Likelihood ratio tests under model misspecification in high dimensions

Nina Dörnemann
Department of Mathematics
Ruhr University Bochum
44801 Bochum, Germany
nina.doernemann@rub.de

March 11, 2022

Abstract

We investigate the likelihood ratio test for a large block-diagonal covariance matrix with an increasing number of blocks under the null hypothesis. While so far the likelihood ratio statistic has only been studied for normal populations, we establish that its asymptotic behavior is invariant under a much larger class of distributions. This implies robustness against model misspecification, which is common in high-dimensional regimes. Demonstrating the flexibility of our approach, we additionally establish asymptotic normality of the log-likelihood ratio test for the equality of many large sample covariance matrices under model uncertainty. A simulation study emphasizes the usefulness of our findings.

Keywords: block-diagonal covariance matrix, central limit theorem, equality of covariance matrices, high-dimensional inference, likelihood ratio test, model misspecification, non-normal population
AMS subject classification: 62H15, 62H10

1 Introduction

Over the last decades, the availability of high-dimensional data sets across diverse disciplines such as as bio statistics, wireless communications and finance has transformed statistical practice (see, e.g., Fan and Li (2006); Johnstone (2006) and references therein). Traditional multivariate analysis, as outlined in the text books of Anderson (1984) or Muirhead (1982), is developed under the paradigm that the dimension is negligible compared to the sample size and breaks down seriously if this assumption is violated. Such problems have spurred the development of new analysis tools, that work for dimensions of the same order as and even larger than the sample size. The literature on these topics is so large, that we can only cite a few illustrative examples, related to the present work: The works of Yamada et al. (2017) and Bodnar et al. (2019) address the question whether a large covariance matrix admits a block-diagonal structure. Tests for independence in various setting are discussed in Han et al. (2017) and Loubaton and Rosuel (2020). Hu et al. (2017) concentrate on tests for the equality of high-dimensional mean vectors, while He et al. (2021) take a broader perspective on high-dimensional testing by investigating a class of U-statistics.

Turning closer to the scope of this work, the likelihood ratio method has received much attention in the literature on high-dimensional statistical inference since the past decade. The starting point for the investigating of various classical testing problems transferred to a high-dimensional setting can be seen in the work of Jiang and Yang (2013) establishing CLTs for the corresponding log-likelihood ratio tests, including the two main testing problems investigated in this work. Jiang and Qi (2015) tried to relax the assumptions on the parameters, while other authors extended these results in various directions. For example, Jiang and Wang (2017) proved a moderate deviation principle for these likelihood ratio tests. More recent generalizations include the works of Qi et al. (2019); Dette and Dörnemann (2020); Guo and Qi (2021). All of these works rely on normally distributed data and the asymptotic behavior of these test statistics under model misspecification has received little attention in the literature on high-dimensional statistics so far. A few works investigating likelihood ratio tests in different settings under model uncertainty include Luo and Tsai (2012); Lemonte (2013, 2016); Strug (2018); Ishii et al. (2021). We add to this line of literature by dropping the restrictive distributional assumption of normality. In particular, we find that the CLTs for the log-likelihood of two specific testing problems remain still valid when only assuming moments of order $(4 + \delta)$ for some $\delta > 0$. Besides the theoretical importance of our findings, these results ensure more robust statistical guarantees for practitioners as the validity of the normal assumption is not a priori clear for high-dimensional data sets.

Interestingly, our results reveal that the limiting distribution of the log-likelihood under the null hypothesis does not depend on specific characteristics of the underlying data generating distribution such as the fourth moment. This observation is illustrated in Figure 1 where we consider the problem of testing whether the covariance matrix of a p-dimensional random vector admits a block-diagonal structure with q blocks. Here, we display three histograms for the corresponding log-likelihood ratio test under the null hypothesis based on three different distributions for the samples. The components of all vectors are independent identically distributed with respect to a standard normal distribution (left column), standardized t-distribution (middle column) and centered exponential distribution (right column), respectively. Thus, the null hypothesis of a block-diagonal covariance matrix is obviously satisfied (in this case, the covariance matrix equals the identity matrix). We observe that the histograms look very similar. This testing problem will be examined in detail in Section 2 of this work.

Before concluding this introduction, we would like to discuss the main ideas for our proofs. Under the normal assumption, the exact distribution of the test statistic is available under the null hypothesis on which proofs of previous works crucially depend on. Equipped with such a knowledge, the moment-generating of the log-likelihood test statistic is investigated (see, e.g., Jiang and Yang, 2013; Qi et al., 2019; Guo and Qi, 2021) or a general central limit theorem is applied (see Dette and Dörnemann, 2020). Obviously we cannot hope for an analogue exact representation without knowing the underlying distribution of the data.

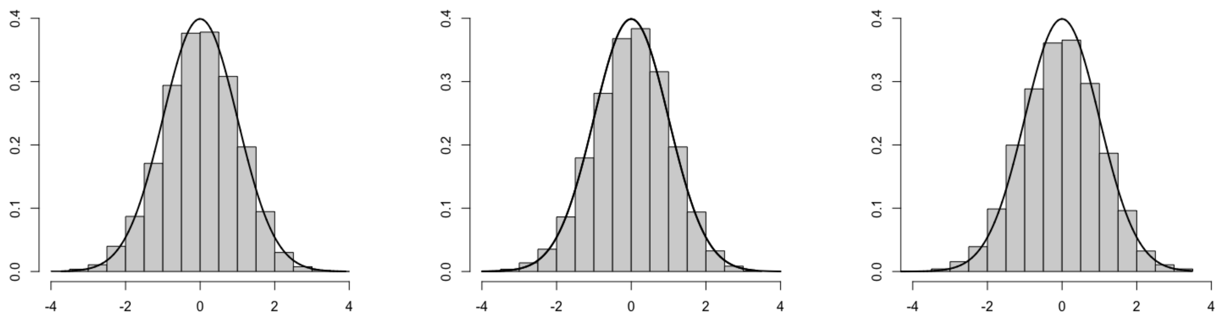

Figure 1: Histograms for the log-likelihood ratio test (3) under the hypothesis (2) of a block diagonal structure of $q = 30$ blocks of equal size 2 in a $p = 60$-dimensional random vector (sample size 100, simulation runs $10,000$). Left column: standard normally distributed data, middle column: standardized t-distributed data with 15 degrees of freedom, right column: centered exponentially distributed data with parameter 1. The grey curve indicates the density of the standard normal distribution.

In order to tackle the difficulties arising in the proof for non-normal populations, we derive a novel representation of the log-likelihood test statistic involving random quadratic forms without imposing restrictive distributional assumptions. For this purpose, we perform a QR-decomposition for the (sub)data matrices. Such QR-decompositions are useful in a broader context: Wang et al. (2018) used this tool in order to derive the logarithmic law of the sample covariance matrix for the case $p/n \to 1$ near singularity, while Heiny and Parolya (2021) investigated the log-determinant of the sample correlation matrix under infinite fourth moment. These papers were partially inspired by works of Nguyen and Vu (2014) and Bao et al. (2015), in which the authors proved Girko's logarithmic law for a general random matrix with independent entries and brought his "method of perpendiculars" (see Girko, 1998) on a mathematically rigorous level. Via our representation, we are in the position to decompose the test statistic into three parts: we will prove that the dominating linear term satisfies a central limit theorem for martingale difference schemes, while the quadratic term converges to constant and the remainders are asymptotically negligible. Heuristically, this decomposition can be motivated by Taylor's expansion $\log(1+x) = x - x^2/2 + \mathcal{O}(x^3)$, though one needs more delicate arguments in order to justify this step mathematically correct.

This work is structured as follows. In Section 2, we present a CLT for the log-likelihood ratio test of a block-diagonal covariance matrix under the null hypothesis. Here, the number of blocks may increase together with dimension of the data and sample size while we do not assume that the data is generated by a normal distribution. As a corollary, the distribution of a test for a diagonal covariance matrix is derived. In Section 3, we apply our method to another classical likelihood ratio test and provide the asymptotic distribution of the log-likelihood ratio test on the equality of many large covariance matrices. The main results of Section 2 and 3 are proven in Section 5. We illustrate our findings with a simulation study in Section 4.

2 Testing for a block-diagonal covariance matrix

In the main part of this work, we revisit a very prominent problem in high-dimensional data analysis, namely a test for uncorrelation of sub-vectors of a multivariate distribution. For normally distributed data, this coincides with a test for independence of these sub-vectors. To be precise, let $\mathbf{y} = \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{x}$ denote a p-dimensional random vector with mean $\boldsymbol{\mu} = \mathbf{0} \in \mathbb{R}^p$ and covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{p\times p}$, where $\mathbf{\Sigma}^{\frac{1}{2}}$ denotes the symmetric square root of $\mathbf{\Sigma}$. We assume that $\mathbf{y}$ is decomposed as

$$\mathbf{y} = \left(\mathbf{y}^{(1)^\top}, \ldots, \mathbf{y}^{(q)^\top}\right)^\top,$$

where $\mathbf{y}^{(i)}$ are vectors of dimension $p_i \in \mathbb{N}$, $1 \leq i \leq q$, such that $\sum_{i=1}^q p_i = p$ for some integer $q \geq 2$. Moreover, we assume that the components of $\mathbf{x}$ are i.i.d. with respect to some centered distribution. Let

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} & \ldots & \mathbf{\Sigma}_{1q} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} & \ldots & \mathbf{\Sigma}_{2q} \\ \vdots & \vdots & & \vdots \\ \mathbf{\Sigma}_{q1} & \mathbf{\Sigma}_{q2} & \ldots & \mathbf{\Sigma}_{qq} \end{pmatrix} \tag{1}$$

denote the corresponding decomposition of the covariance matrix, where $\mathbf{\Sigma}_{ij} := \operatorname{Cov}(\mathbf{y}^{(i)}, \mathbf{y}^{(j)})$. The hypothesis of uncorrelated sub-vectors is formulated as

$$H_0: \ \mathbf{\Sigma}_{ij} = \mathbf{0} \ \text{ for all } i \neq j. \tag{2}$$

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{\text{i.i.d.}}{\sim} \mathbf{x}$ be a sample of independent identically distributed random variables according to $\mathbf{x}$ and denote $\mathbf{y}_i = \mathbf{\Sigma}^{1/2}\mathbf{x}_i$ for $1 \leq i \leq n$. Under the normal assumption $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, the likelihood ratio test is given by

$$\Lambda_n = \frac{|\hat{\mathbf{\Sigma}}|^{n/2}}{\prod_{i=1}^q |\hat{\mathbf{\Sigma}}_{ii}|^{n/2}} = V_n^{\frac{n}{2}}, \tag{3}$$

where

$$\hat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{k=1}^n \mathbf{y}_k \mathbf{y}_k^\top$$

denotes the sample covariance matrix of $\mathbf{y}_1, \ldots, \mathbf{y}_n$ and and $\hat{\mathbf{\Sigma}}_{ij}$ denotes the block in the ith row and jth column of the estimate $\hat{\mathbf{\Sigma}}$ corresponding to the decomposition (1).

