technische universität dortmund

fakultät statistik

Inference for multivariate and high-dimensional data in heterogeneous designs

Dissertation zur Erlangung des Doktorgrades Dr. rer. nat. der Fakultät für Statistik der Technischen Universität Dortmund

Vorgelegt von

Paavo Aljoscha Nanosch Sattler

geboren in Herrenberg

Dortmund, March 2021

Present Dean of the Faculty:
Prof. Dr. Katja Ickstadt

Reviewers:
Prof. Dr. Markus Pauly
Prof. Dr. Philipp Doebler

Date of Defense:

Version February 15, 2022

Abstract

In the presented cumulative thesis, we develop statistical tests to check different hypotheses for multivariate and high-dimensional data. A suitable way to get scalar test statistics for multivariate issues are quadratic forms. The most common are statistics of Wald-type (WTS) or ANOVA-type (ATS) as well as centered and standardized versions of them. Also, Pauly et al. [2015] and Chen and Qin [2010] used such quadratic forms to analyze hypotheses regarding the expectation vector of high-dimensional observations. Thereby, they imposed different assumptions, but both allowed just one respective two groups.
We expand the approach from Pauly et al. [2015] to multiple groups, which leads to a multitude of possible asymptotic frameworks allowing even the number of groups to grow. In the considered split-plot-design with normally distributed data, we investigate the asymptotic distribution of the standardized centred quadratic form under different conditions. In most cases, we could show that between the limit distributions and the specific conditions exists an "if and only if" relation. For the frequently assumed case of equal covariance matrices, we also widen the considered asymptotic frameworks, since also settings with partially fixed sample sizes becomes part. Moreover, we add other cases in which the limit distribution can be calculated. These hold for homoscedasticity of covariance matrices but also for the general case.
This expansion of the asymptotic frameworks is one example on how the assumption of homoscedastic covariance matrices allows widening conclusions. Moreover, assuming equal covariance matrices also simplifies calculations and enables us to use a larger statistical toolbox. For the more general issue of testing hypotheses regarding covariance matrices, existing procedures have strict assumptions (e.g. see Muirhead [1982], Anderson [1984] and Gupta and Xu [2006]), test only special hypotheses (e.g. see Box [1953]), or are known to have low power (e.g. see Zhang and Boos [1993]). We introduce an intuitive approach with fewer restrictions, a multitude of possible null hypotheses, and a convincing small sample approximation. Thereby, nearly every quadratic form known from the mean-based analysis can be used, and two bootstrap approaches are applied to improve their performance. Furthermore, it can be expanded to many other situations like testing hypotheses of correlation matrices or check whether the covariance matrix has a particular structure.
We investigated the type-I-error for all developed tests and the power to detect deviations from the null hypothesis for small sample sizes up to large ones in extensive simulation studies.

Acknowledgements

First of all, I wish to thank my supervisor Professor Markus Pauly for giving me the chance to complete my Ph.D. and for all the support during this special part of my life, which helped me develop, both as a statistician and as a person. I know that I wasn't always the easiest Ph.D. student, so I am thankful for a supervisor who also saw me as an individuum as a whole instead of just a scientist. Further, I want to thank Professor Philipp Doebler for agreeing to be my second supervisor and to invest the related time and effort. The financial support of the DFG gives me the freedom to do research, for which I am also thankful.

I want to thank all of my colleagues who dared together with me the adventure of changing universities from Ulm to Dortmund and also the center of our lives. Mutual support made many things easier and gave the feeling of not being alone with the situation. But without all the nice and helpful new colleagues, students, office neighbors, and in general all the individuals, who welcomed us in this new environment, it would have been less fun and pleasant.

Special thanks go to Regina Stegherr, with who I initially shared an office and later on truly became friends with despite the approximate 500 km distance, which was between us in the second halves of our Ph.D.s.
Her support in all circumstances solved many problems and helped me to make the right decisions. Our deep friendship is one of the most important accomplishments during my time as Ph.D. and I am very glad and thankful for this.

I am blessed with a loving family that always supported me and helped me to find my way, even though this was not always easy. So from the bottom of my heart, I want to thank my parents and my sisters for their unconditional love, which gives me the strength to work on in many difficulties, in particularly when some simulation results were plaguing me.

All that I am, I am because of my mind.

Paavo Nurmi

Contents

Abstract **iii**

Acknowledgements **v**

Notation **xv**

List of Publications **xvii**

I Introduction and Statistical Background **1**

1 Introduction and Motivation **3**

2 Statistical Methods **9**
2.1 Model and Hypotheses 9
2.1.1 Model 9
2.1.2 Hypotheses 11
2.2 High-dimensionality 12
2.3 Quadratic forms 15
2.4 Resampling techniques 19
2.4.1 Parametric bootstrap 21
2.4.2 Wild bootstrap 22
2.4.3 Subsampling 22

3 Summary of the Articles **25**
3.1 Article 1: Inference for high-dimensional split-plot-designs 25
3.2 Extension of Article 1 to different dimensions 29
3.2.1 Appendix 31
3.3 Article 2: Manifold Asymptotics of Quadratic-Form-Based Inference 37
3.4 Article 3: Testing Hypotheses about Covariance Matrices 39
3.5 Extensions and further results based on Article 3 42
3.5.1 Simulating one more setting and another hypothesis . . . 43
3.5.2 Influence of the used hypothesis matrix 47
3.5.3 Testing hypotheses regarding the correlation matrix 56
3.5.4 Testing for covariance patterns 75

3.5.5 Efficient implementation of the developed tests 86
3.5.6 Different dimensions between the groups 89
3.5.7 Appendix 89

4 Discussion and Outlook **97**
4.1 Discussion 97
4.2 Outlook 98

II Publications **101**

Article 1 **103**

Article 2 **167**

Article 3 **192**

Bibliography **233**

List of Tables

2.1 Averaged values depending on the empirical covariance matrix in relation to the exact value, for different dimensions. The estimator is based on $n = 20$ normally distributed observations with $(\mathbf{\Sigma}_1)_{ij} = 0.6^{|i-j|}$ and was averaged over 10,000 simulation runs. . 13

3.1 Simulated type-I-error rates ($\alpha = 5\%$) in scenario A) ($\mathcal{H}_0^v : \mathbf{V}_1 = \mathbf{V}_2$) for ATS, WTS, MATS, Bartletts test and Box's M-test. The observation vectors have dimension 5, covariance matrix $(\mathbf{V})_{ij} = 0.6^{|i-j|}$ and there is always the same relation between group sample sizes with $n_1 := 0.6 \cdot N$ resp. $n_2 := 0.4 \cdot N$. 44

3.2 Simulated type-I-error rates ($\alpha = 5\%$) in scenario A) ($\mathcal{H}_0^v : \mathbf{V}_1 = \mathbf{V}_2$) for ATS, WTS, MATS, Bartletts test and Box's M-test. The observation vectors have dimension 5, covariance matrix $(\mathbf{V})_{ij} = 0.6^{|i-j|}$ and there is always the same relation between group samples size with $n_1 := 0.6 \cdot N$ resp. $n_2 := 0.4 \cdot N$. 45

3.3 Simulated type-I-error rates ($\alpha = 5\%$) in scenario A) ($\mathcal{H}_0^v : \mathbf{V}_1 = \mathbf{V}_2$) for ATS, WTS, MATS, Bartletts test and Box's M-test. The observation vectors have dimension 5, covariance matrix $\mathbf{V} = \mathbf{I}_5 + \mathbf{J}_5$ and there is always the same relation between group samples size with $n_1 := 0.6 \cdot N$ resp. $n_2 := 0.4 \cdot N$. 45

3.4 Simulated type-I-error rates ($\alpha = 5\%$) in scenario A) ($\mathcal{H}_0^v : \mathbf{V}_1 = \mathbf{V}_2$) for ATS, WTS, MATS, Bartletts test and Box's M-test. The observation vectors have dimension 5, covariance matrix $\mathbf{V} = \mathbf{I}_5 + \mathbf{J}_5$ and there is always the same relation between group samples size with $n_1 := 0.6 \cdot N$ resp. $n_2 := 0.4 \cdot N$. 46

3.5 Simulated type-I-error rates ($\alpha = 5\%$) in scenario F) ($\mathcal{H}_0^v : \mathbf{V}_1 = \mathbf{V}$) for ATS, WTS and MATS with 5-dimensional vectors and $(\mathbf{V})_{ij} = 0.6^{|i-j|}$. 48

3.6 Simulated type-I-error rates ($\alpha = 5\%$) in scenario F) ($\mathcal{H}_0^v : \mathbf{V}_1 = \mathbf{V}$) for ATS, WTS and MATS with 5-dimensional vectors and $\mathbf{V} = \mathbf{I}_5 + \mathbf{J}_5$. 49

3.7 Comparison of various ATS-test-statistics based on different hypothesis matrices with distinct number of rows by reference to congruity of the test decisions and different summary statistics for the difference of the p-values. The hypothesis is E) ($\mathcal{H}_0^v : \mathbf{V}_1 = \mathbf{V}_2 = \mathbf{V}_3$) for dimension $d = 5$ and $\boldsymbol{n} = (64, 40, 56)$. 52
3.8 Comparison of various ATS-test-statistics based on different hypothesis matrices with distinct number of rows by reference to congruity of the test decisions and different summary statistics for the difference of the p-values. The hypothesis is B) ($\mathcal{H}_0^v : \mathbf{V}_{111} = ... = \mathbf{V}_{155}$) for dimension $d = 5$ and $n = 50$. 53
3.9 Comparison of various ATS-test-statistics based on different hypothesis matrices with distinct number of rows by their type-I-error rate. The hypothesis is B) ($\mathcal{H}_0^v : \mathbf{V}_{111} = ... = \mathbf{V}_{155}$) for dimension $d = 5$ and $n = 50$ with covariance matrix $(\mathbf{V})_{ij} = 0.6^{|i-j|}$. 54
3.10 Comparison of various ATS-test-statistics based on different hypothesis matrices with distinct number of rows by reference to congruity of the test decisions, the average difference of the p-values, and the maximum difference of the p-values. The hypothesis is B) ($\mathcal{H}_0^v : \mathbf{V}_{111} = ... = \mathbf{V}_{155}$) for dimension $d = 5$ and $n = 50$ with covariance matrix $\mathbf{V} = \mathbf{I}_5 + \mathbf{J}_5$. 55
3.11 Simulated type-I-error rates ($\alpha = 5\%$) in scenario $A_\mathbf{r}$) ($\mathcal{H}_0^v : \mathbf{R}_1 = \mathbf{R}_2$) for ATS, WTS, Steiger's and Jennrich's test. The observation vectors have dimension 5, covariance matrix $(\mathbf{V}_1)_{ij} = 0.6^{|i-j|}$ resp. $\mathbf{V}_2 = \textit{diag}(1, 1.2, ..., 1.8) \cdot \mathbf{V}_1$ and and it always holds $n_1 := 0.6 \cdot N$ resp. $n_2 := 0.4 \cdot N$. 67
3.12 Simulated type-I-error rates ($\alpha = 5\%$) in scenario $A_\mathbf{r}$) ($\mathcal{H}_0^r : \mathbf{R}_1 = \mathbf{R}_2$) for ATS, WTS, Steiger's and Jennrich's test. The observation vectors have dimension 5, covariance matrices $(\mathbf{V}_1) = \mathbf{I}_5 + 0.5 \cdot \mathbf{J}_5$ resp. $\mathbf{V}_2 = \textit{diag}(1, 1.2, ..., 1.8) \cdot \mathbf{V}_1$ and it always holds $n_1 := 0.6 \cdot N$ resp. $n_2 := 0.4 \cdot N$. 68
3.13 Simulated type-I-error rates ($\alpha = 5\%$) in scenario $B_\mathbf{r}$) ($\mathcal{H}_0^r : \mathbf{r} = \mathbf{0}_{10}$) for ATS, WTS, Steiger's and Bartlett's test. The observation vectors have dimension 5 and covariance matrix $\mathbf{V} = \textit{diag}(1, 1.2, 1.4, 1.6, 1.8)$. 69
3.14 Number of observations for the different factor level combinations of sex and diagnosis. 73
3.15 P-values of ATS with parametric bootstrap for testing equality of correlation matrices and equality of covariance matrices. 74

3.16 Simulated type-I-error rates ($\alpha = 5\%$) for testing whether the covariance matrix has a Toeplitz structure, with ATS based on parametric bootstrap and based on Monte-Carlo simulation. The observation vectors have dimension 5, covariance matrix $(\mathbf{V})_{ij} = 1 - |i-j|/5$ and different distributions and sample sizes are considered. 82
3.17 Simulated type-I-error rates ($\alpha = 5\%$) for testing whether the covariance matrix has an autoregressive structure, with ATS based on parametric bootstrap and based on Monte-Carlo simulation. The observation vectors have dimension 5, covariance matrix $(\mathbf{V})_{ij} = 0.6^{|i-j|}$ and different distributions and sample sizes are considered. 82
3.18 P-values of one sample Hotelling's T^2 test and φ^*_{ATS} to check wheater the covariance matrix has a compound symmetry structure resp. a Toeplitz structure. 84
3.19 P-values of one sample Hotelling's T^2 test and φ^*_{ATS} to check wheater the covariance matrix has a compound symmetry structure resp. a Toeplitz structure. 85
3.20 Required time in seconds for various tests statistics and different dimensions for hypothesis C) ($\mathcal{H}_0^v : \mathrm{tr}(\mathbf{V}_1) = \mathrm{tr}(\mathbf{V}_2)$) with a quadratic hypothesis matrix on the left side and a non-quadratic hypothesis matrix on the right sight. Here some methods are used to increase efficiency. 88
3.21 Required time in seconds for various tests statistics and different dimensions for hypothesis A) ($\mathcal{H}_0^v : \{\mathbf{V}_1 = \mathbf{V}_2\}$) with a quadratic hypothesis matrix on the left side and a non-quadratic hypothesis matrix on the right sight. Here some methods are used to increase efficiency. 95
3.22 Required time in seconds for various tests statistics and different dimensions for hypothesis B) ($\mathcal{H}_0^v : \{\mathbf{V}_{111} = ... = \mathbf{V}_{1dd}\}$) with a quadratic hypothesis matrix on the left side and a non-quadratic hypothesis matrix on the right sight. Here some methods are used to increase efficiency. 96
3.23 Required time in seconds for various tests statistics and different dimensions for hypothesis D) ($\mathcal{H}_0^v : \{\mathrm{tr}(\mathbf{V}_1) = \gamma\}$) with a quadratic hypothesis matrix on the left side and a non-quadratic hypothesis matrix on the right sight. Here some methods are used to increase efficiency. 96

List of Figures

1.1 Prostaglandin-D-synthase (ß-trace) of 10 young women during 4 days under different sleep conditions. 3
1.2 Prostaglandin-D-synthase (ß-trace) of 10 young men during 4 days under different sleep conditions. 4

2.1 Number of publications with citation of the initial bootstrap paper of Efron [1979] divided according to the year of publication, retrieved at 04.01.2021. © Copyright Clarivate 2021. All rights reserved [3]. 20

3.1 Simulated power curves of different tests for the hypothesis $A_{\mathbf{r}}$) ($\mathcal{H}_0^r : \mathbf{R}_1 = \mathbf{R}_2$), with $d = 5$, $N = 100$ and $n_1 = 0.6 \cdot N$. The covariance matrix is $(\mathbf{V}_2)_{ij} = 0.6^{|i-j|}$ resp. $\mathbf{V}_1 = \mathbf{V}_2 + \delta \mathbf{J}_5$ and the error terms are based on skew normal distribution (left side) or gamma distribution (right side). 71
3.2 Simulated power curves of different tests for the hypothesis $B_{\mathbf{r}}$ ($\mathcal{H}_0^r : \mathbf{r}_1 = \mathbf{0}_{10}$), with $d = 5$ and $n_1 = 50$. The covariance matrix is $\mathbf{V}_1 = \mathbf{I}_5 + \delta \mathbf{J}_5$ and the error terms are based on skew normal distribution (left side) or gamma distribution (right side). 72

Notation

$[a, b]$	Closed interval, i.e., $t \in [a, b]$ if and only if $a \leqslant t \leqslant b$ with with $-\infty \leqslant a < b \leqslant \infty$
$(a, b]$	Left-open interval, i.e., $t \in (a, b]$ if and only if $a < t \leqslant b$ with $-\infty \leqslant a < b \leqslant \infty$
$\mathbb{N}$	Natural numbers
$\mathbb{N}_k$	$\{1, 2, ..., k\}$, i.e. natural numbers smaller or equal to $k \in \mathbb{N}$
$\boldsymbol{A}^\top$	Transpose of a matrix or vector $\boldsymbol{A}$
$\boldsymbol{A}^-$	Generalized inverse of a matrix $\boldsymbol{A}$
$\boldsymbol{A}^+$	Moore-Penrose inverse of a matrix $\boldsymbol{A}$
$\boldsymbol{A}_0$	Diagonal matrix containing the diagonal elements of the matrix $\boldsymbol{A}$
$\|\|\boldsymbol{v}\|\|$	Euclidean norm of the vector $\boldsymbol{v}$
$(\boldsymbol{v})_\ell$	l-th component of the vector $\boldsymbol{v}$
$\mathbf{I}_d$	$d \times d$ identity matrix, $d \in \mathbb{N}$
$\mathbf{1}_d$	d-dimensional column vector of 1's, $d \in \mathbb{N}$
$\mathbf{J}_d$	$\mathbf{J}_d = \mathbf{1}_d \mathbf{1}_d^\top$, i.e. a $d \times d$ matrix only containing 1's, $d \in \mathbb{N}$
$\mathbf{0}_d$	d-dimensional column vector of 0's, $d \in \mathbb{N}$
$\mathbf{0}_{m \times n}$	$m \times n$ matrix of 0's, $m, n \in \mathbb{N}$
$\mathbf{P}_d$	Centering matrix with dimension d, $\mathbf{P}_d = \mathbf{I}_d - \frac{1}{d}\mathbf{J}_d$, $d \in \mathbb{N}$
$\mathbb{R}_{\neq 0}$	Set of all real numbers without 0
$\otimes$	Kronecker product of matrices
$\oplus$	Direct sum of matrices
eigen()	Amount of eigenvalues of a quadratic matrix
tr()	Trace of a square matrix
rank()	Rank of a matrix
diag$(x_1, ..., x_d)$	d-dimensional diagonal matrix with values $x_1, ..., x_d$ on the diagonal
ker()	Kernel of a matrix

$\mathbb{1}$	Indicator function
$\xrightarrow{\mathcal{D}}$	Convergence in distribution
$\xrightarrow{\mathcal{P}}$	Convergence in probability
$X_n = o_P(1)$	X_n converges in probability to zero
i.i.d.	independent and identically distributed
$\mathbb{E}()$	Expectation value of a random variable
$Var()$	Variance of a random variable
$Cov()$	Covariance between two random variables
S_m	Symmetric group consisting of the permutations of the numbers $1, ..., n$

List of Publications

The following three papers and preprints are part of this cumulative thesis:

Article 1: Sattler, P. and Pauly, M. (2018). Inference for high-dimensional split-plot-designs: An unified approach for small to large numbers of factor levels. Electronic Journal of Statistics 12, 2743-2805, doi.org/10.1214/18-EJS1465.

Contribution of the author:
The author of this thesis conducted the mathematical proofs and implemented the simulations under Prof. Pauly's supervision. He also prepared and structured the manuscript with input from Prof. Pauly.

Article 2: Sattler, P. (2021). A comprehensive treatment of quadratic-form-based inference in repeated measures designs under diverse asymptotics.Electronic Journal of Statistics 15 (1) 3611 - 3634, 2021. doi.org/10.1214/21-EJS1865.

Contribution of the author:
The author of this thesis wrote the paper on his own. All mathematical proofs and simulations were conducted by the author of the thesis.

Article 3: Sattler, P., Bathke, A. C. and Pauly, M. (2022). Testing hypotheses about covariance matrices in general MANOVA designs. Journal of Statistical Planning and Inference. 219:134-146. doi.org/10.1016/j.jspi.2021.12.001

Contribution of the author:
As the first author of the paper, the author of this thesis developed the statistical frameworks based on Prof. Pauly's idea. He conducted the mathematical proofs as well as the extensive simulation studies under the

guidance of Prof. Pauly and Prof. Bathke. The author of this thesis had the leading role in drafting and writing the paper with input from all co-authors.

Part I

Introduction and Statistical Background

1 Introduction and Motivation

Multivariate statistical inference is an essential part of modern statistics. In contrast to univariate approaches, it allows us to consider various variables simultaneously and take their dependency structure into account. A fine example is repeated measure designs, where the same value is measured, i.e., d times, on one test subject or object repetitively. Thus, these d measurements are dependent.

To illustrate such a setting, we consider data from a sleep-laboratory (Jordan et al. [2004]). In this trial, the concentration of an enzyme (prostaglandin-D-synthase) in the blood of ten young men and ten young women was measured. The measurements were conducted every four hours over four days, with different sleep conditions (normal sleep, sleep deprivation, recovery sleep, and REM sleep deprivation) each night. The results of this trial are illustrated in Figure 1.1 (for the women) and Figure 1.2 (for the men).

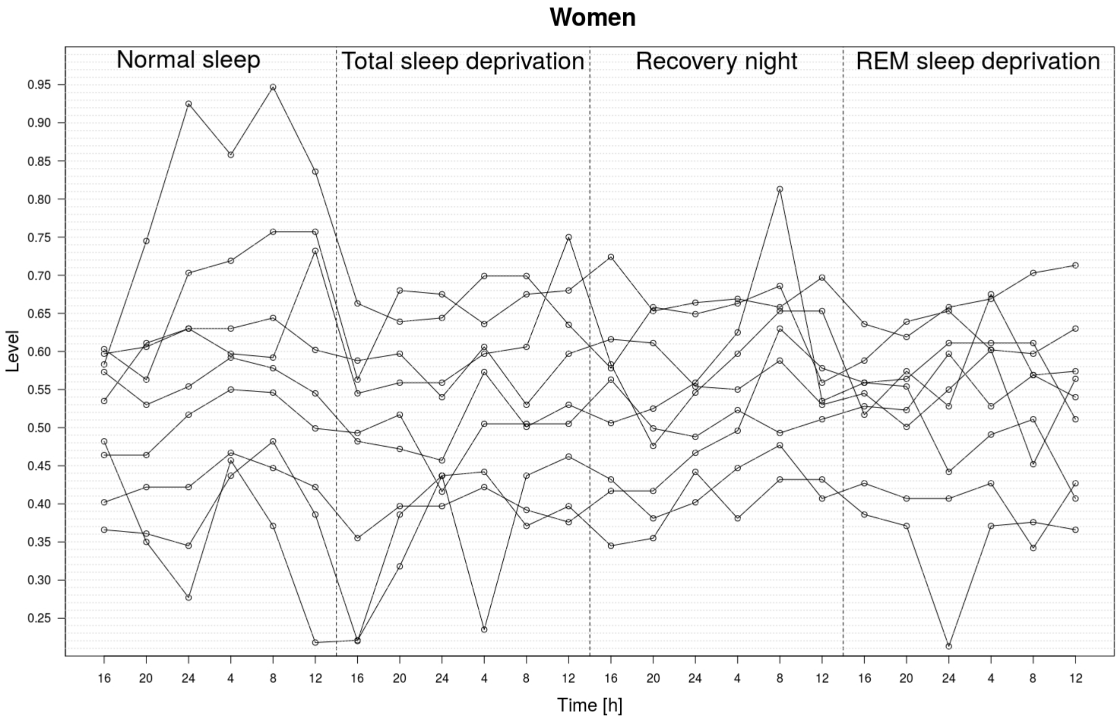


Figure 1.1: Prostaglandin-D-synthase (ß-trace) of 10 young women during 4 days under different sleep conditions.

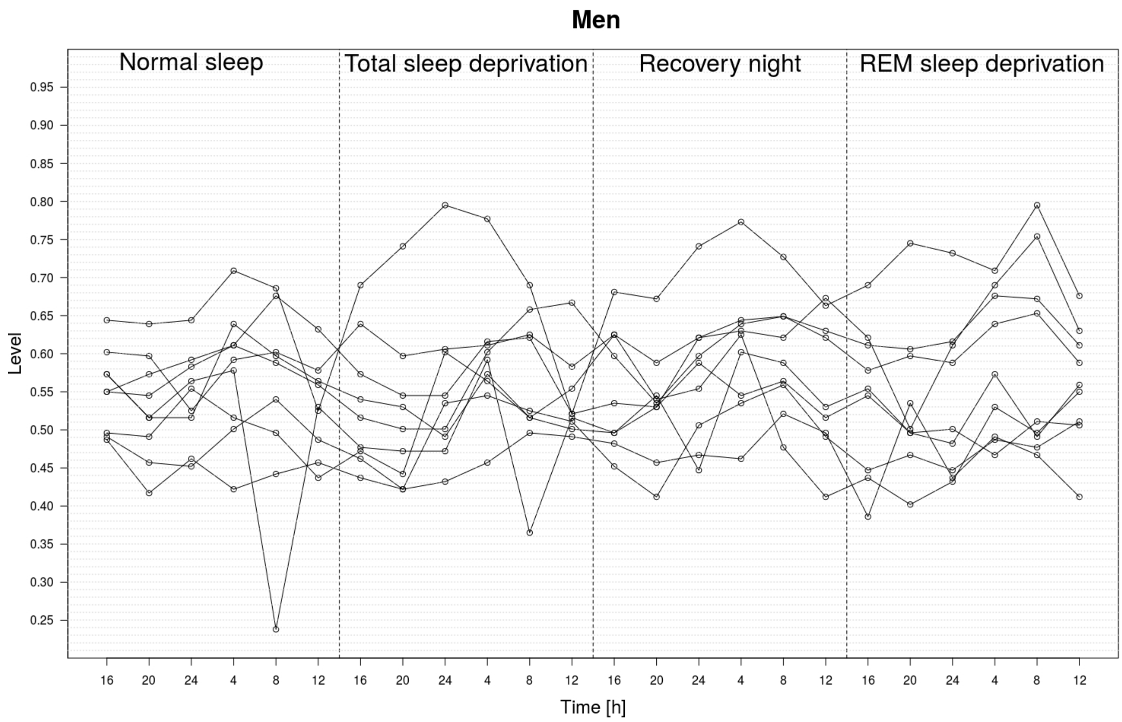


Figure 1.2: Prostaglandin-D-synthase (ß-trace) of 10 young men during 4 days under different sleep conditions.

Here, possible questions of interest are whether there is a gender effect, a time effect, an effect of the sleep condition, or interaction effects between these factors.

In such a repeated measure design, it is often easier or cheaper to increase the number of repetitions instead of the number of test subjects n. But, through the dependency between the measurements, it is frequently insufficient to only increase the number of measures and let the number of test subjects be fixed. Besides, so-called high-dimensional settings with $d > n$ are particularly demanding since usual techniques can often not be applied. A typical problem, in this case, is a singular empirical covariance matrix. Thus, the inverse, as used within Hotelling's T^2 test statistic, is not existing. This is particularly true for the sleep laboratory trial with 24 measurements but only ten test subjects in each group. For asymptotic frameworks with multiple increasing parameters such as sample size and dimension, relations between these parameters, like $d/n \to c \in [0, \infty)$, are often assumed. But such relations between dimension and sample size are usually difficult to verify in practice. In our sleep laboratory trial, a proper asymptotic framework for d/n is unclear. For this reason, conditions on the relation between dimension and sample sizes are often replaced by other parameters.

In Chen and Qin [2010] for this purpose τ_{CQ} is used, which is the relation between traces containing the covariance matrix and is required to go to zero for $n \to \infty$. In Pauly et al. [2015], a promising approach for normally distributed data was introduced, which also has no restriction on the relation between dimension d and sample size n and has lower demands on τ_{CQ}. This work just considered the one group case, while it is often more interesting to compare different groups, such as various treatments or gender. In contrast to Chen and Qin [2010], who consider only the case of two groups, an arbitrary number of groups, say a, should be allowed. Often only small numbers of groups are considered, although, in many areas like genetics or agriculture, comparison of a large number of groups is of great interest. For example, in Omer et al. [2000], seed potato plants infected with *Verticillium dahliae* (resulting in potato early dying) from 39 different isolates were analyzed regarding the aggressiveness of the infection. With nearly 40 groups, this is just one example from this area, where even higher numbers of groups are frequently.

Moreover, a large number of groups occur, among other things, if many different factors are investigated at the same time. Each factor (like gender, preexisting disease, age, or similar potentially influencing factor) allows to separate the examinees more precisely and therefore increases a substantially. Thus, in the below mentioned EEG (**E**lectro**E**ncephalo**G**ram) measurements with two factors, gender and diagnostic, and two resp. four manifestations, there are already eight groups. With the additional differentiation age (younger or older than 70 years), there would be even 16 groups.

For this reason, the number of groups should be allowed to go to infinity instead of being a fixed factor. Such an increasing number of groups was, for example, investigated in Bathke [2002]. With the possibly increasing number of groups, there are various potential asymptotic frameworks where the sample size and at least the dimension or the number of groups go to infinity. These frameworks are a great challenge since many existing asymptotic approaches or estimators are developed for concrete frameworks, while it would be preferable if we could use the same one in each framework.

Equal covariance matrices are a special case, which is frequently considered because this additional condition can simplify some calculations and allow some stronger results. This includes more general asymptotic frameworks, where the sample size does not necessarily have to grow. Furthermore, the inspection of such a more straightforward case allows for developing additional propositions, which can be expanded for the general setting.

Although equal covariances are a widespread assumption, it nevertheless should be verified in practice through a statistical test.
So, test statistics regarding covariance matrices are a reasonable approach to do this. However, in the year 1953, George Box described the role of a preliminary test on the variance as follows:

"To make the preliminary test on variances is rather like putting to sea in a rowingboat to find out whether conditions are sufficiently calm for an ocean liner to leave port".

With this statement, he described such tests' role quite visually: little tools that get less attention compared with the actual test. This role could be one reason why most of the existing methods for testing such hypotheses have limitations. Either they need assumptions that are difficult to justify in practice (see, e.g., Box [1953] or Gupta and Xu [2006]), or only allow single hypotheses like Zhu et al. [2002].
One example for the usage of a preliminary test on the variance is an EEG dataset included in the R package *manova.rm* (Friedrich et al. [2019]). In a study at the University Hospital Salzburg for 160 patients with different diagnoses of impairments, six EEG were recorded in different brain regions. In each of them, different measurements were conducted, like z-scores for brain rate. In Bathke et al. [2018] this dataset was investigated, and therefore the assumption of equal covariance matrices was questioned. Based on the empirical covariance matrix, this assumption seemed to be rather unlikely but was not checked statistically. To do so, an appropriate statistical test should be applied.

If one takes a closer look at this topic, hypotheses regarding covariance matrices offer many opportunities. Not only as a preliminary test for other approaches but also as an autonomous test procedure. For example, on the first day of the sleep laboratory trial, the groups seem to have different variability. This could be investigated by comparing their covariance matrices or employing other effect measurements based on these matrices, such as traces. Thereby, it is reasonable to allow null hypotheses that are as general as possible, similar to mean-based analysis. Thus, a semiparametric model and an approach comparable to the analysis of means is intuitive and simplifies the application. In contrast to the high-dimensional cases, we assume no normal distribution, and the components of the random vectors are allowed to be different kinds of data. Moreover, in this way, most of the techniques known from the mean-based analysis are transferable as quadratic forms as ATS (ANOVA-type-statistic) or bootstrap approaches. Finally, the combination of a rather general kind of null hypotheses and the construction similar to inference for mean vectors allows adaption for many other situations like testing hypotheses regarding correlation matrices or testing for a pattern or structure of the covariance matrix. This versatility allows the application in various settings and questions, especially in the field of psychology, where analysis of covariances is often of interest. For example, to analyze the reliability of a psychological test, there is a multitude of scores based on covariance matrices as Cronbach's α and many others (see, e.g., Cronbach [1951] or Guttman [1945]).

2 Statistical Methods

2.1 Model and Hypotheses

2.1.1 Model

The considered general semiparametric model is based on independent d-dimensional random vectors

$$\mathbf{X}_{ik} = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_{ik}, \tag{2.1}$$

where the index $i = 1, \ldots, a$ refers to the treatment group and $k = 1, \ldots, n_i$ to the individual, on which observations are measured. Within each of these a groups the residuals $\boldsymbol{\epsilon}_{i1}, \ldots, \boldsymbol{\epsilon}_{in_i}$ are i.i.d. (independent identically distributed) with $\mathbb{E}(\boldsymbol{\epsilon}_{i1}) = \mathbf{0}_d$, $Cov(\boldsymbol{\epsilon}_{i1}) = \boldsymbol{\Sigma}_i > 0$ and $\mathbb{E}(||\boldsymbol{\epsilon}_{i1}||^4) < \infty$. We allow for the different groups, different distributions with various parameters. This generality enables manifold application. In contrast,in a parametric repeated measure design all measurements have the same unit and therefore fixed distribution of the residuals $\boldsymbol{\epsilon}_{i1}, \ldots, \boldsymbol{\epsilon}_{in_i}$ is presumed. This assumption of a concrete distribution makes it much easier to obtain mathematical results but can be difficult to verify.

Here, in both models, no structure of the covariance matrices $\boldsymbol{\Sigma}_i$ are assumed, and depending on the specific situation, positive semi-definite covariance matrices can be sufficient.

The number of observation vectors in the i-th group, n_i are allowed to differ, as long as

$$\frac{n_i}{N} \to \kappa_i \in (0, 1] \quad i = 1, ..., a,$$

where $N = \sum_{i=1}^{a} n_i$ denotes the total sample size overall groups. This condition ensures, among other things, that all group sample sizes go to infinity and no single group dominates N. We have $\kappa_1 = 1$ if and only if $a = 1$, as $\kappa_i > 0$ holds for all i. By focussing on subsequences, this assumption could even be weakened to

$$0 < \liminf \left(\frac{n_i}{N}\right) \leqslant \limsup \left(\frac{n_i}{N}\right) \leqslant 1, \quad (i = 1, ..., a).$$

For simplicity, we define $n_{\min} := \min(n_1, ..., n_a)$ and $n_{\max} := \max(n_1, ..., n_a)$. The group specific expectation vector $\boldsymbol{\mu}_i$ can be estimated by the group mean $\overline{\mathbf{X}}_i = n_i^{-1}\sum_{k=1}^{n_i}\mathbf{X}_i$ while an estimator for the covariance matrix $\boldsymbol{\Sigma}_i$ is given by the empirical covariance matrix $\widehat{\boldsymbol{\Sigma}}_i = (n_i - 1)^{-1}\sum_{k=1}^{n_i}(\mathbf{X}_{ik} - \overline{\mathbf{X}}_i)(\mathbf{X}_{ik} - \overline{\mathbf{X}}_i)^\top$. To compare the groups, often the pooled expectation vector $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, ..., \boldsymbol{\mu}_a^\top)^\top$ is considered, as well as pooled mean vector $\overline{\mathbf{X}} = (\overline{\mathbf{X}}_1^\top, ..., \overline{\mathbf{X}}_a^\top)^\top$, while the covariance matrix of $\sqrt{N}\,\overline{\mathbf{X}}$, given through $\boldsymbol{\Sigma} = \oplus_{i=1}^a N/n_i \cdot \boldsymbol{\Sigma}_i$ can be estimated by $\widehat{\boldsymbol{\Sigma}} = \oplus_{i=1}^a N/n_i \cdot \widehat{\boldsymbol{\Sigma}}_i$. For the comparison of dependencies, in case of $\boldsymbol{\Sigma} > 0$ the correlation matrix of the i-th group $\mathbf{R}_i = (\boldsymbol{\Sigma}_i)_0^{-1/2}\boldsymbol{\Sigma}_i(\boldsymbol{\Sigma}_i)_0^{-1/2}$ should be considered. This matrix can be estimated by the empirical correlation matrix $\widehat{\mathbf{R}}_i = (\widehat{\boldsymbol{\Sigma}}_i)_0^{-1/2}\widehat{\boldsymbol{\Sigma}}_i(\widehat{\boldsymbol{\Sigma}}_i)_0^{-1/2}$, if $\boldsymbol{\Sigma}_i^{-1}$ exists. Here, $\boldsymbol{A}_0$ denotes the diagonal matrix given through $\boldsymbol{A}_0 = \mathit{diag}(a_{11}, ..., a_{dd})$ for a matrix $\boldsymbol{A} = (a_{ij})_{ij}^d$. In case of such symmetric $d \times d$ matrices $\boldsymbol{A} = (a_{ij})_{ij}^d$ like covariances, it is preferable to consider the half-vectorization operation vech, which extracts the vector $(a_{11}, a_{12}, ..., a_{1d}, a_{22}, ..., a_{2d}, ..., a_{dd})^\top$ with dimension $p = d(d+1)/2$.

In repeated measure designs, the components of the observation vectors are the different measurements. These take place at different time points, at different places of the body, or under different conditions. Therefore, different scales of the components make no sense. Moreover in the context of repeated measurements at various time points, the expectation vector is often split into its components $\alpha_i \in \mathbb{R}$ for the i-th group effect, $\beta_t \in \mathbb{R}$ for the time effect at time point t and $(\alpha\beta)_{it} \in \mathbb{R}$ for the (i, t)-interaction effect between group and time. Thus, the expectation values are expanded by

$$\mu_{i,t} = \mu + \alpha_i + \beta_t + (\alpha\beta)_{it}, \quad i = 1, \ldots, a;\ t = 1, \ldots, d,$$

with the conditions $\sum_i \alpha_i = \sum_t \beta_t = \sum_{i,t}(\alpha\beta)_{it} = 0$. By further splitting up the indices i in smaller subindices $i_1, i_2, ...$, we can include various other factors in this model. The resulting more complex factorial treatment structure is called split-plot-design and was first introduced in Fisher [1925]. Here, the name split-plot is based on plots from agricultural settings. The easy to change factors are the so-called sub-plots, while the larger factor, which is more difficult to change, is called the whole-plot. For example, this includes settings with combinations of different time points, different areas of the measuring points, and different treatments.

An example of such a setting is the sleep-laboratory-trial, which was presented in the previous section. Since the test subjects are ten women and ten men, there are $a = 2$ groups with $n_1 = n_2 = 10$ people in each of these groups. Measuring every four hours results in six repetitions for each sleep condition (normal sleep, sleep deprivation, recovery sleep, and REM sleep deprivation), which means an overall dimension of $d = 24$. These observation vectors consist of two crossed factors, sleep condition and time, so one measurement per test subject exists for each combination. Usual hypotheses of interest would be whether there is a time effect, a gender effect, or an interaction between gender and time. But the influence of the sleep condition within the group is interesting as well.

2.1.2 Hypotheses

In the classical mean-based analysis, hypotheses of the kind $\mathcal{H}_0 : \mathbf{H}\boldsymbol{\mu} = \mathbf{0}_m$ are considered, where $\mathbf{H} \in \mathbb{R}^{m \times ad}$ with $m \leqslant ad$ is an appropriate hypothesis matrix. Even for simple hypotheses like equality of means in two groups, it is clear that there exist many different possible matrices corresponding to this hypothesis. Therefore, a test decision would potentially depend on the chosen matrix, which essentially reduces the reliability. Fortunately, for this kind of hypothesis, there always exists a unique, symmetric and idempotent matrix $\mathbf{T} \in \mathbb{R}^{ad \times ad}$ with $\mathcal{H}_0 : \mathbf{H}\boldsymbol{\mu} = \mathbf{0}_m \Leftrightarrow \mathcal{H}_0 : \mathbf{T}\boldsymbol{\mu} = \mathbf{0}_{ad}$. This equality of hypotheses was shown, for instance, in Brunner and Puri [2001] and is the reason why it is a convention to use $\mathbf{T}$ to get a trustworthy and comparable result. Using the Moore-Penrose[1] inverse, this projection matrix $\mathbf{T}$ is given through $\mathbf{T} := \mathbf{H}^\top(\mathbf{H}\mathbf{H}^\top)^+\mathbf{H}$, whereas each kind of generalized inverse could be used. This matrix is the same for each kind of generalized inverse as it was shown, for example, in Rao and Mitra [1971]. While the generalized inverse or g-inverse of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is each matrix $\mathbf{A}^- \in \mathbb{R}^{n \times m}$ fulfilling

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A},$$

for the Moore-Penrose inverse $\mathbf{A}^+$ the conditions are stronger through

$$\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A} \quad \mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+ \quad (\mathbf{A}\mathbf{A}^+)^\top = \mathbf{A}\mathbf{A}^+ \quad (\mathbf{A}^+\mathbf{A})^\top = \mathbf{A}^+\mathbf{A}.$$

[1]This inverse is named after E.H. Moore and R. Penrose, who described the concept independently in Moore [1920] resp. Penrose [1955]. Unfortunately, the work of A. Bjerhammar, see for example Bjerhammar [1951], is usually not taken into account.

These additional conditions make the Moore-Penrose inverse a subset of the g-inverses. At the same time, the Moore-Penrose inverse is, in contrast to a g-inverse, unique.

However, we will demonstrate in Section 3.5.2 that for $\mathcal{H}_0 : \mathbf{H}\boldsymbol{\mu} = \boldsymbol{\zeta} \neq \mathbf{0}_m$ the equality of hypotheses is false. In the mean-based analysis, this is irrelevant because, with a subtraction of $\boldsymbol{\zeta}$ from our data, the necessary structure can always be ensured. But for hypotheses of similar types, as we use in Sattler et al. [2022], it is important to keep this in mind. Therein and especially in Section 3.5.2, we investigate the consequences of such non-unique matrices. We will expose that this point can create some difficulties but also provide an opportunity to save computation time.

In factorial repeated measure and split-plot-designs, a common model is given through $\mathbf{H} = \mathbf{H}_W \otimes \mathbf{H}_S$, where $\mathbf{H}_W \in \mathbb{R}^{a \times a}$ refers to the whole-plot and $\mathbf{H}_S \in \mathbb{R}^{d \times d}$ to the subplot, see for example Happ et al. [2016] and Kong and Harrar [2019]. This composition of the hypothesis matrix is plausible because all subplot factors are handled in the same way, and it simplifies several computations without being too restrictive. The corresponding projection matrix in this case is $\mathbf{T} = \mathbf{T}_W \otimes \mathbf{T}_S$, with proper projection matrices $\mathbf{T}_S$ and $\mathbf{T}_W$. These smaller projection matrices can be formed by $\mathbf{T}_S = \mathbf{H}_S^\top(\mathbf{H}_S\mathbf{H}_S^\top)^+\mathbf{H}_S$ resp. $\mathbf{T}_W = \mathbf{H}_W^\top(\mathbf{H}_W\mathbf{H}_W^\top)^+\mathbf{H}_W$ as it was shown in Sattler and Pauly [2018].

2.2 High-dimensionality

There are different opinions on what exactly is the definition of a high-dimensional asymptotic framework. All have in common that not only the sample size is assumed to go to infinity, but also the dimension of the observation vectors. Beyond that accordance, the way how the dimension d goes to infinity is controversial. Frequent assumptions are $d^3/N \to 0$ as in Huber [1973], $d/N \to c \in (0, 1)$ as in Bai and Saranadasa [1996], $d/N \to c \in (0, \infty)$ as in Harrar and Kong [2016], or $N = \mathcal{O}(d^\delta)$ for $1/2 < \delta < 1$ as in Srivastava et al. [2014]. Restrictions of this kind enable less restrictive conditions in other ways, but their justification is difficult in most cases. For this reason some work such as Chen and Qin [2010] or Pauly et al. [2015] completely avoids such assumptions.
In practice, often data sets with smaller sample sizes than dimension were re-

ferred to as high-dimensional data. Approaches such as Hotelling's T^2 test or classical linear regression models can not be used in such settings. For example, the earlier mentioned sleep laboratory trial with $d = 24$ and $n_1 = n_2 = 10$ is high-dimensional. This example demonstrates that high-dimensionality has to be considered even for comparably small dimensions and sample sizes, while the most common examples are genetic data with enormous dimensions.
It has to be taken into account that many common estimators or techniques can not be used in most of these high-dimensional settings. The reason is that mostly their convergence can not be guaranteed for an increasing dimension. A good way to illustrate this is through the empirical covariance matrix $\widehat{\boldsymbol{\Sigma}}$, which is a consistent estimator for the unknown covariance matrix $\boldsymbol{\Sigma}$ for fixed d and $n \to \infty$. Consequently, $\widehat{\boldsymbol{\Sigma}}$ is often used to estimate unknown values depending on this covariance matrix, like eigenvalues or traces. In Table 2.1 some values are calculated, and averaged over 10,000 simulation runs, based on 20 normally distributed observations with $(\boldsymbol{\Sigma}_1)_{ij} = 0.6^{|i-j|}$. The relation to the exact value shows that except for $\operatorname{tr}(\widehat{\boldsymbol{\Sigma}}_1)$, all estimators are quite inappropriate in a high-dimensional setting. As classical plug-in-estimators, these estimators could be used readily in most settings with fixed d and sufficient sample size.

d	$\frac{\operatorname{tr}(\widehat{\boldsymbol{\Sigma}}_1)}{\operatorname{tr}(\boldsymbol{\Sigma}_1)}$	$\frac{\operatorname{tr}\left(\widehat{\boldsymbol{\Sigma}}_1^2\right)}{\operatorname{tr}\left(\boldsymbol{\Sigma}_1^2\right)}$	$\frac{\max\left(\operatorname{eigen}\left(\widehat{\boldsymbol{\Sigma}}_1\right)\right)}{\max\left(\operatorname{eigen}\left(\boldsymbol{\Sigma}_1\right)\right)}$	$\frac{\max\left(\operatorname{eigen}\left(\widehat{\boldsymbol{\Sigma}}_1^2\right)\right)}{\max\left(\operatorname{eigen}\left(\boldsymbol{\Sigma}_1^2\right)\right)}$
5	0.999	1.200	1.072	1.252
10	1.000	1.325	1.203	1.539
25	1.001	1.693	1.624	2.721
50	1.001	2.320	2.264	5.225
100	1.000	3.551	3.355	11.375
200	1.000	6.021	5.298	28.227

Table 2.1: Averaged values depending on the empirical covariance matrix in relation to the exact value, for different dimensions. The estimator is based on $n = 20$ normally distributed observations with $(\boldsymbol{\Sigma}_1)_{ij} = 0.6^{|i-j|}$ and was averaged over 10,000 simulation runs.

This circumstance frequently disables bootstrap usage or other techniques, where parameters such as covariance matrices or eigenvalues thereof are required. For estimators in high-dimensional settings, the property of being dimensional-stable was introduced by Werner [2004] as follows:

Definition 2.2.1:
An array of estimators $\widehat{\theta}_{n,d} \in \mathbb{R}$ *of an unknown quantity* $\theta_{n,d} \in \mathbb{R}_{\neq 0}$ *is called dimensional-stable if the following conditions hold*

1. $\left|\mathbb{E}\left(\frac{\widehat{\theta}_{n,d}}{\theta_{n,d}} - 1\right)\right| \leqslant c_n,$
2. $Var\left(\frac{\widehat{\theta}_{n,d}}{\theta_{n,d}} - 1\right) \leqslant c_n,$

with $c_n \to 0$*, which is uniformly bounded in* d*.*

This property ensures that the estimator works reliably in settings with increasing dimension, although an estimator's consistency does not follow from its dimensional stability. But it directly follows $\widehat{\theta}_{n,d}/\theta_{n,d} \xrightarrow{\mathcal{L}^2} 1$, which we will call ratio-consistency. The necessity of this property was displayed in Table 2.1.

Another challenge dealing with high-dimensional data is the computation time, which is crucial for a large dimension in general. Fundamental mathematical methods like empirical covariance matrices or quadratic forms have a computational complexity of $\mathcal{O}(nd^2)$ resp. of $\mathcal{O}(d^2)$. This clarifies that the dimension is often more relevant for the computational effort than the sample size. Hence, large sample sizes together with high dimension make some adaption of algorithms necessary and have to be considered in the development of, for example, useful estimators.
This holds not only for high-dimensional settings but also for all situations with comparably high dimensions. For example, if vectorized matrices are considered, and therefore the dimension increases quadratically.
Also, settings with an increasing number of groups, $a \to \infty$ are sometimes included in a high-dimensional setting, even if the dimension is fixed. The reason for this is that in such a setting, the usage of pooled vectors like the pooled mean is often almost mandatory. So, in this case, the dimension ad of the pooled vector is increasing, which leads to a high-dimensional setting for this vector. Situations like this are for example investigated in Bathke [2002] or Bathke and Lankowski [2005].
A usual reason for increasing the number of groups can be a large number of crossed factors. Also, in settings where a is not fixed, relations between sample size and the number of groups resp. dimension and number of groups are difficult to justify. Therefore, a high-dimensional setting, where N, d and a are

allowed to increase independently, is preferable. Some limitations are given through $n_i/N \to \kappa_i \in (0,1]$ which was mentioned earlier, but can be relaxed sometimes, for example in case of homogeneity of covariance matrices as it can be seen in Sattler [2021] or in Section 3.2.

2.3 Quadratic forms

A good way to get a univariate test statistic for a hypothesis on multivariate parameter is by using a quadratic form. For a random vector $\mathbf{Y}$ of dimension ad with covariance matrix $\mathbf{V}$, a quadratic form is given through

$$Q = (\mathbf{HY})^\top \mathbf{E}(\mathbf{H},\mathbf{V})(\mathbf{HY}),$$

with $\mathbf{E}(\mathbf{H},\mathbf{V}) \in \mathbb{R}^{m\times m}$ a symmetric matrix which can depend on the matrices $\mathbf{H}$ and $\mathbf{V}$.

One of the central results for quadratic forms in random vectors is a representation theorem, which can be found, for example, in Mathai and Provost [1992] but is repeated here.

Theorem 2.3.1:
Let $\lambda_\ell, \ell = 1,\dots,ad$, *be the eigenvalues of the* $ad \times ad$ *matrix* $\mathbf{V}^{1/2}\mathbf{H}^\top\mathbf{E}(\mathbf{H},\mathbf{V})\mathbf{H}\mathbf{V}^{1/2}$. *Then it holds*

$$Q = \mathbb{E}(\mathbf{Y})^\top\mathbf{H}^\top\mathbf{E}(\mathbf{H},\mathbf{V})\mathbf{H}\mathbb{E}(\mathbf{Y}) + 2\sum_{\ell=1}^{ad}\left(\mathbf{H}^\top\mathbf{E}(\mathbf{H},\mathbf{V})\mathbf{H}\mathbb{E}(\mathbf{Y})\right)_\ell(\mathbf{OZ})_\ell + \sum_{\ell=1}^{ad}\lambda_\ell(\mathbf{OZ})_\ell^2.$$

Thereby, $\mathbf{Z}$ *fulfills* $\mathbf{Y} = \mathbb{E}(\mathbf{Y}) + \mathbf{V}^{1/2}\mathbf{Z}$ *and* $\mathbf{O} \in \mathbb{R}^{ad\times ad}$ *is the orthogonal matrix with* $\mathbf{O}^\top\mathbf{V}^{1/2}\mathbf{H}^\top\mathbf{E}(\mathbf{H},\mathbf{V})\mathbf{H}\mathbf{V}^{1/2}\mathbf{O} = \mathit{diag}(\lambda_1,....,\lambda_{ad})$.

In case of a centred vector $\mathbf{Y}$, this formula can be simplified, which is for example the case for $\mathcal{H}_0 : \mathbf{H}\cdot\mathbb{E}(\mathbf{Y}) = \mathbf{0}_{ad}$. This result is the basis for few results like formulas for expectation value and variance in general settings. For a normally distributed vector $\mathbf{Y}$, even formulas for all moments of quadratic forms can be concluded.

Most frequently, these quadratic forms are used for random vectors $\mathbf{Y}_n$ which fulfill $\mathbf{Y}_n \xrightarrow{\mathcal{D}} \mathcal{N}_{ad}(\mathbf{0}_{ad}, \mathbf{V})$ under the null hypothesis. Therefore we will focus on this case. In this situation, the asymptotic distribution of the quadratic form is a "weighted χ^2-distribution". Thus, it holds $Q \xrightarrow{\mathcal{D}} \sum_{\ell=1}^{ad} \lambda_\ell B_\ell$, where $B_\ell \overset{\text{i.i.d.}}{\sim} \chi_1^2$.

Since, in general, the covariance matrix $\mathbf{V}$ is unknown, an estimator $\widehat{\mathbf{V}}$ is used instead of estimating $\mathbf{E}(\mathbf{H}, \mathbf{V})$, and we obtain the estimated quadratic form $\widehat{Q}$. However, in general, the consistency of $\mathbf{E}(\mathbf{H}, \widehat{\mathbf{V}})$ does not follow from the consistency of $\widehat{\mathbf{V}}$. Therefore in some instances, conditions on the covariance matrix are necessary to get this needed consistency.

One of the most used quadratic forms, named after the statistician Abraham Wald, is the **W**ald-**T**ype-**S**tatistic which is defined through

$$WTS = (\mathbf{HY})^\top \left(\mathbf{HVH}^\top\right)^+ (\mathbf{HY}).$$

Since the matrix $\mathbf{V}^{1/2}\mathbf{H}^\top \left(\mathbf{HVH}^\top\right)^+ \mathbf{HV}^{1/2}$ is a projection matrix in this case, the limit distribution is $\chi^2_{rank(\mathbf{HVH}^\top)}$. In case of $\mathbf{V} > 0$ through $rank(\mathbf{HVH}^\top) = rank(\mathbf{H})$, the distribution is independent of the unknown covariance matrix of the original data. A statistic which distribution not depends on the unknown parameters is called pivot, and helpful for developing tests or confidence regions. This fact is the main reason for the usage of the WTS, but it also has the big advantage of being invariant against scale transformations. For usage of the empirical covariance matrix, conditions like $\mathbf{V} > 0$ are required, and it is a known fact that the WTS leads to highly liberal results except for very large sample sizes, see, e.g., Brunner et al. [1997] or Vallejo et al. [2010]. Therefore, Brunner and Puri [2001] introduced another quadratic form, called **A**nova-**T**ype-**S**tatistic where the matrix $\mathbf{E}(\mathbf{H}, \mathbf{V}) = \mathbf{I}_m / \operatorname{tr}(\mathbf{HVH}^\top)$ is used, which leads to

$$ATS = (\mathbf{HY})^\top (\mathbf{HY}) / \operatorname{tr}(\mathbf{HVH}^\top).$$

The trace as a denominator does not necessarily have to be used. Among other things, this scaled ATS is reasonable because it makes the quadratic form invariant against scalar multiplication. In a repeated measure design where all measurements are done in the same unit, this allows a change of the measuring unit without influencing the test statistic's value. But this is not the case in more general designs, with possibly quite different measuring units.

Although WTS and ATS are the most used quadratic forms by far, there is another interesting one, which can be seen as an intermediate level between them. The so-called MATS for **M**odified-**A**nova-**T**ype-**S**tatistic is given through

$$\mathsf{MATS} = (\mathbf{HY})^\top \left(\mathbf{HV}_0\mathbf{H}^\top\right)^+ (\mathbf{HY}),$$

where $\mathbf{A}_0 = \mathrm{diag}(a_{11}, ..., a_{dd})$ denotes the diagonal matrix containing the diagonal elements of the matrix $\mathbf{A} = (a_{ij})_{ij}^d$. This quadratic form was introduced in Srivastava and Kubokawa [2013] for particular settings and was later extended by Friedrich and Pauly [2017] for a general setting.
In contrast to the WTS, this test statistic is less liberal and requires $\mathbf{V}_0 > 0$ instead of $\mathbf{V} > 0$ for the usage of the empirical covariance matrix. On the other hand, it is invariant against scale transformation, but the asymptotic distribution can not be simplified and is therefore not pivot. As the limit distribution given through $A := \sum_{\ell=1}^{ad} \lambda_\ell B_\ell$ follows no known distribution with quantiles and similar quantities, a Monte-Carlo-based approach should be used to approximate these unknown quantities. To this end, the following steps were done:

1. The unknown covariance matrix $\mathbf{V}$ is estimated, and therefore also $\mathbf{V}^{1/2}\mathbf{H}^\top\mathbf{H}\mathbf{V}^{1/2}/\operatorname{tr}(\mathbf{HVH}^\top)$ resp. $\mathbf{V}^{1/2}\mathbf{H}^\top \left(\mathbf{HV}_0\mathbf{H}^\top\right)^+ \mathbf{H}^\top\mathbf{V}^{1/2}$ is estimated.

2. Based on this matrix the eigenvalues $\widehat{\lambda}_1, ..., \widehat{\lambda}_{ad}$ are estimated.

3. With ad independently generated χ_1^2 random variables one realization A_1 of the weighted sum can be calculated.

4. Repeat step 3 often, say B times, and calculate from these B realizations $A_1, ..., A_B$ the empirical level α quantile or similar quantities.

Usual choices for B are 10,000 while it depends on the dimension of the vector $\mathbf{Y}$, and therefore on the required computation time.

For a matrix $\mathbf{A}$, it holds $\mathrm{eigen}(\mathbf{A}^\top\mathbf{A}) \setminus \{0\} = \mathrm{eigen}(\mathbf{A}\mathbf{A}^\top) \setminus \{0\}$ and also the multiplicity of these eigenvalues is the same. Therefore, for the ATS we know $\sum_{\ell=1}^{ad} \lambda_\ell B_\ell \stackrel{\mathcal{D}}{=} \sum_{\ell=1}^{m} \lambda'_\ell B'_\ell$ with $B'_\ell \stackrel{\text{i.i.d.}}{\sim} \chi_1^2$ and $\lambda'_\ell \in \mathrm{eigen}(\mathbf{HVH}^\top/\operatorname{tr}(\mathbf{HVH}^\top))$. We use the latter version in our simulations because of some advantages in calculating, for example just m eigenvalues and random variables are necessary.

Since for $\sqrt{N}\,\overline{\mathbf{X}}$ the central limit theorem can be used to show the asymptotic normality under the null hypothesis $\mathcal{H}_0 : \mathbf{T}\boldsymbol{\mu} = \mathbf{0}$, this vector is the most frequently used for quadratic forms for tests about means. However, a variety of possible vectors can be used for quadratic forms, like relative effects see e.g. Akritas and Brunner [1997], vectorized Kaplan-Meier estimators in Dobler and Pauly [2020], group-specific survival medians in Ditzhaus et al. [2021] or for vectorized covariance matrices in Sattler et al. [2022].

Another way to develop a test based on the ATS is a Box-type-approximation, see e.g. Box et al. [1954], Brunner [2001] or Happ et al. [2016]. Hereby the estimated scaled ATS, given through $(\mathbf{HY})^\top(\mathbf{HY})/\operatorname{tr}(\mathbf{H}\widehat{\mathbf{V}}\mathbf{H}^\top)$, is approximated by a proper $F(\widehat{f},\widehat{f}_0)$-distribution. The estimators for the degrees of freedom are given by

$$\widehat{f} = \frac{\operatorname{tr}^2\left(\mathbf{T}\widehat{\mathbf{V}}\right)}{\operatorname{tr}\left(\left(\mathbf{T}\widehat{\mathbf{V}}\right)^2\right)} \quad \text{and} \quad \widehat{f}_0 = \frac{\operatorname{tr}^2\left(\mathbf{T}_0\widehat{\mathbf{V}}\right)}{\operatorname{tr}\left(\mathbf{T}_0^2\widehat{\mathbf{V}}^2\boldsymbol{\Lambda}^{-1}\right)}$$

with $\boldsymbol{\Lambda} = \mathit{diag}(n_1 - 1, ..., n_a - 1)$, see Brunner et al. [2019]. Since here the true distribution of the estimated scaled ATS under the null hypothesis is just approximated, we will not use this Box-type-approximation further.

U-Statistics

Since different powers of traces containing the covariance matrix are part of most moments of quadratic forms(see e.g. Mathai and Provost [1992]), it is essential to estimate such values. For this and a variety of other situations, U-statistics are a useful and intuitive approach of estimation and were first introduced by Hoeffding [1948]. The name comes from **U**nbiased, which is just one property of this kind of estimator. Thereby, each U-statistic is based on a real-valued function $h : \mathbb{R}^m \mapsto \mathbb{R}$, which is called the kernel of order $m \in \mathbb{N}$. Let again $X_1, ..., X_n$ be i.i.d. random variables with $\mathbb{E}(|h(X_1, ...X_m)|) < \infty$.
The associated U-statistic for a kernel h is defined through

$$U_n := \frac{(n-m)!}{n!} \sum_{1 \leqslant i_1 \neq i_2 \neq ... \neq i_m \leqslant n} h(X_{i_1}, ..., X_{i_m}).$$

For permutational-symmetric kernels this could be simplified to

$$U_n := \binom{n}{m}^{-1} \sum_{1 \leqslant i_1 < i_2 < ... < i_m \leqslant n} h(X_{i_1}, ..., X_{i_m}),$$

which is important since each kernel can be symmetrized by $h_s : \mathbb{R}^m \mapsto \mathbb{R}$ with

$$h_s(x_1, ..., x_m) = (m!)^{-1} \sum_{\pi \in S_m} h(x_{\pi(1)}, ..., x_{\pi(m)}).$$

Each U-statistic is an unbiased estimator for $\mathbb{E}(h(\mathbf{X}_1, ...\mathbf{X}_m))$ while under some conditions they are even the uniformly minimum-variance unbiased estimator. Many commonly used estimators are U-statistics, for example, the sample mean and the empirical covariance. There exist formulas for the variance of U-statistics, as well as results about their asymptotic distribution. Both can, for example, be found in Sproule [1974]. This definition can, without further ado, be expanded for random vectors $\mathbf{X}_1, ..., \mathbf{X}_n$, wherewith, for example, quadratic forms are possible kernels.
In the case of a permutation-symmetric kernel, the required number of summations corresponds to number of possibilies to choose m different indices from $1, ..., n$. Since the order has no influence, this number is given through $\binom{n}{m}$. Because this is a polynomial in n of degree m, the number can increase very fast. For computation time, this is one of the main challenges that can occur during U-statistics usage. We introduce one solution to this difficulty in the following chapter.

2.4 Resampling techniques

One possibility to handle limit distributions with unknown quantiles is to use resampling procedures, which is also advantageous in several cases. Essentially, two important parts of resampling procedures are used for this, bootstrap techniques and permutation techniques. In this thesis, we focus on the bootstrap approach, which was initially introduced by Efron [1979]. The term bootstrap comes from the part of the boots with the same name and presumably goes back to the phrase "pull oneself up by one's bootstraps". It means to improve his situation without external help [2].

[2] This strongly reminds of an anecdote of "Baron Munchausen", where he pulls himself out of a swamp by his own pigtail.

Since it was first mentioned in 1979, bootstrap became more and more popular, as illustrated in Figure 2.1. In 2005, Efron's paper was almost part of the 25 most cited statistical papers, see Ryan and Woodall [2005]. With, up to now, more than 8900 publications citing [3] the initial paper, bootstrap is an elementary part of modern statistics, whereby the progress towards efficient algorithms and computers had major influence.

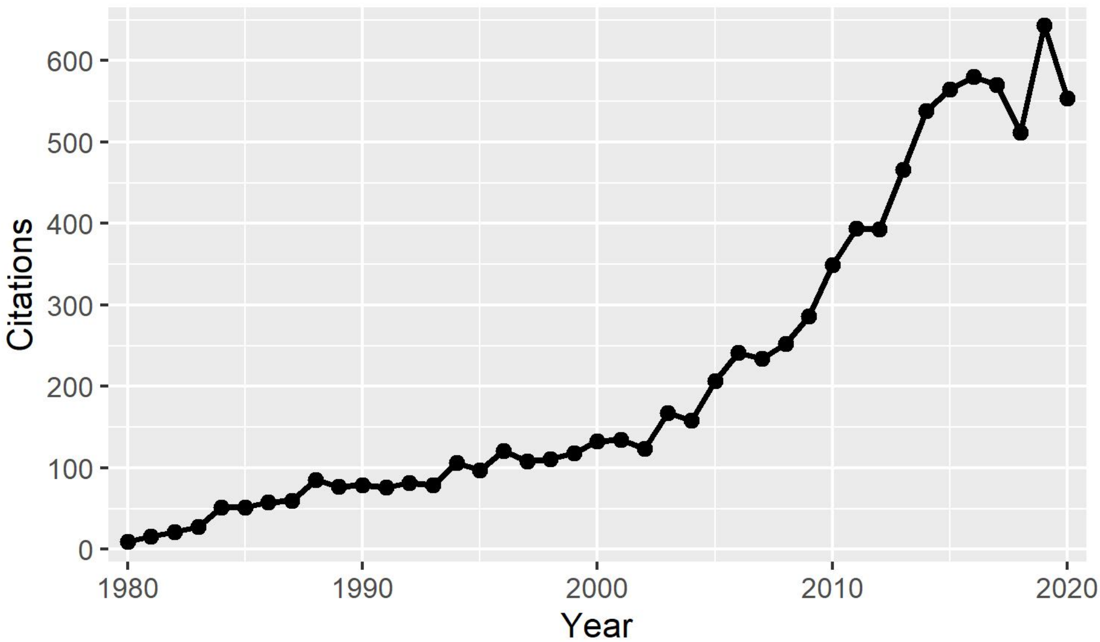


Figure 2.1: Number of publications with citation of the initial bootstrap paper of Efron [1979] divided according to the year of publication, retrieved at 04.01.2021.

The concept behind a bootstrap approach is that critical values based on the available data reflect reality better than critical values of limit distributions, which are independent of the received data.

We will only describe the resampling procedures that we will consider during this thesis: a parametric bootstrap approach, a wild bootstrap approach and a subsampling technique. Other methods like for example, Efron's bootstrap,

[3] The data are from "Web of Science" and were retrieved on 04.01.2021 Web of Science [2021].

will not be dealt with further.
In principle, the procedure is the same in all variations of resampling techniques, and we will take quadratic forms as an example. First, a bootstrap sample is generated, which depends on the real observed data $\mathbf{Y}_1, ..., \mathbf{Y}_n$ and has the same sample size as the original dataset. How to generate this resampling sample is the main difference between the distinct methods, together with the degree of dependence between the generated samples and the original sample. The generated bootstrap sample is used to replace a part of the original test statistic, like the real data, the centred data, or more complex structures depending on the original dataset. The result Q_1 is a bootstrap statistic, based on the data set. With a large number B of independent bootstrap samples and therefore bootstrap statistics $Q_1, ..., Q_B$, critical values can be calculated through empirical quantiles or similar quantities. With the original test statistic Q and this critical value, a resampling-based test can be built. For a useful bootstrap test, it is necessary that for each parameter $\boldsymbol{\nu} \in \mathbb{R}^{ad}$ and $\boldsymbol{\nu}_0 \in \mathbb{R}^{ad}$ fulfilling the null hypothesis $\mathbf{H}\boldsymbol{\nu}_0 = \boldsymbol{\zeta}$ it holds

$$\sup_{x \in \mathbb{R}} \left| P_{\boldsymbol{\nu}}(Q_1 \leqslant x | \mathbf{Y}) - P_{\boldsymbol{\nu}_0}(Q \leqslant x) \right| \xrightarrow{\mathcal{P}} 0.$$

Here, with $P_{\boldsymbol{\nu}}$, we denote the (un)conditional distribution of the quadratic form in the case $\boldsymbol{\nu}$ is the true underlying vector.
Overall, it is recommendable to replace a preferably basic part of the test statistic, like the original data, to mimic the original test statistic's structure best. On the contrary, replacing a more complex part often reduces computation time.
In addition to the calculation of unknown quantities, resampling procedures have one more main advantage. Due to the more direct connection to the real data, bootstrap tests usually require fewer observations than deterministic procedures to get the same quality of results, see, e.g., Friedrich and Pauly [2017].

2.4.1 Parametric bootstrap

As suggested by the name, this approach is mostly used in parametric settings. But it is also possible to use it in semiparametric or non-parametric settings. Here the resampling sample is generated through

$$\mathbf{X}^*_{ik} \sim F_i, i = 1, ..., a, \quad k = 1, ..., n_i,$$

where F_i is an appropriate multivariate distribution which usually depends on parameters of the real observations like the mean $\overline{\mathbf{X}}_i$ or the empirical covariance matrix $\widehat{\boldsymbol{\Sigma}}_i$. The used distribution should correspond as much as possible to the replaced part for good results. Naturally, this is much easier in parametric settings. Otherwise, there are two ways: On the one hand, an arbitrary distribution could be chosen, which is part of the model, whereby some distributions are more recommendable, through practical reasons. If this is impossible for some reason, instead, a part could be replaced by terms from its asymptotic distribution. Thus, the Monte-Carlo approach for the ATS is also a kind of parametric bootstrap, where we replace the quadratic form as a whole.

2.4.2 Wild bootstrap

The basic idea of a wild bootstrap is to generate N random weights, which are independent of the realizations, and use them to weight the original data or a more complex structure depending on the realizations. The used weights W_{ik}, $i = 1, ..., a$ $k = 1, ..., n_i$ are i.i.d. random variables with $\mathbb{E}(W_{11}) = 0$ and $Var(W_{11}) = 1$ while depending on the general setting, further assumptions on moments or similar are made. Each distribution fulfilling these requirements can be chosen for these weights, whereby common distributions are the standard normal distribution or the Rademacher distribution. As the original data's distribution does not have to be taken into account, this approach is especially attractive for non-parametric or semi-parametric settings. Moreover, in contrast to the parametric bootstrap, no estimated parameters are necessary for generating the random weights, which enhances the applicability. Finally, generating a random vector with a corresponding dependency structure is generally much more time-consuming than generating a random weight and multiplicating it with an observation vector. Depending on the sample size, the dimension, and similar factors, this leads to a clear difference in computation time, as can be seen in Section 6 of Sattler et al. [2022].

2.4.3 Subsampling

While for bootstrap and permutation techniques, a usual purpose is to calculate data-based values like quantiles, subsampling is additionally used for a different reason. In most instances, subsampling allows dealing with a data set that would otherwise be too extensive for computation time or memory space.

The idea is based on taking smaller subsamples of a greater superset, and we will introduce it by using a quadratic form as an example. Let U_n be a U-statistic based on a kernel h of order m, while we also consider the realisations of random variables $X_1, ..., X_n$. As mentioned before, the number of necessary summations $\binom{n}{m}$ can be rather high. So it is often not feasible to consider each index combination for the calculation of the U-statistic. Instead of this, m observations were drawn without replacement from the realizations. For these observations, the value of the kernel is calculated and notated as h_1. Then this step is repeated often, say $B \in \mathbb{N}$ times. The mean of these subsampled kernels $h_1, ..., h_B$ is the subsampling version of the U-statistic. If $B \to \infty$ for $N \to \infty$, this subsampling version mostly has the same asymptotic properties as the original one. The desired behavior of the repetitions B can be attained by $B = g \cdot N$ repetitions, with $g \in \mathbb{N}$. Instead of being a polynomial of degree m, the number of summations is then only linear in N. The proper choice of the factor g, or more general B, depends on the available computation time as well as on the number of summations for the deterministic U-statistic. The usage of a subsampling approach is just advisable if B is significantly smaller than $\binom{n}{m}$.

