_1}\left(\frac{p-p_1}{2}\right)\Gamma_{p_1}\left(\frac{n-p+p_1}{2}\right)} \frac{|\mathbf{F}|^{(p-1)/2-p_1}}{|\mathbf{I}_{p_1}+\mathbf{F}|^{n/2}} \quad \text{for} \;\; \mathbf{F} > \mathbf{O},$$

with $\widetilde{\boldsymbol{\Omega}_1}(\mathbf{S}_{11}) = \mathbf{S}_{11}^{1/2}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22\cdot1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{S}_{11}^{1/2}$.

The application of (Muirhead, 1982, Theorem 3.2.17) leads to the density of the eigenvalues of $\mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\hat{\mathbf{T}}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}$, $v_1 \geq v_2 \geq ... \geq v_{p_1} > 0$, given $\mathbf{S}_{11}$ which is expressed as

$$
\begin{aligned}
&g_{v_1,v_2,...,v_{p_1}|\mathbf{S}_{11}}(l_1,l_2,...,l_{p_1}|\mathbf{S}_{11}) = etr\left(-\frac{1}{2}\widetilde{\boldsymbol{\Omega}_1}(\mathbf{S}_{11})\right)\frac{\Gamma_{p_1}\left(\frac{n}{2}\right)}{\Gamma_{p_1}\left(\frac{p-p_1}{2}\right)\Gamma_{p_1}\left(\frac{n-p+p_1}{2}\right)}\frac{\pi^{(p_1)^2/2}}{\Gamma_{p_1}\left(\frac{p_1}{2}\right)}\prod_{i<j}^{p_1}(l_i-l_j)\\
\times &\int\limits_{\mathbf{H}\in\mathcal{O}(p_1)}\frac{|\mathbf{HLH}'|^{(p-1)/2-p_1}}{|\mathbf{I}_{p_1}+\mathbf{HLH}'|^{n/2}}\,{}_1F_1\left(\frac{n}{2};\frac{p-p_1}{2};\frac{1}{2}\widetilde{\boldsymbol{\Omega}_1}(\mathbf{S}_{11})\mathbf{HL}(\mathbf{I}_{p_1}+\mathbf{L})^{-1}\mathbf{H}'\right)\mathrm{d}\mathbf{H}\\
= &\frac{\Gamma_{p_1}\left(\frac{n}{2}\right)}{\Gamma_{p_1}\left(\frac{p-p_1}{2}\right)\Gamma_{p_1}\left(\frac{n-p+p_1}{2}\right)}\frac{\pi^{p_1^2/2}}{\Gamma_{p_1}\left(\frac{p_1}{2}\right)}\frac{|\mathbf{L}|^{(p-1)/2-p_1}}{|\mathbf{I}_{p_1}+\mathbf{L}|^{n/2}}\prod_{i<j}^{p_1}(l_i-l_j)\\
\times &etr\left(-\frac{1}{2}\mathbf{S}_{11}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22\cdot1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\right){}_1F_1^{(p_1)}\left(\frac{n}{2};\frac{p-p_1}{2};\frac{1}{2}\mathbf{S}_{11}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22\cdot1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1},\mathbf{L}(\mathbf{I}_{p_1}+\mathbf{L})^{-1}\right),
\end{aligned}
$$

where the last equality follows from (Gupta and Nagar, 2000, Theorem 1.6.1)

Then the unconditional density of $v_1 \geq v_2 \geq ... \geq v_{p_1} > 0$ is obtained by using (A.8), integrating over $\mathbf{S}_{11}$, applying (Gupta and Nagar, 2000, Lemma 1.5.2) and (A.7). It leads to

$$
\begin{aligned}
&g_{v_1,v_2,...,v_{p_1}}(l_1,l_2,...,l_{p_1}) = \int\limits_{\mathbf{S}_{11}>\mathbf{O}} g_{v_1,v_2,...,v_{p_1}|\mathbf{S}_{11}}(l_1,l_2,...,l_{p_1}|\mathbf{S}_{11})g(\mathbf{S}_{11})\mathrm{d}\mathbf{S}_{11}\\
= &\frac{\Gamma_{p_1}\left(\frac{n}{2}\right)}{\Gamma_{p_1}\left(\frac{p-p_1}{2}\right)\Gamma_{p_1}\left(\frac{n-p+p_1}{2}\right)}\frac{\pi^{p_1^2/2}}{\Gamma_{p_1}\left(\frac{p_1}{2}\right)}\frac{|\mathbf{L}|^{(p-1)/2-p_1}}{|\mathbf{I}_{p_1}+\mathbf{L}|^{n/2}}\prod_{i<j}^{p_1}(l_i-l_j)\\
\times &\frac{1}{2^{p_1n/2}\Gamma_{p_1}\left(\frac{n}{2}\right)}|\boldsymbol{\Sigma}_{11}|^{-n/2}\int\limits_{\mathbf{S}_{11}>\mathbf{O}}|\mathbf{S}_{11}|^{(n-p_1-1)/2}etr\left(-\frac{1}{2}\mathbf{S}_{11}\boldsymbol{\Sigma}_{11}^{-1}\right)\\
\times &etr\left(-\frac{1}{2}\mathbf{S}_{11}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22\cdot1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\right){}_1F_1^{(p_1)}\left(\frac{n}{2};\frac{p-p_1}{2};\frac{1}{2}\mathbf{S}_{11}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22\cdot1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1},\mathbf{L}(\mathbf{I}_{p_1}+\mathbf{L})^{-1}\right)\mathrm{d}\mathbf{S}_{11}\\
= &\frac{\Gamma_{p_1}\left(\frac{n}{2}\right)}{\Gamma_{p_1}\left(\frac{p-p_1}{2}\right)\Gamma_{p_1}\left(\frac{n-p+p_1}{2}\right)}\frac{\pi^{p_1^2/2}}{\Gamma_{p_1}\left(\frac{p_1}{2}\right)}|\mathbf{I}_{p_1}+\boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22\cdot1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1/2}|^{-n/2}\frac{|\mathbf{L}|^{(p-1)/2-p_1}}{|\mathbf{I}_{p_1}+\mathbf{L}|^{n/2}}\prod_{i<j}^{p_1}(l_i-l_j)\\
\times &{}_2F_1^{(p_1)}\left(\frac{n}{2},\frac{n}{2};\frac{p-p_1}{2};(\mathbf{I}_{p_1}+\boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22\cdot1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1/2})^{-1}\boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22\cdot1}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1/2},\mathbf{L}(\mathbf{I}_{p-p_1}+\mathbf{L})^{-1}\right)
\end{aligned}
$$