In the case $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$, Jiang and Yang (2013) and Jiang and Qi (2015) derived a central limit theorem for the corresponding log-likelihood ratio test statistic in a high-dimensional setting when assuming that the number q of blocks is fixed. Several authors such as Qi et al. (2019) and Dette and Dörnemann (2020) demonstrated that such a CLT still holds true if the parameter $q = q_n$ is allowed to increase with sample size and dimension. All of these works rely on normally distributed data. Dropping the normal assumption, the following theorem provides the asymptotic distribution of $\log \Lambda_n$ under the null hypothesis of uncorrelation without assuming a normal distribution for $\mathbf{x}_1, \ldots, \mathbf{x}_n \sim \mathbf{x}$. The proof is deferred to Section 5.1.

Theorem 2.1. *Let the components of* $\mathbf{x}$ *be i.i.d. centered random variables following a continuous distribution with finite* $(4+\delta)$*th moment for some* $\delta > 0$*. Assume that* $q = q_n \geq 2$ *is an possibly increasing integer,* $2 \leq p = p_n < n$ *with* $0 < \inf_{n\in\mathbb{N}} \min_{1\leq i\leq q}(p_i q)/n \leq \sup_{n\in\mathbb{N}} p/n < 1$ *and* $\max_{1\leq i\leq q} p_i \leq \eta p$ *for each* $n \in \mathbb{N}$ *and some fixed* $\eta \in (0,1)$*. Then, it holds under the null hypothesis in* (2) *that*

$$\frac{\log V_n - \mu_n}{\sigma_n} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

where

$$\mu_n = \sum_{i=1}^{q} \left(n - p_i - \frac{1}{2}\right) \log\left(1 - \frac{p_i}{n}\right) - \left(n - p - \frac{1}{2}\right) \log\left(1 - \frac{p}{n}\right),$$
$$\sigma_n^2 = 2\left\{\sum_{i=1}^{q} \log\left(1 - \frac{p_i}{n}\right) - \log\left(1 - \frac{p}{n}\right)\right\}.$$

Remark 2.1. *Choose* $\alpha \in (0,1)$*. We propose to reject the null hypothesis whenever*

$$\log V_n \leq \sigma_n u_\alpha + \mu_n \ , \tag{4}$$

where u_α *denotes the* α*-quantile of the standard normal distribution. Thus, we have by Theorem 2.1*

$$\lim_{n\to\infty} \mathbb{P}_{H_0}\left(\log V_n \leq \sigma_n u_\alpha + \mu_n\right) = \mathbb{P}(\mathcal{N}(0,1) \leq u_\alpha) = \alpha,$$

which means that the test keeps asymptotically its nominal level α*.*

As a noteworthy-by-product of Theorem 2.1, we are able to construct a test for a diagonal covariance matrix based on the sample correlation matrix.

Remark 2.2 (Testing for a diagonal covariance matrix)**.** 1. *As an application, we consider the special case of testing for a diagonal covariance matrix which coincides with complete independence of the* p *components of* $\mathbf{x}$ *in the normal case. In this case, the test in* (2) *is equivalent to*

$$H_0 : \mathbf{R} = \mathbf{I}_p, \tag{5}$$

where

$$\mathbf{R} = \operatorname{diag}(\mathbf{\Sigma})^{-\frac{1}{2}} \mathbf{\Sigma} \operatorname{diag}(\mathbf{\Sigma})^{-\frac{1}{2}}$$

denotes the population correlation matrix of $\mathbf{y}$*. Then, the statistic* V_n *defined in* (3) *can be written as the determinant of the sample correlation matrix, that is,*

$$V_n = |\hat{\mathbf{R}}|,$$

where

$$\hat{\mathbf{R}} = \operatorname{diag}(\hat{\mathbf{\Sigma}})^{-\frac{1}{2}} \hat{\mathbf{\Sigma}} \operatorname{diag}(\hat{\mathbf{\Sigma}})^{-\frac{1}{2}}$$

denotes the sample correlation matrix of $\mathbf{y}_1, \ldots, \mathbf{y}_n$*. Several authors investigated tests for the hypothesis given in* (5) *in different frameworks (e.g., see Jiang and Yang (2013), Jiang and Qi (2015), Gao et al. (2017), Mestre and Vallet (2017), Qi et al. (2019), Parolya et al. (2021), Heiny and Parolya (2021)). We observe that testing for* (5) *is a special case of testing for* (2) *by letting* $q = p$ *and* $p_1 = \ldots = p_q = 1$*. Then, Theorem 2.1 gives us the following result.*

Corollary 2.1. *Let the components of* $\mathbf{x}$ *be i.i.d. centered random variables following a continuous distribution with finite* $(4+\delta)$*th moment for some* $\delta > 0$*. Assume that* $2 \leq p = p_n < n$ *and* $0 < \inf_{n\in\mathbb{N}} p/n \leq \sup_{n\in\mathbb{N}} p/n < 1$*. Then, it holds under the null hypothesis in* (5) *that*

$$\frac{\log|\mathbf{R}| - \overline{\mu}_n}{\overline{\sigma}_n} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

where

$$\overline{\mu}_n = p\left(n - \frac{3}{2}\right)\log\left(1 - \frac{1}{n}\right) - \left(n - p - \frac{1}{2}\right)\log\left(1 - \frac{p}{n}\right),$$
$$\overline{\sigma}_n^2 = 2\left\{p\log\left(1 - \frac{1}{n}\right) - \log\left(1 - \frac{p}{n}\right)\right\}.$$

2. *Note that Parolya et al. (2021) investigated the log-determinant of the sample correlation matrix in a more general context. In fact, they cover the case* $\mathbf{R} \neq \mathbf{I}$*, while Corollary 2.1 is formulated under the null hypothesis* $\mathbf{R} = \mathbf{I}$*. If we assume that* $p/n \to \gamma \in (0,1)$*, then Corollary 2.1 yields a special version of their Theorem 2.1, since*

$$\overline{\mu}_n = -\left(n - p - \frac{1}{2}\right)\log\left(1 - \frac{p-1}{n}\right) - (p-1) + \frac{p}{n} + o(1),$$
$$\overline{\sigma}_n^2 = -2\left\{\frac{p}{n} - \log\left(1 - \frac{p-1}{n}\right)\right\} + o(1),$$

which coincides with mean and variance given in their Theorem 2.1 in the case $\mathbf{R} = \mathbf{I}$*. In a follow-up work, Heiny and Parolya (2021) showed that the CLT for the sample correlation matrix in the case* $\mathbf{R} = \mathbf{I}$ *still holds true under infinite fourth moment.*

3 Testing for equality of covariance matrices

We expect that our method for proving a CLT as given in Theorem 2.1 can be adapted to the investigation of other classical likelihood ratio tests in a non-normal setting. In order to demonstrate this adaption, we consider in this section the comparison of q centered distributions with covariance matrices $\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_q \in \mathbb{R}^{p\times p}$ and generic elements $\mathbf{y}_1 = \mathbf{\Sigma}_1^{1/2}\mathbf{x}, \ldots, \mathbf{y}_q = \mathbf{\Sigma}_q^{1/2}\mathbf{x}$. We assume that for each group j a sample of size n_j is available, $j \in \{1, \ldots, q\}$. When considering asymptotics, the dimension p and the number q of groups increase with the (sub)sample sizes. As before, we assume that the components of $\mathbf{x}$ are i.i.d. with respect to some centered distribution.

An important assumption for multivariate analysis of variance (MANOVA) is that of equal covariances in the different groups, which motivates our interest in testing the hypothesis

$$H_0 : \mathbf{\Sigma}_1 = \ldots = \mathbf{\Sigma}_q. \tag{6}$$

This problem has been considered by several authors in the context of high-dimensional inference (see O'Brien (1992), Schott (2007), Srivastava and Yanagihara (2010), Jiang and Yang (2013), Jiang and Qi (2015), Dette and Dörnemann (2020) and Guo and Qi (2021) among others). In this section, we add to this line of literature and investigate the asymptotic distribution of the likelihood ratio test based on samples of independent distributed observations $\mathbf{y}_{ji} \overset{\text{i.i.d.}}{\sim} \mathbf{y}_j$, $1 \leq i \leq n_j$, $1 \leq j \leq q$. To be precise, let $n = \sum_{j=1}^q n_j$ be the total sample size, then the likelihood ratio test for the hypothesis (6) under the normal assumption $x \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is given by

$$\Lambda_{n,2} = \frac{\prod_{j=1}^{q} |\mathbf{A}_j / n_j|^{\frac{1}{2} n_j}}{|\mathbf{A}/n|^{\frac{1}{2} n}}, \tag{7}$$

where the $p \times p$ matrices $\mathbf{A}_j$ and $\mathbf{A}$ are defined as

$$\mathbf{A}_j = \sum_{k=1}^{n_j} \mathbf{y}_{jk} \mathbf{y}_{jk}^\top \,, \quad \mathbf{A} = \sum_{j=1}^{q} \mathbf{A}_j.$$

In the case $\mathbf{y}_j \sim \mathcal{N}(\boldsymbol{\mu}_j, \mathbf{\Sigma}_j)$, $1 \leq j \leq q$, Jiang and Yang (2013) and Jiang and Qi (2015) proved asymptotic normality of the corresponding log-likelihood ratio test statistic under the null hypothesis if the number q of groups is fixed. These results were generalized by Dette and Dörnemann (2020) and Guo and Qi (2021) for the case of an increasing number $q = q_n$ of groups. All of these works dealt only with normally distributed data. In the following theorem, we provide the limiting distribution of $\log \Lambda_{n,2}$ under the null hypothesis without imposing a normal assumption on $\mathbf{y}_1, \ldots, \mathbf{y}_q$ in a high-dimensional setting, where the number of groups is allowed to increase.