In general, we assume that the importance of subsampling will increase over the next years due to ever-growing data sets. The development in the last years further suggests this. In 2010 there were about 7.800 citations in Google Scholar [2021] of subsampling, in 2015 the number were about 12,000 and in 2020 the number of citations passed 17,000 [4].

[4] These numbers were retrieved at 04.01.2021 .

3 Summary of the Articles

3.1 Article 1: Inference for high-dimensional split-plot-designs: A unified approach for small to large numbers of factor levels

Let's look at the sleep laboratory trial from the beginning, where the data are displayed in Figure 1.1 and Figure 1.2. There are two groups, ten subjects, and six measurements under each sleep condition, which results in 24 time points for each person. Because of the two crossed factors, time and sleep condition, questions of interest are whether there is a gender effect, an effect of time, an effect of sleep condition, or corresponding interactions. This situation is a classical high-dimensional setting where many usual techniques like Hotellings T^2 can not be used to investigate these hypotheses.
Moreover, the data set and the experimental setup give no indications on the relation between sample size and dimension. This relation is necessary for most of the existing approaches, in addition to requirements on the structure of the covariance matrix.

One exception is the test of Chen and Qin [2010], which allows different distributions but requires $\tau_{CQ} = \operatorname{tr}((\mathbf{T\Sigma})^4)/\operatorname{tr}^2((\mathbf{T\Sigma})^2) \to 0$, and just considers the hypothesis of equal expectation vectors between two groups. However, for the most interesting hypothesis of no gender effect, $\tau_{CQ} \to 1$ seems more viable. In the case of normally distributed observations in Pauly et al. [2015], an approach handling $\tau_{CQ} \to 0$ and $\tau_{CQ} \to 1$ was introduced. The assumption of normally distributed observations is a common restriction to do without other ones and is, for example, also used in Harrar and Kong [2016] or Happ et al. [2016].

So we extend the approach of Pauly et al. [2015] to settings with two or more groups and consider even an increasing number of groups a in our asymptotic frameworks, which are then given through

$$\begin{aligned} a \in \mathbb{N} \text{ fixed} &\text{ and } \min(d, N) \to \infty, \\ d \in \mathbb{N} \text{ fixed} &\text{ and } \min(a, N) \to \infty, \\ &\text{ or } \min(a, d, N) \to \infty, \end{aligned}$$

with a as the number of independent groups, n_i the sample size of the i-th group and d-dimensional observation vectors.

These frameworks allow using our test for situations with many groups and therefore expand the applicability. The considered parametric model is given through independent random vectors $\mathbf{X}_{ik} \sim \mathcal{N}_d(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ with $\boldsymbol{\Sigma}_i > 0$ for $i = 1, ..., a$, and $k = 1, ..., n_i$. To develop an asymptotic level α-test, we consider a standardized quadratic form based on the pooled mean, which is given by

$$W_N = \frac{N \cdot \overline{\mathbf{X}}^\top \mathbf{T} \overline{\mathbf{X}} - \operatorname{tr}(\mathbf{T}\boldsymbol{\Sigma})}{\sqrt{2 \operatorname{tr}\left((\mathbf{T}\boldsymbol{\Sigma})^2\right)}}.$$

The asymptotic distribution of W_N under the null hypothesis was figured out in Theorem 3.1, where decreasing ordered standardized eigenvalues β_i of $\mathbf{T}\boldsymbol{\Sigma}\mathbf{T}$ are of particular specific importance. It holds that $W_N \xrightarrow{\mathcal{D}} Z \sim \mathcal{N}(0,1)$ if and only if $\beta_1 \to 0$, and $W_N \xrightarrow{\mathcal{D}} Z \sim (\chi_1^2 - 1)/\sqrt{2}$ if and only if $\beta_1 \to 1$. In extension of Pauly et al. [2015], we could prove the equivalence between the convergence of the largest standardized eigenvalue and the asymptotic distribution of the standardized quadratic form. This equivalence confirms that β_1 is a decisive value.

For this parameter β_1 it holds $\tau_{CQ} \to 0 \Leftrightarrow \beta_1 \to 0 \Leftrightarrow \tau_P \to 0$ and $\tau_{CQ} \to 1 \Leftrightarrow \beta_1 \to 1 \Leftrightarrow \tau_P \to 1$, with $\tau_P = \operatorname{tr}^3((\mathbf{T}\boldsymbol{\Sigma})^2)/\operatorname{tr}^2((\mathbf{T}\boldsymbol{\Sigma})^3)$. The case of $\tau_{CQ} \to 1$ is an important extension since in Pauly et al. [2015] it was shown that its behaviour depends partially just on the way how the dimension is increased.
Finally, Theorem 3.1 also includes the case where $\beta_i \to b_i \in [0,1]\ \forall i \in \mathbb{N}_{ad}$, with $\sum_{i=1}^\infty b_i^2 = 1$ and that the limit distribution in this case is an infinite sum of standardized and centred χ_1^2-variables with the weights given through b_i.

For this theorem's usage, it is necessary to develop proper estimators for the unknown traces, while the required computation time should not be too high for practical use. The development of estimators usable in each of our different asymptotic frameworks was one of the main difficulties for working in a multiple group setting. To this aim, we develop trace estimators based on symmetrized U-statistics, which are ratio-consistent in all of our asymptotic frameworks, regardless if the null hypothesis is true. Here, it is useful to estimate the different summands of the traces, not only for simplicity but also for practical reasons. Because the asymptotic frameworks include an increasing number of groups, it is reasonable to use an estimator allowing for additional groups without calculating everything new.

Concerning the potentially high computation time, we use a subsampling approach to take care of this. Thereby, the U-statistic is not calculated based on all index combinations but on a random subset of them.
The number of elements in this subset, which we denote as B, allows controlling the number of summations. As a consequence, it influences the needed time, and therefore should be chosen suitable for the situation.

This technique is especially important for developing an appropriate estimator for τ_{CQ} or τ_P, which is, among other things, necessary to examine the behavior of τ_{CQ}. Moreover, Pauly et al. [2015] showed that critical values based on $K_f = (\chi_f^2 - f)\sqrt{2f}$ improve the small sample properties, if $\beta_1 \to \{0,1\}$ and with meaningful degrees of freedom f. One choice is $f_{CQ} = \tau_{CQ}^{-1}$, while they used only $f_P = \tau_P^{-1}$. With f_P there is a concordance of the first three moments. Both f make a choice unnecessary whether a quantile of a standard normal distribution or a standardized χ_1^2 distribution is better for the test. Since in some tests f_P leads to better results, we also used it, and only shortly mentioned τ_{CQ} but never used it further.

An extensive simulation study with $N = 25$ (small), $N = 50$ (medium) and $N = 125$ (large) for increasing dimension d between 5 and 800 shows that in case of the normal distribution our test based on K_{f_P} performs mostly better than the test of Chen and Qin [2010] in case of $\tau_{CQ} \to 0$, $\tau_{CQ} \to 1$ or even $\tau_{CQ} \to b \in (0,1)$. Hereby, different covariance matrices and hypotheses were used, while the groups were unbalanced.

In summary, we developed estimators and an asymptotic test, usable for general split-plot-designs with heteroscedastic covariance settings for normally distributed observations. Here an extraordinary combination of asymptotic frameworks allows us to use the test in many situations without assumptions on the relation between the three parameters, the sample size N, the dimension d, and the number of groups a. The usage of a critical value based on K_{f_p} leads to a good small sample approximation and makes the distinction between $\beta_1 \to 0$ and $\beta_1 \to 1$ unnecessary. Finally, also for $\beta_1 \to b_1 \in (0,1)$, the type-I-error rate in simulations was good, while there is no theoretical evidence.

3.2 Extension of Article 1 to different dimensions

The model of Sattler and Pauly [2018] allows testing of hypotheses regarding different groups of repeated measurements with the same dimension. However, there are also situations where not all groups have the same number of measurements and, therefore, the same dimension. So in two studies, there could be different measurement points in the sleep laboratory or different numbers of subplot factors. This setting also includes questionnaires with group-specific questions and, therefore, different lengths. Moreover, this allows investigating the influence of the number of repetitions on the data. The impact of a questionnaire's length on the results is a popular topic, see, e.g., Roszkowski and Bean [1990] or Hallal et al. [2004]. But it also enables, for example, to investigate whether the number of animals in a litter has an influence on the respective development status. It is clear that in such situations, some hypotheses like equality of means make no sense.

We want to adapt the model and the results of Sattler and Pauly [2018] for the more general setting, also considered in Friedrich et al. [2017]. So we assume

$$\mathbf{X}_{i,j} = (X_{i,j,1}, \ldots, X_{i,j,d_i})^\top \overset{ind}{\sim} \mathcal{N}_{d_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}) \;\; j = 1, \ldots, n_i, \;\; i = 1, \ldots, a,$$

and define the dimension of the pooled mean vector by $D = \sum_{i=1}^{a} d_i$. Since a Kronecker product of wholeplot and subplot matrix can not be used in a model allowing groups with different dimension, the considered hypothesis matrix and therefore the null hypotheses changes. Therefore, a block matrix

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_{11} & ... & \mathbf{T}_{1a} \\ \vdots & \ddots & \vdots \\ \mathbf{T}_{a1} & \cdots & \mathbf{T}_{aa} \end{pmatrix} \in \mathbb{R}^{D \times D}$$

which is idempotent and symmetric with components $\mathbf{T}_{ij} \in \mathbb{R}^{d_i \times d_j}$ is used to formulate our null hypothesis through $\mathcal{H}_0 : \mathbf{T}\boldsymbol{\mu} = \mathbf{0}_D$. The parts of the hypothesis matrix need not be quadratic, neither do they have to be idempotent or symmetric. However, through the symmetry of $\mathbf{T}$ it holds $\mathbf{T}_{ij} = \mathbf{T}_{ji}^\top$ for $i, j \in \mathbb{N}_a$. With the notations from Sattler and Pauly [2018] we find $\sqrt{N}\,\mathbf{T}\overline{\mathbf{X}} \overset{H_0}{\sim} \mathcal{N}_D(\mathbf{0}_D, \mathbf{T}\mathbf{V}_N\mathbf{T})$ and define the standardized quadratic form $\widetilde{W}_N$. In contrast to Sattler and Pauly [2018] for estimation of the unknown expectation value and variance of the quadratic form we will not consider the individual traces, which are part of $\operatorname{tr}(\mathbf{T}\mathbf{V}_N\mathbf{T})$ resp. $\operatorname{tr}\left((\mathbf{T}\mathbf{V}_N\mathbf{T})^2\right)$. The main reason is as previously mentioned, not all matrices $\mathbf{T}_{ij}$ are quadratic, which was part of many proofs in Sattler and Pauly [2018] to verify the properties of the used estimators. Instead we estimated the whole trace, similarly as it was done for $\mathbf{C}_1$ in Sattler and Pauly [2018]. To this aim we again use the D-dimensional random vectors Z,

$$\mathbf{Z}_{(\ell_1,\ell_2,\dots,\ell_{2a})} := \left(\sqrt{\frac{N}{n_1}}\left(\mathbf{X}_{1,\ell_1}-\mathbf{X}_{1,\ell_2}\right)^\top,\dots,\sqrt{\frac{N}{n_a}}\left(\mathbf{X}_{a,\ell_{2a-1}}-\mathbf{X}_{a,\ell_{2a}}\right)^\top\right)^\top,$$

for $\ell_1 \neq \ell_2 \in \mathbb{N}_{n_1}, \dots, \ell_{2a-1} \neq \ell_{2a} \in \mathbb{N}_{n_a}$. Then

$$C_1 = \sum_{\substack{\ell_{1,1},\ell_{2,1}=1\\ \ell_{1,1}\neq\ell_{2,1}}}^{n_1} \cdots \sum_{\substack{\ell_{1,a},\ell_{2,a}=1\\ \ell_{1,a}\neq\ell_{2,a}}}^{n_a} \frac{\mathbf{Z}_{(\ell_{1,1},\ell_{2,1},\dots,\ell_{1,a},\ell_{2,a})}^\top \mathbf{T}\mathbf{Z}_{(\ell_{1,1},\ell_{2,1},\dots,\ell_{1,a},\ell_{2,a})}}{2\cdot\prod_{i=1}^{a}\frac{n_i!}{(n_i-2)!}}$$

and

$$C_2 = \sum_{\substack{\ell_{1,1},\dots,\ell_{4,1}=1\\ \ell_{1,1}\neq\dots\neq\ell_{4,1}}}^{n_1} \cdots \sum_{\substack{\ell_{1,a},\dots,\ell_{4,a}=1\\ \ell_{1,a}\neq\dots\neq\ell_{4,a}}}^{n_a} \frac{\left[\mathbf{Z}_{(\ell_{1,1},\ell_{2,1},\dots,\ell_{1,a},\ell_{2,a})}^\top \mathbf{T}\mathbf{Z}_{(\ell_{3,1},\ell_{4,1},\dots,\ell_{3,a},\ell_{4,a})}\right]^2}{4\cdot\prod_{i=1}^{a}\frac{n_i!}{(n_i-4)!}}$$

are our new estimators used for the expectation respective the variance.

The assumption of $q > 0$ with $n_{\min} = \mathcal{O}(a^q)$ or similar conditions make sure that $\prod_{i=1}^{a}\frac{(n_i-2)!\cdot(n_i-2)!}{n_i!(n_i-4)!} \to 1$ holds. Then, these estimators have the same properties as $\widehat{E_{H_0}}(Q_N)$ resp. A_4 from Sattler and Pauly [2018]. The proof of this statement can be found in the appendix together with the subsampling version of these estimators. Again a too high number of necessary summations for C_1 and C_2 can make the usage of subsampling attractive.

So the substitution of the estimators used in Sattler and Pauly [2018] by these new ones makes all their results valid for this changed setting.

Finally, it remains to reconsider the so far used asymptotic frameworks for the case of different dimensions between the groups. An important element of the used approach is that the dimension of the pooled mean vector goes to infinity.

While for equal dimensions, this was ad and therefore needed either the dimension or number of groups to go to infinity, in this new setting ad is replaced by $D = \sum_{i=1}^{a} d_i$. Thus, the number of groups or at least one dimension has to go to infinity. This allows for very unbalanced dimension and also settings with fixed dimension in some groups and increasing dimension in other ones, which can be seen as semi-high-dimensional. This way, data sets from trials with fixed dimensions can be compared with high-dimensional data sets. To our knowledge, such a comparison has not been part of other papers yet.

But this is not the only way where the original setting from Sattler and Pauly [2018] can be generalized. As mentioned at the beginning, a usual condition for designs with several groups is

$$\frac{n_i}{N} \to \kappa_i \in (0,1) \quad i = 1, ..., a,$$

which is required to get a reasonable covariance matrix for the limit distribution of $\sqrt{N}\,\overline{\mathbf{X}}$. Sattler and Pauly [2018] do not need this assumption with the usage of the standardized quadratic form and the representation theorem. For the convergence of the resulting weighted sum, only the behavior of β_i $(i = 1, ..., ad)$ is of importance. Further, the sample size of each group must go to infinity.
So the new, more general frameworks are

$$\begin{aligned}
& a \in \mathbb{N} \text{ fixed} && \text{and} && \min(\max(d_1, ..., d_a), n_1, ..., n_a) \to \infty,\\
\forall i \in \mathbb{N}_a \quad & d_i \in \mathbb{N} \text{ fixed} && \text{and} && \min(a, n_1, ..., n_a)) \to \infty,\\
& && \text{or} && \min(a, \max(d_1, ..., d_a), n_1, ..., n_a) \to \infty,
\end{aligned}$$

especially including the semi-high-dimensional settings.

3.2.1 Appendix

The new estimators C_1 and C_2 can be seen as a combination of the old ones $A_{i,1}$ and $A_{i,3}$ with the pooled vector $\mathbf{Z}$ and the approach introduced for C_5. Similar to C_5 and as mentioned in Section 2.3(U-Statistics) the number of necessary summations, namely $\prod_{i=1}^{a} \frac{n_i!}{(n_i-2)!}$ resp. $\prod_{i=1}^{a} \frac{n_i!}{(n_i-4)!}$, can increase really fast. Therefore, the simulation becomes comparatively time consuming and the usage of a subsampling approach, as introduced in Section 2.4.3, is reasonable.

This technique was already used for some of the other estimators in Sattler and Pauly [2018] and needs B independently drawn random subsamples. For each $i = 1, ..., a$ and $b = 1, ..., B$ these subsamples $\sigma_{1i}(b), \sigma_{2i}(b)$ of length two or $\sigma_{1i}(b), ..., \sigma_{4i}(b)$ of length four are drawn without replacement from n_i and used to define vectors $\boldsymbol{\sigma}(b) = (\sigma_{11}(b), ..., \sigma_{2a}(b))$ resp. $\boldsymbol{\sigma}(b) = (\sigma_{11}(b), ..., \sigma_{4a}(b))$. With this subsampling vectors the subsampling version can be defined.

Theorem 3.2.1:

a) *With the two kernels*

$$\Lambda_4(\ell_{1,1}, \ldots, \ell_{2,a}) = \mathbf{Z}_{(\ell_{1,1},\ell_{2,1},\ldots,\ell_{1,a},\ell_{2,a})}{}^\top \mathbf{T}\mathbf{Z}_{(\ell_{1,1},\ell_{2,1},\ldots,\ell_{1,a},\ell_{2,a})},$$
$$\Lambda_5(\ell_{1,1}, \ldots, \ell_{4,a}) = \left[\mathbf{Z}_{(\ell_{1,1},\ell_{2,1},\ldots,\ell_{1,a},\ell_{2,a})}{}^\top \mathbf{T}\mathbf{Z}_{(\ell_{3,1},\ell_{4,1},\ldots,\ell_{3,a},\ell_{4,a})}\right]^2,$$

based on quadratic forms, we define the corresponding U-statistics through

$$C_1 = \sum_{\substack{\ell_{1,1},\ell_{2,1}=1\\ \ell_{1,1}\neq\ell_{2,1}}}^{n_1} \cdots \sum_{\substack{\ell_{1,a},\ell_{2,a}=1\\ \ell_{1,a}\neq\ell_{2,a}}}^{n_a} \frac{\Lambda_4(\ell_{1,1}, \ldots, \ell_{2,a})}{2 \cdot \prod_{i=1}^{a} \frac{n_i!}{(n_i-2)!}}$$

and

$$C_2 = \sum_{\substack{\ell_{1,1},\ldots,\ell_{4,1}=1\\ \ell_{1,1}\neq\ldots\neq\ell_{4,1}}}^{n_1} \cdots \sum_{\substack{\ell_{1,a},\ldots,\ell_{4,a}=1\\ \ell_{1,a}\neq\ldots\neq\ell_{4,a}}}^{n_a} \frac{\Lambda_5(\ell_{1,1}, \ldots, \ell_{4,a})}{4 \cdot \prod_{i=1}^{a} \frac{n_i!}{(n_i-4)!}}.$$

b) *In addition the subsampling version of these estimators are given by*

$$C_1^\star(B) = \frac{1}{2 \cdot B} \sum_{b=1}^{B} \Lambda_4(\boldsymbol{\sigma}(b, 2)),$$

and

$$C_2^\star(B) = \frac{1}{4 \cdot B} \sum_{b=1}^{B} \Lambda_5(\boldsymbol{\sigma}(b, 4)),$$

with $B \in \mathbb{N}$ *the number of subsampling steps.*
Then, if some $q > 0$ *exists with* $n_{\min} = \mathcal{O}(a^q)$ *and* $B \to \infty$, C_1 *and* $C_1^\star$ *are*

unbiased and dimensional-stable estimators for $\operatorname{tr}(\mathbf{TV}_N)$. *Under these conditions also* C_2 *and* $C_2^\star$ *are unbiased and dimensional-stable estimators for* $\operatorname{tr}\left((\mathbf{TV}_N)^2\right)$.

Proof:

a) With the formulas for moments of the quadratic forms from Sattler et al. [2022] and an adaption of estimators therein we calculate

$$\begin{aligned}
\mathbb{E}(C_1) &= \frac{1}{2\cdot\prod\limits_{i=1}^{a}\frac{n_i!}{(n_i-2)!}} \sum_{\substack{\ell_{1,1},\ell_{2,1}=1\\ \ell_{1,1}\neq\ell_{2,1}}}^{n_1} \cdots \sum_{\substack{\ell_{1,a},\ell_{2,a}=1\\ \ell_{1,a}\neq\ell_{2,a}}}^{n_a} \mathbb{E}\left(\Lambda_4(\ell_{1,1},\dots,\ell_{2,a})\right)\\
&= \frac{1}{2\cdot\prod\limits_{i=1}^{a}\frac{n_i!}{(n_i-2)!}} \sum_{\substack{\ell_{1,1},\ell_{2,1}=1\\ \ell_{1,1}\neq\ell_{2,1}}}^{n_1} \cdots \sum_{\substack{\ell_{1,a},\ell_{2,a}=1\\ \ell_{1,a}\neq\ell_{2,a}}}^{n_a} \operatorname{tr}\left(2\mathbf{TV}_N\mathbf{T}\right)\\
&= \operatorname{tr}\left(\mathbf{TV}_N\mathbf{T}\right)
\end{aligned}$$

and

$$\begin{aligned}
&\mathit{Var}(C_1)\\
=\ &\sum_{\substack{\ell_{1,1},\ell_{2,1}=1\\ \ell_{1,1}\neq\ell_{2,1}}}^{n_1} \cdots \sum_{\substack{\ell_{1,a},\ell_{2,a}=1\\ \ell_{1,a}\neq\ell_{2,a}}}^{n_a} \sum_{\substack{\ell'_{1,1},\ell'_{2,1}=1\\ \ell'_{1,1}\neq\ell'_{2,1}}}^{n_1} \cdots \sum_{\substack{\ell'_{1,a},\ell'_{2,a}=1\\ \ell'_{1,a}\neq\ell'_{2,a}}}^{n_a} \frac{\mathit{Cov}\left(\Lambda_4(\ell_{1,1},\dots,\ell_{2,a}),\Lambda_4(\ell'_{1,1},\dots,\ell'_{2,a})\right)}{\left(2\cdot\prod\limits_{i=1}^{a}\frac{n_i!}{(n_i-2)!}\right)^2}\\
\leqslant\ &\frac{\prod\limits_{i=1}^{a}\frac{n_i!}{(n_i-2)!}-\prod\limits_{i=1}^{a}\frac{(n_i-2)!}{(n_i-4)!}}{2\cdot\prod\limits_{i=1}^{a}\frac{n_i!}{(n_i-2)!}}\,\mathit{Var}\left(\Lambda_4(1,\dots,2)\right)\\
=\ &\frac{\prod\limits_{i=1}^{a}\frac{n_i!}{(n_i-2)!}-\prod\limits_{i=1}^{a}\frac{(n_i-2)!}{(n_i-4)!}}{\prod\limits_{i=1}^{a}\frac{n_i!}{(n_i-2)!}}\cdot\mathcal{O}\left(\operatorname{tr}^2\left(\mathbf{TV}_N\right)\right).
\end{aligned}$$

Similar we get

$$\begin{aligned}
\mathbb{E}(C_2) &= \frac{1}{4\cdot\prod\limits_{i=1}^{a}\frac{n_i!}{(n_i-4)!}} \sum_{\substack{\ell_{1,1},\dots,\ell_{4,1}=1\\ \ell_{1,1}\neq\dots\neq\ell_{4,1}}}^{n_1} \cdots \sum_{\substack{\ell_{1,a},\dots,\ell_{4,a}=1\\ \ell_{1,a}\neq\dots\neq\ell_{4,a}}}^{n_a} \mathbb{E}\left(\Lambda_5(\ell_{1,1},\dots,\ell_{2,a})\right)\\
&= \frac{1}{4\cdot\prod\limits_{i=1}^{a}\frac{n_i!}{(n_i-4)!}} \sum_{\substack{\ell_{1,1},\dots,\ell_{4,1}=1\\ \ell_{1,1}\neq\dots\neq\ell_{4,1}}}^{n_1} \cdots \sum_{\substack{\ell_{1,a},\dots,\ell_{4,a}=1\\ \ell_{1,a}\neq\dots\neq\ell_{4,a}}}^{n_a} \operatorname{tr}\left((2\mathbf{TV}_N\mathbf{T})^2\right)\\
&= \operatorname{tr}\left((\mathbf{TV}_N\mathbf{T})^2\right)
\end{aligned}$$

and

$$
\begin{aligned}
& Var\,(C_2) \\
= \; & \sum_{\substack{\ell_{1,1},\dots,\ell_{4,1}=1\\ \ell_{1,1}\neq\dots\neq\ell_{4,1}}}^{n_1} \cdots \sum_{\substack{\ell_{1,a},\dots,\ell_{4,a}=1\\ \ell_{1,a}\neq\dots\neq\ell_{4,a}}}^{n_a} \sum_{\substack{\ell'_{1,1},\dots,\ell'_{4,1}=1\\ \ell'_{1,1}\neq\dots\neq\ell'_{4,1}}}^{n_1} \cdots \sum_{\substack{\ell'_{1,a},\dots,\ell'_{4,a}=1\\ \ell'_{1,a}\neq\dots\neq\ell'_{4,a}}}^{n_a} \frac{Cov\left(\Lambda_5(\ell_{1,1},\dots,\ell_{4,a}),\Lambda_5(\ell'_{1,1},\dots,\ell'_{4,a})\right)}{\left(4\cdot\prod\limits_{i=1}^{a}\frac{n_i!}{(n_i-4)!}\right)^2} \\
\leqslant \; & \frac{\prod\limits_{i=1}^{a}\frac{n_i!}{(n_i-4)!}-\prod\limits_{i=1}^{a}\frac{(n_i-4)!}{(n_i-8)!}}{4\cdot\prod\limits_{i=1}^{a}\frac{n_i!}{(n_i-4)!}}\, Var\,(\Lambda_5(1,\dots,4)) \\
= \; & \frac{\prod\limits_{i=1}^{a}\frac{n_i!}{(n_i-4)!}-\prod\limits_{i=1}^{a}\frac{(n_i-4)!}{(n_i-8)!}}{\prod\limits_{i=1}^{a}\frac{n_i!}{(n_i-4)!}}\cdot \mathcal{O}\left(\mathrm{tr}^2\left((\mathbf{TV}_N)^2\right)\right).
\end{aligned}
$$

b) For the subsampling version, we take the same steps as for the comparable estimators from Sattler and Pauly [2018] and use some results shown therein. Denote with $\mathcal{F}(\sigma_i(B,m))$ the smallest σ-field, which contains $\sigma_i(b,m)$ $\forall b \in B$. Then we get

$$
\begin{aligned}
\mathbb{E}\,(C_1^\star(B)) &= \frac{1}{2B}\sum_{b=1}^{B}\mathbb{E}\,(\Lambda_4(\boldsymbol{\sigma}(b,2))) \\
&= \frac{1}{2B}\sum_{b=1}^{B}\mathbb{E}\,(\Lambda_4(\ell_{1,1},\dots,\ell_{2,a})) \\
&= \frac{1}{2B}\sum_{b=1}^{B}\mathrm{tr}\,(2\mathbf{TV}_N) = \mathrm{tr}\,(\mathbf{TV}_N)
\end{aligned}
$$

and

$$
Var\,(\mathbb{E}\,(C_1^\star(B)|\mathcal{F}(\boldsymbol{\sigma}(B,2)))) = Var\,(\mathrm{tr}\,((\mathbf{TV}_N))) = 0.
$$

With this and $M(B,\boldsymbol{\sigma}(b,2))$ as the notation of the amount of pairs $(k,\ell) \in \mathbb{N}_B \times \mathbb{N}_B$, which fulfill that $\boldsymbol{\sigma}(k,2)$ and $\boldsymbol{\sigma}(\ell,2)$ have totally different elements, we get

$$
\begin{aligned}
Var\,(C_1^\star(B)) &= 0 + \mathbb{E}\,(Var\,(C_1^\star(B)|\mathcal{F}(\boldsymbol{\sigma}(B,2)))) \\
&\leqslant \frac{1}{4B^2}\mathbb{E}\left(\sum_{(j,\ell)\in\mathbb{N}_B\times\mathbb{N}_B\setminus M(B,\boldsymbol{\sigma}(b,2))} Var\,(\Lambda_4(\boldsymbol{\sigma}(j,2))|\mathcal{F}(\boldsymbol{\sigma}(B,2)))\right) \\
&= \frac{\mathbb{E}(|\mathbb{N}_B\times\mathbb{N}_B\setminus M(B,\boldsymbol{\sigma}(b,2))|)}{B^2}\cdot\frac{Var\left(\mathbf{Z}_{(1,2)}{}^\top\mathbf{T}\mathbf{Z}_{(1,2)}\right)}{4} \\
&\leqslant \left(1-\left(1-\frac{1}{B}\right)\cdot\prod_{i=1}^{a}\frac{\binom{n_i-2}{2}}{\binom{n_i}{2}}\right)\cdot\mathrm{tr}\,((\mathbf{TV}_N)).
\end{aligned}
$$

The same calculations with $M(B, \boldsymbol{\sigma}(b,4))$ for our other estimator lead to

$$\begin{aligned}\mathbb{E}\left(C_2^\star(B)\right) &= \frac{1}{4B}\sum_{b=1}^{B}\mathbb{E}\left(\Lambda_5(\boldsymbol{\sigma}(b,4))\right)\\ &= \frac{1}{4B}\sum_{b=1}^{B}\mathbb{E}\left(\Lambda_5(\ell_{1,1},\dots,\ell_{4,a})\right).\\ &= \frac{1}{4B}\sum_{b=1}^{B}\operatorname{tr}\left((2\mathbf{T}\mathbf{V}_N)^2\right) = \operatorname{tr}\left((\mathbf{T}\mathbf{V}_N)^2\right)\end{aligned}$$

as well as

$$\mathit{Var}\left(\mathbb{E}\left(C_2^\star(B)|\mathcal{F}(\boldsymbol{\sigma}(B,4))\right)\right) = \mathit{Var}\left(\operatorname{tr}\left((\mathbf{T}\mathbf{V}_N)^2\right)\right) = 0$$

and finally to

$$\begin{aligned}\mathit{Var}\left(C_2^\star(B)\right) &= 0 + \mathbb{E}\left(\mathit{Var}\left(C_2^\star(B)|\mathcal{F}(\boldsymbol{\sigma}(B,4))\right)\right)\\ &\leqslant \frac{1}{16B^2}\mathbb{E}\left(\sum_{(j,\ell)\in\mathbb{N}_B\times\mathbb{N}_B\setminus M(B,\boldsymbol{\sigma}(b,4))}\mathit{Var}\left(\Lambda_5(\boldsymbol{\sigma}(j,4))|\mathcal{F}(\boldsymbol{\sigma}(B,4))\right)\right)\\ &= \frac{\mathbb{E}(|\mathbb{N}_B\times\mathbb{N}_B\setminus M(B,\boldsymbol{\sigma}(b,4))|)}{B^2}\cdot\frac{\mathit{Var}\left(\left[\mathbf{Z}_{(1,2)}{}^\top\mathbf{T}\mathbf{Z}_{(3,4)}\right]^2\right)}{16}\\ &\leqslant \left(1-\left(1-\frac{1}{B}\right)\cdot\prod_{i=1}^{a}\frac{\binom{n_i-4}{4}}{\binom{n_i}{4}}\right)\cdot 27\operatorname{tr}^2\left((\mathbf{T}\mathbf{V}_N)^2\right).\end{aligned}$$

These results show that the estimators are unbiased and dimensional-stable if

$$\left(1-\left(1-\frac{1}{B}\right)\cdot\prod_{i=1}^{a}\frac{\binom{n_i-2}{2}}{\binom{n_i}{2}}\right)\quad \mathit{resp.}\quad \left(1-\left(1-\frac{1}{B}\right)\cdot\prod_{i=1}^{a}\frac{\binom{n_i-4}{4}}{\binom{n_i}{4}}\right)$$

goes asymptotically to zero. Hence, it is necessary that $B\to\infty$ as well as the second part of the respective product goes to 1. While the first point is easy to fulfill, the second one is done through requirements on the relation between samples sizes and number of groups.

□

Remark 3.2.1:
As an alternative to the existence of $q > 0$ *with* $n_{\min} = \mathcal{O}(a^q)$ *also other conditions are possible which make sure* $\prod_{i=1}^{a} \frac{(n_i-2)!\cdot(n_i-2)!}{n_i!(n_i-4)!} \to 1$*. This condition is also sufficient for the corresponding condition for* C_2 *resp* C_5*. If all of these options are not fulfilled, both estimators* C_1 *and* C_2 *could be adjusted as it was done in Sattler and Pauly [2018] with* C_7 *for the estimator* C_5*.*

3.3 Article 2: Manifold Asymptotics of Quadratic-Form-Based Inference in Repeated Measures Designs

In Sattler and Pauly [2018], three different asymptotic frameworks were investigated, which all need each group's sample size to go to infinity. It is of interest to determine under which conditions our frameworks can be extended, for example, containing $a \to \infty$ while n_i and d are fixed as considered in screening experiments. On a closer inspection, the theoretical results reveal that the increasing sample size is only required to estimate the traces containing the single groups' covariance matrices and not for the test statistic's asymptotic distribution.

Thus, under the assumption of equal covariance matrices, it is sufficient if at least one group sample size goes to infinity, or if the size of all groups is limited, but the number of groups increases. Therefore, the different considered asymptotic frameworks for homogenous covariance matrices are

$$\begin{aligned} a &\to \infty, \\ \min(a, d) &\to \infty, \\ \min(a, n_{\max}) &\to \infty, \\ \min(d, n_{\max}) &\to \infty, \\ \min(a, d, n_{\max}) &\to \infty, \end{aligned}$$

while the statistical model from Sattler and Pauly [2018] is simplified through $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$. For using the standardized quadratic form again as the test statistic, estimators of the unknown traces under the assumption of equal covariance matrices in all groups are necessary. In their development, we take into account that our asymptotic frameworks allow strongly unbalanced sample sizes in the individual groups. As we allow single groups with limited sample size, while at the same time the sample size of other groups can go to infinity, the equivalent assessment of all groups might reduce the quality of our estimators. Therefore, we developed estimators, where the relation between the group sample size and the total sample size is used to weight the appropriate estimator from this group. This choice allows using estimators based on individual groups, which is the most intuitive approach. Unfortunately, for these symmetrized

U-statistics, the number of potential index combinations is again too high. The quite unbalanced sample sizes have to be taken into account for using a subsampling approach, making the task more demanding. This fact is also the reason why we do not consider estimators using observations from different groups. The required subsampling combined with the unbalanced groups would make this kind of estimator quite complicated in practice. The results of Theorem 3.1 from Sattler and Pauly [2018] do not only hold for the more general asymptotic frameworks and the new estimators but also could be expanded to a more general case:

> If and only if all $\beta_i \to b_i$, where only a finite number r of b_i is unequal zero, then $W_N \xrightarrow{\mathcal{D}} Z \sim \sum_{i=1}^r b_i(C_i - 1)/\sqrt{2} + \sqrt{1 - \sum_{i=1}^r b_i^2} \cdot B$, with $C_i \overset{i.i.d.}{\sim} \chi_1^2$ and $B \sim \mathcal{N}(0, 1)$.

This includes the cases with $\beta_1 \to b_1 \in \{0, 1\}$, as special cases. Moreover it makes clearer, how the different limit distributions from Theorem 3.1 are connected. Through this new theorem, we not only expand the situations where the asymptotic distribution is known but also have an equivalence in cases where it so far was only a one direction relation.

Since the asymptotic distribution does not require equal covariances or particular asymptotic frameworks, this result also holds in the setting of Sattler and Pauly [2018], which allows for generalizing their results. Simulations for an increasing number of unbalanced groups and fixed sample size and dimension show good results for the type-I-error and the power to detect deviation from the null hypothesis.

The critical values based on a random variable K_f can also be used in this setting, with the same theoretical results. It was a bit surprising, but through the simplified structure of f_P, we could show that in some cases, the concrete asymptotic distribution only depends on the setting and the hypotheses and not on the considered data. Consequently, in these cases, the uncertainty of whether the limit of τ_{CQ} is in $\{0, 1\}$ or not can be avoided.

All results also hold under the weaker condition of $\mathbf{T}_S \boldsymbol{\Sigma}_i \mathbf{T}_S$ being the same in all groups $i = 1, ..., a$, which depends on the considered hypothesis.

3.4 Article 3: Testing Hypotheses about Covariance Matrices in General MANOVA Designs

The assumption of equal covariance matrices between some or all groups is often used and can simplify many calculations. Nevertheless, it should be statistically justified. For such a preliminary test on variances, it is impractical to need a lot of additional requirements. To this end, it is of great importance to develop a test with nearly no distributional requirements. Therefore, we use the semiparametric model (2.1) from Section 2.1.1, with $\boldsymbol{\Sigma}_i \geqslant 0$. Here, beyond the assumption of finite fourth moments, which is difficult to avoid while handling the analysis of covariance matrices, all other assumptions are common and less restrictive. This fact allows using our approach for many different kinds of data, such as continuous or discrete data, as long as the empirical covariance matrix is meaningfully defined. Existing tests, unfortunately, often have assumptions on the density function or complicated moments that are difficult to verify. One exception is the test of Zhang and Boos [1992], which has the same requirement as our approach but is known to have low power. To this aim, another bootstrap approach is introduced in their paper, with more power but which also requires severe restrictions through assuming $\mathbb{E}\left(\left[\operatorname{vech}(\boldsymbol{\epsilon}_{11}\boldsymbol{\epsilon}_{11}^\top)\right]\left[\operatorname{vech}(\boldsymbol{\epsilon}_{11}\boldsymbol{\epsilon}_{11})^\top\right]^\top\right) = \mathbb{E}\left(\left[\operatorname{vech}(\boldsymbol{\epsilon}_{21}\boldsymbol{\epsilon}_{21}^\top)\right]\left[\operatorname{vech}(\boldsymbol{\epsilon}_{21}\boldsymbol{\epsilon}_{21}^\top)\right]^\top\right)$. This assumption changes the hypothesis of equal covariance matrices to the subset of equal covariance matrices and equality of this special fourth moment. Due to this smaller hypothesis, the power increases, but rejection of this null hypothesis allows no conclusions on the covariance matrices.

To allow a variety of possible hypotheses in addition to homoscedasticity of covariance matrices of multiple groups, we consider $\mathcal{H}_0 : \mathbf{C}\boldsymbol{\nu} = \boldsymbol{\zeta}$ with $\mathbf{C} \in \mathbb{R}^{m \times ap}$ for $m \leqslant p$. This model is comparable to the ANOVA-based analysis, in which usual idempotent and symmetric quadratic hypothesis matrices are used. It allows to compare covariance matrices of multiple groups as well as parts of them and moreover comparisons of effect measurements like the trace of the covariance matrix. For one group, this model contains testing for a given trace of the covariance matrix, for a given covariance matrix and whether components are uncorrelated. Also, testing equality of all diagonal elements and many other interesting hypotheses is part of the null hypothesis class. The procedure to check all the considered hypothesis is based on the asymptotic distribution of the vectorized empirical covariance matrix $\operatorname{vech}(\widehat{\boldsymbol{\Sigma}}_i)$, which is given in Theorem 3.1 through

$$\sqrt{n_i}(\operatorname{vech}(\widehat{\boldsymbol{\Sigma}}_i) - \operatorname{vech}(\boldsymbol{\Sigma}_i)) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}_p, \mathbf{V}_i).$$

Hereby, $\mathbf{V}_i = Cov(\operatorname{vech}((\mathbf{X}_{i1} - \mu_i)(\mathbf{X}_{i1} - \mu_i)^\top)$, which makes it harder to check important properties such as positive definiteness. Based on this asymptotic distribution, all quadratic forms known from the mean-based analysis can be used if their requirements are fulfilled. With $\widehat{\mathbf{V}}_i$ as the empirical covariance matrix of vech $\left([\mathbf{X}_{ik} - \overline{\mathbf{X}}_i][\mathbf{X}_{ik} - \overline{\mathbf{X}}_i]^\top\right)$ and $\mathbf{V} = \oplus_{i=1}^a N/n_i \mathbf{V}_i$ estimated through $\widehat{\mathbf{V}} = \oplus_{i=1}^a N/n_i \widehat{\mathbf{V}}_i$ quadratic forms given by

$$\widehat{Q}_v = N \left[\mathbf{C} \begin{pmatrix} \operatorname{vech}(\widehat{\boldsymbol{\Sigma}}_1) \\ \vdots \\ \operatorname{vech}(\widehat{\boldsymbol{\Sigma}}_a) \end{pmatrix} - \zeta \right]^\top \mathbf{E}(\mathbf{C}, \widehat{\mathbf{V}}) \left[\mathbf{C} \begin{pmatrix} \operatorname{vech}(\widehat{\boldsymbol{\Sigma}}_1) \\ \vdots \\ \operatorname{vech}(\widehat{\boldsymbol{\Sigma}}_a) \end{pmatrix} - \zeta \right],$$

are used. Here, $\mathbf{E}(\mathbf{C}, \widehat{\mathbf{V}})$ is the quadratic form defining matrix, where we study the specific choices $\mathbf{E}(\mathbf{C}, \widehat{\mathbf{V}}) = (\mathbf{C}\widehat{\mathbf{V}}\mathbf{C}^\top)^+$ for the WTS, $\mathbf{E}(\mathbf{C}, \widehat{\mathbf{V}}) = \mathbf{I}_m / \operatorname{tr}(\mathbf{C}\widehat{\mathbf{V}}\mathbf{C}^\top)$ for the ATS and $\mathbf{E}(\mathbf{C}, \widehat{\mathbf{V}}) = (\mathbf{C}\widehat{\mathbf{V}}_0\mathbf{C}^\top)^+$ for the MATS.

Since some of these quadratic forms are asymptotic non-pivot or exhibit bad small sample approximations, resampling techniques are essential to solve both difficulties. We consider a parametric bootstrap approach on the one hand and a wild bootstrap approach on the other hand. Hereby, some adaptions are necessary due to the focus on the covariance matrix instead of the expectation value. One other challenge was the adaption of the class of considered null hypotheses from mean-based analysis to vectorized covariance matrices. For hypotheses $\mathbf{C}v = \zeta$, where ζ is allowed to have nonzero values, the frequently used unique hypothesis matrix does not have to exist. As a consequence, the choice of the appropriate hypothesis matrix turned out to be of particular importance. In particular, a closer look at this topic reveals that the choice of hypothesis matrix can save an essential amount of computation time.

Since the dimension p of the vectorized covariance matrix grows quadratically in dimension d of the observation vector, small sample behavior and computation time are essential aspects of a test statistic. Therefore an extensive simulation is conducted to check these and other properties. Here, various hypotheses are considered for one, two, and three groups and four different distributions together with two different covariance matrices. To see the properties for different combinations, we consider dimensions 5 and 7, which means that the vectorized matrix has dimensions 15 and 21. Sample sizes depending on the dimension, such as 5d and 50d, enable comparability.

Different quadratic forms with both bootstrap approaches are simulated and compared with existing procedures regarding their type-I-error and their power to detect deviations from the null hypotheses. Hereby, the ATS given through $\mathbf{E}(\mathbf{C}, \widehat{\mathbf{V}}) = \mathbf{I}_p / \operatorname{tr}(\mathbf{C}\widehat{\mathbf{V}}\mathbf{C}^\top)$ shows the best results in terms of type-I error control and power to detect deviations from the null hypothesis. The ATS with parametric bootstrap has the best properties, while a Monte-Carlo-based approach is hardly less favorable but has clear advantages regarding computation time. In almost every setting, both are better than the two tests from Zhang and Boos [1992], which also allow other hypotheses than equal covariance matrices but give no information about the required test statistics. Even the very popular Box's M-test, which is only applicable for normal distribution, has for the normal distribution no better large sample approximation than our ATS with parametric bootstrap, but worse small sample approximation.

Since our approach is similar to the classical mean-based analysis in many aspects, it is intuitive and understandable for potential users. Furthermore, the requirements are comparatively low, and the considered null hypotheses are rather general. Together with the convincing simulation results, this work is preferable to most of the existing procedures and enables a variety of expansions and applications.

3.5 Extensions and further results based on Article 3

In Sattler et al. [2022] many hypotheses were considered, extensive simulations were done, and further questions like saving time through the choice of an appropriate hypotheses matrix were investigated. Nevertheless, a few issues were left pending. We want to take a closer look at several of these issues, which are:

1. For testing equality of covariance matrices between multiple groups, we had convincing simulation results in the case of equal distributions between groups. It is of great interest to investigate the performance in settings where the groups have different distributions since this violates the conditions of one of the tests from Zhang and Boos [1993]. Moreover, we want to examine the type-I-error rate for the challenging hypothesis of testing for a given covariance matrix, which was, e.g., investigated in Gupta and Xu [2006]. Such hypotheses are rare, but as we will see afterwards, there exist some useful applications.

2. It was mentioned that the hypothesis matrices are, in general, not unique. Therefore, the qualities of alternative non-quadratic hypothesis matrices were pointed out. This makes it necessary to investigate the influence of the chosen hypothesis matrix on the test decision.

3. After developing a test to examine hypotheses regarding the covariance matrix, it is an enticing project to adapt the approach for testing hypotheses regarding correlation matrices. They contain much information about the dependency structure of the underlying data set without being influenced by most multiplications. Thus, depending on the situation, it is more suitable to consider them instead of the covariance matrix.

4. In many situations, the covariance matrix structure (like being an autoregressive matrix, a compound symmetry matrix, or a Toeplitz matrix) is of greater interest than, for example, comparing the covariance matrix with another matrix. Since the pattern of a covariance matrix contains lots of information about the dependency structure and, therefore, about an appropriate model, this extends the test procedure's applicability substantially.

5. As could be seen in Section 6 of Sattler et al. [2022], computation time is an essential factor for choosing the adequate test statistic for a hypothesis in

the individual situation. This aspect is, in particular relevant for hypotheses regarding the covariance matrix, since the dimension of the vectorized covariance matrix grows quadratically in d.
In this section, it was also shown that a non-quadratic hypothesis matrix could save a substantial amount of time in many cases. The influence of this change is investigated in Section 3.5.2. Therefore, we want to introduce some tricks and techniques to further reduce the computation time and use the capability of the non-quadratic matrix without lowering the quality of our approach.

6. Finally, it is of interest whether the assumptions of the statistical model can be reduced. Since there are few requirements, the only condition which can be relaxed is the equal dimension in each group.

3.5.1 Simulating one more setting and another hypothesis

First, we want to analyze how the different test statistics perform for the hypothesis of equal covariance matrices of two groups if the distributions are from two different families. This setting is more challenging and also violates the additional requirement for the pooled bootstrap from Zhang and Boos [1993]. To investigate the behavior in this setting, we build all different pairs from our four distributions, which are based on t_9, normal distribution, skew normal distribution, and gamma distribution. As in Sattler et al. [2022] we use two different covariance matrices $(\mathbf{V})_{ij} = 0.6^{|i-j|}$ and $\mathbf{V} = \mathbf{I}_5 + \mathbf{J}_5$. The results are displayed in Tables 3.1-3.4, where values in the 95% binomial interval $[0.047, 0.053]$ are printed bold.

It is apparent that for all our differing test-statistics, the results are comparable with the results from observations with the same distribution, and in some cases, even slightly better. This result is of great importance because such mixed distributions appear to be more demanding and allow us to contrast different kinds of data.

The ATS with parametric bootstrap and our Monte-Carlo ATS have convincing results even for the smaller sample size of $N = 100$, while for $N = 50$, the better small sample approximation of the parametric bootstrap ATS can be seen. Also, the MATS with parametric bootstrap shows good results for some combinations. Interestingly, the Box's M tests perform somewhat better for t_9 vs.

Normal and Normal vs. Skew Normal, where at least a trend can be seen. But in all combinations with the gamma distribution, the error rate is higher than 17%, even if the normal distribution is involved. These values show one more time that both groups have to be normally distributed for Box's M-test.

	t_9 vs. Normal				t_9 vs. Skew normal				t_9 vs. Gamma			
N	50	100	250	500	50	100	250	500	50	100	250	500
ATS-Para	.0540	**.0501**	**.0488**	**.0484**	.0539	**.0516**	**.0504**	**.0497**	.0570	**.0500**	**.0493**	**.0493**
ATS-Wild	.0803	.0647	.0562	**.0526**	.0887	.0676	.0575	.0551	.0961	.0752	.0637	.0571
ATS	.0589	**.0514**	**.0489**	**.0483**	.0602	**.0525**	**.0503**	**.0498**	.0621	**.0507**	**.0492**	**.0491**
WTS-Para	.0613	.0567	**.0514**	.0454	.0641	.0616	.0554	.0537	.0831	.0904	.0899	.0786
WTS-Wild	.0911	.0753	.0618	**.0508**	.0984	.0852	.0670	.0617	.1330	.1298	.1106	.0899
WTS-χ^2_{15}	.4877	.2011	.0888	.0630	.5006	.2140	.0973	.0716	.5445	.2682	.1419	.0976
MATS-Para	.0613	.0540	**.0520**	**.0506**	.0631	.0588	.0539	**.0524**	.0729	.0665	.0625	.0590
MATS-Wild	.0846	.0653	.0558	**.0530**	.0900	.0716	.0607	.0566	.1053	.0857	.0717	.0657
Bartlett-S	.0147	.0381	**.0494**	**.0481**	.0161	.0420	.0537	**.0522**	.0235	.0536	.0653	.0602
Bartlett-P	.0213	.0298	.0354	.0347	.0232	.0361	.0417	.0412	.0353	**.0524**	.0628	.0634
Box's M-χ^2_{15}	.0914	.0798	.0824	.0764	.1167	.1098	.1119	.1104	.2098	.2216	.2433	.2488
Box's M-F	.0865	.0788	.0822	.0763	.1098	.1083	.1117	.1103	.2015	.2197	.2429	.2488

Table 3.1: Simulated type-I-error rates ($\alpha = 5\%$) in scenario $A)$ ($\mathcal{H}_0^v : \mathbf{V}_1 = \mathbf{V}_2$) for ATS, WTS, MATS, Bartletts test and Box's M-test. The observation vectors have dimension 5, covariance matrix $(\mathbf{V})_{ij} = 0.6^{|i-j|}$ and there is always the same relation between group sample sizes with $n_1 := 0.6 \cdot N$ resp. $n_2 := 0.4 \cdot N$.

Now it remains to inspect the performance of both Bartlett test-statistics, where we expect an effect, at least for the pooled bootstrap. For the Bartlett test-statistic with separate bootstrap, these mixed distributions' performance is a bit worse than for groups from the same distributional family. For example, this can be seen through the number of error-rates within the 95% binomial interval $[0.047, 0.053]$. This number is identical to Sattler et al. [2022] even though the number of different distributions was one and a half times larger. As a comparison for φ^*_{ATS}, this number increased from 16 to 27, which is even more than the factor 1.5 .

But for the pooled bootstrap, the consequences are significantly stronger than for the separate bootstrap. For some of the distribution pairs like t_9 vs. Normal, the rate is lower than 3.6% over all sample sizes. In most other distributional settings, the error rate started lower and ended higher than the theoretical 5%-level, up to values of 8.2%.

	Normal vs. Skew Normal				Normal vs. Gamma				Skew Normal vs. Gamma			
N	50	100	250	500	50	100	250	500	50	100	250	500
ATS-Para	.0615	.0568	**.0516**	**.0524**	.0637	.0552	**.0514**	**.0502**	.0577	**.0526**	**.0504**	**.0518**
ATS-Wild	.0906	.0706	.0579	.0553	.1002	.0768	.0651	.0577	.0979	.0786	.0631	.0586
ATS	.0661	.0581	**.0520**	**.0513**	.0700	.0569	**.0505**	**.0497**	.0634	.0549	**.0511**	**.0505**
WTS-Para	.0781	.0831	.0689	.0662	.1139	.1381	.1269	.1072	.1025	.1142	.1032	.0928
WTS-Wild	.1183	.1079	.0802	.0727	.1711	.1800	.1501	.1203	.1543	.1558	.1250	.1058
WTS-χ^2_{15}	.5226	.2456	.1144	.0845	.5941	.3316	.1850	.1309	.5759	.3003	.1564	.1132
MATS-Para	.0719	.0647	.0576	.0556	.0857	.0795	.0705	.0637	.0764	.0751	.0643	.0605
MATS-Wild	.0970	.0771	.0616	.0591	.1197	.0971	.0787	.0693	.1100	.0945	.0736	.0671
Bartlett-S	.0143	.0391	**.0483**	**.0498**	.0220	**.0519**	.0612	.0585	.0231	.0539	.0631	.0613
Bartlett-P	.0283	.0423	**.0507**	.0539	.0425	.0655	.0777	.0821	.0347	.0556	.0672	.0720
Box's M-χ^2_{15}	.0888	.0785	.0757	.0729	.1775	.1795	.1975	.2018	.1971	.2093	.2187	.2291
Box's M-F	.0826	.0775	.0755	.0729	.1692	.1780	.1971	.2018	.1885	.2070	.2183	.2291

Table 3.2: Simulated type-I-error rates ($\alpha = 5\%$) in scenario A) ($\mathcal{H}_0^v : \mathbf{V}_1 = \mathbf{V}_2$) for ATS, WTS, MATS, Bartletts test and Box's M-test. The observation vectors have dimension 5, covariance matrix $(\mathbf{V})_{ij} = 0.6^{|i-j|}$ and there is always the same relation between group samples size with $n_1 := 0.6 \cdot N$ resp. $n_2 := 0.4 \cdot N$.

	t_9 vs. Normal				t_9 vs. Skew normal				t_9 vs. Gamma			
N	50	100	250	500	50	100	250	500	50	100	250	500
ATS-Para	.0594	**.0524**	**.0509**	**.0498**	.0584	.0542	**.0512**	**.0504**	.0594	.0535	.0533	**.0511**
ATS-Wild	.0815	.0620	.0551	.0531	.0856	.0688	.0575	.0538	.0919	.0744	.0646	.0566
ATS	.0641	**.0525**	**.0504**	**.0494**	.0639	.0571	**.0519**	**.0496**	.0635	.0546	.0531	**.0502**
WTS-Para	.0616	.0567	**.0511**	.0455	.0641	.0616	.0564	.0544	.0823	.0903	.0894	.0795
WTS-Wild	.0911	.0753	.0618	**.0508**	.0984	.0852	.0670	.0617	.1330	.1298	.1106	.0899
WTS-χ^2_{15}	.4877	.2011	.0888	.0630	.5006	.2140	.0973	.0716	.5445	.2682	.1419	.0976
MATS-Para	.0651	.0554	**.0521**	**.0514**	.0653	.0601	.0560	.0532	.0760	.0695	.0633	.0594
MATS-Wild	.0830	.0646	.0554	.0537	.0867	.0701	.0604	.0550	.0982	.0825	.0703	.0640
Bartlett-S	.0145	.0383	**.0495**	**.0481**	.0160	.0423	.0534	**.0524**	.0233	.0534	.0649	.0600
Bartlett-P	.0213	.0298	.0354	.0347	.0232	.0361	.0417	.0412	.0353	**.0524**	.0628	.0634
Box's M-χ^2_{15}	.0914	.0798	.0824	.0764	.1167	.1098	.1119	.1104	.2098	.2216	.2433	.2488
Box's M-F	.0865	.0788	.0822	.0763	.1098	.1083	.1117	.1103	.2015	.2197	.2429	.2488

Table 3.3: Simulated type-I-error rates ($\alpha = 5\%$) in scenario A) ($\mathcal{H}_0^v : \mathbf{V}_1 = \mathbf{V}_2$) for ATS, WTS, MATS, Bartletts test and Box's M-test. The observation vectors have dimension 5, covariance matrix $\mathbf{V} = \mathbf{I}_5 + \mathbf{J}_5$ and there is always the same relation between group samples size with $n_1 := 0.6 \cdot N$ resp. $n_2 := 0.4 \cdot N$.

	Normal vs. Skew Normal				Normal vs. Gamma				Skew Normal vs. Gamma			
N	50	100	250	500	50	100	250	500	50	100	250	500
ATS-Para	.0646	.0584	.0534	**.0521**	.0667	.0573	**.0523**	**.0497**	.0608	.0559	**.0528**	**.0521**
ATS-Wild	.0858	.0685	.0577	.0544	.0982	.0766	.0622	.0569	.0952	.0777	.0627	.0595
ATS	.0696	.0598	.0535	**.0513**	.0730	.0587	**.0524**	**.0501**	.0663	.0574	**.0525**	**.0514**
WTS-Para	.0799	.0837	.0693	.0656	.1151	.1369	.1259	.1068	.0999	.1146	.1025	.0929
WTS-Wild	.1183	.1079	.0802	.0727	.1711	.1800	.1501	.1203	.1543	.1558	.1250	.1058
WTS-χ^2_{15}	.5226	.2456	.1144	.0845	.5941	.3316	.1850	.1309	.5759	.3003	.1564	.1132
MATS-Para	.0726	.0657	.0575	.0548	.0885	.0760	.0693	.0615	.0782	.0747	.0644	.0608
MATS-Wild	.0893	.0741	.0606	.0560	.1103	.0891	.0763	.0657	.1031	.0895	.0709	.0655
Bartlett-S	.0144	.0394	**.0484**	**.0498**	.0219	**.0517**	.0617	.0585	.0229	.0540	.0628	.0614
Bartlett-P	.0283	.0423	**.0507**	.0539	.0425	.0655	.0777	.0821	.0347	.0556	.0672	.0720
Box's M-χ^2_{15}	.0888	.0785	.0757	.0729	.1775	.1795	.1975	.2018	.1971	.2093	.2187	.2291
Box's M-F	.0826	.0775	.0755	.0729	.1692	.1780	.1971	.2018	.1885	.2070	.2183	.2291

Table 3.4: Simulated type-I-error rates ($\alpha = 5\%$) in scenario $A)$ ($\mathcal{H}_0^v : \mathbf{V}_1 = \mathbf{V}_2$) for ATS, WTS, MATS, Bartletts test and Box's M-test. The observation vectors have dimension 5, covariance matrix $\mathbf{V} = \mathbf{I}_5 + \mathbf{J}_5$ and there is always the same relation between group samples size with $n_1 := 0.6 \cdot N$ resp. $n_2 := 0.4 \cdot N$.

This is not surprising, since their method requires the additional condition of $\mathbb{E}\left([\text{vech}(\boldsymbol{\epsilon}_{11}\boldsymbol{\epsilon}_{11}^\top)][\text{vech}(\boldsymbol{\epsilon}_{11}\boldsymbol{\epsilon}_{11}^\top)]^\top\right) = \mathbb{E}\left([\text{vech}(\boldsymbol{\epsilon}_{21}\boldsymbol{\epsilon}_{21}^\top)][\text{vech}(\boldsymbol{\epsilon}_{21}\boldsymbol{\epsilon}_{21}^\top)]^\top\right)$.
Through this condition, they test a considerably larger null hypothesis, which is rejected if the additional condition is violated, as well as if the covariance matrices differ. And due to the larger null hypothesis, the power increases, but rejection allows no conclusions on the homoscedasticity of covariance matrices. This fact, together with the simulation results, clearly demonstrates that the pooled bootstrap approach from Zhang and Boos [1993] should be used with caution.

The above results show that φ^*_{ATS} and φ_{ATS} are appropriate procedures for situations with possibly different kinds of distributions in the groups. Thereby, φ^*_{ATS} shows better results, especially for smaller sample sizes, but needs more computation time, see Section 3.5.5 and Sattler et al. [2022]. Overall, this is one more argument to favor these tests over existing procedures, in addition to generally better approximation for small sample sizes and the wide variety of possible null hypotheses.

Moreover we want to investigate another hypothesis:

F) $a = 1 \quad \mathcal{H}_0^v : \mathbf{V}_1 = \mathbf{V}$ for given $\mathbf{V}$,

where also scenario F) can be formulated with an idempotent symmetric matrix $\mathbf{C}(F) = \mathbf{I}_{15}$.
The new hypothesis of testing for a given matrix is particularly interesting because it seems quite challenging. The values in Table 3.5 and Table 3.6 confirmed this assumption, especially for the WTS, which always had type-I-error rates higher than 15%. The other tests were too liberal as well but to a much lesser extent. Considering the demanding nature of this hypothesis, the type-I-error rates of φ^*_{ATS} seem to be acceptable for higher sample sizes as $n_1 = 125$ or $n_1 = 250$ as they fulfill, for example, Bradley's liberal criterion from Bradley [1978]. In this work, Bradley considered a statistical test as robust, if the empirical type I error rate is between 0.5α and 1.5α [1]. This kind of hypothesis is comparably rare, but we consider similar hypotheses in Section 3.5.3 and Section 3.5.4.

3.5.2 Influence of the used hypothesis matrix

To investigate the influence of the chosen hypothesis matrix on the test result, we use tests with different matrices for the same data with the same seed to strip out all other influences on the test decision. For this simulation, we use 1,000 bootstrap runs for the tests based on parametric and wild bootstrap and 10,000 runs for the Monte-Carlo test. To get results that are as generic as possible, we use 10,000 repetitions, each with four distributions (based on t_9-distribution, Normal-distribution, Skew Normal-distribution, and Gamma-distribution) and two covariance matrices ($(\mathbf{V})_{ij} = 0.6^{|i-j|}$ and $\mathbf{V} = \mathbf{I}_5 + \mathbf{J}_5$). The considered dimension was $d = 5$ and the sample size depending on the hypothesis was $n = 50$, $\mathbf{n} = (60, 40)$ or $\mathbf{n} = (64, 40, 56)$. We calculate the congruity of the test decisions, i.e., the proportion of both tests resulting in the same test decision. Moreover, we consider different summary statistics for the difference of the p-values like quantiles or the maximum. We do not differentiate between the different distributions or covariance matrices for all these values but consider the whole set of all 80,000 test decisions.

[1] This criterion was described by himself as "The most liberal criterion that I am able to take seriously" (Bradley [1978], page 146).

	t_9				Normal			
N	25	50	125	250	25	50	125	250
ATS-Para	.0795	.0668	.0548	.0560	.0804	.0673	.0609	**.0512**
ATS-Wild	.1166	.0905	.0656	.0617	.1125	.0840	.0690	.0563
ATS	.0851	.0696	.0552	.0559	.0874	.0698	.0618	**.0512**
WTS-Para	.6914	.5576	.3256	.2077	.6152	.4438	.2413	.1429
WTS-Wild	.7383	.6020	.3581	.2275	.6582	.4857	.2622	.1542
WTS-χ^2_{15}	.9813	.8014	.4445	.2608	.9673	.7269	.3556	.1929
MATS-Para	.2191	.1542	.1007	.0812	.1786	.1224	.0850	.0643
MATS-Wild	.2679	.1812	.1103	.0878	.2167	.1425	.0925	.0684
	Skew Normal				Gamma			
N	25	50	125	250	25	50	125	250
ATS-Para	.0816	.0657	.0585	**.0522**	.0910	.0735	.0537	**.0517**
ATS-Wild	.1226	.0877	.0692	.0587	.1418	.1098	.0716	.0626
ATS	.0871	.0678	.0585	**.0531**	.0978	.0744	.0543	**.0521**
WTS-Para	.6623	.5217	.3012	.1876	.7911	.6948	.4613	.3028
WTS-Wild	.7060	.5639	.3298	.2053	.8234	.7392	.5047	.3330
WTS-χ^2_{15}	.9759	.7772	.4186	.2375	.9889	.8831	.5737	.3587
MATS-Para	.2163	.1463	.0956	.0750	.3110	.2280	.1407	.1059
MATS-Wild	.2620	.1707	.1044	.0806	.3687	.2675	.1616	.1172

Table 3.5: Simulated type-I-error rates ($\alpha = 5\%$) in scenario F) ($\mathcal{H}_0^v : \mathbf{V}_1 = \mathbf{V}$) for ATS, WTS and MATS with 5-dimensional vectors and $(\mathbf{V})_{ij} = 0.6^{|i-j|}$.

	t_9				Normal			
N	25	50	125	250	25	50	125	250
ATS-Para	.0940	.0759	.0625	.0600	.0962	.0763	.0644	.0555
ATS-Wild	.1283	.0966	.0702	.0659	.1241	.0889	.0690	.0589
ATS	.1006	.0777	.0617	.0603	.1029	.0786	.0643	.0553
WTS-Para	.6924	.5585	.3247	.2085	.6145	.4451	.2414	.1435
WTS-Wild	.7383	.6020	.3581	.2275	.6582	.4857	.2622	.1542
WTS-χ^2_{15}	.9813	.8014	.4445	.2608	.9673	.7269	.3556	.1929
MATS-Para	.2177	.1534	.0986	.0810	.1870	.1241	.0866	.0668
MATS-Wild	.2546	.1744	.1066	.0843	.2168	.1373	.0929	.0688
	Skew Normal				Gamma			
N	25	50	125	250	25	50	125	250
ATS-Para	.0952	.0758	.0631	.0561	.1041	.0827	.0607	.0554
ATS-Wild	.1294	.0952	.0709	.0602	.1476	.1138	.0767	.0656
ATS	.1015	.0786	.0624	.0548	.1100	.0857	.0611	.0547
WTS-Para	.6620	.5226	.3021	.1879	.7902	.6947	.4618	.3032
WTS-Wild	.7060	.5639	.3298	.2053	.8234	.7392	.5047	.3330
WTS-χ^2_{15}	.9759	.7772	.4186	.2375	.9889	.8831	.5737	.3587
MATS-Para	.2133	.1468	.0938	.0756	.2932	.2165	.1379	.1028
MATS-Wild	.2474	.1667	.1017	.0799	.3367	.2454	.1514	.1090

Table 3.6: Simulated type-I-error rates ($\alpha = 5\%$) in scenario F) ($\mathcal{H}_0^v : \mathbf{V}_1 = \mathbf{V}$) for ATS, WTS and MATS with 5-dimensional vectors and $\mathbf{V} = \mathbf{I}_5 + \mathbf{J}_5$.

Obviously, all tests based on bootstrap or Monte-Carlo-simulation have a slightly random behavior due to the influence of the used realizations of considered random variables. To analyze this potential randomness's influence, we also simulate our bootstrap-based test statistic for the same data but with another seed. Overall considered distributions, hypotheses, and test statistics for 1,000 bootstrap runs, the test decisions' congruity is in this case always about 99.1%. For the Monte-Carlo-simulation with 10,000 simulation runs, it is about 99.7%. Due to the higher number of runs and, therefore, reduced importance of the individual realizations, this second value is higher. For 500 bootstrap runs, the congruity is about 98.8%, and for 5,000 bootstrap runs, it is about 99.6%. We should keep this behavior in mind for the decision of the used number of bootstrap runs.

In scenario C) ($\mathcal{H}_0^v : \operatorname{tr}(\mathbf{V}_1) = \operatorname{tr}(\mathbf{V}_2)$) the only reasonable way to get hypothesis matrices with less rows, is to remove the rows which only contain zeros. It is not surprising that this does not change the decision of our tests. For scenario A) ($\mathcal{H}_0^v : \mathbf{V}_1 = \mathbf{V}_2$) there is also just one way to reduce the hypothesis matrix, by removing the second row of $\mathbf{H}_W$. Multiplied with the mean vector this just simplifies the vector $(\boldsymbol{v}_1 - \boldsymbol{v}_2, -(\boldsymbol{v}_1 - \boldsymbol{v}_2))^\top$ so we get $\boldsymbol{v}_1 - \boldsymbol{v}_2$ and identical for the matrix $\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^\top$. This repetition does not influence the quadratic form's value, so the chosen hypothesis matrix without this repetition again has no impact.

There is only one exception for all these hypotheses where the used matrix has a small impact, the Monte-Carlo test in the ATS. This effect presumably has numerical reasons. Although the reduction of the size of the hypothesis matrix does not change the non-zero eigenvalues of the matrix $\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^\top$, it eliminates the eigenvalues that are zero. From a theoretical point of view, this does not influence the considered weighted sum of independent χ_1^2 random variables. However, these values are estimated for this test's application and therefore are close to zero but not exactly zero. Dependent on the χ_1^2 random variables, this can have a slight influence on the weighted sum, and therefore on the test decision. In our simulation, this difference in the weighted sum influences the test decision in about 0.3% of the cases. As this is comparatively rare and probably can be remedied through smart coding, for all these hypotheses and also for comparable hypotheses, the matrices with fewer rows can be used without worrying.
For other hypotheses, partially there are much more useful hypothesis matrices.

We want to examine two of these hypotheses. On one hand, there is equality of the covariance matrices of three groups E) ($\mathcal{H}_0^v : \mathbf{V}_1 = \mathbf{V}_2 = \mathbf{V}_3$). Here, different options for a hypothesis matrix are given by

$$\mathbf{C}_1(\mathrm{E}) = \mathbf{P}_3 \otimes \mathbf{I}_5, \quad \mathbf{C}_2(\mathrm{E}) = \frac{1}{2} \cdot \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \end{pmatrix} \otimes \mathbf{I}_5, \quad \mathbf{C}_3(\mathrm{E}) = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix} \otimes \mathbf{I}_5.$$

It can be seen that there is an increased focus on the first group in $\mathbf{C}_2(\mathrm{E})$ and $\mathbf{C}_3(\mathrm{E})$. Similarly, we could define matrices with a focus on the second or third group, which is, in this case, only a permutation of columns. Since the influence of such permutation is investigated subsequently for scenario B), we renounce to do this here.

On the other hand, there is the hypothesis from scenario B) ($\mathcal{H}_0^v : \mathbf{V}_{111} = ... = \mathbf{V}_{155}$), which enables many different matrices. Here we considered the following potential hypothesis matrices:

$$\mathbf{C}_1(\mathrm{B}) = \mathit{diag}(\mathbf{h}_d) - \mathbf{h}_d \cdot \mathbf{h}_d^\top / d,$$

$$\mathbf{C}_2(\mathrm{B}) = (\mathbf{1}_{d-1}, \mathbf{0}_{d-1\times d-1}, -\boldsymbol{e}_1, \mathbf{0}_{d-1\times d-2}, -\boldsymbol{e}_2, ..., \mathbf{0}_{d-1}, \boldsymbol{e}_{d-1}) \in \mathbb{R}^{d-1\times p},$$

$$\mathbf{C}_3(\mathrm{B}) = \begin{pmatrix} \mathbf{C}_1(\mathrm{B})_{1\bullet} \\ \mathbf{C}_1(\mathrm{B})_{6\bullet} \\ \mathbf{C}_1(\mathrm{B})_{10\bullet} \\ \mathbf{C}_1(\mathrm{B})_{13\bullet} \end{pmatrix}, \quad \mathbf{C}_4(\mathrm{B}) = \begin{pmatrix} \mathbf{C}_1(\mathrm{B})_{1\bullet} \\ \mathbf{C}_1(\mathrm{B})_{6\bullet} \\ \mathbf{C}_1(\mathrm{B})_{10\bullet} \\ \mathbf{C}_1(\mathrm{B})_{15\bullet} \end{pmatrix}.$$

We focus on φ^*_{ATS}, $\varphi^\star_{ATS}$ and φ_{ATS}, because for the WTS and the MATS, the congruity between all these different hypotheses matrices was 100%. The results of these simulations can be seen in Table 3.7 and Table 3.8 where the hypothesis with a different seed value is marked by $\mathbf{C}_1'(\mathrm{B})$ resp. $\mathbf{C}_1'(\mathrm{E})$.
In Table 3.7 it can be seen that for all hypotheses with identical seed again, the φ^*_{ATS} and φ_{ATS} have very similar results. Moreover, the results for all pairs of covariance matrices for hypothesis E) are comparable. The values of the wild bootstrap approach are always worse than the other two. Therefore, it could be concluded that the applied hypothesis matrix has more effect on the wild bootstrap. But it is more likely to be due to the more liberal behavior of the bootstrap. A liberal test has clearly less congruity than a more conservative test.