□

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 eigenvectors of $\boldsymbol{\Omega}_1(\mathbf{S}_{11})$ and of $\widetilde{\mathbf{R}} = \boldsymbol{\Sigma}_{22\cdot1}^{-1/2}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{S}_{11}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22\cdot1}^{-1/2}$, respectively, for any fixed dimension p and sample size n. Consequently, without loss of generality both matrices $\boldsymbol{\Omega}_1(\mathbf{S}_{11})$ and of $\widetilde{\mathbf{R}}$ can be taken as diagonal. These simplifies 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 $[0,\tilde{b}]$. We consider

$$
(p-p_1)\left(m_{F^{\mathbf{WT}^{-1}}}(z)-s_n(z)\right) = (p-p_1)\left(m_{F^{\mathbf{WT}^{-1}}}(z)-s_n^*(z)\right)+(p-p_1)\left(s_n^*(z)-s_n(z)\right) \quad (A.9)
$$

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

$$zs_n(z) = -1 + \int \frac{tdH_n(t)}{t - z(1 + \gamma_{2,n} zs_n(z))} \tag{A.10}$$

$$zs_n^*(z) = -1 + \int \frac{tdF_n^{\mathbf{W}}(t)}{t - z(1 + \gamma_{2,n} zs_n^*(z))} \tag{A.11}$$

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 et al. (2015), we get that the first summand $(p-p_1)\left(m_{F^{\mathbf{WT}^{-1}}}(z) - s_n^*(z)\right)$ in (A.9) conditionally on the matrix $\mathbf{W}$ converges to a Gaussian process $M_1(z)$ with the mean function

$$\mathrm{E}(M_1(z)) = \frac{\gamma_2 b^3(z)}{z^2 q^2(z)} \int \frac{tdH(t)}{(t/z - b(z))^3} = \frac{1}{2}(log(q(z))' \tag{A.12}$$

and the covariance function

$$Cov(M_1(z_1), M_1(z_2)) = 2\frac{(z_1 b(z_1))'(z_2 b(z_2))'}{(z_1 b(z_1) - z_2 b(z_2))^2} = 2\frac{\partial log((z_1 b(z_1) - z_2 b(z_2)))}{\partial z_1 \partial z_2}\,, \tag{A.13}$$

where

$$\begin{aligned} b(z) &= 1 + \gamma_2 zs(z), \\ q(z) &= 1 - \gamma_2 \int \frac{b^2(z)dH(t)}{(t/z - b(z))^2} \end{aligned} \tag{A.14}$$

for z_1 and z_2 from $\mathcal{C}$. Since all quantities in (A.12)-(A.14) do not depend on the condition $\mathbf{W}$, we get that this is also the unconditional distribution and both summands in (A.9) are independent.

Next, we derive the asymptotic distribution of the second summand $(p-p_1)\left(s_n^*(z) - s_n(z)\right)$ in (A.9). 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.10), and (A.11) we get

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

Hence,

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

where

$$\frac{b_n(z)}{1-\gamma_{2,n} z m_{F_n^{\mathbf{W}}}(zb_n^*(z)) - b_n(z)\gamma_{2,n} z^2 \int \frac{dF_n^W(t)}{(t-zb_n^*(z))(t-zb_n(z))}}$$
$$\xrightarrow{a.s.} \quad \theta_{b,H}(z) = \frac{b(z)}{1-\gamma_2 z m_H(zb(z)) - b(z)\gamma_2 z^2 \int \frac{dH(t)}{(t-zb(z))^2}} = \frac{b^2(z)}{q(z)}\,,$$

where the last equality follows from (A.14) and

$$\gamma_2 z b(z) m_H(zb(z)) = b(z) - 1\,. \tag{A.15}$$

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

$$\begin{aligned}(p-p_1)(m_{F_n^{\mathbf{W}}}(zb_n(z)) - m_{H_n}(zb_n(z))) &= (p-p_1)(m_{F_n^{\mathbf{W}}}(zb_n(z)) - m_{H_n^{\mathbf{S}_{11}}}(zb_n(z))) &(A.16)\\ &+ (p-p_1)(m_{H_n^{\mathbf{S}_{11}}}(zb_n(z)) - m_{H_n}(zb_n(z))) &(A.17)\end{aligned}$$

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

$$\begin{aligned}\frac{m_{H_n^{\mathbf{S}_{11}}}(z)}{(1+\gamma_{1,n} m_{H_n^{\mathbf{S}_{11}}}(z))} &= \int \frac{dF_n^{\widetilde{\mathbf{R}}}(t)}{t-(1+\gamma_{1,n} m_{H_n^{\mathbf{S}_{11}}}(z))\left[(1+\gamma_{1,n} m_{H_n^{\mathbf{S}_{11}}}(z))z-(1-\gamma_{1,n})\right]} &(A.18)\\ \frac{m_{H_n}(z)}{(1+\gamma_{1,n} m_{H_n}(z))} &= \int \frac{d\tilde{H}_n(t)}{t-(1+\gamma_{1,n} m_{H_n}(z))\left[(1+\gamma_{1,n} m_{H_n}(z))z-(1-\gamma_{1,n})\right]}\,, &(A.19)\end{aligned}$$

where $\widetilde{\mathbf{R}} = 1/p_1 \mathbf{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{S}_{11}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot 1}^{-1/2}$, $\tilde{H}_n(t)$ 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.17). 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.18) and (A.19) one can write

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

Rearranging terms, we get

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

where

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

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

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

and the covariance function

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

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 $\mathbf{R}=\mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22\cdot1}^{-1/2}$.