Theorem 3.1. *Let the components of* $\mathbf{x}$ *be i.i.d. centered random variables following a continuous distribution with finite* $(4 + \delta)$*th moment for some* $\delta > 0$*. Assume that* $q = q_n \geq 2$ *is a possibly increasing integer,* $n_j = n_j(n) > p = p_n$ *for every* $n \in \mathbb{N}$ *and* $0 < \inf_{n \in \mathbb{N}} p/n \leq \sup_{n \in \mathbb{N}} \max_{1 \leq j \leq q} p/n_j < 1$*. Then it holds under the null hypothesis* (6)

$$\frac{2 \left(\log \Lambda_{n,2} - \mu_n \right)}{n \sigma_n} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

where

$$\begin{aligned} \mu_n =& n \left(n - p - \frac{1}{2} \right) \log \left(1 - \frac{p}{n} \right) - \sum_{j=1}^{q} n_j \left(n_j - p - \frac{1}{2} \right) \log \left(1 - \frac{p}{n_j} \right), \\ \sigma_n^2 =& \log \left(1 - \frac{p}{n} \right) - \sum_{j=1}^{q} \left(\frac{n_j}{n} \right)^2 \log \left(1 - \frac{p}{n_j} \right). \end{aligned} \tag{8}$$

The proof is provided is Section 5.3. Similarly to Remark 2.1, an asymptotic level α for the hypothesis (6) can be constructed using Theorem 3.1. For the sake of brevity, we omit the details.

4 Finite-sample properties

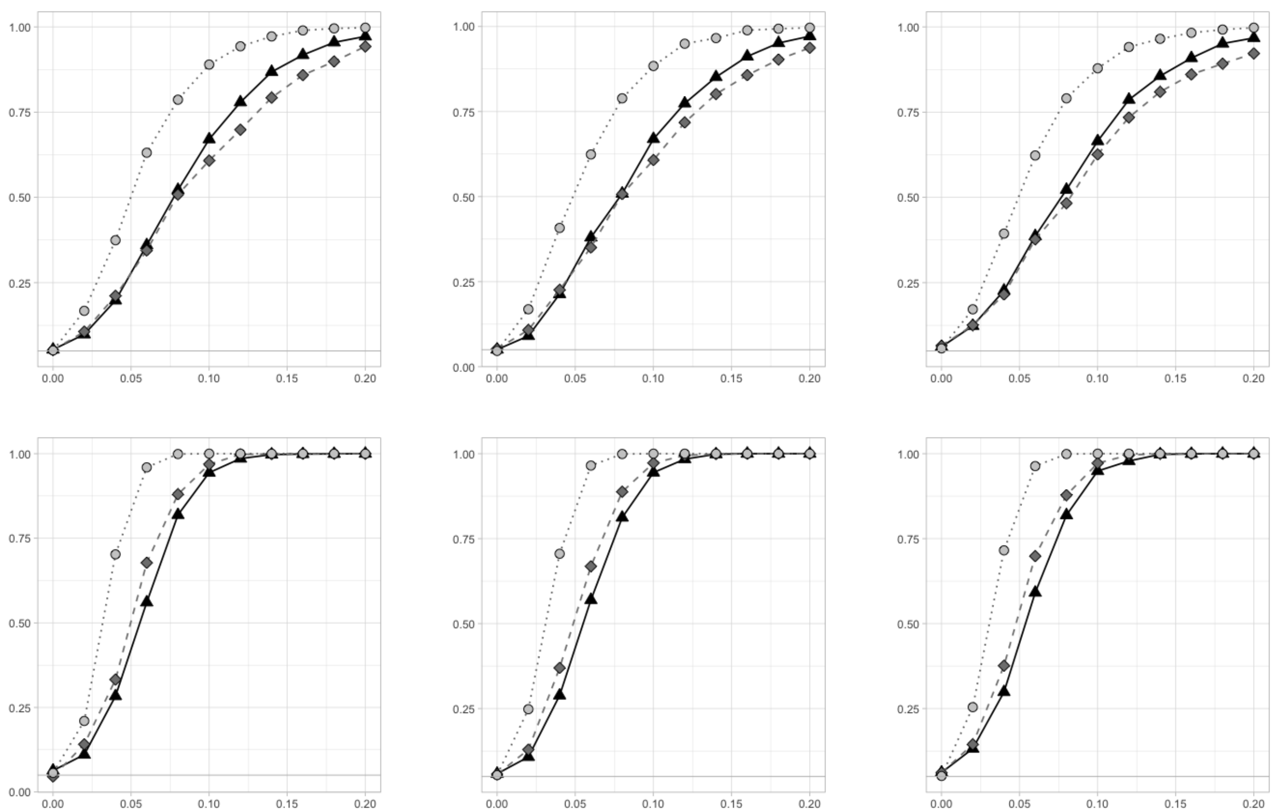

Figure 2: Rejection rates for the log-likelihood ratio test (4) for scenario 1 (first row) and scenario 2 (second row) under (9) based on 2,000 simulation runs. First column: standard normally distributed data, second column: standardized t-distributed data with 15 degrees of freedom, third column: centered exponentially distributed data with parameter 1. The triangle indicates $n = 100, p = 60$, the square $n = 120, p = 90$ and the circle $n = 180, p = 120$. The vertical grey line in each figure defines the nominal level $\alpha = 5\%$.

In this section, we investigate the finite-sample properties of the test (4) under both the hypothesis and the alternative. Following Qi et al. (2019), we consider the following alternative

$$\mathbf{\Sigma} = (1 - \delta)\mathbf{I} + \delta\mathbf{1}, \tag{9}$$

where $\mathbf{1}$ denotes the $p \times p$ matrix filled with ones and $\mathbf{I}$ denotes the $p \times p$ identity matrix. Here, the parameter $\delta \geq 0$ determines the "distance" to the null hypothesis (2) (note that the choice $\delta = 0$ corresponds to the null hypothesis (2)). In Figure 2, we display the empirical rejection rates of the test (4) for different choices of δ, n, p, q, p_i and different distributions for the random vector $\mathbf{x}$. All results are based on $2,000$ simulation runs and the components of $\mathbf{x}$ are independent identically distributed with respect to a standard normal distribution (first column), standardized t-distribution (second column) and centered exponential distribution (third column), respectively. The vertical grey line in each figure defines the nominal level $\alpha = 5\%$. For the choice of the different groups, we consider the following two scenarios:

1. $q = 3$, $p_1 = p_2 = p_3 = p/3$,
2. $q = p/2$, $p_1 = \ldots = p_{q-1} = 1, p_q = q + 1$.

We observe a good approximation of the nominal level in all cases under consideration. Moreover the power increases reasonably as δ increases. It should be noted that the increase in power is a bit stronger for scenario 2 than for scenario 1. The finite-sample properties do not significantly differ for the three underlying data generating distributions as indicated by the asymptotic result provided in Theorem 2.1. Overall, the test admits a desirable performance for finite-sample sizes, both for a large number and relatively small number of groups as covered by scenarios 1 and 2, and the accuracy improves for large sample size and dimension.

5 Proofs

In Section 5.1, we provide the proof of Theorem 2.1 and use auxiliary results given in Section 5.2. We conclude with the proof of Theorem 3.1 in Section 5.3. In the following, we make use of the notation $a \lesssim b$ which means that a is less than or equal to b up to a positive constant, that is, there exists some $C > 0$ independent of $n \in \mathbb{N}$ such that $a \leq Cb$.

5.1 Proof of Theorem 2.1

Note that, under the null hypothesis (2), we have

$$V_n = \frac{|\mathbf{\Sigma}||\hat{\mathbf{I}}|}{\prod\limits_{i=1}^{q} \left(|\mathbf{\Sigma}_{ii}||\hat{\mathbf{I}}_{ii}|\right)} = \frac{|\hat{\mathbf{I}}|}{\prod\limits_{i=1}^{q} |\hat{\mathbf{I}}_{ii}|}, \tag{10}$$

where

$$\begin{aligned}
\hat{\mathbf{I}} =& \frac{1}{n}\mathbf{X}_n\mathbf{X}_n^\top, \\
\hat{\mathbf{I}}_{ii} =& \frac{1}{n}\mathbf{X}_{n,i}\mathbf{X}_{n,i}^\top, \\
\mathbf{X}_n =& (\mathbf{x}_1, \ldots \mathbf{x}_n) = (\mathbf{b}_1, \ldots, \mathbf{b}_p)^\top, \\
\mathbf{X}_{n,i} =& (\mathbf{b}_{p_{i-1}^\star+1}, \ldots, \mathbf{b}_{p_i^\star})^\top, \\
p_i^\star =& \sum_{j=1}^{i} p_i,
\end{aligned}$$

for $1 \leq i \leq q$, where we set $p_0^\star = 0$. In order to establish a more handy representation for determinants of the sample covariance matrix, we proceed with a QR decomposition of $\mathbf{X}_n^\top$ and $\mathbf{X}_{n,i}^\top$ as explained in detail in Section 2 of Wang et al. (2018) and get

$$\hat{\mathbf{I}} = \frac{1}{n} \prod_{i=1}^{p} \mathbf{b}_i^\top \mathbf{P}(i-1)\mathbf{b}_i,$$

$$\hat{\mathbf{I}}_{ii} = \frac{1}{n} \prod_{j=p_{i-1}^\star+1}^{p_i^\star} \mathbf{b}_j^\top \mathbf{P}(p_{i-1}^\star+1; j-1)\mathbf{b}_j, \ 1 \leq i \leq q,$$

where

$$\begin{aligned}&\mathbf{P}(p_{i-1}^\star+1; j-1)\\ &= \mathbf{I} - \mathbf{X}_n(p_{i-1}^\star+1; j-1)^\top \left(\mathbf{X}_n(p_{i-1}^\star+1; j-1)\mathbf{X}_n(p_{i-1}^\star+1; j-1)^\top\right)^{-1} \mathbf{X}(p_{i-1}^\star+1; j-1),\end{aligned} \tag{11}$$

and

$$\mathbf{P}(j) = \mathbf{P}(1; j),$$

denote the projection matrices on the orthogonal complements of span$(\mathbf{b}_{p_{i-1}^\star+1}, \ldots \mathbf{b}_{j-1})$ and span$(\mathbf{b}_1, \ldots \mathbf{b}_j)$, respectively. Here, we denote

$$\mathbf{X}_n(i; j) = (\mathbf{b}_i, \ldots, \mathbf{b}_j)^\top, \ 1 \leq i \leq j \leq p$$

and $\mathbf{P}(0) = \mathbf{I} = \mathbf{P}(i; j)$ for $j > i$. This implies

$$\begin{aligned}\log V_n &= \sum_{i=1}^{p} \log\left(\mathbf{b}_i^\top \mathbf{P}(i-1)\mathbf{b}_i\right) - \sum_{i=1}^{q} \sum_{j=p_{i-1}^\star+1}^{p_i^\star} \log\left(\mathbf{b}_j^\top \mathbf{P}(p_{i-1}^\star+1; j-1)\mathbf{b}_j\right)\\ &= \sum_{i=p_1+1}^{p} \log\left(\mathbf{b}_i^\top \mathbf{P}(i-1)\mathbf{b}_i\right) - \sum_{i=2}^{q} \sum_{j=p_{i-1}^\star+1}^{p_i^\star} \log\left(\mathbf{b}_j^\top \mathbf{P}(p_{i-1}^\star+1; j-1)\mathbf{b}_j\right),\end{aligned}$$

where we used $\mathbf{P}(i-1) = \mathbf{P}(1; i-1)$ for $1 \leq i \leq p_1$. In the following, we will make use of Stirling's formula

$$\log n! = n \log n - n + \frac{1}{2}\log(2\pi n) + \frac{1}{12n} + \mathcal{O}\left(n^{-3}\right), \ n \to \infty.$$