	Congruity			25% Quantile			Median		
	ATS-Para	ATS-Wild	ATS	ATS-Para	ATS-Wild	ATS	ATS-Para	ATS-Wild	ATS
$\mathbf{C}_2(\mathrm{E})$ vs. $\mathbf{C}_3(\mathrm{E})$	.961	.949	.961	.030	.031	.030	.070	.075	.070
$\mathbf{C}_1(\mathrm{E})$ vs. $\mathbf{C}_2(\mathrm{E})$	.967	.955	.966	.029	.029	.029	.067	.072	.067
$\mathbf{C}_1(\mathrm{E})$ vs. $\mathbf{C}_3(\mathrm{E})$	.963	.947	.962	.032	.032	.032	.073	.078	.074
$\mathbf{C}_1(\mathrm{E})$ vs. $\mathbf{C}_1'(\mathrm{E})$	.992	.990	.997	.005	.005	.002	.012	.011	.004
	Mean			75% Quantile			Maximum		
	ATS-Para	ATS-Wild	ATS	ATS-Para	ATS-Wild	ATS	ATS-Para	ATS-Wild	ATS
$\mathbf{C}_2(\mathrm{E})$ vs. $\mathbf{C}_3(\mathrm{E})$	.086	.095	.086	.127	.140	.127	.457	.580	.449
$\mathbf{C}_1(\mathrm{E})$ vs. $\mathbf{C}_2(\mathrm{E})$	.083	.091	.083	.121	.134	.121	.463	.564	.461
$\mathbf{C}_1(\mathrm{E})$ vs. $\mathbf{C}_3(\mathrm{E})$	.089	.098	.089	.132	.145	.132	.467	.511	.445
$\mathbf{C}_1(\mathrm{E})$ vs. $\mathbf{C}_1'(\mathrm{E})$	.015	.014	.005	.021	.020	.007	.098	.095	.026

Table 3.7: Comparison of various ATS-test-statistics based on different hypothesis matrices with distinct number of rows by reference to congruity of the test decisions and different summary statistics for the difference of the p-values. The hypothesis is $\mathrm{E})$ ($\mathcal{H}_0^v : \mathbf{V}_1 = \mathbf{V}_2 = \mathbf{V}_3$) for dimension $d = 5$ and $\mathbf{n} = (64, 40, 56)$.

Similar for the differences of the p-values. For all three pairs with congruity values between 0.947 and 0.967, the matrix's impact is quite high, compared with the seed's effect. Moreover, summary statistics with average differences of p-values over 0.08 and maximal differences over 0.44 show that different test decisions appear not only in situations with pretty close p-values.

For hypothesis $\mathrm{B})$ again Table 3.8 shows worse results for the wild bootstrap, for the same reason as before. It should be noted that $\mathbf{C}_1(\mathrm{B})$ vs. $\mathbf{C}_3(\mathrm{B})$ and $\mathbf{C}_1(\mathrm{B})$ vs. $\mathbf{C}_4(\mathrm{B})$ have better values than the other pairs. Since $\mathbf{C}_3(\mathrm{B})$ and $\mathbf{C}_4(\mathrm{B})$ emerge from $\mathbf{C}_1(\mathrm{B})$ by removing all-zero rows and for each of them one other row, this makes sense.

It is interesting that $\mathbf{C}_3(\mathrm{B})$ vs. $\mathbf{C}_4(\mathrm{B})$ has, with congruity value between 0.945 and 0.959, clearly lower values than $\mathbf{C}_1(\mathrm{B})$ vs. $\mathbf{C}_3(\mathrm{B})$. In particular, the structure of both matrices $\mathbf{C}_1(\mathrm{B})$ and $\mathbf{C}_3(\mathrm{B})$ is very similar. Therefore, we might have expected that the results are better than for other more different hypotheses, but the comparison $\mathbf{C}_3(\mathrm{B})$ vs. $\mathbf{C}_4(\mathrm{B})$ shows the highest maximal difference in p-values. With values up to 0.839, this is nearly the whole range of p-values, while all other summary statistics, in this case, are less extreme compared to the other pairs of hypothesis matrices.

	Congruity			25% Quantile			Median		
	ATS-Para	ATS-Wild	ATS	ATS-Para	ATS-Wild	ATS	ATS-Para	ATS-Wild	ATS
$\mathbf{C}_1(\text{B})$ vs. $\mathbf{C}_2(\text{B})$	.942	.927	.940	.046	.044	.046	.107	.110	.107
$\mathbf{C}_2(\text{B})$ vs. $\mathbf{C}_3(\text{B})$	.942	.924	.938	.044	.043	.044	.105	.109	.105
$\mathbf{C}_3(\text{B})$ vs. $\mathbf{C}_4(\text{B})$	.959	.945	.957	.031	.032	.031	.082	.087	.082
$\mathbf{C}_1(\text{B})$ vs. $\mathbf{C}_3(\text{B})$	.974	.963	.972	.027	.027	.027	.061	.065	.062
$\mathbf{C}_1(\text{B})$ vs. $\mathbf{C}_4(\text{B})$	.976	.968	.975	.024	.024	.024	.055	.058	.055
$\mathbf{C}_1(\text{B})$ vs. $\mathbf{C}_1'(\text{B})$	.992	.990	.997	.005	.005	.002	.012	.011	.004

	Mean			75% Quantile			Maximum		
	ATS-Para	ATS-Wild	ATS	ATS-Para	ATS-Wild	ATS	ATS-Para	ATS-Wild	ATS
$\mathbf{C}_1(\text{B})$ vs. $\mathbf{C}_2(\text{B})$	.123	.132	.123	.184	.199	.185	.581	.839	.570
$\mathbf{C}_2(\text{B})$ vs. $\mathbf{C}_3(\text{B})$	.127	.136	.127	.192	.206	.192	.593	.780	.575
$\mathbf{C}_3(\text{B})$ vs. $\mathbf{C}_4(\text{B})$	.121	.130	.121	.177	.190	.177	.703	.839	.695
$\mathbf{C}_1(\text{B})$ vs. $\mathbf{C}_3(\text{B})$	.081	.087	.081	.108	.116	.108	.612	.651	.615
$\mathbf{C}_1(\text{B})$ vs. $\mathbf{C}_4(\text{B})$	.074	.079	.074	.097	.105	.098	.585	.659	.586
$\mathbf{C}_1(\text{B})$ vs. $\mathbf{C}_1'(\text{B})$	.014	.014	.005	.021	.020	.007	.104	.097	.029

Table 3.8: Comparison of various ATS-test-statistics based on different hypothesis matrices with distinct number of rows by reference to congruity of the test decisions and different summary statistics for the difference of the p-values. The hypothesis is B) ($\mathcal{H}_0^v : \mathbf{V}_{111} = ... = \mathbf{V}_{155}$) for dimension $d = 5$ and $n = 50$.

In comparison to the values of $\mathbf{C}_1(\mathrm{B})$ vs. $\mathbf{C}_3(\mathrm{B})$ and $\mathbf{C}_1(\mathrm{B})$ vs. $\mathbf{C}_4(\mathrm{B})$ this suggests more fluctuation in the p-values.
Finally, the pairs $\mathbf{C}_1(\mathrm{B})$ vs. $\mathbf{C}_2(\mathrm{B})$ and $\mathbf{C}_2(\mathrm{B})$ vs. $\mathbf{C}_3(\mathrm{B})$ have similar values, although $\mathbf{C}_2(\mathrm{B})$ in contrast to the other matrices uses pairwise comparison of the diagonal elements with the first diagonal element. Compared with hypothesis $\mathrm{E})$, hypothesis $\mathrm{B})$ shows clearly more concordance except for $\mathbf{C}_1(\mathrm{B})$ vs. $\mathbf{C}_3(\mathrm{B})$ and $\mathbf{C}_1(\mathrm{B})$ vs. $\mathbf{C}_4(\mathrm{B})$. Again, the reason is the more conservative behavior of hypothesis $\mathrm{E})$ compared to hypothesis $\mathrm{B})$. All simulations were also conducted with $n = 125$ resp. $\mathbf{n} = (160, 100, 140)$ which lead to very similar results.

For congruity with values partially under 0.95, it remains the question whether all tests show similar type-I-error rates or if some matrices lead to more conservative or liberal behavior. To investigate this issue, we compare the error rates of three tests based on different matrices, but again with the same seed and the same data in Table 3.9 and 3.10. Again the error rates in the 95% binomial interval are printed bold, while for 10,000 runs, the interval is given through $[0.0458, 0.0543]$.

	ATS-Para			ATS-Wild			ATS		
	$\mathbf{C}_1(\mathrm{B})$	$\mathbf{C}_2(\mathrm{B})$	$\mathbf{C}_3(\mathrm{B})$	$\mathbf{C}_1(\mathrm{B})$	$\mathbf{C}_2(\mathrm{B})$	$\mathbf{C}_3(\mathrm{B})$	$\mathbf{C}_1(\mathrm{B})$	$\mathbf{C}_2(\mathrm{B})$	$\mathbf{C}_3(\mathrm{B})$
t_9	.039	.061	.041	.057	.074	.058	.042	.064	.044
Normal	**.050**	.066	**.050**	.061	.069	.060	**.054**	.069	**.052**
Skew normal	.045	.063	**.049**	.064	.077	.066	**.047**	.067	**.053**
Gamma	.032	.060	.036	.063	.087	.064	.036	.064	.038

Table 3.9: Comparison of various ATS-test-statistics based on different hypothesis matrices with distinct number of rows by their type-I-error rate. The hypothesis is $\mathrm{B})$ ($\mathcal{H}_0^v : \mathbf{V}_{111} = ... = \mathbf{V}_{155}$) for dimension $d = 5$ and $n = 50$ with covariance matrix $(\mathbf{V})_{ij} = 0.6^{|i-j|}$.

The results from Table 3.9 and Table 3.10 display that tests based on $\mathbf{C}_1(\mathrm{B})$ and $\mathbf{C}_3(\mathrm{B})$ have almost the same values, while $\mathbf{C}_2(\mathrm{B})$ has a more liberal behavior. For $\varphi^\star_{ATS}$, which is always liberal, this makes the test presumable to liberal. On the other hand φ^*_{ATS} and φ_{ATS} with a slightly conservative behavior for $\mathbf{C}_1(\mathrm{B})$, using $\mathbf{C}_2(\mathrm{B})$ could be reasonable. Overall distributions, the level seems to be comparable, but either negligible conservative or liberal.
Although $\mathrm{B})$ and $\mathrm{E})$ are hypotheses with $\zeta = \mathbf{0}$ and, therefore, can be formulated with unique projection matrices, the results show an apparent influence

	ATS-Para			ATS-Wild			ATS		
	$\mathbf{C}_1(\mathrm{B})$	$\mathbf{C}_2(\mathrm{B})$	$\mathbf{C}_3(\mathrm{B})$	$\mathbf{C}_1(\mathrm{B})$	$\mathbf{C}_2(\mathrm{B})$	$\mathbf{C}_3(\mathrm{B})$	$\mathbf{C}_1(\mathrm{B})$	$\mathbf{C}_2(\mathrm{B})$	$\mathbf{C}_3(\mathrm{B})$
t_9	.037	.057	.038	.058	.074	.055	.040	.062	.041
Normal	**.049**	.064	**.048**	.060	.071	.058	**.052**	.067	**.050**
Skew normal	.043	.063	**.046**	.064	.077	.066	**.046**	.066	**.049**
Gamma	.032	**.054**	.033	.061	.086	.065	.034	.058	.037

Table 3.10: Comparison of various ATS-test-statistics based on different hypothesis matrices with distinct number of rows by reference to congruity of the test decisions, the average difference of the p-values, and the maximum difference of the p-values. The hypothesis is B) ($\mathcal{H}_0^v : \mathbf{V}_{111} = ... = \mathbf{V}_{155}$) for dimension $d = 5$ and $n = 50$ with covariance matrix $\mathbf{V} = \mathbf{I}_5 + \mathbf{J}_5$.

of the chosen hypothesis matrix. This behavior causes problems in the analysis of data for hypotheses with $\zeta \neq \mathbf{0}$. Because this never happens in classical mean-based analysis, there are no conventions on how to choose the matrix. For testing for a given covariance matrix, it seems obvious to choose the identity matrix. But if we are interested, for example, in testing for a given trace, it could be less plausible which matrix to choose. So until there are clear conventions, in situations with $\zeta \neq \mathbf{0}$, the hypothesis matrices should be chosen consciously and should be mentioned in the statistical interpretation. This also holds for situations with $\zeta = \mathbf{0}$, where non-quadratic matrices are used. For example, non-quadratic matrices, as we introduced for hypothesis E), are often used in the profile analysis. Thus, results from this area could be influenced by the use of a non-unique hypothesis matrix. Users should know this issue and use some kind of unique matrix as far as possible. Moreover, the choice of a non-unique hypothesis matrix gives an option to adapt the test to get a more conservative or more liberal test. The fact that WTS and MATS have the same results for all hypotheses matrices in our simulations is useful even if especially the WTS has comparably worse simulation results.

Furthermore, we can conclude from our simulations together with Section 6 from Sattler et al. [2022], that the traditional way how for $\zeta = \mathbf{0}$ unique hypothesis matrices are chosen should be adjusted to optimize the computation time.
Usually, the unique projection matrix is chosen and used to calculate all quadratic forms without considering the number of zero rows or rows that are just mul-

tiple rows. Our simulations showed that removing such rows does not change the result but considerably reduces the computation time, especially for higher dimensions. Thus, it would be desirable that the unique projection matrix is not used in such a situation, but the used matrix is formed by removing the previously mentioned rows. Through this clear procedure, uniqueness is not violated. In this way for the example in B) instead of a unique $p \times p$ matrix or a not unique $(d-1) \times p$ matrix we would use a unique $d \times p$ matrix. This choice leads to considerable time reductions, although the test decision is the same as for the unique projection matrix.

At last, we want to deal with the question whether a projection matrix $\mathbf{P} \in \mathbb{R}^{p \times p}$ does always exist to formulate a hypothesis like $\mathbf{P}\boldsymbol{\nu} = \boldsymbol{\zeta}$ or similar. Such matrices exist always, independent of the vector $\boldsymbol{\zeta}$. If for any matrix $\mathbf{C} \in \mathbb{R}^{p \times p}$ it holds $\mathbf{C}\boldsymbol{\nu} = \tilde{\boldsymbol{\zeta}}$ we also know $\mathbf{C}^\top(\mathbf{C}\mathbf{C}^\top)^+\mathbf{C}\boldsymbol{\nu} = \mathbf{C}^\top(\mathbf{C}\mathbf{C}^\top)^+\tilde{\boldsymbol{\zeta}}$ with usage of the Moore-Penrose-inverse. Obviously with $\mathbf{P} := \mathbf{C}^\top(\mathbf{C}\mathbf{C}^\top)^+\mathbf{C}$ and $\boldsymbol{\zeta} := \mathbf{C}^\top(\mathbf{C}\mathbf{C}^\top)^+\tilde{\boldsymbol{\zeta}}$ it fulfills $\mathbf{P}\boldsymbol{\nu} = \boldsymbol{\zeta}$, and $\mathbf{P}$ is a projection matrix. For $\tilde{\boldsymbol{\zeta}} = \mathbf{0}_p$ the other direction is also true, but in general for $\tilde{\boldsymbol{\zeta}} \neq \mathbf{0}_p$ the hypothesis using $\mathbf{C}^\top(\mathbf{C}\mathbf{C}^\top)^+\mathbf{C}$ is larger than the original one. An example of this can be found in the appendix. So in some situations choosing a hypothesis matrix bears the risk of accidentally expanding the null hypothesis and should be done with the requisite care.

3.5.3 Testing hypotheses regarding the correlation matrix

Although the covariance matrix contains much information about the variation of random vectors or similar aspects, there are some disadvantages if we want to analyze the dependency structure of the underlying random vectors. For example, a simple change of the measuring unit can completely change the matrix. This fact is one reason why it is useful to consider the correlation matrix if we are more interested in the dependency structure, which is, for example, an essential part of a statistical model. Therefore, we want to develop test statistics to examine hypotheses regarding the correlation matrix.

To this aim, we first remember our general semiparametric model, given by independent d-dimensional random vectors of the shape

$$\mathbf{X}_{ik} = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_{ik},$$

where the index $i = 1, \dots, a$ refers to the treatment group and $k = 1, \dots, n_i$ to the individual, on which d-variate observations are measured. Of course, it is useless to analyze a scalar's correlation, so we assume $d \geqslant 2$.
The residuals $\boldsymbol{\epsilon}_{i1}, \dots, \boldsymbol{\epsilon}_{in_i}$ are assumed to be centred $\mathbb{E}(\boldsymbol{\epsilon}_{i1}) = \mathbf{0}_d$ and i.i.d. within each group, with finite fourth moment $\mathbb{E}(||\boldsymbol{\epsilon}_{i1}||^4) < \infty$. Finally, we need again:

(A1) $\frac{n_i}{N} \to \kappa_i \in (0, 1]$, $i = 1, ..., a$ for $\min(n_1, \dots, n_a) \to \infty$ with $N = \sum_{i=1}^{a} n_i$.

The only difference to the conditions for testing hypotheses regarding the covariance is that the covariance matrix here has to be strictly positive definite and not just semidefinite. So we require $Cov(\boldsymbol{\epsilon}_{i1}) = \mathbf{V}_i > 0$, $i = 1, \dots, a$. While we used the so-call half-vectorization operation vech for the covariance matrix, this is not the best choice here. For a correlation matrix, the diagonal elements are always one and therefore contain no information. Hence, a new vectorization operation vech^- is defined, which we will call the upper-half-vectorization. With $\mathbf{R}_i$ the correlation matrix for the i-th group, this vectorization operation allows us to define

$$\mathbf{r}_i = \text{vech}^-(\mathbf{R}_i) = (r_{i12}, \dots, r_{i1d}, r_{i23}, \dots, r_{i2d}, \dots, r_{i(d-1)d})^\top, \ i = 1, \dots, a,$$

containing just the upper triangular entries of $\mathbf{R}_i$ which are not on the diagonal. The resulting vector has the dimension $p_u = d(d-1)/2$ which is substantially smaller than p. Therewith, we formulate hypotheses in terms of the pooled correlation vector $\mathbf{r} = (\mathbf{r}_1^\top, \dots, \mathbf{r}_a^\top)^\top$ as

$$\mathcal{H}_0^r : \mathbf{C}\mathbf{r} = \boldsymbol{\zeta}, \tag{3.1}$$

with a proper hypothesis matrix $\mathbf{C} \in \mathbb{R}^{m \times a p_u}$ and a vector $\boldsymbol{\zeta} \in \mathbb{R}^m$. Hypotheses which are part of this model are among others:
(a) **Testing Homogeneity of correlation matrices:**

$$\mathcal{H}_0^r : \mathbf{R}_1 = \cdots = \mathbf{R}_a, \text{ resp. } \mathcal{H}_0^r : \mathbf{r}_1 = \cdots = \mathbf{r}_a,$$

as for example investigated in Jennrich [1970].
For $d = 2$ this includes the problem of testing the null hypothesis

$$\mathcal{H}_0^r : \rho_1 = \rho_2 = \cdots = \rho_a$$

of equal correlations $\rho_i = \text{Corr}(X_{i11}, X_{i12}), i = 1, \ldots, a$ within (3.1), which contains testing equality of correlations between two groups. See e.g. Gupta and Xu [2006] or Omelka and Pauly [2012].

(b) **Testing a diagonal correlation matrix:**

$$\mathcal{H}_0^r : \mathbf{R}_1 = \mathbf{I}_d, \text{ resp. } \mathcal{H}_0^r : \mathbf{r}_1 = \mathbf{0}_{p_u}.$$

A test, whether the correlations in a single group are zero, was analyzed, e.g., introduced in Bartlett [1951]. With this hypothesis, we can evaluate the suitability of a dataset or correlation matrix for factor analysis.

More hypotheses on the structure of the covariance matrix can be found for example in Joereskog [1978], Steiger [1980] and Wu et al. [2018].

(c) **Testing for a given correlation:** Let $\mathbf{R}$ be a given correlation matrix, like an autoregressive or compound symmetry matrix. For $a = 1$, we then also cover testing the null hypothesis

$$\mathcal{H}_0^r : \mathbf{R}_1 = \mathbf{R} \text{ resp. } \mathbf{r}_1 = \text{vech}^-(\mathbf{R}) \text{ for a given matrix } \mathbf{R}.$$

For $d = 2$, this also contains the issue of testing the null hypothesis

$$\mathcal{H}_0^r : \rho_1 = 0$$

of uncorrelated random variables with $\rho_1 = \text{Corr}(X_{111}, X_{112})$, see e.g. Aitkin et al. [1968].

As explained in Sattler et al. [2022], $\mathbf{C}$ does not have to be a projection matrix as ζ is allowed to be different from the zero vector.

Asymptotics regarding the vectorized correlation

To check null hypotheses of the kind $\mathcal{H}_0^r : \mathbf{C}\mathbf{r} = \zeta$, it is necessary to first investigate the asymptotic distribution of $\mathbf{C}\widehat{\mathbf{r}}$.
To this end, a result from Sattler et al. [2022] is shortly repeated first. For $\boldsymbol{v} = (\boldsymbol{v}_1^\top, ..., \boldsymbol{v}_a^\top)^\top = (\text{vech}(\mathbf{V}_1)^\top, ..., \text{vech}(\mathbf{V}_a)^\top)^\top$ and $\widehat{\boldsymbol{v}}$ as the empirical coun-

terpart, it holds

$$\sqrt{N}\mathbf{C}(\widehat{\boldsymbol{v}} - \boldsymbol{v}) \xrightarrow{\mathcal{D}} \mathcal{N}_{ap}\left(\mathbf{0}_{ap}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}^\top\right),$$

where $\boldsymbol{\Sigma} = \bigoplus_{i=1}^{a} \frac{1}{k_i}\boldsymbol{\Sigma}_i$ and $\boldsymbol{\Sigma}_i = Cov(\text{vech}(\boldsymbol{\epsilon}_{i1}\boldsymbol{\epsilon}_{i1}^\top)^\top)$ for $i = 1, \dots, a$.

To use this result for the correlation, first, some matrices have to be defined. Let $\boldsymbol{e}_{k,p} = (\delta_{k\ell})_{\ell=1}^{p}$ define the p-dimensional vector which contains a one in the k-th component and zeros elsewhere. Moreover, we need a d-dimensional auxiliary vector $\boldsymbol{a} = (a_1, ..., a_d)$, given through $a_k = 1 + \sum_{j=1}^{k-1}(d + 1 - j)$, $k = 1, ..., d$. It contains the position of components in the half-vectorized matrix which are the diagonal elements of the original matrix. In accordance to this, we define the p_u-dimensional vector $\boldsymbol{b}$ which contains the numbers from one to p in ascending order without the elements from $\boldsymbol{a}$. This vector $\boldsymbol{b}$ contains the position of components in the half-vectorized matrix which are non diagonal elements. With these vectors we are able to define a $d \times d$ matrix $\mathbf{H} = \mathbf{1}_d \cdot \boldsymbol{a}$ and the vectors $\boldsymbol{h}_1 = \text{vech}^-(\mathbf{H})$ and $\boldsymbol{h}_2 = \text{vech}^-(\mathbf{H}^\top)$. Finally, we can define the matrices

$$\mathbf{M}_1 = \sum_{\ell=1}^{p_u} \boldsymbol{e}_{\ell,p_u} \cdot (\boldsymbol{e}_{\boldsymbol{h}_{1\ell},p} + \boldsymbol{e}_{\boldsymbol{h}_{2\ell},p})^\top \quad \text{and} \quad \mathbf{L}_p^u = \sum_{\ell=1}^{p_u} \boldsymbol{e}_{\ell,p_u} \cdot \boldsymbol{e}_{\boldsymbol{b}_{\ell},p}^\top.$$

This allows us to formulate a connection between the vech operator and the vech$^-$ operator, since the matrix $\mathbf{L}_p^u$ fulfills $\mathbf{L}_p^u\,\text{vech}(\boldsymbol{A}) = \text{vech}^-(\boldsymbol{A})$ for each arbitrary matrix $\boldsymbol{A} \in \mathbb{R}^{p\times p}$. This matrix is comparable to the elimination matrix from Magnus and Neudecker [1980] and adapted to this special kind of half-vectorization.

With all these matrices, a conjunction can be found between $\sqrt{n_i}(\widehat{\boldsymbol{v}}_i - \boldsymbol{v}_i)$ and $\sqrt{n_i}(\widehat{\boldsymbol{r}}_i - \boldsymbol{r}_i)$, which allows to get the requested result by applying Theorem 3.1 from Sattler et al. [2022]. The approach to connect vectorized correlation and vectorized covariance is based on Browne and Shapiro [1986] and Nel [1985], and adapted to the setting of our work.

Theorem 3.5.1:
With the previously defined matrices $\mathbf{L}_p^u$, $\mathbf{M}_1$ *and*

$$\mathbf{M}(\boldsymbol{v}_i, \boldsymbol{r}_i) := \left[\mathbf{L}_p^u - \frac{1}{2}diag(\boldsymbol{r}_i)\mathbf{M}_1\right] \cdot diag(\text{vech}((v_{i11}, ..., v_{idd})^\top \cdot (v_{i11}, ..., v_{idd})))^{-\frac{1}{2}},$$

it holds

$$\sqrt{n_i}(\widehat{\mathbf{r}}_i - \mathbf{r}_i) = \mathbf{M}(\boldsymbol{\nu}_i, \mathbf{r}_i) \cdot \sqrt{n_i}(\widehat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_i) + o_P(1).$$

Thus

$$\sqrt{n_i}(\widehat{\mathbf{r}}_i - \mathbf{r}_i) \xrightarrow{\mathcal{D}} \mathbf{Z}_i \sim \mathcal{N}_{p_u}\left(\mathbf{0}_{p_u}, \underbrace{\mathbf{M}(\boldsymbol{\nu}_i, \mathbf{r}_i)\boldsymbol{\Sigma}_i\mathbf{M}(\boldsymbol{\nu}_i, \mathbf{r}_i)^\top}_{=:\boldsymbol{\Upsilon}_i}\right)$$

and

$$\sqrt{N}(\widehat{\mathbf{r}} - \mathbf{r}) \xrightarrow{\mathcal{D}} \mathbf{Z} \sim \mathcal{N}_{ap_u}\left(\mathbf{0}_{ap_u}, \bigoplus_{i=1}^{a} \frac{1}{\kappa_i}\boldsymbol{\Upsilon}_i\right) = \mathcal{N}_{ap_u}(\mathbf{0}_{ap_u}, \boldsymbol{\Upsilon}).$$

To use this result, we have to estimate the matrices $\boldsymbol{\Upsilon}_1, ..., \boldsymbol{\Upsilon}_a$, which is done by using

$$\widehat{\boldsymbol{\Upsilon}}_i = \mathbf{M}(\widehat{\boldsymbol{\nu}}_i, \widehat{\mathbf{r}}_i)\widehat{\boldsymbol{\Sigma}}_i\mathbf{M}(\widehat{\boldsymbol{\nu}}_i, \widehat{\mathbf{r}}_i)^\top,$$

and $\widehat{\boldsymbol{\Upsilon}} := \bigoplus_{i=1}^{a} \frac{n_i}{N}\widehat{\boldsymbol{\Upsilon}}_i$. It is obvious that these estimators are consistent, since they consist of consistent estimators and continous functions applied to them.

With this asymptotic result, test statistics based on quadratic forms can be formulated through:

Theorem 3.5.2:
Let $\mathbf{E}(\mathbf{C}, \widehat{\boldsymbol{\Upsilon}}) \in \mathbb{R}^{m \times m}$ be some symmetric matrix which can be written as a function of the hypothesis matrix $\mathbf{C} \in \mathbb{R}^{m \times ap_u}$ and the covariance matrix estimator $\widehat{\boldsymbol{\Upsilon}} \in \mathbb{R}^{ap_u \times ap_u}$. Additionally, it holds $\mathbf{E}(\mathbf{C}, \widehat{\boldsymbol{\Upsilon}}) \xrightarrow{\mathcal{P}} \mathbf{E}(\mathbf{C}, \boldsymbol{\Upsilon})$. Then, under the null hypothesis $\mathcal{H}_0^r : \mathbf{C}\mathbf{r} = \boldsymbol{\zeta}$, the quadratic form $\widehat{Q}_r$ defined by

$$\widehat{Q}_{\mathbf{r}} = N\,[\mathbf{C}\widehat{\mathbf{r}} - \boldsymbol{\zeta}]^\top\,\mathbf{E}(\mathbf{C}, \widehat{\boldsymbol{\Upsilon}})\,[\mathbf{C}\widehat{\mathbf{r}} - \boldsymbol{\zeta}]$$

has asymptotically a "weighted χ^2-distribution", i.e. for $N \to \infty$ it holds that

$$\widehat{Q}_{\mathbf{r}} \xrightarrow{\mathcal{D}} \sum_{\ell=1}^{ap_u} \lambda_\ell B_\ell,$$

where $B_\ell \overset{i.i.d.}{\sim} \chi_1^2$ and $\lambda_\ell, \ell = 1, \ldots, ap_u$, are the eigenvalues of $\boldsymbol{\Upsilon}^{1/2}\mathbf{C}^\top\mathbf{E}(\mathbf{C}, \boldsymbol{\Upsilon})\mathbf{C}\boldsymbol{\Upsilon}^{1/2}$.

So all quadratic forms from Sattler et al. [2022] can also be formulated for the correlation, for example

$$WTS_{\mathbf{r}} = N\left[\mathbf{C}(\widehat{\mathbf{r}}-\mathbf{r})\right]^\top \left(\mathbf{C}\widehat{\boldsymbol{\Upsilon}}\mathbf{C}^\top\right)^+ \left[\mathbf{C}(\widehat{\mathbf{r}}-\mathbf{r})\right],$$

$$ATS_{\mathbf{r}} = N\left[\mathbf{C}(\widehat{\mathbf{r}}-\mathbf{r})\right]^\top \left[\mathbf{C}(\widehat{\mathbf{r}}-\mathbf{r})\right] / \operatorname{tr}\left(\mathbf{C}\widehat{\boldsymbol{\Upsilon}}\mathbf{C}^\top\right),$$

or

$$MATS_{\mathbf{r}} = N\left[\mathbf{C}(\widehat{\mathbf{r}}-\mathbf{r})\right]^\top \left(\mathbf{C}\widehat{\boldsymbol{\Upsilon}}_0\mathbf{C}^\top\right)^+ \left[\mathbf{C}(\widehat{\mathbf{r}}-\mathbf{r})\right],$$

while $\mathbf{A}_0$ denotes a matrix which only contains the diagonal elements of $\mathbf{A}$. This leads to $\varphi_{WTS_{\mathbf{r}}} = \mathbb{1}\{WTS_{\mathbf{r}} \notin (-\infty, \chi^2_{rank(\mathbf{C});1-\alpha}]\}$, which needs $\boldsymbol{\Upsilon} > 0$. The simulation results from Sattler et al. [2022] as well as from Section 3.5.1 suggest to use a Monte-Carlo version of the $ATS_{\mathbf{r}}$ given by $\varphi_{ATS_{\mathbf{r}}} := \mathbb{1}\{ATS_{\mathbf{r}} \notin (-\infty, q^{MC}_{1-\alpha}]\}$.

It would be more convenient to find a direct connection between $\widehat{\mathbf{r}} - \mathbf{r}$ and $\operatorname{vech}^-(\widehat{\mathbf{V}} - \mathbf{V})$, instead of having a result for the half vectorization and using this for developing results for the upper half-vectorization. Unfortunately, this has not been possible so far, although a result for $\sqrt{N}\operatorname{vech}^-(\widehat{\mathbf{V}} - \mathbf{V})$ could be developed analogue to Theorem 3.1 from Sattler et al. [2022]. However, it has to be taken into account that we need the covariance matrix's diagonal elements to calculate the correlation matrix. Moreover, dependencies between components, as well as the structure of $\widehat{\mathbf{R}} - \mathbf{R}$, make this task quite challenging. Therefore, this kind of workaround has to be done, which has just minimal impact on the computation time or similar aspects.

Resampling Procedures

Again a resampling procedure may be useful, on the one hand, for a better small sample approximation and, on the other hand, for quadratic forms with critical values that are difficult to calculate. Since the simulations from Sattler et al. [2022] showed clear advantages of the parametric bootstrap, we only consider this approach.

Thus, for every group with realisations $\mathbf{X}_{i1}, ..., \mathbf{X}_{in_i}$ we calculate the covariance matrix $\widehat{\boldsymbol{\Upsilon}}_i$. With this covariance matrix we generate random vectors $\mathbf{Y}^*_{i1}, ..., \mathbf{Y}^*_{in_i} \stackrel{i.i.d}{\sim} \mathcal{N}_{p_u}(\mathbf{0}_{p_u}, \widehat{\boldsymbol{\Upsilon}}_i)$ which are independent from the realisations and calculate their sample covariance $\widehat{\boldsymbol{\Upsilon}}^*_i$ respectively $\widehat{\boldsymbol{\Upsilon}}^* := \bigoplus_{i=1}^{a} \frac{1}{\kappa_i}\widehat{\boldsymbol{\Upsilon}}^*_i$. For these random vectors we now consider the asymptotic distribution.

Theorem 3.5.3:
If Assumption (A1) is fulfilled, it holds:
(a) For $i = 1, ..., a$*, the conditional distribution of* $\sqrt{N}\, \overline{\mathbf{Y}}_i^*$*, given the data, converges weakly to* $\mathcal{N}_{p_u}\left(\mathbf{0}_{p_u}, \kappa_i^{-1} \cdot \boldsymbol{\Upsilon}_i\right)$ *in probability. Since we have* $\widehat{\boldsymbol{\Upsilon}}_i^* \to \boldsymbol{\Upsilon}_i$ *in probability, the unknown covariance matrix* $\boldsymbol{\Upsilon}_i$ *can be estimated through* $\widehat{\boldsymbol{\Upsilon}}_i^*$*.*
(b) The conditional distribution of $\sqrt{N}\, \overline{\mathbf{Y}}^*$*, given the data, converges weakly to* $\mathcal{N}_{ap_u}(\mathbf{0}_{ap_u}, \boldsymbol{\Upsilon})$ *in probability. Since we have* $\widehat{\boldsymbol{\Upsilon}}^* \to \boldsymbol{\Upsilon}$ *in probability, the unknown covariance matrix* $\boldsymbol{\Upsilon}$ *can be estimated through* $\widehat{\boldsymbol{\Upsilon}}^*$*.*

In consequence of Theorem 3.5.3 it is reasonable to calculate the bootstrap version of the previous quadratic forms, which are

$$Q_{\mathbf{r}}^* = N \left[\mathbf{C}\overline{\mathbf{Y}}^*\right]^\top \mathbf{E}(\mathbf{C}, \widehat{\boldsymbol{\Upsilon}}^*) \left[\mathbf{C}\, \overline{\mathbf{Y}}^*\right].$$

Similar to Sattler et al. [2022], two important quadratic forms are given by

$$\begin{aligned} ATS_{\mathbf{r}}^* &= N \left[\mathbf{C}\overline{\mathbf{Y}}^*\right]^\top \left[\mathbf{C}\overline{\mathbf{Y}}^*\right] / \operatorname{tr}\left(\mathbf{C}\widehat{\boldsymbol{\Upsilon}}^*\mathbf{C}^\top\right), \\ WTS_{\mathbf{r}}^* &= N \left[\mathbf{C}\overline{\mathbf{Y}}^*\right]^\top \left(\mathbf{C}\widehat{\boldsymbol{\Upsilon}}^*\mathbf{C}^\top\right)^+ \left[\mathbf{C}\overline{\mathbf{Y}}^*\right]. \end{aligned}$$

The bootstrap versions approximate the null distribution of $\widehat{Q}_{\mathbf{r}}$, as established below.

Corollary 3.5.1:
For each parameter vector $\mathbf{r} \in \mathbb{R}^{ap_u}$ *and* $\mathbf{r}_0 \in \mathbb{R}^{ap_u}$ *with* $\mathbf{C}\mathbf{r}_0 = \boldsymbol{\zeta}$*, under Assumption (A1) we have*

$$\sup_{x \in \mathbb{R}} \left| P_{\mathbf{r}}(Q_{\mathbf{r}}^* \leqslant x | \mathbf{X}) - P_{\mathbf{r}_0}(\widehat{Q}_{\mathbf{r}} \leqslant x) \right| \xrightarrow{\mathcal{P}} 0,$$

where $P_{\mathbf{r}}$ *denotes the (un)conditional distribution of the test statistic when* $\mathbf{r}$ *is the true underlying vector.*

This motivates the definition of various bootstrap tests like $\varphi_{ATS_{\mathbf{r}}}^* := \mathbb{1}\{ATS_{\mathbf{r}} \notin (-\infty, c_{ATS_{\mathbf{r}}^*, 1-\alpha}]\}$ as asymptotic level α test, with $c_{ATS_{\mathbf{r}}^*, 1-\alpha}$ the conditional quantile of $ATS_{\mathbf{r}}^*$ given the data and similar for WTS and MATS.

As it could be seen in Sattler et al. [2022], in case of just one group it could be useful to generate the bootstrap observations with a $\mathcal{N}_m\left(\mathbf{0}_m, \mathbf{C}\widehat{\boldsymbol{\Upsilon}}_1\mathbf{C}^\top\right)$ distribution and adapt the quadratic form for this. From a theoretical point of view, it changes nothing, but depending on the hypothesis matrix's dimension, this positively affects the computation time. This will be treated in more detail in Section 3.5.5.

In the analysis of correlation matrices, Fisher z-transformed vectors are often used instead of the original vectorized correlation matrices. Although the root of this approach is the distribution of the Fisher z-transformed correlation in the case of normally distributed observations, it is also used for tests without this distributional restriction, see, e.g., Steiger [1980]. So we could also consider our tests together with the transformed vector. This approach assumes that all components of ζ differ from one, which is always possible to ensure. We can define tests based on the transformation for each of our quadratic forms, including the tests based on bootstrap or Monte-Carlo simulations.
Our simulations showed that the tests based on this transformation have more liberal behavior than the original one. Since all of our test statistics were already a bit liberal, it is not useful to consider these versions further.

Simulations

To investigate the performance we analyze the type-I-error rate of the following two hypotheses

$A_\mathbf{r}$) Homogeneity of correlation matrices: $\mathcal{H}_0^\mathbf{r}: \mathbf{R}_1 = \mathbf{R}_2$,

$B_\mathbf{r}$) Diagonal structure of the covariance matrix $\mathcal{H}_0^\mathbf{r}: \mathbf{R}_1 = \mathbf{I}_p$ resp. $\mathbf{r}_1 = \mathbf{0}_{p_u}$,

with $\alpha = 0.05$. The hypothesis matrices are chosen as the projection matrices $\mathbf{C}(A_\mathbf{r}) = \mathbf{P}_2 \otimes \mathbf{I}_{p_u}$ and $\mathbf{C}(B_\mathbf{r}) = \mathbf{I}_{p_u}$ while ζ is in both cases a zero vector with appropriate dimension.

The setting is very similar to Sattler et al. [2022] so we used d=5 and therefore $p_u = 10$, while for one group we have $n \in \{25, 50, 125, 250\}$ and for two groups we have $n_1 = 0.6 \cdot N$ and $n_2 = 0.4 \cdot N$ with $N \in \{50, 100, 250, 500\}$. The used error terms are based on

- a standard normal distribution, i.e. $Z_{ikj} \sim \mathcal{N}(0, 1)$.
- a standardized centred gamma distribution i.e. $(\sqrt{2}Z_{ikj} + 2) \sim \mathcal{G}(2, 1)$
- a standardized centred skew normal distribution with location parameter $\xi = 0$, scale parameter $\omega = 1$ and $\alpha = 4$. The density of a skew normal distribution is given through $\frac{2}{\omega}\varphi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\left(\frac{x-\xi}{\omega}\right)\right)$, where φ denotes the density of a standard normal distribution and Φ the according distribution function.
- a standardized centred t-distribution with 9 degrees of freedom,

together with different covariance matrices for the varying hypotheses. For $A_{\mathbf{r}}$) we use $(\mathbf{V}_1)_{ij} = 0.6^{|i-j|}$ resp. $\mathbf{V}_1 = \mathbf{I}_d + 0.5\mathbf{J}_d$ for the first group and for the second group we multiply these covariance matrices with $diag(1, 1.2, ..., 1.8)$. Here we have a setting where the covariance matrices are different but the correlation matrices are equal. To investigate $B_{\mathbf{r}}$) we just consider one matrix, given by $\mathbf{V}_3 = diag(1, 1.2, ..., 1.8)$.

For both hypotheses there exist already tests, while some of them are part of the R-package *psych* by Revelle [2019]. We want to compare the type-I-error rate of our tests with $\varphi_{\texttt{Jennrich}}$ from Jennrich [1970] and $\varphi_{\texttt{Steiger}}$ resp. $\varphi_{\texttt{Steiger}_{\texttt{Fz}}}$ from Steiger [1980] for equality of correlation matrices. Hereby $\varphi_{\texttt{Steiger}_{\texttt{Fz}}}$ is the same test statistic as $\varphi_{\texttt{Steiger}}$ but uses a Fisher z-transformation on the vectorized correlation matrices.

Testing wheter the correlation matrix is equal to the identity matrix can be investigated with $\varphi_{\texttt{Bartlett}}$ from Bartlett [1951] and again with $\varphi_{\texttt{Steiger}}$ resp. $\varphi_{\texttt{Steiger}_{\texttt{Fz}}}$.

We use 1,000 bootstrap steps for our parametric bootstrap, 10,000 simulation steps for the Monte-Carlo approach and 10,000 runs for all tests to get reliable results. Hereby, the actual test statistic is multiplied with the factor $(N-3)/N$ for $\varphi_{\texttt{Steiger}}$ and $\varphi_{\texttt{Steiger}_{\texttt{Fz}}}$. This approach is based on a specific result of the Fisher z-transformation of the correlation vector of normal distributed random vectors. This multiplication's main purpose seems to be the less liberal behavior of the test, while it has asymptotically no effect. To get a better impression of the impact of such a multiplication, we also include our ATS with parametric boot-

strap using such multiplication and denote this with an m for multiplication. This also simplifies the comparison of the tests under equal conditions.
The results of hypothesis A_r) can be seen in Table 3.11 and Table 3.12 for the different covariance matrices. It is interesting to note that the type-I-error rate of $\varphi_{\text{Steifer}_{\text{Fz}}}$ differs more and more from the 5% rate for increasing sample sizes. Therefore, this test should not be used, at least for our setting. However, without the Fisher z-transformation, Steiger's test is way too conservative, especially for larger sample sizes. In contrast, φ_{ATS^*} and φ_{ATS} are too liberal but show the best type-I-error rate for N greater than 50. Moreover, these tests are the only ones of the considered tests which fulfill Bradley's liberal criterion (from Bradley [1978]) stably for N larger than 100. Similar to the results for covariance matrices, the bootstrap version has slightly better results than Monte-Carlo-based tests, while the error rates get closer for greater sample sizes. Again the WTS is way too liberal and needs large sample sizes, despite the bootstrap approach. The less known MATS statistic based on a parametric bootstrap is better but clearly worse than both tests based on the ATS. Hence as well as the WTS, this test seems not to be recommendable for testing equality of correlation matrices. However, except for the gamma-distribution, it fulfills Bradley's liberal criterion for N larger than 100. At last, the test of Jennrich is even more liberal than the ATS-based test. For $(\boldsymbol{V})_{ij} = 0.6^{|i-j|}$ the error rates are always higher than 24% even for the sample size $N = 500$. All in all, the only tests among the considered ones that should be used for this hypothesis are φ_{ATS} and φ_{ATS^*}. It can be seen that with a correction factor, like $(N-3)/N$, their small sample performance could be clearly improved. Even without any correction, they are preferable to tests based on Steiger [1980], using this factor.

For hypothesis $B_{\mathbf{r}}$), the results are included in Table 3.13. Again φ_{ATS_r} and $\varphi_{ATS_r^*}$ have the best results of all our test statistics, while the results are considerably better than for hypothesis A_r). For example, Bradley's liberal criterion holds for all sample sizes except for $n_1 = 25$. Moreover, the better performance in hypothesis $B_{\mathbf{r}}$) than hypothesis $A_{\mathbf{r}}$) can be observed through the number of values in the 95% binomial interval $[0.0458, 0.0543]$. With the correction factor, this test has a better type-I-error rate than φ_{Steiger} and comparable to $\varphi_{\text{Steiger}_{\text{Fz}}}$. Nevertheless, $\varphi_{\text{Bartlett}}$ is a test only developed for this one hypothesis and therefore has an excellent error rate through all distributions.

Both hypotheses show that our developed tests are useful in many situations, although partwise large sample sizes are necessary to get good results. This is a known fact for testing hypotheses regarding correlation matrices, which was, for example, mentioned in Steiger [1980].

In addition to the type-I-error rate, the ability to detect deviations from the null hypothesis is an important criterion of a test. To this aim, we also investigate the power of some of the tests mentioned above. We choose a quite simple kind of alternative, which is suitable for our situation. As covariance matrix we consider $\mathbf{V}_1 + \delta \cdot \mathbf{J}_d$ for $\delta \in [0, 2.5]$ in hypothesis A_r) and for $\delta \in [0, 0.65]$ in hypothesis $B_\mathbf{r}$). The reason for this considerable difference in the δ range is that for hypothesis $B_\mathbf{r}$), the addition changes the setting from uncorrelated to correlated. For hypothesis $A_\mathbf{r}$), it just increases the correlations, which is clearly more challenging to detect.
Due to computation time, we simulate only one sample size, which is $N = 250$ resp. $n_1 = 125$ and consider error terms based on the skew normal distribution and the Gamma distribution. The Monte-Carlo steps, bootstrap steps and the simulation runs are the same as before. We simulate only the test with good results for its type-I-error rate, which was for $A_\mathbf{r}$) $\varphi_{ATS_r^*}$, and as comparisons $\varphi_{Steiger}$ and $\varphi_{Jennrich}$.
Based on the results from Table 3.13 for hypothesis $B_\mathbf{r}$) we only consider φ_{ATS^*} as well as $\varphi_{SteigerFz}$ and $\varphi_{Bartlett}$ while the setting is the same. Because of the similarity of the results from the parametric bootstrap and the Monte-Carlo-based approach, we use just one. But for both hypotheses, we also investigate the ATS with parametric bootstrap and the multiplied factor to see the influence of the multiplication again.
For hypothesis $A_\mathbf{r}$) Figure 3.1 shows that despite the very liberal behaviour of $\varphi_{Jennrich}$, for $\delta \geqslant 1$ the power is nearly the same as $\varphi_{ATS_r^*}$. Given the fact that $\varphi_{Jennrich}$ has a type-I-error rate of about 0.28, which is more than 0.2 higher than both tests based on the ATS. With such a high type-I-error rate, the test based on Jennrich [1970] can not be recommended. On the contrary, $\varphi_{Steiger}$ has little power due to its conservative behavior and therefore needs δ to be larger than 2.5 to reach power close to one. For all of the considered tests, both tests based on the ATS, show by far the best power because of the relatively steep slope. This slope allows detecting deviations from the null hypothesis without being too liberal. The multiplication makes the test slightly less liberal and therefore lowers the power.

	t_9				Normal			
N	50	100	250	500	50	100	250	500
ATS-Para	.1096	.0821	.0589	.0562	.1085	.0739	.0579	.0566
ATS-Wild	.1327	.0938	.0638	.0576	.1309	.0830	.0607	.0587
ATS-Para-m	.0932	.0761	.0566	.0556	.0918	.0659	.0547	.0553
ATS	.1166	.0832	.0590	.0544	.1160	.0746	.0585	.0560
ATSFz	.1092	.0809	.0588	**.0537**	.1080	.0715	.0565	.0560
ATSFz-m	.0929	.0732	.0559	**.0529**	.0900	.0662	**.0541**	.0545
WTS	.4700	.2443	.1118	.0861	.4257	.2061	.0988	.0769
Steiger	.0108	.0206	.0224	.0245	.0109	.0176	.0215	.0287
SteigerFz	.0606	.0821	.0873	.0928	.0601	.0744	.0861	.0954
Jennrich	.3341	.2837	.2533	.2582	.3306	.2808	.2515	.2472
	Skew Normal				Gamma			
N	50	100	250	500	50	100	250	500
ATS-Para	.1104	.0804	.0613	.0590	.1230	.0899	.0632	.0596
ATS-Wild	.1329	.0898	.0640	.0610	.1462	.1030	.0689	.0619
ATS-Para-m	.0941	.0736	.0583	.0577	.1054	.0829	.0598	.0585
ATS	.1179	.0822	.0593	.0588	.1293	.0920	.0633	.0583
ATSFz	.1117	.0800	.0586	.0580	.1243	.0901	.0621	.0582
ATSFz-m	.0934	.0727	.0560	.0563	.1055	.0813	.0593	.0573
WTS	.4563	.2364	.1186	.0853	.5257	.3074	.1498	.1028
Steiger	.0101	.0187	.0222	.0294	.0098	.0183	.0208	.0261
SteigerFz	.0569	.0803	.0872	.0952	.0654	.0828	.0894	.0929
Jennrich	.3375	.2858	.2590	.2476	.3424	.2942	.2651	.2538

Table 3.11: Simulated type-I-error rates ($\alpha = 5\%$) in scenario A_r) ($\mathcal{H}_0^v : \mathbf{R}_1 = \mathbf{R}_2$) for ATS, WTS, Steiger's and Jennrich's test. The observation vectors have dimension 5, covariance matrix $(\mathbf{V}_1)_{ij} = 0.6^{|i-j|}$ resp. $\mathbf{V}_2 = diag(1, 1.2, ..., 1.8) \cdot \mathbf{V}_1$ and and it always holds $n_1 := 0.6 \cdot N$ resp. $n_2 := 0.4 \cdot N$.

	t_9				Normal			
N	50	100	250	500	50	100	250	500
ATS-Para	.1180	.0847	.0596	.0546	.1102	.0714	.0556	.0601
ATS-Wild	.1590	.1024	.0669	.0596	.1482	.0881	.0611	.0616
ATS-Para-m	.0959	.0759	.0575	**.0532**	.0897	.0645	**.0533**	.0588
ATS	.1262	.0872	.0595	.0550	.1170	.0730	.0558	.0601
ATSFz	.1193	.0859	.0589	.0545	.1132	.0709	.0552	.0593
ATSFz-m	.0966	.0757	.0562	**.0521**	.0929	.0633	**.0531**	.0577
WTS	.5194	.2776	.1249	.0912	.4784	.2339	.1063	.0814
Steiger	.0099	.0181	.0242	.0276	.0097	.0167	.0205	.0287
SteigerFz	.0431	.0648	.0689	.0749	.0410	.0562	.0658	.0768
Jennrich	.1341	.0977	.0805	.0810	.1319	.0983	.0767	.0805

	Skew Normal				Gamma			
N	50	100	250	500	50	100	250	500
ATS-Para	.1168	.0842	.0606	.0593	.1312	.0959	.0612	.0595
ATS-Wild	.1549	.1016	.0675	.0620	.1746	.1203	.0703	.0667
ATS-Para-m	.0973	.0749	.0574	.0580	.1079	.0872	.0586	.0586
ATS	.1233	.0851	.0595	.0590	.1384	.0992	.0614	.0591
ATSFz	.1201	.0828	.0588	.0592	.1332	.0973	.0615	.0589
ATSFz-m	.0983	.0741	.0561	.0577	.1108	.0870	.0579	.0574
WTS	.5103	.2648	.1272	.0891	.5720	.3299	.1586	.1040
Steiger	.0080	.0186	.0242	.0297	.0061	.0167	.0199	.0243
SteigerFz	.0432	.0626	.0698	.0768	.0445	.0624	.0661	.0754
Jennrich	.1275	.0975	.0812	.0806	.1277	.0961	.0768	.0782

Table 3.12: Simulated type-I-error rates ($\alpha = 5\%$) in scenario A_r) ($\mathcal{H}_0^r : \mathbf{R}_1 = \mathbf{R}_2$) for ATS, WTS, Steiger's and Jennrich's test. The observation vectors have dimension 5, covariance matrices $(\mathbf{V}_1) = \mathbf{I}_5 + 0.5 \cdot \mathbf{J}_5$ resp. $\mathbf{V}_2 = diag(1, 1.2, ..., 1.8) \cdot \mathbf{V}_1$ and it always holds $n_1 := 0.6 \cdot N$ resp. $n_2 := 0.4 \cdot N$.

	t_9				Normal			
ATS-Para	.0928	.0639	**.0515**	**.0543**	.0798	.0587	**.0486**	**.0526**
ATS-Wild	.1859	.1128	.0768	.0662	.1584	.0984	.0644	.0579
ATS-Para-m	**.0507**	.0450	.0445	**.0502**	.0430	.0415	.0423	**.0479**
ATS	.1050	.0681	**.0531**	**.0532**	.0893	.0617	**.0480**	**.0514**
ATSFz	.0989	.0642	**.0516**	**.0526**	.0860	.0589	**.0467**	**.0508**
ATSFz-m	**.0539**	.0432	.0451	**.0486**	**.0464**	.0419	.0413	**.0467**
ATS-PCov	.0755	**.0541**	**.0464**	**.0482**	.0733	**.0531**	.0436	**.0470**
ATS-WCov	.0826	.0666	.0628	.0598	.0798	.0673	.0557	.0562
ATS-Cov	.0878	.0562	**.0461**	**.0476**	.0806	.0544	.0440	**.0464**
WTS	.8423	.5194	.2332	.1393	.8142	.4794	.2021	.1173
Steiger	.0245	.0343	.0436	**.0481**	.0223	.0330	.0419	**.0458**
SteigerFz	.0546	**.0499**	**.0509**	**.0528**	**.0528**	**.0491**	**.0485**	**.0488**
Bartlett	**.0543**	**.0493**	**.0520**	**.0519**	**.0516**	**.0482**	**.0467**	**.0488**
	Skew Normal				Gamma			
ATS-Para	.0872	.0630	.0568	**.0516**	.1066	.0740	.0573	**.0519**
ATS-Wild	.1686	.1079	.0792	.0601	.2132	.1342	.0810	.0663
ATS-Para-m	.0457	.0444	**.0489**	**.0488**	.0582	.0559	**.0495**	**.0479**
ATS	.0966	.0674	.0575	**.0507**	.1169	.0776	.0563	**.0511**
ATSFz	.0936	.0639	.0564	**.0497**	.1059	.0710	.0549	**.0504**
ATSFz-m	**.0493**	.0445	**.0492**	**.0461**	.0572	**.0540**	**.0483**	**.0474**
ATS-PCov	.0715	.0564	**.0507**	**.0458**	.0760	.0595	**.0477**	**.0464**
ATS-WCov	.0785	.0672	.0634	.0566	.0714	.0668	.0620	.0569
ATS-Cov	.0819	.0586	**.0516**	**.0461**	.0899	.0639	**.0482**	**.0469**
WTS	.8206	.4994	.2252	.1260	.8518	.5661	.2680	.1544
Steiger	.0226	.0349	**.0492**	.0455	.0246	.0385	**.0473**	**.0487**
SteigerFz	**.0532**	**.0515**	.0555	**.0494**	.0574	.0547	**.0534**	**.0506**
Bartlett	**.0479**	**.0504**	.0544	**.0500**	.0551	**.0515**	**.0541**	**.0487**

Table 3.13: Simulated type-I-error rates ($\alpha = 5\%$) in scenario $B_{\mathbf{r}}$) ($\mathcal{H}_0^{\mathbf{r}} : \mathbf{r} = \mathbf{0}_{10}$) for ATS, WTS, Steiger's and Bartlett's test. The observation vectors have dimension 5 and covariance matrix $\mathbf{V} = diag(1, 1.2, 1.4, 1.6, 1.8)$.

In Figure 3.2 for hypothesis B_r) it can be seen that $\varphi_{ATS_r^*}$ has clearly more power than $\varphi_{Bartlett}$ which even grows faster. While Bartlett's test is preferable with regard to the type-I-error rate, our tests are favorable while concerning the power. The test based on the Fisher z-transformation has a similar slope to our test but is less liberal. However, it could be seen, that with such a factor, our test have the best type-I-error rate. The slope of the corresponding power curve is similar to $\varphi_{ATS_r^*}$ and $\varphi_{SteigerFz}$. For power, the chosen distribution hardly has any influence.

All in all, the simulation showed that our developed test has better performance than the existing test in hypothesis A_r), while for B_r), there is no clear choice between our test and the test based on Bartlett [1951]. It depends on whether type-I-error or the capability of detecting deviation from the null hypothesis has a higher priority.

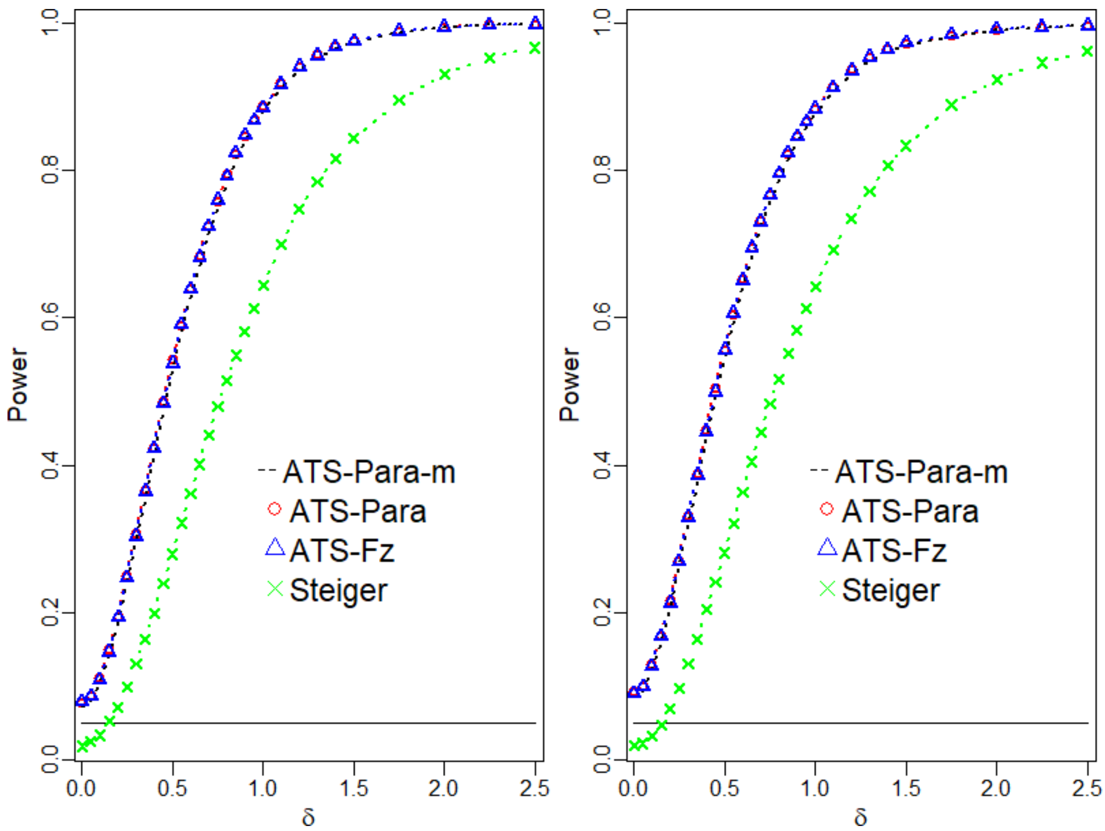


Figure 3.1: Simulated power curves of different tests for the hypothesis A_r) ($\mathcal{H}_0^r : \mathbf{R}_1 = \mathbf{R}_2$), with $d = 5$, $N = 100$ and $n_1 = 0.6 \cdot N$. The covariance matrix is $(\mathbf{V}_2)_{ij} = 0.6^{|i-j|}$ resp. $\mathbf{V}_1 = \mathbf{V}_2 + \delta \mathbf{J}_5$ and the error terms are based on skew normal distribution (left side) or gamma distribution (right side).

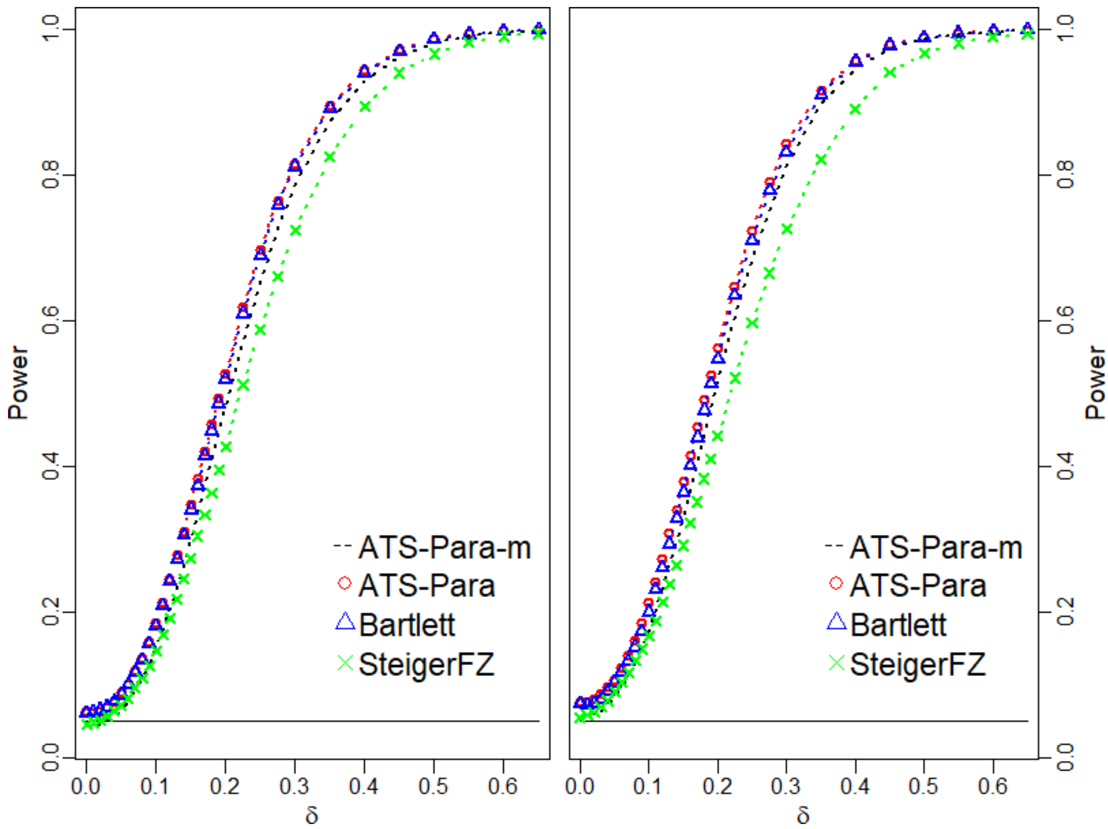


Figure 3.2: Simulated power curves of different tests for the hypothesis $B_{\mathbf{r}}$ ($\mathcal{H}_0^{\mathbf{r}}$: $\mathbf{r}_1 = \mathbf{0}_{10}$), with $d = 5$ and $n_1 = 50$. The covariance matrix is $\mathbf{V}_1 = \mathbf{I}_5 + \delta\mathbf{J}_5$ and the error terms are based on skew normal distribution (left side) or gamma distribution (right side).

Illustrative Data Analysis

After using the proposed method in a simulation study, we apply it to a real data set. To this aim, we take a closer look at the EEG data set from the R-package *manova.rm* by Friedrich et al. [2019], which was already mentioned in the first section. In this study from Staffen et al. [2014], conducted at the University Clinic of Salzburg (Department of Neurology), electroencephalography (EEG) data from 160 patients with different diagnoses of impairments were measured. These are Alzheimer's disease (AD), mild cognitive impairment (MCI), and subjective cognitive complaints (SCC). Thereby, this last diagnosis is differentiated between subjective cognitive complaints with minimal cognitive dysfunction (SCC+) and without (SCC-).

The number of patients divided by sex and diagnosis can be found in Table 3.14. Since in Bathke et al. [2018] and Sattler [2021] there was no distinction between SCC+ and SCC-, we consider both together as diagnosis SCC.

Table 3.14: Number of observations for the different factor level combinations of sex and diagnosis.

	AD	MCI	SCC+	SCC-
male	12	27	14	6
female	24	30	31	16

With two kinds of measurements (z-score for brain rate and Hjorth complexity) and three different electrode positions (frontal, temporal and central), the observation vector's dimension is $d = 6$ and therefore $p_u = 15$. In relation to this dimension, all sample sizes are rather small, which we should keep in mind for the evaluation of our results.
The considered hypotheses, are:

a) Homogeneity of correlation matrices between different diagnoses,

b) Homogeneity of correlation matrices between different sexes,

while we will denote the corresponding hypothesis regarding the covariance matrix with $\mathcal{H}_0^v$.
In Sattler [2021] homogeneity of covariance matrices between different diagnoses as well as different sexes were investigated. Here, we consider the more general hypothesis of equal correlation matrices between the diagnoses and the sexes. Thereby it is of interest to compare the results from homogeneity of covariance matrices with them from testing homogeneity of correlation matrices. We expect higher p-values for equality of correlation through the larger hypothesis, but each rejection of equal correlation matrices directly allows us to reject the corresponding equality of covariance matrices. In Table 3.15 for both hypotheses, the p-values for the ATS with parametric bootstrap are displayed, while for both bootstrap tests, 10,000 bootstrap runs are done.

It is interesting that for two hypotheses, the p-value of equal correlation matrices are rejected at level 5%, while we could not reject the smaller hypothesis of equal covariance matrices. But for both hypotheses, the sample sizes are rather small with $N < 40$. Our simulation results for $d = 5$ showed that the ATS with

		ATS-Para for $\mathcal{H}_0^v$ p-value	ATS-Para for $\mathcal{H}_0^r$ p-value
male	AD vs. MCI	.1000	**.0389**
male	AD vs. SCC	**.0452**	**.0073**
male	MCI vs. SCC	**.0289**	.0753
female	AD vs. MCI	.0613	.3601
female	AD vs. SCC	**.0128**	.4882
female	MCI vs. SCC	.5656	.8799
AD	male vs. female	.1008	**.0346**
MCI	male vs. female	.2455	.6703
SCC	male vs. female	.2066	.1748

Table 3.15: P-values of ATS with parametric bootstrap for testing equality of correlation matrices and equality of covariance matrices.

parametric bootstrap is too liberal for small sample sizes, which might be the reason why the larger hypotheses can be rejected, and the smaller ones can not. Moreover, it can be seen that the difference between some hypotheses is relatively small, like for the first three hypotheses, but it can also be quite large as for the comparison of women with AD and with SCC. This shows that from a rejection of $\mathcal{H}_0^v$ no conclusion on $\mathcal{H}_0^r$ can be drawn.
Through the small sample size in relation to the dimension of the vectorized correlation matrix $p_u = 15$, the rejections are not as reliable as for the covariance matrix. Still, they are helpful in selecting hypotheses for further analysis.

Conclusion

In the previous section, a series of new test statistics was developed. They could be used for many different kinds of null hypotheses with hardly any restrictions and could be expanded easily, for example, by using a Fisher z-transformation. In our simulation study, it could be seen that our tests based on the ATS are appropriate for many different hypotheses. This holds for critical values based on a parametric bootstrap as well as on a Monte-Carlo simulation. The developed tests outperform existing procedures for some hypotheses, while they offer good and interesting alternatives for other ones. One more time, these results show the flexibility of the approach introduced in Sattler et al. [2022] and the multitude of possible applications.

3.5.4 Testing for covariance patterns

In Sattler et al. [2022], some hypotheses, together with suitable hypothesis matrices, were introduced. But because of the very general model, there are various other possible hypotheses. For example, these include a pattern in the covariance matrix or a particular type of covariance matrix. We will focus on the latter while applying our approach for these hypotheses gives a good impression of how it can be used for other patterns. This topic is of great interest, and therefore there are already several more or less detailed approaches. For instance, for high-dimensional data, Zhong et al. [2017] allows general hypotheses under normality while for non-high-dimensional data, Gupta and Xu [2006] checks for sphericity, but with some conditions on the characteristic function. Finally, Wakaki et al. [1990] allows for testing all of the existing structures with fewer distributional restrictions. Unfortunately, their procedure is quite complex since many parameters need to be calculated or estimated. Together with the fact that the concrete test is never mentioned, this approach is challenging to use in practice. This is also mentioned in Yuan [2005] or Herzog et al. [2007]. Thus, our new test based on Sattler et al. [2022] should allow tests for a wide range of possible structures, with less distributional conditions and comparatively clear and intuitive usage.

This section will shortly introduce the most common covariance structures, together with an appropriate hypothesis matrix for the particular case. For selecting the covariance structures, we follows Kincaid [2005], which gives a good overview of the most important models without being too detailed.

It is quite clear that the correlation matrix structure could be of interest besides the covariance matrix structure. We do not treat this because of the direct connection between hypotheses regarding the correlation matrix and regarding the covariance matrix from Theorem 3.5.1. Thereby it can be easily adapted from the following.