In order to derive the asymptotic distribution of the first summand in (A.16), we use the results in Yao (2013) to the conditional distribution of $(p-p_1)(m_{F_n^{\mathbf{W}}}(zb_n(z))-m_{F_n^{\mathbf{S}_{11}}}(zb_n(z)))$ 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

$$\widetilde{\mathbf{W}}=(\frac{1}{\sqrt{p_1}}\mathbf{X}+\frac{1}{\sqrt{p_1}}\mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{M})(\frac{1}{\sqrt{p_1}}\mathbf{X}+\frac{1}{\sqrt{p_1}}\mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{M})^\top.$$

with $\mathbf{M}=\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{S}_{11}^{1/2}$. Furthermore, following Lemma 1 it is enough to consider the case where $\mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{M}\mathbf{M}^T\mathbf{\Sigma}_{22\cdot1}^{-1/2}$ is diagonal and, consequently, $\mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{M}$ 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 $Cov(M_2(z_1), M_2(z_2))$ 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 $E(M_2(z))$ and $Cov(M_2(z_1), M_2(z_2))$ given by*

$$E(M_2(z)) \quad = \quad B(zb(z))$$

and the covariance function

$$Cov(M_2(z_1), M_2(z_2)) = 2\frac{\partial^2 \log(z_1b(z_1)\eta(z_1b(z_1)) - z_2b(z_2)\eta(z_2b(z_2)))}{\partial(z_1b(z_1))\partial(z_2b(z_2))}$$

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.*

Thus, merging the results for the independent asymptotic processes $M_2(z)$ and $M_3(z)$, we get

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

i.e., converges to a Gaussian process with mean and covariance functions given by

$$\theta_{b,H}(z)\left(B(zb(z)) + \theta_{\tilde{b},\tilde{H}}(zb(z))\frac{c_1^2\int \underline{m}_{\tilde{H}}^3(zb(z))t^2(c_1+t\underline{m}_{\tilde{H}}(zb(z)))^{-3}dG(t)}{(1-c_1\int \underline{m}_{\tilde{H}}^2(zb(z))t^2(c_1+t\underline{m}_{\tilde{H}}(zb(z)))^{-2}dG(t))^2}\right) \tag{A.20}$$

and

$$\begin{aligned}&2\theta_{b,H}(z_1)\theta_{b,H}(z_2)\left[\frac{\partial^2 \log(z_1b(z_1)\eta(z_1b(z_1)) - z_2b(z_2)\eta(z_2b(z_2)))}{\partial(z_1b(z_1))\partial(z_2b(z_2))}\right.\\ &+ \left.\theta_{\tilde{b},\tilde{H}}(z_1b(z_1))\theta_{\tilde{b},\tilde{H}}(z_2b(z_2))\left(\frac{\frac{\partial}{\partial(z_1b(z_1))}\underline{m}_{\tilde{H}}(z_1b(z_1))\frac{\partial}{\partial(z_1b(z_1))}\underline{m}_{\tilde{H}}(z_2b(z_2))}{(\underline{m}_{\tilde{H}}(z_1b(z_1)) - \underline{m}_{\tilde{H}}(z_2b(z_2)))^2} - \frac{1}{(z_1b(z_1)-z_2b(z_2))^2}\right)\right],\end{aligned} \tag{A.21}$$

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

$$\begin{aligned}&\theta_{b,H}(z_1)\theta_{b,H}(z_2)\frac{\partial^2 \log(z_1b(z_1)\eta(z_1b(z_1)) - z_2b(z_2)\eta(z_2b(z_2)))}{\partial(z_1b(z_1))\partial(z_2b(z_2))}\\ =\ &\frac{b^2(z_1)}{q(z_1)(z_1b(z_1))'}\frac{b^2(z_2)}{q(z_2)(z_2b(z_2))'}\frac{\partial^2 \log(z_1b(z_1)\eta(z_1b(z_1)) - z_2b(z_2)\eta(z_2b(z_2)))}{\partial z_1\partial z_2}\\ =\ &\frac{\partial^2 \log(z_1b(z_1)\eta(z_1b(z_1)) - z_2b(z_2)\eta(z_2b(z_2)))}{\partial z_1\partial z_2},\end{aligned} \tag{A.22}$$

where the last equality in (A.22) follows from (A.15) and

$$q(z)(zb(z))' = \left(1-\gamma_2(b(z)z)^2\frac{m_H'(zb(z))}{(zb(z))'}\right)(zb(z))' = (zb(z))' - (zb(z))^2(-\frac{1}{z^2} + \frac{(zb(z))'}{(zb(z))^2}) = b^2(z)\,.$$

Similarly, we get

$$
\begin{aligned}
&\theta_{b,H}(z_1)\theta_{b,H}(z_2)\left(\frac{\frac{\partial}{\partial z_1 b(z_1)}\underline{m}_{\tilde{H}}(z_1 b(z_1))\frac{\partial}{\partial z_2 b(z_2)}\underline{m}_{\tilde{H}}(z_2 b(z_2))}{(\underline{m}_{\tilde{H}}(z_1 b(z_1))-\underline{m}_{\tilde{H}}(z_2 b(z_2)))^2}-\frac{1}{(z_1 b(z_1)-z_2 b(z_2))^2}\right)\\
=\;& \frac{b^2(z_1)}{q(z_1)(z_1 b(z_1))'}\frac{b^2(z_2)}{q(z_2)(z_2 b(z_2))'}\frac{\partial^2 \log\left(\frac{\underline{m}_{\tilde{H}}(z_1 b(z_1))-\underline{m}_{\tilde{H}}(z_2 b(z_2))}{z_1 b(z_1)-z_2 b(z_2)}\right)}{\partial z_1 \partial z_2}\\
=\;& \frac{\partial^2 \log\left(\frac{\underline{m}_{\tilde{H}}(z_1 b(z_1))-\underline{m}_{\tilde{H}}(z_2 b(z_2))}{z_1 b(z_1)-z_2 b(z_2)}\right)}{\partial z_1 \partial z_2}\,. \qquad \text{(A.23)}
\end{aligned}
$$