As a preparation, we note that

$$\begin{aligned}&\sum_{i=p_1+1}^{p} \log(n-i+1) - \sum_{i=2}^{q} \sum_{j=p_{i-1}^\star+1}^{p_i^\star} \log(n-j+1+p_{i-1}^\star)\\ &= \log\frac{(n-p_1)!}{(n-p)!} - \sum_{i=2}^{q} \log\frac{n!}{(n-p_i)!}\end{aligned}$$

$$=\sum_{i=2}^{q}\left(n-p_i+\frac{1}{2}\right)\log\left(1-\frac{p_i}{n}\right)-\left(n-p+\frac{1}{2}\right)\log\left(1-\frac{p}{n}\right)-\sum_{i=2}^{q}\left(\frac{1}{12n}-\frac{1}{12(n-p_i)}\right)+o(1)$$

$$=\mu_n+\frac{\sigma_n^2}{2}-\frac{1}{12}\sum_{i=2}^{q}\frac{p_i}{n(n-p_i)}+o(1)=\mu_n+\frac{\sigma_n^2}{2}+o(1), \tag{12}$$

where we used the fact

$$\min_{1\le i\le q}(n-p_i)\to\infty, \tag{13}$$

which is a consequence of our assumptions. Defining for $p_1+1\le i\le p, 2\le j\le q$

$$X_i=\frac{\mathbf{b}_i^\top\mathbf{P}(i-1)\mathbf{b}_i-(n-i+1)}{n-i+1},\ X_{j,i}=\frac{\mathbf{b}_i^\top\mathbf{P}(p_{j-1}^\star+1;i-1)\mathbf{b}_i-(n-i+1+p_{j-1}^\star)}{n-i+1+p_{j-1}^\star},$$

$$Y_i=\log(1+X_i)-\left(X_i-\frac{X_i^2}{2}\right),\ Y_{j,i}=\log(1+X_{j,i})-\left(X_{j,i}-\frac{X_{j,i}^2}{2}\right),$$

we decompose using (12)

$$\begin{aligned}&\log V_n-\mu_n\\ =&\sum_{i=p_1+1}^{p}X_i-\sum_{j=2}^{q}\sum_{i=p_{j-1}^\star+1}^{p_j^\star}X_{j,i}-\left(\sum_{i=p_1+1}^{p}\frac{X_i^2}{2}-\sum_{j=2}^{q}\sum_{i=p_{j-1}^\star+1}^{p_j^\star}\frac{X_{j,i}^2}{2}-\frac{\check{\sigma}_n^2}{2}\right)\\ &+\sum_{i=p_1+1}^{p}Y_i-\sum_{j=2}^{q}\sum_{i=p_{j-1}^\star+1}^{p_j^\star}Y_{j,i}+o(1).\end{aligned}$$

The assertion of Theorem 2.1 is then implied by the following auxiliary results, which are proven in Sections 5.2.1, 5.2.2 and 5.2.3.

Lemma 5.1. *Under the assumptions of Theorem 2.1, it holds*

$$\frac{\sum_{i=p_1+1}^{p}X_i-\sum_{j=2}^{q}\sum_{i=p_{j-1}^\star+1}^{p_j^\star}X_{j,i}}{\sigma_n}\xrightarrow{\mathcal{D}}\mathcal{N}(0,1),\ n\to\infty.$$

Lemma 5.2. *Under the assumptions of Theorem 2.1, it holds*

$$\frac{\sum_{i=p_1+1}^{p}\frac{X_i^2}{2}-\sum_{j=2}^{q}\sum_{i=p_{j-1}^\star+1}^{p_j^\star}\frac{X_{j,i}^2}{2}-\frac{\sigma_n^2}{2}}{\sigma_n}\xrightarrow{\mathbb{P}}0,\ n\to\infty.$$

Lemma 5.3. *Under the assumptions of Theorem 2.1, it holds*

$$\sum_{i=p_1+1}^{p}\frac{Y_i}{\sigma_n}\xrightarrow{\mathbb{P}}0, \tag{14}$$

$$\sum_{j=2}^{q} \sum_{i=p_{j-1}^\star+1}^{p_j^\star} \frac{Y_{j,i}}{\sigma_n} \xrightarrow{\mathbb{P}} 0, \tag{15}$$

as $n \to \infty$.

5.2 Auxiliary results for the proof of Theorem 2.1

5.2.1 Proof of Lemma 5.1

Let $\mathcal{F}_i = \sigma(\{\mathbf{b}_1, \ldots, \mathbf{b}_i\})$ denote the σ-field generated by $\mathbf{b}_1, \ldots, \mathbf{b}_i$ for $1 \le i \le p$. We write

$$\sum_{j=p_1+1}^{p} X_j - \sum_{i=2}^{q} \sum_{j=p_{i-1}^\star+1}^{p_i^\star} X_{i,j} = \sum_{i=2}^{q} \sum_{j=p_{i-1}^\star+1}^{p_i^\star} (X_j - X_{i,j}) = \sum_{i=p_1+1}^{p} Z_i,$$

where

$$Z_i = X_i - X_{g(i),i}, \; p_1 + 1 \le i \le p,$$

where $g(i) = k$ if $\mathbf{x}^{(i)}$ belongs to the kth group, that is, if $p_{k-1} + 1 \le i \le p_k$. We observe that

$$\mathbb{E}\left[Z_i | \mathcal{F}_{i-1}\right] = 0$$

and that Z_i is measurable with respect to $\mathcal{F}_i$ for $p_1 + 1 \le i \le p$. Thus, we conclude that $(Z_i)_{p_1+1\le i \le p}$ forms a martingale difference scheme with respect to the filtration $(\mathcal{F}_i)_{p_1+1\le i\le p}$ scheme for every $n \in \mathbb{N}$. We aim to apply a central limit theorem for this dependency structure. In order to calculate the limiting variance, we write

$$\begin{aligned} Z_i = & \mathbf{b}_i^\top \frac{(n-i+1+p_{g(i)-1}^\star)\mathbf{P}(i-1) - (n-i+1)\mathbf{P}(p_{g(i)-1}^\star + 1; i-1)}{(n-i+1)(n-i+1+p_{g(i)-1}^\star)} \mathbf{b}_i \\ & - \operatorname{tr}\left(\frac{(n-i+1+p_{g(i)-1}^\star)\mathbf{P}(i-1) - (n-i+1)\mathbf{P}(p_{g(i)-1}^\star + 1; i-1)}{(n-i+1)(n-i+1+p_{g(i)-1}^\star)} \right), \; p_1 + 1 \le i \le p, \end{aligned}$$

and use the fact

$$\mathbb{E}\left(\mathbf{b}_i^\top \mathbf{A} \mathbf{b}_i - \operatorname{tr}\mathbf{A}\right)^2 = 2\operatorname{tr}\mathbf{A}^2 + (\nu_4 - 3)\operatorname{tr}\left(\mathbf{A}^{\odot 2}\right), \; 1 \le i \le p, \tag{16}$$

for any non-random matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, where $\nu_4 = \mathbb{E}[b_{11}^4]$. Here, $\odot$ denotes the Hadamard product of matrices (entry-wise multiplication) and we use the notation

$$\mathbf{A} \odot \mathbf{A} = \mathbf{A}^{\odot 2}.$$

Consequently, we observe

$$\mathbb{E}[Z_i^2 | \mathcal{F}_{i-1}] = \mathbb{E}[(X_i - X_{g(i),i})^2 | \mathcal{F}_{i-1}]$$

$$=2\left(\frac{\operatorname{tr}\mathbf{P}(i-1)^2}{(n-i+1)^2}+\frac{\operatorname{tr}\mathbf{P}(p^\star_{g(i)-1}+1;i-1)^2}{(n-i+1+p^\star_{g(i)-1})^2}-2\frac{\operatorname{tr}\left(\mathbf{P}(p^\star_{g(i)-1}+1;i-1)\mathbf{P}(i-1)\right)}{(n-i+1+p^\star_{g(i)-1})(n-i+1)}\right)$$

$$+(\nu_4-3)\Bigg(\frac{\operatorname{tr}(\mathbf{P}(i-1)^{\odot 2})}{(n-i+1)^2}+\frac{\operatorname{tr}\left(\mathbf{P}(p^\star_{g(i)-1}+1;i-1)^{\odot 2}\right)}{(n-i+1+p^\star_{g(i)-1})^2}$$

$$-2\frac{\operatorname{tr}\left(\mathbf{P}(p^\star_{g(i)-1}+1;i-1)\odot\mathbf{P}(i-1)\right)}{(n-i+1+p^\star_{g(i)-1})(n-i+1)}\Bigg)$$

$$=\sigma^2_{n,1,i}+(\nu_4-3)\sigma^2_{n,2,i},$$

where

$$\sigma^2_{n,1,i}=2\left(\frac{1}{n-i+1}-\frac{1}{n-i+1+p^\star_{g(i)-1}}\right),$$

$$\sigma^2_{n,2,i}=\frac{\operatorname{tr}(\mathbf{P}(i-1)^{\odot 2})}{(n-i+1)^2}+\frac{\operatorname{tr}\left(\mathbf{P}(p^\star_{g(i)-1}+1;i-1)^{\odot 2}\right)}{(n-i+1+p^\star_{g(i)-1})^2}$$

$$-2\frac{\operatorname{tr}\left(\mathbf{P}(p^\star_{g(i)-1}+1;i-1)\odot\mathbf{P}(i-1)\right)}{(n-i+1+p^\star_{g(i)-1})(n-i+1)}. \tag{17}$$