Many authors mention the hypothesis regarding structures or patterns of covariance/correlation matrix without explaining their test's concrete usage for this hypothesis more precisely, for example, by introducing the necessitated hypothesis matrices. So, e.g., Steiger [1980] simulated for the hypothesis whether a correlation matrix has the structure of a Toeplitz matrix. But neither the representation of this hypothesis in the underlying model nor the hypothesis matrix was mentioned. In most cases, this increases the readability and avoids the need for extensive definitions. But unfortunately, at the same time, this complicates the application of the procedure by users. Therefore, the following part aims to provide the necessary hypothesis matrices for the most common covariance structures. To this end, we define some general matrices that are part of different null hypotheses and limit ourselves to auxiliary vectors introduced in Section 3.5.3. On the one hand, we use the vector $\boldsymbol{a}$ containing the indices of components in the half-vectorized matrix which belong to diagonal elements of the original matrix. These indices are given through $a_k = 1 + \sum_{j=1}^{k-1}(d+1-j)$, $k = 1, ..., d$, so the vector is defined as $\boldsymbol{a} = (a_1, ..., a_d)$. On the other hand, the p_u-dimensional vector $\boldsymbol{b}$ containing all other indices in ascending order, which belong to non-diagonal elements of the original matrix.

All of the following hypothesis matrices are formulated most comprehensively and intuitively, without using projection matrices. If $\boldsymbol{\zeta} = \mathbf{0}$ holds (which will be true except for the autoregressive structure), a unique projection matrix exists. Based on the results from Section 3.5.2 we recommend for these matrices to remove zero rows and to use the so formed matrices.

Sphericity

The sphericity of a covariance matrix denotes that the covariance matrix is the product of a scalar and the identity matrix. This is a necessary assumption in many repeated measurement approaches, like the ANOVA. It can be seen as a special case of a compound symmetry matrix with the additional requirement that the non-diagonal elements are equal and have the value 0. One way to formulate this hypothesis would be to expand the hypothesis matrix of the compound symmetry structure. But it is more consistent to treat this structure as the combination of equal diagonal elements and zeros elsewhere which leads to $\mathbf{C}_S = (\mathbf{C}_{S1}^\top, \mathbf{C}_{S2}^\top)^\top$ with

$$\mathbf{C}_{S1} = \sum_{k=1}^{d-1} \boldsymbol{e}_{k,d-1} \cdot (\boldsymbol{e}_{a_k,p_u+1}^\top - \boldsymbol{e}_{a_{k+1},p}^\top) \quad \text{and } \mathbf{C}_{S2} = \sum_{\ell=1}^{p_u} \boldsymbol{e}_{\ell,p_u} \cdot \boldsymbol{e}_{b_\ell,p}^\top.$$

Thereby with $\mathcal{H}_0^{\boldsymbol{v}}(S) : \mathbf{C}_S \boldsymbol{v} = \mathbf{0}_{p-1}$ we express the hypothesis of $\mathcal{H}_0$: {sphericity of the covariance matrix $\boldsymbol{V}$}.

Diagonality

As well as the sphericity of a matrix, the more general hypothesis of a diagonal covariance matrix is of great interest. It shows that all components are uncorrelated, which allows many conclusions on the underlying model. The hypothesis of diagonality, also known as variance components, was considered in Section 3.5.3 as $B_{\mathbf{r}}$) and was given through $\mathcal{H}_0^r(D) : \mathbf{r} = \mathbf{0}_{p_u}$.

Compound Symmetry

This widespread covariance matrix structure, which is especially known from split-plot-designs, is characterized by two conditions: the equality of all diagonal elements and all non-diagonal elements' equality. As a consequence, the appropriate hypothesis matrix is also composed of two parts through

$$\mathbf{C}_{CS} = \begin{pmatrix} \mathbf{C}_{S1} \\ \mathbf{C}_{CS1} \end{pmatrix} \quad \text{with } \mathbf{C}_{CS1} = \sum_{\ell=1}^{p_u-1} \boldsymbol{e}_{\ell,p_u-1} \cdot (\boldsymbol{e}_{\mathfrak{b}_{\ell},p}^{\top} - \boldsymbol{e}_{\mathfrak{b}_{\ell+1},p}^{\top}).$$

With this matrix we can formulate $\mathcal{H}_0$: {$\mathbf{V}$ is a compound symmetry matrix} through $\mathcal{H}_0^v(CS) : \mathbf{C}_{CS}\boldsymbol{v} = \mathbf{0}_{p-2}$.
A variation of this structure is the so-called Heterogenous Compound Symmetry. Hereby, the classical compound symmetry matrix is multiplied from both sides with a diagonal matrix $diag(\sigma_1, ..., \sigma_d)$ with all components being positive real numbers. Fortunately, the hypothesis $\mathcal{H}_0^v$: {$\mathbf{V}$ is a heterogenous compound symmetry matrix} can be formulated with the correlation matrix as $\mathcal{H}_0^r(HCS) : \mathbf{P}_{p_u}\mathbf{r} = \mathbf{0}_{p_u}$.

Toeplitz

This matrix is defined through the fact that the diagonal elements are equal as well as all the secondary diagonals, which is why it is also called a diagonal-constant matrix. In contrast to the compound symmetry matrix, not all non-diagonal elements need to have the same value, just within the individual secondary diagonals, they have to be equal. For the covariance matrix construction, we use the vector $\mathbf{h}_1$ from Section 3.5.3. With this and

$$\mathbf{C}_{T1} = \sum_{k=1}^{d-2} \sum_{\ell=1}^{d-k-1} \boldsymbol{e}_{h_{1,k}-d-k+\ell,p_u-d+1} \cdot (\boldsymbol{e}_{\mathfrak{a}_{k+1}+\ell,p}^{\top} - \boldsymbol{e}_{\mathfrak{b}_{\ell},p}^{\top}),$$

we get $\mathbf{C}_T = (\mathbf{C}_{S1}^\top, \mathbf{C}_{T1}^\top)^\top$. In this case, the hypothesis of $\mathcal{H}_0^v$: {$\mathbf{V}$ is a Toeplitz matrix} can be formulated as $\mathcal{H}_0^v(T) : \mathbf{C}_T\boldsymbol{v} = \mathbf{0}_{p_u}$. Again, there exists a heterogenous Toeplitz matrix which can be tested through

$$\mathbf{C}_{HT} = \sum_{k=1}^{d-2}\sum_{\ell=1}^{d-k-1} \boldsymbol{e}_{h_{1,k}-d-k+\ell,p_u-d+1} \cdot (\boldsymbol{e}^\top_{\boldsymbol{a}_{k+1}+\ell-k-1,p} - \boldsymbol{e}^\top_{\boldsymbol{b}_\ell-1,p})$$

and $\mathcal{H}_0^r(HT) : \mathbf{C}_{HT}\mathbf{r} = \mathbf{0}_{p_u-d+1}$.

Autoregressive

This last structure can be seen as a special case of a Toeplitz matrix. However, there is a kind of proportionality between the different secondary diagonals for an autoregressive covariance matrix. The whole matrix depends just on one parameter $\sigma \in (0,1)$ and is given through $(\mathbf{V})_{ij} = \sigma^{|i-j|}$, which shows that the correlation of the components decreases exponentially with the distance. This is useful, for example, if consecutive components belong to neighboring measure points. Moreover, this structure is often used for repeated measurements at different time points, since it is reasonable that measurements that are further apart in time have a smaller correlation. While the equal secondary diagonals could be tested similar to the case of a Toeplitz matrix, the proportionality, as mentioned above, is much more complicated. Each kind of proportionality can just be considered if all components are different from zero. Therefore, this test can not be used if one component of the empirical covariance matrix has zero value. If a component is zero, the null hypothesis of an autoregressive covariance matrix should be received.
First, we define the function $f : \mathbb{R}_{\neq 0}{}^p \to \mathbb{R}^p$, $(x_1, ..., x_p) \mapsto (\ln(|x_1|), ..., \ln(|x_p|))$, where $\mathbb{R}_{\neq 0}{}^p$ denotes the real values p-dimensional vectors with only positive components. With the δ-method it holds

$$\sqrt{N}\,(f(\boldsymbol{v}) - f(\widehat{\boldsymbol{v}})) = \sqrt{N}(\ln(|\boldsymbol{v}|) - \ln(|\widehat{\boldsymbol{v}}|)) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}_p, \mathit{diag}(\boldsymbol{v})^{-1}\boldsymbol{\Sigma}\,\mathit{diag}(\boldsymbol{v})^{-1}).$$

The absolute values ensure that the logarithm can be calculated, even if the estimated covariance vector has negative components. Moreover, we define

$$\mathbf{C}_{AR1} = \sum_{k=1}^{d-1}\sum_{\ell=1}^{d-k} \boldsymbol{e}_{a_k-k+\ell,p} \cdot (\boldsymbol{e}_{a_k+\ell-1,p} - \cdot\boldsymbol{e}_{a_k+\ell,p})^\top + \sum_{k=1}^{d} \boldsymbol{e}_{p_u+k,p} \cdot \boldsymbol{e}^\top_{a_k,p}$$

as the first part of our hypothesis matrix. The first sum builds differences between components, corresponding to elements of the covariance matrix, which are neighbouring and from the same row. The second sum picks the components corresponding to diagonal elements of the original covariance matrix, to verify that $\mathbf{V}_{11} = \mathbf{V}_{22} = ... = \mathbf{V}_{aa} = 1$.

Under the null hypothesis of an autoregressive covariance matrix with unknown parameter σ, it holds $\mathbf{C}_{AR1}f(\boldsymbol{\nu}) = (\ln(\sigma)\mathbf{I}_{p_u}, \mathbf{0}_d^\top)^\top$. To check whether the first p_u components of this vector are equal, we use $\mathbf{C}_{AR2} = \mathbf{P}_{p_u} \oplus \mathbf{I}_d$ and define $\mathbf{C}_{AR} = \mathbf{C}_{AR2}\mathbf{C}_{AR1}$. With this matrix, we formulate our null hypothesis through $\mathcal{H}_0^{\nu}(AR) : \mathbf{C}_{AR}f(\boldsymbol{\nu}) = \mathbf{0}_p$. With $\mathbf{C}_{AR}\nabla f(\widehat{\boldsymbol{\nu}})\widehat{\boldsymbol{\Sigma}}\nabla f(\widehat{\boldsymbol{\nu}})^\top\mathbf{C}_{AR}^\top$ as a consistent estimator for the unknown covariance matrix and the asymptotic distribution, it is possible to define the usual quadratic forms. The Monte-Carlo based ATS can be directly formulated while the parametric bootstrap can be adapted easily.

For the sake of completeness, we want to present another function that also could be used to check this null hypothesis. To this end, we define the continuous function

$$g : \mathbb{R}_{\neq 0}{}^p \to \mathbb{R}^p,\ \boldsymbol{x} \mapsto \left(\frac{x_{a_1+1}}{x_{a_1}}, ..., \frac{x_{a_1+d-1}}{x_{a_1+d-2}}, \frac{x_{a_2+1}}{x_{a_2}}, ..., \frac{x_{a_2+d-2}}{x_{a_2+d-3}}, ..., \frac{x_{a_{d-1}}}{x_{a_{d-1}+1}}, x_{a_1}, ..., x_{a_d}\right).$$

For the first p_u components the relation between the components and following components is calculated, except this corresponds to a relation between two components, which are in different rows in the original matrix. The last d components are the identity for the diagonal elements of the original matrix. Again, because of the δ-method it holds

$$\sqrt{N}\left(g(\boldsymbol{\nu}) - g(\widehat{\boldsymbol{\nu}})\right) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}_p, \nabla g(\boldsymbol{\nu})\boldsymbol{\Sigma}\nabla g(\boldsymbol{\nu})^\top),$$

with

$$\begin{aligned}\nabla g(\boldsymbol{x}) = \ & \sum_{k=1}^{d-1}\sum_{\ell=1}^{d-k} \boldsymbol{e}_{a_k-k+\ell,p} \cdot \left(\frac{1}{x_{a_k+\ell-1}} \cdot \boldsymbol{e}_{a_k+\ell,p} - \frac{x_{a_k+\ell}}{x^2_{a_k+\ell-1}} \cdot \boldsymbol{e}_{a_k+\ell-1,p}\right)^\top \\ & + \sum_{k=1}^{d} \boldsymbol{e}_{p_u+k,p} \cdot \boldsymbol{e}_{a_k,p}^\top.\end{aligned}$$

It remains to test whether the first p_u components of $g(\boldsymbol{\nu})$ are equal and the last d components have the value one. Because of this we build the hypothesis matrix through $\widetilde{\mathbf{C}}_{AR} = \mathbf{P}_{p_u} \oplus \mathbf{I}_d$. Hence, under the null hypothesis of an autoregressive covariance matrix, it holds

$$\sqrt{N}\left(\widetilde{\mathbf{C}}_{AR}\, g(\widehat{\boldsymbol{\nu}}) - \begin{pmatrix}\mathbf{0}_{p_u}\\ \mathbf{1}_d\end{pmatrix}\right) \xrightarrow{\mathcal{D}} \mathcal{N}_p(\mathbf{0}_p, \widetilde{\mathbf{C}}_{AR}\nabla g(\boldsymbol{\nu})\boldsymbol{\Sigma}\nabla g(\boldsymbol{\nu})^\top \widetilde{\mathbf{C}}_{AR}^\top),$$

whereby $\widetilde{\mathbf{C}}_{AR}\nabla g(\widehat{\boldsymbol{\nu}})\widehat{\boldsymbol{\Sigma}}\nabla g(\widehat{\boldsymbol{\nu}})^\top \widetilde{\mathbf{C}}_{AR}^\top$ is a consistent estimator for the unknown covariance matrix. Based on this, once more quadratic forms like the ATS can be defined to examine the hypothesis $\mathcal{H}_0^{\boldsymbol{\nu}}(AR) : \widetilde{\mathbf{C}}_{AR}\boldsymbol{\nu} = (\mathbf{0}_{p_u}^\top, \mathbf{1}_d^\top)^\top$.
Also, for this structure, a heterogeneous version exists, which can be tested with similar matrices and the vectorized correlation matrix. Then, the second part of the matrices $\mathbf{C}_{AR1}$ and $\mathbf{C}_{AR2}$ is unnecessary. The first part can be reduced because the correlation matrix's diagonal does not have to be compared regarding the proportionality. Similarly, the function g, as well as the corresponding hypothesis matrix, could be simplified.
Finally, there exists one more version of the autoregressive structure, called first order autoregressive. Here, the elements are given through $(\mathbf{V})_{ij} = \rho \cdot \sigma^{|i-j|}$, with $\rho, \sigma > 0$. This is a more general case, since the diagonal elements have to be equal but are allowed to have other values than one. To check whether the covariance matrix has this structure, we replace $\mathbf{C}_{AR2}$ with $\mathbf{P}_{p_u} \oplus \mathbf{P}_d$. Under the null hypothesis of a first order autoregressive structure, it holds $\mathbf{C}_{AR1} f(\boldsymbol{\nu}) = (\ln(\sigma)\mathbf{I}_{p_u}, \ln(\rho)\mathbf{I}_d)^\top$ and therefore $\mathbf{C}_{AR2}\mathbf{C}_{AR1} f(\boldsymbol{\nu}) = \mathbf{0}_p$.

Simulations

Finally, we want to investigate these tests' performance through their type-I-error rate. Diagonality was already tested in Section 3.5.3, so we test for an autoregressive structure and a Toeplitz matrix, based on the above-introduced matrices. For the Toeplitz matrix, we use the hypothesis matrix, which is formed by removing zero rows from the existing unique projection matrix. Here, we have one structure based on just one parameter and one based on five parameters, which is interesting as a comparison. In Herzog et al. [2007], they assumed a relation between the number of parameters and the required sample size for a sufficient approximation of the asymptotic distribution.
For the autoregressive structure, we chose the parameter $\sigma = 0.6$, and for the Toeplitz matrix, we use $(\mathbf{V})_{ij} = 1 - |i-j|/d$. We simulate for dimension five and error terms based on

- a standard normal distribution, i.e. $Z_{ikj} \sim \mathcal{N}(0,1)$,
- a standardized centered gamma distribution i.e. $(\sqrt{2}Z_{ikj} + 2) \sim \mathcal{G}(2,1)$,
- a standardized centered skew normal distribution with location parameter $\xi = 0$, scale parameter $\omega = 1$ and $\alpha = 4$. The density of a skew normal distribution is given through $\frac{2}{\omega}\varphi\left(\frac{x-\xi}{\omega}\right)\Phi\left(\alpha\left(\frac{x-\xi}{\omega}\right)\right)$, where φ denotes the densitiy of a standard normal distribution and Φ the according distribution function,
- a standardized centered t-distribution with 9 degrees of freedom,

while the sample size are $\mathbf{n} = (25, 50, 100, 250)$. On the basis of the previous results, we only consider the ATS with parametric bootstrap and with Monte-Carlo-based critical values. Here we use 1,000 bootstrap runs and 10,000 Monte-Carlo steps. The type-I-error rates, based on 20,000 simulation runs, can be seen in Table 3.16 and Table 3.17 for $\alpha = 5\%$. For the autoregressive structure both approaches, based on the function f and the function g, are used. Finally, for the Toeplitz structure, we use the hypothesis matrix, which is formed by removing zero rows from the unique projection matrix.

For testing whether the covariance matrix is a Toeplitz matrix, the parametric bootstrap has a better small sample performance than the Monte-Carlo approach. For larger sample sizes, their type-I-error rates approach. Both tests fulfill Bradley's liberal criterion in all cases and have small error rates, especially for the t_9 distribution.

For the considerably more challenging hypothesis of an autoregressive structure, both tests are quite conservative. The liberal criterion is only fulfilled for $n_1 = 250$ for most of the considered distributions. For the normal distribution, even higher sample sizes are needed. This result is not surprising because the proportionality, which is the only difference to the Toeplitz matrix, is difficult to check between all secondary diagonals. In contrast to all other hypotheses, here, the Monte-Carlo approach has a better small sample performance than the parametric bootstrap ATS. This holds for both approaches based on the function f and g.

While for $n_1 < 250$ the function f seems more recommendable overall, for $n_1 = 250$ the function g has a better type-I-error control in most cases. These

	ATS-Para				ATS			
N	25	50	100	250	25	50	100	250
t_9	**.0493**	**.0481**	**.0472**	**.0505**	.0569	**.0513**	**.0474**	**.0508**
Normal	.0556	**.0525**	.0544	**.0524**	.0635	.0555	.0551	**.0521**
Skew normal	.0557	.0531	**.0514**	**.0491**	.0639	.0559	**.0515**	**.0496**
Gamma	.0458	.0443	.0435	**.0485**	**.0522**	.0466	.0444	**.0492**

Table 3.16: Simulated type-I-error rates ($\alpha = 5\%$) for testing whether the covariance matrix has a Toeplitz structure, with ATS based on parametric bootstrap and based on Monte-Carlo simulation. The observation vectors have dimension 5, covariance matrix $(\mathbf{V})_{ij} = 1 - |i - j|/5$ and different distributions and sample sizes are considered.

	ATS-Para-f				ATS-f			
N	25	50	100	250	25	50	100	250
t_9	.0127	.0133	.0197	.0252	.0137	.0138	.0198	.0252
Normal	.0073	.0099	.0161	.0222	.0083	.0105	.0163	.0221
Skew normal	.0109	.0122	.0187	.0260	.0121	.0131	.0188	.0256
Gamma	.0254	.0261	.0298	.0328	.0285	.0264	.0304	.0328
	ATS-Para-g				ATS-g			
N	25	50	100	250	25	50	100	250
t_9	.0094	.0111	.0155	.0278	.0106	.0121	.0162	.0270
Normal	.0093	.0113	.0197	.0260	.0105	.0122	.0198	.0254
Skew normal	.0085	.0116	.0185	.0286	.0102	.0125	.0184	.0289
Gamma	.0107	.0126	.0199	.0280	.0124	.0131	.0202	.0275

Table 3.17: Simulated type-I-error rates ($\alpha = 5\%$) for testing whether the covariance matrix has an autoregressive structure, with ATS based on parametric bootstrap and based on Monte-Carlo simulation. The observation vectors have dimension 5, covariance matrix $(\mathbf{V})_{ij} = 0.6^{|i-j|}$ and different distributions and sample sizes are considered.

values, together with the easier application, make the f function a more reasonable choice. Simultaneously, these are just some proposals since there are many other possible functions and matrices to check this hypothesis.
Since our tests are quite conservative for this hypothesis, it could be useful to apply the wild bootstrap approach, which had a more liberal behavior than the parametric bootstrap in Sattler et al. [2022], and therefore could balance the performance.

This section shows that with our approach, hypotheses regarding the structure or a pattern of the covariance matrix can be checked with a suitable hypothesis matrix. For some more complex structures, some adaption has to be done, and a large sample size is recommended to get reliable results.

Illustrative Data Analysis

After introducing these tests for covariance matrix patterns, we want to illustrate their application and the resulting conclusions. To this aim, we reconsider the EEG-data set from Friedrich et al. [2019], but this time focus on another aspect.
The z-score of the brain rate and the Hjorth complexity are measured at three different locations, namely temporal, frontal and central. Here we want to investigate whether the position of the measuring points influences the measured values. Similar questions are often considered in repeated measure designs, where the repetitions have a temporal context, to investigate wheater there is a time effect. One way to check such an impact is to compare the means of the three locations and use thereto, for example, a one-sample Hotelling's T^2 test Anderson [2003]. But this is not the only way how the position of the measurement points can influence the measurements. It could also have an effect on the variance of the individual measure points, as well as on the dependency structure between them.
Therefore, we also want to consider the covariance matrix and investigate which conclusions can be drawn out of it. This is done by testing whether the covariance matrix has a compound symmetry structure. A rejection of this structure means that the variances are different or the correlation between the locations is different. Therefore, the locations are not exchangeable, which shows an influence of the measuring point's position. For the sake of completeness, we are also testing whether the covariance matrix has a Toeplitz structure. In contrast

to a compound symmetry matrix, this would mean that there are systematical differences in the correlations. This would make sense to represent the distance between the measurement points, which is not useful in our setting since all locations are neighboring.
Through the low dimension $d = 3$ and therefore $p = 6$, we expect reliable results of our test, even for the sample size of only 12 observations.

The one sample Hotelling's T^2 test, based on an χ^2 distribution, is for example part of the R-package *ICSNP*(Nordhausen et al. [2018]). To apply this test, we use the matrix $\mathbf{C}_{HT2} \in \mathbb{R}^{2\times 3}$, given through

$$\mathbf{C}_{HT2} = 1/2 \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}.$$

For both structures, we use the ATS with parametric bootstrap based on 10,000 bootstrap runs and calculate the p-values. The results, together with the results of the one-sample Hotelling's T^2 test, are displayed in Table 3.18 and Table 3.19.

Brain rate		Hotelling's T^2 p-value	ATS-Para for $\mathcal{H}_0^{\mathcal{V}}(CS)$ p-value	ATS-Para for $\mathcal{H}_0^{\mathcal{V}}(T)$ p-value
male	AD	0.9881	0.4056	0.4883
male	MCI	0.7472	0.4869	0.5882
male	SCC	**0.0162**	0.2380	0.2395
female	AD	0.6483	0.5845	0.5553
female	MCI	0.9261	0.8014	0.7572
female	SCC	0.9391	0.6938	0.6634

Table 3.18: P-values of one sample Hotelling's T^2 test and φ^*_{ATS} to check wheater the covariance matrix has a compound symmetry structure resp. a Toeplitz structure.

For the brain rate in the group of men with SCC, a difference in the mean can be verified at level 5%, while for the covariance, no structure can be rejected. In contrast, for the Hjorth complexity, for women with SCC, the location's influence can be proven for the mean and the covariance since both structures are rejected. The influence on the covariance matrix here is clearer than on the mean and could also be verified for level 1%.

Hjorth complexity		Hotelling's T^2 p-value	ATS-Para for $\mathcal{H}_0^v(CS)$ p-value	ATS-Para for $\mathcal{H}_0^v(T)$ p-value
male	AD	0.4372	0.4029	0.4110
male	MCI	0.1276	0.1113	0.1142
male	SCC	0.1273	0.1412	0.1453
female	AD	0.3139	0.3809	0.3491
female	MCI	0.9328	0.1172	0.1207
female	SCC	**0.0213**	**0.0073**	**0.0079**

Table 3.19: P-values of one sample Hotelling's T^2 test and φ^*_{ATS} to check wheater the covariance matrix has a compound symmetry structure resp. a Toeplitz structure.

To investigate this group in more detail, further hypotheses could be tested, as equal variances of all components. One could thereto apply the test from Sattler [2021]. Rejection of this or similar larger null hypotheses would allow us better to understand the location's influence on the covariance matrix. Since the compound symmetry structure is a special case of the Toeplitz structure, it has lower p-values in most groups. But for these 3×3 covariance matrices, the difference in structure is relatively small, and therefore this relation does not hold for all groups.
All in all, the noticeable difference in p-values between mean and covariance matrix structure shows, that for the verification of an effect, both aspects should be considered, the mean and the covariance matrix. This example illustrates, how both kinds of hypotheses can be used to investigate two aspects of the same question.

Conclusion

It could be seen that the hypothesis matrices for testing the presented structures differ in complexity and the matrix rank. Interestingly, this has no relation to the number of parameters from which the covariance matrix is built.

Some less common patterns are treated in other works, e.g. Steiger [1980], where a so-called circumplex hypothesis is checked, which tests for a Toeplitz matrix with diagonal elements one and a so-called equicorrelation hypotheses, which means that all non-diagonal elements are equal. The example shows that most of the not considered patterns are related to the treated structures and, therefore, can be investigated similarly. Moreover, while we focused on covariance matrices' structure, a correlation matrix structure could also be of great interest. However, through the heterogeneous version of our hypotheses, these hypotheses' adaption for correlations was already done.

We are convinced that nearly every kind of covariance matrix or correlation matrix pattern can be tested based on the approach published in Sattler et al. [2022] resp. Section 3.5.3. Partially the usage is a bit more complicated or needs techniques like the δ-method, comparable to the autoregressive structure.

3.5.5 Efficient implementation of the developed tests

Computation time is an essential factor for choosing the adequate test statistic for a hypothesis in the individual situation. This is particularly relevant for hypotheses regarding the covariance matrix, since the dimension of the vectorized covariance matrix grows quadratically in d.
As seen in Section 6 of Sattler et al. [2022], in many cases the choice of a non-quadratic hypothesis matrix could save a substantial amount of time. The influence of this change on the test result was investigated in Section 3.5.2. This passage showed which kind of non-quadratic matrix does not change the test result. The saved time through such non-quadratic matrices varied widely. In detail, the percentage of saved time for hypothesis $D)(\mathcal{H}_0^\gamma : \operatorname{tr}(\mathbf{V}_1) = \gamma)$ is significantly greater than for $C)(\mathcal{H}_0^\gamma : \operatorname{tr}(\mathbf{V}_1) = \operatorname{tr}(\mathbf{V}_2))$.
One reason for this is a trick that is usable for one group. Our goal is to generalize it for more groups if the hypothesis matrix has a special structure. We also found other ways to make the calculation more efficient for non-quadratic matrices without changing the results' validity.

Consider hypotheses of the kind $\mathbf{C} = \mathbf{C}_W \otimes \mathbf{C}_S$ with $\mathbf{C}_W \in \mathbb{R}^{m_1 \times a}$ and $\mathbf{C}_S \in \mathbb{R}^{m_2 \times p}$, while $m_1 \in \mathbb{N}_a$ and $m_2 \in \mathbb{N}_p$. This kind of hypothesis is very popular, for example, in the context of repeated measurement. All hypotheses simulated in Sattler et al. [2022] can be written in this way. Moreover, each hypothesis for just one group is part of this model.
In many settings estimating the variance of $\mathbf{C}_s \operatorname{vech}((\mathbf{X}_{i1} - \boldsymbol{\mu}_i)(\mathbf{X}_{i1} - \boldsymbol{\mu}_i)^\top)\mathbf{C}_S^\top$ is more efficient than estimating $\boldsymbol{\Sigma}_i$ and afterwards calculate $\mathbf{C}_S \widehat{\boldsymbol{\Sigma}}_i \mathbf{C}_S^\top$, especially for $m_2 < p$. More details on when this approach is useful can be found in the

appendix of this section.

Since it holds

$$\mathbf{C}\widehat{\boldsymbol{\Sigma}}\mathbf{C}^\top = (\mathbf{C}_W \otimes \mathbf{I}_{m_2}) \left(\bigoplus_{i=1}^{a} \mathbf{C}_S \widehat{\boldsymbol{\Sigma}}_i \mathbf{C}_S^\top \right) (\mathbf{C}_W \otimes \mathbf{I}_{m_2})^\top,$$

with an efficient calculation of $\mathbf{C}_S\widehat{\boldsymbol{\Sigma}}_i\mathbf{C}_S$, the computation of the matrix can be simplified.

In a similar way the generation of the bootstrap samples can be transformed for the parametric bootstrap approach as well as the wild bootstrap. In Sattler et al. [2022] we generated $\mathbf{Y}^*_{1n_i}, ..., \mathbf{Y}^*_{in_i} \sim \mathcal{N}_p(\mathbf{0}_p, \widehat{\boldsymbol{\Sigma}}_i)$ for the parametric bootstrap and used them to calculate quadratic forms in $\mathbf{C}\overline{\mathbf{Y}}^*$ with $\overline{\mathbf{Y}}^* = (\overline{\mathbf{Y}}_1^{*\top}, ..., \overline{\mathbf{Y}}_a^{*\top})^\top$. Instead of these bootstrap variables, $\mathbf{Z}^*_{1n_i}, ..., \mathbf{Z}^*_{in_i} \sim \mathcal{N}_{m_2}(\mathbf{0}_{m_2}, \mathbf{C}_S\widehat{\boldsymbol{\Sigma}}_i\mathbf{C}_S^\top)$ can be generated and used to calculate quadratic forms in $(\mathbf{C}_W \otimes \mathbf{I}_{m_2})\overline{\mathbf{Z}}^*$. This does not only save time through generation of smaller random variables, but above all through use of the empirical covariance of $\mathbf{Z}^*_{i1}, ..., \mathbf{Z}^*_{in_i}$ which is denoted as $\widehat{\boldsymbol{\Sigma}}_i^*(\mathbf{Z})$. Hence, in ATS, WTS, and MATS we can replace $\mathbf{C}\widehat{\boldsymbol{\Sigma}}^*\mathbf{C}^\top$ by $(\mathbf{C}_W \otimes \mathbf{I}_{m_2}) \left(\bigoplus_{i=1}^{a} \widehat{\boldsymbol{\Sigma}}_i^*(\mathbf{Z}) \right) (\mathbf{C}_W \otimes \mathbf{I}_{m_2})^\top$. Hereby for each m_1 and m_2, the number of necessary multiplications can be reduced considerably, which is also shown more detailed in the appendix.
Since the bootstrap part is repeated B times, the time reduction in this part greatly influences the total calculation time. Although the distribution is the same, this other kind of bootstrap variables can influence the test's behavior. This was checked, and the differences were negligible.

For the wild bootstrap approach this can be done in a similar manner, by using $\mathbf{Z}^\star_{ik} = W_{ik} \cdot \mathbf{C}_S \left[\operatorname{vech}(\widetilde{\mathbf{X}}_{ii}\widetilde{\mathbf{X}}_{ik}^\top) - n_i^{-1} \sum_{\ell=1}^{n_i} \operatorname{vech}(\widetilde{\mathbf{X}}_{ii}\widetilde{\mathbf{X}}_{ik}^\top) \right]$ and the empirical covariance matrices based on $\mathbf{Z}^\star_{i1}, ..., \mathbf{Z}^\star_{in_i}$.
Since these alternative bootstrap approaches do not change the distribution of the vectors used for the quadratic forms, all results from Sattler et al. [2022] remain valid.

To illustrate the influence of these and other small changes, we repeat the time computation from Sattler et al. [2022] for all hypotheses and quadratic hypothesis matrices as well as non-quadratic, implemented in the previously explained way. The results of hypothesis C)$(\mathcal{H}_0^\gamma : \operatorname{tr}(\mathbf{V}_1) = \operatorname{tr}(\mathbf{V}_2))$ can be seen in Table 3.20

while for other hypotheses it can be found in the appendix of this section.

	$\mathbf{C}(C)$				$\widetilde{\mathbf{C}}(C)$			
d	2	5	10	20	2	5	10	20
ATS-Para	0.476	0.862	17.620	141.163	0.409	0.408	0.411	0.449
ATS-Wild	0.415	0.536	10.370	111.671	0.394	0.393	0.396	0.435
ATS	0.085	0.195	0.388	1.177	0.032	0.034	0.034	0.042
WTS-Para	0.585	1.074	30.496	336.538	0.492	0.491	0.496	0.519
WTS-Wild	0.531	0.709	21.997	300.008	0.487	0.489	0.488	0.529
WTS-χ^2	0.002	0.003	0.040	0.343	0.002	0.002	0.003	0.028

Table 3.20: Required time in seconds for various tests statistics and different dimensions for hypothesis C) ($\mathcal{H}_0^v$: $\operatorname{tr}(\mathbf{V}_1) = \operatorname{tr}(\mathbf{V}_2)$) with a quadratic hypothesis matrix on the left side and a non-quadratic hypothesis matrix on the right sight. Here some methods are used to increase efficiency.

It can be seen for the projection matrix that the alternative approach is faster than the earlier one, especially for the parametric bootstrap approaches. Here, sometimes 40% or more can be saved, while for the WTS with a wild bootstrap, nearly no time is saved. However, in comparison to the non-quadratic matrix, the required computation time is enormous. Simultaneously, the result is identical if the non-quadratic matrix is chosen in the right way, as it could be seen in Section 3.5.2. It is clear, that with $m_1 = 1$ and $m_2 = 1$ this is an extreme example but there exist several other hypotheses with $m_2 < p$ as it could be seen in Sattler et al. [2022]. Moreover, other measurements based on the covariance matrix exist, which have small values for m_2, like the determinant or parameter to rate the quality of psychological questionnaires, see Pauly et al. [2016] for the latter. In this context, the saved time through the non-quadratic matrix and the other techniques is crucial since a dimension higher than 20 is often the case.
So for hypotheses which can be written as $\mathbf{C} = \mathbf{C}_W \otimes \mathbf{C}_S$, the parametric bootstrap approaches are usable even for a larger dimension as $d = 5$. The right choice between the parametric bootstrap and Monte-Carlo approach depends on the hypotheses and the used hypothesis matrix, as well as the available time and the required accuracy. Also, as can be seen in the appendix, the saved computation time is not always that high, each possibility to reduce the time without loss of quality should be used. And for our preferable test, the ATS with parametric bootstrap, the saved proportion is often up to 50%, like in the

test for equal covariance matrices, which is the most important hypothesis.

3.5.6 Different dimensions between the groups

Similar to the situation of Section 3.2 it could also be of interest to compare parts of covariance matrices with different dimensions. As a direct comparison between covariance matrices makes less sense, such parts could be single components or terms which directly depend on the covariance matrix. For example, the trace as an effect measure for the total variance could be used to compare multiple groups with different dimensions.

The adaptions of Sattler et al. [2022] are minimal, because with d_i as the dimension of the i-th group and $p_i = d_i(d_i+1)/2$ all results for the single groups are valid. And with $p = p_1 + ... + p_a$ the dimension of the pooled vector of covariances and appropriate hypothesis matrix $\mathbf{C} \in \mathbb{R}^{m \times p}$ this holds for all results from Sattler et al. [2022] involving more than one group.
The same holds for testing hypotheses regarding the correlation matrix, as is was done in Section 3.5.3.

3.5.7 Appendix

Influence of the used hypothesis matrix

Consider for the case of one group and dimension d=3 the matrices

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{C}^\top(\mathbf{C}\mathbf{C}^\top)^+ = \frac{1}{3}\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

For $\tilde{\boldsymbol{\zeta}} = (0,1,0,0,0,0)^\top$ it follows $\boldsymbol{\zeta} = \mathbf{0}_6$ and the equation $\mathbf{C}^\top(\mathbf{C}\mathbf{C}^\top)^+\mathbf{C}\boldsymbol{v} = \mathbf{0}$ has $\ker(\mathbf{C}^\top(\mathbf{C}\mathbf{C}^\top)^+\mathbf{C})$ as solution, which is a 5-dimensional subspace. On the contrary $\mathbf{C}\boldsymbol{v} = \tilde{\boldsymbol{\zeta}}$ has no solution, which shows impressively the difference between both hypotheses.

Testing hypotheses regarding the correlation matrix

Proof of Theorem 3.5.1: This proof is based on the proof from Browne and Shapiro [1986] and Nel [1985], where a similar situation is considered. Because we are just interested in the matrix's upper triangular, some adaptions have to be done.

With $\mathbf{\Delta}_i := \sqrt{n_i}(\widehat{\mathbf{V}}_i - \mathbf{V}_i)$ and $\mathbf{U}_i = \mathbf{V}_{i,0}^{-1/2}\mathbf{\Delta}_i\mathbf{V}_{i,0}^{-1/2}$, it can be calculated

$$\begin{aligned}
\widehat{\mathbf{R}}_i &= \widehat{\mathbf{V}}_{i,0}^{-1/2}\widehat{\mathbf{V}}_i\widehat{\mathbf{V}}_{i,0}^{-1/2} \\
&= \left(\mathbf{V}_{i,0} + \tfrac{1}{\sqrt{n_i}}\mathbf{\Delta}_{i,0}\right)^{-1/2}\left(\mathbf{V}_i + \tfrac{1}{\sqrt{n_i}}\mathbf{\Delta}_i\right)\left(\mathbf{\Sigma}_0 + \tfrac{1}{\sqrt{n_i}}\mathbf{\Delta}_{i,0}\right)^{-1/2} \\
&= \left(\mathbf{I}_p + \tfrac{1}{\sqrt{n_i}}\mathbf{U}_{i,0}\right)^{-1/2}\left(\mathbf{R}_i + \tfrac{1}{\sqrt{n_i}}\mathbf{U}_i\right)\left(\mathbf{I}_p + \tfrac{1}{\sqrt{n_i}}\mathbf{U}_{i,0}\right)^{-1/2}.
\end{aligned}$$

The Taylor series of $x \mapsto x^{-1/2}$ in point 1 leads to $x^{-1/2} = 1-(x-1)/2+\mathcal{O}((x-1)^2)$. For using this for the diagonal matrices with $x = 1 + n_i^{-1/2}\mathbf{U}_{i,0}$, we first consider the corresponding remainder. Since $\mathbf{U}_{i,0}$ converges to a normally distributed random variable from Slutzky's theorem we know that the remainder is $o_P\left(n_i^{-1/2}\right)$. This leads to

$$\left(\mathbf{I}_p + \frac{1}{\sqrt{n_i}}\mathbf{U}_{i,0}\right)^{-1/2} = \left(\mathbf{I}_p - \frac{1}{2\sqrt{n_i}}\mathbf{U}_{i,0}\right) + o_P\left(n_i^{-1/2}\right)$$

and hence

$$\begin{aligned}
\widehat{\mathbf{R}}_i &= \left(\mathbf{I}_p - \tfrac{1}{2\sqrt{n_i}}\mathbf{U}_{i,0}\right)\left(\mathbf{R}_i + \tfrac{1}{\sqrt{n_i}}\mathbf{U}_i\right)\left(\mathbf{I}_p - \tfrac{1}{2\sqrt{n_i}}\mathbf{U}_{i,0}\right) \\
&\quad + o_P\left(n_i^{-1/2}\right)\cdot\left[\left(\mathbf{I}_p - \tfrac{1}{2\sqrt{n_i}}\mathbf{U}_{i,0}\right)\left(\mathbf{R}_i + \tfrac{1}{\sqrt{n_i}}\mathbf{U}_i\right)\right] \\
&\quad + o_P\left(n_i^{-1/2}\right)\cdot\left[\left(\mathbf{R}_i + \tfrac{1}{\sqrt{n_i}}\mathbf{U}_i\right)\left(\mathbf{I}_p - \tfrac{1}{2\sqrt{n_i}}\mathbf{U}_{i,0}\right)\right] \\
&\quad + o_P\left(n_i^{-1/2}\right)\cdot\left[o_P\left(n_i^{-1/2}\right)\left(\mathbf{R}_i + \tfrac{1}{\sqrt{n_i}}\mathbf{U}_i\right)\right] \\
&= \left(\mathbf{I}_p - \tfrac{1}{2\sqrt{n_i}}\mathbf{U}_{i,0}\right)\left(\mathbf{R}_i + \tfrac{1}{\sqrt{n_i}}\mathbf{U}_i\right)\left(\mathbf{I}_p - \tfrac{1}{2\sqrt{n_i}}\mathbf{U}_{i,0}\right) + o_P\left(n_i^{-1/2}\right),
\end{aligned}$$

where again Slutzky's theorem is used. Multiplication with $\sqrt{n_i}$ leads to

$$
\begin{aligned}
&\sqrt{n_i}\,\widehat{\mathbf{R}}_i \\
= \;& \sqrt{n_i}\,\mathbf{R}_i + \left[\mathbf{U}_i - \tfrac{1}{2}\left(\mathbf{U}_{i,0}\mathbf{R}_i + \mathbf{R}_i\mathbf{U}_0\right)\right] + \sqrt{n_i}\cdot o_P\left(n_i^{-1/2}\right) \\
= \;& \sqrt{n_i}\,\mathbf{R}_i + \left[\mathbf{U}_i - \tfrac{1}{2}\left(\mathbf{U}_{i,0}\mathbf{R}_i + \mathbf{R}_i\mathbf{U}_{i,0}\right)\right] + o_P\left(1\right).
\end{aligned}
$$

We define

$$\mathbf{M}_1 = \sum_{\ell=1}^{p_u} \boldsymbol{e}_{\ell,p_u} \cdot \left(\boldsymbol{e}_{\mathbf{h}_{1\ell},p} + \boldsymbol{e}_{\mathbf{h}_{2\ell},p}\right)^\top$$

and

$$\mathbf{M}_2 := \sum_{\ell=1}^{p} \boldsymbol{e}_{\ell,p} \cdot \boldsymbol{e}_{\mathbf{h}_{4\ell},p}^\top \quad \mathbf{M}_3 := \sum_{\ell=1}^{p} \boldsymbol{e}_{\ell,p} \cdot \boldsymbol{e}_{\mathbf{h}_{3\ell},p}^\top$$

$$\mathbf{M}_4 := \mathbf{M}_2 + \mathbf{M}_3 \quad \mathbf{M}_5 := \mathit{diag}(\mathrm{vech}(\mathbf{I}_d)).$$

With these matrices it is easy to check that the following equations

$$\mathrm{vech}(\mathbf{U}_{i,0}\mathbf{R}_i) = \mathit{diag}(\mathrm{vech}(\mathbf{R}_i)) \cdot \mathbf{M}_2 \cdot \mathrm{vech}(\mathbf{U}_{i,0}),$$

$$\mathrm{vech}(\mathbf{R}_i\mathbf{U}_{i,0}) = \mathit{diag}(\mathrm{vech}(\mathbf{R}_i)) \cdot \mathbf{M}_3 \cdot \mathrm{vech}(\mathbf{U}_{i,0})$$

and

$$\mathrm{vech}(\mathbf{U}_{i,0}\mathbf{R}_i) + \mathrm{vech}(\mathbf{R}_i\mathbf{U}_{i,0}) = \mathrm{vech}(\mathbf{R}_i)) \cdot (\mathbf{M}_2 + \mathbf{M}_3)$$

hold, and therefore with $\mathrm{vech}(\mathbf{U}_{i,0}) = \mathbf{M}_5\,\mathrm{vech}(\mathbf{U}_i)$

$$
\begin{aligned}
&\sqrt{n_i}\,\mathrm{vech}(\widehat{\mathbf{R}}_i - \mathbf{R}_i) \\
= \;& \left[\mathrm{vech}(\mathbf{U}_i) - \tfrac{1}{2}\left(\mathrm{vech}(\mathbf{U}_{i,0}\mathbf{R}_i) + \mathrm{vech}(\mathbf{R}_i\mathbf{U}_{i,0})\right)\right] + o_P(1) \\
= \;& \left[\mathrm{vech}(\mathbf{U}_i) - \tfrac{1}{2}\,\mathit{diag}(\mathrm{vech}(\mathbf{R}_i)) \cdot (\mathbf{M}_2 + \mathbf{M}_3)\,\mathrm{vech}(\mathbf{U}_{i,0})\right] + o_P(1) \\
= \;& \left[\mathrm{vech}(\mathbf{U}_i) - \tfrac{1}{2}\,\mathit{diag}(\mathrm{vech}(\mathbf{R}_i)) \cdot (\mathbf{M}_2 + \mathbf{M}_3)\,\mathbf{M}_5\;\mathrm{vech}(\mathbf{U}_i)\right] + o_P(1) \\
= \;& \left[\mathbf{I}_p - \tfrac{1}{2}\,\mathit{diag}(\mathrm{vech}(\mathbf{R}_i) \cdot (\mathbf{M}_2 + \mathbf{M}_3)\,\mathbf{M}_5\right]\;\mathrm{vech}(\mathbf{U}_i) + o_P(1).
\end{aligned}
$$

Multiplication with the matrix $\mathbf{M}_5$ changes nothing in this case because it just picks the columns which are unequal to zero and drops the rest. So all in all with $\mathbf{M}_4 = \mathbf{M}_2 + \mathbf{M}_3$ it holds

$$\sqrt{n_i}\,\mathrm{vech}(\widehat{\mathbf{R}}_i - \mathbf{R}_i) = \left[\mathbf{I}_p - \tfrac{1}{2}\,diag(\mathrm{vech}(\mathbf{R}_i)) \cdot \mathbf{M}_4\right] \mathrm{vech}(\mathbf{U}_i) + o_P(1)$$

Now to adapt this result for the upper-half-vectorization, we use the special elimination matrix $\mathbf{L}_p^u$ which gives a connection between vech and vech$^-$

$$\begin{aligned} & \sqrt{n_i}(\widehat{\mathbf{r}}_i - \mathbf{r}_i) \\ = \; & \mathbf{L}_p^u \left[\mathbf{I}_p - \tfrac{1}{2}\,diag(\mathrm{vech}(\mathbf{R}_i)) \cdot \mathbf{M}_4\right] \mathrm{vech}(\mathbf{U}_i) + o_P(1) \\ = \; & \left[\mathbf{L}_p^u - \tfrac{1}{2}\,diag(\mathbf{r}_i) \cdot \mathbf{L}_p^u \cdot \mathbf{M}_4\right] \mathrm{vech}(\mathbf{U}_i) + o_P(1) \\ = \; & \left[\mathbf{L}_p^u - \tfrac{1}{2}\,diag(\mathbf{r}_i) \cdot \mathbf{M}_1\right] \mathrm{vech}(\mathbf{U}_i) + o_P(1). \end{aligned}$$

Here, we used the relation $\mathbf{L}_p^u \cdot \mathbf{M}_4 = \mathbf{M}_1$ and because of

$$\mathrm{vech}(\mathbf{U}_i) = diag(\mathrm{vech}((\nu_{i11}, ..., \nu_{idd})^\top \cdot (\nu_{i11}, ..., \nu_{idd})))^{-\frac{1}{2}}\,\mathrm{vech}(\boldsymbol{\Delta}_i)$$

it is useful to define

$$\mathbf{M}(\boldsymbol{\nu}_i, \mathbf{r}_i) = \left[\mathbf{L}_p^u - \frac{1}{2}\,diag(\mathbf{r}_i)\mathbf{M}_1\right] diag(\mathrm{vech}((\nu_{i11}, ..., \nu_{idd})^\top \cdot (\nu_{i11}, ..., \nu_{idd})))^{-\frac{1}{2}}.$$

Therefore, it holds

$$\sqrt{n_i}(\widehat{\mathbf{r}}_i - \mathbf{r}_i)\mathbf{M}(\boldsymbol{\nu}_i, \mathbf{r}_i)\,\mathrm{vech}(\boldsymbol{\Delta}_i) + o_P(1)$$

and because of Theorem 1 from Sattler et al. [2022] it follows

$$\sqrt{n_i}(\widehat{\mathbf{r}}_i - \mathbf{r}_i) \xrightarrow{\mathcal{D}} \mathbf{Z}_i \sim \mathcal{N}_{p_u}\left(\mathbf{0}_{p_u}, \underbrace{\mathbf{M}(\boldsymbol{\nu}_i, \mathbf{r}_i)\boldsymbol{\Sigma}_i\mathbf{M}(\boldsymbol{\nu}_i, \mathbf{r}_i)^\top}_{=:\boldsymbol{\Upsilon}_i}\right).$$

We could get the same result by using the δ-method on the results for the vectorized covariance matrices. In our opinion, the approach of Browne and Shapiro [1986], together with Nel [1985], is preferable due to its stepwise structure. Therefore it is more suitable to get an understanding of the used matrices. □

Proof of Theorem 3.5.2: With the result from Theorem 3.5.1, the asymptotic distribution of the quadratic form follows exactly from Theorem 2 from Sattler et al. [2022]. □

Proof of Theorem 3.5.3: We prove only the first part because the second part follows directly from the single groups' result.
For an application of the multivariate Lindeberg-Feller-Theorem (given the data), we need to check the conditions. As $\mathbf{Y}^*_{ik}$ under $\mathbf{X}$ is p_u-dimensional normal distributed with expectation $\mathbf{0}_{p_u}$ and variance $\widehat{\boldsymbol{\Upsilon}}_i$:

$$1.)\quad \sum_{k=1}^{n_i} \mathbb{E}\left(\frac{\sqrt{N}}{n_i}\mathbf{Y}^*_{ik}\middle|\mathbf{X}\right) = \sum_{k=1}^{n_i} \frac{\sqrt{N}}{n_i} \cdot \mathbb{E}\left(\mathbf{Y}^*_{ik}\middle|\mathbf{X}\right) = \mathbf{0}.$$

$$2.)\quad \sum_{k=1}^{n_i} Cov\left(\frac{\sqrt{N}}{n_i}\mathbf{Y}^*_{ik}\middle|\mathbf{X}\right) = \sum_{k=1}^{n_i} \frac{N}{n_i^2}\widehat{\boldsymbol{\Upsilon}}_i \xrightarrow{\mathcal{P}} \frac{1}{\kappa_i}\boldsymbol{\Upsilon}_i.$$

$$\begin{aligned} 3).\quad & \lim_{N\to\infty} \sum_{k=1}^{n_i} \mathbb{E}\left(\left|\left|\frac{\sqrt{N}}{n_i}\mathbf{Y}^*_{ik}\right|\right|^2 \cdot \mathbb{1}_{\left|\left|\frac{\sqrt{N}}{n_i}\mathbf{Y}^*_{ik}\right|\right|>\delta} \middle|\mathbf{X}\right) \\ = \; & \lim_{N\to\infty} \frac{N}{n_i^2} \sum_{k=1}^{n_i} \mathbb{E}\left(\left|\left|\mathbf{Y}^*_{i1}\right|\right|^2 \cdot \mathbb{1}_{||\mathbf{Y}^*_{i1}||>\delta\frac{n_i}{\sqrt{N}}} \middle|\mathbf{X}\right) \\ = \; & \frac{1}{\kappa_i} \cdot \lim_{N\to\infty} \mathbb{E}\left(\left|\left|\mathbf{Y}^*_{i1}\right|\right|^2 \cdot \mathbb{1}_{||\mathbf{Y}^*_{i1}||>\delta\frac{n_i}{\sqrt{N}}} \middle|\mathbf{X}\right) \\ \leqslant \; & \frac{1}{\kappa_i} \cdot \lim_{N\to\infty} \sqrt{\mathbb{E}\left(||\mathbf{Y}^*_{i1}||^4 \,|\mathbf{X}\right)} \cdot \sqrt{\mathbb{E}\left(\mathbb{1}_{||\mathbf{Y}^*_{i1}||>\delta\frac{n_i}{\sqrt{N}}} \middle|\mathbf{X}\right)} = 0. \end{aligned}$$

For the last part, we used the Cauchy-Bunjakowski-Schwarz-Inequality and that we know $\mathbb{E}\left(\left|\left|\mathbf{Y}^*_{i1}\right|\right|^4 |\mathbf{X}\right) < \infty$. Finally through $n_i/N \to \kappa_i$ and therefore $\delta \cdot n_i/\sqrt{N} \to \infty$, it holds $P\left(\left|\left|\mathbf{Y}^*_{i1}\right|\right| > \delta \cdot n_i/\sqrt{N}\right) \to 0$, which leads to the result.

Given the data $\mathbf{X}$, it follows that $\sqrt{N}\,\overline{\mathbf{Y}}_i^*$ converges in distribution to $\mathbf{Z}_i \sim \mathcal{N}_{p_u}(\mathbf{0}_{p_u}, 1/\kappa_i \cdot \boldsymbol{\Upsilon}_i)$ and therefore because of independence of groups $\sqrt{N}\,\overline{\mathbf{Y}}^*$ converges in distribution to $\mathbf{Z} \sim \mathcal{N}_{ap_u}(\mathbf{0}_{ap_u}, \boldsymbol{\Upsilon})$.

Because the empirical covariance matrix of the bootstrap sample is consistent, it follows $\widehat{\boldsymbol{\Upsilon}}_i^* \xrightarrow{\mathcal{P}} \widehat{\boldsymbol{\Upsilon}}_i$. From the construction of $\widehat{\boldsymbol{\Upsilon}}_i$ and the properties of $\widehat{\boldsymbol{\Sigma}}_i$ it is clear that $\widehat{\boldsymbol{\Upsilon}}_i \xrightarrow{\mathcal{P}} \boldsymbol{\Upsilon}_i$. The result follows with the triangle inequality and the continuous mapping theorem. □

Efficient implementation of the developed tests

For testing hypotheses regarding the covariance matrix, we use quadratic forms. Here, we want to investigate more detailed the required number of multiplications for their calculation in different ways.
Thereby, the presented numbers should be seen as heuristic and can potentially be reduced through smart implementation.

For the classical way to calculate $\mathbf{C}_S\widehat{\boldsymbol{\Sigma}}_i\mathbf{C}_S^\top$, the calculation of the empirical covariance needs $n_i p^2$ multiplications while for the calculation of the matrix product $p^2 m_2 + m_2^2 p$ multiplications are needed. Thereby, this "classical" approach needs $n_i p^2 + p^2 m_2 + m_2^2 p$ multiplications.
For the alternative approach, first $\mathbf{C}_S\mathbf{X}_{i1}, ..., \mathbf{C}_S\mathbf{X}_{in_i}$, and afterwards, the empirical covariance matrix of these vectors have to be calculated. All in all, these are $n_i p m_2 + n_i m_2^2$ multiplications. For $n_i \leqslant p$, our alternative is more efficient while for $n_i > p$ it depends on the concrete setting.
Since for the bootstrap approaches, the bootstrap sample has not to be multiplied with $\mathbf{C}_S$, the number is reduced to $n_i m_2^2$, which is clearly lower than for the classical approach, even for $p = m_2$.

Finally, it remains to check, wheter $(\mathbf{C}_W \otimes \mathbf{I}_{m_2})\left(\bigoplus_{i=1}^a \mathbf{C}_S\widehat{\boldsymbol{\Sigma}}_i\mathbf{C}_S^\top\right)(\mathbf{C}_W \otimes \mathbf{I}_{m_2})^\top$ or $\mathbf{C}\widehat{\boldsymbol{\Sigma}}\mathbf{C}^\top$ need more multiplications, because the latter representation is necessary for our alternative approach. For the classical approach, we need $p^2 a^2 m_1 m_2 + m_1^2 m_2^2 + m_1^2 m_2^2 p a$ multiplications while the alternative approach requires $m_2^3 m_1 a + m_1 m_2^2 a^2 \cdot 2$ multiplications. Since $p^2 a^2 m_1 m_2 > m_1 m_2^2 a \cdot 2$ and $m_1^2 m_2^2 p a > m_2^3 m_1 a$ the required number of multiplications can be substantially reduced, which makes this way clearly faster.

The following tables show the required computation times for a time measurement with the same setting as in Section 6 in Sattler et al. [2022]. We will show the hypothesis matrices for the used hypotheses again.

A) Equal Covariance Matrices: Testing the hypothesis $\mathcal{H}_0^v : \{\mathbf{V}_1 = \mathbf{V}_2\} = \{\mathbf{C}(A)\boldsymbol{v} = \mathbf{0}\}$ is most of the time described by $\mathbf{C}(A) = \mathbf{P}_2 \otimes \mathbf{I}_p$, but an alternative choice would be $\widetilde{\mathbf{C}}(A) = (1, -1) \otimes \mathbf{I}_p \in \mathbb{R}^{p \times 2p}$.

B) Equal Diagonal Elements: The hypothesis $\mathcal{H}_0^v : \{\mathbf{V}_{111} = ... = \mathbf{V}_{1dd}\} = \{\mathbf{C}(B)\boldsymbol{v} = \mathbf{0}\}$ can, e.g., be described by $\mathbf{C}(B) = \mathit{diag}(\mathbf{h}_d) - \mathbf{h}_d \cdot \mathbf{h}_d^\top / d$, but the same hypothesis is tested by chosing the a_1-th row up to the a_d-th row of $\mathbf{C}(B)$ and build $\widetilde{\mathbf{C}}(B)$ this way. Here again for $k = 1, ..., d$, we use $a_k = 1 + \sum_{j=1}^{k-1}(d + 1 - j)$.

D) Test for a given trace: $\mathcal{H}_0^v : \{tr(\mathbf{V}_1) = \gamma\}\{\mathbf{C}(D)\boldsymbol{v} = \mathbf{h}_d \cdot \gamma\}$ for a given value $\gamma \in \mathbb{R}$ can be described by $\mathbf{C}(D) = [\mathbf{h}_d \cdot \mathbf{h}_d^\top]/d$ or $\widetilde{\mathbf{C}}(D) = \mathbf{h}_d^\top/d \in \mathbb{R}^{1\times p}$ through $\mathcal{H}_0^v : \{tr(\mathbf{V}_1) = \gamma\}\{\widetilde{\mathbf{C}}(D)\boldsymbol{v} = \gamma\}$.

In contrast to the investigation of the computation time from Sattler et al. [2022], for hypothesis B) we use another matrix with d rows instead of d−1 rows. Here, just the zero-rows are removed, which does not change the result.

Again, we take the average time of eight different configurations, which consist of four distributions (t_9, normal, skew normal, and gamma) and two covariance matrices ($(\mathbf{V}_1)_{ij} = 0.6^{|i-j|}$ and $\mathbf{V}_2 = \mathbf{I}_d + \mathbf{J}_d$), to get more reliable results. Moreover, this was calculated 100 times and afterwards averaged, while 1,000 bootstrap steps resp. 10,000 Monte-Carlo repetitions were used.

	$\mathbf{C}(A)$				$\widetilde{\mathbf{C}}(A)$			
d	2	5	10	20	2	5	10	20
ATS-Para	0.525	0.905	17.560	142.577	0.521	0.884	16.902	114.385
ATS-Wild	0.410	0.540	10.144	112.359	0.410	0.521	9.493	85.180
ATS	0.084	0.196	0.387	1.199	0.058	0.153	0.275	0.691
WTS-Para	0.607	1.202	34.023	384.415	0.589	1.078	24.477	196.349
WTS-Wild	0.539	0.867	26.929	356.703	0.514	0.743	17.366	168.426
WTS-χ^2	0.002	0.003	0.043	0.408	0.002	0.003	0.035	0.207

Table 3.21: Required time in seconds for various tests statistics and different dimensions for hypothesis A) ($\mathcal{H}_0^v : \{\mathbf{V}_1 = \mathbf{V}_2\}$) with a quadratic hypothesis matrix on the left side and a non-quadratic hypothesis matrix on the right sight. Here some methods are used to increase efficiency.

This way, a remarkable time reduction can be achieved, allowing bootstrap approaches for higher dimensions.

	$\mathbf{C}(\mathrm{B})$				$\widetilde{\mathbf{C}}(\mathrm{B})$			
d	2	5	10	20	2	5	10	20
ATS-Para	0.250	0.402	5.424	39.496	0.241	0.302	0.850	1.337
ATS-Wild	0.231	0.279	1.404	12.660	0.228	0.259	0.646	0.845
ATS	0.027	0.087	0.313	1.077	0.020	0.038	0.111	0.181
WTS-Para	0.330	0.525	11.842	101.937	0.319	0.386	1.075	2.055
WTS-Wild	0.291	0.393	8.610	84.960	0.286	0.335	0.896	1.584
WTS-χ^2	0.001	0.001	0.015	0.100	0.001	0.001	0.004	0.006

Table 3.22: Required time in seconds for various tests statistics and different dimensions for hypothesis B) ($\mathcal{H}_0^v : \{\mathbf{V}_{111} = ... = \mathbf{V}_{1dd}\}$) with a quadratic hypothesis matrix on the left side and a non-quadratic hypothesis matrix on the right sight. Here some methods are used to increase efficiency.

	$\mathbf{C}(\mathrm{D})$				$\widetilde{\mathbf{C}}(\mathrm{D})$			
d	2	5	10	20	2	5	10	20
ATS-Para	0.314	0.527	9.348	57.335	0.298	0.350	0.435	0.634
ATS-Wild	0.275	0.336	0.861	40.259	0.271	0.285	0.305	0.373
ATS	0.047	0.152	0.257	0.626	0.037	0.068	0.123	0.163
WTS-Para	0.401	0.677	13.245	104.695	0.383	0.455	0.585	0.937
WTS-Wild	0.351	0.471	8.939	83.664	0.345	0.377	0.442	0.669
WTS-χ^2	0.001	0.002	0.018	0.104	0.001	0.001	0.002	0.003

Table 3.23: Required time in seconds for various tests statistics and different dimensions for hypothesis D) ($\mathcal{H}_0^v : \{\mathrm{tr}(\mathbf{V}_1) = \gamma\}$) with a quadratic hypothesis matrix on the left side and a non-quadratic hypothesis matrix on the right sight. Here some methods are used to increase efficiency.

4 Discussion and Outlook

4.1 Discussion

This thesis expands and complements existing procedures to analyze repeated measure and general MANOVA designs. Thereby, a specific focus is on high-dimensional repeated measure designs, for which especially the manifold asymptotic frameworks are exceptional for heterogeneous covariance matrices as well as for homogeneous ones. Moreover, it was possible to expand our approach for groups with different numbers of repetitions, where equal covariance matrices would make no sense. The normality assumption in high-dimensional data is rather restrictive and needs to be verified. However, it allows us to work without many other requirements, which usually are even harder to justify. Under the additional assumption of equal covariance matrices, we considered more asymptotic frameworks and relinquished some conditions. In some settings, we could even work without estimating τ_P, since $\tau_P \to 0$ can be conducted from known values.

The validation of the used homoscedasticity of covariance matrices leads to the topic of preliminary tests on variances and more general hypotheses regarding covariance matrices. Here, innovative considerations were made to develop a family of test statistics with few requirements and a wide field of possible applications. The simulation results were compelling, especially for more groups, higher dimensions, or cases where the considered groups have different distributions. The latter one is more challenging than equal distributions and violates the assumption $\mathbb{E}\left([\text{vech}(\epsilon_{11}\epsilon_{11}^\top)][\text{vech}(\epsilon_{11}\epsilon_{11})^\top]^\top\right) = ... = \mathbb{E}\left([\text{vech}(\epsilon_{a1}\epsilon_{a1})^\top)][\text{vech}(\epsilon_{a1}\epsilon_{a1}^\top)]^\top\right)$. The equality of this moment in all groups is part of some tests from Zhang and Boos [1992]. This violation has considerable effects on these tests and makes them nearly unusable, while even their bootstrap approach without this condition performs clearly worse in this setting.

Since not always a unique hypothesis matrix could be chosen in the considered model, we investigated how this influences the test's final result. It turned out that the usual way to choose the unique projection matrix as the hypothesis matrix can, in some cases, be improved with regard to the required time. This fact needs to be kept in mind while working with quadratic forms in vectors with high dimension, regardless of whether they are means, vectorized matrices, or something else.

Due to our general model, it was possible to expand our approach to the related testing of hypotheses formulated in correlation matrices. A correlation matrix is focused on the underlying dependency structure and therefore is sometimes more suitable. It performed quite well compared to existing procedures, while it is more flexible and allows various applications.

Moreover, with this additional result, it is possible to examine whether the covariance or correlation matrix has a particular structure, like an autoregressive or a compound symmetry matrix. More generally, many different kinds of patterns can be checked by choosing a proper hypothesis matrix. Simulations showed our test's good properties, while for testing the diagonality of the covariance matrix, the test of Bartlett [1951] had even better values. Since this test is designed only for this concrete structure and the number of other tests tackling this problem is limited, our test is very useful for modeling. It allows checking for given data, whether the covariance matrix has a specific structure, corresponding with some model assumptions. For example, an autoregressive matrix or a Toeplitz matrix shows that all components have the same variance and the covariance between components only depends on the distance between them. This conforms to a repeated measure design, where all time points have the same variance, and only the time difference is relevant for the covariance of two components, not the concrete time. Finally, it was essential to investigate how the required computation time for testing hypotheses regarding the covariance matrix could be reduced. The presented possibilities are applicable to all the expansions of the original approach. Besides, they are substantially in the development of an R-package to make the tests accessible for users, see the end of 4.2 below.

4.2 Outlook

In Theorem 3.1 a) from Sattler [2021] there is an "if and only if"-relation while part b) is only if. Finding expansion for it to be "if and only if" will complete the theorem and is another future research point. A solution was not found yet, and it remains an open task. Also, to further decompose part b) and prove it for these smaller parts might be useful. Other goals are to widen the applicability

of this theorem further, include more settings of the standardized eigenvalues β_i, and investigate the behavior of K_{f_P} in settings with $\beta_1 \in (0, 1)$.
Moreover, the work of Chen and Qin [2010] is, in our opinion, one of the most important papers in the area of inferring means of high-dimensional data since they do not have any condition on the relation between sample size and dimension and further, have a semi-parametric model. On the other hand, they consider only two groups and need $\beta_1 \to 0$. This restriction on two groups follows from the fact that hypotheses for two groups always can be expressed as subtraction of the means of the groups. Since this subtraction of two means is a key component of their work, it is interesting to expand this to more groups. Also, their test statistic's behavior in the case of $\beta_1 \to 1$ should be investigated to allow the application in many more situations.

Our approach for testing hypotheses regarding the covariance matrix is quite general and can be used for various other situations. Testing hypotheses regarding the correlation matrix or testing for the structure of the covariance matrix is, in our opinion, just the beginning. There is a variety of situations for which this approach can be adapted. For example, there are some measurements based on the covariance matrix respective correlation matrix, like Cronbach's alpha. With this value, the dependency between questions in questionnaires is evaluated. In Pauly et al. [2016], a permutation-based approach for comparing two Cronbach's alphas was introduced, so it is of interest for future research to compare their performance with an adaption of our bootstrap-based procedure.
Moreover, many existing issues can be reformulated regarding a covariance matrix. For example, the occurrence of a random block effect can be investigated by analyzing corresponding covariance matrices.
As has been mentioned several times, the computation time is a usual difficulty in analyzing vectors with a larger dimension. This led to a closer look at a topic otherwise given relatively little attention: choosing the hypothesis matrix. Since there is the necessity of well-founded statements on the influence of the chosen hypothesis matrix on the test statistic, further considerations will be made.

It is especially interesting to combine both presented main topics to find an approach for testing hypotheses regarding covariance and correlation matrices of high-dimensional observations. This is rather difficult for two main reasons.

First, since the empirical covariance matrix performs not reliable in situations with increasing dimension (which was illustrated in Table 2.1) an expansion of Theorem 1 from Sattler et al. [2022] for increasing dimension will be challenging. Under additional conditions on the covariance matrix $\mathbf{\Sigma}_i$, covariance estimators for high-dimensional settings like banding estimators, were developed in Bickel and Levina [2008] and Cai et al. [2016]. Such an estimator $\widetilde{\mathbf{\Sigma}}_i$ could be used to replace $\widehat{\mathbf{\Sigma}}_i$ and therefore to investigate the asymptotic distribution of $\sqrt{N}(\text{vech}(\mathbf{\Sigma}_i) - \text{vech}(\widetilde{\mathbf{\Sigma}}_i))$ under the null hypothesis.
The second difficulty is the subsequent estimation of $\mathit{Cov}(\text{vech}(\widetilde{\mathbf{\Sigma}}_i))$, which is necessary for building test statistics. Again, under additional requirements on $\mathit{Cov}(\text{vech}(\widetilde{\mathbf{\Sigma}}_i))$ existing estimator for structured high-dimensional covariance can be applied. But, since we expect a quite complex structure of $\mathit{Cov}(\text{vech}(\widetilde{\mathbf{\Sigma}}_i))$, comparable to $\mathbf{V}_i = \mathit{Cov}(\text{vech}([\mathbf{X}_{i1} - \mathbb{E}(\mathbf{X}_{i1})][\mathbf{X}_{i1} - \mathbb{E}(\mathbf{X}_{i1})]^\top))$, all assumptions here are difficult to verify. Nevertheless, this should be further investigated in future research. An interesting competitor would be Bai et al. [2019], which have a quite general high-dimensional semi-parametric setting, but allow only hypotheses regarding linear combinations of covariance matrices.

Since we were able to expand the results from Sattler and Pauly [2018] and Sattler et al. [2022] for groups with different dimensions, it remains an open question whether it is possible to adapt these approaches for the case of observations with different dimensions within the groups. This kind of data is sometimes called clustered data[1]. For mean-based analysis, we are currently working on an approach based on a permutation technique that can handle increasing dimensions under some conditions.

Finally, for the usage of our approaches from Sattler and Pauly [2018] and Sattler et al. [2022] in practice, it is helpful to have an R-package, including the corresponding functions. We plan such a package for the future. It makes sense to include the test for correlation matrices, introduced in Section 3.5.3, or even the test for a particular structure of the covariance/correlation matrix from Section 3.5.4.

[1]This should not be confused with the cluster-analysis, which is, for example, used to group big sets of objects.

Part II

Publications

Article 1

Sattler, P. and Pauly, M. (2018). Inference for high-dimensional split-plot-designs: An unified approach for small to large numbers of factor levels. Electronic Journal of Statistics 12, 2743-2805, doi.org/10.1214/18-EJS1465

Electronic Journal of Statistics
Vol. 12 (2018) 2743–2805
ISSN: 1935-7524
https://doi.org/10.1214/18-EJS1465

Inference for high-dimensional split-plot-designs: A unified approach for small to large numbers of factor levels

Paavo Sattler and Markus Pauly

Institute of Statistics, Ulm University, Helmholtzstr. 20, 89081 Ulm, Germany.
e-mail: paavo.sattler@uni-ulm.de; markus.pauly@uni-ulm.de

Abstract: Statisticians increasingly face the problem to reconsider the adaptability of classical inference techniques. In particular, diverse types of high-dimensional data structures are observed in various research areas; disclosing the boundaries of conventional multivariate data analysis. Such situations occur, e.g., frequently in life sciences whenever it is easier or cheaper to repeatedly generate a large number d of observations per subject than recruiting many, say N, subjects. In this paper, we discuss inference procedures for such situations in general heteroscedastic split-plot designs with a independent groups of repeated measurements. These will, e.g., be able to answer questions about the occurrence of certain time, group and interactions effects or about particular profiles.