At last, combining the results (A.12), (A.13), (A.20), (A.21) together with (A.22) and (A.23) we get that the process $(p-p_1)\left(m_{F^{\mathbf{WT}^{-1}}}(z)-s_n(z)\right)$ is asymptotically Gaussian with mean and covariance functions given by

$$
\frac{1}{2}\,\mathrm{dlog}(q(z))+\theta_{b,H}(z)\left(B(zb(z)+\theta_{\tilde{b},\tilde{H}}(zb(z))\frac{c_1^2\int \underline{m}^3_{\tilde{H}}(zb(z))t^2(c_1+t\underline{m}_{\tilde{H}}(zb(z)))^{-3}dG(t)}{(1-c_1\int \underline{m}^2_{\tilde{H}}(zb(z))t^2(c_1+t\underline{m}_{\tilde{H}}(zb(z)))^{-2}dG(t))^2})\right)
$$

and

$$
2\left[\frac{\partial^2 \log(z_1 b(z_1)\eta(z_1 b(z_1))-z_2 b(z_2)\eta(z_2 b(z_2)))}{\partial z_1 \partial z_2}+\theta_{\tilde{b},\tilde{H}}(z_1 b(z_1))\theta_{\tilde{b},\tilde{H}}(z_2 b(z_2))\left(\frac{\partial^2 \log\left(\frac{\underline{m}_{\tilde{H}}(z_1 b(z_1))-\underline{m}_{\tilde{H}}(z_2 b(z_2))}{z_1 b(z_1)-z_2 b(z_2)}\right)}{\partial z_1 \partial z_2}\right.\right.
$$

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

$$
\sum_{i=1}^{p-p_1} f(\lambda_i)-(p-p_1)\int f(x)F_n(dx)=-\frac{1}{2\pi i}\oint f(z)(p-p_1)(m_{F^{\mathbf{WT}^{-1}}}(z)-s_n(z))dz\,, \quad \text{(A.24)}
$$

where λ_i is the ith eigenvalue of the matrix $\mathbf{WT}^{-1}$ and f is an arbitrary analytic function with the support containing the interval $[0,r]$, which is the asymptotic support of the matrix $\mathbf{WT}^{-1}$. The application of (A.24) to our process together with some elementary calculus lead to the result of the theorem. □

Proof of Lemma 2

Proof. Let

$$
\begin{aligned}
T_n(z) &= \left(\frac{1}{1+\delta_n(z)}\mathbf{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{M}\mathbf{M}^\top\mathbf{\Sigma}_{22\cdot 1}^{-1/2}-z(1+\tilde{\delta}_n(z))\mathbf{I}_{p-p_1}\right)^{-1}\\
\tilde{T}_n(z) &= \left(\frac{1}{1+\tilde{\delta}_n(z)}\mathbf{M}^\top\mathbf{\Sigma}_{22\cdot 1}^{-1}\mathbf{M}-z(1+\delta_n(z))\mathbf{I}_{p_1}\right)^{-1},
\end{aligned}
$$

where $\delta_n(z)$ and $\tilde{\delta}_n(z)$ are the unique solutions of the following system of equations

$$
\delta_n(z) = \frac{1}{p_1}\mathrm{tr}\Big(T_n(z)\Big), \quad \tilde{\delta}_n(z)=\frac{1}{p_1}\mathrm{tr}\left(\tilde{T}_n(z)\right)
$$

in the class of Stieltjes transforms of non-negative measures[3] with support in $\mathbb{R}^+$.

[3]In fact, δ_n is the Stieltjes transform of a measure with total mass equal to $\frac{p-p_1}{p_1}$ while $\tilde{\delta}_n$ is the Stieltjes transform of a measure with total mass equal to 1 (see, Hachem et al. (2012))

The functions $T_n(z)$ and $\tilde{T}_n(z)$ are the deterministic approximations of the resolvents

$$Q_n(z) = \left((\mathbf{X} + \mathbf{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{M})(\mathbf{X} + \mathbf{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{M})^\top - z\mathbf{I}_{p-p_1}\right)^{-1},$$

$$\tilde{Q}_n(z) = \left((\mathbf{X} + \mathbf{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{M})^\top(\mathbf{X} + \mathbf{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{M}) - z\mathbf{I}_{p_1}\right)^{-1},$$

respectively, in the sense that

$$\frac{1}{p-p_1}\mathrm{tr}(Q_n(z) - T_n(z)) \xrightarrow{a.s.} 0 \quad \text{and} \quad \frac{1}{p_1}\mathrm{tr}(\tilde{Q}_n(z) - \tilde{T}_n(z)) \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty\,.$$

First, we find the connection between $\delta_n(z)$ and the Stieltjes transform $m_H(z)$, where H is the limiting spectral distribution of $\mathbf{W}$. For that reason, we consider the asymptotic values of $\delta_n(z)$ and $\tilde{\delta}_n(z)$ given by

$$\begin{aligned}
\delta_n(z) &= \frac{p-p_1}{p_1}(1+\delta_n(z))\frac{1}{p-p_1}\mathrm{tr}\left(\mathbf{\Sigma}_{22\cdot 1}^{-1/2}\mathbf{M}\mathbf{M}^\top\mathbf{\Sigma}_{22\cdot 1}^{-1/2} - z(1+\tilde{\delta}_n(z))(1+\delta_n(z))\mathbf{I}_{p-p_1}\right)^{-1} \\
&= (1+\delta_n(z))\frac{p-p_1}{p_1}\int\frac{d\tilde{H}_n(t)}{t - z\eta_n(z)} \longrightarrow \delta(z) = (1+\delta(z))\gamma_1 m_{\tilde{H}}(z\eta(z)) \\
\tilde{\delta}_n(z) &= \frac{1}{p_1}(1+\tilde{\delta}_n(z))\mathrm{tr}\left(\mathbf{M}^\top\mathbf{\Sigma}_{22\cdot 1}^{-1}\mathbf{M} - z(1+\tilde{\delta}_n(z))(1+\delta_n(z))\mathbf{I}_{p_1}\right)^{-1} \\
&= (1+\tilde{\delta}_n(z))\frac{1}{p_1}\int\frac{d\underline{\tilde{H}}_n(t)}{t - z\eta(z)} \longrightarrow \tilde{\delta}(z) = (1+\tilde{\delta}(z))m_{\underline{\tilde{H}}}(z\eta_n(z)) \\
&= (1+\tilde{\delta}(z))\left(-\frac{1-\gamma_1}{z\eta(z)} + \gamma_1 m_{\tilde{H}}(z\eta(z))\right)
\end{aligned}$$