Note that $\sum_{i=p_1+1}^{p}\sigma^2_{n,2,i}=o_{\mathbb{P}}(1)$ by Lemma 5.5. For the term $\sigma^2_{n,1,i}$, we used that $\mathbf{P}(i-1)\mathbf{P}(p^\star_{g(i)-1}+1;i-1)=\mathbf{P}(i-1)$. Thus, we have for this term

$$\sum_{i=p_1+1}^{p}\sigma^2_{n,1,i}=2\left(\log(n-p_1)-\log(n-p)-\sum_{i=2}^{q}\sum_{j=p^\star_{i-1}+1}^{p^\star_i}\frac{1}{n-j+1+p^\star_{i-1}}\right)+o(1)$$

$$=2\Bigg\{\log(n-p_1)-\log(n-p)-\sum_{i=2}^{q}\left\{\log(n)-\log(n-p^\star_i+p^\star_{i-1})\right\}$$

$$-\sum_{i=2}^{q}\left(\frac{1}{2n}-\frac{1}{2(n-p_i)}\right)\Bigg\}+o(1)$$

$$=2\left(\log\left(1-\frac{p_1}{n}\right)-\log\left(1-\frac{p}{n}\right)+\sum_{i=2}^{q}\log\left(1-\frac{p_i}{n}\right)+\frac{1}{2}\sum_{i=2}^{q}\frac{p_i}{n(n-p_i)}\right)+o(1)$$

$$=2\left(\sum_{i=1}^{q}\log\left(1-\frac{p_i}{n}\right)-\log\left(1-\frac{p}{n}\right)\right)+o(1)=\sigma^2_n+o(1), \tag{18}$$

where we used (13), $\sum_{i=2}^{q}p_i\le p$ and the expansion for the partial sums of the harmonic series

$$\sum_{k=1}^{n}\frac{1}{k}=\log n+\gamma+\frac{1}{2n}+\mathcal{O}\left(n^{-2}\right),\ n\to\infty. \tag{19}$$

Here, γ denotes the Euler-Mascheroni constant. Note that the term in (18) is bounded away from zero for all $n \in \mathbb{N}$. More precisely, we have applying inequality (33) of Qi et al. (2019)

$$\sigma_n^2 \geq 2 \left\{ \sum_{i=1}^{q} \log\left(1 - \frac{p_i}{n}\right) - \log\left(1 - \frac{p}{n}\right) \right\} \geq -2 \left(\log\left(1 - \frac{p}{n}\right) + \frac{p}{n} \right) (1-\eta) > 0 \tag{20}$$

uniformly over $n \in \mathbb{N}$ (recall that $\inf_{n\in\mathbb{N}} p/n > 0$ and $\max_{1\leq i\leq q} p_i \leq \eta p$). These considerations imply

$$\sum_{i=p_1+1}^{p} \mathbb{E}\left[\frac{Z_i^2}{\sigma_n^2} \Big| \mathcal{F}_{i-1} \right] \stackrel{\mathbb{P}}{\to} 1, \; n \to \infty. \tag{21}$$

Let $\varepsilon > 0$. Then, using Lemma B.26 in Bai and Silverstein (2010) and recalling that $\mathbb{E}|x_{11}|^{4+\delta} < \infty$, we get

$$\begin{aligned}
&\sum_{i=p_1+1}^{p} \mathbb{E}[Z_i^2 I\{|Z_i| > \varepsilon\}] \leq \frac{1}{\varepsilon^{\frac{\delta}{2}}} \sum_{i=p_1+1}^{p} \mathbb{E}[|Z_i|^{2+\frac{\delta}{2}}] \\
&\lesssim \sum_{i=p_1+1}^{p} \frac{1}{(n-i+1)^{2+\frac{\delta}{2}}} \mathbb{E}\left| \mathbf{b}_i^\top \mathbf{P}(i-1)\mathbf{b}_i - (n-i+1) \right|^{2+\frac{\delta}{2}} \\
&\quad + \sum_{i=2}^{q} \sum_{j=p_{i-1}^\star+1}^{p_i^\star} \frac{1}{\left(n-j+1+p_{i-1}^\star\right)^{2+\frac{\delta}{2}}} \mathbb{E}\left| \mathbf{b}_j^\top \mathbf{P}(p_{i-1}^\star+1; i-1)\mathbf{b}_j - (n-j+1+p_{i-1}^\star) \right|^{2+\frac{\delta}{2}} \\
&\lesssim \sum_{i=p_1+1}^{p} \frac{1}{(n-i+1)^{1+\frac{\delta}{4}}} + \sum_{i=2}^{q} \sum_{j=p_{i-1}^\star+1}^{p_i^\star} \frac{1}{\left(n-j+1+p_{i-1}^\star\right)^{1+\frac{\delta}{4}}} = o(1), \; n \to \infty.
\end{aligned}$$

Using (20) and the fact that $(\sigma_n^2)_{n\in\mathbb{N}}$ is bounded, we see that $(Z_i/\sigma_n)_{p_1+1\leq i\leq p}$ satisfies the following Lindeberg condition for all $\varepsilon > 0$:

$$\sum_{i=p_1+1}^{p} \mathbb{E}\left[\frac{Z_i^2}{\sigma_n^2} I\{|Z_i/\sigma_n| > \varepsilon\} \right] = o(1), \; n \to \infty. \tag{22}$$

Since (21) and (22) hold true, we may apply a CLT for martingale difference schemes (e.g., see Corollary 3.1 in Hall and Heyde (1980)) and the proof of Lemma 5.1 concludes.

5.2.2 Proof of Lemma 5.2

Note that for each $n \in \mathbb{N}$ both $(X_i)_{p_1+1\leq i\leq p}$ and $(X_{g(i),i})_{p_1+1\leq i\leq p}$ form a martingale difference scheme with respect to the filtration $(\mathcal{F}_i)_{p_1+1\leq i\leq p}$ defined in Section 5.2.1. We obtain from the proof of Lemma 5.1 in Section 5.2.1

$$\sum_{i=p_1+1}^{p} \mathbb{E}[X_i^2 | \mathcal{F}_{i-1}] = \check{\sigma}_{n,1}^2 + o_{\mathbb{P}}(1), \; n \to \infty, \tag{23}$$

and

$$\sum_{i=2}^{q} \mathbb{E}\left[X_{g(i),i}^2 \middle| \mathcal{F}_{i-1}\right] = \sum_{i=2}^{q} \sum_{j=p_{i-1}^{\star}+1}^{p_i^{\star}} \mathbb{E}[X_{i,j}^2 | \mathcal{F}_{i-1}] = \check{\sigma}_{n,2}^2 + o_{\mathbb{P}}(1), \ n \to \infty, \tag{24}$$

where we define

$$\check{\sigma}_{n,1}^2 = -2 \log\left(1 - \frac{p}{n}\right) + (\nu_4 - 3) \sum_{i=p_1+1}^{p} \frac{\operatorname{tr}(\mathbf{P}(i-1)^{\odot 2})}{(n-i+1)^2},$$

$$\check{\sigma}_{n,2}^2 = -2 \sum_{i=1}^{q} \log\left(1 - \frac{p_i}{n}\right) + (\nu_4 - 3) \sum_{i=p_1+1}^{p} \frac{\operatorname{tr}\left(\mathbf{P}(p_{g(i)-1}^{\star}+1; i-1)^{\odot 2}\right)}{(n-i+1+p_{g(i)-1}^{\star})^2}.$$

Recalling that $0 < \inf_{n\in\mathbb{N}} \min_{1\le i\le q}(p_i q)/n \le \sup_{n\in\mathbb{N}} p/n < 1$ and using the inequality $\log(1 + x) \le x$ for $x > -1$, we note that

$$0 < \inf_{n\in\mathbb{N}} \check{\sigma}_{n,1}^2 \le \sup_{n\in\mathbb{N}} \check{\sigma}_{n,1}^2 < \infty, \ 0 < \inf_{n\in\mathbb{N}} \check{\sigma}_{n,2}^2 \le \sup_{n\in\mathbb{N}} \check{\sigma}_{n,2}^2 < \infty. \tag{25}$$

Taking a closer look at the proof of (22), we observe that both schemes satisfy the Lindeberg condition, that is, we have for $\varepsilon > 0$

$$\sum_{i=p_1+1}^{p} \mathbb{E}\left[\frac{X_i^2}{\check{\sigma}_{n,1}^2} I\{|X_i/\check{\sigma}_{n,1}| > \varepsilon\}\right] = o(1), \tag{26}$$

$$\sum_{i=p_1+1}^{p} \mathbb{E}\left[\frac{X_{g(i),i}^2}{\check{\sigma}_{n,2}^2} I\{|X_{g(i),i}/\check{\sigma}_{n,2}| > \varepsilon\}\right] = o(1). \tag{27}$$

By the proof of Corollary 3.1 in Hall and Heyde (1980), we see that (23), (24), (26) and (27) imply that the conditional variance can be approximated by the sum of squares, that is,

$$\sum_{i=p_1+1}^{p} \frac{X_i^2 - \mathbb{E}[X_i^2|\mathcal{F}_{i-1}]}{\check{\sigma}_{n,1}^2} \xrightarrow{\mathbb{P}} 0, \ \sum_{i=p_1+1}^{p} \frac{X_{g(i),i}^2 - \mathbb{E}\left[X_{g(i),i}^2|\mathcal{F}_{i-1}\right]}{\check{\sigma}_{n,2}^2} \xrightarrow{\mathbb{P}} 0, \ n \to \infty,$$

Combining these observations with (23), (24) and (25), we get

$$\sum_{i=p_1+1}^{p} X_i^2 - \check{\sigma}_{n,1}^2 \xrightarrow{\mathbb{P}} 0, \ \sum_{i=p_1+1}^{p} X_{g(i),i}^2 - \check{\sigma}_{n,2}^2 \xrightarrow{\mathbb{P}} 0.$$

Using (20) and $\check{\sigma}_{n,1}^2 - \check{\sigma}_{n,2}^2 = \sigma_n^2 + o_{\mathbb{P}}(1)$ by Lemma 5.4, the proof of Lemma 5.2 concludes.