The test procedures are based on standardized quadratic forms involving suitably symmetrized U-statistics-type estimators which are robust against an increasing number of dimensions d and/or groups a. We then discuss their limit distributions in a general asymptotic framework and additionally propose improved small sample approximations. Finally, the small sample performance is investigated in simulations and applicability is illustrated by a real data analysis.

Keywords and phrases: Approximations, high-dimensional data, quadratic forms, repeated measures, split-plot designs.

Received October 2017.

Contents

1 Introduction 2744
2 Statistical model and hypotheses 2745
3 The test statistic and its asymptotics 2746
3.1 Symmetrized trace estimators 2749
4 Better approximations 2751
5 Simulations 2754
5.1 Asymptotic distribution and type-I error control 2755
5.2 Power performance 2760
6 Analysis of a sleep laboratory data set 2761
7 Conclusion & outlook 2765
Acknowledgement 2766

A Basics . . . 2766
B Proofs of Section 3 . . . 2772
C Proofs of Section 4 . . . 2787
D On the asymptotic distribution in our simulation designs . . . 2798
E On the Chen-Qin-Condition . . . 2798
References . . . 2802

1. Introduction

In our current century of data, statisticians increasingly face the problem to reconsider the adaptability of classical inferential techniques. In particular, diverse types of high-dimensional data structures are observed in various research areas; disclosing the boundaries of conventional multivariate data analysis. Here, the *curse of high dimensionality* or the *large d small N problem* is especially encountered in life sciences whenever it is easier (or cheaper) to repeatedly generate a large number d of observations per subject than recruiting many, say N, subjects. Similar observations can be made in industrial sciences with subjects replaced by units. Such designs, where experimental units are repeatedly observed under different conditions or at different time points, are called *repeated measures designs* or (if two or more groups are observed) *split-plot designs*. In these trials, one likes to answer questions about the occurrence of certain group or time effects or about particular profiles. Conventionally, for $d < N$, corresponding null hypotheses are inferred with Hotelling's T^2 (one or two sample case) or Wilks's Λ, see e.g. Davis [14][Section 4.3] or Johnson & Wichern [24] [Section 6.8]. Besides normality, these procedures heavily rely on the assumption of equal covariance matrices and particularly break down in high-dimensional settings with $N < d$. While there exist several promising approaches to adequately deal with the problem of covariance heterogeneity in the classical case with $d < N$ (see e.g. Box [6], Geisser & Greenhouse [17], Greenhouse & Geisser [18], Huynh & Feldt [23], Lecoutre [30], Vallejo & Ato [40], Ahmad et al. [1], Kenward & Roger [27], Brunner et al. [9], Pesarin & Salmaso [35], Skene & Kenward [38], Konietschke et al. [29], Happ et al. [20], Harden [21], Friedrich et al. [16]) most procedures for high-dimensional repeated measures designs rely on certain sparsity conditions (see e.g. Bai & Saranadasa [2], Chen & Qin [11], Katayama et al. [26], Nishiyama et al. [33], Secchi et al. [37], Cai et al. [10], Harrar & Kong [22] and the references cited therein). In particular, in an asymptotic $(d, N) \to \infty$ framework, typical assumptions restrict the way the sample size N and/or various powers of traces of the underlying covariances increase with respect to d. These type of sparsity conditions guarantee central limit theorems that lead to approximations of underlying test statistics by a fixed limit distribution. However, as illustrated in Pauly et al. [34] for one-sample repeated measures these conditions can in general not be regarded as regularity assumptions. In particular, they may even fail for classical covariance structures. To this end, the authors proposed a novel approximation technique that showed considerably accurate results and investigated its asymptotic behavior in a flexible and non-restrictive $(d, N) \to \infty$ framework. Here, no assumptions regarding the dependence between d and N or the covariance matrix were made. In the current paper, we follow this approach and extend the results of Pauly et al. [34] to general heteroscedastic split-plot designs with a independent groups of repeated measurements. To even allow a large number of groups as in Bathke & Harrar [3], Bathke et al. [4] or Zhan & Hart [43], we do not only consider the case with a fixed number $a \in \mathbb{N}$ of samples but additionally allow for situations with $a \to \infty$. The latter case is of particular interest if most groups are rather small (as in screening trials) such that a classical test would essentially possess no power for fixed a. Here increasing the number of groups implies increasing the total sample size from which a power increase might be expected as well. This leads to one of the following asymptotic frameworks

$$\begin{aligned} a \in \mathbb{N} \ \text{fixed} \quad &\text{and} \quad (d, N) \to \infty, \\ d \in \mathbb{N} \ \text{fixed} \quad &\text{and} \quad (a, N) \to \infty, \\ &\text{or} \quad (a, d, N) \to \infty \end{aligned}$$

which we handle simultaneously in the sequel. For all considerations, the adequate and dimension-stable estimation of traces of certain powers of combined covariances turned out to be a major problem. It is tackled by introducing symmetrized estimates of U-statistics-type which possess nice asymptotic properties under all asymptotic frameworks given above.

The paper is organized as follows. The statistical model together with the considered hypotheses of interest are introduced in Section 2. The test statistic and its asymptotic behavior is investigated in Section 3, where also novel dimension-stable trace estimators are introduced. Additional approximations for small sample sizes are theoretically discussed in Section 4 and their performance is studied in simulations in Section 5. Afterwards, the new methods will be applied to analyze a high-dimensional data set from a sleep-laboratory trial in Section 6. The paper closes with a discussion and an outlook. All proofs in this paper are shifted to the Appendix.

2. Statistical model and hypotheses

We consider a split-plot design given by a independent groups of d-dimensional random vectors

$$\boldsymbol{X}_{i,j} = (X_{i,j,1}, \dots, X_{i,j,d})^\top \overset{ind}{\sim} \mathcal{N}_d(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \quad j = 1, \dots, n_i, \quad i = 1, \dots, a \qquad (1)$$

with mean vectors $\mathbb{E}(\boldsymbol{X}_{i,1}) = \boldsymbol{\mu}_i = (\mu_{i,t})_{t=1}^d \in \mathbb{R}^d$ and positive definite covariance matrices $Cov(\boldsymbol{X}_{i,1}) = \boldsymbol{\Sigma}_i$. Here $j = 1, \dots, n_i$ denotes the individual subject or unit in group $i = 1, \dots, a$, $n_i, a \in \mathbb{N}$, where no specific structure of the group-specific covariance matrices $\boldsymbol{\Sigma}_i$ is assumed. In particular, they are even allowed to differ completely. Altogether we have a total number of $N = \sum_{i=1}^a n_i$ random vectors representing observations from independent subjects. Within this framework, a factorial structure on the factors group or time can be incorporated by splitting up indices. Also, a group-specific random subject effect can be incorporated as outlined in Pauly et al. [34][Equation (2.2)].

Writing $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \dots, \boldsymbol{\mu}_a^\top)^\top$, linear hypotheses of interest in this general split-plot model are formulated as

$$H_0(\boldsymbol{H}) : \boldsymbol{H}\boldsymbol{\mu} = \boldsymbol{0} \tag{2}$$

for a proper hypothesis matrix $\boldsymbol{H}$. It is of the form $\boldsymbol{H} = \boldsymbol{H}_W \otimes \boldsymbol{H}_S$, where $\boldsymbol{H}_W$ and $\boldsymbol{H}_S$ refer to **w**hole-plot (group) and/or **s**ubplot (time) effects. For theoretical considerations it is often more convenient to reformulate $H_0(\boldsymbol{H})$ by means of the corresponding projection matrix $\boldsymbol{T} = \boldsymbol{H}^\top[\boldsymbol{H}\boldsymbol{H}^\top]^-\boldsymbol{H}$, see e.g. Pauly et al. [34]. Here $(\cdot)^-$ denotes some generalized inverse of the matrix and $H_0(\boldsymbol{H})$ can equivalently be written as $H_0(\boldsymbol{T}) : \boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0}$. It is a simple exercise to prove that the matrix $\boldsymbol{T}$ is of the form $\boldsymbol{T} = \boldsymbol{T}_W \otimes \boldsymbol{T}_S$ for projection matrices $\boldsymbol{T}_W$ and $\boldsymbol{T}_S$, see Lemma A.1 (p.2766) in the Appendix. Typical examples are given by

(a) No group effect: $H_0^a : \left(\boldsymbol{P}_a \otimes \frac{1}{d}\boldsymbol{J}_d\right) \boldsymbol{\mu} = \boldsymbol{0}$,
(b) No time effect: $H_0^b : \left(\frac{1}{a}\boldsymbol{J}_a \otimes \boldsymbol{P}_d\right) \boldsymbol{\mu} = \boldsymbol{0}$,
(c) No interaction effect between time and group: $H_0^{ab} : (\boldsymbol{P}_a \otimes \boldsymbol{P}_d)\, \boldsymbol{\mu} = \boldsymbol{0}$,

where $\boldsymbol{J}_d$ is the d-dimensional matrix only containing 1s and $\boldsymbol{P}_d := \boldsymbol{I}_d - \boldsymbol{J}_d/d$ is the centring matrix. For interpretational purposes it is sometimes helpful to decompose the component-wise means as

$$\mu_{i,t} = \mu + \alpha_i + \beta_t + (\alpha\beta)_{it}, \qquad i = 1, \dots, a,\ t = 1, \dots, d,$$

where $\alpha_i \in \mathbb{R}$ represents the i-th group effect, $\beta_t \in \mathbb{R}$ the time effect at time point t and $(\alpha\beta)_{it} \in \mathbb{R}$ the (i,t)-interaction effect between group and time with the usual side conditions $\sum_i \alpha_i = \sum_t \beta_t = \sum_{i,t}(\alpha\beta)_{it} = 0$. With this notation the above null hypothesis can be rewritten as (a) $H_0^a : \alpha_i \equiv 0$ for all i, (b) $H_0^b : \beta_t \equiv 0$ for all t and (c) $H_0^{ab} : (\alpha\beta)_{it} \equiv 0$ for all i, t, respectively.

These and other hypotheses will be utilized in the data analysis Section 6.

3. The test statistic and its asymptotics

We derive appropriate inference procedures for $H_0(\boldsymbol{T})$ and analyze their asymptotic properties under the following asymptotic frameworks

$$a \in \mathbb{N} \text{ fixed} \quad \text{and} \quad \min(d, n_1, \dots, n_a) \to \infty, \tag{3}$$
$$d \in \mathbb{N} \text{ fixed} \quad \text{and} \quad \min(a, n_1, \dots, n_a) \to \infty, \tag{4}$$
$$\text{or} \quad \min(a, d, n_1, \dots, n_a) \to \infty, \tag{5}$$

as $N \to \infty$. Here, no dependency on how the dimension $d = d(N)$ in (3) and (5) or the number of groups $a = a(N)$ in (4)–(5) converges to infinity with respect to the sample sizes n_i and N is postulated. In particular, we cover high-dimensional ($d > n_i$ or even $d > N$) as well as low-dimensional settings. For a lucid presentation of subsequent results and proofs we additionally assume throughout that

$$\frac{n_i}{N} \to \rho_i \in (0,1), \quad i = 1, \dots, a. \tag{6}$$

However, by turning to convergent subsequences, all main results can be shown to hold under the more general condition

$$0 < \liminf n_i/N \leq \limsup n_i/N < 1, \quad (i = 1, \dots, a).$$

It is convenient to measure deviations from the null hypothesis $H_0(\boldsymbol{T}) : \boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0}$ by means of the quadratic form

$$Q_N = N \cdot \overline{\boldsymbol{X}}^\top \boldsymbol{T} \overline{\boldsymbol{X}}, \tag{7}$$

where $\overline{\boldsymbol{X}}^\top = (\overline{\boldsymbol{X}}_1^\top, \dots \overline{\boldsymbol{X}}_a^\top)$ with $\overline{\boldsymbol{X}}_i = n_i^{-1} \sum_{j=1}^{n_i} \boldsymbol{X}_{i,j}, i = 1, \dots, a$, denotes the vector of pooled group means.

Since Q_N is in general asymptotically degenerated under (3)–(5) we study its standardized version. To this end, note that under the null hypothesis it holds that

$$\sqrt{N} \cdot \boldsymbol{T}\overline{\boldsymbol{X}} \overset{H_0}{\sim} \mathcal{N}_{ad}\left(\boldsymbol{0}_{ad}, \boldsymbol{T}\left[\bigoplus_{i=1}^{a} \frac{N}{n_i}\boldsymbol{\Sigma}_i\right]\boldsymbol{T}\right),$$

due to assumption (1). Thus, it follows from classical theorems about moments of quadratic forms, see e.g. Mathai & Provost [32] or Theorem A.4 in the Appendix, that its mean and variance under the null hypothesis can be expressed as

$$\mathbb{E}_{H_0}(Q_N) = \operatorname{tr}\left(\boldsymbol{T}\left[\bigoplus_{i=1}^{a} \frac{N}{n_i}\boldsymbol{\Sigma}_i\right]\right) = \sum_{i=1}^{a} \frac{N}{n_i} (\boldsymbol{T}_W)_{ii} \operatorname{tr}(\boldsymbol{T}_S \boldsymbol{\Sigma}_i), \tag{8}$$

$$\begin{aligned} Var_{H_0}(Q_N) &= 2 \operatorname{tr}\left(\left(\boldsymbol{T}\left[\bigoplus_{i=1}^{a} \frac{N}{n_i}\boldsymbol{\Sigma}_i\right]\right)^2\right) && (9) \\ &= 2 \sum_{i=1}^{a} \sum_{r=1}^{a} \frac{N^2}{n_i n_r} (\boldsymbol{T}_W)_{ir} (\boldsymbol{T}_W)_{ri} \operatorname{tr}(\boldsymbol{T}_S \boldsymbol{\Sigma}_i \boldsymbol{T}_S \boldsymbol{\Sigma}_r) \\ &= 2 \sum_{i=1}^{a} \sum_{r=1}^{a} \frac{N^2}{n_i n_r} {(\boldsymbol{T}_W)_{ir}}^2 \operatorname{tr}(\boldsymbol{T}_S \boldsymbol{\Sigma}_i \boldsymbol{T}_S \boldsymbol{\Sigma}_r) && (10) \\ &= 4 \sum_{i,r=1, r<i}^{a} \frac{N^2}{n_i n_r} {(\boldsymbol{T}_W)_{ir}}^2 \operatorname{tr}(\boldsymbol{T}_S \boldsymbol{\Sigma}_i \boldsymbol{T}_S \boldsymbol{\Sigma}_r) \\ &\quad + 2 \sum_{i=1}^{a} \frac{N^2}{n_i^2} {(\boldsymbol{T}_W)_{ii}}^2 \operatorname{tr}\left((\boldsymbol{T}_S \boldsymbol{\Sigma}_i)^2\right). \end{aligned}$$

Henceworth we investigate the asymptotic behaviour (under $H_0(\boldsymbol{T})$) of the standardized quadratic form $\widetilde{W}_N = \{Q_N - \mathbb{E}_{H_0}(Q_N)\} / \, Var_{H_0}(Q_N)^{1/2}$. Denoting by $\boldsymbol{V}_N := \bigoplus_{i=1}^{a} \frac{N}{n_i} \boldsymbol{\Sigma}_i$ the inversely weighted combined covariance matrix, the representation theorem for quadratic forms given in Mathai & Provost [32][p.90], implies that

$$\widetilde{W}_N \quad = \quad \frac{Q_N - \mathbb{E}_{H_0}(Q_N)}{Var_{H_0}(Q_N)^{1/2}} \stackrel{\mathcal{D}}{=} \sum_{s=1}^{ad} \frac{\lambda_s}{\sqrt{\sum_{\ell=1}^{ad} \lambda_\ell^2}} \left(\frac{C_s - 1}{\sqrt{2}}\right). \tag{11}$$

Here '$\stackrel{\mathcal{D}}{=}$' denotes equality in distribution, λ_s are the eigenvalues of $\boldsymbol{T}\boldsymbol{V}_N\boldsymbol{T}$ in decreasing order, and $(C_s)_s$ is a sequence of independent χ_1^2-distributed random variables. Note, that the eigenvalues λ_s also depend on the dimension d and the sample sizes n_i. Transferring the results of [34] for the one-group design with $a = 1$ to our general setting, we obtain the subsequent asymptotic null distributions of the standardized quadratic form for all asymptotic settings (3)–(5).

Theorem 3.1. *Let* $\beta_s = \lambda_s \Big/ \sqrt{\sum_{\ell=1}^{ad} \lambda_\ell^2}$ *for* $s = 1, \ldots, ad$*. Then* $\widetilde{W}_N$ *has, under* $H_0(\boldsymbol{T})$*, and one of the frameworks* (3)–(5) *asymptotically*

a) a standard normal distribution if and only if

$$\beta_1 = \max_{s \leq ad} \beta_s \to 0 \qquad as \ \ N \to \infty,$$

b) a standardized $\left(\chi_1^2 - 1\right)/\sqrt{2}$ *distribution if and only if*

$$\beta_1 \to 1 \qquad as \ \ N \to \infty,$$

c) the same distribution as the random variable $\sum_{s=1}^{\infty} b_s \left(C_s - 1\right)/\sqrt{2}$*, if*

$$for\ all\ s \in \mathbb{N} \qquad \beta_s \to b_s \qquad as \quad N \to \infty,$$

for a decreasing sequence $(b_s)_s$ *in* $[0, 1]$ *with* $\sum_{s=1}^{\infty} b_s^2 = 1$.

It is worth to note that the influence of the different asymptotic frameworks is hidden in the corresponding conditions on the sequence of standardized eigenvalues $(\beta_s)_s$, which depend on both, a and d.

Moreover, for the specific one-group case with $a = 1$ the equivalent statements in a) and b) even complement the results of Pauly et al. [34] who only proved the sufficient part.

While Theorem 3.1 studies the asymptotic null distribution of $\widetilde{W}_N$, it is of additional interest to study its behaviour under local alternatives. To this end, we adopt two local situations already considered in Chen & Qin [11] for the case $a = 2$ and $H_0 = \boldsymbol{P}_2 \otimes \frac{1}{d}\boldsymbol{J}_d$ to our present design.

Theorem 3.2.

i) Under the local alternative $H_1(\boldsymbol{T}) : \boldsymbol{T}\boldsymbol{\mu} \neq \boldsymbol{0}_{ad}$ *it holds with* $N \cdot \boldsymbol{\mu}^\top \boldsymbol{T}\boldsymbol{V}_N\boldsymbol{T}\boldsymbol{\mu} \in \mathcal{o}\left(\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)\right)$ *that*

$$\widetilde{W}_N \stackrel{\mathcal{D}}{=} W_N(H_0) + \frac{N \cdot \boldsymbol{\mu}^\top \boldsymbol{T}\boldsymbol{\mu}}{\sqrt{2\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}} + \mathcal{o}_{\mathcal{P}}(1).$$

Here, $W_N(H_0)$ denotes a statistic that possesses the same distribution as $\widetilde{W}_N$ under H_0, i.e. $\mathcal{L}(W_N(H_0)) = \mathcal{L}(\widetilde{W}_N|H_0)$.

ii) Under the local alternative $H_1(\boldsymbol{T}) : \boldsymbol{T}\boldsymbol{\mu} \neq \boldsymbol{0}_{ad}$ it holds with $N \cdot \boldsymbol{\mu}^\top \boldsymbol{T}\boldsymbol{\mu} \in \mathcal{O}\left(\sqrt{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right)$ and $\beta_1 \to 0$, that

$$\widetilde{W}_N \overset{\mathcal{D}}{=} \sqrt{1 + 2N \frac{\boldsymbol{\mu}^\top \boldsymbol{T}\boldsymbol{V}_N \boldsymbol{T}\boldsymbol{\mu}}{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}} \cdot W_N(H_0) + \frac{N \cdot \boldsymbol{\mu}^\top \boldsymbol{T}\boldsymbol{\mu}}{\sqrt{2\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}} + o_{\mathcal{P}}(1).$$

Consulting the results of Theorems 3.1 and 3.2 it is easy to calculate asymptotic power functions of $\widetilde{W}_N$-tests. In particular, for $a = 2$, $H_0 = \boldsymbol{P} \otimes \frac{1}{d}\boldsymbol{J}_d$ and $\beta_1 \to 0$ we obtain the power functions stated in Chen & Qin [11]; noting that their asymptotic framework is contained in ours if $\beta_1 \to 0$.

Since the eigenvalues λ_s and standardized eigenvalues β_s are unknown in general we cannot apply the result directly. In particular, we are not even able to calculate the test statistic $\widetilde{W}_N$, not to mention to choose its correct limit distribution. To this end, we first introduce novel unbiased estimates of the unknown traces involved in (8)–(10) and discuss their mathematical properties. Plugging them into (8)–(10) leads to the calculation of adequately standardized test statistics. Finally, the choice of proper critical values is discussed in Section 4.

3.1. Symmetrized trace estimators

Here we derive unbiased and ratio-consistent estimates for the unknown traces $\operatorname{tr}(\boldsymbol{T}_S\boldsymbol{\Sigma}_i), \operatorname{tr}((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^2)$ and $\operatorname{tr}(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\boldsymbol{T}_S\boldsymbol{\Sigma}_r), i \neq r$, given in (8)–(10). Since it is not obvious that the usual plug-in estimates that are based on empirical covariance matrices are useful in high-dimensional settings we follow the approach of Brunner et al. [8] and Pauly et al. [34] and directly estimate the traces. Different to the one-sample design studied therein, we face the problem of additional nuisance parameters – the mean vectors $\boldsymbol{\mu}_i$. To avoid their estimation we adopt Tyler's symmetrization trick from M-estimates of scatter (see e.g. Croux et al. [13], Dümbgen [15] or Tyler et al. [39]) to the present situation, see also Brunner [7] and Harden [21]. In particular, we consider differences of observation pairs $(\ell_1, \ell_2), \ell_1 \neq \ell_2$, from the same group which fulfill $(\boldsymbol{X}_{i,\ell_1} - \boldsymbol{X}_{i,\ell_2}) \sim \mathcal{N}_d(\boldsymbol{0}_d, 2\boldsymbol{\Sigma}_i)$ and introduce the following novel estimators for $i = 1, \ldots, a$:

$$A_{i,1} = \frac{1}{2 \cdot \binom{n_i}{2}} \sum_{\substack{\ell_1,\ell_2=1 \\ \ell_1>\ell_2}}^{n_i} (\boldsymbol{X}_{i,\ell_1} - \boldsymbol{X}_{i,\ell_2})^\top \boldsymbol{T}_S (\boldsymbol{X}_{i,\ell_1} - \boldsymbol{X}_{i,\ell_2}), \tag{12}$$

$$A_{i,r,2} = \frac{1}{4 \cdot \binom{n_i}{2}\binom{n_r}{2}} \sum_{\substack{\ell_1,\ell_2=1 \\ \ell_1>\ell_2}}^{n_i} \sum_{\substack{k_1,k_2=1 \\ k_1>k_2}}^{n_r} \left[(\boldsymbol{X}_{i,\ell_1} - \boldsymbol{X}_{i,\ell_2})^\top \boldsymbol{T}_S (\boldsymbol{X}_{r,k_1} - \boldsymbol{X}_{r,k_2})\right]^2, \tag{13}$$

$$A_{i,3} = \frac{1}{24\binom{n_i}{4}} \sum_{\substack{\ell_1,\ell_2=1\\ \ell_1>\ell_2}}^{n_i} \sum_{\substack{k_2=1\\ k_2\neq \ell_1,\ell_2}}^{n_i-1} \sum_{\substack{k_1=k_2+1\\ \ell_2,\ell_1\neq k_1}}^{n_i} \Big[(\boldsymbol{X}_{i,\ell_1} - \boldsymbol{X}_{i,\ell_2})^\top \boldsymbol{T}_S (\boldsymbol{X}_{i,k_1} - \boldsymbol{X}_{i,k_2})\Big]^2, \tag{14}$$

$$A_4 = \sum_{i=1}^{a} \left(\frac{N}{n_i}\right)^2 (\boldsymbol{T}_W)_{ii}{}^2 A_{i,3} + 2\sum_{i=1}^{a-1} \sum_{r=i+1}^{a} \frac{N^2}{n_i n_r} (\boldsymbol{T}_W)_{ir}{}^2 A_{i,r,2}. \tag{15}$$

Here and throughout the paper expressions of the kind $a \neq b \neq c$ mean that the indices are pairwise different. In this sense all estimators (12)–(15) are *symmetrized U-statistics*, where the kernel is given by a specific quadratic or bilinear form. Their properties are analyzed below.

Lemma 3.3. *For any $\boldsymbol{\mu} \in \mathbb{R}^{ad}$ and $i \neq r = 1, \ldots, a$ it holds that*

1. *$\widehat{\mathbb{E}_{H_0}}(Q_N) := \sum_{i=1}^{a} \frac{N}{n_i} (\boldsymbol{T}_W)_{ii} A_{i,1}$ is an unbiased and ratio-consistent estimator for $\mathbb{E}_{H_0}(Q_N)$.*
2. *A_4 is an unbiased and ratio-consistent estimator for $\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)$.*
3. *$A_{i,1}, A_{i,r,2}$ and $A_{i,3}$ are unbiased and ratio-consistent estimators for $\operatorname{tr}(\boldsymbol{T}_S\boldsymbol{\Sigma}_i)$, $\operatorname{tr}(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\boldsymbol{T}_S\boldsymbol{\Sigma}_r)$ and $\operatorname{tr}\left((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^2\right)$, respectively.*

Remark 3.4. *(a) Recall that an $\mathbb{R}$-valued estimator $\widehat{\theta}_N$ is ratio-consistent for a sequence of real parameters θ_N if $\widehat{\theta}_N/\theta_N \to 1$ in probability as $N \to \infty$. Here the estimators and parameters may depend on $a = a(N)$ and/or $d = d(N)$.*

(b) Studying the proof of Lemma 3.3 given in the Appendix, we see that all these estimators are even (dimension-)stable in the sense of Brunner et al. [8], i.e. they fulfill $|\mathbb{E}(\widehat{\theta}_N/\theta_N - 1)| \leq b_N$ and $Var(\widehat{\theta}_N/\theta_N) \leq c_N$ for sequences $b_N, c_N \downarrow 0$ not depending on a and d.

It follows from Lemma 3.3 that

$$\widehat{Var_{H_0}}(Q_N) := 2\sum_{i=1}^{a} \left(\frac{N}{n_i}\right)^2 (\boldsymbol{T}_W)_{ii}{}^2 A_{i,3} + 4\sum_{i=1}^{a-1} \sum_{r=i+1}^{a} \frac{N^2}{n_i n_r} (\boldsymbol{T}_W)_{ir}{}^2 A_{i,r,2} = 2A_4$$

is an unbiased estimator of $Var_{H_0}(Q_N)$. This motivates to study the standardized quadratic form

$$W_N = \frac{Q_N - \widehat{\mathbb{E}_{H_0}}(Q_N)}{\widehat{Var_{H_0}}(Q_N)^{1/2}}$$

for testing $H_0(T)$. Its asymptotic behaviour (under $H_0(T) : \boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0}_{ad}$) is summarized below.

Theorem 3.5.

a) Under $H_0(T) : \boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0}_{ad}$ and one of the frameworks (3)–(5) the statistic W_N has the same asymptotic limit distribution as $\widetilde{W}_N$, if the respective conditions (a)–(c) from Theorem 3.1 are fulfilled.

b) Under the asymptotic frameworks (3)–(5) the statistic W_N has the same asymptotic limit distribution as $\widetilde{W}_N$, if the respective local alternative condition a) or b) from Theorem 3.2 is fulfilled.

The result shows that it is not reasonable to approximate the unknown distribution of the test statistic with a fixed distribution to obtain a valid test procedure. For example, choosing $z_{1-\alpha}$, the $(1-\alpha)$-quantile of the standard-normal distribution ($\alpha \in (0,1)$), as critical value would lead to a valid asymptotic level α test $\psi_z = \mathbf{1}\{W_N > z_{1-\alpha}\}$ in case of $\beta_1 \to 0$, i.e. $\mathbb{E}_{H_0}(\psi_z) \to \alpha$. However, for $\beta_1 \to 1$ we would obtain $\mathbb{E}_{H_0}(\psi_z) \to P(\chi_1^2 > \sqrt{2}z_{1-\alpha} + 1)$ which may lead to an asymptotically liberal ($\alpha = 0.01$ or 0.05) or conservative ($\alpha = 0.1$) test decision, see Table 1. Contrary, choosing $c_{1-\alpha} = (\chi^2_{1;1-\alpha} - 1)/\sqrt{2}$ as critical value (where $\chi^2_{1;1-\alpha}$ denotes the $(1-\alpha)$-quantile of the χ_1^2-distribution) for the test $\psi_\chi = \mathbf{1}\{W_N > c_{1-\alpha}\}$, it follows that $\mathbb{E}_{H_0}(\psi_\chi) \to \alpha$ if $\beta_1 \to 1$ but $\mathbb{E}_{H_0}(\psi_\chi) \to 1 - \Phi(c_{1-\alpha})$ for $\beta_1 \to 0$, where Φ denotes the cumulative distribution function of $\mathcal{N}(0,1)$. Again we obtain an asymptotically liberal ($\alpha = 0.1$) or extremely conservative ($\alpha = 0.05$ or 0.01) test decision, see the last column of Table 1.

Table 1
Asymptotic levels of the tests ψ_z and ψ_χ with fixed critical values under the null hypothesis and all asymptotic frameworks (3)–(5).

chosen	True asymptotic level of the test			
level α	ψ_z $(\beta_1 \to 0)$	ψ_z $(\beta_1 \to 1)$	ψ_χ $(\beta_1 \to 0)$	ψ_χ $(\beta_1 \to 1)$
0.10	0.10	0.09354	0.11391	0.10
0.05	0.05	0.06819	0.02226	0.05
0.01	0.01	0.03834	0.00003	0.01

Hence, an indicator (i.e. estimator) for whether $\beta_1 \to 0$, $\beta_1 \to 1$ or betwixt would be desirable. Nevertheless, even if the tests with fixed critical values are asymptotically correct (ψ_z in case of $\beta_1 \to 0$ or ψ_χ in case of $\beta_1 \to 1$), their true type-I error control may be poor for small sample sizes, see the simulations in Section 5.1.

Thus, in any case, it seems more appropriate to approximate W_N by a sequence of standardized distributions as already advocated in Pauly et al. [34] for the case of $a = 1$. We will propose such approximations in the next sections, where also a check criterion for $\beta_1 \to 0$ or $\beta_1 \to 1$ is presented.

4. Better approximations

To motivate the subsequent approximation, recall from (11) that $\widetilde{W}_N$ is of weighted χ_1^2-form. Following Zhang [44] it is reasonable to approximate statistics of this form by a standardized $(\chi_f^2 - 1)/\sqrt{2f}$-distribution, while f is selected such that the first three moments coincide. Straightforward calculations show that this is achieved by approximating with

$$K_{f_P} = \frac{\chi^2_{f_P} - f_P}{\sqrt{2f_P}} \quad \text{such that} \quad f_P = \frac{\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}, \tag{16}$$

where f_P is called the Pearson approximation. In case of $a = 1$ this simplifies to the method presented in Pauly et al. [34]. There it has already been seen that the approximation (16) performs much better for smaller sample sizes and/or dimensions than the above approaches with a fixed distribution. We will later rediscover this observation in Section 5 for our present design with general a. The next theorem gives a mathematical reason for this approximation.

Theorem 4.1. *Under the conditions of Lemma 3.1 and one of the frameworks (3)–(5) we have that K_{f_P} given in (16) has, under $H_0 : \boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0}_{ad}$, asymptotically*

a) a standard normal distribution if $\beta_1 \to 0$ as $N \to \infty$,
b) a standardized $\left(\chi^2_1 - 1\right)/\sqrt{2}$ distribution if $\beta_1 \to 1$ as $N \to \infty$.

Thus, compared to the approximation with a fixed limit distribution, the K_{f_P}-approach would at least be asymptotically correct whenever $\beta_1 \to \gamma \in \{0, 1\}$, while always providing a three moment approximation to the test statistic. To apply this result, an estimator for f in (16) is needed. Since we have already found A_4 as unbiased and ratio-consistent estimator for $\operatorname{tr}((\boldsymbol{T}\boldsymbol{V}_N)^2)$, it remains to find an adequate one for $\operatorname{tr}((\boldsymbol{T}\boldsymbol{V}_N)^3)$. A combination of both will then lead to a proper estimator for f_P and $\tau_P = f_P{}^{-1}$, respectively. Again we prefer a direct estimation of the involved traces. To this end, we introduce random vectors

$$\boldsymbol{Z}_{(\ell_1,\ell_2,\ldots,\ell_{2a})} := \left(\sqrt{\frac{N}{n_1}}\,(\boldsymbol{X}_{1,\ell_1} - \boldsymbol{X}_{1,\ell_2})^\top, \ldots, \sqrt{\frac{N}{n_a}}\,(\boldsymbol{X}_{a,\ell_{2a-1}} - \boldsymbol{X}_{a,\ell_{2a}})^\top\right)^\top$$

with $1 \le \ell_{2i-1} \ne \ell_{2i} \le n_i$ for all $i = 1\ldots, a$. Note, that this vectors are multivariate normally distributed with $\mathbb{E}(\boldsymbol{Z}_{(\ell_1,\ell_2,\ldots,\ell_{2a-1},\ell_{2a})}) = \boldsymbol{0}_{ad}$ and covariance matrix $Cov\left(\boldsymbol{Z}_{(\ell_1,\ell_2,\ldots,\ell_{2a-1},\ell_{2a})}\right) = 2\bigoplus_{i=1}^a \frac{N}{n_i}\boldsymbol{\Sigma}_i = 2\boldsymbol{V}_N$. Utilizing their particular form, it is shown in the Appendix, that a cyclic combination of these random vectors yields an unbiased estimator for $\operatorname{tr}((\boldsymbol{T}\boldsymbol{V}_N)^3)$. In particular, writing $\boldsymbol{Z}_{(\ell_1,\ell_2)}$ for $\boldsymbol{Z}_{(\ell_1,\ell_2,\ell_1,\ell_2,\ldots,\ell_1,\ell_2)}$ we have

$$\mathbb{E}\left(\boldsymbol{Z}_{(1,2)}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\boldsymbol{Z}_{(3,4)}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(5,6)}\boldsymbol{Z}_{(5,6)}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(1,2)}\right) = 8\operatorname{tr}((\boldsymbol{T}\boldsymbol{V}_N)^3). \tag{17}$$

This motivates the definition of (for $n_i \ge 6$)

$$C_5 = \sum_{\substack{\ell_{1,1},\ldots,\ell_{6,1}=1\\ \ell_{1,1}\ne\cdots\ne\ell_{6,1}}}^{n_1} \cdots \sum_{\substack{\ell_{1,a},\ldots,\ell_{6,a}=1\\ \ell_{1,a}\ne\cdots\ne\ell_{6,a}}}^{n_a} \frac{\prod_{m=1}^3 \Lambda_m(\ell_{1,1},\ldots,\ell_{6,a})}{8\cdot\prod_{i=1}^a \frac{n_i!}{(n_i-6)!}}, \tag{18}$$

where

$$\Lambda_1(\ell_{1,1},\ldots,\ell_{6,a}) = \boldsymbol{Z}_{(\ell_{1,1},\ell_{2,1},\ldots,\ell_{1,a},\ell_{2,a})}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(\ell_{3,1},\ell_{4,1},\ldots,\ell_{3,a},\ell_{4,a})},$$
$$\Lambda_2(\ell_{1,1},\ldots,\ell_{6,a}) = \boldsymbol{Z}_{(\ell_{3,1},\ell_{4,1},\ldots,\ell_{3,a},\ell_{4,a})}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(\ell_{5,1},\ell_{6,1},\ldots,\ell_{5,a},\ell_{6,a})},$$

$$\Lambda_3(\ell_{1,1},\ldots,\ell_{6,a}) = \boldsymbol{Z}_{(\ell_{5,1},\ell_{6,1},\ldots,\ell_{5,a},\ell_{6,a})}{}^\top \boldsymbol{T}\boldsymbol{Z}_{(\ell_{1,1},\ell_{2,1},\ldots,\ell_{1,a},\ell_{2,a})}.$$

Its properties together with a consistent estimator for f_P are summarized below.

Lemma 4.2. *(a) The estimator C_5 given in (18) is unbiased for $\operatorname{tr}((\boldsymbol{T}\boldsymbol{V}_N)^3)$.*

(b) Suppose that $a \in \mathbb{N}$ is fixed. Then $\widehat{\tau}_P := C_5^2/A_4^3$ is a consistent estimator for $\tau_P = 1/f_P$ as $\min(d, n_1, \ldots, n_a) \to \infty$, i.e. we have convergence in probability

$$\widehat{\tau}_P - \tau_P = \frac{C_5^2}{A_4^3} - \frac{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}{\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} \xrightarrow{\mathcal{P}} 0. \tag{19}$$

(c) Now suppose that $a \to \infty$ and that there exists some $q > 1$ which fulfills $\min(n_1, \ldots, n_a) = \mathcal{O}(a^q)$. Then (19) even holds under the asymptotic frameworks (4) - (5).

Theorem 4.3. *Suppose (19). Then, Theorem 4.1 remains valid if we replace f_P by its estimator $\widehat{f}_P = 1/\widehat{\tau}_P$.*

Remark 4.4. *(a) Using similar arguments as in the proof of Lemma 8.1. of Pauly et al. [34] we obtain the equivalences $\beta_1 \to 0 \Leftrightarrow \tau_P \to 0$ and $\beta_1 \to 1 \Leftrightarrow \tau_P \to 1$. Thus, $\widehat{\tau}_P$ can also be used as check criterion for these two cases.*

(b) It is also possible to derive a consistent estimator for $\tau_{CQ} = 1/f_{CQ} = \operatorname{tr}((\boldsymbol{T}\boldsymbol{V}_N)^4)/\operatorname{tr}^2((\boldsymbol{T}\boldsymbol{V}_N)^2)$, a key quantity in Chen & Qin [11], see the Appendix for details concerning the estimator. The corresponding approximation by the sequence $K_{f_{CQ}}$ even shares the same asymptotic properties of the Pearson approximation (16) stated in Theorem 4.1 and Theorem 4.3. However, it only provides a two moment approximation which turned out to perform worse in simulations (results not shown).

(c) In the Appendix, we additionally present an unbiased estimator C_7 for $\operatorname{tr}((\boldsymbol{T}\boldsymbol{V}_N)^3)$ such that C_7^2/A_4^3 is consistent for τ_P in all asymptotic frameworks (3) - (5). Particularly, the extra condition $\min(n_1, \ldots, n_a) = \mathcal{O}(a^q)$ is not needed. However, it is computationally more expensive compared to C_5 and thus omitted here.

In practical applications, the computation costs for C_5 are nevertheless rather high. This leads to disproportional waiting times for p-values of the corresponding approximate test $\varphi_N = \mathbf{1}\{W_N > K_{\widehat{f}_P;1-\alpha}\}$, where the critical value is given as $(1-\alpha)$-quantile of $K_{\widehat{f}_P}$. Therefore, we propose a certain subsampling-type method. Since the unbiasedness of C_5 clearly stems from (17), it seems reasonable to proceed as follows: For each $i = 1, \ldots, a$ and $b = 1, \ldots, B$ we independently draw random subsamples $\{\sigma_{1i}(b), \ldots, \sigma_{6i}(b)\}$ of length 6 from $\{1, \ldots, n_i\}$ and store them in a joint random vector $\boldsymbol{\sigma}(b) = (\sigma_{11}(b), \ldots, \sigma_{6a}(b))$. Then, a subsampling-version of the estimator C_5 is given by

$$C_5^\star = C_5^\star(B) = \frac{1}{8 \cdot B} \sum_{b=1}^{B} \Lambda_1(\boldsymbol{\sigma}(b)) \cdot \Lambda_2(\boldsymbol{\sigma}(b)) \cdot \Lambda_3(\boldsymbol{\sigma}(b)).$$

Letting $B = B(N) \to \infty$ as $N \to \infty$ it is easy to see (cf. the Appendix for details), that $C_5^\star$ has the same asymptotic properties as C_5. In particular, it is stated in the Appendix that $\widehat{\tau}_P^\star := 1/\widehat{f}_P^\star := C_5^{\star 2}/A_4^3$ is a consistent estimator for τ_P and that the approximation $K_{\hat{f}_P^\star}$ has the same weak limits as $K_{\hat{f}_P}$ stated in Theorem 4.3. This leads to $\varphi_N^\star = \mathbf{1}\{W_N > K_{\hat{f}_P^\star;1-\alpha}\}$ which is an asymptotically exact test whenever $\beta_1 \to \gamma \in \{0,1\}$. The finite sample, dimension and group size performance of this approximation are investigated in the subsequent section.

5. Simulations

In the previous sections, we considered the asymptotic properties of the proposed inference methods which are valid for large sample and fixed or possibly large dimension and/or group sizes. Here we investigate the small sample properties of our proposed approximation procedure $\varphi_N^\star = \mathbf{1}\{W_N > K_{\hat{f}_P^\star;1-\alpha}\}$ in comparison to the statistical tests $\psi_z = \mathbf{1}\{W_N > z_{1-\alpha}\}$ and $\psi_\chi = \mathbf{1}\{W_N > c_{1-\alpha}\}$ based on fixed critical values.

Furthermore, we consider versions of the Chen & Qin [11] test $\psi_{CQ} = \mathbf{1}\{T_{CQ}/\widehat{\sigma} > z_{1-\alpha}\}$ which was originally only developed for the high-dimensional two-sample mean comparison. Their procedure is based on the test statistic

$$T_{CQ} = \frac{\sum_{\ell_1 \neq \ell_2}^{n_1} \boldsymbol{X}_{1\ell_1}^\top \boldsymbol{X}_{1\ell_2}}{n_1(n_1-1)} + \frac{\sum_{k_1 \neq k_2}^{n_2} \boldsymbol{X}_{2k_1}^\top \boldsymbol{X}_{2k_2}}{n_2(n_2-1)} - 2\frac{\sum_{\ell=1}^{n_1}\sum_{k=1}^{n_2} \boldsymbol{X}_{1\ell}^\top \boldsymbol{X}_{2k}}{n_1 n_2},$$

and the variance estimator

$$\widehat{\sigma} = \frac{2}{n_1(n_1-1)}\widehat{\operatorname{tr}(\boldsymbol{\Sigma}_1^2)} + \frac{2}{n_2(n_2-1)}\widehat{\operatorname{tr}(\boldsymbol{\Sigma}_2^2)} + \frac{4}{n_1 n_2}\widehat{\operatorname{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)}$$

using

$$\widehat{\operatorname{tr}(\boldsymbol{\Sigma}_i^2)} = \frac{1}{n_i(n_i-1)} \cdot \operatorname{tr}\left(\sum_{j\neq k}^{n_i} (\boldsymbol{X}_{ij} - \overline{\boldsymbol{X}}_{i(j,k)})\boldsymbol{X}_{ij}^\top (\boldsymbol{X}_{ik} - \overline{\boldsymbol{X}}_{i(j,k)})\boldsymbol{X}_{ik}^\top\right),$$

$$\widehat{\operatorname{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2)} = \frac{1}{n_1 n_2} \cdot \operatorname{tr}\left(\sum_{\ell=1}^{n_1}\sum_{k=1}^{n_2} (\boldsymbol{X}_{1\ell} - \overline{\boldsymbol{X}}_{i(\ell)})\boldsymbol{X}_{1\ell}^\top (\boldsymbol{X}_{2k} - \overline{\boldsymbol{X}}_{2(k)})\boldsymbol{X}_{2k}^\top\right).$$

Here, $\overline{\boldsymbol{X}}_{i(j,k)}$ denotes the i-th sample mean after excluding $\boldsymbol{X}_{ij}$ and $\boldsymbol{X}_{ik}$, and $\boldsymbol{X}_{i(\ell)}$ is the i-th sample mean without $\boldsymbol{X}_{i\ell}$.

It is apparent, that ψ_{CQ} and ψ_z use the same critical z-value. In particular, Chen & Qin [11] have proven that ψ_{CQ} is asymptotically valid if $\beta_1 \to 0$, i.e. in the same situation as ψ_z. Its behaviour has, however, not been investigated in the case of $\beta_1 \nrightarrow 0$. As the enumerator T_{CQ} of the Chen-Qin test statistic is basically ours (with $\boldsymbol{T} = \boldsymbol{P}_2 \otimes \frac{1}{2}\boldsymbol{J}_d$) after subtracting the mixed terms $\sum_{\ell_1=1}^{n_1} \boldsymbol{X}_{1\ell_1}^\top \boldsymbol{X}_{1\ell_1}, \sum_{k_1=1}^{n_2} \boldsymbol{X}_{2k_1}^\top \boldsymbol{X}_{2k_1}$, the key difference is the choice of variance estimator. While ours is of symmetrized U-statistics-type, $\widehat{\sigma}$ is more of a jackknife-type estimator and it is of interest to see how both compare in our general setting.

In particular, we below compare all testing procedures in simulation studies with respect to

(a) their type-I error rate control under the null hypothesis (Section 5.1) and
(b) their power behaviour under various alternatives (Section 5.2).

All simulations were performed with the help of the R computing environment (R Development Core Team, 2013), each with $n_{sim} = 10^4$ simulation runs.

5.1. Asymptotic distribution and type-I error control

First, we study the speed of convergence, i.e. type-I error control, of the three different tests under the null hypothesis. To be in line with the simulation results presented in Pauly et al. [34] for the case $a = 1$ we also multiplied the statistic W_N by $\sqrt{N/(N-1)}$ to avoid a slightly liberal behaviour.

Due to the abundance of different split-plot designs and the more methodological focus of the paper, we restrict our simulation study to three specific null hypotheses and a high dimensional and heteroscedastic two-sample setting.

In particular, we investigate the type-I error behaviour of all four tests for the null hypotheses

- $H_0^a : \left(\boldsymbol{P}_2 \otimes \frac{1}{d}\boldsymbol{J}_d\right)\boldsymbol{\mu} = \boldsymbol{0}_{2d}$,
- $H_0^b : \left(\frac{1}{2}\boldsymbol{J}_2 \otimes \boldsymbol{P}_d\right)\boldsymbol{\mu} = \boldsymbol{0}_{2d}$ and
- $H_0^{ab} : \left(\boldsymbol{P}_a \otimes \boldsymbol{P}_d\right)\boldsymbol{\mu} = \boldsymbol{0}_{2d}$.

Since the Chen & Qin [11] test ψ_{CQ} is only applicable for H_0^a, we additionally translate their procedure to also test the other two hypotheses H_0^b and H_0^{ab}. This is possible by recognizing that $H_0^b : \left(\frac{1}{a}\boldsymbol{J}_a \otimes \boldsymbol{P}_d\right)\boldsymbol{\mu} = \boldsymbol{0}_{2\cdot d}$ can be written as $\mathbb{E}(\boldsymbol{P}_d\boldsymbol{X}_{11}) = \mathbb{E}(\boldsymbol{P}_d\boldsymbol{X}_{21})$ while $H_0^{ab} : \left(\boldsymbol{P}_a \otimes \boldsymbol{P}_d\right)\boldsymbol{\mu} = \boldsymbol{0}_{2\cdot d}$ can be expressed by $\mathbb{E}(\boldsymbol{P}_d\boldsymbol{X}_{11}) = -\mathbb{E}(\boldsymbol{P}_d\boldsymbol{X}_{21})$. Thus, carrying out ψ_{CQ} in the transformed vectors $\boldsymbol{Y}_{ik} = \boldsymbol{P}_d\boldsymbol{X}_{ik}$ (for H_0^b) and $\boldsymbol{Y}_{1k} = \boldsymbol{P}_d\boldsymbol{X}_{1k}$, $\boldsymbol{Y}_{2k} = \boldsymbol{P}_d\boldsymbol{X}_{2k}$ (for H_0^{ab}), $k = 1, \ldots, n_i$, $i = 1, 2$, respectively, allows us to also use their procedure for testing H_0^b and H_0^{ab}. The resulting test will again be denoted as ψ_{CQ}.

In all cases sample sizes were chosen from $n_1 \in \{10, 20, 50\}$ and $n_2 \in \{15, 30, 75\}$ combined with various choices of dimensions $d \in \{5, 10, 20, 40, 70, 100, 150, 200, 300, 450, 600, 800\}$. For the covariance matrices a heteroscedastic setting with autoregressive structures $(\boldsymbol{\Sigma}_1)_{i,j} = 0.6^{|i-j|}$ and $(\boldsymbol{\Sigma}_2)_{i,j} = 0.65^{|i-j|}$ was chosen and for each simulation run $B(N) = 500 \cdot N$, $N = n_1 + n_2$, subsamples were drawn.

Note that these settings imply $\beta_1 \to 1$ for H_0^a and $\beta_1 \to 0$ for H_0^b, H_0^{ab}, see the Appendix for details.

Thus, $\varphi_N^\star$ is asymptotically exact in both cases while ψ_χ and ψ_z possess the asymptotic behaviour given in Table 1. In particular, the z-test ψ_z should be rather liberal for testing for H_0^a and ψ_χ strongly conservative for H_0^b. All these theoretical findings can be recovered in our simulations: The results for H_0^a, displayed in Figure 1, show an inflated type-I error level control of ψ_z around 8% for smaller samples sizes ($N = 25$). For larger sample sizes ($N = 125$) it stabilizes

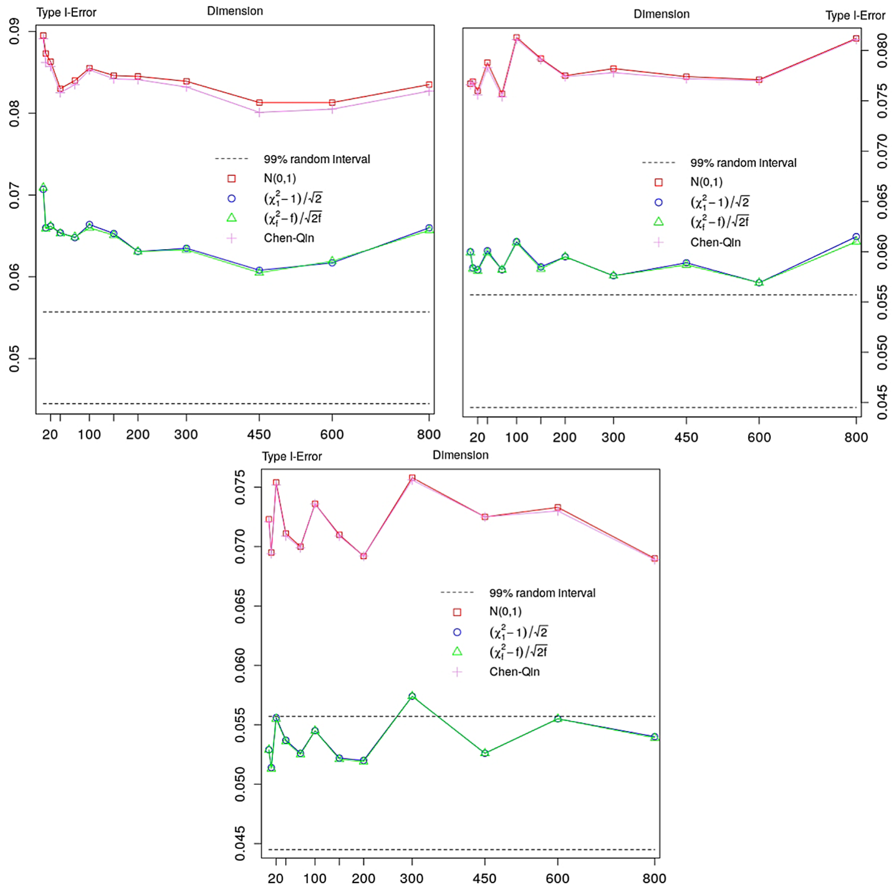


FIG 1. *Simulated type-I error rates ($\alpha = 5\%$) for the statistic $W_N \cdot \sqrt{N/(N-1)}$ compared with the critical values of a standard normal, standardized χ_1^2 and K_f-distribution and the test ψ_{CQ} of Chen & Qin under the null hypothesis $H_0^a : \left(\boldsymbol{P}_2 \otimes \frac{1}{d}\boldsymbol{J}_d\right)\boldsymbol{\mu} = \boldsymbol{0}$ for increasing dimension and covariance matrices $(\boldsymbol{\Sigma}_1)_{i,j} = 0.6^{|i-j|}$ and $(\boldsymbol{\Sigma}_2)_{i,j} = 0.65^{|i-j|}$. The sample sizes are increased from left ($n_1 = 10, n_2 = 15$) to right ($n_1 = 20, n_2 = 30$) to bottom ($n_1 = 50, n_2 = 75$).*

in the region of its asymptotic level of $7.2\% \pm 0.3\%$. The other z-test ψ_{CQ} leads to nearly the same results. For both tests, the error control is only slightly affected by the varying dimensions under investigation. In comparison, (in this situation) the two asymptotically correct tests $\varphi_N^\star$ and ψ_χ are slightly liberal for smaller sample sizes and more or less asymptotically correct for moderate ($N = 50$) to larger sample sizes. Here, it is astonishing that both procedures are nearly superposable, suggesting a fast convergence of the degrees of freedom estimator $\widehat{f_P}$.

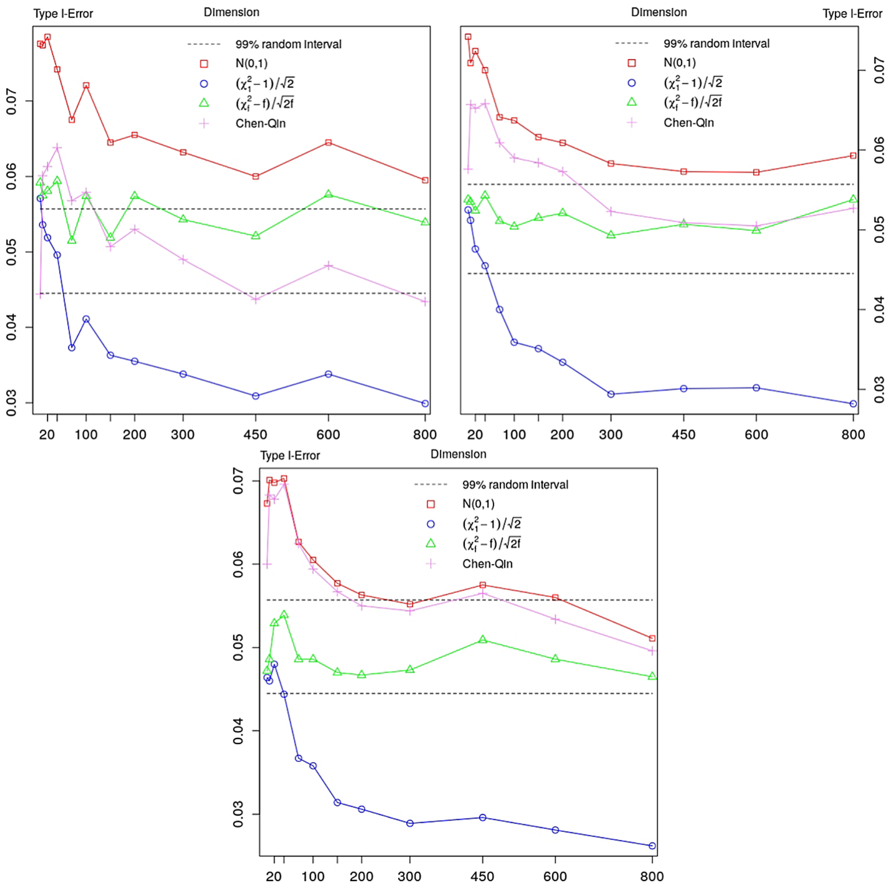


FIG 2. *Simulated type-I error rates ($\alpha = 5\%$) for the statistic $W_N \cdot \sqrt{N/(N-1)}$ compared with the critical values of a standard normal, standardized χ_1^2 and K_f-distribution and the test ψ_{CQ} of Chen & Qin under the null hypothesis $H_0^b : \left(\frac{1}{2}\boldsymbol{J}_2 \otimes \boldsymbol{P}_d\right)\boldsymbol{\mu} = \mathbf{0}$ for increasing dimension and covariance matrices $(\boldsymbol{\Sigma}_1)_{i,j} = 0.6^{|i-j|}$ and $(\boldsymbol{\Sigma}_2)_{i,j} = 0.65^{|i-j|}$. The sample sizes are increased from left ($n_1 = 10, n_2 = 15$) to right ($n_1 = 20, n_2 = 30$) to bottom ($n_1 = 50, n_2 = 75$).*

The results for H_0^b, presented in Figure 2, are slightly different. In particular, all the tests ψ_χ, ψ_z and ψ_{CQ} depending on fixed critical values are more affected by the underlying dimension: For smaller $d < 100$ the true level is considerably larger than their asymptotic level given in Table 1; resulting in a rather liberal behaviour of ψ_z and ψ_{CQ} and close to exact type-I error control for ψ_χ. This effect is decreased with increasing sample sizes with clear advantages for ψ_{CQ} over ψ_z. Moreover, for larger dimension ($d \geq 200$) all tests approach their asymptotic level. In comparison, the procedure $\varphi_N^\star$ based on the $K_{\hat{f}^\star}$ approximation shows a fairly good α level control through all dimension and sample size settings.

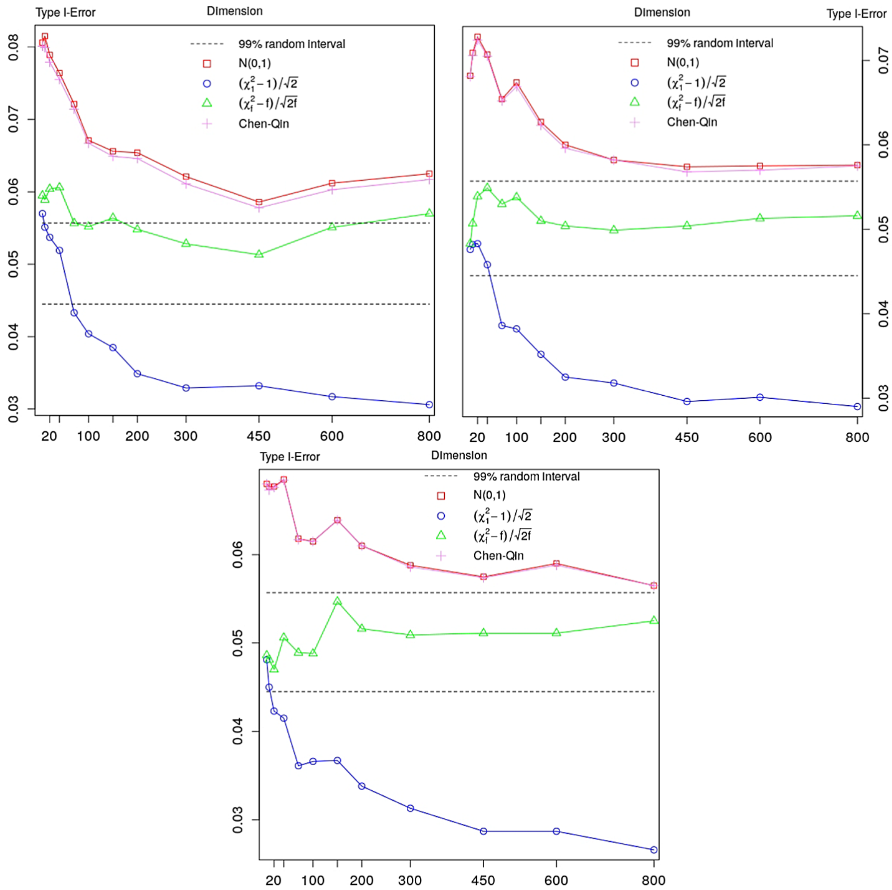


FIG 3. *Simulated type-I error rates ($\alpha = 5\%$) for the statistic $W_N \cdot \sqrt{N/(N-1)}$ compared with the critical values of a standard normal, standardized χ_1^2 and K_f-distribution and the test ψ_{CQ} of Chen & Qin, under the null hypothesis $H_0^{ab} : \left(\frac{1}{2}\boldsymbol{J}_2 \otimes \boldsymbol{P}_d\right)\boldsymbol{\mu} = \boldsymbol{0}$ for increasing dimension and covariance matrices $(\boldsymbol{\Sigma}_1)_{i,j} = 0.6^{|i-j|}$ and $(\boldsymbol{\Sigma}_2)_{i,j} = 0.65^{|i-j|}$. The sample sizes are increased from left ($n_1 = 10, n_2 = 15$) to right ($n_1 = 20, n_2 = 30$) to bottom ($n_1 = 50, n_2 = 75$).*

In case of the interaction hypothesis H_0^{ab} (Figure 3) similar observations can be made: The proposed approximation test $\varphi_N^\star$ controls the type-I error level fairly well over all settings while ψ_χ exhibits a rather conservative behaviour, particularly for increasing d. The behaviour of the two z-tests ψ_z and ψ_{CQ} is now almost equal: Both show a quite liberal behaviour for smaller dimensions d which decreases for larger d. To sum up, judging from Figures 1–3, $\varphi_N^\star$ seems to be the method of choice regardless of whether $\beta_1 \to 0$ or $\beta_1 \to 1$.

To also get an idea about the behaviour of all procedures in between those two cases we finally investigate a situation with $\beta_1 \to b_1 \notin \{0, 1\}$. To this

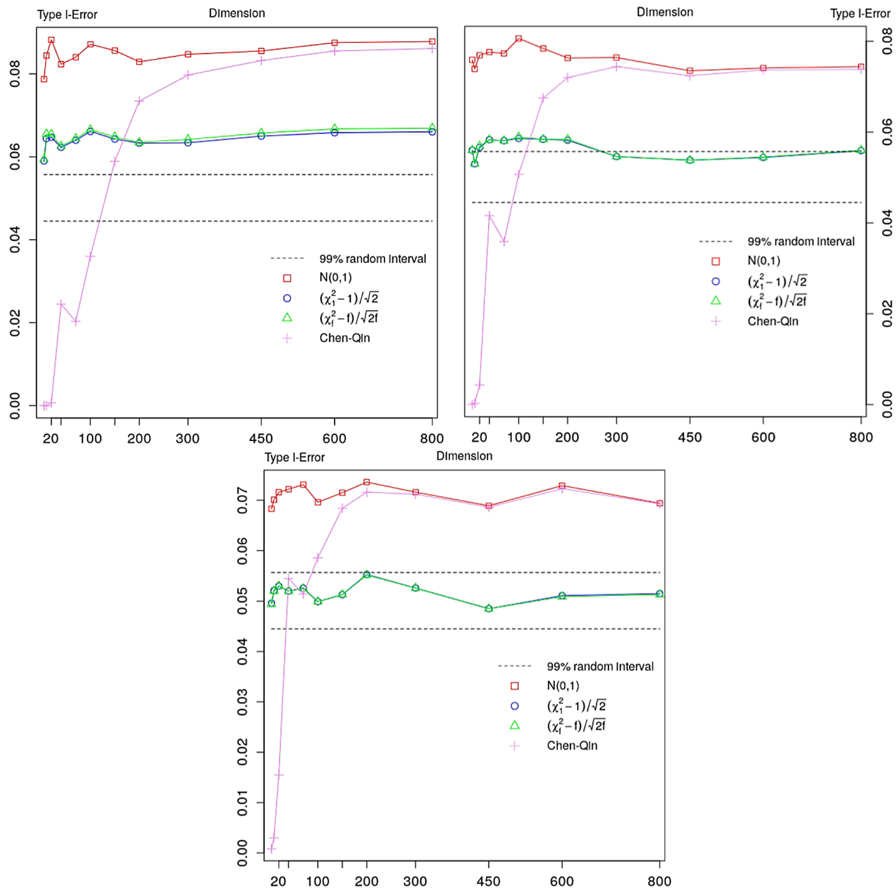


FIG 4. *Simulated type-I error rates ($\alpha = 5\%$) for the statistic $W_N \cdot \sqrt{N/(N-1)}$ compared with the critical values of a standard normal, standardized χ_1^2 and K_f-distribution and the test ψ_{CQ} of Chen & Qin, under the null hypothesis $H_0^b : \left(\frac{1}{2}\boldsymbol{J}_2 \otimes \boldsymbol{P}_d\right)\boldsymbol{\mu} = \boldsymbol{0}$ for increasing dimension and covariance matrices $(\boldsymbol{\Sigma}_1)_{i,j} = 0.6^{|i-j|/d}$ and $(\boldsymbol{\Sigma}_2)_{i,j} = 0.65^{|i-j|/d}$. The sample sizes are increased from left ($n_1 = 10, n_2 = 15$) to right ($n_1 = 20, n_2 = 30$) to bottom ($n_1 = 50, n_2 = 75$).*

end, we again test for the hypothesis H_0^b but now consider covariance matrices $(\boldsymbol{\Sigma}_1)_{i,j} = 0.6^{|i-j|/d}$ and $(\boldsymbol{\Sigma}_2)_{i,j} = 0.65^{|i-j|/d}$ for the two groups. Here, $b_1 \approx 0.76$, see Table 5 in the Appendix for details.

The simulation results are displayed in Figure 4. It is apparent that the behaviour of the two z-tests ψ_z and ψ_{CQ} is now considerably different for $d \leq 200$: While ψ_z behaves fairly liberal for all dimensions and sample size settings with error rates between 6.8% and 8.5% ($d \leq 50$), ψ_{CQ} is pretty conservative for smaller dimensions ($d \leq 100$) with error rates close to 0% ($d \leq 20$) and finally coincides with ψ_z for larger $d > 200$. This large differences for smaller

d may be explained by the different variance estimators involved in W_N and ψ_{CQ}. In contrast, $\varphi_N^\star$ and ψ_χ exhibit close to identical error rates for all choices of d and sample sizes. While both are slightly liberal for the smallest sample sizes the type-I error rate is close to the asymptotic level for $N = 50$ and even improves with increasing dimension and sample size. Because of this, we can also recommend $\varphi_N^\star$ in this situation.

5.2. Power performance

For ease of presentation and due to its favorable type-I error control we only examined the power of $\varphi_N^\star$ based on the test statistic W_N and estimated critical values from K_{f_P}.

Again a heteroscedastic two group split-plot design with autoregressive covariance structures ($(\boldsymbol{\Sigma}_1)_{i,j} = 0.6^{|i-j|}$ and $(\boldsymbol{\Sigma}_2)_{i,j} = 0.65^{|i-j|}$) was selected. The alpha level (5%) and the null hypotheses were restricted to $H_0^a : \left(\boldsymbol{P}_2 \otimes \frac{1}{d}\boldsymbol{J}_d\right)\boldsymbol{\mu} = \mathbf{0}$ and $H_0^b : \left(\frac{1}{2}\boldsymbol{J}_2 \otimes \boldsymbol{P}_d\right)\boldsymbol{\mu} = \mathbf{0}$. The investigated alternatives were

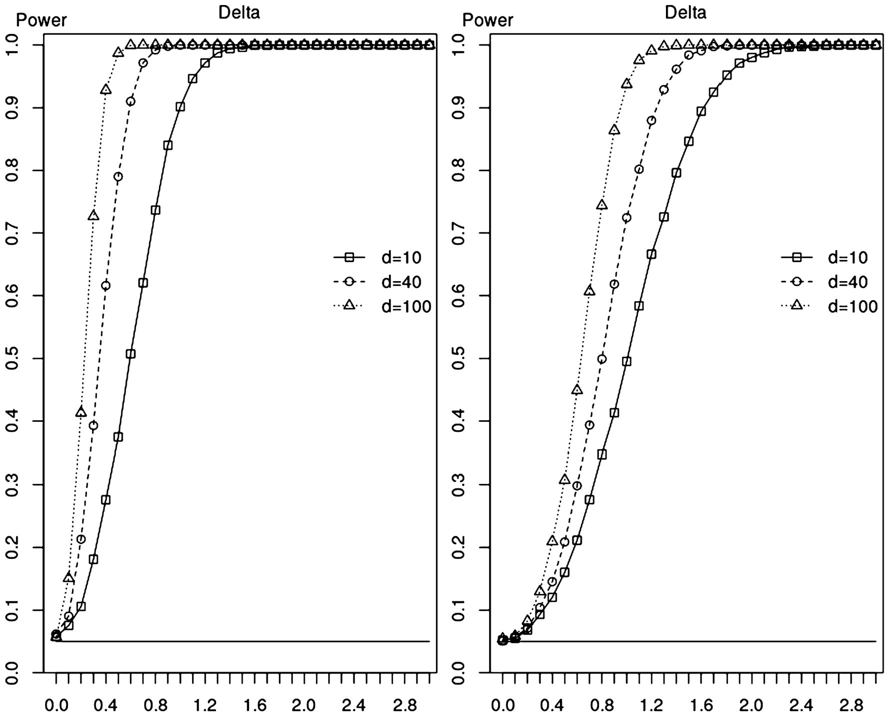
- a trend alternative for both hypotheses with $\boldsymbol{\mu}_2 = \mathbf{0}_d$ and $\mu_{1,t} = t\cdot\delta/d, 1 \le t \le d$ and additionally
- a shift alternative for H_0^a with $\boldsymbol{\mu}_2 = \mathbf{0}_d$ and $\boldsymbol{\mu}_1 = \mathbf{1}_d \cdot \delta$ and
- a one-point alternative for H_0^a and H_0^b, with $\boldsymbol{\mu}_2 = \mathbf{0}_d$ and $\boldsymbol{\mu}_1 = \boldsymbol{e}_1 \cdot \delta$,

each with increased $\delta \in [0, 3]$. Moreover, we only considered the moderate sample size setting with $n_1 = 20$ and $n_2 = 30$ together with three choices of dimensions $d = \{10, 40, 100\}$. Because of this sample sizes, a critical value based on f_P is chosen and the results can be found in Figures 5–7. It can be readily seen that the power depends on the type of alternative: For the trend (Figure 5) and the shift alternative (Figure 7) the power gets larger with increasing dimension. This is essentially apparent for the shift alternative, where the power increases considerably from $d = 10$ to $d = 40$. Contrary, for the one-point alternative the power becomes smaller for higher dimensions d (Figure 6). However, this is as expected since a difference in one single component can be detected more easily for smaller d.

Especially for testing H_0^a in the one-point alternative the power is poor even for $d = 10$. However this is completely in line with the result from Theorem 3.2: Calculating the corresponding values involved in the local alternative we get

- $\frac{N\cdot\boldsymbol{\mu}^\top\boldsymbol{T}\boldsymbol{T}\boldsymbol{\mu}}{\sqrt{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}} = \mathcal{O}\left(\frac{N}{d^2}\right)$ for H_0^a and
- $\frac{N\cdot\boldsymbol{\mu}^\top\boldsymbol{T}\boldsymbol{T}\boldsymbol{\mu}}{\sqrt{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}} = \mathcal{O}\left(\frac{N}{\sqrt{d}}\right)$ for H_0^b.