with

$$\eta_n(z) = (1+\delta_n(z))(1+\tilde{\delta}_n(z)) \quad \text{and} \quad \eta(z) = (1+\delta(z))(1+\tilde{\delta}(z)).$$

Equivalently, we have

$$\frac{\delta(z)}{1+\delta(z)} = \gamma_1 m_{\tilde{H}}\big(z\eta(z)\big) \quad \text{and} \quad \frac{\tilde{\delta}(z)}{1+\tilde{\delta}(z)} = m_{\underline{\tilde{H}}}(z\eta(z)) = -\frac{1-\gamma_1}{z\eta(z)} + \gamma_1 m_{\tilde{H}}\big(z\eta(z)\big), \tag{A.25}$$

which leads to

$$\tilde{\delta}(z) = -\frac{1-\gamma_1}{z} + \delta(z)\,. \tag{A.26}$$

We claim that in fact we have

$$\delta(z) = \gamma_1 m_H(z) \tag{A.27}$$

and, consequently, $\tilde{\delta}(z) = -\frac{1-\gamma_1}{z} + \gamma_1 m_H(z)$. In order to prove (A.27), we plug $\delta(z) = \gamma_1 m_H(z)$ into (A.25) and use (A.26). It leads to

$$\begin{aligned}
\frac{m_H(z)}{1+\gamma_1 m_H(z)} &= m_{\tilde{H}}\left(z[1+\gamma_1 m_H(z)]\left[1+\gamma_1 m_H(z) - \frac{1-\gamma_1}{z}\right]\right) \\
&= m_{\tilde{H}}\Big([1+\gamma_1 m_H(z)][z(1+\gamma_1 m_H(z)) - (1-\gamma_1)]\Big),
\end{aligned}$$

which is exactly (4.1). From the uniqueness of the solution the claim (A.27) follows. Thus, in light of Theorem 2 we get as $n \to \infty$

$$\begin{aligned}
\delta_n(z) &\longrightarrow \gamma_1 m_H(z),\\
\tilde{\delta}_n(z) &\longrightarrow -\frac{1-\gamma_1}{z} + \gamma_1 m_H(z) = m_{\underline{H}}(z)\,.
\end{aligned}$$

For all $z \in \mathbb{C}^+$, we define

$$\begin{aligned}
\Delta_n(z) &= \left(1 - \frac{\frac{1}{p_1}\mathrm{tr}[T_n^2(z)\mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{M}\mathbf{M}^\top\mathbf{\Sigma}_{22\cdot1}^{-1/2}]}{(1+\delta_n(z))^2}\right)^2 - z^2\xi_n(z)\tilde{\xi}_n(z)\\
\Psi_n(z) &= \left(1 - z\xi_n(z) - \frac{\frac{1}{p_1}\mathrm{tr}\left(T_n^2(z)\mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{M}\mathbf{M}^\top\mathbf{\Sigma}_{22\cdot1}^{-1/2}\right)}{(1+\delta_n(z))^2}\right)^{-1}\\
\omega_n(z) &= \frac{1}{p_1}\sum_{j=1}^{p_1} z^2\tilde{t}_{jj}^2\\
\zeta_n(z) &= \frac{1}{p-p_1}\sum_{\substack{k=1\\k\neq l}}^{p-p_1}\sum_{l=1}^{p-p_1}(\mathbf{m}_k^\top T_n(z)\mathbf{m}_l)^2
\end{aligned}$$

with $\tilde{t}_{jj}$ being the diagonal elements of the matrix $\tilde{T}_n$ and $\mathbf{m}_k$ - the kth column of matrix $\mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{M}$, while

$$\xi_n(z_1,z_2) = \frac{1}{p_1}\mathrm{tr}(T_n(z_1)T_n(z_2)), \quad \tilde{\xi}_n(z_1,z_2) = \frac{1}{p_1}\mathrm{tr}(\tilde{T}_n(z_1)\tilde{T}_n(z_2)) \qquad \text{(A.28)}$$

and, obviously, $\xi_n(z) \equiv \xi_n(z,z)$ and $\tilde{\xi}_n(z) \equiv \tilde{\xi}_n(z,z)$.

Next, we simplify the above expressions. In using (A.25), we get

$$\begin{aligned}
\xi_n(z_1,z_2) &= \frac{p-p_1}{p_1}\frac{(1+\delta_n(z_1))(1+\delta_n(z_2))}{p-p_1}\\
&\times \mathrm{tr}\left(\left[\mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{M}\mathbf{M}^\top\mathbf{\Sigma}_{22\cdot1}^{-1/2} - z_1\eta_n(z_1)\mathbf{I}_{p-p_1}\right]^{-1}\left[\mathbf{\Sigma}_{22\cdot1}^{-1/2}\mathbf{M}\mathbf{M}^\top\mathbf{\Sigma}_{22\cdot1}^{-1/2} - z_2\eta_n(z_2)\mathbf{I}_{p-p_1}\right]^{-1}\right)\\
&\longrightarrow \xi(z_1,z_2) = \gamma_1(1+\delta(z_1))(1+\delta(z_2))\int\frac{d\tilde{H}(t)}{(t-z_1\eta(z_1))(t-z_2\eta(z_2))}\\
&= \gamma_1(1+\delta(z_1))(1+\delta(z_2))\frac{m_{\tilde{H}}(z_1\eta(z_1)) - m_{\tilde{H}}(z_2\eta(z_2))}{z_1\eta(z_1)-z_2\eta(z_2)}\\
&= \frac{\delta(z_1)-\delta(z_2)}{z_1\eta(z_1)-z_2\eta(z_2)}\,.
\end{aligned}$$