5.2.3 Proof of Lemma 5.3

In the following, we will show that the convergence in (14) holds true. Then, the assertion (15) can be shown similarly.
Let $0 < \varepsilon < 1$. Then, we estimate for $1 + p_1 \le i \le p$ using Taylor's expansion

$$\mathbb{E}\left[|Y_i| I\{|X_i| \le 1-\varepsilon\}\right] \lesssim \mathbb{E}\left[|X_i|^3 I\{|X_i| \le 1-\varepsilon\}\right] \lesssim \mathbb{E}|X_i|^{2+\frac{\delta}{2}}.$$

We also have

$$\begin{aligned}
&\mathbb{E}\left[|Y_i| I\{|X_i| > 1-\varepsilon\}\right] \\
\leq &\mathbb{E}\left[|\log(1+X_i)| I\{|X_i| > 1-\varepsilon\}\right] + \mathbb{E}\left[|X_i| I\{|X_i| > 1-\varepsilon\}\right] + \mathbb{E}\left[X_i^2 I\{|X_i| > 1-\varepsilon\}\right] \\
\lesssim &\mathbb{E}\left[|X_i| I\{|X_i| > 1-\varepsilon\}\right] + \mathbb{E}\left[X_i^2 I\{|X_i| > 1-\varepsilon\}\right] \lesssim \mathbb{E}|X_i|^{2+\frac{\delta}{2}},
\end{aligned}$$

where we used the inequality $\log(1+x) \leq x$ for all $x > -1$. These two estimates imply

$$\sum_{i=p_1+1}^{p} \mathbb{E}|Y_i| \lesssim \sum_{i=p_1+1}^{p} \mathbb{E}|X_i|^{2+\frac{\delta}{2}} = o(1), \ n \to \infty,$$

where we used Lemma B.26 in Bai and Silverstein (2010) as in the proof of (22). Thus, we obtain

$$\sum_{i=p_1+1}^{p} Y_i \xrightarrow{\mathbb{P}} 0, \ n \to \infty,$$

which implies the assertion of Lemma 5.3 recalling $\inf_{n\in\mathbb{N}} \sigma_n^2 > 0$.

5.2.4 Further auxiliary results for the proof of Theorem 2.1

Lemma 5.4. *It holds, as $n \to \infty$,*

$$\sum_{i=p_1+1}^{p} \left(\frac{\operatorname{tr}\left(\mathbf{P}(i-1)^{\odot 2}\right)}{(n-i+1)^2} - \frac{1}{n} \right) \xrightarrow{\mathbb{P}} 0, \tag{28}$$

$$\sum_{i=p_1+1}^{p} \left(\frac{\operatorname{tr}\left(\mathbf{P}(p^\star_{g(i)-1}+1; i-1)^{\odot 2}\right)}{(n-i+1+p^\star_{g(i)-1})^2} - \frac{1}{n} \right) \xrightarrow{\mathbb{P}} 0, \tag{29}$$

where the projection matrices are defined in (11).

Proof of Lemma 5.4. As a preparation, we will first show that for any sequence $(i_n)_{n\in\mathbb{N}}$ such that $2 \leq i_n \leq p_n$ for all $n \in \mathbb{N}$ and the limit $\lim_{n\to\infty} i_n/n \in [0,1)$ exists, it holds

$$a_{i_n,n} - b_{i_n,n} \xrightarrow{\mathbb{P}} 0, \ n \to \infty, \tag{30}$$

where we define for $2 \leq i \leq p$

$$a_{i,n} = \frac{\operatorname{tr}\left(\mathbf{P}(i-1)^{\odot 2}\right)}{n-i+1}, \ b_{i,n} = \left(1 - \frac{i}{n}\right), \quad c_n = \sum_{i=p_1+1}^{p} \frac{a_{i,n} - b_{i,n}}{n-i+1}.$$

In the following, we denote the diagonal entries of $\mathbf{P}(i_n - 1)$ by p_{ii} $(1 \leq i \leq n)$. First, we consider the case $\lim_{n\to\infty} i_n/n = 0$. For this case, we note that

$$\frac{\operatorname{tr}\left(\mathbf{P}(i_n-1)^{\odot 2}\right)}{n} = \frac{1}{n}\sum_{i=1}^{n} (1-p_{ii})^2 - 1 + \frac{2}{n}\sum_{i=1}^{n} p_{ii} = \frac{2(n-i_n+1)}{n} - 1 + o_{\mathbb{P}}(1)$$

$$= 1 + o_{\mathbb{P}}(1),\ n \to \infty,$$

where we used

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(1-p_{ii})^2 \le \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[1-p_{ii}] = \frac{1}{n}\operatorname{tr}(\mathbf{I}-\mathbf{P}) = \frac{i_n-1}{n} = o(1),\ n\to\infty.$$

In this case, we conclude

$$a_{i_n,n} - b_{i_n,n} = \frac{\operatorname{tr}\left(\mathbf{P}(i_n-1)^{\odot 2}\right)}{n} - 1 + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1),\ n\to\infty.$$

Now consider the case $\lim_{n\to\infty} i_n/n = \gamma \in (0,1)$. Then we have from Theorem 3.2 in Anatolyev and Yaskov (2017)

$$\frac{1}{n}\sum_{i=1}^{n}(1-p_{ii}-\gamma)^2 \xrightarrow{\mathbb{P}} 0,\ n\to\infty,$$

which implies

$$\begin{aligned}\frac{\operatorname{tr}\left(\mathbf{P}(i_n-1)^{\odot 2}\right)}{n} &= \frac{1}{n}\sum_{i=1}^{n}(1-p_{ii}-\gamma)^2 - (1-\gamma)^2 + \frac{2(1-\gamma)}{n}\sum_{i=1}^{n}p_{ii} \\ &= \frac{2(1-\gamma)(n-i_n+1)}{n} - (1-\gamma)^2 + o_{\mathbb{P}}(1) = (1-\gamma)^2 + o_{\mathbb{P}}(1),\ n\to\infty,\end{aligned}$$

which implies that (30) holds also true in this case. We continue with a proof of (28) by showing that any subsequence of $(c_n)_{n\in\mathbb{N}}$ admits a further subsequence converging in probability to 0. Let $(c_{n_j})_{j\in\mathbb{N}}$ be an arbitrary subsequence of $(c_n)_{n\in\mathbb{N}}$. We choose

$$i_{n_j} \in \operatorname*{arg\,max}_{p_1+1\le i\le p}\left(a_{i,n_j} - b_{i,n_j}\right).$$

Not that there exists a subsequence $(i_{n_{j_k}})_{k\in\mathbb{N}}$ of $(i_{n_j})_{j\in\mathbb{N}}$ which admits a limit $\lim_{k\to\infty} i_{n_{j_k}}/k \in [0,1)$ (that is, this subsequence satisfies the assumption for (30)). Then, it holds using (30)

$$c_{n_{j_k}} \lesssim \max_{p_1+1\le i\le p}\left(a_{i,n_{j_k}} - b_{i,n_{j_k}}\right) = a_{i_{n_{j_k}},n_{j_k}} - b_{i_{n_{j_k}},n_{j_k}} \xrightarrow{\mathbb{P}} 0,\ n\to\infty.$$

This implies the convergence $c_n \xrightarrow{\mathbb{P}} 0$ of the whole sequence $(c_n)_{n\in\mathbb{N}}$ for $n\to\infty$ and thus, the convergence in (33) holds true. The second assertion (29) of Lemma 5.4 can be shown similarly. □

Lemma 5.5. *It holds*

$$\sum_{i=p_1+1}^{p}\sigma^2_{n,2,i} \xrightarrow{\mathbb{P}} 0,\ n\to\infty,$$

where the term $\sigma^2_{n,2,i}$ *is defined in* (17).

Proof of Lemma 5.5. Recalling the definition of $\sigma^2_{n,2,i}$ and using Lemma 5.4, it suffices to show

$$\sum_{i=p_1+1}^{p} \left(\operatorname{tr} \mathbf{A}_i \odot \mathbf{B}_i - \frac{1}{n} \right) \xrightarrow{\mathbb{P}} 0,\ n \to \infty,$$

where we denote

$$\mathbf{A}_i = \frac{\mathbf{P}(i-1)}{n-i+1}, \quad \mathbf{B}_i = \frac{\mathbf{P}(p^\star_{g(i)-1}+1; i-1)}{n-i+1+p^\star_{g(i)-1}},\ p_1+1 \le i \le p.$$

Note that $\operatorname{tr} \mathbf{A}_i = \operatorname{tr} \mathbf{B}_i = 1$ for all $p_1+1 \le i \le p$. This gives

$$\begin{aligned}
\sum_{i=p_1+1}^{p} \left(\operatorname{tr} \mathbf{A}_i \odot \mathbf{B}_i - \frac{1}{n} \right) &= \sum_{i=p_1+1}^{p} \operatorname{tr} \left\{ \left(\mathbf{A}_i - \frac{1}{n}\mathbf{I} \right) \odot \left(\mathbf{B}_i - \frac{1}{n}\mathbf{I} \right) \right\} \\
&\le \sum_{i=p_1+1}^{p} \left\{ \operatorname{tr} \left(\mathbf{A}_i - \frac{1}{n}\mathbf{I} \right)^{\odot 2} \operatorname{tr} \left(\mathbf{B}_i - \frac{1}{n}\mathbf{I} \right)^{\odot 2} \right\}^{\frac{1}{2}} \\
&\le \left\{ \sum_{i=p_1+1}^{p} \operatorname{tr} \left(\mathbf{A}_i - \frac{1}{n}\mathbf{I} \right)^{\odot 2} \sum_{i=p_1+1}^{p} \operatorname{tr} \left(\mathbf{B}_i - \frac{1}{n}\mathbf{I} \right)^{\odot 2} \right\}^{\frac{1}{2}} \\
&= \left\{ \sum_{i=p_1+1}^{p} \left(\operatorname{tr} \mathbf{A}_i^{\odot 2} - \frac{1}{n} \right) \sum_{i=p_1+1}^{p} \left(\operatorname{tr} \mathbf{B}_i^{\odot 2} - \frac{1}{n} \right) \right\}^{\frac{1}{2}} \xrightarrow{\mathbb{P}} 0,\ n \to \infty,
\end{aligned}$$

where we applied the Cauchy-Schwarz inequality twice and Lemma 5.4. □

5.3 Proof of Theorem 3.1

For proving Theorem 5.6, we need the following properties of the variance.