This explains the power decrease with increasing dimension which is more pronounced when testing H_0^a in comparison to testing for H_0^b.



Power curves for a trend alternative and $\mathrm{H}_0^a : \left(\mathbf{P}_2 \otimes \frac{1}{\mathrm{d}}\mathbf{J}_\mathrm{d}\right) \boldsymbol{\mu} = \mathbf{0}$. Power curves for a trend alternative and $\mathrm{H}_0^b : \left(\frac{1}{2}\mathbf{J}_2 \otimes \mathbf{P}_\mathrm{d}\right) \boldsymbol{\mu} = \mathbf{0}$.

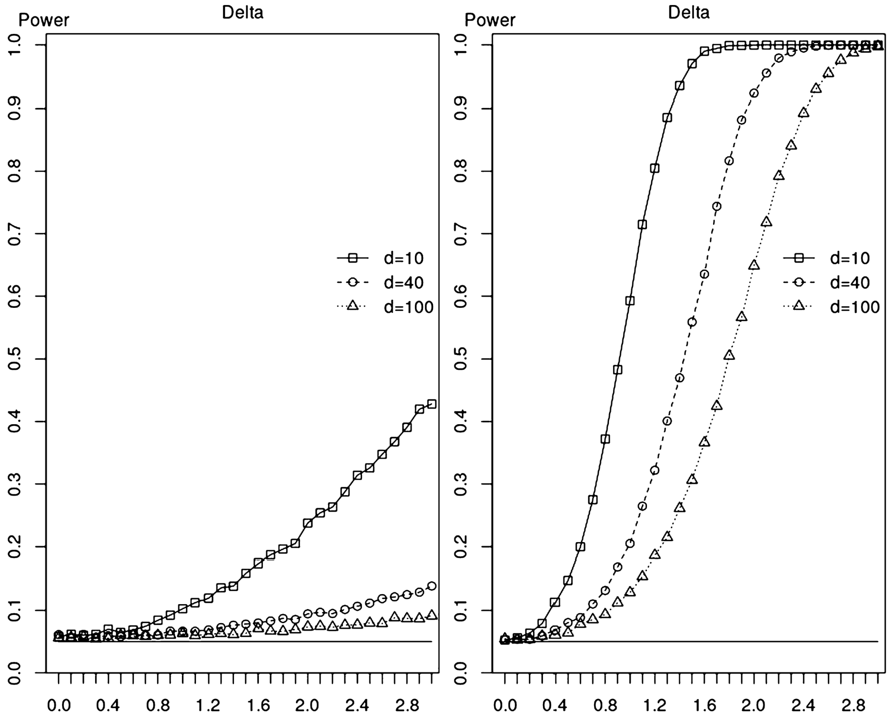
FIG 5. *Simulated power curves for the statistic* $W_N \cdot \sqrt{(N-1)/N}$ *in* 10^4 *simulation runs for different dimensions with* $n_1 = 20, n_2 = 30$ *and an autoregressive structure(*$(\boldsymbol{\Sigma}_1)_{i,j} = 0.6^{|i-j|}$ *and* $(\boldsymbol{\Sigma}_2)_{i,j} = 0.65^{|i-j|}$*).*

6. Analysis of a sleep laboratory data set

Finally, the new methods are exemplified on the sleep laboratory trial reported in Jordan et al. [25]. In this two-armed repeated measures trial, the activity of prostaglandin-D-synthase (β-trace) was measured every 4 hours over a period of 4 days. The grouping factor was gender and the above $d = 24$ repeated measures were observed on $n_i = 10$ young healthy women (group $i = 1$) and men (group $i = 2$). Since each day presented a certain sleep condition the repeated measures are structured by two crossed fixed factors:

- intervention (with 4 levels: normal sleep, sleep deprivation, recovery sleep and REM sleep deprivation) and
- time (with the 6 levels/time points $24h, 4h, 8h, 12h, 16h$ and $20h$).

Due to $d > n_i$ we are thus dealing with a high-dimensional split-plot design with $a = 2$ groups and $d = 24$ repeated measurements. The time profiles of each



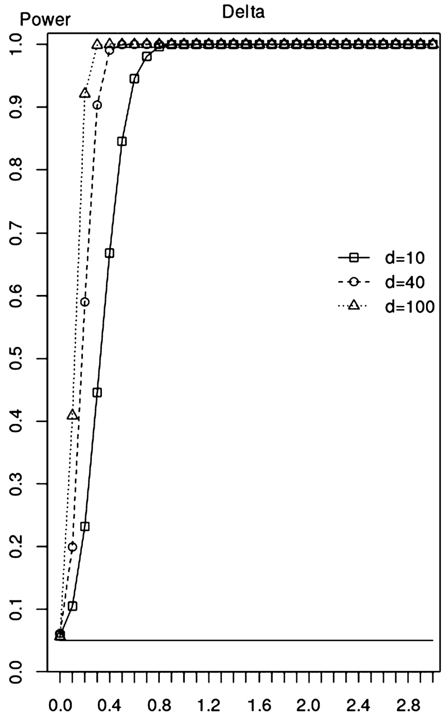
Power curves for an one-point alternative $\mathrm{H}_0^{\mathrm{a}} : \left(\mathbf{P}_2 \otimes \frac{1}{\mathrm{d}}\mathbf{J}_{\mathrm{d}}\right)\boldsymbol{\mu} = \mathbf{0}$.

Power curves for an one-point alternative and $\mathrm{H}_0^{\mathrm{b}} : \left(\frac{1}{2}\mathbf{J}_2 \otimes \mathbf{P}_{\mathrm{d}}\right)\boldsymbol{\mu} = \mathbf{0}$.

FIG 6. *Simulated power curves for the statistic* $W_N \cdot \sqrt{(N-1)/N}$ *in* 10^4 *simulation runs for different dimensions with* $n_1 = 20, n_2 = 30$ *and an autoregressive structure*$((\boldsymbol{\Sigma}_1)_{i,j} = 0.6^{|i-j|}$ *and* $(\boldsymbol{\Sigma}_2)_{i,j} = 0.65^{|i-j|}$*).*

subject are displayed in Figure 8 (for the female group 1) and Figure 9 (for the male group 2). We note, that group-specific profile analysis could already be performed by the methods given in Pauly et al. [34]. In particular, they found a significant intervention and a borderline time effect for the male group. For the current two-sample design additional questions concern (1) whether there is a gender effect, i.e. the time profiles of the groups differ, and if so (2) whether they differ with respect to certain interventions. Moreover, investigations regarding (3) a general effect of time and (4) interactions between the different factors are of equal interest. Utilizing the notation from Section 2, the corresponding null hypotheses can be formalized via adequate contrast matrices. In particular, we are interested in testing the null hypotheses

(a) No gender effect: $H_0^a : \left(\boldsymbol{P}_2 \otimes \frac{1}{24}\boldsymbol{J}_{24}\right)\boldsymbol{\mu} = \mathbf{0}$,
(b) No time effect: $H_0^b : \left(\frac{1}{2}\boldsymbol{J}_2 \otimes \boldsymbol{P}_{24}\right)\boldsymbol{\mu} = \mathbf{0}$,
(c) No interaction effect between time and group: $H_0^{ab} : (\boldsymbol{P}_2 \otimes \boldsymbol{P}_{24})\,\boldsymbol{\mu} = \mathbf{0}$,



Power curves for a shift alternative and $H_0^a : \left(\mathbf{P}_2 \otimes \frac{1}{d}\mathbf{J}_d\right) \boldsymbol{\mu} = \mathbf{0}$.

FIG 7. *Simulated power curves for the statistic* $W_N \cdot \sqrt{(N-1)/N}$ *in* 10^4 *simulation runs for different dimensions with* $n_1 = 20, n_2 = 30$ *and an autoregressive structure*$((\boldsymbol{\Sigma}_1)_{i,j} = 0.6^{|i-j|}$ *and* $(\boldsymbol{\Sigma}_2)_{i,j} = 0.65^{|i-j|}$*)*.

(d) No time effect for intervention ℓ, $\ell \in \{1, \ldots, 4\}$:
$H_0^{t\ell} : \left(\boldsymbol{P}_2 \otimes \left(\left(\boldsymbol{e}_l \cdot \boldsymbol{e}_l^\top\right) \otimes \boldsymbol{P}_6\right)\right) \boldsymbol{\mu} = \mathbf{0}$,

(e) No effect between interventions ℓ and k, $\ell \neq k \in \{1, \ldots, 4\}$:
$H_0^{\ell \times k} : \left(\boldsymbol{P}_2 \otimes \left(\left(\boldsymbol{e}_\ell \cdot \boldsymbol{e}_\ell^\top - \boldsymbol{e}_\ell \cdot \boldsymbol{e}_k^\top\right) \otimes \frac{1}{6}\boldsymbol{J}_6\right)\right) \boldsymbol{\mu} = \mathbf{0}$,

where $\boldsymbol{e}_\ell$ denotes the $\ell - th$ d-dimensional unit vector with all entries zero but the ℓ-th one. Applying the test $\varphi_N^\star$ based on the standardized quadratic form W_N as test statistic and the proposed $K_{\hat{f}_P^\star}$-approximation with $B = 50000 \cdot N = 100,000$ subsamples we obtain the results summarized in Table 2.

There it can be readily seen that most hypotheses cannot be rejected at level $\alpha = 5\%$. In particular, there is no evidence for an overall gender effect, so that we have not performed post-hoc analyses on the interventions. Only a highly significant time effect, as well as a significant effect between the first two interventions (normal sleep and sleep deprivation), could be detected. However, applying a multiplicity adjustment (Bonferroni or Holm) only the time effect remained significant.

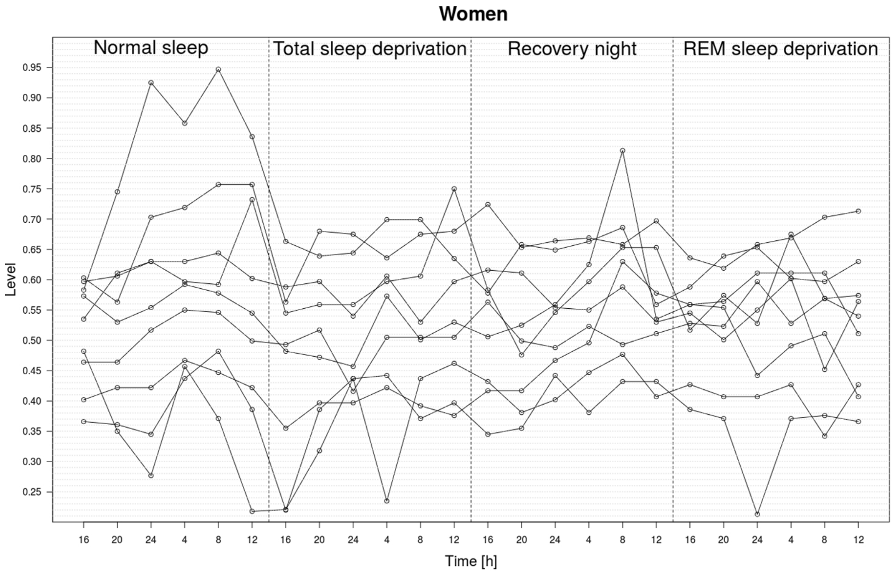


Fig 8. *Prostaglandin-D-synthase (ß-trace) of 10 young women during 4 days under different sleep conditions.*

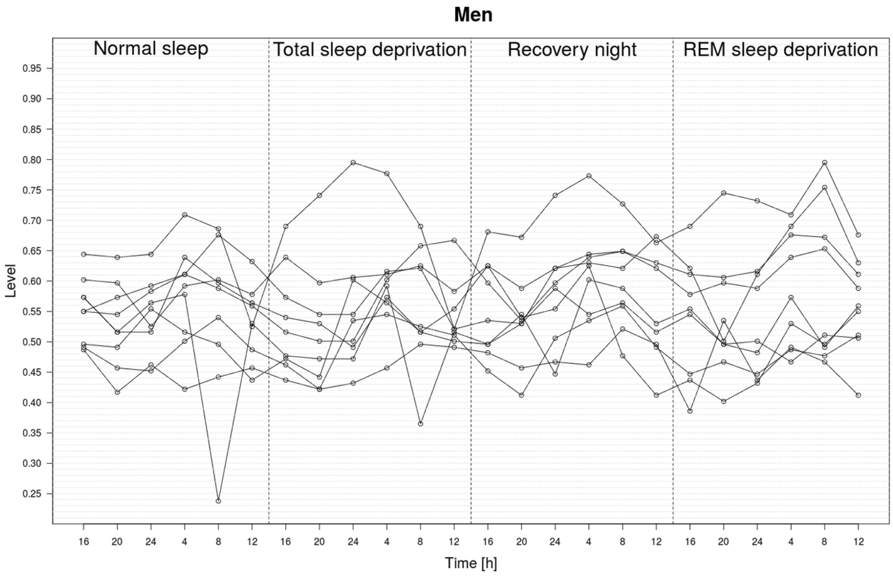


Fig 9. *Prostaglandin-D-synthase (ß-trace) of 10 young men during 4 days under different sleep conditions.*

TABLE 2
Analysis of the sleep lab trial from Figures 8–9: Shown are the values of the test statistic W_N and the estimator $\hat{f}_P^\star$ as well as the p-values of the test $\varphi_N^\star = \mathbf{1}\{W_N > K_{\hat{f}_P^\star;1-\alpha}\}$ for different null hypotheses of interest.

Hypothesis	W_N^A	$\hat{f}_P^\star$	p-value
H_0^a	-0.45671	1.19030	0.55832
H_0^b	6.24114	7.07832	**0.00008**
H_0^{ab}	0.74578	7.21217	0.20120
H_0^{t1}	-0.795083	461.874	0.784463
H_0^{t2}	-0.591851	360.048	0.71764
H_0^{t3}	-0.43381	223.24000	0.65845
H_0^{t4}	-1.18382	426.083	0.88385
$H_0^{1\times 2}$	2.37921	155.89025	**0.01285**
$H_0^{1\times 3}$	0.23757	156.64141	0.39240
$H_0^{1\times 4}$	–0.49984	143.57718	0.68099
$H_0^{2\times 3}$	-0.72716	91.83337	0.75968
$H_0^{2\times 4}$	-0.56510	79.78169	0.70183
$H_0^{3\times 4}$	-0.66704	130.56430	0.74046

7. Conclusion & outlook

In this paper we have investigated inference procedures for general split-plot models, allowing for unbalanced and/or heteroscedastic covariance settings as well as a factorial structure on the whole- and sub-plot factors. Inspired by the work of Pauly et al. [34] for one group repeated measures designs the test statistics were based on standardized quadratic forms. However, different to their work novel symmetrized U-statistics were introduced to adequately handle the problem of additional nuisance parameters in the multiple sample case.

To jointly cover low and highdimensional models as well as situations with a small or large number of groups, we conducted an in-depth study of their asymptotic behaviour under a unified asymptotic framework. In particular, the number of groups a and dimensions d may be fixed as in classical asymptotic settings, or even converge to infinity. Here we do neither postulate any assumptions on how d and/or a and the underlying sample sizes converge to infinity nor any sparsity conditions on the covariance structures since such assumptions are usually hard to check for a practical data set at hand. As a consequence, it turned out that the test statistic possess a whole continuum of asymptotic limits that depends on the eigenvalues of the underlying covariances. We thus argued that an approximation by a fixed critical value is not adequate and proposed an approximation by a sequence of standardized χ^2-distributions with estimated degrees of freedom. For computational efficiency, we additionally provided a subsampling-type version of the degrees of freedom estimator. Our approach provides a reasonably good three-moment approximation of the test statistic and is even asymptotically exact if the influence of the largest eigenvalue is negligible (leading to a standard normal limit) or decisive (leading to a standardized χ_1^2 limit).

Apart from these asymptotic considerations, we evaluated the finite sample and dimension performance of our approximation technique. In particular, for varying combinations of sample sizes and dimensions, we compared its power and type-I error control with test procedures based on fixed critical values. In all designs it showed a quite accurate error control over all low- ($d \leq 10$) to highdimensional situations (with up to $d = 800$). In comparison, its performance was considerably better than that of the other tests which partially disclosed a rather liberal or conservative behaviour.

In future research, we like to extend the current results to general highdimensional MANOVA designs, where we also like to relax the involved assumption of multivariate normality and/or even test simultaneously for mean and covariance effects as recently proposed in Liu et al. [31]. These investigations, however, require completely different (e.g., martingale) techniques and estimators of the involved traces. Moreover, we also plan to conduct more detailed simulations (especially for larger group sizes a and other covariance matrices) in a more applied paper.

Acknowledgement

The authors would like to thank Edgar Brunner for helpful discussions. This work was supported by the German Research Foundation project DFG-PA 2409/4-1.

Appendix A: Basics

In Section 2 of the main paper we claimed that the unique projection matrix $\boldsymbol{T}$ which describes the equivalent null hypotheses as $\boldsymbol{H} = \boldsymbol{H}_S \otimes \boldsymbol{H}_W$ is given by the product of two projection matrices $\boldsymbol{T}_S \otimes \boldsymbol{T}_W$. We start with the proof of this claim:

Lemma A.1. *Let be* $\boldsymbol{H} = \boldsymbol{H}_W \otimes \boldsymbol{H}_S$ *with* $\boldsymbol{H} \in \mathbb{R}^{ad\times ad}, \boldsymbol{H}_W \in \mathbb{R}^{a\times a}, \boldsymbol{H}_S \in \mathbb{R}^{d\times d}$. *For each hypothesis* $\boldsymbol{H}\boldsymbol{\mu} = \boldsymbol{0}_{ad}$ *with such a matrix* $\boldsymbol{H}$ *exist projectors* $\boldsymbol{T} \in \mathbb{R}^{ad\times ad}, \boldsymbol{T}_W \in \mathbb{R}^{a\times a}, \boldsymbol{T}_S \in \mathbb{R}^{d\times d}$ *which can be used to formulate the same null hypothesis* $\boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0}_{ad}$ *with* $\boldsymbol{T} = \boldsymbol{T}_W \otimes \boldsymbol{T}_S$.

Proof. It is known that the projector $\boldsymbol{T} = \boldsymbol{H}^\top[\boldsymbol{H}\boldsymbol{H}^\top]^-\boldsymbol{H}$ fulfills $\boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0}_{ad} \iff \boldsymbol{H}\boldsymbol{\mu} = \boldsymbol{0}_{ad}$. For this reason and utilizing well known rules (see for example Rao & Mitra [36]) for generalized inverses we obtain

$$\begin{aligned}
\boldsymbol{T} &= \boldsymbol{H}^\top[\boldsymbol{H}\boldsymbol{H}^\top]^-\boldsymbol{H} \\
&= (\boldsymbol{H}_W \otimes \boldsymbol{H}_S)^\top[(\boldsymbol{H}_W \otimes \boldsymbol{H}_S)(\boldsymbol{H}_W \otimes \boldsymbol{H}_S)^\top]^-(\boldsymbol{H}_W \otimes \boldsymbol{H}_S) \\
&= (\boldsymbol{H}_W^\top \otimes \boldsymbol{H}_S^\top)[(\boldsymbol{H}_W \otimes \boldsymbol{H}_S)(\boldsymbol{H}_W^\top \otimes \boldsymbol{H}_S^\top)]^-(\boldsymbol{H}_W \otimes \boldsymbol{H}_S) \\
&= (\boldsymbol{H}_W^\top \otimes \boldsymbol{H}_S^\top)[(\boldsymbol{H}_W\boldsymbol{H}_W^\top) \otimes (\boldsymbol{H}_S\boldsymbol{H}_S^\top)]^-(\boldsymbol{H}_W \otimes \boldsymbol{H}_S) \\
&= (\boldsymbol{H}_W^\top \otimes \boldsymbol{H}_S^\top)([\boldsymbol{H}_W\boldsymbol{H}_W^\top)]^- \otimes [\boldsymbol{H}_S\boldsymbol{H}_S^\top]^-)(\boldsymbol{H}_W \otimes \boldsymbol{H}_S) \\
&= (\boldsymbol{H}_W^\top \otimes \boldsymbol{H}_S^\top)([\boldsymbol{H}_W\boldsymbol{H}_W^\top)]^-\boldsymbol{H}_W \otimes [\boldsymbol{H}_S\boldsymbol{H}_S^\top]^-\boldsymbol{H}_S) \\
&= \boldsymbol{H}_W^\top[\boldsymbol{H}_W\boldsymbol{H}_W^\top]^-\boldsymbol{H}_W \otimes \boldsymbol{H}_S^\top[\boldsymbol{H}_S\boldsymbol{H}_S^\top]^-\boldsymbol{H}_S
\end{aligned}$$

$$= \boldsymbol{T}_W \otimes \boldsymbol{T}_S.$$

Thus, $\boldsymbol{T}_W := \boldsymbol{H}_W^\top[\boldsymbol{H}_W\boldsymbol{H}_W^\top]^-\boldsymbol{H}_W$ and $\boldsymbol{T}_S := \boldsymbol{H}_S^\top[\boldsymbol{H}_S\boldsymbol{H}_S^\top]^-\boldsymbol{H}_S$ are projectors, i.e. idempotent and symmetric. □

For proofing our main results we have to compare various traces of powers of combinations underlying covariance matrices. To this end, we will particularly apply the following inequalities:

Lemma A.2. *For positive real numbers a,b and a symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ it holds*

$$\operatorname{tr}^2\left(\boldsymbol{A}^{a+b}\right) \leq \operatorname{tr}\left(\boldsymbol{A}^{2a}\right)\operatorname{tr}\left(\boldsymbol{A}^{2b}\right).$$

For $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ symmetric with eigenvalues $\lambda_1, \ldots, \lambda_d \geq 0$ it holds that

$$\operatorname{tr}\left(\boldsymbol{A}^2\right) \leq \operatorname{tr}^2\left(\boldsymbol{A}\right).$$

If $\boldsymbol{\Sigma}_i \in \mathbb{R}^{d\times d}$ is positive definite and symmetric and $\boldsymbol{T} \in \mathbb{R}^{d\times d}$ is idempotent and symmetric it holds for every $k \in \mathbb{N}$ that

$$\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{\Sigma}_i)^{2k}\right) \leq \operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{\Sigma}_i)^{k}\right).$$

Proof. The first part is an application of the Cauchy–Bunyakovsky–Schwarz inequality, with the Frobenius inner product. Therefore

$$\begin{aligned}
\operatorname{tr}^2\left(\boldsymbol{A}^{a+b}\right) &= \operatorname{tr}^2\left(\boldsymbol{A}^a\boldsymbol{A}^b\right) = \operatorname{tr}^2\left(\boldsymbol{A}^a\boldsymbol{A}^{b^\top}\right) \\
&\leq \left(\sqrt{\operatorname{tr}\left(\boldsymbol{A}^a\boldsymbol{A}^{a^\top}\right)}\cdot\sqrt{\operatorname{tr}\left(\boldsymbol{A}^b\boldsymbol{A}^{b^\top}\right)}\right)^2 = \operatorname{tr}\left(\boldsymbol{A}^a\boldsymbol{A}^a\right)\cdot\operatorname{tr}\left(\boldsymbol{A}^b\boldsymbol{A}^b\right) \\
&= \operatorname{tr}\left(\boldsymbol{A}^{2a}\right)\operatorname{tr}\left(\boldsymbol{A}^{2b}\right).
\end{aligned}$$

The second part just uses the binomial theorem together with the condition $\lambda_t \geq 0$ for $t = 1, \ldots, d$:

$$\operatorname{tr}(\boldsymbol{A}^2) = \sum_{t=1}^{d}\lambda_t^2 \leq \sum_{t_1=1}^{d}\lambda_{t_1}^2 + \sum_{t_1=1}^{d}\sum_{t_2=1,t_2\neq t_1}^{d}\lambda_{t_1}\lambda_{t_2} = \left(\sum_{t=1}^{d}\lambda_t\right)^2 = \operatorname{tr}^2(\boldsymbol{A}).$$

Finally, the last inequality follows from the second one, if we show that all conditions are fulfilled. With idempotence of $\boldsymbol{T}$ and invariance of the trace under cyclic permutations, it follows for all $k \in \mathbb{N}$ that

$$\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{\Sigma}_i)^{2k}\right) = \operatorname{tr}\left(\boldsymbol{T}^2\boldsymbol{\Sigma}_i\cdot\cdots\cdot\boldsymbol{T}^2\boldsymbol{\Sigma}_i\right) = \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T})^{2k}\right).$$

Thus, it is sufficient to consider this term. Since $\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}$ is symmetric all powers are symmetric too and it follows with $k' = \lfloor k/2 \rfloor$ that

$$\forall \boldsymbol{x} \in \mathbb{R}^d : \quad \boldsymbol{x}^\top \left(\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\right)^k \boldsymbol{x} = \boldsymbol{x}^\top \left(\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\right)^{k'} \boldsymbol{T}\boldsymbol{\Sigma}_i^{k-2k'}\boldsymbol{T} \left(\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\right)^{k'} \boldsymbol{x}$$
$$= \left[\boldsymbol{T}\left(\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\right)^{k'} \boldsymbol{x}\right]^\top \boldsymbol{\Sigma}_i^{k-2k'} \left[\boldsymbol{T}\left(\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\right)^{k'} \boldsymbol{x}\right] \geq 0$$

since $\boldsymbol{\Sigma}_i$ and $\boldsymbol{I}_d$ are positive definite and $k - 2k' \in \{0, 1\}$. So both conditions of the second inequation are shown and

$$\operatorname{tr}\left(\left(\boldsymbol{T}\boldsymbol{\Sigma}_i\right)^{2k}\right) = \operatorname{tr}\left(\left[\left(\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\right)^k\right]^2\right) \leq \operatorname{tr}^2\left(\left(\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\right)^k\right) = \operatorname{tr}^2\left(\left(\boldsymbol{T}\boldsymbol{\Sigma}_i\right)^k\right). \quad \square$$

Furthermore, an inequality for traces which contain $\boldsymbol{\Sigma}_i$ and $\boldsymbol{\Sigma}_r$ is needed.

Lemma A.3. *Let $\boldsymbol{\Sigma}_i, \boldsymbol{\Sigma}_r \in \mathbb{R}^{d\times d}$ be positive definite and symmetric matrices and suppose that $\boldsymbol{T} \in \mathbb{R}^{d\times d}$ is idempotent and symmetric. Then it holds for $i \neq r$ that*

$$\operatorname{tr}\left(\left(\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\boldsymbol{\Sigma}_r\right)^2\right) \leq \operatorname{tr}^2\left(\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\boldsymbol{\Sigma}_r\right).$$

Proof. As shown before $\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}$ and $\boldsymbol{T}\boldsymbol{\Sigma}_r\boldsymbol{T}$ are symmetric and positive semidefinite. For this reason, it exists a symmetric matrix $\boldsymbol{W}$ with $\boldsymbol{W}\boldsymbol{W} = \boldsymbol{T}\boldsymbol{\Sigma}_r\boldsymbol{T}$. Due to the fact that all matrices are symmetric, it holds

$$\left(\boldsymbol{W}\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\boldsymbol{W}\right)^\top = \boldsymbol{W}^\top\boldsymbol{T}^\top\boldsymbol{\Sigma}_i^\top\boldsymbol{T}^\top\boldsymbol{W}^\top = \boldsymbol{W}\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\boldsymbol{W}$$

and because $\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}$ is positive semidefinite also

$$\forall \boldsymbol{x} \in \mathbb{R}^d \qquad \boldsymbol{x}^\top\boldsymbol{W}\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\boldsymbol{W}\boldsymbol{x} = (\boldsymbol{W}\boldsymbol{x})^\top\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}(\boldsymbol{W}\boldsymbol{x}) = \boldsymbol{y}^\top\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\boldsymbol{y} \geq 0.$$

This allows to use the inequalities from above for this matrix, and again utilizing the invariance of the trace under cyclic permutations we obtain

$$\begin{aligned}
&\operatorname{tr}\left(\left(\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\boldsymbol{\Sigma}_r\right)^2\right) \\
&\quad= \operatorname{tr}\left(\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\boldsymbol{T}\boldsymbol{\Sigma}_r\boldsymbol{T}\cdot\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\boldsymbol{T}\boldsymbol{\Sigma}_r\boldsymbol{T}\right) = \operatorname{tr}\left(\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\boldsymbol{W}\boldsymbol{W}\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\boldsymbol{W}\boldsymbol{W}\right) \\
&\quad= \operatorname{tr}\left(\boldsymbol{W}\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\boldsymbol{W}\boldsymbol{W}\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\boldsymbol{W}\right) = \operatorname{tr}\left(\left(\boldsymbol{W}\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\boldsymbol{W}\right)^2\right) \\
&\quad\leq \operatorname{tr}^2\left(\boldsymbol{W}\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\boldsymbol{W}\right) = \operatorname{tr}^2\left(\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\boldsymbol{W}\boldsymbol{W}\right) = \operatorname{tr}^2\left(\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\boldsymbol{T}\boldsymbol{\Sigma}_r\boldsymbol{T}\right) \\
&\quad= \operatorname{tr}^2\left(\boldsymbol{T}\boldsymbol{\Sigma}_i\boldsymbol{T}\boldsymbol{\Sigma}_r\right). \qquad \square
\end{aligned}$$

To standardize the quadratic form we also have to calculate its moments. Here, the following theorem helps:

Theorem A.4. *Let $\boldsymbol{T} \in \mathbb{R}^{d\times d}$ be a symmetric matrix and $\boldsymbol{X} \sim \mathcal{N}_d\left(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X\right)$, where $\boldsymbol{\Sigma}_X$ is positive definite. Then with $r \in \mathbb{N}$ it holds,*

$$\mathbb{E}\left(\left(\boldsymbol{X}^\top\boldsymbol{T}\boldsymbol{X}\right)^r\right) = \sum_{r_1=0}^{r-1}\binom{r-1}{r_1} g^{(r-1-r_1)} \sum_{r_2=0}^{r_1-1}\binom{r_1-1}{r_2} g^{(r_1-1-r_2)} \cdots$$

with $g^{(k)} = 2^k k! \left[\operatorname{tr}\left((\boldsymbol{T\Sigma})^{k+1} \right) + (k+1)\, \boldsymbol{\mu}_X (\boldsymbol{T\Sigma})^k \boldsymbol{T}\boldsymbol{\mu}_X \right]$ *for* $k \in \mathbb{N}$ *and* $g^{(0)} = \operatorname{tr}(\boldsymbol{T\Sigma}_X) + \boldsymbol{\mu}_X{}^\top \boldsymbol{T}\boldsymbol{\mu}_X$.

Proof. The proof can be found on page 53 in Mathai & Provost [32]. □

Corollary A.5. *Let* $\boldsymbol{T} \in \mathbb{R}^{d\times d}$ *be a symmetric matrix and* $\boldsymbol{X} \sim \mathcal{N}_d(\boldsymbol{0}_d, \boldsymbol{\Sigma}_X)$ *and* $\boldsymbol{Y} \sim \mathcal{N}_d(\boldsymbol{0}_d, \boldsymbol{\Sigma}_Y)$ *independent, where* $\boldsymbol{\Sigma}_X, \boldsymbol{\Sigma}_Y \in \mathbb{R}^{d\times d}$ *are positive definite. Then we have for all* $n_i, n_r, N \in \mathbb{N}$ *that*

$$\begin{aligned}
\mathbb{E}\left(\left(\boldsymbol{X}^\top \boldsymbol{T}\boldsymbol{X}\right)^1\right) &= \operatorname{tr}(\boldsymbol{T\Sigma}_X),\\
\mathbb{E}\left(\left(\boldsymbol{X}^\top \boldsymbol{T}\boldsymbol{X}\right)^2\right) &= 2\operatorname{tr}\left((\boldsymbol{T\Sigma}_X)^2\right) + \operatorname{tr}^2(\boldsymbol{T\Sigma}_X) \overset{A.2}{=} \mathcal{O}\left(\operatorname{tr}^2(\boldsymbol{T\Sigma}_X)\right),\\
Var\left(\boldsymbol{X}^\top \boldsymbol{T}\boldsymbol{X}\right) &= \mathcal{O}\left(\operatorname{tr}^2(\boldsymbol{T\Sigma}_X)\right),\\
\mathbb{E}\left(\left(\boldsymbol{X}^\top \boldsymbol{T}\boldsymbol{Y}\right)^1\right) &= 0,\\
\mathbb{E}\left(\left(\boldsymbol{X}^\top \boldsymbol{T}\boldsymbol{Y}\right)^2\right) &= \operatorname{tr}(\boldsymbol{T\Sigma}_X\boldsymbol{T\Sigma}_Y),\\
\mathbb{E}\left(\left(\boldsymbol{X}^\top \boldsymbol{T}\boldsymbol{Y}\right)^3\right) &= 0,\\
\mathbb{E}\left(\left(\boldsymbol{X}^\top \boldsymbol{T}\boldsymbol{Y}\right)^4\right) &= 6\operatorname{tr}\left((\boldsymbol{T\Sigma}_X\boldsymbol{T\Sigma}_Y)^2\right) + 3\operatorname{tr}^2(\boldsymbol{T\Sigma}_X\boldsymbol{T\Sigma}_Y),\\
Var\left(\boldsymbol{X}^\top \boldsymbol{T}\boldsymbol{Y}\right) &= \operatorname{tr}(\boldsymbol{T\Sigma}_X\boldsymbol{T\Sigma}_Y),\\
Var\left(\left(\boldsymbol{X}^\top \boldsymbol{T}\boldsymbol{Y}\right)^2\right) &= 6\operatorname{tr}\left((\boldsymbol{T\Sigma}_X\boldsymbol{T\Sigma}_Y)^2\right) + 2\operatorname{tr}^2(\boldsymbol{T\Sigma}_X\boldsymbol{T\Sigma}_Y),\\
\frac{4N}{n_i^2 n_r^2}\, Var\left(\left(\boldsymbol{X}^\top \boldsymbol{T}\boldsymbol{Y}\right)^2\right) &\overset{A.3}{=} \mathcal{O}\left(\operatorname{tr}^2\left(\left(\frac{N}{n_i}\boldsymbol{T\Sigma}_X \cdot \frac{N}{n_r}\boldsymbol{T\Sigma}_Y\right)^2\right)\right).
\end{aligned}$$

Moreover, for $\boldsymbol{\Sigma}_X = \boldsymbol{\Sigma}_Y$

$$\begin{aligned}
Var\left(\boldsymbol{X}^\top \boldsymbol{T}\boldsymbol{Y}\right) &= \operatorname{tr}(\boldsymbol{T\Sigma}_X\boldsymbol{T\Sigma}_X) = \mathcal{O}\left(\operatorname{tr}^2(\boldsymbol{T\Sigma}_X\boldsymbol{T\Sigma}_X)\right),\\
Var\left(\left(\boldsymbol{X}^\top \boldsymbol{T}\boldsymbol{Y}\right)^2\right) &\overset{A.2}{=} \mathcal{O}\left(\operatorname{tr}^2(\boldsymbol{T\Sigma}_X\boldsymbol{T\Sigma}_X)\right).
\end{aligned}$$

Proof. Using the inequalities for traces and with the bilinear form written as

$$\boldsymbol{X}^\top \boldsymbol{T}\boldsymbol{Y} = \frac{1}{2}\begin{pmatrix}\boldsymbol{X}\\ \boldsymbol{Y}\end{pmatrix}^\top \begin{pmatrix}0 & \boldsymbol{T}\\ \boldsymbol{T} & 0\end{pmatrix}\begin{pmatrix}\boldsymbol{X}\\ \boldsymbol{Y}\end{pmatrix},$$

$$\begin{pmatrix}\boldsymbol{X}\\ \boldsymbol{Y}\end{pmatrix} \sim \mathcal{N}_{2d}\left(\begin{pmatrix}\boldsymbol{\mu}_X\\ \boldsymbol{\mu}_Y\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_X & \boldsymbol{\Sigma}_{XY}\\ \boldsymbol{\Sigma}_{XY} & \boldsymbol{\Sigma}_Y\end{pmatrix}\right)$$

all equations follows with the previous theorem. □

Lemma A.6. *Let $X_n \in \mathcal{L}^2$ be a real random variable with $\mathbb{E}(X_n) = \mu$, $b_{n,d}$ a sequence with $\lim_{n,d\to\infty} b_{n,d} = 0$, and furthermore $c_{a,d,n_{\min}}$ a sequence with $\lim_{a,d,n_{\min}\to\infty} c_{a,d,n_{\min}} = 0$ then it holds*

- *$Var\,(X_n) \leq b_{n,d} \Rightarrow X_n$ is an consistent estimator for μ, if $n,d \to \infty$,*
- *$Var\,(X_n) \leq c_{a,d,n_{\min}} \Rightarrow X_n$ is an consistent estimator for μ, if $a, d, n_{\min} \to \infty$.*

For $\mu \neq 0$ they are especially ratio-consistent.

Proof. For arbitrary $\epsilon > 0$ the Tschebyscheff inequality leads to

$$\mathbb{P}\left(|X_n - \mu| \geq \epsilon\right) \leq \frac{\mathbb{E}\left(|X_n - \mu|^2\right)}{\epsilon^2} = \frac{Var\,(X_n)}{\epsilon^2} \leq \frac{b_{n,d}}{\epsilon^2}.$$

Consider the limit for $n, d \to \infty$ justifies the consistency and using this for X_n/μ leads to ratio-consistency. The second part follows identically. □

This result is especially true if $b_{n,d}$ or $c_{a,d,n_{\min}}$ only depends on n resp. $n_{\min}$. For completeness we state a straightforward application of the Cauchy–Bunyakovsky–Schwarz inequality:

Lemma A.7. *For real random variables $X, Y \in \mathcal{L}^2$ it holds*

$$Cov\,(X,Y) \leq \sqrt{Var\,(X)}\sqrt{Var\,(Y)}$$

and so for X, Y identically distributed

$$Cov\,(X,Y) \leq Var\,(X).$$

The next result gives equivalent conditions for $\beta_1 \to \gamma \in \{0,1\}$:

Lemma A.8. *Let be λ_ℓ again the eigenvalues of $\boldsymbol{T}\boldsymbol{V}_N\boldsymbol{T}$ sorted so that λ_1 is the biggest one. Then it follows*

$$\lim_{N,d\to\infty} \beta_1 = 1 \;\Leftrightarrow\; \lim_{N,d\to\infty} \frac{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}{\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} = 1 \;\Leftrightarrow\; \lim_{N,d\to\infty} \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} = 1,$$

$$\lim_{N,d\to\infty} \beta_1 = 0 \;\Leftrightarrow\; \lim_{N,d\to\infty} \frac{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}{\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} = 0 \;\Leftrightarrow\; \lim_{N,d\to\infty} \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} = 0.$$

Moreover we know $0 \leq \frac{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}{\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} = \tau_P \leq 1$. This Lemma also holds if $\lim_{N,d\to\infty}$ is replaced by $\lim_{a,N} \to \infty$ or $\lim_{a,d,N\to\infty}$.

Proof. This follows from Lemma 8.1 given in the supplement in Pauly et al. [34][page 21] since their result does not depend on the concrete matrix, i.e. can be directly applied for $\boldsymbol{V}_N$. Moreover, the different asymptotic frameworks do not influence the proof since they are hidden within the above convergences. □

To prove the properties of the subsampling-type estimators some auxiliaries are needed. In particular, the following lemma allows us to decompose the variances and to use conditional terms for the calculation.

Lemma A.9. *Let X be a real random variable and denote by $\mathcal{F}$ a σ-field. Then it holds that*

$$Var(X) = \mathbb{E}\left(Var\left(X|\mathcal{F}\right)\right) + Var\left(\mathbb{E}\left(X|\mathcal{F}\right)\right).$$

Proof. With the rules for conditional expectations we calculate

$$\mathbb{E}\left(Var\left(X|\mathcal{F}\right)\right) = \mathbb{E}\left(\mathbb{E}\left(X^2|\mathcal{F}\right)\right) - \mathbb{E}\left(\left[\mathbb{E}\left(X|\mathcal{F}\right)\right]^2\right) = \mathbb{E}\left(X^2\right) - \mathbb{E}\left(\left[\mathbb{E}\left(X|\mathcal{F}\right)\right]^2\right),$$
$$Var\left(\mathbb{E}\left(X|\mathcal{F}\right)\right) = \mathbb{E}\left(\left[\mathbb{E}\left(X|\mathcal{F}\right)\right]^2\right) - \left[\mathbb{E}\left(\mathbb{E}\left(X|\mathcal{F}\right)\right)\right]^2 = \mathbb{E}\left(\left[\mathbb{E}\left(X|\mathcal{F}\right)\right]^2\right) - \left[\mathbb{E}\left(X\right)\right]^2.$$

The result follows by sum up this both parts. ☐

We will apply the result for certain amounts (i.e. numbers) of pairs below. There, for each $i = 1, \dots, a$ and $b = 1, \dots, B$ we independently draw random subsamples $\{\sigma_{1i}(b), \dots, \sigma_{mi}(b)\}$ of length m from $\{1, \dots, n_i\}$ and store them in a joint random vector $\boldsymbol{\sigma}(b,m) = (\boldsymbol{\sigma}_1(b,m), \dots, \boldsymbol{\sigma}_a(b,m)) = (\sigma_{11}(b), \dots, \sigma_{ma}(b))$. Besides we define $\mathbb{N}_k = \{1, \dots, k\}$.

Lemma A.10. *Let $M(B, \boldsymbol{\sigma}(b,m))$ be the amount of pairs $(k,\ell) \in \mathbb{N}_B^2$, which fulfill $\boldsymbol{\sigma}_i(k,m)$ and $\boldsymbol{\sigma}_i(\ell,m)$ have totally different elements for all $i = 1, ..., a$ and analogue $M(B, \boldsymbol{\sigma}_i(b,m))$. As long as $m \leq n_i$ for all $i \in \mathbb{N}_a$, it holds*

$$\frac{\mathbb{E}\left(|\mathbb{N}_B^2 \setminus M(B, \boldsymbol{\sigma}(b,m))|\right)}{B^2} = 1 - \left(1 - \frac{1}{B}\right) \cdot \prod_{i=1}^{a} \frac{\binom{n_i - m}{m}}{\binom{n_i}{m}}$$

and

$$\frac{\mathbb{E}\left(|\mathbb{N}_B^2 \setminus M(B, \boldsymbol{\sigma}_i(b,m))|\right)}{B^2} = 1 - \left(1 - \frac{1}{B}\right) \cdot \frac{\binom{n_i - m}{m}}{\binom{n_i}{m}}$$

where $|\cdot|$ denotes the number of elements.

Let $M(B, (\boldsymbol{\sigma}_i(b,m), \boldsymbol{\sigma}_r(b,m)))$ be the amount of pairs $(k,\ell) \in \mathbb{N}_B^2$ fulfilling $\boldsymbol{\sigma}_i(k,m)$ and $\boldsymbol{\sigma}_i(\ell,m)$ and moreover $\boldsymbol{\sigma}_r(k,m)$ and $\boldsymbol{\sigma}_r(\ell,m)$ have totally different elements. If $m \leq n_i$ it holds

$$\frac{\mathbb{E}\left(|\mathbb{N}_B^2 \setminus M(B, (\boldsymbol{\sigma}_i(b,m), \boldsymbol{\sigma}_r(b,m)))|\right)}{B^2} = 1 - \left(1 - \frac{1}{B}\right) \cdot \frac{\binom{n_i - m}{m}}{\binom{n_i}{m}} \cdot \frac{\binom{n_r - m}{m}}{\binom{n_r}{m}}.$$

Proof. Because $M(B, \boldsymbol{\sigma}(b,m))$ never contains pairs of the kind (k,k) the maximal number of elements is $B^2 - B$. The fact that two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$ have no element in common, even at different components, is denoted as $\boldsymbol{a} \neq! \, \boldsymbol{b}$.

The number of totally different pairs can be seen as a binomial distribution with $B^2 - B$ elements, and to calculate the necessary probability independence is used. With the fact that all combinations in this situation have the same probability it follows that

$$\begin{aligned}\mathbb{P}\left(\boldsymbol{\sigma}(k,m) \neq! \boldsymbol{\sigma}(\ell,m)\right) &= \mathbb{P}\left(\bigcap_{i=1}^{a} (\boldsymbol{\sigma}_i(k,m) \neq! \boldsymbol{\sigma}_i(\ell,m))\right) \\ &= \prod_{i=1}^{a} \mathbb{P}\left(\boldsymbol{\sigma}_i(k,m) \neq! \boldsymbol{\sigma}_i(\ell,m)\right) = \prod_{i=1}^{a} \frac{\binom{n_i}{m} \cdot \binom{n_i-m}{m}}{\binom{n_i}{m}^2} = \prod_{i=1}^{a} \frac{\binom{n_i-m}{m}}{\binom{n_i}{m}}.\end{aligned}$$

If two times m elements are picked from $\mathbb{N}_{n_i}$ there are $\binom{n_i}{m}^2$ possibilities, where in $\binom{n_i}{m} \cdot \binom{n_i-m}{m}$ of them both m-tuples are totally different. This leads to the stated probability and with the mean of the binomial distribution we get

$$\mathbb{E}\left(|M(B, \boldsymbol{\sigma}(b,m))|\right) = (B^2 - B) \cdot \prod_{i=1}^{a} \frac{\binom{n_i-m}{m}}{\binom{n_i}{m}}.$$

All in all we calculate

$$\begin{aligned}\frac{\mathbb{E}\left(|\mathbb{N}_B^2 \setminus M(B, \boldsymbol{\sigma}(b,m))|\right)}{B^2} &= \frac{|\mathbb{N}_B^2| - \mathbb{E}\left(|M(B, \boldsymbol{\sigma}(b,m))|\right)}{B^2} \\ &= 1 - \left(1 - \frac{1}{B}\right) \cdot \prod_{i=1}^{a} \frac{\binom{n_i-m}{m}}{\binom{n_i}{m}}.\end{aligned}$$

For $M(B, (\boldsymbol{\sigma}_i(b,m), \boldsymbol{\sigma}_r(b,m)))$ and $M(B, \boldsymbol{\sigma}_i(b,m))$ less multiplications are necessary, so the results follow. □

If $B(N) \to \infty$ (for example B could be chosen proportional to N) these terms converge to zero, disregarding the number of groups or of m.

Appendix B: Proofs of Section 3

Proof of Theorem 3.1 (p.2748). The proof of this lemma is very similar to the one from Pauly et al. [34][Theorem 2.1]. Due to the fact that a finite sum of multivariate normally distributed random variables is again multivariate normally distributed, the representation theorem can be used to (distributionally equivalently) express the quadratic form as $W_N = \sum_{s=1}^{ad} \frac{\lambda_s}{\sqrt{\sum_{\ell=1}^{ad} \lambda_\ell^2}} \left(\frac{C_s - 1}{\sqrt{2}}\right)$.

The only differences to Pauly et al. [34][Theorem 2.1] are that in the case of more groups the eigenvalues do not only depend on d but also on the n_i and a and that there are more terms to sum. The first point has only an influence on the limit of the β_s. The higher number of summands does not matter because we observe the asymptotic under the asymptotic frameworks (4)–(5), for which at least a or d converge to infinity. The proofs from Pauly et al. [34][Theorem 2.1] only need the representation from above, a number of summations which goes to infinity and the conditions on the limits of the β_s. Since these are fulfilled the proof can be conducted in the same way.

Finally, it remains to prove the if and only if result stated in a) and b) for which we underline the dependence of β_i on N by writing $\beta_i(N)$.

Part (a) Suppose that $Q_N \xrightarrow{\mathcal{D}} Z \sim \mathcal{N}(0,1)$. Then this convergence also holds for all subsequences N' of N, i.e. $Q_{N'} \xrightarrow{\mathcal{D}} Z$, for all $N' \to \infty$. Now we consider $\beta_1(N)$. Due to $\beta_1(N) \in [0,1]$ there exists an arbitrary convergent subsequence which we denote as $\beta_1(N') \to b_1 \in [0,1]$.

We define $Z'(N') := Q_{N'} - \beta_1(N') \cdot (C_1 - 1)/\sqrt{2}$. From Lévy's continuity theorem it follows that $\varphi_{Q_{N'}}(t) \to \varphi_Z(t)$ for all $t \in \mathbb{R}$ for the corresponding characteristic function. Due to independence we calculate for all $t \in \mathbb{R}$:

$$\varphi_{Q_{N'}}(t) = \varphi_{\beta_1(N')\cdot(C_1-1)/\sqrt{2}+Z'(N')}(t) = \varphi_{\beta_1(N')\cdot(C_1-1)/\sqrt{2}}(t) + \varphi_{Z'(N')}(t).$$

Because $\varphi_{Q_{N'}}(t) \to \varphi_Z(t)$ and $\varphi_{\beta_1(N')\cdot(C_1-1)/\sqrt{2}}(t) \to \varphi_{b_1\cdot(C_1-1)/\sqrt{2}}(t)$ holds for all $t \in \mathbb{R}$, we also know that $\varphi_{Z'(N')}(t)$ converges to some $\varphi_\Upsilon(t)$. Moreover there exists a random variable Υ with the characteristic function $\varphi_\Upsilon(t)$ and therefore $Z'(N') \xrightarrow{\mathcal{D}} \Upsilon$. All in all we have

$$Q_{N'} \xrightarrow{\mathcal{D}} b_1 \cdot (C_1 - 1)/\sqrt{2} + \Upsilon \quad \text{and} \quad Q_{N'} \xrightarrow{\mathcal{D}} Z \sim \mathcal{N}(0,1)$$

while $b_1 \cdot (C_1 - 1)/\sqrt{2}$ and Υ are independent. With Cramér's Theorem (see Cramér [12]), the sum of a scaled standardized χ_1^2-distributed random variable and another independent random variable can never be normally distributed. Therefore $b_1 = 0$ follows for all convergent subsequences of $\beta_1(N)$ and so $\beta_1(N) \to 0$.

Part (b) Now assume that for $N \to \infty$, we have $Q_N \xrightarrow{\mathcal{D}} (C_1 - 1)/\sqrt{2}$ with $C_1 \sim \chi_1^2$. Then we can obvious exclude $\beta_1(N)^2 \to 0$, because in this case the asymptotic distribution of the quadratic form would be a standard normal distribution by part (a). The characteristic function of $W_N = \frac{Q_N - \operatorname{tr}(\boldsymbol{TV}_N)}{\sqrt{2\operatorname{tr}\left((\boldsymbol{TV}_N)^2\right)}}$ is, e.g., given in Witting & Müller-Funke [42], Section 5. With the help of Lévy's continuity theorem this leads for all $t \in \mathbb{R}$ to

$$\begin{aligned}\varphi_{W_N}(t) &= \prod_{\ell=1}^{ad}\left(1 - \frac{2i\beta_\ell(N)t}{\sqrt{2}}\right)^{-1/2} \exp\left(-it\frac{\beta_\ell(N)}{\sqrt{2}}\right) \\ &\to \left(1 - \frac{2it}{\sqrt{2}}\right)^{-1/2} \exp\left(-\frac{it}{\sqrt{2}}\right) = \varphi_{(C_1-1)/\sqrt{2}}(t).\end{aligned}$$

Thus, applying the continuous mapping theorem we have for all $t \in \mathbb{R}$

$$\begin{aligned}&\Big|\prod_{\ell=1}^{ad}\left(1 - \frac{2i\beta_\ell(N)t}{\sqrt{2}}\right)^{-1/2} \exp\left(-\frac{i\beta_\ell(N)t}{\sqrt{2}}\right)\Big|^{-4} = \prod_{\ell=1}^{ad}\Big|1 - \frac{2i\beta_\ell(N)t}{\sqrt{2}}\Big|^2 \\ =&\prod_{\ell=1}^{ad}\left(1 + \frac{4\beta_\ell(N)^2t^2}{2}\right) \to 1 + \frac{4}{2}t^2 = \Big|\left(1 - \frac{2i}{\sqrt{2}}t\right)^{-1/2} \exp\left(-\frac{i}{\sqrt{2}}t\right)\Big|^{-4}.\end{aligned}$$

In the special case $t = 1$ this means

$$\prod_{\ell=1}^{ad} (1 + 2\beta_\ell(N)^2) \to 3.$$

But we can size up the product by

$$\begin{aligned}\prod_{\ell=1}^{ad}(1 + 2\beta_\ell(N)^2) &\geq 1 + 2 \cdot \sum_{\ell=1}^{ad} \beta_\ell(N)^2 + 4\beta_1(N)^2 \left(\sum_{\ell=2}^{ad} \beta_\ell(N)^2\right) \\ &= 1 + 2 \cdot 1 + 4\beta_1(N)^2 \left(1 - \beta_1(N)^2\right) \\ &= 3 + 4\beta_1(N)^2 \left(1 - \beta_1(N)^2\right) \geq 3.\end{aligned}$$

Now we again consider an arbitrary convergent subsequence $\beta_1(N') \to b_1 \in (0, 1]$. Since the above inequality, also holds for all subsequences, the product only converges if $\lim_{N\to\infty} \beta_1(N')^2(1 - \beta_1(N')^2) = b_1^2(1 - b_1^2) = 0$, which implies $b_1 = 1$. Due to $\beta_1(N) \in [0, 1]$ we deduce $\beta_1(N) \to 1$. □

Proof of Theorem 3.2 (p.2748). First we consider the distribution of the standardized quadratic form $\widetilde{W}_N$ under $H_1 : \boldsymbol{T\mu} \neq 0$ with $\boldsymbol{Z} \sim \mathcal{N}_{ad}(\boldsymbol{0}, \boldsymbol{V}_N)$

$$\begin{aligned}Q_N &= N\overline{\boldsymbol{X}}^\top \boldsymbol{T}\overline{\boldsymbol{X}} = N(\overline{\boldsymbol{X}} - \boldsymbol{\mu} + \boldsymbol{\mu})^\top \boldsymbol{T}(\overline{\boldsymbol{X}} - \boldsymbol{\mu} + \boldsymbol{\mu}) \\ &\stackrel{\mathcal{D}}{=} \boldsymbol{Z}^\top \boldsymbol{TZ} + \boldsymbol{Z}^\top \sqrt{N}\boldsymbol{T\mu} + \sqrt{N}\boldsymbol{\mu}^\top \boldsymbol{TZ} + N\boldsymbol{\mu}^\top \boldsymbol{T}^\top \boldsymbol{T\mu}.\end{aligned}$$

For part a) we calculate

$$\widetilde{W}_N \stackrel{\mathcal{D}}{=} \frac{\boldsymbol{Z}^\top \boldsymbol{TZ} + 2\sqrt{N}\boldsymbol{\mu}^\top \boldsymbol{TZ} + N\boldsymbol{\mu}^\top \boldsymbol{T\mu} - \operatorname{tr}(\boldsymbol{TV}_N)}{\sqrt{2\operatorname{tr}\left((\boldsymbol{TV}_N)^2\right)}}.$$

The second summand fulfills

$$\begin{aligned}\mathbb{E}\left(\frac{2\sqrt{N}\boldsymbol{\mu}^\top \boldsymbol{TZ}}{\sqrt{2\operatorname{tr}\left((\boldsymbol{TV}_N)^2\right)}}\right) &= 0, \\ Var\left(\frac{2\sqrt{N}\boldsymbol{\mu}^\top \boldsymbol{TZ}}{\sqrt{2\operatorname{tr}\left((\boldsymbol{TV}_N)^2\right)}}\right) &= 2\frac{N\boldsymbol{\mu}^\top \boldsymbol{TV}_N\boldsymbol{T\mu}}{\operatorname{tr}\left((\boldsymbol{TV}_N)^2\right)} \in \mathcal{O}(1)\end{aligned}$$

under the given local alternative. Thus, by Tschebyscheff inequality this means

$$\widetilde{W}_N \stackrel{\mathcal{D}}{=} \frac{\boldsymbol{Z}^\top \boldsymbol{TZ} - \operatorname{tr}(\boldsymbol{TV}_N)}{\sqrt{2\operatorname{tr}\left((\boldsymbol{TV}_N)^2\right)}} + \frac{N \cdot \boldsymbol{\mu}^\top \boldsymbol{T\mu}}{\sqrt{2\operatorname{tr}\left((\boldsymbol{TV}_N)^2\right)}} + \mathcal{O}_{\mathcal{P}}(1).$$

Now the first part has exactly the same distribution as the standardized quadratic form $\widetilde{Q}_N$ under the null hypothesis and therefore the result follows.

For part b) we consider again the quadratic form and calculate with Mathai & Provost [32]

$$Q_N \stackrel{\mathcal{D}}{=} \sum_{\ell=1}^{ad} \lambda_\ell \widetilde{C}_\ell \quad \widetilde{C}_\ell \sim \chi_1^2(\underbrace{(\sqrt{N}\boldsymbol{O}_N \boldsymbol{V}_N^{-1/2}\boldsymbol{T}\boldsymbol{\mu})_\ell^2}_{:=\delta_\ell^2}),$$

where $\boldsymbol{O}_N$ is the orthogonal matrix which diagonalizes $\boldsymbol{V}_N^{1/2}\boldsymbol{T}\boldsymbol{V}_N^{1/2}$ and λ_ℓ are the eigenvalues of $\boldsymbol{V}_N^{1/2}\boldsymbol{T}\boldsymbol{V}_N^{1/2}$ in decreasing order. The involved non-central chi-square distributed random variables have expectation $\mathbb{E}(\widetilde{C}_\ell) = 1 + \delta_\ell^2$ and variance $Var(\widetilde{C}_\ell) = 2(1 + 2\delta_\ell^2)$. Defining $\widetilde{\lambda}_\ell = \lambda_\ell \sqrt{1 + 2\delta_\ell^2}$ and $\widetilde{\beta}_\ell = \widetilde{\lambda}_\ell / \sqrt{\sum_{k=1}^{ad} \widetilde{\lambda}_k^2}$ we calculate

$$\begin{aligned}
\widetilde{W}_N &= \sum_{\ell=1}^{ad} \frac{\lambda_\ell}{\sqrt{\sum_{k=1}^{ad} \lambda_k^2}} \left(\frac{\widetilde{C}_\ell - (1 + \delta_\ell^2)}{\sqrt{2}} \right) + \sum_{\ell=1}^{ad} \frac{\lambda_\ell}{\sqrt{\sum_{k=1}^{ad} \lambda_k^2}} \left(\frac{\delta_\ell^2}{\sqrt{2}} \right) \\
&= \sqrt{\frac{\sum_{k=1}^{ad} \widetilde{\lambda}_k^2}{\sum_{k=1}^{ad} \lambda_k^2}} \cdot \sum_{\ell=1}^{ad} \frac{\lambda_\ell \cdot \sqrt{1 + 2\delta_\ell^2}}{\sqrt{\sum_{k=1}^{ad} \widetilde{\lambda}_k^2}} \left(\frac{\widetilde{C}_\ell - (1 + \delta_\ell^2)}{\sqrt{2} \cdot \sqrt{1 + 2\delta_\ell^2}} \right) + \sum_{\ell=1}^{ad} \beta_\ell \left(\frac{\delta_\ell^2}{\sqrt{2}} \right) \\
&= \sqrt{1 + 2N \frac{\boldsymbol{\mu}^\top \boldsymbol{T}\boldsymbol{V}_N \boldsymbol{T}\boldsymbol{\mu}}{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2 \right)}} \cdot \sum_{\ell=1}^{ad} \widetilde{\beta}_\ell \left(\frac{\widetilde{C}_\ell - (1 + \delta_\ell^2)}{\sqrt{2} \cdot \sqrt{1 + 2\delta_\ell^2}} \right) + \frac{N \cdot \boldsymbol{\mu}^\top \boldsymbol{T}\boldsymbol{\mu}}{\sqrt{2 \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2 \right)}}.
\end{aligned}$$

Now, if $\beta_1 \to 0 \Leftrightarrow \beta_\ell \to 0 \;\; \forall l \in \mathbb{N}_{ad}$ it holds for arbitrary $\widetilde{\beta}_\ell^2$ that

$$\begin{aligned}
0 \le \widetilde{\beta}_\ell^2 &= \frac{\lambda_\ell^2 (1 + 2\delta_\ell^2)}{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2 \right) + 2 \sum_{k=1}^{ad} \lambda_k^2 \delta_k^2} \\
&\le \beta_\ell^2 + 2 \frac{\lambda_\ell^2 \delta_\ell^2}{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2 \right)} = \beta_\ell^2 + 2\beta_\ell \frac{\lambda_\ell \delta_\ell^2}{\sqrt{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2 \right)}} \\
&\le \beta_\ell^2 + 2\beta_\ell \frac{\sum_{\ell=1}^{ad} \lambda_\ell \delta_\ell^2}{\sqrt{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2 \right)}} = \beta_\ell^2 + 2\beta_\ell \frac{N \cdot \boldsymbol{\mu}^\top \boldsymbol{T}\boldsymbol{\mu}}{\sqrt{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2 \right)}} \to 0.
\end{aligned}$$

Because all requirements are fulfilled we can use Theorem 1 from Hajek et al. [19] to deduce the asymptotic distribution of

$$\sum_{\ell=1}^{ad} \widetilde{\beta}_\ell \left(\frac{\widetilde{C}_\ell - (1 + \delta_\ell^2)}{\sqrt{2} \cdot \sqrt{1 + 2\delta_\ell^2}} \right)$$

as before. This evidently leads to

$$\widetilde{W}_N \stackrel{\mathcal{D}}{=} \sqrt{1 + 2N\frac{\boldsymbol{\mu}^\top \boldsymbol{T}\boldsymbol{V}_N\boldsymbol{T}\boldsymbol{\mu}}{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}} \cdot Z + \frac{N \cdot \boldsymbol{\mu}^\top \boldsymbol{T}\boldsymbol{\mu}}{\sqrt{2\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}} + \mathcal{O}_{\mathcal{P}}(1)$$

for a normally distributed random variable $Z \sim \mathcal{N}(0,1)$. For $\beta_1 \to 0$ we know that $W_N(H_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$ and therefore the result follows. □

Proof of Lemma 3.3 (p.2750). Remember that with $\boldsymbol{Y}_{i,\ell,k} := \boldsymbol{T}_S(\boldsymbol{X}_{i,\ell} - \boldsymbol{X}_{i,k})$ and $i \neq r \in \mathbb{N}_a$, $a > 1$ trace estimators were defined by

$$A_{i,1} = \frac{1}{2 \cdot \binom{n_i}{2}} \sum_{\substack{\ell_1,\ell_2=1 \\ \ell_1>\ell_2}}^{n_i} (\boldsymbol{X}_{i,\ell_1} - \boldsymbol{X}_{i,\ell_2})^\top \boldsymbol{T}_S (\boldsymbol{X}_{i,\ell_1} - \boldsymbol{X}_{i,\ell_2}),$$

$$A_{i,r,2} = \frac{1}{4 \cdot \binom{n_i}{2}\binom{n_r}{2}} \sum_{\substack{\ell_1,\ell_2=1 \\ \ell_1>\ell_2}}^{n_i} \sum_{\substack{k_1,k_2=1 \\ k_1>k_2}}^{n_r} \left[(\boldsymbol{X}_{i,\ell_1} - \boldsymbol{X}_{i,\ell_2})^\top \boldsymbol{T}_S (\boldsymbol{X}_{r,k_1} - \boldsymbol{X}_{r,k_2})\right]^2,$$

$$A_{i,3} = \frac{1}{4 \cdot 6\binom{n_i}{4}} \sum_{\substack{\ell_1,\ell_2=1 \\ \ell_1>\ell_2}}^{n_i} \sum_{\substack{k_2=1 \\ k_2\neq\ell_1,\ell_2}}^{n_i-1} \sum_{\substack{k_1=k_2+1 \\ k_1\neq\ell_1,\ell_2}}^{n_i} \times \left[(\boldsymbol{X}_{i,\ell_1} - \boldsymbol{X}_{i,\ell_2})^\top \boldsymbol{T}_S (\boldsymbol{X}_{i,k_1} - \boldsymbol{X}_{i,k_2})\right]^2,$$

$$A_4 = \sum_{i=1}^{a} \left(\frac{N}{n_i}\right)^2 (\boldsymbol{T}_W)_{ii}{}^2 A_{i,3} + 2\sum_{i=1}^{a-1}\sum_{r=i+1}^{a} \frac{N^2}{n_i n_r} (\boldsymbol{T}_W)_{ir}{}^2 A_{i,r,2}.$$

For $\ell \neq k$ we know $\boldsymbol{Y}_{i,\ell,k} \sim \mathcal{N}(\boldsymbol{0}_d, 2\boldsymbol{T}_S\boldsymbol{\Sigma}_i\boldsymbol{T}_S)$ and for totally different indices the $\boldsymbol{Y}_{i,\ell,k}$ are statistically independent. So the previous lemmata can be used to calculate the moments. The unbiasedness can be shown by calculating the expectation values for each estimator

$$\mathbb{E}(A_{i,1}) = \frac{1}{2 \cdot \binom{n_i}{2}} \sum_{\substack{\ell_1,\ell_2=1 \\ \ell_1>\ell_2}}^{n_i} \mathbb{E}\left[\boldsymbol{Y}_{i,\ell_1,\ell_2}{}^\top \boldsymbol{Y}_{i,\ell_1,\ell_2}\right] \stackrel{A.5}{=} \operatorname{tr}(\boldsymbol{T}_S\boldsymbol{\Sigma}_i).$$

The following argument will be used several times in this work with small differences, so incidentally it will be more detailed.

We recognize first that $Cov\left[\boldsymbol{Y}_{i,\ell_1,\ell_2}{}^\top \boldsymbol{Y}_{i,\ell_1,\ell_2}; \boldsymbol{Y}_{i,\ell_1',\ell_2'}{}^\top \boldsymbol{Y}_{i,\ell_1',\ell_2'}\right]$ is 0 if all indices are totally different, so just $\binom{n_i}{2}\left(\binom{n_i}{2} - \binom{n_i-2}{2}\right)$ combinations remain. Instead of calculating the covariances of the remaining quadratic forms it is easier to use lemmata from above. By using the fact that all quadratic forms are identically distributed, we can calculate the variances which are all the same so it is just the number of remaining combinations multiplied with the variances. This leads to:

$$\begin{aligned}
Var\left(A_{i,1}\right) &= \frac{1}{4\cdot\binom{n_i}{2}^2}\sum_{\substack{\ell_1,\ell_2=1\\ \ell_1>\ell_2}}^{n_i}\sum_{\substack{\ell_1',\ell_2'=1\\ \ell_1'>\ell_2'}}^{n_i} Cov\left[\boldsymbol{Y}_{i,\ell_1,\ell_2}{}^\top\boldsymbol{Y}_{i,\ell_1,\ell_2};\boldsymbol{Y}_{i,\ell_1',\ell_2'}{}^\top\boldsymbol{Y}_{i,\ell_1',\ell_2'}\right]\\
&\overset{A.7}{\leq} \frac{\binom{n_i}{2}-\binom{n_i-2}{2}}{4\binom{n_i}{2}}\ Var\left[\boldsymbol{Y}_{i,1,2}{}^\top\boldsymbol{Y}_{i,1,2}\right] + \frac{\binom{n_i-2}{2}}{4\binom{n_i}{2}}\cdot 0\\
&\overset{A.5}{=} \frac{\binom{n_i}{2}-\binom{n_i-2}{2}}{4\binom{n_i}{2}}\mathcal{O}\left(\operatorname{tr}^2\left(2\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)\right)\\
&= \mathcal{O}\left(n_i^{-1}\right)\cdot\mathcal{O}\left(\operatorname{tr}^2\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)\right).
\end{aligned}$$

With these values we know for $\boldsymbol{V}_N = \bigoplus_{i=1}^a \frac{N}{n_i}\boldsymbol{\Sigma}_i$ that

$$\mathbb{E}\left(\sum_{i=1}^a \frac{N}{n_i}(\boldsymbol{T}_W)_{ii}A_{i,1}\right) = \sum_{i=1}^a \frac{N}{n_i}(\boldsymbol{T}_W)_{ii}\mathbb{E}\left(A_{i,1}\right) = \operatorname{tr}\left(\boldsymbol{T}\boldsymbol{V}_N\right)$$

and

$$\begin{aligned}
Var\left(\frac{\sum\limits_{i=1}^a \frac{N}{n_i}(\boldsymbol{T}_W)_{ii}A_{i,1}}{\mathbb{E}\left(\sum\limits_{i=1}^a \frac{N}{n_i}(\boldsymbol{T}_W)_{ii}A_{i,1}\right)}\right) &= \frac{\sum\limits_{i=1}^a \frac{N^2}{n_i^2}(\boldsymbol{T}_W)_{ii}{}^2\ Var(A_{i,1})}{\operatorname{tr}^2\left(\boldsymbol{T}\boldsymbol{V}_N\right)}\\
&\leq \frac{\sum\limits_{i=1}^a \mathcal{O}\left(n_i^{-1}\right)\cdot\mathcal{O}\left(\operatorname{tr}^2\left(\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)\right)}{\operatorname{tr}^2\left(\boldsymbol{T}\boldsymbol{V}_N\right)}\\
&\leq \frac{\mathcal{O}\left(\frac{1}{n_{\min}}\right)\cdot\mathcal{O}\left(\sum\limits_{i=1}^a \operatorname{tr}^2\left(\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)\right)}{\operatorname{tr}^2\left(\boldsymbol{T}\boldsymbol{V}_N\right)}\\
&\leq \frac{\mathcal{O}\left(\frac{1}{n_{\min}}\right)\cdot\mathcal{O}\left(\operatorname{tr}^2\left(\sum\limits_{i=1}^a \frac{N}{n_i}(\boldsymbol{T}_W)_{ii}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)\right)}{\operatorname{tr}^2\left(\boldsymbol{T}\boldsymbol{V}_N\right)}\\
&= \mathcal{O}\left(\frac{1}{n_{\min}}\right).
\end{aligned}$$

So the conditions for an unbiased and ratio-consistent estimator are fulfilled.