In the case of $z_1 = z_2 = z$, we obtain

$$\begin{aligned}
\xi_n(z) &\longrightarrow \xi(z) = \gamma_1(1+\delta(z))^2\int\frac{d\tilde{H}(t)}{(t-z\eta(z))^2}\\
&= \gamma_1\frac{m'_{\tilde{H}}(z\eta(z))}{z\eta'(z)+\eta(z)}(1+\delta(z))^2 = \frac{\delta'(z)}{(z\eta(z))'}\,.
\end{aligned}$$

Similarly, using (A.25) we get for $\tilde{\xi}_n(z_1, z_2)$, i.e.,

$$\begin{aligned}
\tilde{\xi}_n(z_1, z_2) \quad \longrightarrow \quad & \tilde{\xi}(z_1, z_2) = \frac{\tilde{\delta}(z_1) - \tilde{\delta}(z_2)}{z_2\eta(z_2) - z_1\eta(z_1)} = (1-\gamma_1)\frac{z_1 - z_2}{(z_1\eta(z_1) - z_2\eta(z_2))z_1z_2} + \frac{\delta(z_1) - \delta(z_2)}{z_1\eta(z_1) - z_2\eta(z_2)} \\
= \quad & \frac{(1-\gamma_1)}{z_1z_2}\frac{z_1 - z_2}{z_1\eta(z_1) - z_2\eta(z_2)} + \xi(z_1, z_2) \\
\tilde{\xi}_n(z) \quad \longrightarrow \quad & \tilde{\xi}(z) = \frac{\left(\frac{1-\gamma_1}{z^2} + \delta'(z)\right)}{(z\eta(z))'} = \frac{(1-\gamma_1)}{z^2(z\eta(z))'} + \xi(z)\,.
\end{aligned}$$

In using these results as well as

$$\begin{aligned}
m'_{\tilde{H}}(z\eta(z)) \quad &= \quad \frac{\partial}{\partial z}\int \frac{d\tilde{H}(t)}{(t - z\eta(z))} = \int \frac{d\tilde{H}(t)}{(t - z\eta(z))^2}(z\eta(z))', && \text{(A.29)} \\
\gamma_1 m'_{\tilde{H}}(z\eta(z)) \quad &= \quad \frac{\delta'(z)}{(1+\delta)^2} && \text{(A.30)}
\end{aligned}$$

and applying

$$\begin{aligned}
\gamma_1\int \frac{d\tilde{H}(t)}{(t - z\eta(z))^2} \quad &= \quad \frac{\delta'(z)}{(1+\delta(z))^2}\frac{1}{(z\eta(z))'} = \frac{\xi(z)}{(1+\delta(z))^2}, \\
\gamma_1\int \frac{td\tilde{H}(t)}{(t - z\eta(z))^2} \quad &= \quad \gamma_1\int \frac{\tilde{H}(t)}{(t - z\eta(z))} + \gamma_1 z\eta(z)\int \frac{\tilde{H}(t)}{(t - z\eta(z))^2} = \frac{\delta(z)}{1+\delta(z)} + \frac{\xi(z)}{(1+\delta(z))^2}z\eta(z)\,,
\end{aligned}$$

we get

$$\begin{aligned}
\Delta_n(z) \quad \longrightarrow \quad \Delta(z) &= \left(1 - \gamma_1\int \frac{t\tilde{H}(t)}{(t - z\eta(z))^2}\right)^2 - z^2\xi(z)\tilde{\xi}(z) \\
&= \quad \left(\frac{1}{1+\delta(z)} - \frac{z\eta(z)}{(1+\delta(z))^2}\xi(z)\right)^2 - z^2\xi(z)\tilde{\xi}(z) \\
&= \quad \left(\frac{1}{1+\delta(z)} - z\xi(z) + \frac{1-\gamma_1}{1+\delta(z)}\xi(z)\right)^2 - z^2\xi^2(z) - \frac{1-\gamma_1}{(z\eta(z)')}\xi(z) \qquad \text{(A.31)}
\end{aligned}$$

Moreover, the term $(z\eta(z))'$ can be rewritten further as follows

$$\begin{aligned}
(z\eta(z))' \quad &= \quad (z(1+\delta(z))(1+\tilde{\delta}(z)))' = \left(z(1+\delta(z))\left(1 + \delta(z) - \frac{1-\gamma_1}{z}\right)\right)' \\
&= \quad (z(1+\delta(z))^2 - (1-\gamma_1)(1+\delta(z)))' = (1+\delta(z))^2 + 2(1+\delta(z))\delta'(z)z - (1-\gamma_1)\delta'(z) \\
&= \quad (1+\delta(z))^2 + 2(z\eta(z))'(1+\delta(z))\xi(z)z - (1-\gamma_1)(z\eta(z))'\xi(z)\,,
\end{aligned}$$

which yields to

$$\frac{1}{(z\eta(z))'} = \frac{1}{1+\delta(z)}\left(\frac{1}{1+\delta(z)} - 2\xi(z)z + \frac{1-\gamma_1}{1+\delta(z)}\xi(z)\right)$$