Lemma 5.6. *Under the assumptions of Theorem 3.1, we have*

$$0 < \inf_{n \in \mathbb{N}} \sigma_n^2 \le \sup_{n \in \mathbb{N}} \sigma_n^2 < \infty,$$

where σ_n^2 denotes the variance defined in (8)*.*

Proof of Lemma 5.6 . Define the functions

$$\xi(x) = -\left(\log(1-x) + x \right),\ \eta(x) = \frac{\xi(x)}{x^2},$$

where $x \in (0,1)$. Note that η is a monotone increasing function with $\eta(x) \ge 1/2$. Using $\sum_{j=1}^{q} n_j = n$ and the definition $n_{\max} = \max_{1 \le j \le q} n_j$, we obtain the estimate

$$n^2 \sigma_n^2 = -\sum_{j=1}^{q} n_j^2 \xi\left(\frac{p}{n_j}\right) + n^2 \xi\left(\frac{p}{n}\right) = p^2 \left\{ \sum_{j=1}^{q} \eta\left(\frac{p}{n_j}\right) - \eta\left(\frac{p}{n}\right) \right\}$$

$$\geq p^2 \left\{ q\eta\left(\frac{p}{n_{max}}\right) - \eta\left(\frac{p}{n_{\max}}\right)\right\} = p^2(q-1)\eta\left(\frac{p}{n_{max}}\right).$$

Using $\inf_{n\in\mathbb{N}} p/n > 0$, we conclude $\inf_{n\in\mathbb{N}} \sigma_n^2 > 0$. Moreover, from our assumption on the dimension-to-subsample-size ratios p/n_j, we obtain

$$\sup_{n\in\mathbb{N}} \sigma_n^2 \leq \sup_{n\in\mathbb{N}} \max_{1\leq j\leq q} \left(-\log\left(1-\frac{p}{n_j}\right)\right) < \infty.$$

□

In order to prove Theorem 3.1, we follow the same strategy as in the proof of Theorem 2.1 and concentrate on discussing the main steps. Recalling the definition (7) of the likelihood ratio test, we note that under the null hypothesis (6)

$$\begin{aligned} 2\log\Lambda_{n,1} &= \sum_{j=1}^{q} n_j \log|\mathbf{A}_j| - n\log|\mathbf{A}| + pn\log n - \sum_{j=1}^{q} n_j p \log n_j \\ &= \sum_{j=1}^{q} n_j \log|n_j\hat{\mathbf{I}}_j| - n\log|n\hat{\mathbf{I}}| + pn\log n - \sum_{j=1}^{q} n_j p\log n_j, \end{aligned}$$

where $\hat{\mathbf{I}}$ is defined in the proof of Theorem 2.1 and

$$\hat{\mathbf{I}}_j = \frac{1}{n_j}\sum_{k=1}^{n_j} \mathbf{x}_{jk}\mathbf{x}_{jk}^\top.$$

Applying the QR-procedure to the matrices $\hat{\mathbf{I}}$ and $\hat{\mathbf{I}}_j$ $(1\leq j\leq q)$, we obtain for their determinants

$$|n\hat{\mathbf{I}}| = \prod_{i=1}^{p} \mathbf{b}_i^\top \mathbf{P}(i-1)\mathbf{b}_i,$$

$$|n_j\hat{\mathbf{I}}_j| = \prod_{i=1}^{p} \mathbf{b}_{ji}^\top \mathbf{P}(j;i-1)\mathbf{b}_{ji},$$

where

$$\mathbf{b}_i = (\mathbf{b}_{1i},\ldots,\mathbf{b}_{qi})^\top \in \mathbb{R}^n, \mathbf{b}_{ji}\in\mathbb{R}^{n_j},\ 1\leq j\leq q,$$

and throughout this proof, $\mathbf{P}(j;i-1)\in\mathbb{R}^{n_j\times n_j}$ denotes the projection matrix on the orthogonal complement of $\text{span}\{\mathbf{b}_{j1},\ldots,\mathbf{b}_{j,i-1}\}$ for $1\leq i\leq p,\ 1\leq j\leq q$ (note that we have a different definition than in the proof of Theorem 2.1). We set $\mathbf{P}(j;0)=\mathbf{I}\in\mathbb{R}^{n_j\times n_j}$. The remaining quantities are defined as in the proof of Theorem 2.1. With a slight abuse of notation, we define for $1\leq i\leq p,\ 1\leq j\leq q$

$$X_i = \frac{\mathbf{b}_i^\top\mathbf{P}(i-1)\mathbf{b}_i - (n-i+1)}{n-i+1},\ X_{j,i} = \frac{\mathbf{b}_{ji}^\top\mathbf{P}(j;i-1)\mathbf{b}_{ji} - (n_j-i+1)}{n_j-i+1},$$

$$Y_i = \log(1+X_i) - \left(X_i - \frac{X_i^2}{2}\right), \; Y_{j,i} = \log(1+X_{j,i}) - \left(X_{j,i} - \frac{X_{j,i}^2}{2}\right).$$

Similarly to (12), we obtain using Stirling's formula

$$\begin{aligned}
&\sum_{i=1}^{p}\sum_{j=1}^{q} n_j \log(n_j - i + 1) - n \sum_{i=1}^{p} \log(n-i+1) + pn\log n - \sum_{j=1}^{q} n_j p \log n_j \\
=&\sum_{j=1}^{q} n_j \log\left(\frac{n_j!}{(n_j-p)!}\right) - n\log\left(\frac{n!}{(n-p)!}\right) + pn\log n - \sum_{j=1}^{q} n_j p \log n_j \\
=&n\left(n-p+\frac{1}{2}\right)\log\left(1-\frac{p}{n}\right) - \sum_{j=1}^{q} n_j\left(n_j - p + \frac{1}{2}\right)\log\left(1-\frac{p}{n_j}\right) \\
&+\sum_{j=1}^{q}\frac{n_j}{12}\left(\frac{1}{n_j} - \frac{1}{n_j-p}\right) + o(n) \\
=&n\left(n-p+\frac{1}{2}\right)\log\left(1-\frac{p}{n}\right) - \sum_{j=1}^{q} n_j\left(n_j - p + \frac{1}{2}\right)\log\left(1-\frac{p}{n_j}\right) \\
&-\sum_{j=1}^{q}\frac{p}{12(n_j-p)} + o(n) \\
=&n\left(n-p+\frac{1}{2}\right)\log\left(1-\frac{p}{n}\right) - \sum_{j=1}^{q} n_j\left(n_j - p + \frac{1}{2}\right)\log\left(1-\frac{p}{n_j}\right) + o(n) \qquad (31) \\
=&\mu_n + \frac{\breve{\sigma}_n^2}{2} + o(n), \; n \to \infty,
\end{aligned}$$

where

$$\breve{\sigma}_n^2 = 2\left\{n\log\left(1-\frac{p}{n}\right) - \sum_{j=1}^{q} n_j \log\left(1-\frac{p}{n_j}\right)\right\}.$$

For (31), we used that under the assumptions of Theorem 3.1

$$\sum_{j=1}^{q}\frac{p}{12n(n_j-p)} \lesssim \frac{q}{n} \le \frac{1}{\min_j n_j} = o(1), \; n \to \infty. \qquad (32)$$

Consequently, we may decompose

$$\begin{aligned}
&2\left(\log\Lambda_{n,1} - \mu_n\right) \\
=&\sum_{j=1}^{q}\sum_{i=1}^{p} n_j X_{j,i} - n\sum_{i=1}^{p} X_i - \left(\sum_{j=1}^{q}\sum_{i=1}^{p} n_j \frac{X_{j,i}^2}{2} - n\sum_{i=1}^{p}\frac{X_i^2}{2} - \frac{\breve{\sigma}_n^2}{2}\right) \\
&+\sum_{j=1}^{q}\sum_{i=1}^{p} n_j Y_{j,i} - n\sum_{i=1}^{p} Y_i + o(n), \; n \to \infty.
\end{aligned}$$

Note that $(W_i)_{1\le i\le p}$ with

$$W_i = \sum_{j=1}^{q} n_j X_{j,i} - nX_i,\ 1 \le i \le p,$$

forms a martingale difference scheme with respect to the filtration $(\mathcal{A}_i)_{1\le i\le p}$, where the σ-field $\mathcal{A}_i$ is generated by the random variables $\mathbf{b}_1, \ldots, \mathbf{b}_i$ for $1 \le i \le p$. One can show that

$$\frac{\sum_{j=1}^{q}\sum_{i=1}^{p} n_j \frac{X_{j,i}^2}{2} - n \sum_{i=1}^{p} \frac{X_i^2}{2} - \frac{\breve{\sigma}_n^2}{2}}{n\sigma_n} \xrightarrow{\mathbb{P}} 0,\ n \to \infty, \tag{33}$$

and

$$\frac{\sum_{j=1}^{q}\sum_{i=1}^{p} n_j Y_{j,i} - n \sum_{i=1}^{p} Y_i}{n\sigma_n} \xrightarrow{\mathbb{P}} 0,\ n \to \infty. \tag{34}$$

For the sake of brevity, we omit the proofs of (33) and (34) as they are very similar to the proofs of Lemma 5.2 and Lemma 5.3. We continue with a proof of the asymptotic normality of the scheme $(W_i/(n\sigma_n))_{1\le i\le p}$. To begin with, we show that

$$\sum_{i=1}^{p} \mathbb{E}\left[\left(\frac{W_i}{n\sigma_n} \right)^2 \middle| \mathcal{A}_{i-1} \right] = 1 + o_{\mathbb{P}}(1),\ n \to \infty. \tag{35}$$