The same steps with a different number of remaining combinations leads to

$$\begin{aligned}
\mathbb{E}\left(A_{i,3}\right) &= \frac{1}{4\cdot 6\binom{n_i}{4}}\sum_{\substack{\ell_1,\ell_2=1\\ \ell_1>\ell_2}}^{n_i}\sum_{\substack{k_2=1\\ k_2\neq\ell_1,\ell_2}}^{n_i-1}\sum_{\substack{k_1=k_2+1\\ k_1\neq\ell_1,\ell_2}}^{n_i}\mathbb{E}\left(\left[\boldsymbol{Y}_{i,\ell_1,\ell_2}{}^\top\boldsymbol{Y}_{i,k_1,k_2}\right]^2\right)\\
&\overset{A.5}{=} \frac{1}{4\cdot 6\binom{n_i}{4}}\cdot 6\binom{n_i}{4}\cdot\operatorname{tr}\left(4\cdot\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)^2\right) = \operatorname{tr}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)^2\right),
\end{aligned}$$

$$
\begin{aligned}
Var\left(A_{i,3}\right) &= \sum_{\substack{\ell_1,\ell_2=1\\ \ell_1>\ell_2}}^{n_i} \sum_{\substack{k_1,k_2=1\\ \ell_2,\ell_1\neq k_1,k_2}}^{n_i} \sum_{\substack{\ell_1',\ell_2'=1\\ \ell_1'>\ell_2'}}^{n_i} \sum_{\substack{k_1',k_2'=1\\ \ell_2',\ell_1'\neq k_1',k_2'}}^{n_i} \\
&\quad \times \frac{Cov\left([\boldsymbol{Y}_{i,\ell_1,\ell_2}{}^\top \boldsymbol{Y}_{i,k_1,k_2}]^2 ; [\boldsymbol{Y}_{i,\ell_1',\ell_2'}{}^\top \boldsymbol{Y}_{i,k_1',k_2'}]^2\right)}{4^2\cdot 6^2\cdot \binom{n_i}{4}^2} \\
&\overset{\text{A.7}}{\leq} \frac{6\binom{n_i}{4}-6\binom{n_i-4}{4}}{4^2\cdot 6\cdot\binom{n_i}{4}}\, Var\left(\left[\boldsymbol{Y}_{i,1,2}{}^\top \boldsymbol{Y}_{i,3,4}\right]^2\right) \\
&\overset{\text{A.5}}{=} \frac{\binom{n_i}{4}-\binom{n_i-4}{4}}{16\binom{n_i}{4}}\mathcal{O}\left(\operatorname{tr}^2\left((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^2\right)\right) \\
&= \mathcal{O}\left(n_i^{-1}\right)\cdot \mathcal{O}\left(\operatorname{tr}^2\left((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^2\right)\right),
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left(A_{i,r,2}\right) &= \frac{1}{4\cdot\binom{n_i}{2}\binom{n_r}{2}} \sum_{\substack{\ell_1,\ell_2=1\\ \ell_1>\ell_2}}^{n_i} \sum_{\substack{k_1,k_2=1\\ k_1>k_2}}^{n_r} \mathbb{E}\left(\left[\boldsymbol{Y}_{i,\ell_1,\ell_2}{}^\top \boldsymbol{Y}_{r,k_1,k_2}\right]^2\right) \\
&\overset{\text{A.5}}{=} \frac{1}{4\cdot\binom{n_i}{2}\binom{n_r}{2}}\cdot\binom{n_i}{2}\cdot\binom{n_r}{2}\cdot\operatorname{tr}\left(4\cdot\boldsymbol{\Sigma}_i\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right) \\
&= \operatorname{tr}\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right),
\end{aligned}
$$

$$
\begin{aligned}
Var\left(\frac{2N^2}{n_in_r}A_{i,r,2}\right) &= \frac{4N^4}{n_i^2n_r^2} \sum_{\substack{\ell_1,\ell_2=1\\ \ell_1>\ell_2}}^{n_1} \sum_{\substack{k_1,k_2=1\\ k_1>k_2}}^{n_2} \sum_{\substack{\ell_1',\ell_2'=1\\ \ell_1'>\ell_2'}}^{n_i} \sum_{\substack{k_1',k_2'=1\\ k_1'>k_2'}}^{n_r} \\
&\quad\times\frac{Cov\left(\left[\boldsymbol{Y}_{i,\ell_1,\ell_2}{}^\top \boldsymbol{Y}_{r,k_1,k_2}\right]^2 ; \left[\boldsymbol{Y}_{i,\ell_1',\ell_2'}{}^\top \boldsymbol{Y}_{r,k_1',k_2'}\right]^2\right)}{16\cdot\binom{n_i}{2}^2\binom{n_r}{2}^2} \\
&\overset{\text{A.7}}{\leq} \frac{4N^4}{n_i^2n_r^2}\frac{\binom{n_i}{2}\binom{n_r}{2}-\binom{n_i-2}{2}\binom{n_r-2}{2}}{16\cdot\binom{n_i}{2}\binom{n_r}{2}}\, Var\left(\left[\boldsymbol{Y}_{i,1,2}{}^\top \boldsymbol{Y}_{r,1,2}\right]^2\right) \\
&\overset{\text{A.5}}{\leq} \frac{\binom{n_i}{2}\binom{n_r}{2}-\binom{n_i-2}{2}\binom{n_r-2}{2}}{\binom{n_i}{2}\binom{n_r}{2}}\cdot\mathcal{O}\left(\operatorname{tr}^2\left(\frac{N}{n_i}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\frac{N}{n_r}\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right)\right) \\
&\leq \mathcal{O}\left(\frac{1}{n_{\min}}\right)\cdot\mathcal{O}\left(\operatorname{tr}^2\left(\frac{N}{n_i}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\frac{N}{n_r}\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right)\right).
\end{aligned}
$$

Finally, the conditions for A_4 have to be checked. With the expectation values from above we calculate

$$
\begin{aligned}
&\mathbb{E}\left(A_4\right) \\
&\quad= \sum_{i=1}^{a}\frac{N^2}{n_i^2}(\boldsymbol{T}_W)_{ii}{}^2\mathbb{E}(A_{i,3}) + 2\sum_{i=1}^{a-1}\sum_{r=i+1}^{a}\frac{N^2}{n_in_r}(\boldsymbol{T}_W)_{ir}{}^2\mathbb{E}\left(A_{i,r,2}\right) \\
&\quad= \sum_{i=1}^{a}\frac{N^2}{n_i^2}(\boldsymbol{T}_W)_{ii}{}^2\operatorname{tr}\left((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^2\right) + \sum_{i=1}^{a-1}\sum_{r=i+1}^{a}\frac{2N^2}{n_in_r}(\boldsymbol{T}_W)_{ir}{}^2\operatorname{tr}\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right)
\end{aligned}
$$

$$= \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right).$$

To calculate the variances the following additional inequalities are needed:

$$\frac{Var\left(\sum_{i=1}^{a}\left(\frac{N}{n_i}\right)^2 (\boldsymbol{T}_W)_{ii}{}^2 A_{i,3}\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}$$

$$= \sum_{i=1}^{a} \frac{Var\left(\left(\frac{N}{n_i}\right)^2 (\boldsymbol{T}_W)_{ii}{}^2 A_{i,3}\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}$$

$$\leq \sum_{i=1}^{a} \mathcal{O}\left(n_i^{-1}\right) \cdot \frac{\mathcal{O}\left((\boldsymbol{T}_W)_{ii}{}^4 \operatorname{tr}^2\left(\left(\boldsymbol{T}_S \frac{N}{n_i}\boldsymbol{\Sigma}_i\right)^2\right)\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}$$

$$\leq \mathcal{O}\left(\frac{1}{n_{\min}}\right) \frac{\mathcal{O}\left(\operatorname{tr}^2\left(\sum_{i=1}^{a} (\boldsymbol{T}_W)_{ii}{}^2 \left(\boldsymbol{T}_S \frac{N}{n_i}\boldsymbol{\Sigma}_i\right)^2\right)\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} \operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right) = \mathcal{O}\left(\frac{1}{n_{\min}}\right)$$

and

$$\frac{Var\left(2\sum_{r<i\in\mathbb{N}_a} \frac{N^2}{n_i n_r}(\boldsymbol{T}_W)_{ir}{}^2 A_{i,r,2}\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}$$

$$\overset{\text{A.7}}{\leq} 4 \sum_{i<r\in\mathbb{N}_a} \sum_{h<g\in\mathbb{N}_a} \frac{\sqrt{Var\left(\frac{N^2}{n_i n_r}(\boldsymbol{T}_W)_{ir} A_{i,r,2}\right)}\sqrt{Var\left(\frac{N^2}{n_h n_g}(\boldsymbol{T}_W)_{gh} A_{h,g,2}\right)}}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}$$

$$\leq \left(\sum_{i\neq r\in\mathbb{N}_a} \frac{\sqrt{\mathcal{O}\left(\frac{1}{n_{\min}}\right)}(\boldsymbol{T}_W)_{ir}{}^2 \operatorname{tr}\left(\boldsymbol{T}_S \frac{N}{n_i}\boldsymbol{\Sigma}_i \boldsymbol{T}_S \frac{N}{n_r}\boldsymbol{\Sigma}_r\right)}{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right)^2$$

$$\leq \mathcal{O}\left(\frac{1}{n_{\min}}\right) \left(\frac{\mathcal{O}\left(\sum_{i\neq r\in\mathbb{N}_a} (\boldsymbol{T}_W)_{ir}{}^2 \operatorname{tr}\left(\boldsymbol{T}_S \frac{N}{n_i}\boldsymbol{\Sigma}_i \boldsymbol{T}_S \frac{N}{n_r}\boldsymbol{\Sigma}_r\right)\right)}{\sum_{i,r\in\mathbb{N}_a} (\boldsymbol{T}_W)_{ir}{}^2 \operatorname{tr}\left(\boldsymbol{T}_S \frac{N}{n_i}\boldsymbol{\Sigma}_i \frac{N}{n_r}\boldsymbol{T}_S \boldsymbol{\Sigma}_r\right)}\right)^2 \leq \mathcal{O}\left(\frac{1}{n_{\min}}\right).$$

Together this leads to

$$Var\left(\frac{A_4}{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right)$$

$$\overset{A.7}{\leq} \left[\sqrt{\frac{Var\left(2\sum_{r<i\in\mathbb{N}_a}\frac{N^2}{n_i n_r}(\boldsymbol{T}_W)_{ir}{}^2 A_{i,r,2}\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}} + \sqrt{\frac{Var\left(\sum_{i=1}^{a}\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}{}^2 A_{i,3}\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}}\right]^2$$

$$= \left[\sqrt{\mathcal{O}\left(\frac{1}{n_{\min}}\right)} + \sqrt{\mathcal{O}\left(\frac{1}{n_{\min}}\right)}\right]^2 = \mathcal{O}\left(\frac{1}{n_{\min}}\right)$$

and therefore A_4 is an unbiased and ratio-consistent estimator of $\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)$. Moreover, we want to stress that the zero sequences used as upper border for $\widehat{\mathbb{E}_{H_0}}(Q_N)$ and A_4 do not depend on the number of groups or dimensions, so this estimators can be also used for increasing number of groups.

With the expectation values and variances from the beginning it follows directly that $A_{i,1}, A_{i,r,2}, A_{i,3}, A_4$ are unbiased, ratio-consistent estimators of $\operatorname{tr}(\boldsymbol{T}_S\boldsymbol{\Sigma}_i)$,$\operatorname{tr}(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\boldsymbol{T}_S\boldsymbol{\Sigma}_r)$, $\operatorname{tr}\left((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^2\right)$ and $\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)$.

It is worth to note that all of this estimators also consistent estimators which are even dimension-stable in the sense of Brunner et al. [8]. □

For $A_{i,r,2}$ there exists an alternative form which can be implemented substantially more efficient and was considered in Brunner et al. [9]. It uses matrices of the form $\widehat{\boldsymbol{M}}_{i,r} = \boldsymbol{P}_{n_i}\left(\boldsymbol{T}_S\boldsymbol{X}_{i,1},\ldots,\boldsymbol{T}_S\boldsymbol{X}_{i,n_i}\right)^\top \cdot \left(\boldsymbol{T}_S\boldsymbol{X}_{r,1},\ldots,\boldsymbol{T}_S\boldsymbol{X}_{r,n_r}\right)\boldsymbol{P}_{n_r}^\top$. Recalling that $\mathbf{1}_n$ is the vector of ones and $\#$ denotes the Hadamard-Schur-Product, it can be seen that

$$A_{i,r,2} = \frac{\mathbf{1}_{n_i}{}^\top\left(\widehat{\boldsymbol{M}_{i,r}}\#\widehat{\boldsymbol{M}_{i,r}}\right)\mathbf{1}_{n_r}}{(n_i-1)(n_r-1)}.$$

For $A_{i,3}$ there also exists an alternative formula, which expands much longer, but is more efficient:

$$\begin{aligned}
A_{i,3} = &\sum_{\substack{\ell_1,\ell_2=1\\ \ell_1\neq\ell_2}}^{n_i}\frac{\left[\boldsymbol{X}_{i,\ell_1}{}^\top\boldsymbol{T}_S\boldsymbol{X}_{i,\ell_2}\right]^2}{n_i(n_i-1)} - \sum_{\substack{\ell_1,\ell_2,\ell_3=1\\ \ell_3\neq\ell_1,\ell_2}}^{n_i}\frac{\left[\boldsymbol{X}_{i,\ell_1}^\top\boldsymbol{T}_S\boldsymbol{X}_{i,\ell_3}\boldsymbol{X}_{i,\ell_2}^\top\boldsymbol{T}_S(\boldsymbol{X}_{i,\ell_3}+\boldsymbol{X}_{i,\ell_1})\right]}{n_i(n_i-1)(n_i-2)(n_i-3)}\\
&+\sum_{\substack{\ell_1,\ell_2,\ell_3=1\\ \ell_1\neq\ell_2\neq\ell_3}}^{n_i}\frac{\left[\boldsymbol{X}_{i,\ell_1}^\top\boldsymbol{T}_S\boldsymbol{X}_{i,\ell_3}\boldsymbol{X}_{i,\ell_2}^\top\boldsymbol{T}_S\boldsymbol{X}_{i,\ell_2}\right]+(2n_i+5)\cdot\left[\boldsymbol{X}_{i,\ell_1}^\top\boldsymbol{T}_S\boldsymbol{X}_{i,\ell_2}\boldsymbol{X}_{i,\ell_1}^\top\boldsymbol{T}_S\boldsymbol{X}_{i,\ell_3}\right]}{n_i(n_i-1)(n_i-2)(n_i-3)}\\
&-\sum_{\substack{\ell_1,\ell_2,\ell_3=1\\ \ell_1\neq\ell_2}}^{n_i}\frac{\left[\boldsymbol{X}_{i,\ell_1}^\top\boldsymbol{T}_S\boldsymbol{X}_{i,\ell_2}\boldsymbol{X}_{i,\ell_2}^\top\boldsymbol{T}_S\boldsymbol{X}_{i,\ell_3}\right]}{n_i(n_i-1)(n_i-2)(n_i-3)}\\
&-\frac{n_i^2\left[\overline{\boldsymbol{X}}_i^\top\boldsymbol{T}_S\overline{\boldsymbol{X}}_i\right]\left(n_i^2\overline{\boldsymbol{X}}_i^\top\boldsymbol{T}_S\overline{\boldsymbol{X}}_i-\sum_{\ell_1=1}^{n_i}\left[\boldsymbol{X}_{i,\ell_1}^\top\boldsymbol{T}_s\boldsymbol{X}_{i,\ell_1}\right]\right)}{n_i(n_i-1)(n_i-2)(n_i-3)}.
\end{aligned}$$

To finally prove Theorem 3.5 (p.2750) we need another lemma.

Lemma B.1. *For the previously defined estimators it holds for $n_{\min} \to \infty$ that*

$$\frac{\sum_{i=1}^{a} \frac{N}{n_i} (\boldsymbol{T}_W)_{ii} A_{i,1} - \sum_{i=1}^{a} \frac{N}{n_i} (\boldsymbol{T}_W)_{ii} \operatorname{tr}(\boldsymbol{T}_S \boldsymbol{\Sigma}_i)}{\sqrt{2 \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}} \xrightarrow{\mathcal{P}} 0 \quad \textit{independent of } d \textit{ or } a.$$

Proof. We know that

$$\mathbb{E}\left(\sum_{i=1}^{a} \frac{\frac{N}{n_i} (\boldsymbol{T}_W)_{ii} ((A_{i,1}) - \operatorname{tr}(\boldsymbol{T}_S \boldsymbol{\Sigma}_i))}{\sqrt{2 \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}}\right) = \sum_{i=1}^{a} \frac{\frac{N}{n_i} (\boldsymbol{T}_W)_{ii} (\mathbb{E}(A_{i,1}) - \operatorname{tr}(\boldsymbol{T}_S \boldsymbol{\Sigma}_i))}{\sqrt{2 \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}} = 0.$$

Thus,

$$\begin{aligned} &\mathit{Var}\left(\sum_{i=1}^{a} \frac{\frac{N}{n_i} (\boldsymbol{T}_W)_{ii} (A_{i,1} - \operatorname{tr}(\boldsymbol{T}_S \boldsymbol{\Sigma}_i))}{\sqrt{2 \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}}\right) \\ &= \frac{\sum_{i=1}^{a} \frac{N^2}{n_i^2} {(\boldsymbol{T}_W)_{ii}}^2 \mathit{Var}(A_{i,1})}{2 \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} \\ &\overset{\text{Proof of 3.3}}{\leq} \mathcal{O}\left(\frac{1}{n_{\min}}\right) \frac{\sum_{i=1}^{a} \frac{N^2}{n_i^2} {(\boldsymbol{T}_W)_{ii}}^2 \operatorname{tr}\left((2\boldsymbol{T}_S \boldsymbol{\Sigma}_i)^2\right)}{2 \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} = \mathcal{O}\left(\frac{1}{n_{\min}}\right). \end{aligned}$$

In the last step we used the fact that all terms are non-negative and applied the binomial theorem in the last inequality. It is a zero sequence which only depends on $n_{\min}$, so again with Lemma A.6 (p.2770) the result is proved. ☐

Proof of Theorem 3.5 (p.2750). From Lemma A.6 it follows independent of a or d for $n_{\min} \to \infty$ that $A_4 / \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right) \xrightarrow{\mathcal{P}} 1$ and therefore $\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right) / A_4 \xrightarrow{\mathcal{P}} 1$. Moreover, it also follows that $\sqrt{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right) / A_4} \xrightarrow{\mathcal{P}} 1$ and with Lemma B.1 we deduce $\frac{\sum_{i=1}^{a} \frac{N}{n_i} (\boldsymbol{T}_W)_{ii} A_{i,1} - \operatorname{tr}(\boldsymbol{T}\boldsymbol{V}_N)}{\sqrt{2 \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}} \xrightarrow{\mathcal{P}} 0$.

Thus, we can finally calculate the standardized quadratic form as

$$W_N = \frac{Q_N - \sum_{i=1}^{a} \frac{N}{n_i} (\boldsymbol{T}_W)_{ii} A_{i,1}}{\sqrt{2A_4}}$$

$$= \left(\frac{Q_N - \operatorname{tr}(\boldsymbol{T}\boldsymbol{V}_N)}{\sqrt{2\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}} - \frac{\sum_{i=1}^a \frac{N}{n_i}(\boldsymbol{T}_W)_{ii} A_{i,1} - \operatorname{tr}(\boldsymbol{T}\boldsymbol{V}_N)}{\sqrt{2\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}} \right) \cdot \sqrt{\frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}{A_4}}$$

$$= \left(\frac{Q_N - \operatorname{tr}(\boldsymbol{T}\boldsymbol{V}_N)}{\sqrt{2\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}} - o_{\mathcal{P}}(1) \right) \cdot (1 + o_{\mathcal{P}}(1))$$

$$= \widetilde{W}_N + \widetilde{W}_N \cdot o_{\mathcal{P}}(1) - o_{\mathcal{P}}(1) - o_{\mathcal{P}}(1) \cdot o_{\mathcal{P}}(1).$$

The last two parts converge in probability to zero, so also in distribution and with Slutzky $\widetilde{W}_N \cdot o_{\mathcal{P}}(1)$ converges in distribution to zero if one of the conditions of Theorem 3.1 is fulfilled. Thereby W_N has asymptotical the same distribution as $\widetilde{W}_N$.

Replacing the traces by their estimators in the above calculation, it follows with the same arguments that the asymptotic distribution in both cases of local alternatives does not change, since the estimators are also consistent under the alternative. □

For large numbers of groups many estimators $A_{i,1}, A_{i,r,2}$ and $A_{i,3}$ and have to be calculated which leads to long computation time. In this cases it is better to again use subsamling-type estimators which leads to $A^\star_{i,1}, A^\star_{i,r,2}, A^\star_{i,3}$ and therefore to $A^\star_4$.

Lemma B.2. *With the definitions from above let be*

$$A^\star_{i,1}(B) = \frac{1}{2 \cdot B} \sum_{b=1}^{B} \boldsymbol{Y}_{i,\sigma_{i1}(b),\sigma_{i2}(b)}^\top \boldsymbol{Y}_{i,\sigma_{i1}(b),\sigma_{i2}(b)},$$

$$A^\star_{i,r,2}(B) = \frac{1}{4 \cdot B} \sum_{b=1}^{B} \left[\boldsymbol{Y}_{i,\sigma_{i1}(b),\sigma_{i2}(b)}^\top \boldsymbol{Y}_{r,\sigma_{r1}(b),\sigma_{r2}(b)} \right]^2,$$

$$A^\star_{i,3}(B) = \frac{1}{4 \cdot B} \sum_{b=1}^{B} \left[\boldsymbol{Y}_{i,\sigma_{i1}(b),\sigma_{i2}(b)}^\top \boldsymbol{Y}_{i,\sigma_{i3}(b),\sigma_{i4}(b)} \right]^2,$$

$$A^\star_4(B) = \sum_{i=1}^{a} \frac{N^2}{n_i^2} {(\boldsymbol{T}_W)_{ii}}^2 A^\star_{i,3}(B) + 2 \sum_{i=1}^{a} \sum_{r=1, r<i}^{a} \frac{N^2}{n_i n_r} {(\boldsymbol{T}_W)_{ir}}^2 A^\star_{i,r,2}(B).$$

If $B(N) \to \infty$, this estimators and $\sum_{i=1}^a A^\star_{i,1}$ have the same properties as $A_{i,1}, A_{i,r,2}, A_{i,3}$, A_4 and $\sum_{i=1}^a A_{i,1}$ which were defined in Lemma 3.3 (p.2750).

Proof. For $A^\star_{i,1}(B)$, this lemma will be proved in detail. For all other terms only the major steps are shown.

The unbiasedness is clear because the random variables $\sigma_{i1}(b), \sigma_{i2}(b)$ have no influence on the number of terms of the sum and also the terms are identically distributed. Hence,

$$\mathbb{E}\left(A^{\star}_{i,1}(B)\right) = \frac{1}{2\cdot B}\sum_{b=1}^{B}\mathbb{E}\left(\boldsymbol{Y}_{i,\sigma_{i1}(b),\sigma_{i2}(b)}{}^{\top}\boldsymbol{Y}_{i,\sigma_{i1}(b),\sigma_{i2}(b)}\right)$$
$$= \frac{1}{2\cdot B}\sum_{b=1}^{B}\mathbb{E}\left(\boldsymbol{Y}_{i,1,2}{}^{\top}\boldsymbol{Y}_{i,1,2}\right) \overset{A.5}{=} \operatorname{tr}(\boldsymbol{T}_S\boldsymbol{\Sigma}_i).$$

The second part is more complicated. Let $\mathcal{F}(\boldsymbol{\sigma}_i(B,m))$ be the smallest σ-field which contains $\boldsymbol{\sigma}_i(b,m)$ $\forall b \in B$, so obvious $M(B,\boldsymbol{\sigma}_i(b))$ is $\mathcal{F}(\boldsymbol{\sigma}_i(B))$ - measurable. Identical for $\mathcal{F}(\boldsymbol{\sigma}_i(B,m),\boldsymbol{\sigma}_r(B,m))$ and $\mathcal{F}(\boldsymbol{\sigma}(B,m))$. Similar to the previous part, the distribution of the bilinear form does not depend on the index combination. Together with the independence of the normally distributed vectors and $\sigma_{i1}(b),\sigma_{i2}(b)$ this leads to

$$Var\left(\mathbb{E}\left(A^{\star}_{i,1}(B)\middle|\mathcal{F}(\boldsymbol{\sigma}_i(B,2))\right)\right) = Var\left(\operatorname{tr}\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)\right) = 0.$$

With Lemma A.9 (p.2771) we thus obtain

$$Var\left(A^{\star}_{i,1}(B)\right) = 0 + \mathbb{E}\left(Var\left(A^{\star}_{i,1}(B)|\mathcal{F}(\boldsymbol{\sigma}_i(B,2))\right)\right).$$

For the calculation of the conditional variance of the sum, it would be useful finding an upper bound that is based on the variance instead of calculate the covariances. To achieve this, we calculate the number of index combinations which leads to a covariance which is zero. This amount is non-deterministic and we recognize it contains the amount $M(B,\boldsymbol{\sigma}_i(b,2))$ which was considered before.

Again not the amount is important but the number of elements which are contained in $M(B,\boldsymbol{\sigma}_i(b,2))$ since the bilinear forms are identically distributed. Therefore the condition of the variance of the bilinear form disappears since the random indices have no influence on the variance. With the $\mathcal{F}(\boldsymbol{\sigma}_i(B,2))$-measurability of $M(B,\boldsymbol{\sigma}_i(b,2))$ it thus follows that

$$Var\left(A^{\star}_{i,1}(B)\right) = 0 + \mathbb{E}\left(Var\left(A^{\star}_{i,1}(B)|\mathcal{F}(\boldsymbol{\sigma}_i(B,2))\right)\right)$$
$$\overset{A.7}{\leq} \mathbb{E}\left(\sum_{(j,\ell)\in\mathbb{N}_B^2\setminus M(B,(\boldsymbol{\sigma}_i(b,2)))}\frac{Var\left(\boldsymbol{Y}_{i,\sigma_{i1}(j),\sigma_{i2}(j)}{}^{\top}\boldsymbol{Y}_{i,\sigma_{i1}(j),\sigma_{i2}(j)}\middle|\mathcal{F}(\boldsymbol{\sigma}_i(B,2))\right)}{4B^2}\right)$$
$$= \frac{1}{4B^2}\mathbb{E}\left(\sum_{(j,\ell)\in\mathbb{N}_B^2\setminus M(B,(\boldsymbol{\sigma}_i(b,2)))} Var\left(\boldsymbol{Y}_{i,1,2}{}^{\top}\boldsymbol{Y}_{i,1,2}\right)\right)$$
$$\overset{A.5}{=} \frac{\mathbb{E}\left(|\mathbb{N}_B^2\setminus M(B,(\boldsymbol{\sigma}_i(b,2)))|\right)}{B^2}\cdot\frac{\mathcal{O}\left(\operatorname{tr}^2\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)\right)}{4}$$
$$\overset{A.10}{=} \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_i-2}{2}}{\binom{n_i}{2}}\right)\cdot\mathcal{O}\left(\operatorname{tr}^2\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)\right).$$

The other values are calculated in a similar way.

$$\mathbb{E}\left(A^\star_{i,r,2}(B)\right) = \frac{1}{4\cdot B}\sum_{b=1}^{B}\mathbb{E}\left(\left[\boldsymbol{Y}_{i,\sigma_{i1}(b),\sigma_{i2}(b)}{}^\top \boldsymbol{Y}_{r,\sigma_{r1}(b),\sigma_{r2}(b)}\right]^2\right)$$
$$= \frac{1}{4\cdot B}\sum_{b=1}^{B}\mathbb{E}\left(\left[\boldsymbol{Y}_{i,1,2}{}^\top \boldsymbol{Y}_{r,1,2}\right]^2\right) \overset{\text{A.5}}{=} \operatorname{tr}(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\boldsymbol{T}_S\boldsymbol{\Sigma}_r).$$
$$Var\left(\mathbb{E}\left(A^\star_{i,r,2}(B)|\mathcal{F}(\boldsymbol{\sigma}_i(B,2),\boldsymbol{\sigma}_r(B,2))\right)\right) = Var\left(\operatorname{tr}\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right)\right) = 0.$$

$$Var\left(A^\star_{i,r,2}(B)\right) = 0 + \mathbb{E}\left(Var\left(A^\star_{i,r,2}(B)|\mathcal{F}(\boldsymbol{\sigma}_i(B),\boldsymbol{\sigma}_r(B,2))\right)\right)$$
$$\leq \frac{\mathbb{E}\left(|\mathbb{N}_B^2 \setminus M(B,\boldsymbol{\sigma}_i(b,2),\boldsymbol{\sigma}_r(b,2))|\right)}{B^2}\cdot Var\left(\left[\boldsymbol{Y}_{i,1,2}{}^\top \boldsymbol{Y}_{r,1,2}\right]^2\right)$$
$$\overset{\text{A.5}}{\leq} \frac{\mathbb{E}\left(|\mathbb{N}_B^2 \setminus M(B,\boldsymbol{\sigma}_i(b,2),\boldsymbol{\sigma}_r(b,2))|\right)}{B^2}\cdot \mathcal{O}\left(\operatorname{tr}^2\left(\frac{N}{n_i}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\frac{N}{n_r}\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right)\right)$$
$$\overset{\text{A.10}}{=} \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_i-2}{2}\cdot\binom{n_r-2}{2}}{\binom{n_i}{2}\cdot\binom{n_r}{2}}\right)\cdot \mathcal{O}\left(\operatorname{tr}^2\left(\frac{N}{n_i}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\frac{N}{n_r}\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right)\right)$$
$$\leq \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_{\min}-2}{2}^2}{\binom{n_{\min}}{2}^2}\right)\cdot \mathcal{O}\left(\operatorname{tr}^2\left(\frac{N}{n_i}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\frac{N}{n_r}\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right)\right).$$

$$\mathbb{E}\left(A^\star_{i,3}(B)\right) = \frac{1}{4\cdot B}\sum_{b=1}^{B}\mathbb{E}\left(\left[\boldsymbol{Y}_{i,\sigma_{i1}(b),\sigma_{i2}(b)}{}^\top \boldsymbol{Y}_{i,\sigma_{i3}(b),\sigma_{i4}(b)}\right]^2\right)$$
$$= \frac{1}{4\cdot B}\sum_{b=1}^{B}\mathbb{E}\left(\left[\boldsymbol{Y}_{i,1,2}{}^\top \boldsymbol{Y}_{i,1,2}\right]^2\right) \overset{\text{A.5}}{=} \operatorname{tr}\left((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^2\right).$$
$$Var\left(\mathbb{E}\left(A^\star_{i,3}(B)|\mathcal{F}(\boldsymbol{\sigma}_i(B,4))\right)\right) = Var\left(\operatorname{tr}\left((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^2\right)\right) = 0.$$

$$Var\left(A^\star_{i,3}(B)\right)$$
$$= 0 + \mathbb{E}\left(Var\left(A^\star_{i,3}(B)|\mathcal{F}(\boldsymbol{\sigma}_i(B,4))\right)\right)$$
$$\overset{\text{A.7}}{\leq} \mathbb{E}\left(\sum_{(j,\ell)\in\mathbb{N}_B^2\setminus M(B,\boldsymbol{\sigma}_i(b,4))}\frac{Var\left(\left[\boldsymbol{Y}_{i,\sigma_{i1}(j),\sigma_{i2}(j)}{}^\top \boldsymbol{Y}_{i,\sigma_{i3}(j),\sigma_{i4}(j)}\right]^2\Big|\mathcal{F}(\boldsymbol{\sigma}_i(B,4))\right)}{16B^2}\right)$$
$$\overset{\text{A.5}}{\leq} \frac{\mathbb{E}\left(|\mathbb{N}_B^2\setminus M(B,\boldsymbol{\sigma}_i(b,4))|\right)}{B^2}\cdot\frac{\mathcal{O}\left(\operatorname{tr}^2\left((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^2\right)\right)}{16}$$
$$\overset{\text{A.10}}{=} \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_i-4}{4}}{\binom{n_i}{4}}\right)\cdot\mathcal{O}\left(\operatorname{tr}^2\left((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^2\right)\right).$$

$$\mathbb{E}\left(\sum_{i=1}^{a}\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}A^\star_{i,1}\right) = \sum_{i=1}^{a}\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}\mathbb{E}\left(A^\star_{i,1}\right) = \sum_{i=1}^{a}\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}\operatorname{tr}\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right).$$

$$
\begin{aligned}
&Var\left(\frac{\sum\limits_{i=1}^{a}\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}A^{\star}_{i,1}}{\operatorname{tr}\left(\boldsymbol{T}\boldsymbol{V}_N\right)}\right)\\
&=\frac{\sum\limits_{i=1}^{a}\frac{N^2}{n_i^2}{(\boldsymbol{T}_W)_{ii}}^2\ Var\left(A^{\star}_{i,1}\right)}{\operatorname{tr}^2\left(\boldsymbol{T}\boldsymbol{V}_N\right)}\\
&=\frac{\sum\limits_{i=1}^{a}{(\boldsymbol{T}_W)_{ii}}^2\left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_i-2}{2}}{\binom{n_i}{2}}\right)\cdot\mathcal{O}\left(\operatorname{tr}^2\left(\boldsymbol{T}_S\frac{N}{n_i}\boldsymbol{\Sigma}_i\right)\right)}{\operatorname{tr}^2\left(\boldsymbol{T}\boldsymbol{V}_N\right)}\\
&\leq\frac{\sum\limits_{i=1}^{a}{(\boldsymbol{T}_W)_{ii}}^2\left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_{\min}-2}{2}}{\binom{n_{\min}}{2}}\right)\cdot\mathcal{O}\left(\operatorname{tr}^2\left(\boldsymbol{T}_S\frac{N}{n_i}\boldsymbol{\Sigma}_i\right)\right)}{\operatorname{tr}^2\left(\boldsymbol{T}\boldsymbol{V}_N\right)}\\
&\leq\left(1-\frac{\left(1-\frac{1}{B}\right)\cdot\binom{n_{\min}-2}{2}}{\binom{n_{\min}}{2}}\right)\cdot\frac{\mathcal{O}\left(\operatorname{tr}^2\left(\sum\limits_{i=1}^{a}\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)\right)}{\operatorname{tr}^2\left(\boldsymbol{T}\boldsymbol{V}_N\right)}\\
&=\left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_{\min}-2}{2}}{\binom{n_{\min}}{2}}\right)\cdot\mathcal{O}\left(1\right).
\end{aligned}
$$

For $B(N)\to\infty$ the first factor is a zero sequence and therefore $\sum\limits_{i=1}^{a}\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}A^{\star}_{i,1}$ a ratio-consistent, unbiased estimator of $\operatorname{tr}\left(\boldsymbol{T}\boldsymbol{V}_N\right)$.

$$
\begin{aligned}
&\mathbb{E}\left(\sum_{i=1}^{a}\frac{N^2}{n_i^2}{(\boldsymbol{T}_W)_{ii}}^2A^{\star}_{i,3}+\sum_{i\neq r\in\mathbb{N}_a}\frac{N^2}{n_in_r}{(\boldsymbol{T}_W)_{ir}}^2A^{\star}_{i,r,2}\right)\\
&=\sum_{i=1}^{a}\frac{N^2}{n_i^2}{(\boldsymbol{T}_W)_{ii}}^2\mathbb{E}\left(A^{\star}_{i,3}\right)+\sum_{i\neq r\in\mathbb{N}_a}\frac{N^2}{n_in_r}{(\boldsymbol{T}_W)_{ir}}^2\mathbb{E}\left(A^{\star}_{i,r,2}\right)=\operatorname{tr}\left(\left(\boldsymbol{T}\boldsymbol{V}_N\right)^2\right).
\end{aligned}
$$

$$
\begin{aligned}
&Var\left(\frac{\sum\limits_{i=1}^{a}\frac{N^2}{n_i^2}{(\boldsymbol{T}_W)_{ii}}^2A^{\star}_{i,3}}{\operatorname{tr}\left(\left(\boldsymbol{T}\boldsymbol{V}_N\right)^2\right)}\right)\\
&=\frac{\sum\limits_{i=1}^{a}Var\left(\frac{N^2}{n_i^2}{(\boldsymbol{T}_W)_{ii}}^2A^{\star}_{i,3}\right)}{\operatorname{tr}^2\left(\left(\boldsymbol{T}\boldsymbol{V}_N\right)^2\right)}\\
&\leq\frac{\sum\limits_{i=1}^{a}{(\boldsymbol{T}_W)_{ii}}^4\left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_i-4}{4}}{\binom{n_i}{4}}\right)\cdot\mathcal{O}\left(\operatorname{tr}^2\left(\left(\boldsymbol{T}_S\frac{N}{n_i}\boldsymbol{\Sigma}_i\right)^2\right)\right)}{\operatorname{tr}^2\left(\left(\boldsymbol{T}\boldsymbol{V}_N\right)^2\right)}
\end{aligned}
$$

$$\leq \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_{\min}-4}{4}}{\binom{n_{\min}}{4}}\right)\cdot\frac{\sum_{i=1}^{a}(\boldsymbol{T}_W)_{ii}{}^4\mathcal{O}\left(\operatorname{tr}^2\left(\left(\boldsymbol{T}_S\frac{N}{n_i}\boldsymbol{\Sigma}_i\right)^2\right)\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}$$

$$\leq \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_{\min}-4}{4}}{\binom{n_{\min}}{4}}\right)\cdot\frac{\mathcal{O}\left(\operatorname{tr}^2\left(\left(\sum_{i=1}^{a}\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)^2\right)\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}$$

$$\leq \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_{\min}-4}{4}}{\binom{n_{\min}}{4}}\right)\cdot\mathcal{O}(1).$$

$$Var\left(\frac{\sum_{i\neq r\in\mathbb{N}_a}\frac{N^2}{n_in_r}(\boldsymbol{T}_W)_{ir}{}^2A^\star_{i,r,2}}{\operatorname{tr}(\boldsymbol{T}\boldsymbol{V}_N)}\right)$$

$$\leq \left(\sum_{i\neq r\in\mathbb{N}_a}\frac{\sqrt{Var\left(\frac{N^2}{n_in_j}(\boldsymbol{T}_W)_{ir}{}^2A^\star_{i,r,2}\right)}}{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right)^2$$

$$\leq \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_{\min}-2}{2}^2}{\binom{n_{\min}}{2}^2}\right)\cdot\left(\frac{\sum_{i\neq r\in\mathbb{N}_a}(\boldsymbol{T}_W)_{ir}{}^2\sqrt{\mathcal{O}\left(\operatorname{tr}^2\left(\frac{N}{n_i}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\frac{N}{n_r}\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right)\right)}}{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right)^2$$

$$\leq \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_{\min}-2}{2}^2}{\binom{n_{\min}}{2}^2}\right)\cdot\left(\frac{\sum_{i\neq r\in\mathbb{N}_a}\mathcal{O}\left((\boldsymbol{T}_W)_{ir}{}^2\operatorname{tr}\left(\boldsymbol{T}_S\frac{N}{n_i}\boldsymbol{\Sigma}_i\boldsymbol{T}_S\frac{N}{n_r}\boldsymbol{\Sigma}_r\right)\right)}{\sum_{i,r\in\mathbb{N}_a}(\boldsymbol{T}_W)_{ir}{}^2\operatorname{tr}\left(\boldsymbol{T}_S\frac{N}{n_i}\boldsymbol{\Sigma}_i\frac{N}{n_r}\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right)}\right)^2$$

$$\leq \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_{\min}-2}{2}^2}{\binom{n_{\min}}{2}^2}\right)\cdot\mathcal{O}(1).$$

$$Var\left(\frac{\sum_{i=1}^{a}\frac{N^2}{n_i^2}(\boldsymbol{T}_W)_{ii}{}^2A^\star_{i,3}+\sum_{i\neq r\in\mathbb{N}_a}\frac{N^2}{n_in_r}(\boldsymbol{T}_W)_{ir}{}^2A^\star_{i,r,2}}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right)$$

$$\overset{\text{A.7}}{\leq}\left[\sqrt{\frac{Var\left(2\sum_{r<i\in\mathbb{N}_a}\frac{N^2}{n_in_r}(\boldsymbol{T}_W)_{ir}{}^2A^\star_{i,r,2}\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}}+\sqrt{\frac{Var\left(\sum_{i=1}^{a}\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}{}^2A^\star_{i,3}\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}}\right]^2$$

$$\leq \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_{\min}-2}{2}^2}{\binom{n_{\min}}{2}^2}\right)\cdot\mathcal{O}(1).$$

So again this is a zero sequence, and $A_4^\star$ is an unbiased and dimensional stable (i.e. also ratio consistent) estimator of $\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)$. □

Appendix C: Proofs of Section 4

Lemma C.1. *For*

$$\Lambda_1(\ell_{1,1},\ldots,\ell_{6,a}) = \boldsymbol{Z}_{(\ell_{1,1},\ell_{2,1},\ldots,\ell_{1,a},\ell_{2,a})}{}^\top \boldsymbol{T}\boldsymbol{Z}_{(\ell_{3,1},\ell_{4,1},\ldots,\ell_{3,a},\ell_{4,a})},$$

$$\Lambda_2(\ell_{1,1},\ldots,\ell_{6,a}) = \boldsymbol{Z}_{(\ell_{3,1},\ell_{4,1},\ldots,\ell_{3,a},\ell_{4,a})}{}^\top \boldsymbol{T}\boldsymbol{Z}_{(\ell_{5,1},\ell_{6,1},\ldots,\ell_{5,a},\ell_{6,a})},$$

$$\Lambda_3(\ell_{1,1},\ldots,\ell_{6,a}) = \boldsymbol{Z}_{(\ell_{5,1},\ell_{6,1},\ldots,\ell_{5,a},\ell_{6,a})}{}^\top \boldsymbol{T}\boldsymbol{Z}_{(\ell_{1,1},\ell_{2,1},\ldots,\ell_{1,a},\ell_{2,a})},$$

we define

$$C_5 = \sum_{\substack{\ell_{1,1},\ldots,\ell_{6,1}=1\\ \ell_{1,1}\neq\cdots\neq\ell_{6,1}}}^{n_1} \cdots \sum_{\substack{\ell_{1,a},\ldots,\ell_{6,a}=1\\ \ell_{1,a}\neq\cdots\neq\ell_{6,a}}}^{n_a} \frac{\prod_{m=1}^3 \Lambda_m(\ell_{1,1},\ldots,\ell_{6,a})}{8\cdot\prod_{i=1}^a \frac{n_i!}{(n_i-6)!}}.$$

With this notation it follows that

$$\mathbb{E}(C_5) = \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right), \quad Var(C_5) \le \frac{\left(\prod\limits_{i=1}^a \binom{n_i}{6} - \prod\limits_{i=1}^a \binom{n_i-6}{6}\right)}{\prod\limits_{i=1}^a \binom{n_i}{6}} \cdot 27\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right).$$

Proof. Set

$$\widetilde{\boldsymbol{Z}}_{(\ell_{3,1},\ell_{4,1},\ldots,\ell_{3,a},\ell_{4,a})} := \left(\sqrt{2}\boldsymbol{V}_N^{1/2}\right)^{-1} \boldsymbol{Z}_{(\ell_{3,1},\ell_{4,1},\ldots,\ell_{3,a},\ell_{4,a})} \sim \mathcal{N}_{ad}(\boldsymbol{0}_{ad}, \boldsymbol{I}_{ad}).$$

It then follows that

$$\begin{aligned}
&\mathbb{E}\left(\boldsymbol{T}\boldsymbol{Z}_{(\ell_{3,1},\ell_{4,1},\ldots,\ell_{3,a},\ell_{4,a})}\cdot \boldsymbol{Z}_{(\ell_{3,1},\ell_{4,1},\ldots,\ell_{3,a},\ell_{4,a})}{}^\top \boldsymbol{T}^\top\right)\\
&= \mathbb{E}\left(\left(\sqrt{2}\boldsymbol{T}\boldsymbol{V}_N^{1/2}\widetilde{Z}_{(\ell_{3,1},\ell_{4,1},\ldots,\ell_{3,a},\ell_{4,a})}\right)\left(\sqrt{2}\boldsymbol{T}\boldsymbol{V}_N^{1/2}\widetilde{\boldsymbol{Z}}_{(\ell_{3,1},\ell_{4,1},\ldots,\ell_{3,a},\ell_{4,a})}\right)^\top\right)\\
&= 2\boldsymbol{T}\boldsymbol{V}_N^{1/2}\mathbb{E}\left(\widetilde{\boldsymbol{Z}}_{(\ell_{3,1},\ell_{4,1},\ldots,\ell_{3,a},\ell_{4,a})}\widetilde{\boldsymbol{Z}}^\top_{(\ell_{3,1},\ell_{4,1},\ldots,\ell_{3,a},\ell_{4,a})}\right)\boldsymbol{V}_N^{1/2\top}\boldsymbol{T}\\
&= 2\boldsymbol{T}\boldsymbol{V}_N^{1/2}\boldsymbol{I}_{ad}\boldsymbol{V}_N^{1/2\top}\boldsymbol{T} = 2\boldsymbol{T}\boldsymbol{V}_N\boldsymbol{T}.
\end{aligned}$$

With the rules for conditional expectation and the involved independence it follows that

$$\mathbb{E}(C_5) = \sum_{\substack{\ell_{1,1},\ldots,\ell_{6,1}=1\\ \ell_{1,1}\neq\cdots\neq\ell_{6,1}}}^{n_1} \cdots \sum_{\substack{\ell_{1,a},\ldots,\ell_{6,a}=1\\ \ell_{1,a}\neq\cdots\neq\ell_{6,a}}}^{n_a}$$

$$
\begin{aligned}
&\times \frac{\mathbb{E}\left(\Lambda_1(\ell_{1,1},\ldots,\ell_{6,a})\cdot\Lambda_2(\ell_{1,1},\ldots,\ell_{6,a})\cdot\Lambda_3(\ell_{1,1},\ldots,\ell_{6,a})\right)}{8\cdot\prod_{i=1}^{a}\frac{n_i!}{(n_i-6)!}}\\
&= \sum_{\substack{\ell_{1,1},\ldots,\ell_{6,1}=1\\ \ell_{1,1}\neq\cdots\neq\ell_{6,1}}}^{n_1}\cdots\sum_{\substack{\ell_{1,a},\ldots,\ell_{6,a}=1\\ \ell_{1,a}\neq\cdots\neq\ell_{6,a}}}^{n_a}\\
&\quad\times \frac{\mathbb{E}\left(\boldsymbol{Z}_{(1,2)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\cdot\boldsymbol{Z}_{(3,4)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(5,6)}\cdot\boldsymbol{Z}_{(5,6)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(1,2)}\right)}{8\cdot\prod_{i=1}^{a}\frac{n_i!}{(n_i-6)!}}\\
&= \frac{1}{8}\mathbb{E}\left(\boldsymbol{Z}_{(1,2)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\cdot\boldsymbol{Z}_{(3,4)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(5,6)}\cdot\boldsymbol{Z}_{(5,6)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(1,2)}\right)\\
&= \frac{1}{8}\mathbb{E}\left(\mathbb{E}\left(\boldsymbol{Z}_{(1,2)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\cdot\boldsymbol{Z}_{(3,4)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(5,6)}\cdot\boldsymbol{Z}_{(5,6)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(1,2)}\mid \boldsymbol{Z}_{(1,2)}\right)\right)\\
&= \frac{1}{8}\mathbb{E}\left(\boldsymbol{Z}_{(1,2)}{}^{\top}\mathbb{E}\left(\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\cdot\boldsymbol{Z}_{(3,4)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(5,6)}\cdot\boldsymbol{Z}_{(5,6)}{}^{\top}\boldsymbol{T}\right)\boldsymbol{Z}_{(1,2)}\right)\\
&= \frac{4}{8}\mathbb{E}\left(\boldsymbol{Z}_{(1,2)}{}^{\top}\boldsymbol{T}\boldsymbol{V}_N\boldsymbol{T}\boldsymbol{T}\boldsymbol{V}_N\boldsymbol{T}\boldsymbol{Z}_{(1,2)}\right)\\
&= \frac{1}{2}\operatorname{tr}((\boldsymbol{T}\boldsymbol{V}_N\boldsymbol{T}\boldsymbol{T}\boldsymbol{V}_N\boldsymbol{T})2\boldsymbol{V}_N) = \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right).
\end{aligned}
$$

Due to the fact that all $\boldsymbol{X}_{i,j}$ are identically distributed we can neglect the concrete indices, as long as we maintain the structure of dependence of the bilinear forms. The last term fulfills the requirements from Korollar A.5 (p.2769) with $\boldsymbol{Z}_{(1,2)} \sim \mathcal{N}(\boldsymbol{0}_{ad}, 2\boldsymbol{V}_N)$ and the matrix $\boldsymbol{T}\boldsymbol{V}_N\boldsymbol{T}\boldsymbol{T}\boldsymbol{V}_N\boldsymbol{T}$.

For the calculation of the variance it is useful to diagonalize the matrix ${\boldsymbol{V}_N^{1/2}}^{\top}\boldsymbol{T}\boldsymbol{V}_N^{1/2}$: There exists an orthogonal matrix $\boldsymbol{P}$ with $\boldsymbol{P}{\boldsymbol{V}_N^{1/2}}^{\top}\boldsymbol{T}\boldsymbol{V}_N^{1/2}\boldsymbol{P}^{\top} = \boldsymbol{D} = diag(\lambda_1,\ldots,\lambda_{ad})$, where λ_i are the eigenvalues of ${\boldsymbol{V}_N^{1/2}}^{\top}\boldsymbol{T}\boldsymbol{V}_N^{1/2}$. We define $\boldsymbol{J}_i := \boldsymbol{P}\widetilde{\boldsymbol{Z}}_{(i,j)}$ so with the properties of the standard normal distribution $\boldsymbol{J}_i \sim \mathcal{N}_{ad}(\boldsymbol{0}_{ad}, \boldsymbol{I}_{ad})$, where the $\boldsymbol{J}_i$ are independent for different indices. Thus, we can rewrite

$$
\begin{aligned}
\boldsymbol{Z}_{(1,2)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(3,4)} &= \widetilde{\boldsymbol{Z}}_{(1,2)}^{\top}2{\boldsymbol{V}_N^{1/2}}^{\top}\boldsymbol{T}\boldsymbol{V}_N^{1/2}\widetilde{\boldsymbol{Z}}_{(3,4)}\\
&= 2\widetilde{\boldsymbol{Z}}_{(1,2)}^{\top}\boldsymbol{P}^{\top}\boldsymbol{D}\boldsymbol{P}\widetilde{\boldsymbol{Z}}_{(3,4)} = 2\boldsymbol{J}_1^{\top}\boldsymbol{D}\boldsymbol{J}_3.
\end{aligned}
$$

With this argument for all three random variables it follows for the second moment that

$$
\begin{aligned}
&\mathbb{E}\left(\left[\boldsymbol{J}_1^{\top}\boldsymbol{D}\boldsymbol{J}_3\boldsymbol{J}_3^{\top}\boldsymbol{D}\boldsymbol{J}_5\boldsymbol{J}_5^{\top}\boldsymbol{D}\boldsymbol{J}_1\right]^2\right)\\
&\quad= \mathbb{E}\left(\left[\sum_{i=1}^{ad}\lambda_i J_{1i}J_{3i}\right]^2\left[\sum_{j=1}^{ad}\lambda_j J_{3j}J_{5j}\right]^2\left[\sum_{\ell=1}^{ad}\lambda_\ell J_{5\ell}J_{1\ell}\right]^2\right)
\end{aligned}
$$

$$= \sum_{\substack{i_1,j_1,\ell_1=1\\ i_2,j_2,\ell_2=1}}^{ad} \lambda_{i_1}\lambda_{i_2}\lambda_{j_1}\lambda_{j_2}\lambda_{\ell_1}\lambda_{\ell_2}$$
$$\times\, \mathbb{E}\left(J_{1i_1}J_{3i_1}J_{1i_2}J_{3i_2}J_{3j_1}J_{5j_1}J_{3j_2}J_{5j_2}J_{5\ell_1}J_{1\ell_1}J_{5\ell_2}J_{1\ell_2}\right).$$

Now we consider the expectation value for the different combinations. If all indices are equal, it is given by

$$\mathbb{E}\left(J_{11}^4 J_{31}^4 J_{51}^4\right) = 3^3 = 27.$$

Moreover, for $i_1 = i_2 \neq \ell_1 = \ell_2$ and $\ell_2 \neq j_1 = j_2 \neq i_1$ it holds that

$$\mathbb{E}\left(J_{11}^2 J_{31}^2 J_{32}^2 J_{52}^2 J_{13}^2 J_{53}^2\right) = 1^6 = 1.$$

Next, the case $i_1 = i_2 = j_1 = j_2 \neq \ell_1 = \ell_2$ is considered (noting this result can also be used for both analogue combinations):

$$\mathbb{E}\left(J_{11}^2 J_{31}^4 J_{51}^2 J_{12}^2 J_{52}^2\right) = 3^1 \cdot 1^4 = 3.$$

Finally, we consider the combination $i_1 = j_1 = \ell_1 \neq i_2 = j_2 = \ell_2$ and obtain

$$\mathbb{E}\left(\left[J_{11}J_{31}J_{12}J_{32}J_{51}J_{52}\right]^2\right) = \prod_{i=1}^{2} \mathbb{E}\left(J_{1i}^2\right)\mathbb{E}\left(J_{3i}^2\right)\mathbb{E}\left(J_{5i}^2\right) = 1^{3^2}.$$

This is also true for $i_1 = j_2 = \ell_1 \neq i_2 = j_1 = \ell_2$ and the analogue combinations, so, all in all, we have 4 combinations of this kind. All other index combinations lead to expectation zero because in this combinations at least one index appears just one time in the product. Thus, due to independence and the fact that all random variables J_i are centered, it follows that

$$\mathbb{E}\left(\left[\boldsymbol{J}_1^\top \boldsymbol{D}\boldsymbol{J}_3\boldsymbol{J}_3^\top \boldsymbol{D}\boldsymbol{J}_5\boldsymbol{J}_5^\top \boldsymbol{D}\boldsymbol{J}_1\right]^2\right)$$
$$= \sum_{i=1}^{ad} \lambda_i^6 \cdot 27 + \sum_{\substack{i,j=1\\ i\neq j}}^{ad} \lambda_i^3\lambda_j^3 \cdot 1 \cdot 4 + \sum_{\substack{i,j=1\\ i\neq j}}^{d} \lambda_i^2\lambda_j^4 \cdot 9 + \sum_{\substack{i,j,\ell=1\\ i\neq j\neq \ell}}^{ad} \lambda_i^2\lambda_j^2\lambda_\ell^2$$
$$= 23\sum_{i=1}^{ad} \lambda_i^6 + 4\left(\sum_{\substack{i,j=1\\ i\neq j}}^{ad} \lambda_i^3\lambda_j^3 + \sum_{i=j=1}^{ad} \lambda_i^3\lambda_j^3\right) + 9\sum_{\substack{i,j=1\\ i\neq j}}^{ad} \lambda_i^2\lambda_j^4 + \sum_{\substack{i,j,\ell=1\\ i\neq j\neq \ell}}^{ad} \lambda_i^2\lambda_j^2\lambda_\ell^2$$
$$= 17\sum_{i=1}^{ad} \lambda_i^6 + 4\sum_{i,j=1}^{ad} \lambda_i^3\lambda_j^3 + 3\sum_{\substack{i,j=1\\ i\neq j}}^{ad} \lambda_i^2\lambda_j^4$$
$$+ 6\left(\sum_{\substack{i,j=1\\ i\neq j}}^{ad} \lambda_i^2\lambda_j^4 + \sum_{i=1}^{ad} \lambda_i^6\right) + \sum_{\substack{i,j,\ell=1\\ i\neq j\neq \ell}}^{ad} \lambda_i^2\lambda_j^2\lambda_\ell^2$$

$$
\begin{aligned}
&= 17\sum_{i=1}^{ad}\lambda_i^6 + 4\operatorname{tr}^2\left((\boldsymbol{TV}_N)^3\right) + 3\sum_{\substack{i,j=1\\ i\neq j}}^{ad}\lambda_i^2\lambda_j^4 + 6\sum_{i,j=1}^{ad}\lambda_i^2\lambda_j^4 + \sum_{\substack{i,j,\ell=1\\ i\neq j\neq \ell}}^{ad}\lambda_i^2\lambda_j^2\lambda_\ell^2 \\
&\overset{A.2}{\leq} 21\operatorname{tr}^2\left((\boldsymbol{TV}_N)^3\right) + 3\sum_{\substack{i,j=1\\ i\neq j}}^{ad}\lambda_i^2\lambda_j^4 + 6\operatorname{tr}\left((\boldsymbol{TV}_N)^4\right)\operatorname{tr}\left((\boldsymbol{TV}_N)^2\right) \\
&\quad + \sum_{\substack{i,j,\ell=1\\ i\neq j\neq \ell}}^{ad}\lambda_i^2\lambda_j^2\lambda_\ell^2 \\
&\overset{A.2}{\leq} 21\operatorname{tr}^2\left((\boldsymbol{TV}_N)^3\right) + 3\sum_{\substack{i,j=1\\ i\neq j}}^{ad}\lambda_i^2\lambda_j^4 + 6\operatorname{tr}^3\left((\boldsymbol{TV}_N)^2\right) + \sum_{\substack{i,j,\ell=1\\ i\neq j\neq \ell}}^{ad}\lambda_i^2\lambda_j^2\lambda_\ell^2 \\
&\overset{A.2}{\leq} 20\operatorname{tr}^2\left((\boldsymbol{TV}_N)^3\right) + 6\operatorname{tr}^3\left((\boldsymbol{TV}_N)^2\right) \\
&\quad + \left(\sum_{\substack{i,j,\ell=1\\ i\neq j\neq \ell}}^{ad}\lambda_i^2\lambda_j^2\lambda_\ell^2 + 3\sum_{\substack{i,j=1\\ i\neq j}}^{ad}\lambda_i^2\lambda_j^4 + \sum_{i=1}^{ad}\lambda_i^6\right) \\
&= 20\operatorname{tr}^2\left((\boldsymbol{TV}_N)^3\right) + 7\operatorname{tr}^3\left((\boldsymbol{TV}_N)^2\right) \\
&\overset{A.2}{\leq} 20\operatorname{tr}\left((\boldsymbol{TV}_N)^4\right)\operatorname{tr}\left((\boldsymbol{TV}_N)^2\right) + 7\operatorname{tr}^3\left((\boldsymbol{TV}_N)^2\right) \\
&\overset{A.2}{\leq} 27\operatorname{tr}^3\left((\boldsymbol{TV}_N)^2\right).
\end{aligned}
$$

So we can control the variance by

$$
\begin{aligned}
&Var(C_5) \\
&\overset{A.7}{\leq} \frac{Var\left(\Lambda_1(1,2,3,4,5,6,\ldots,5,6)\cdot\Lambda_2(1,2,3,4,5,6,\ldots,5,6)\cdot\Lambda_3(1,2,3,4,5,6,\ldots,5,6)\right)}{64\cdot\prod_{i=1}^{a}\binom{n_i}{6}\cdot\left(\prod_{i=1}^{a}\binom{n_i}{6}-\prod_{i=1}^{a}\binom{n_i-6}{6}\right)^{-1}} \\
&\leq \frac{\mathbb{E}\left([\Lambda_1(1,2,3,4,5,6,\ldots,5,6)\cdot\Lambda_2(1,2,3,4,5,6,\ldots,5,6)\cdot\Lambda_3(1,2,3,4,5,6,\ldots,5,6)]^2\right)}{64\cdot\prod_{i=1}^{a}\binom{n_i}{6}\cdot\left(\prod_{i=1}^{a}\binom{n_i}{6}-\prod_{i=1}^{a}\binom{n_i-6}{6}\right)^{-1}} \\
&= \frac{\mathbb{E}\left(\left[2^3\cdot\boldsymbol{J}_1^\top\boldsymbol{D}\boldsymbol{J}_3\boldsymbol{J}_3^\top\boldsymbol{D}\boldsymbol{J}_5\boldsymbol{J}_5^\top\boldsymbol{D}\boldsymbol{J}_1\right]^2\right)}{64\cdot\prod_{i=1}^{a}\binom{n_i}{6}\cdot\left(\prod_{i=1}^{a}\binom{n_i}{6}-\prod_{i=1}^{a}\binom{n_i-6}{6}\right)^{-1}} \\
&\leq \frac{\left(\prod_{i=1}^{a}\binom{n_i}{6}-\prod_{i=1}^{a}\binom{n_i-6}{6}\right)}{\prod_{i=1}^{a}\binom{n_i}{6}}\cdot 27\operatorname{tr}^3\left((\boldsymbol{TV}_N)^2\right). \qquad \square
\end{aligned}
$$

With this result, we can construct an estimator for τ_P step by step:

Lemma C.2. *For C_5 as previously defined, it holds for fixed a that*

$$\frac{C_5}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} - \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} \xrightarrow{\mathcal{P}} 0 \qquad \min(d, n_{\min}) \to \infty.$$

It even holds in the asymptotic frameworks (4)–(5) if $q > 1$ exists with $n_{\min} = \mathcal{O}(a^q)$.

Proof. From the previous lemma, we know that

$$\begin{aligned}
&\mathbb{E}\left(\frac{C_5}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} - \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right) \\
&\quad = \mathbb{E}\left(\frac{C_5}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right) - \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} = 0, \\
&\mathit{Var}\left(\frac{C_5}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} - \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right) \\
&\quad = \frac{\mathit{Var}(C_5)}{\operatorname{tr}^{3}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} \overset{\text{C.1}}{\leq} 27 \cdot \frac{\left(\prod\limits_{i=1}^{a}\binom{n_i}{6} - \prod\limits_{i=1}^{a}\binom{n_i-6}{6}\right)}{\prod\limits_{i=1}^{a}\binom{n_i}{6}}.
\end{aligned}$$

For fixed a this is a zero sequence. If we consider $a \to \infty$ we need the existence of $q > 1$ and $n_{\min} = \mathcal{O}(a^q)$ to guarantee that the upper border is a zero sequence. So in both cases Lemma A.6 (p.2770) can be used. □

Lemma C.3. *Moreover C_5 holds for fixed a*

$$\frac{C_5^2}{\operatorname{tr}^{3}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} - \tau_P \xrightarrow{\mathcal{P}} 0 \qquad d, n_{\min} \to \infty.$$

If $q > 1$ exists with $n_{\min} = \mathcal{O}(a^q)$, the convergence even holds in the asymptotic frameworks (4)–(5).

Proof. With the last lemma it follows for both cases that

$$\begin{aligned}
\frac{C_5^2}{\operatorname{tr}^{3}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} - \tau_P &= \left(\frac{C_5}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right)^2 - \left(\frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right)^2 \\
&= \left[\frac{C_5}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} - \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right]
\end{aligned}$$

$$\times \left[\frac{C_5}{\operatorname{tr}^{3/2}\left((\boldsymbol{TV}_N)^2\right)} + \frac{\operatorname{tr}\left((\boldsymbol{TV}_N)^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{TV}_N)^2\right)} \right]$$
$$= \mathcal{O}_{\mathcal{P}}(1) \cdot \left[\frac{C_5}{\operatorname{tr}^{3/2}\left((\boldsymbol{TV}_N)^2\right)} - \frac{\operatorname{tr}\left((\boldsymbol{TV}_N)^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{TV}_N)^2\right)} + 2\sqrt{\tau_P} \right]$$
$$= \mathcal{O}_{\mathcal{P}}(1) \cdot \left[\mathcal{O}_{\mathcal{P}}(1) + 2\sqrt{\tau_P}\right] = \mathcal{O}_{\mathcal{P}}(1).$$

For the last step we used that $\tau_P \in [0,1]$ which is known from Lemma A.8 (p.2770) and hence $\operatorname{tr}\left((\boldsymbol{TV}_N)^3\right) \Big/ \operatorname{tr}^{3/2}\left((\boldsymbol{TV}_N)^2\right) = \sqrt{\tau_P} \in [-1,1]$. As a product of a bound term and a term which converges to zero in probability, it also converges to zero in probability and with Slutzky's Lemma the result follows. □

Proof of Lemma 4.2. From Lemma 3.3 (p.2750) together with Lemma A.6 (p.2770) it follows

$$\frac{A_4}{\operatorname{tr}\left((\boldsymbol{TV}_N)^2\right)} \xrightarrow{\mathcal{P}} 1 \quad \text{and therefore} \quad \frac{\operatorname{tr}^3\left((\boldsymbol{TV}_N)^2\right)}{A_4^3} \xrightarrow{\mathcal{P}} 1 \quad \text{for} \quad n_{\min} \to \infty,$$

independent of d or a. With Lemma C.3 (p.2791) it follows

$$\frac{C_5^2}{\operatorname{tr}^3\left((\boldsymbol{TV}_N)^2\right)} - \tau_P \xrightarrow{\mathcal{P}} 0 \qquad \text{for} \quad d, n_{\min} \to \infty$$

or under the additional condition also in the asymptotic frameworks (4)–(5).

With these limits in both cases we can calculate

$$\frac{C_5^2}{A_4^3} - \tau_P = \frac{C_5^2}{\operatorname{tr}^3\left((\boldsymbol{TV}_N)^2\right)} \cdot \frac{\operatorname{tr}^3\left((\boldsymbol{TV}_N)^2\right)}{A_4^3} - \tau_P$$
$$= \frac{C_5^2}{\operatorname{tr}^3\left((\boldsymbol{TV}_N)^2\right)} \cdot (1 + \mathcal{O}_{\mathcal{P}}(1)) - \tau_P$$
$$= \frac{C_5^2}{\operatorname{tr}^3\left((\boldsymbol{TV}_N)^2\right)} - \tau_P + \left(\frac{C_5^2}{\operatorname{tr}^3\left((\boldsymbol{TV}_N)^2\right)} - \tau_P + \tau_P \right) \cdot \mathcal{O}_{\mathcal{P}}(1)$$
$$= \mathcal{O}_{\mathcal{P}}(1) + \mathcal{O}_{\mathcal{P}}(1) \cdot \mathcal{O}_{\mathcal{P}}(1) + \tau_P \cdot \mathcal{O}_{\mathcal{P}}(1) = \mathcal{O}_{\mathcal{P}}(1).$$

As in the previous lemma we used $\tau_P \in [0,1]$ and Slutzky. □

For $C_5^\star$ the properties are shown in a similar way as in Lemma B.2 (p.2782).

Lemma C.4. *For*

$$\Lambda_1(\ell_{1,1},\ldots,\ell_{6,a}) = \boldsymbol{Z}_{(\ell_{1,1},\ell_{2,1},\ldots,\ell_{1,a},\ell_{2,a})}{}^\top \boldsymbol{TZ}_{(\ell_{3,1},\ell_{4,1},\ldots,\ell_{3,a},\ell_{4,a})},$$
$$\Lambda_2(\ell_{1,1},\ldots,\ell_{6,a}) = \boldsymbol{Z}_{(\ell_{3,1},\ell_{4,1},\ldots,\ell_{3,a},\ell_{4,a})}{}^\top \boldsymbol{TZ}_{(\ell_{5,1},\ell_{6,1},\ldots,\ell_{5,a},\ell_{6,a})},$$

$$\Lambda_3(\ell_{1,1},\ldots,\ell_{6,a}) = \boldsymbol{Z}_{(\ell_{5,1},\ell_{6,1},\ldots,\ell_{5,a},\ell_{6,a})}{}^\top \boldsymbol{T}\boldsymbol{Z}_{(\ell_{1,1},\ell_{2,1},\ldots,\ell_{1,a},\ell_{2,a})},$$

define

$$C_5^\star(B) = \frac{1}{8\cdot B}\sum_{b=1}^{B} \Lambda_1(\boldsymbol{\sigma}(b,6))\cdot\Lambda_2(\boldsymbol{\sigma}(b,6))\cdot\Lambda_3(\boldsymbol{\sigma}(b,6)).$$

Then it holds

$$\mathbb{E}\left(C_5^\star(B)\right) = \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right),$$
$$Var\left(C_5^\star(B)\right) \le \left(1-\left(1-\frac{1}{B}\right)\cdot\prod_{i=1}^{a}\frac{\binom{n_i-6}{6}}{\binom{n_i}{6}}\right)\cdot 27\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right).$$

Proof. With the same steps as in the previous lemma and by using the fact that expectation and variance do not depend on the concrete indices but rather on the structure of independences we get

$$\begin{aligned}
\mathbb{E}\left(C_5^\star(B)\right) &= \frac{1}{8B}\sum_{b=1}^{B}\mathbb{E}\left(\Lambda_1(\boldsymbol{\sigma}(b,6))\cdot\Lambda_2(\boldsymbol{\sigma}(b,6))\cdot\Lambda_3(\boldsymbol{\sigma}(b,6))\right)\\
&= \frac{1}{8B}\sum_{b=1}^{B}\mathbb{E}\left(\Lambda_1(\ell_{1,1},\ldots,\ell_{6,a})\cdot\Lambda_2(\ell_{1,1},\ldots,\ell_{6,a})\cdot\Lambda_3(\ell_{1,1},\ldots,\ell_{6,a})\right).\\
&\overset{C.1}{=} \frac{1}{8B}\sum_{b=1}^{B}\operatorname{tr}\left((2\boldsymbol{T}\boldsymbol{V}_N)^3\right) = \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right).
\end{aligned}$$

$$Var(\mathbb{E}(C_5^\star(B)|\mathcal{F}(\boldsymbol{\sigma}(B,6)))) = Var\left(\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)\right) = 0.$$

$$\begin{aligned}
Var\left(C_5^\star(B)\right) &= 0 + \mathbb{E}\left(Var\left(C_5^\star(B)|\mathcal{F}(\boldsymbol{\sigma}(B,6))\right)\right)\\
&\overset{A.7}{\le} \mathbb{E}\Bigg(\sum_{(j,\ell)\in\mathbb{N}_B^2\setminus M(B,\boldsymbol{\sigma}(b,6))}\\
&\qquad \times \frac{Var\left(\Lambda_1(\boldsymbol{\sigma}(j,6))\Lambda_2(\boldsymbol{\sigma}(j,6))\Lambda_3(\boldsymbol{\sigma}(j,6))|\mathcal{F}(\boldsymbol{\sigma}(B,6))\right)}{64B^2}\Bigg)\\
&= \frac{\mathbb{E}\left(|\mathbb{N}_B^2\setminus M(B,\boldsymbol{\sigma}(b,6))|\right)}{B^2}\\
&\qquad \cdot \frac{Var\left(\boldsymbol{Z}_{(1,2)}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\cdot\boldsymbol{Z}_{(3,4)}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(5,6)}\cdot\boldsymbol{Z}_{(5,6)}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(1,2)}\right)}{64}\\
&\overset{C.1}{\le} \left(1-\left(1-\frac{1}{B}\right)\cdot\prod_{i=1}^{a}\frac{\binom{n_i-6}{6}}{\binom{n_i}{6}}\right)\cdot 27\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right).
\end{aligned}$$
□

Proof of Theorem 4.3 (p.2753). With Lemma C.4 we recognize $\tau_P \to 1 \Leftrightarrow \widehat{\tau_P} \xrightarrow{\mathcal{P}} 1$ and $\tau_P \to 0 \Leftrightarrow \widehat{\tau_P} \xrightarrow{\mathcal{P}} 0$. Therefore $f_P \to 1 \Leftrightarrow \widehat{f_P} \xrightarrow{\mathcal{P}} 1$ and $f_P \to \infty \Leftrightarrow \widehat{f}_P \xrightarrow{\mathcal{P}} \infty$. This is the only condition needed for the proof of Pauly et al. [34][Theorem 3.1], so the result follows. □

Although $n_{\min} = \mathcal{O}(a^q)$ with $q > 1$ is not too critical in most settings we additionally developed an estimator which can be used without any restrictions.