Similarly,

$$\begin{aligned}
\Psi_n^{-1}(z) \longrightarrow \Psi^{-1}(z) \quad &= \quad \frac{1}{1+\delta(z)} - \frac{z\eta(z)}{(1+\delta(z))^2}\xi(z) - z\xi(z) \\
&= \quad \frac{1}{1+\delta(z)} - 2\xi(z)z + \frac{1-\gamma_1}{1+\delta(z)}\xi(z)\,, && \text{(A.32)}
\end{aligned}$$

which is exactly equal to $(1+\delta(z))/(z\eta(z))'$. Now, (A.32) and (A.31) lead to

$$\begin{aligned}\Delta(z) &= \left(\frac{1}{1+\delta(z)}+\frac{1-\gamma_1}{1+\delta(z)}\xi(z)\right)\left(\frac{1}{1+\delta(z)}-2\xi(z)z+\frac{1-\gamma_1}{1+\delta(z)}\xi(z)\right)-\frac{1-\gamma_1}{(z\eta(z))'}\xi(z)\\ &= \left(\frac{1}{1+\delta(z)}+\frac{1-\gamma_1}{1+\delta(z)}\xi(z)\right)\Psi^{-1}(z)-\frac{1-\gamma_1}{1+\delta(z)}\xi(z)\Psi^{-1}(z)\\ &= \frac{1}{1+\delta(z)}\Psi^{-1}(z) \qquad \text{(A.33)}\end{aligned}$$

From Lemma 1 we get that the matrices $T_n(z)$ and $\tilde{T}_n(z)$ could be chosen without loss of generality as diagonal matrices, which implies

$$\begin{aligned}\omega_n(z) &= z^2\frac{1}{p_1}\sum_{j=1}^{p_1}\tilde{t}_{jj}^2 = z^2\mathrm{tr}(\tilde{T}_n^2(z)) \to z^2\tilde{\delta}^2(z),\\ \zeta_n(z) &= \frac{1}{p-p_1}\sum_{\substack{k=1\\k\neq l}}^{p-p_1}\sum_{l=1}^{p-p_1}(\mathbf{m}_k^\top T_n(z)\mathbf{m}_l)^2 = 0\,.\end{aligned}$$

Now, Theorems 2.2.1 and 2.2.2 by Yao (2013) reveal that $M_{2,n}(z) = (p-p_1)(m_F\mathbf{w}(z) - m_{H_n}(z))$ converges to a Gaussian process $M_2(z)$ with mean function and covariance function given by

$$\begin{aligned}\mathrm{E}(M_2(z)) &= \frac{\Psi_n(z)}{\Delta_n(z)}\left(z^2\tilde{\xi}_n(z)\frac{1}{p_1}\mathrm{tr}\big(T_n^3(z)\big)+\zeta_n(z)\frac{1}{p_1}\mathrm{tr}\big(T_n^3(z)\big)\right.\\ &+ 2\frac{\frac{1}{p_1}\mathrm{tr}\big(\boldsymbol{\Sigma}_{22\cdot1}^{-1/2}\mathbf{M}\mathbf{M}^\top\boldsymbol{\Sigma}_{22\cdot1}^{-1/2}T_n^3(z)\big)}{(1+\delta_n(z))^2}\left(1-\frac{\frac{1}{p_1}\mathrm{tr}\left(\boldsymbol{\Sigma}_{22\cdot1}^{-1/2}\mathbf{M}\mathbf{M}^\top\boldsymbol{\Sigma}_{22\cdot1}^{-1/2}T_n^2(z)\right)}{(1+\delta_n(z))^2}\right)\\ &+ \left.\frac{\omega_n(z)}{(1+\delta_n(z))^2}\left(\frac{1}{p_1}\mathrm{tr}\big(\boldsymbol{\Sigma}_{22\cdot1}^{-1/2}\mathbf{M}\mathbf{M}^\top\boldsymbol{\Sigma}_{22\cdot1}^{-1/2}T_n^3(z)\big)\xi_n(z)-\frac{1}{p_1^2}\mathrm{tr}\big(\boldsymbol{\Sigma}_{22\cdot1}^{-1/2}\mathbf{M}\mathbf{M}^\top\boldsymbol{\Sigma}_{22\cdot1}^{-1/2}T_n^2(z)\big)\mathrm{tr}\big(T_n^3(z)\big)\right)\right)\\ \mathrm{Cov}(M_2(z_1),M_2(z_2)) &= 2\frac{(z_1\eta_n(z_1))'(z_2\eta_n(z_2))'}{(z_1\eta_n(z_1)-z_2\eta_n(z_2))^2}\,.\end{aligned}$$

Since $\eta_n(z)\to\eta(z)$, we get that

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

Furthermore,

$$\begin{aligned}\mathrm{E}(M_2(z)) \longrightarrow B(z) &= \frac{\Psi(z)}{\Delta(z)}\left(z^2\tilde{\xi}(z)\gamma_1\int\frac{d\tilde{H}(t)}{(t-z\eta(z))^3}(1+\delta(z))^3+2\gamma_1\frac{\int\frac{td\tilde{H}(t)}{(t-z\eta(z))^3}(1+\delta(z))^3}{(1+\delta(z))^2}\left[\Psi^{-1}(z)+z\xi(z)\right]\right.\\ &+ \left.\frac{z^2\tilde{\delta}^2(z)}{(1+\delta(z))^2}\left(\gamma_1\int\frac{td\tilde{H}(t)}{(t-z\eta(z))^3}(1+\delta(z))^3\xi_n(z)-\gamma_1^2\int\frac{td\tilde{H}(t)}{(t-z\eta(z))^2}\int\frac{d\tilde{H}(t)}{(t-z\eta(z))^3}(1+\delta(z))^5\right)\right),\end{aligned}$$