As a preparation for (35), note that

$$\begin{aligned}
&\sum_{i=1}^{p}\sum_{j=1}^{q} n_j n \mathbb{E}[X_i X_{j,i} | \mathcal{A}_{i-1}] \\
&= \sum_{i=1}^{p}\sum_{j=1}^{q} n n_j \left\{ \sum_{k=n_{j-1}^{\star}+1}^{n_j^{\star}} (\nu_4 - 3) \frac{(\mathbf{P}(j;i-1))_{kk}\, (\mathbf{P}(i-1))_{kk}}{(n-i+1)(n_j-i+1)} + 2\frac{\operatorname{tr}\left(\mathbf{P}(j-1)\tilde{\mathbf{P}}(j;i-1)\right)}{(n-i+1)(n_j-i+1)} \right\} \\
&= \sum_{i=1}^{p}\sum_{j=1}^{q} n n_j \left\{ (\nu_4 - 3) \frac{\operatorname{tr}\left(\tilde{\mathbf{P}}(j;i-1) \odot \mathbf{P}(i-1)\right)}{(n-i+1)(n_j-i+1)} + 2\frac{\operatorname{tr}\tilde{\mathbf{P}}(j;i-1)}{(n_j-i+1)(n-i+1)} \right\} \\
&= (\nu_4 - 3) \sum_{i=1}^{p}\sum_{j=1}^{q} n n_j \sum \frac{\operatorname{tr}\left(\tilde{\mathbf{P}}(j;i-1) \odot \mathbf{P}(i-1)\right)}{(n-i+1)(n_j-i+1)} + \sum_{i=1}^{p} \frac{2n^2}{(n-i+1)},
\end{aligned}$$

where $\tilde{\mathbf{P}}(j;i-1)$ denotes the $(n \times n)$ dimensional embedded matrix of $\mathbf{P}(j;i-1) \in \mathbb{R}^{n_j \times n_j}$, that is,

$$\left(\tilde{\mathbf{P}}(j;i-1)\right)_{kl} = \begin{cases} (\mathbf{P}(j;i-1))_{kl} & : n_{j-1}^{\star} + 1 \le k, l \le n_j^{\star}, \\ 0 & : \text{else}, \end{cases}$$

for $1 \leq k, l \leq n,\ 1 \leq j \leq q,\ 1 \leq i \leq p$. In order to prove (35), we calculate

$$\begin{aligned}\sum_{i=1}^{p} \mathbb{E}\left[W_i^2 \middle| \mathcal{A}_{i-1}\right] =& n^2 \sum_{i=1}^{p} \mathbb{E}[X_i^2|\mathcal{A}_{i-1}] + \sum_{i=1}^{p}\sum_{j=1}^{q} n_j^2 \mathbb{E}[X_{j,i}^2|\mathcal{A}_{i-1}] - 2\sum_{i=1}^{p}\sum_{j=1}^{q} n_j n \mathbb{E}[X_i X_{j,i}|\mathcal{A}_{i-1}] \\ =& 2\sum_{j=1}^{q}\sum_{i=1}^{p} n_j^2 \frac{1}{n_j - i + 1} - 2n^2 \sum_{i=1}^{p} \frac{1}{n-i+1} + \tilde{\sigma}_{n,2}^2,\end{aligned}$$

where we used the fact $\mathbb{E}[X_{j,i}X_{j',i}|\mathcal{A}_{i-1}] = 0$ for different groups $j, j' \in \{1, \ldots, q\},\ j \neq j'$ and (16) and we define

$$\begin{aligned}\tilde{\sigma}_{n,2}^2 =& (\nu_4 - 3)\Bigg\{ n^2 \sum_{i=1}^{p} \frac{\operatorname{tr} \mathbf{P}(i-1)^{\odot 2}}{(n-i+1)^2} + \sum_{i=1}^{p}\sum_{j=1}^{q} n_j^2 \frac{\operatorname{tr} \mathbf{P}(j; i-1)^{\odot 2}}{(n_j - i+1)^2} \\ & - 2\sum_{i=1}^{p}\sum_{j=1}^{q} n n_j \frac{\operatorname{tr}\left(\tilde{\mathbf{P}}(j; i-1) \odot \mathbf{P}(i-1)\right)}{(n-i+1)(n_j-i+1)} \Bigg\}.\end{aligned}$$

Note that $\tilde{\sigma}_{n,2}^2/n^2 = o_{\mathbb{P}}(1)$ as $n \to \infty$ (similarly to Lemma 5.4 and Lemma 5.5). Using (19) and (32), we obtain

$$\begin{aligned}\sum_{i=1}^{p} \mathbb{E}\left[\left(\frac{W_i}{n}\right)^2 \middle| \mathcal{A}_{i-1}\right] =& \sum_{i=1}^{p} \log\left(1 - \frac{p}{n}\right) - \sum_{i=1}^{p}\sum_{j=1}^{q} \left(\frac{n_j}{n}\right)^2 \log\left(1 - \frac{p}{n_j}\right) \\ & + \sum_{j=1}^{q} \left(\frac{n_j}{n}\right)^2 \left\{\frac{1}{2n_j} - \frac{1}{2(n_j - p)}\right\} + \frac{\tilde{\sigma}_{n,2}^2}{n^2} + o(1) \\ =& \sigma_n^2 + o_{\mathbb{P}}(1),\ n \to \infty,\end{aligned}$$

which implies (35) by an application of Lemma 5.6. Note that the Lindeberg condition for the scheme $(W_i/(n\sigma_n))_{1 \leq i \leq p}$ can be shown similarly to (22) using Lemma 5.6. Combining (33) and (34), we conclude

$$\frac{2\left(\log \Lambda_{n,1} - \mu_n\right)}{n\sigma_n} = \frac{\sum\limits_{j=1}^{q}\sum\limits_{i=1}^{p} n_j X_{j,i} - n \sum\limits_{i=1}^{p} X_i}{n\sigma_n} + o_{\mathbb{P}}(1),\ n \to \infty.$$

By an application of Corollary 3.1 of Hall and Heyde (1980), we obtain

$$\frac{\sum\limits_{j=1}^{q}\sum\limits_{i=1}^{p} n_j X_{j,i} - n \sum\limits_{i=1}^{p} X_i}{n\sigma_n} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),\ n \to \infty.$$

The proof Theorem 3.1 concludes.

Acknowledgements. This work was supported by the German Research Foundation (DFG Research Unit 1735, DE 502/26-2, RTG 2131, *High-dimensional Phenomena in Probability - Fluctuations and Discontinuity* and project number 460867398, DFG Research unit 5381, *Mathematical Statistics in the Information Age*). The author would like to thank Tim Kutta for his helpful comments on an earlier version of this manuscript.

References

Anatolyev, S. and Yaskov, P. (2017). Asymptotics of diagonal elements of projection matrices under many instruments/regressors. *Econometric Theory*, 33(3):717–738.

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, second edition.

Bai, Z. and Silverstein, J. W. (2010). *Spectral analysis of large dimensional random matrices*, volume 20. Springer.

Bao, Z., Pan, G., and Zhou, W. (2015). The logarithmic law of random determinant. *Bernoulli*, 21(3):1600–1628.

Bodnar, T., Dette, H., and Parolya, N. (2019). Testing for independence of large dimensional vectors. *The Annals of Statistics*, 47(5):2977–3008.

Dette, H. and Dörnemann, N. (2020). Likelihood ratio tests for many groups in high dimensions. *Journal of Multivariate Analysis*, 178:104605.

Fan, J. and Li, R. (2006). Statistical challenges with high dimensionality: Feature selection in knowledge discovery. *Proceedings of the International Congress of Mathematicians, Madrid*, 3.

Gao, J., Han, X., Pan, G., and Yang, Y. (2017). High dimensional correlation matrices: The central limit theorem and its applications. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):677–693.

Girko, V. L. (1998). A refinement of the central limit theorem for random determinants. *Theory of Probability & Its Applications*, 42(1):121–129.

Guo, W. and Qi, Y. (2021). Asymptotic distributions for likelihood ratio tests for the equality of covariance matrices. *arXiv preprint arXiv:2110.02384*.

Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application.* Academic press.

Han, F., Chen, S., and Liu, H. (2017). Distribution-free tests of independence in high dimensions. *Biometrika*, 104(4):813–828.

He, Y., Xu, G., Wu, C., and Pan, W. (2021). Asymptotically independent u-statistics in high-dimensional testing. *The Annals of Statistics*, 49(1):154–181.

Heiny, J. and Parolya, N. (2021). Log determinant of large correlation matrices under infinite fourth moment. *arXiv preprint arXiv:2112.15388*.

Hu, J., Bai, Z., Wang, C., and Wang, W. (2017). On testing the equality of high dimensional mean vectors with unequal covariance matrices. *Annals of the Institute of Statistical Mathematics*, 69(2):365–387.

Ishii, R., Maruo, K., Noma, H., and Gosho, M. (2021). Statistical inference based on accelerated failure time models under model misspecification and small samples. *Statistics in Biopharmaceutical Research*, 13(4):384–394.

Jiang, H. and Wang, S. (2017). Moderate deviation principles for classical likelihood ratio tests of high-dimensional normal distributions. *Journal of Multivariate Analysis*, 156:57–69.

Jiang, T. and Qi, Y. (2015). Likelihood ratio tests for high-dimensional normal distributions. *Scandinavian Journal of Statistics*, 42(4):988–1009.

Jiang, T. and Yang, F. (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *The Annals of Statistics*, 41(4):2029–2074.

Johnstone, I. M. (2006). High dimensional statistical inference and random matrices. *Proceedings of the International Congress of Mathematicians, Madrid.*

Lemonte, A. (2016). *The gradient test: another likelihood-based test.* Academic Press.

Lemonte, A. J. (2013). On the gradient statistic under model misspecification. *Statistics & Probability Letters*, 83(1):390–398.

Loubaton, P. and Rosuel, A. (2020). Large random matrix approach for testing independence of a large number of gaussian time series. *arXiv preprint arXiv:2007.08806*.

Luo, X. and Tsai, W. Y. (2012). A proportional likelihood ratio model. *Biometrika*, 99(1):211–222.

Mestre, X. and Vallet, P. (2017). Correlation tests and linear spectral statistics of the sample correlation matrix. *IEEE Transactions on Information Theory*, 63(7):4585–4618.

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

Nguyen, H. H. and Vu, V. (2014). Random matrices: Law of the determinant. *The Annals of Probability*, 42(1):146–167.

O'Brien, P. (1992). Robust procedures for testing equality of covariance matrices. *Biometrics*, 48(3):819–827.

Parolya, N., Heiny, J., and Kurowicka, D. (2021). Logarithmic law of large random correlation matrix. *arXiv preprint arXiv:2103.13900*.

Qi, Y., Wang, F., and Zhang, L. (2019). Limiting distributions of likelihood ratio test for independence of components for high-dimensional normal vectors. *Annals of the Institute of Statistical Mathematics*, 71(4):911–946.

Schott, J. R. (2007). A test for the equality of covariance matrices when the dimension is large relative to the sample sizes. *Computational Statistics and Data Analysis*, 51:6535–6542.

Srivastava, M. S. and Yanagihara, H. (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *Journal of Multivariate Analysis*, 101:1319–1329.

Strug, L. J. (2018). The evidential statistical paradigm in genetics. *Genetic Epidemiology*, 42(7):590–607.

Wang, X., Han, X., and Pan, G. (2018). The logarithmic law of sample covariance matrices near singularity. *Bernoulli*, 24(1):80–114.

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.