where from (A.33) it follows that

$$\frac{\Psi(z)}{\Delta(z)} = (1+\delta(z))\Psi^2(z)\,.$$

Because of (A.29), (A.30) and

$$\begin{aligned}
\frac{\partial^2}{\partial z^2}\int \frac{td\tilde{H}(t)}{(t-z\eta(z))} &= 2\int \frac{d\tilde{H}(t)}{(t-z\eta(z))^3}\left[(z\eta(z))^{\top}\right]^2 + \int \frac{d\tilde{H}(t)}{(t-z\eta(z))^2}(z\eta(z))'' \\
&= 2\int \frac{d\tilde{H}(t)}{(t-z\eta(z))^3}\left[(z\eta(z))'\right]^2 + m'_{\tilde{H}}(z\eta(z))\frac{(z\eta(z))''}{(z\eta(z))'} \\
&= 2\int \frac{d\tilde{H}(t)}{(t-z\eta(z))^3}\left[(z\eta(z))'\right]^2 + \gamma^{-1}\frac{\delta'(z)}{(1+\delta(z))^2}\frac{(z\eta(z))''}{(z\eta(z))'}, \\
\gamma_1 m''_{\tilde{H}}(z\eta(z)) &= \frac{\delta''(z)}{(1+\delta(z)^2} - 2\frac{\delta'^{\,2}(z)}{(1+\delta(z))^3},
\end{aligned}$$

we obtain

$$\gamma_1\int \frac{d\tilde{H}(t)}{(t-z\eta(z))^3} = \frac{\frac{\delta''(z)}{(1+\delta(z))^2} - 2\frac{\delta'^{\,2}(z)}{(1+\delta(z))^3} - \frac{\delta'(z)}{(1+\delta(z))^2}\frac{(z\eta(z))''}{(z\eta(z))'}}{2(z\eta(z))'^{\,2}}.$$

On the other hand, it holds that

$$\xi'(z) = \frac{\delta''(z)}{(z\eta(z))'} - \delta'(z)\frac{(z\eta(z))''}{(z\eta(z))'^{\,2}} = \frac{\delta''(z)}{(z\eta(z))'} - \xi(z)\frac{(z\eta(z))''}{(z\eta(z))'}.$$

Thus, we have

$$\begin{aligned}
\gamma_1\int \frac{d\tilde{H}(t)}{(t-z\eta(z))^3} &= \frac{\xi'(z)}{2(1+\delta(z))^2(z\eta(z))'} - \frac{\xi^2(z)}{(1+\delta(z))^3} \\
&= \frac{1}{(1+\delta(z))^3}\left(\frac{\xi'(z)\Psi^{-1}(z)}{2} - \xi^2(z)\right).
\end{aligned}$$

and

$$\begin{aligned}
\gamma_1\int \frac{td\tilde{H}(t)}{(t-z\eta(z))^3} &= \gamma_1\int \frac{d\tilde{H}(t)}{(t-z\eta(z))^2} + \gamma_1 z\eta(z)\int \frac{d\tilde{H}(t)}{(t-z\eta(z))^3} \\
&= \frac{1}{(1+\delta(z))^2}\left(\xi(z) + \left(z - \frac{(1-\gamma_1)}{1+\delta(z)}\right)\left(\frac{\xi'(z)\Psi^{-1}(z)}{2} - \xi^2(z)\right)\right).
\end{aligned}$$

As a result, it holds that

$$
\begin{aligned}
B(z) &= (1+\delta(z))\Psi^2(z)\left(\left(z^2\xi(z)+\frac{1-\gamma_1}{1+\delta(z)}\Psi^{-1}(z)\right)\left(\frac{\xi^\top(z)\Psi^{-1}(z)}{2}-\xi^2(z)\right)\right.\\
&+ \frac{2}{1+\delta(z)}\left(\xi(z)+\left(z-\frac{1-\gamma_1}{1+\delta(z)}\right)\left(\frac{\xi'(z)\Psi^{-1}(z)}{2}-\xi^2(z)\right)\right)\left(\Psi^{-1}(z)+z\xi(z)\right)\\
&+ \frac{\omega(z)}{(1+\delta(z))}\left(\xi(z)\left(\xi(z)+\left(z-\frac{1-\gamma_1}{1+\delta(z)}\right)\left(\frac{\xi'(z)\Psi^{-1}(z)}{2}-\xi^2(z)\right)\right)\right.\\
&- \left.\left.(\delta(z)+\xi(z)(z(1+\delta(z))-(1-\gamma_1)))\left(\frac{\xi'(z)\Psi^{-1}(z)}{2}-\xi^2(z)\right)\right)\right)\\
&= (1+\delta(z))\Psi^2(z)\left[\widetilde{\omega}(z)N(z)+\frac{2}{1+\delta(z)}\left(\xi(z)+\left(z-\frac{1-\gamma_1}{1+\delta(z)}\right)N(z)\right)\left[\Psi^{-1}(z)+z\xi(z)\right]\right.\\
&+ \left.\frac{\omega(z)}{1+\delta(z)}\left(\xi^2(z)-\delta(z)N(z)\left(z-\frac{1-\gamma_1}{1+\delta(z)}+1\right)\right)\right]\\
&= (1+\delta(z))\Psi^2(z)\left[\widetilde{\omega}(z)N(z)+\frac{2}{1+\delta(z)}\left(-\widetilde{\omega}(z)N(z)+\xi(z)(\Psi^{-1}(z)+z\xi(z))\right.\right.\\
&+ \left.z\underbrace{\left(2z\xi(z)-\frac{1-\gamma_1}{1+\delta(z)}\xi(z)-\frac{1}{1+\delta(z)}\right)}_{-\Psi^{-1}(z)}N(z)+\frac{1}{1+\delta(z)}N(z)+z\Psi^{-1}(z)N(z)\right)\\
&+ \left.\frac{\omega(z)}{1+\delta(z)}\left(\xi^2(z)-\delta(z)N(z)\left(z-\frac{1-\gamma_1}{1+\delta(z)}+1\right)\right)\right]\\
&= (1+\delta(z))\Psi^2(z)\left[\widetilde{\omega}(z)N(z)\frac{\delta(z)-1}{1+\delta(z)}\right.\\
&+ \frac{1}{1+\delta(z)}\left(\frac{1}{(1+\delta(z))}N(z)+\xi(z)\left(\Psi^{-1}(z)+z\xi(z)\right)\right)\\
&+ \left.\frac{\omega(z)}{1+\delta(z)}\left(\xi^2(z)-\delta(z)N(z)\left(z-\frac{1-\gamma_1}{1+\delta(z)}+1\right)\right)\right]
\end{aligned}
$$

with

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

$\square$