For this estimator another random vector has to be introduced: The random vector $\pi_{j,i}$ represents a random permutation of the numbers $1, \ldots, n_i$, where $\pi_{j,i}$ are independent for different i or j and $\pi_{j,i}(l)$ denotes its l-th element. Then we define

$$C_7(w) = \frac{1}{w}\sum_{j=1}^{w}\sum_{\ell_1\neq\cdots\neq\ell_6=1}^{n_{\min}} \frac{\Lambda_4(j;\ell_1,\ldots,\ell_6)\cdot\Lambda_5(j;\ell_1,\ldots,\ell_6)\cdot\Lambda_6(j;\ell_1,\ldots,\ell_6)}{8\cdot\frac{n_{\min}!}{(n_{\min}-6)!}}$$

with

$$\begin{aligned}\Lambda_4(j;\ell_1,\ldots,\ell_6) &= {\boldsymbol{Z}^{\boldsymbol{\pi}_j}_{(\ell_1,\ell_2)}}^\top \boldsymbol{T}\boldsymbol{Z}^{\boldsymbol{\pi}_j}_{(\ell_3,\ell_4)},\\ \Lambda_5(j;\ell_1,\ldots,\ell_6) &= {\boldsymbol{Z}^{\boldsymbol{\pi}_j}_{(\ell_3,\ell_4)}}^\top \boldsymbol{T}\boldsymbol{Z}^{\boldsymbol{\pi}_j}_{(\ell_5,\ell_6)},\\ \Lambda_6(j;\ell_1,\ldots,\ell_6) &= {\boldsymbol{Z}^{\boldsymbol{\pi}_j}_{(\ell_5,\ell_6)}}^\top \boldsymbol{T}\boldsymbol{Z}^{\boldsymbol{\pi}_j}_{(\ell_1,\ell_2)}.\end{aligned}$$

and

$$\boldsymbol{Z}^{\boldsymbol{\pi}_j}_{(\ell_1,\ell_2)} := \boldsymbol{Z}_{(\pi_{j,1}(\ell_1),\pi_{j,1}(\ell_2),\pi_{j,2}(\ell_1),\ldots,\pi_{j,a}(\ell_1),\pi_{j,a}(\ell_2))}$$

This estimator again uses $\boldsymbol{Z}$, but different to C_5 the indices are the same for all groups. However the highest index is $n_{\min}$ and some index combinations are unachievable. For this reason, the above random permutations were used. So first the observations in each group were rearranged randomly and with this rearranged samples we calculated the sum of the used terms. Thereafter, we again rearrange the observations and the same terms as before are calculated. If these values were summed up and divided by the number of rearrangements we get an alternative for C_5 which is shown in the following lemma.

Lemma C.5. *For C_7 as defined before it holds*

$$\begin{aligned}\mathbb{E}(C_7(w)) &= \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)\\ Var(C_7(w)) &\leq \left(\frac{\frac{n_{\min}!}{(n_{\min}-6)!} - \frac{(n_{\min}-6)!}{(n_{\min}-12)!}}{\frac{n_{\min}!}{(n_{\min}-6)!}}\right)\cdot\mathcal{O}\left(\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)\right).\end{aligned}$$

Proof. Again we calculate

$$\begin{aligned}\mathbb{E}(C_7(w)) &= \frac{1}{w}\sum_{j=1}^{w}\sum_{\ell_1\neq\cdots\neq\ell_6=1}^{n_{\min}} \frac{\mathbb{E}\left(\prod_{m=4}^{6}\Lambda_m(j;\ell_1,\ldots,\ell_6)\right)}{8\cdot\frac{n_{\min}!}{(n_{\min}-6)!}}\\ &= \frac{1}{w}\sum_{j=1}^{w}\sum_{\ell_1\neq\cdots\neq\ell_6=1}^{n_{\min}} \frac{\mathbb{E}\left(\prod_{m=4}^{6}\Lambda_m(j;1,\ldots,6)\right)}{8\cdot\frac{n_{\min}!}{(n_{\min}-6)!}} = \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right).\end{aligned}$$

Because of the fact that all groups use the same indices, the number of remaining index combinations simplifies and we receive

$$\begin{aligned}
&Var\left(\sum_{\ell_1\neq\cdots\neq\ell_6=1}^{n_{\min}}\frac{\prod_{m=4}\Lambda_m\left(j;\ell_1,\ldots,\ell_6\right)}{8\cdot\frac{n_{\min}!}{(n_{\min}-6)!}}\right)\\
&\leq\frac{\frac{n_{\min}!}{(n_{\min}-6)!}-\frac{(n_{\min}-6)!}{(n_{\min}-12)!}}{\frac{n_{\min}!}{(n_{\min}-6)!}}\cdot Var\left(\prod_{m=4}^{6}\Lambda_m\left(j;\ell_1,\ldots,\ell_6\right)\right)\\
&\leq\frac{\frac{n_{\min}!}{(n_{\min}-6)!}-\frac{(n_{\min}-6)!}{(n_{\min}-12)!}}{\frac{n_{\min}!}{(n_{\min}-6)!}}\cdot\mathcal{O}\left(\operatorname{tr}^3\left(\left(\boldsymbol{T}\boldsymbol{V}_N\right)^2\right)\right).
\end{aligned}$$

For the sum this leads to

$$\begin{aligned}
&Var\left(C_7\left(w\right)\right)\\
&=Var\left(\frac{1}{w}\sum_{j=1}^{w}\sum_{\ell_1\neq\cdots\neq\ell_6=1}^{n_{\min}}\frac{\prod_{m=4}^{6}\Lambda_m\left(j;\ell_1,\ldots,\ell_6\right)}{8\cdot\frac{n_{\min}!}{(n_{\min}-6)!}}\right)\\
&\overset{A.7}{\leq}\frac{1}{w^2}\sum_{j_1,j_2=1}^{w}Var\left(\sum_{\ell_1\neq\cdots\neq\ell_6=1}^{n_{\min}}\frac{\prod_{m=4}^{6}\Lambda_m\left(j;\ell_1,\ldots,\ell_6\right)}{8\cdot\frac{n_{\min}!}{(n_{\min}-6)!}}\right)\\
&\leq\frac{1}{w^2}\sum_{j_1,j_2=1}^{w}\left(\frac{\frac{n_{\min}!}{(n_{\min}-6)!}-\frac{(n_{\min}-6)!}{(n_{\min}-12)!}}{\frac{n_{\min}!}{(n_{\min}-6)!}}\right)\cdot\mathcal{O}\left(\operatorname{tr}^3\left(\left(\boldsymbol{T}\boldsymbol{V}_N\right)^2\right)\right)\\
&=\left(\frac{\frac{n_{\min}!}{(n_{\min}-6)!}-\frac{(n_{\min}-6)!}{(n_{\min}-12)!}}{\frac{n_{\min}!}{(n_{\min}-6)!}}\right)\cdot\mathcal{O}\left(\operatorname{tr}^3\left(\left(\boldsymbol{T}\boldsymbol{V}_N\right)^2\right)\right). \qquad\square
\end{aligned}$$

Simulations (not shown here) show that higher values for w lead to better estimations.

Lemma C.6. *For C_7 as previously defined, it holds*

$$\frac{C_7^2}{\operatorname{tr}^3\left(\left(\boldsymbol{T}\boldsymbol{V}_N\right)^2\right)}-\tau_P\xrightarrow{\mathcal{P}}0\qquad\text{for}\quad n_{\min}\to\infty,$$

independent of a or d. Therefore this holds for the asymptotic frameworks (3)–(5).

Proof. With the previous lemma we know

$$\mathbb{E}\left(\frac{C_7(w)-\operatorname{tr}\left(\left(\boldsymbol{T}\boldsymbol{V}_N\right)^3\right)}{\operatorname{tr}^{3/2}\left(\left(\boldsymbol{T}\boldsymbol{V}_N\right)^2\right)}\right)$$

$$= \mathbb{E}\left(\frac{C_7(w)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right) - \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} = 0,$$

$$Var\left(\frac{C_7(w) - \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right)$$

$$= \frac{Var\left(C_7(w)\right)}{\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} \leq \left(\frac{\frac{n_{\min}!}{(n_{\min}-6)!} - \frac{(n_{\min}-6)!}{(n_{\min}-12)!}}{\frac{n_{\min}!}{(n_{\min}-6)!}}\right) \cdot \mathcal{O}(1).$$

So exactly the same steps as in the proof of Lemma 4.2, which in this case uses that the zero sequence not depends on a or d, leads to the result. □

But for the calculation of this estimator we need $w \cdot n_{\min}!/(n_{\min} - 6)!$ summations. Thus, a subsampling-type version of C_7 is necessary which is now defined.

Lemma C.7. *For each* $b = 1, \ldots, B$ *we independently draw random subsamples* $\boldsymbol{\sigma}_0(b, 6)$ *of length* 6 *from* $\{1, \ldots, n_{\min}\}$ *and define*

$$C_7^\star(w, B) = \sum_{j=1}^{w}\sum_{b=1}^{B} \frac{\Lambda_4\left(j; \boldsymbol{\sigma}_0(b,6)\right)\Lambda_5\left(j; \boldsymbol{\sigma}_0(b,6)\right)\Lambda_6\left(j; \boldsymbol{\sigma}_0(b,6)\right)}{8wB}$$

which holds

$$\mathbb{E}\left(C_7^\star(w,B)\right) = \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)\right),$$

$$Var\left(C_7^\star(w,B)\right) = \left(1 - \left(1 - \frac{1}{B}\right)\frac{\binom{n_{\min}-6}{6}}{\binom{n_{\min}}{6}}\right) 27 \operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right).$$

Proof. The proof for this subsampling-type estimator takes the same steps as before, with another amount $M(B, \boldsymbol{\sigma}_0(b, 6))$. At the beginning we calculate expectation value and an upper bound for the variance of the inner sum. We get

$$\mathbb{E}\left(\sum_{b=1}^{B} \frac{\prod_{m=4}^{6}\Lambda_m\left(j; \boldsymbol{\sigma}_0(b,6)\right)}{8B}\right) = \sum_{b=1}^{B}\frac{\mathbb{E}\left(\prod_{m=4}^{6}\Lambda_m\left(j; 1, \ldots, 6\right)\right)}{8B}$$

$$= \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right).$$

$$Var\left(\mathbb{E}\left(\sum_{b=1}^{B}\frac{\prod_{m=4}^{6}\Lambda_m\left(j; \boldsymbol{\sigma}_0(b,6)\right)}{8B}\Big|\mathcal{F}\left(\boldsymbol{\sigma}_0(B)\right)\right)\right) = Var\left(\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)\right) = 0.$$

$$Var\left(\sum_{b=1}^{B}\frac{\prod_{m=4}^{6}\Lambda_m\left(j; \boldsymbol{\sigma}_0(b,6)\right)}{8B}\right)$$

$$
\begin{aligned}
&= 0 + \mathbb{E}\left(Var\left(\sum_{b=1}^{B} \frac{\prod_{m=4}^{6} \Lambda_m\left(j; \boldsymbol{\sigma}_0(b,6)\right)}{8B} \Big| \mathcal{F}\left(\boldsymbol{\sigma}_0(B)\right) \right) \right) \\
&\overset{A.7}{\leq} \mathbb{E}\left(\sum_{(b_1,b_2)\in\mathbb{N}_B^2 \setminus M(B,\boldsymbol{\sigma}_0(b,6))} Var\left(\prod_{m=4}^{6} \Lambda_m\left(j; \boldsymbol{\sigma}_0(b_1,6)\right) \Big| \mathcal{F}\left(\boldsymbol{\sigma}_0(B)\right) \right) \cdot \frac{1}{64B^2} \right) \\
&= \frac{\mathbb{E}\left(|\mathbb{N}_B^2 \setminus M\left(B, \boldsymbol{\sigma}_0(b,6)\right)| \right)}{B^2} \\
&\quad \cdot \frac{Var\left(\Lambda_4\left(j;1,\ldots,6\right) \cdot \Lambda_5\left(j;1,\ldots,6\right) \cdot \Lambda_6\left(j;1,\ldots,6\right)\right)}{64} \\
&\overset{C.1}{\leq} \left(1 - \left(1 - \frac{1}{B}\right) \cdot \frac{\binom{n_{\min}-6}{6}}{\binom{n_{\min}}{6}} \right) \cdot 27 \operatorname{tr}^3\left((\boldsymbol{TV}_N)^2 \right).
\end{aligned}
$$

With these values we can consider the whole estimator

$$
\begin{aligned}
\mathbb{E}\left(C_7^\star(w,B)\right) &= \frac{1}{w} \sum_{j=1}^{w} \mathbb{E}\left(\sum_{b=1}^{B} \frac{\prod_{m=4}^{6} \Lambda_m\left(j; \boldsymbol{\sigma}_0(b,6)\right)}{8B} \Big| \mathcal{F}\left(\boldsymbol{\sigma}_0(B)\right) \right) \\
&= \operatorname{tr}\left((\boldsymbol{TV}_N)^3 \right), \\
Var\left(C_7^\star(w,B)\right) &\leq \frac{1}{w^2} \left(\sum_{j=1}^{w} \sqrt{ Var\left(\sum_{b=1}^{B} \frac{\prod_{m=4}^{6} \Lambda_m\left(j; \boldsymbol{\sigma}_0(b,6)\right)}{8B} \right) } \right)^2 \\
&\leq \frac{1}{w^2} \left(\sum_{j=1}^{w} \sqrt{ \left(1 - \left(1 - \frac{1}{B}\right) \cdot \frac{\binom{n_{\min}-6}{6}}{\binom{n_{\min}}{6}} \right) \cdot 27 \operatorname{tr}^3\left((\boldsymbol{TV}_N)^2 \right) } \right)^2 \\
&= \left(1 - \left(1 - \frac{1}{B}\right) \cdot \frac{\binom{n_{\min}-6}{6}}{\binom{n_{\min}}{6}} \right) \cdot 27 \operatorname{tr}^3\left((\boldsymbol{TV}_N)^2 \right). \qquad \square
\end{aligned}
$$

The next lemma shows that the version of the estimators with random indices has all the properties the classical ones possess.

Lemma C.8. *The statements of Lemma B.1, Lemma C.2, Lemma C.3, Lemma 4.2 and Lemma C.6 are also true, if all or only a part of the estimators are replaced by the subsampling-type estimators.*

Moreover, Theorem 3.1, Theorem 3.5 and Theorem 4.3 hold, if all or only a part of the estimators are replaced by the subsampling-type estimators.

Proof. For the proofs of the classical estimators from the first paragraph, only the expectation values are used together with upper bounds for the variances which are zero sequences. With random indices, the expectation is the same and for the variance, all traces are the same but the zero sequence changes. So the proofs of the subsampling-type estimators work identically.

For the second paragraph, only some convergences are necessary, which the subsampling-type estimators also fulfills. □

Appendix D: On the asymptotic distribution in our simulation designs

To determine the asymptotic distribution of our test statistic (corresponding to validity of the different tests) in our simulation settings, the asymptotic behaviour of β_1 has to be investigated. Due to equivalence we calculate the value of $\tau_P = \operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)\Big/\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)$. This is sufficient since $\boldsymbol{V}_N$ is known, i.e. no estimation is needed. The ratio n_1/N and n_2/N are the same for all our sample sizes, so the different numbers n_1, n_2 have no influence on the values of τ_P. Results for different choices of $\boldsymbol{T}$ and $\boldsymbol{\Sigma}_i, i = 1, 2$, corresponding to the simulation settings from Section 5 are displayed in Tables 3–5. It can be seen that for H_0^a (Table 3) we have $\tau_P \to 1$ and thus $\beta_1 \to 1$ by Lemma A.8. For H_0^b (Table 4) we have $\tau_P \to 0$ and thus $\beta_1 \to 0$; and in case of the autoregressive covariance matrices with correlation factor depending on the the dimension, we seem to have $\beta_1 \to b_1 \approx 0.7589$.

Table 3
τ_P for $\boldsymbol{T} = \left(\boldsymbol{P}_2 \otimes \frac{1}{d}\boldsymbol{J}_d\right)$ with $(\boldsymbol{\Sigma}_1)_{i,j} = 0.6^{|i-j|}$ and $(\boldsymbol{\Sigma}_2)_{i,j} = 0.65^{|i-j|}$

d	5	10	20	40	70	100	150	200	300	450	600	800
τ_P	1	1	1	1	1	1	1	1	1	1	1	1

Table 4
τ_P for $\boldsymbol{T} = \left(\frac{1}{2}\boldsymbol{J}_2 \otimes \boldsymbol{P}_d\right)$ and $\boldsymbol{T} = (\boldsymbol{P}_2 \otimes \boldsymbol{P}_d)$ with $(\boldsymbol{\Sigma}_1)_{i,j} = 0.6^{|i-j|}$ and $(\boldsymbol{\Sigma}_2)_{i,j} = 0.65^{|i-j|}$

d	5	10	20	40	70	100	150	200	300	450	600	800
τ_P	.49	.35	.21	.11	.061	.043	.029	.021	.014	.0095	.0071	.0053

Table 5
τ_P and β_1 for $\boldsymbol{T} = \left(\frac{1}{2}\boldsymbol{J}_2 \otimes \boldsymbol{P}_d\right)$ with $(\boldsymbol{\Sigma}_1)_{i,j} = 0.6^{|i-j|/d}$ and $(\boldsymbol{\Sigma}_2)_{i,j} = 0.65^{|i-j|/d}$

d	5	10	20	40	70	100	150	200	300	450	800
τ_P	.9311	.9408	.9444	.9454	.9457	.9457	.9458	.9458	.9458	.9458	.9458
β_1	.7082	.7392	.7534	.7575	.7584	.7587	.7588	.7588	.7589	.7589	.7589

Appendix E: On the Chen-Qin-Condition

We can also develop an estimator for $\tau_{CQ} = \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right)/\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right) = 1/f_{CQ}$ on an analogical way as before. This leads to:

Lemma E.1. *Let be*

$$C_6 = \sum_{\substack{\ell_{1,1},\dots,\ell_{8,1}=1\\ \ell_{1,1}\neq\dots\neq\ell_{8,1}}}^{n_1} \cdots \sum_{\substack{\ell_{1,a},\dots,\ell_{8,a}=1\\ \ell_{1,a}\neq\dots\neq\ell_{8,a}}}^{n_a} \left[\frac{1}{6}\frac{\Lambda_7(\ell_{1,1},\dots,\ell_{8,a})}{16\cdot\prod\limits_{i=1}^{a}\frac{n_i!}{(n_i-8)!}} - \frac{1}{2}\frac{\Lambda_8(\ell_{1,1},\dots,\ell_{8,a})}{16\cdot\prod\limits_{i=1}^{a}\frac{n_i!}{(n_i-8)!}}\right]$$

with

$$\Lambda_7(\ell_{1,1},\dots,\ell_{8,a}) = \left[\boldsymbol{Z}^\top_{(\ell_{1,1},\ell_{2,1},\dots,\ell_{2,a})}\boldsymbol{T}\boldsymbol{Z}_{(\ell_{3,1},\ell_{4,1},\dots\ell_{4,a})}\right]^4,$$

$$\Lambda_8(\ell_{1,1},\dots,\ell_{8,a}) = \left[\sqrt{\Lambda_7(\ell_{1,1},\dots,\ell_{8,a})}\cdot \boldsymbol{Z}^\top_{(\ell_{5,1},\ell_{6,1},\dots,\ell_{6,a})}\boldsymbol{T}\boldsymbol{Z}_{(\ell_{7,1},\ell_{8,1},\dots,\ell_{8,a})}\right]^2.$$

Then we know

$$\mathbb{E}(C_6) = \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right) \qquad Var(C_6) \le \frac{\prod\limits_{i=1}^a \binom{n_i}{8} - \prod\limits_{i=1}^a \binom{n_i-8}{8}}{16^2\cdot \prod\limits_{i=1}^a \binom{n_i}{8}}\mathcal{O}\left(\operatorname{tr}^4\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)\right).$$

Proof.

$$\begin{aligned}\mathbb{E}(C_6) &= \frac{\mathbb{E}\left(\left[\boldsymbol{Z}^\top_{(1,2)}\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\right]^4\right)}{6\cdot 16} - \frac{\mathbb{E}\left(\left[\boldsymbol{Z}^\top_{(1,2)}\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\right]^2\left[\boldsymbol{Z}^\top_{(5,6)}\boldsymbol{T}\boldsymbol{Z}_{(7,8)}\right]^2\right)}{2\cdot 16}\\ &\overset{A.4}{=} \frac{1}{6\cdot 16}\left(6\operatorname{tr}\left((2\boldsymbol{T}\boldsymbol{V}_N)^4\right) + 3\operatorname{tr}^2\left((2\boldsymbol{T}\boldsymbol{V}_N)^2\right)\right) - \frac{1}{2\cdot 16}\operatorname{tr}^2\left((2\boldsymbol{T}\boldsymbol{V}_N)^2\right)\\ &= \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right)\end{aligned}$$

For the second inequality, the variance of parts is calculated. Like before with Lemma A.2 (p.2767) and Theorem A.4 (p.2768) we calculate

$$Var\left(\frac{1}{6}\left[\boldsymbol{Z}_{(1,2)}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\right]^4\right) = \mathcal{O}\left(\operatorname{tr}^4\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)\right)$$

and

$$\begin{aligned}&Var\left(\frac{1}{2}\left[\boldsymbol{Z}_{(1,2)}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\right]^2\left[\boldsymbol{Z}_{(5,6)}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(7,8)}\right]^2\right)\\ &\le \frac{1}{4}\cdot\mathbb{E}\left(\left[\boldsymbol{Z}_{(1,2)}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\right]^4\left[\boldsymbol{Z}_{(5,6)}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(7,8)}\right]^4\right)\\ &= \frac{1}{4}\left(6\operatorname{tr}\left((2\boldsymbol{T}\boldsymbol{V}_N)^4\right) + 3\operatorname{tr}^2\left((2\boldsymbol{T}\boldsymbol{V}_N)^2\right)\right)^2 = \mathcal{O}\left(\operatorname{tr}^4\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)\right).\end{aligned}$$

With Lemma A.7 (p.2770) it is known

$$Var(B) \le Var(A) + Var(B) + 2|\,Cov(A,B)| \le \left(\sqrt{Var(A)} + \sqrt{Var(B)}\right)^2$$

and therefore

$$Var(C_6) \le \frac{\prod\limits_{i=1}^a \binom{n_i}{8} - \prod\limits_{i=1}^a \binom{n_i-8}{8}}{16^2\cdot \prod\limits_{i=1}^a \binom{n_i}{8}}\, Var\left(\frac{1}{6}\Lambda_7(1,\dots,8) - \frac{1}{2}\Lambda_8(1,\dots,8)\right)$$

$$\leq \frac{\prod_{i=1}^{a}\binom{n_i}{8} - \prod_{i=1}^{a}\binom{n_i-8}{8}}{16^2 \cdot \prod_{i=1}^{a}\binom{n_i}{8}}$$

$$\times \left(\sqrt{\mathcal{O}\left(\operatorname{tr}^4\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)\right)} + \sqrt{\mathcal{O}\left(\operatorname{tr}^4\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)\right)}\right)^2$$

$$= \frac{\prod_{i=1}^{a}\binom{n_i}{8} - \prod_{i=1}^{a}\binom{n_i-8}{8}}{16^2 \cdot \prod_{i=1}^{a}\binom{n_i}{8}} \mathcal{O}\left(\operatorname{tr}^4\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)\right). \qquad \square$$

Lemma E.2. *With the estimators introduced in the previous lemmata it holds for fixed a*

$$\frac{C_6}{A_4^2} - \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} \xrightarrow{\mathcal{P}} 0 \qquad \text{for} \quad d, n_{\min} \to \infty.$$

If $q > 1$ exists with $n_{\min} = \mathcal{O}(a^q)$, the convergence even holds in the asymptotic frameworks (4)–(5).

Proof. Again we first consider the parts:

$$\mathbb{E}\left(\frac{C_6}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} - \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right) = \frac{\mathbb{E}\left(C_6\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} - \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} = 0.$$

$$Var\left(\frac{C_6}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} - \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right)$$

$$\leq \frac{\prod_{i=1}^{a}\binom{n_i}{8} - \prod_{i=1}^{a}\binom{n_i-8}{8}}{16^2 \cdot \prod_{i=1}^{a}\binom{n_i}{8}} \frac{\mathcal{O}\left(\operatorname{tr}^4\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)\right)}{\operatorname{tr}^4\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}$$

$$\leq \frac{\prod_{i=1}^{a}\binom{n_i}{8} - \prod_{i=1}^{a}\binom{n_i-8}{8}}{\prod_{i=1}^{a}\binom{n_i}{8}} \cdot \mathcal{O}(1).$$

So with Lemma A.6 (p.2770) for fixed a and $d, n_{\min} \to \infty$ and moreover if the additional condition is fulfilled even for the asymptotic frameworks (4)–(5), it follows

$$\frac{C_6}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} - \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} \xrightarrow{\mathcal{P}} 0.$$

Analogue to the proof of Lemma 4.2 it follows $\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)\Big/A_4^2 \xrightarrow{\mathcal{P}} 1.$

Together this leads to

$$\frac{C_6}{A_4^2} - \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} = \frac{C_6}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} \cdot \frac{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}{A_4^2} - \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}$$

$$= \frac{C_6}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} \cdot (1 + \mathcal{O}_{\mathcal{P}}(1)) - \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} = \mathcal{O}_{\mathcal{P}}(1) + \mathcal{O}_{\mathcal{P}}(1) = \mathcal{O}_{\mathcal{P}}(1). \qquad \square$$

Again in most cases, the subsampling-type version of this estimator should be used.

Lemma E.3. *Let be*

$$C_6^\star(B) = \frac{1}{16B}\sum_{b=1}^{B}\left(\frac{\Lambda_7(\boldsymbol{\sigma}(b,8))}{6} - \frac{\Lambda_8(\boldsymbol{\sigma}(b,8))}{2}\right).$$

Then it holds

$$\mathbb{E}\left(C_6^\star(B)\right) = \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right),$$

$$Var\left(C_6^\star(B)\right) \le \left(1 - \left(1 - \frac{1}{B}\right)\cdot\prod_{i=1}^{a}\frac{\binom{n_i-8}{8}}{\binom{n_i}{6}}\right)\cdot\mathcal{O}\left(\operatorname{tr}^4\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)\right).$$

Proof. By using the same steps as before it holds

$$\begin{aligned}
&\mathbb{E}\left(C_6^\star(B)\right)\\
&= \frac{1}{16B}\sum_{b=1}^{B}\mathbb{E}\left(\frac{\Lambda_7(\ell_{1,1},\ldots,\ell_{8,a})}{6} - \frac{\Lambda_8(\ell_{1,1},\ldots,\ell_{8,a})}{2}\right)\\
&= \frac{1}{16B}\sum_{b=1}^{B}\mathbb{E}\\
&\quad\times\left(\left[\boldsymbol{Z}_{(1,2)}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\right]^2\cdot\left(\frac{\left[\boldsymbol{Z}_{(1,2)}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\right]^2}{6} - \frac{\left[\boldsymbol{Z}_{(5,6)}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(7,8)}\right]^2}{2}\right)\right)\\
&\overset{\text{E.1}}{=} \frac{1}{16B}\sum_{b=1}^{B}\operatorname{tr}\left((2\boldsymbol{T}\boldsymbol{V}_N)^4\right) = \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right).
\end{aligned}$$

$$Var\left(\mathbb{E}\left(C_6^\star(B)|\mathcal{F}(\boldsymbol{\sigma}(B,8))\right)\right) = Var\left(\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right)\right) = 0.$$

$$\begin{aligned}
&Var\left(C_6^\star(B)\right)\\
&= 0 + \mathbb{E}\left(Var\left(C_6^\star(B)|\mathcal{F}(\boldsymbol{\sigma}(B,8))\right)\right)\\
&\overset{\text{A.7}}{\le} \frac{1}{16^2B^2}\mathbb{E}
\end{aligned}$$

$$\times \left(\sum_{(j,\ell)\in\mathbb{N}_B^2 \setminus M(B,\boldsymbol{\sigma}(b,8))} Var\left(\frac{\Lambda_7(\boldsymbol{\sigma}(j,8))}{6} - \frac{\Lambda_8(\boldsymbol{\sigma}(j,8))}{2} \middle| \mathcal{F}(\boldsymbol{\sigma}(B,8)) \right) \right)$$

$$= \frac{Var\left(\frac{\Lambda_7(\ell_{1,1},\dots,\ell_{8,a})}{6} - \frac{\Lambda_8(\ell_{1,1},\dots,\ell_{8,a})}{2} \right)}{16^2 B \cdot \left(\mathbb{E}\left(|\mathbb{N}_B^2 \setminus M(B,\boldsymbol{\sigma}(b,8))| \right) \right)^{-1}}$$

$$\overset{\text{E.1}}{\leq} \left(1 - \left(1 - \frac{1}{B} \right) \cdot \prod_{i=1}^{a} \frac{\binom{n_i-8}{8}}{\binom{n_i}{8}} \right) \cdot \mathcal{O}\left(\operatorname{tr}^4\left((\boldsymbol{T}\boldsymbol{V}_N)^2 \right) \right). \qquad \square$$

With Lemma C.7 we get an estimator for τ_{CQ} with $\widehat{\tau_{CQ}}(C_6^\star, A_4) = C_6^\star / A_4^2$ and once more for a large number of groups $A_4^\star$ should be used.

Lemma E.4. *Theorem 4.1 is also valid if f_P is replaced by f_{CQ} or by $(\widehat{\tau_{CQ}}(C_6, A_4))^{-1}$. Using $C_6^\star$ or $A_4^\star$ also doesn't change the result. Identical the result of Lemma E.2 remains true if one or all estimators are replaced by their subsampling version.*

Proof. With Lemma A.8 we know $f_P \to 1 \Leftrightarrow f_{CQ} \to 1$ and $f_P \to 0 \Leftrightarrow f_{CQ} \to 0$ so in both cases K_{f_P} is asymptotically identic with $K_{f_{CQ}}$.

From Lemma E.2 we know that $\widehat{\tau_{CQ}} - \tau_{CQ}$ converges in probability to zero so this result follows identically to Theorem 4.1. At last the subsampling versions have the same properties as the standard estimators. $\square$

Therefore this is a second way to test the hypotheses and moreover, it provides an indicator for the choice of the limit distribution, because of Lemma A.8. For situation c) from Theorem 3.1 there is no proof that this approach can be used but in the case of just one group, it leads to good results.

References

[1] Ahmad, M. R., Werner, C. and Brunner, E. (2008). Analysis of High Dimensional Repeated Measures Designs: The One Sample Case. *Computational Statistics and Data Analysis* **53** 416–427. MR2649096

[2] Bai, Z. and Saranadasa, H. (1996). Effect of highdimension: by an example of a two sample problem. *Statistica Sinica* **6** 311–329. MR1399305

[3] Bathke, A. C. and Harrar, S. W. (2008). Nonparametric methods in multivariate factorial designs for large number of factor levels. *Journal of Statistical Planning and Inference* **138** 588–610. MR2382556

[4] Bathke, A. C., Harrar, S. W. and Madden, L. V. (2008). How to compare small multivariate samples using nonparametric tests. *Computational Statistics and Data Analysis* **52** 4951–4965. MR2526205

[5] Billingsley, P. (1968). *Convergence of probability measures.* John Wiley & Sons, New York. MR0233396

[6] Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics* **25** 290–302. MR0061787

[7] Brunner, E. (2009). Repeated measures under non-sphericity. Proceedings of the 6th St. Petersburg Workshop on Simulation.
[8] Brunner, E., Becker, B. and Werner, C. (2010). Approximate distributions of quadratic forms in high-dimensional repeated-measures designs. Technical Report, Department Medizinische Statistik Georg-August-Universität Göttingen
[9] Brunner, E., Bathke, A. C. and Placzek, M. (2012). Estimation of Box's ϵ for low- and high-dimensional repeated measures designs with unequal covariance matrices. *Biometrical Journal* **54** 301–316. MR2931076
[10] Cai, T., Liu, W. and Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B* **76** 349–372. MR3164870
[11] Chen, S. X. and Qin, Y.-L. (2010). A Two-Sample Test for High-Dimensional Data with Applications to Gene-Set Testing. *The Annals of Statistics* **38** 808–835. MR2604697
[12] Cramér, H. (1936). Ueber eine Eigenschaft der normalen Verteilungsfunktion. *Mathematische Zeitschrift* **41** 405–414. MR1545629
[13] Croux, C., Rousseeuw, P. J. and Hössjer, O. (1994). Generalized S-estimators. *Journal of the American Statistical Association* **89** 1271–1281. MR1310221
[14] Davis, C. S. (2002). *Statistical Methods for the Analysis of Repeated Measurements.* Springer, New York. MR1883764
[15] Dümbgen, L. (1998). On Tyler's M-Functional of Scatter in High Dimension. *Annals of the Institute of Statistical Mathematics* **50** 471–491. MR1664575
[16] Friedrich, S., Brunner, E. and Pauly, M. (2017). Permuting longitudinal data in spite of the dependencies. *Journal of Multivariate Analysis* **153** 255–265. MR3578850
[17] Geisser, S. and Greenhouse, S. W. (1958). An Extension of Box's Result on the Use of the F Distribution in Multivariate Analysis. *Annals of Mathematical Statistics* **29** 885–891. MR0096331
[18] Greenhouse, S. W. and Geisser, S. (1959). On Methods in the Analysis of Profile Data. *Psychometrika* **24** 95–112. MR0103783
[19] Hájek, J., Šidak, Z. and Sen, P. K. (1999). *Theory of Rank Tests.* Academic Press, San Diego. MR1680991
[20] Happ, M., Harrar, S. W. and Bathke, A. C. (2016). Inference for low- and high-dimensional multigroup repeated measures designs with unequal covariance matrices. *Biometrical Journal* **58** 810–830. MR3527417
[21] Harden, M. (2012). Das Behrens-Fisher-Problem für hochdimensionale Split-Plot-Designs. Master thesis, University Göttingen. http://www.ams.med.uni-goettingen.de/download/diplom/Master-Harden.pdf
[22] Harrar, S. W. and Kong, X. (2016). High-dimensional multivariate repeated measures analysis with unequal covariance matrices. *Journal of Multivariate Analysis* **145** 1–21. MR3459934

[23] Huynh, H. and Feldt, L. S. (1976). Estimation of the Box Correction for Degrees of Freedom From Sample Data in Randomized Block and Split-Plot Designs. *Journal of Educational Statistics* **1** 69–82.

[24] Johnson, R. and Wichern, D. (2007). *Applied multivariate statistical analysis*. 6th Edition, Prentice Hall. MR2372475

[25] Jordan, W., Tumani, H., Cohrs, S., Eggert, S., Rodenbeck, A., Brunner, E., Rüther, E., Hajak, G. (2004). Prostaglandin-D-synthase (beta-trace) in healthy human sleep. *Sleep* **27** 867–874.

[26] Katayama, S., Kano, Y. and Srivastava, M. S. (2013). Asymptotic distributions of some test criteria for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis* **116** 410–421. MR3049913

[27] Kenward, M. G. and Roger, J. H. (2009). An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics & Data Analysis* **53** 2583–2595. MR2665909

[28] Keselman, H. J., Algina, J. and Kowalchuk, R. K. (2001). The analysis of repeated measures designs. *British Journal of Mathematical and Statistical Psychology* **54** 1–20.

[29] Konietschke, F., Bathke, A. C., Harrar, S. W. and Pauly, M. (2015). Parametric and Nonparametric Bootstrap Methods for General MANOVA. *Journal of Multivariate Analysis* **140** 291–301. MR3372569

[30] Lecoutre, B. (1991). A Correction for the $\widetilde{\epsilon}$: Approximative Test in Repeated Measures Designs With Two or More Independent Groups. *Journal of Educational Statistics* **16**, 371–372.

[31] Liu, Z., Liu, B., Zheng, S. and Shi, N.-Z. (2017). Simultaneous testing of mean vector and covariance matrix for high-dimensional data. *Journal of Statistical Planning and Inference* **188** 82–93. MR3648319

[32] Mathai, A. M. and Provost, S. B. (1992). *Quadratic forms in random variables*. Marcel Dekker Inc., New York. MR1192786

[33] Nishiyama, T., Hyodo, M., Seo, T. and Pavlenko, T. (2013). Testing linear hypotheses of mean vectors for high-dimension data with unequal covariance matrices. *Journal of Statistical Planning and Inference* **143** 1898–1911. MR3095080

[34] Pauly, M., Ellenberger, D. and Brunner, E. (2015). Analysis of high-dimensional one group repeated measures designs. *Statistics* **49**(6) 1243–1261. MR3415218

[35] Pesarin, F. and Salmaso, L. (2012). A review and some new results on permutation testing for multivariate problems. *Statistics and Computing* **22** 639–646. MR2865041

[36] Rao, C. R. and Mitra, S. K. (1971). *Generalized Inverse of Matrices and Its Applications*. Wiley, New York. MR0338013

[37] Secchi, P., Stamm, A. and Vantini, S. (2013). Inference for the mean of large p small n data: A finite-sample high-dimensional generalization of Hotelling's theorem. *Electronic Journal of Statistics* **7** 2005–2031. MR3085016

[38] Skene, S. S. and Kenward, M. G. (2010). The analysis of very small samples of repeated measurements II: A modified Box correction. *Statistics in Medicine* **29** 2838–2856. MR2758474

[39] Tyler, D. E., Critchley, F., Dümbgen L. and Oja, H. (2009). Invariant Co-ordinate Selection (with discussion). *Journal of the Royal Statistical Society, Series B* **71** 549–592. MR2749907

[40] Vallejo, G. and Ato, M. (2006). Modified Brown-Forsythe procedure for testing interaction effects in split-plot designs. *Multivariate Behavioral Research* **41** 549–578.

[41] Werner, C. (2004). Dimensionsstabile Approximation für Verteilungen von quadratischen Formen im Repeated-Measures-Design. Diploma Thesis, University of Göttingen. http://www.ams.med.uni-goettingen.de/download/TECHREPORT.PDF

[42] Witting, H. and Müller-Funke, U. (1995). Asymptotische Statistik: Parametrische Modelle und nicht-parametrische Funktionale, Stuttgart: Teubner.

[43] Zhan, D. and Hart, J. D. (2014). Testing equality of a large number of densities. *Biometrika* **101.2** 449–464. MR3215359

[44] Zhang, J. T. (2005). Approximate and asymptotic distributions of chi-squared–type mixtures with applications. *Journal of the American Statistical Association* **100** 273–285. MR2156837

Article 2

Sattler, P. (2021). A comprehensive treatment of quadratic-form-based inference in repeated measures designs under diverse asymptotics.Electronic Journal of Statistics 15 (1) 3611 - 3634, 2021. doi.org/10.1214/21-EJS1865.

Electronic Journal of Statistics
Vol. 15 (2021) 3611–3634
ISSN: 1935-7524
https://doi.org/10.1214/21-EJS1865

A comprehensive treatment of quadratic-form-based inference in repeated measures designs under diverse asymptotics*

Paavo Sattler

Faculty of Statistics,
TU Dortmund University, 44221 Dortmund, Germany.
e-mail: paavo.sattler@tu-dortmund.de

Abstract: Split-Plot or Repeated Measures Designs with multiple groups occur naturally in sciences. Their analysis is usually based on the classical Repeated Measures ANOVA. Roughly speaking, the latter can be shown to be asymptotically valid for large sample sizes n_i assuming a fixed number of groups a and time points d. However, for high-dimensional settings with $d > n_i$, this argument breaks down and statistical tests are often based on (standardized) quadratic forms. Furthermore, analysis of their limit behaviour is usually based on certain assumptions on how d converges to ∞ with respect to n_i. As this may be hard to argue in practice, we do not want to make such restrictions. Moreover, sometimes also the number of groups a may be large compared to d or n_i. To also have an impression about the behaviour of (standardized) quadratic forms as test statistic, we analyze their asymptotics under diverse settings on a, d and n_i. In fact, we combine all kinds of combinations, where they diverge or are bounded in a unified framework. To this aim, we assume equal covariance matrices between all groups. Studying the limit distributions in detail, we follow Sattler and Pauly (2018) and propose an approximation to obtain critical values. The resulting test and its approximation approach are investigated in an extensive simulation study focusing on the exceptional asymptotic frameworks that are the main focus of this work.

Keywords and phrases: Multivariate data, equal covariance matrices, high-dimensions, repeated measures.

Received March 2020.

Contents

1 Motivation and introduction 3612
2 Statistical model and hypotheses 3613
3 Test statistics and their asymptotics 3614
4 Simulation 3619
 4.1 Type-I error 3619
  4.1.1 Power 3621
5 Conclusion 3625

*This work was supported by the German Research Foundation project DFG-PA2409/4-1.

6 Appendix . . . 3626
Acknowledgments . . . 3632
References . . . 3632

1. Motivation and introduction

In many studies, it is possible to conduct and handle a large number of measurements, which makes high-dimensionality an increasingly important topic. In fact, high-dimensional repeated measure designs or split-plot designs for multiple groups are the objectives of many analyses in science. This is the case in life science, where test persons were examined multiple times during a study, or in the industry where some parameters are measured on a nearly continuous basis. Therein we consider d measurements from N subjects, which are divided into a independent and generally unbalanced groups where the i-th group contains n_i observations. Moreover, factor levels on the groups or repeated measures are possible. For independent d-dimensional observation vectors $\boldsymbol{X}_{ik} \sim \mathcal{N}_d(\mu_i, \Sigma_i)$ null hypotheses regarding $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_a)^\top$ are investigated, where popular hypotheses are the existence of a group effect, a time effect as well as an interaction effect between time and group. For a classical repeated measures ANOVA design with $d \leq n_i$, this was treated for example in [6]. But in many cases, it is easier, cheaper, or ethically more justifiable to increase the number of repetitions rather than increasing the sample size. Therefore techniques are needed, which can handle the case of $d > n_i$.

In the particular case with just two groups but a general distributional setting and without restriction on the dimension d, this was treated in [7]. For more groups and a more general setting regarding hypotheses, [9] uses a classical ANOVA F test statistic, which has just an exact F-distribution for very special covariance matrices. So under some conditions on n_i/d or the relation between the dimension and some power of traces containing the covariance matrix, they developed a decent approximation for the test statistic.

In [10] they handle several cases with an increasing number of groups under some requirements on the covariance matrices and the relation between sample sizes and the number of factor levels. In contrast, [17] investigated the case with just one normally distributed group, but fewer assumptions on the covariance matrix and no specific relation between sample size and dimension.

[18] expand these results especially for a larger number of groups, which is also allowed to approach infinity, together with the sample sizes and the dimension. As a result of this, no restrictions on their respective convergence rate were made. However, this does not treat the small n large a case which was, e.g., treated by [2] or [3] for fixed dimensions d and balanced designs $n_i \equiv n$.

The importance of such large a small n cases increased in the last years, for example, through more interest for *personalized medicine*, as mentioned in [1]. Here the idea is to develop treatments adapted to the properties of the patients, see for example [11]. A similar idea is in stratified medicine, where depending on common biological or other characteristics, appropriate therapies are developed for groups of patients. Therefore it is necessary to divide existing groups into subgroups with smaller numbers of subjects. Also, in other areas like insurance, there is a trend for more personalized products. Together with the frequent use of high-dimensional data, there is a demand for more comprehensive asymptotic frameworks.

Therefore, in addition to the large a small n case, we include the large d small n case, further combining both and developing a technique that can be used in each of these settings. To this end, we follow the same approach as [12] and assume homogenous covariance matrices with $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} > 0$, again with no further assumptions on the structure of the covariance matrix $\boldsymbol{\Sigma}$. The homoscedastic setting allows some generalizations as well as a smaller number of other requirements on the underlying statistical model.

This paper is organized as follows. Section 2 introduces the statistical model, the investigated hypotheses, and the notations used in the paper's remainder. In Section 3, the test statistic is presented, as well as their asymptotic behavior and an alternative small sample approximation. Section 4 contains simulations regarding the type-I-error rate and the tests' power, introduced in the previous chapters. The paper closes with a short conclusion. For brevity and readability, all proofs are shifted to the appendix.

2. Statistical model and hypotheses

We consider a homogenous split-plot design given by a independent and unbalanced groups of d-dimensional random vectors

$$\boldsymbol{X}_{i,j} = (X_{i,j,1}, \dots, X_{i,j,d})^\top \overset{ind}{\sim} \mathcal{N}_d(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}) \quad j = 1, \dots, n_i, \quad i = 1, \dots, a, \tag{1}$$

whereby each vector represents the measurement of one independent subject. It is assumed that mean vectors $E(\boldsymbol{X}_{i,1}) = \boldsymbol{\mu}_i = (\mu_{i,t})_{t=1}^d \in \mathbb{R}^d$ and one positive definite covariance matrix $Cov(\boldsymbol{X}_{i,1}) = \boldsymbol{\Sigma} > 0$ exist. As usual $j = 1, \dots, n_i$ denotes the individual subjects or units in group $i = 1, \dots, a$, $a, n_i \in \mathbb{N}$, so we have a total number of $N = \sum_{i=1}^a n_i$ random vectors. This framework allows a factorial structure regarding time, group or both, by splitting up the indices, accordingly, see [13] for example.

Within this model linear hypotheses of repeated measures ANOVA, formulated as

$$\mathcal{H}_0(\boldsymbol{H}) : \boldsymbol{H}\boldsymbol{\mu} = \boldsymbol{0} \quad \boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \dots, \boldsymbol{\mu}_a^\top)^\top, \tag{2}$$

are investigated. Here, $\boldsymbol{H} = \boldsymbol{H}_W \otimes \boldsymbol{H}_S$ denots a proper hypothesis matrix, where $\boldsymbol{H}_W$ and $\boldsymbol{H}_S$ refer to **w**hole-plot (group) and/or **s**ubplot (time) effects, while $\otimes$ denotes the Kronecker product.

For theoretical considerations it is often more convenient to reformulate $\mathcal{H}_0(\boldsymbol{H})$ through a corresponding projection matrix $\boldsymbol{T} = \boldsymbol{H}^\top[\boldsymbol{H}\boldsymbol{H}^\top]^-\boldsymbol{H}$, see e.g. [17]. Here $(\cdot)^-$ denotes some generalized inverse of the matrix and $\mathcal{H}_0(\boldsymbol{H})$ can equivalently be written as $\mathcal{H}_0(\boldsymbol{T}) : \boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0}$. As discussed in [18], $\boldsymbol{T}$ has the form $\boldsymbol{T} = \boldsymbol{T}_W \otimes \boldsymbol{T}_S$ for projection matrices $\boldsymbol{T}_W$ and $\boldsymbol{T}_S$. Now hypotheses of interest are for example given by

(a) No group effect:
$\mathcal{H}_0^a : \left(\boldsymbol{P}_a \otimes \frac{1}{d}\boldsymbol{J}_d\right)\boldsymbol{\mu} = \boldsymbol{0}$,
(b) No time effect:
$\mathcal{H}_0^b : \left(\frac{1}{a}\boldsymbol{J}_a \otimes \boldsymbol{P}_d\right)\boldsymbol{\mu} = \boldsymbol{0}$,
(c) No interaction effect between time and group:
$\mathcal{H}_0^{ab} : (\boldsymbol{P}_a \otimes \boldsymbol{P}_d)\,\boldsymbol{\mu} = \boldsymbol{0}$.

Here, $\boldsymbol{J}_d$ is the d-dimensional matrix only containing 1s and $\boldsymbol{P}_d := \boldsymbol{I}_d - 1/d \cdot \boldsymbol{J}_d$ is the centering matrix.

It is often useful to split the expectation vector into its components to simplify the interpretation. With the common conditions $\sum_i \alpha_i = \sum_t \beta_t = \sum_{i,t}(\alpha\beta)_{it} = 0$, this can be done by expanding

$$\mu_{i,t} = \mu + \alpha_i + \beta_t + (\alpha\beta)_{it}, \quad i = 1, \ldots, a;\ t = 1, \ldots, d.$$

Here, $\alpha_i \in \mathbb{R}$ describes the i-th group effect, $\beta_t \in \mathbb{R}$ the time effect at time point t and $(\alpha\beta)_{it} \in \mathbb{R}$ the (i,t)-interaction effect between group and time. Thereby the above hypotheses can alternatively be formulated through

(a) $H_0^a : \alpha_i \equiv 0$ for all i,
(b) $H_0^b : \beta_t \equiv 0$ for all t,
(c) $H_0^{ab} : (\alpha\beta)_{it} \equiv 0$ for all i, t.

3. Test statistics and their asymptotics

In this work, we consider the following five different asymptotic frameworks, which are:

$$a \to \infty, \tag{I}$$
$$a, d \to \infty, \tag{II}$$
$$a, n_{\max} \to \infty, \tag{III}$$
$$d, n_{\max} \to \infty, \tag{IV}$$
$$a, d, n_{\max} \to \infty. \tag{V}$$

This great diversity is exceptional and distinguishes the present proposal from nearly all other approaches. Most of the existing procedures just consider special cases of one of these cases (for example [7] (IV) with $a = 1$ or [17] (IV) with $a = 2$). Others allow for only one as [9] for (IV) or [2] for (I).

In contrast, our framework allows the combination of any of these assumptions. However, $d \to \infty$ alone is not included as this would not allow the construction of consistent trace estimators of covariances which are later needed for inference. Moreover, the case $n_{\max} = \max(n_1, ..., n_a) \to \infty$ with fixed a and d has already been studied in detail in the literature and is thus excluded here, see, e.g., [8] or [4] and the references cited therein.

It is apparent that in contrast to [18] and other papers, the common conditions of $\frac{n_i}{N} \to \kappa_i \in (0,1)$ are missing. This is relevant because it allows an appreciably larger amount of settings, especially for $a \to \infty$. But it also clearly generalizes the model for the case of fixed a, e.g. in unbalanced settings, where we only let some group sample sizes converge to ∞.

To examine the validity of the null hypothesis $H_0(\boldsymbol{T}) : \boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0}$ unattached from the asymptotic framework, we use $Q_N = N \cdot \overline{\boldsymbol{X}}^\top \boldsymbol{T} \overline{\boldsymbol{X}}$. Here $\overline{\boldsymbol{X}} = (\overline{\boldsymbol{X}}_1^\top, \dots \overline{\boldsymbol{X}}_a^\top)^\top$ with $\overline{\boldsymbol{X}}_i = n_i^{-1} \sum_{j=1}^{n_i} \boldsymbol{X}_{i,j}, i = 1, \dots, a$, denotes the vector of pooled group means. Unfortunately for many covariance matrices $\boldsymbol{\Sigma}$, the random variable Q_N tends to converge to infinity, for $d \to \infty$ or $a \to \infty$. To avoid this behaviour the standardized quadratic form given by

$$\widetilde{W}_N = \frac{Q_N - \mathbb{E}_{\mathcal{H}_0}(Q_N)}{\sqrt{Var_{\mathcal{H}_0}(Q_N)}},$$

is used, which also enables us to evaluate all limit distributions in detail.

For normal distributed observations the expectation and variance of the quadratic form is known and it follows that

$$\begin{aligned} \mathbb{E}(Q_N) \quad &= \operatorname{tr}(\boldsymbol{T}_S \boldsymbol{\Sigma}) \cdot \sum_{i=1}^{a} \tfrac{N}{n_i} (\boldsymbol{T}_W)_{ii} \\ Var(Q_N) \quad &= 2 \cdot \operatorname{tr}((\boldsymbol{T}_S \boldsymbol{\Sigma})^2) \cdot \sum_{i=1}^{a} \sum_{r=1}^{a} \tfrac{N^2}{n_i n_r} (\boldsymbol{T}_W)_{ir}^2. \end{aligned}$$

Observe, that for both values only the first factor $\operatorname{tr}(\boldsymbol{T}_S\boldsymbol{\Sigma})$ resp. $\operatorname{tr}((\boldsymbol{T}_S\boldsymbol{\Sigma})^2)$ depends on the unknown covariance matrix, while all other quantities are known from the test setting.

Applying the representation theorem for quadratic forms in normaly distributed random vectors from [16] we can rewrite the standardized statistic $\widetilde{W}_N$ as

$$\widetilde{W}_N = \frac{Q_N - \mathbb{E}_{H_0}(Q_N)}{Var_{H_0}(Q_N)^{1/2}} \overset{\mathcal{D}}{=} \sum_{s=1}^{ad} \frac{\lambda_s}{\sqrt{\sum_{\ell=1}^{ad} \lambda_\ell^2}} \left(\frac{C_s - 1}{\sqrt{2}} \right). \tag{3}$$

Here λ_s are the eigenvalues of $\boldsymbol{T}\boldsymbol{V}_N\boldsymbol{T}$ in decreasing order, $\boldsymbol{V}_N = \bigoplus_{i=1}^{a} \frac{N}{n_i} \boldsymbol{\Sigma}$ and $(C_s)_s$ is a sequence of independent χ_1^2-distributed random variables. As a consequence, the asymptotic behaviour of the eigenvalues, determine the asymptotic limit distribution of $\widetilde{W}_N$. In fact, we obtain in generalization of [17] and [18]:

Theorem 1. *Let* $\beta_s = \lambda_s \Big/ \sqrt{\sum_{\ell=1}^{ad} \lambda_\ell^2}$ *for* $s = 1, \dots, ad$. *Then* $\widetilde{W}_N$ *has, under* $H_0(\boldsymbol{T})$*, and one of the frameworks* (I)-(V) *asymptotically*

a) a distribution of the form $\sum_{s=1}^{r} b_s (C_s - 1) / \sqrt{2} + \sqrt{1 - \sum_{s=1}^{r} b_s^2} \cdot Z$*, if and only if*

$$\text{for all } s \in \mathbb{N} \quad \beta_s \to b_s \quad \text{as} \quad N \to \infty,$$

for a decreasing sequence $(b_s)_s$ *in* $[0,1]$ *with* $r := \#\{b_i \neq 0\}$*, while* $C_i \overset{i.i.d.}{\sim} \chi_1^2$*,* $Z \sim \mathcal{N}(0,1)$.

b) a distribution of the form $\sum_{s=1}^{\infty} b_s\,(C_s-1)/\sqrt{2}$, *if*

$$\text{for all } s\in\mathbb{N} \quad \beta_s\to b_s \quad \text{as} \quad N\to\infty,$$

for a decreasing sequence $(b_s)_s$ *in* $(0,1)$ *with* $\sum_{s=1}^{\infty} b_s^2=1$ *and* $C_i \overset{i.i.d.}{\sim} \chi_1^2$.

Putting the results into context. [7] only considered case a) with $r=0$. [18] at least found asymptotic results in case b) but for case a) they need $\boldsymbol{r\in\{0,1\}}$. So this theorem is not only distinct from other results through the variety of asymptotic settings. It also considerably enhances the continuum of limit distributions through a mixture of normal distribution and finite sums of weighted standardized χ_1^2-distributed random variables. Furthermore, the if and only if relation shows the importance of the demands for the standardized eigenvalues and that it isn't possible to relax them.

To use this test statistic, it is necessary to construct proper estimators with the necessary properties. One of these is ratio-consistency, where we call an estimator $\widehat{\theta}_{n,d}$ for θ ratio-consistent, if it holds $\widehat{\theta}_{n,d}/\theta \xrightarrow{\mathcal{P}} 1$. To get such estimators, we define

$$A_1=\frac{1}{\sum_{i=1}^a (n_i-1)n_i}\sum_{i=1}^{a}\sum_{\ell_1<\ell_2=1}^{n_i}(\boldsymbol{X}_{i,\ell_1}-\boldsymbol{X}_{i,\ell_2})^\top\boldsymbol{T}_S(\boldsymbol{X}_{i,\ell_1}-\boldsymbol{X}_{i,\ell_2}$$

and

$$A_2=\sum_{i=1}^{a}\sum_{\substack{\ell_1,\ell_2=1\\ \ell_1>\ell_2}}^{n_i}\sum_{\substack{k_2=1\\ k_2\neq\ell_1\neq\ell_2}}^{n_i}\sum_{\substack{k_1=1\\ \ell_2\neq\ell_1\neq k_1>k_2}}^{n_i}\frac{\left[\left(\boldsymbol{X}_{i,\ell_1}-\boldsymbol{X}_{i,\ell_2}\right)^\top\boldsymbol{T}_S\left(\boldsymbol{X}_{i,k_1}-\boldsymbol{X}_{i,k_2}\right)\right]^2}{4\cdot 6\sum_{i=1}^a\binom{n_i}{4}}.$$

Below we prove that they are unbiased and ratio consistent estimators for $\operatorname{tr}(\boldsymbol{T}_S\boldsymbol{\Sigma})$ and $\operatorname{tr}\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^2\right)$, respectivly, under both, the nullhypothesis and the alternative. This allows us to define the estimated version of our test statistic by

$$W_N=\frac{Q_N-A_1\cdot\sum_{i=1}^a\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}}{\sqrt{2\cdot A_2\cdot\sum_{i=1}^a\sum_{r=1}^a\frac{N^2}{n_i n_r}(\boldsymbol{T}_W)_{ir}^2}}.$$

The following Lemma justifies the usage of the estimated version instead of the exact one.

Theorem 2. *Under* $H_0(T):\boldsymbol{T\mu}=\boldsymbol{0}_{ad}$ *and one of the frameworks* (I)-(V) *the statistic* W_N *has the same asymptotic limit distributions as* $\widetilde{W}_N$, *if the respective conditions (a)-(b) from Theorem 1 are fulfilled.*

Unfortunately, the calculation of the standardized eigenvalues β_s is generally not simplified through homogeneity. Therefore it is nearly impossible to find an appropriate estimator which can be used in all our frameworks. Moreover, simulations showed that large sample sizes, dimensions or number of groups are necessary for a good approximation, which make quantiles based on Theorem 1 a) difficult to apply. For similar reasons, in [17] and [18] they used the quantiles of a random variable of the kind

$$K_f=(\chi_f^2-f)/\sqrt{2f}, \tag{4}$$

in case of $\beta_1 \to \{0,1\}$. The choice of $f_P = \mathrm{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)/\mathrm{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)$ for the degrees of freedom lead to a third moment approximation. In our homoscedastic model the usage of this random variable K_{f_P} is based on the following theorem.

Theorem 3. *Under the conditions of Theorem 1 and one of the frameworks* (I)-(V) *the random variable K_{f_P} has, under $H_0 : \boldsymbol{T}\boldsymbol{\mu} = \boldsymbol{0}_{ad}$, asymptotically*

a) a standard normal distribution if $\beta_1 \to 0$ as $N \to \infty$,
b) a standardized $\left(\chi_1^2 - 1\right)/\sqrt{2}$ distribution if $\beta_1 \to 1$ as $N \to \infty$.

With the well known rules for the kronecker product and traces we can decompose the parameter f_P by

$$f_P = \frac{\mathrm{tr}^3\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^2\right)}{\mathrm{tr}^2\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^3\right)}\cdot\frac{\mathrm{tr}^3\left([\mathrm{diag}(N/n_1,...,N/n_a)\cdot\boldsymbol{T}_W]^2\right)}{\mathrm{tr}^2\left([\mathrm{diag}(N/n_1,...,N/n_a)\cdot\boldsymbol{T}_W]^3\right)} =: \frac{\mathrm{tr}^3\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^2\right)}{\mathrm{tr}^2\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^3\right)}\cdot\eta_{N,a}.$$

The connection between f_P and β_1 in the two extreme cases, i.e. $\beta_1 \to 0$ if and only $f_P \to \infty$ and $\beta_1 \to 1$ if and only if $f_P \to 1$, have been investigated in [17] for the case of $a = 1$ but also translate to the present framework.

Here we have to estimate the first part, while the second one $\eta_{N,a}$ just depends on the asymptotic setting and therefore is known. This allows us to use the same estimated traces for different hypothesis which differ only in $\boldsymbol{T}_W$.

Moreover, for $\eta_{N,a} \to \infty$, we also have $f_P \to \infty$, without estimation, because $\mathrm{tr}^3\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^2\right)/\mathrm{tr}^2\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^3\right) \geq 1$. Otherwise, however, the behaviour of f_P is unclear and we have to find consistent estimators for $\mathrm{tr}\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^3\right)$ in all our different frameworks. This achieved by considering the class of estimators

$$C_{i,1} := \frac{1}{8}\sum_{\ell_1\neq\ldots\neq\ell_6=1}^{n_i} \boldsymbol{Y}_{i,\ell_1,\ell_2}^\top\boldsymbol{Y}_{i,\ell_3,\ell_4}\boldsymbol{Y}_{i,\ell_3,\ell_4}^\top\boldsymbol{Y}_{i,\ell_5,\ell_6}\boldsymbol{Y}_{i,\ell_5,\ell_6}^\top\boldsymbol{Y}_{i,\ell_1,\ell_2},$$

with $\boldsymbol{Y}_{i,\ell_1,\ell_2} := \boldsymbol{T}_S(\boldsymbol{X}_{i,\ell_1} - \boldsymbol{X}_{i,\ell_2})$. These are based on suitable symmetrized U-statistics, while $\ell_1 \neq \ell_2 \neq ... \neq \ell_6$ means that all indices are different.

Afterwards these estimators for each individual group are combined, to get an estimator which uses the observations of each group, given by

$$C_1 := \frac{1}{6!\cdot\sum_{j=1}^a\binom{n_j}{6}}\sum_{i=1}^a C_{i,1}.$$

Together with the estimators from above, we can construct a consistent estimator for f_P by $\widehat{f}_P := A_2^3/C_1^2\cdot\eta_{N,a}$.

Theorem 4. *In all our frameworks* (I)-(V)*, it holds that*

i) C_1 is an unbiased estimator for $\mathrm{tr}\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^3\right)$,
ii) $\left(\widehat{f}_P\right)^{-1} - (f_P)^{-1} \xrightarrow{\mathcal{P}} 0$,

where $\mathcal{P}$ denotes convergence in propabilty.

Through the usage of U-statistics with a kernel of order 6, for each estimator $C_{1,i}$, $6! \cdot \binom{n_i}{6}$ summations have to be done. In contrast, estimators based on observations from all groups would require much higher numbers. For example in [18] $\prod_{i=1}^{a} 6! \cdot \binom{n_i}{6}$ summations are necessary. Due to homogeneity, we don't need this kind of estimator, but C_1 also requires $6! \cdot \sum_{j=1}^{a} \binom{n_j}{6}$ summations, which is already really high, even for comparatively small samples sizes or numbers of groups. Thus, as in [18], the usage of subsampling versions of our estimators is reasonable to make them applicable in practice. Instead of summing up all possible index combinations of one group, the underlying idea is only to do this for a randomly chosen subset of combinations.

To define the subsampling version, it is first necessary to introduce some definitions and notations. A parameter $\upsilon \in (0, \infty)$ is chosen and used to define $w_i = \lceil \upsilon \cdot \binom{n_i}{6} \rceil, i = 1, ..., a$ as the number of subsampling repetitions done for the i-th group. It is clear that the choice of υ has a great influence on the calculation time and accuracy, so it should be chosen suitable for the situation.

Then, random subsamples $\boldsymbol{\sigma}_i(b) = \{\sigma_{1i}(b), \ldots, \sigma_{6i}(b)\}$ of length 6 from $\{1, \ldots, n_i\}$ are drawn independently for each $i = 1, \ldots, a$ and $b = 1, \ldots, w_i$, to define the subsampling version of $C_{i,1}$ by

$$C_{i,1}^{\star} = C_{i,1}^{\star}(w_i) = \sum_{b=1}^{w_i} \Lambda_1(\boldsymbol{\sigma}_i(b)) \cdot \Lambda_2(\boldsymbol{\sigma}_i(b)) \cdot \Lambda_3(\boldsymbol{\sigma}_i(b)).$$

Here

$$\begin{aligned}\Lambda_1(\ell_1, \ell_2, \ell_3, \ell_4, \ell_5, \ell_6) &= \boldsymbol{Y}_{i,\ell_1,\ell_2}^{\top} \boldsymbol{Y}_{i,\ell_3,\ell_4},\\ \Lambda_2(\ell_1, \ell_2, \ell_3, \ell_4, \ell_5, \ell_6) &= \boldsymbol{Y}_{i,\ell_3,\ell_4}^{\top} \boldsymbol{Y}_{i,\ell_5,\ell_6},\\ \Lambda_3(\ell_1, \ell_2, \ell_3, \ell_4, \ell_5, \ell_6) &= \boldsymbol{Y}_{i,\ell_5,\ell_6}^{\top} \boldsymbol{Y}_{i,\ell_1,\ell_2}.\end{aligned}$$

Combining them, allows to define the subsamling version of C_1 by

$$C_1^{\star} := \frac{1}{8 \cdot \sum_{j=1}^{a} w_i} \cdot \sum_{i=1}^{a} C_{i,1}^{\star}(w_i).$$

Theorem 5. *For $\sum_{i=1}^{a} w_i \to \infty$, if $N \to \infty$ (which includes frameworks (I)-(V)) it holds:*

a) $C_1^{\star}$ is an unbiased estimator for $\operatorname{tr}\left((\boldsymbol{T}_S \boldsymbol{\Sigma})^3\right)$.

b) $\widehat{f}_P^{\star} := \frac{A_2^3}{(C_1^{\star})^2} \cdot \eta_{N,a}$ fullfilles $\left(\widehat{f}_P^{\star}\right)^{-1} - (f_P)^{-1} \xrightarrow{\mathcal{P}} 0$.

This way of defining the number of subsampling repetitions w_i, guarantees that the relation between the subsampled parts $C_{1,i}^{\star}$ resembles the relation between the original $C_{1,i}$. Although this can lead to great differences between the subsampling sizes for the different groups, it ensures that single groups' influence is not too big.

These results allow formulating a more useable version of K_{f_P} through the following theorem.

Theorem 6. *The results of Theorem 3 remains valid if f_P is replaced by $\widehat{f}_P$ or $\widehat{f}_P^\star$.*

For the estimation of the unknown traces, it would also be possible to construct estimators that use observations from different groups. This is feasible and seems reasonable, but in practice, we would again need subsampling versions of these estimators, which take care of the dataset's structure. This is really complicated and therefore not usable in practice. So we avoid these difficulties by using estimators for the separate groups and combine them afterward.

Relaxing the assumption of homogeneous covariance matrices to $\boldsymbol{X}_{ij} \sim \mathcal{N}_d(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ with $\boldsymbol{T}_S\boldsymbol{\Sigma}_1 = \boldsymbol{T}_S\boldsymbol{\Sigma}_2 = ... = \boldsymbol{T}_S\boldsymbol{\Sigma}_a$, which is essentially easier to fulfill, wouldn't change the validity of the previous results. From a theoretical point of view it would be even sufficient to assume $\operatorname{tr}((\boldsymbol{T}_S\boldsymbol{\Sigma}_1)^j) = \operatorname{tr}((\boldsymbol{T}_S\boldsymbol{\Sigma}_2)^j) = ... = \operatorname{tr}((\boldsymbol{T}_S\boldsymbol{\Sigma}_a)^j)$ for $j \in \{1, 2, 3\}$, but this is nearly impossible to justify in practice.

Remark 1. a) The equality of the covariance matrices is a central condition of our approach. Otherwise the structure of $\mathbb{E}(Q_N)$ and $Var(Q_N)$ changes considerably, and properties of all estimators holds no longer. The consquences of a violation strongly depends on the setting and are difficult to assess. So if there exists an $i \in \mathbb{N}_a$ with $\boldsymbol{\Sigma}_i \neq \boldsymbol{\Sigma}$ then $\operatorname{tr}((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^j)/\operatorname{tr}(\boldsymbol{T}_S\boldsymbol{\Sigma})^j)$ can be close to one for $j = 1, 2$ but stronlgy influence $\widehat{Q}_N$, depending on the interplay, the sample size and the asymptotic framework.

b) Therefore, in the frameworks (III)-(V), it is preferable to use the approach from [18], if the condition seems less plausible.

c) All our introduced estimators are composed from estimators for the single groups. This allows to recognize groups, whose traces vary widley from the others, and therefore deteced groups with other covariance matrix and assess their influence.

4. Simulation

For an evaluation of the finite sample behavior of the introduced method, we have conducted extensive simulations regarding

(i) their ability in keeping the nominal significance level and
(ii) their power to detect certain alternatives in various scenarios.

Here we focus on the frameworks (I) and (II), which are the most interesting ones because they don't require the usual condition of increasing sample sizes. Therefore they are a strict expansion of the settings considered in [18].

4.1. Type-I error

To check the type-I error rate for $\alpha = 5\%$ we consider small($d = 5$, $d = 50$), moderate($d = 200$) and large dimension($d = 600$) and increasing the number of groups from 2 to 12. The sample sizes are fixed in a quite unbalanced setting given by $\boldsymbol{n} = (n_1, ..., n_{12}) = (15, 15, 20, 35, 25, 20, 30, 30, 35, 20, 15, 25)$. We used 10,000 simulation runs and chose $\upsilon = 0.05$ for our subsampling type estimators. Thereby, the number of subsamling draws are between 251 and 81,158, one basis of the quite unbalanced setting. Higher values for υ would increase the accurancy but noticeable extend the computation time.

Two different null hypotheses are investigated to have a situation with $\beta_1 \to 0$ as well as with $\beta_1 \to 1$. These hypotheses are

- $\mathcal{H}_0^a : (\boldsymbol{P}_a \otimes \boldsymbol{P}_d)\, \boldsymbol{\mu} = \boldsymbol{0}$,
- $\mathcal{H}_0^b : \left(\frac{1}{a}\boldsymbol{J}_a \otimes \frac{1}{d}\boldsymbol{J}_d\right) \boldsymbol{\mu} = \boldsymbol{0}$.

For both hypotheses the same distributional setting is choosen, with $\boldsymbol{\Sigma}$ as a autoregressive covariance matrix with parameter 0.6 e.g. $(\boldsymbol{\Sigma})_{i,j} = 0.6^{|i-j|}$ and $\boldsymbol{\mu}_i = \boldsymbol{0}_d$ for $i = 1, ..., a$, to achieve better comparabilty. For $\mathcal{H}_0^b$ it holds $\tau_P \equiv 1$ while the values for $\mathcal{H}_0^a$ can be seen in Table 1

Table 1
τ_P for $\boldsymbol{T} = \frac{1}{a}\boldsymbol{J}_a \otimes \frac{1}{d}\boldsymbol{J}_d$ and $(\boldsymbol{\Sigma})_{ij} = 0.6^{|j-i|}$ with different dimension and numbers of groups.

τ_P	a=2	a=3	a=4	a=5	a=6	a=7	a=8	a=9	a=10	a=11	a=12
d=5	.524	.268	.189	.146	.122	.105	.097	.092	.080	.074	.070
d=50	.100	.051	.036	.028	.023	.020	.019	.018	.015	.014	.013
d=200	.025	.013	.009	.007	.006	.005	.005	.004	.004	.004	.003
d=600	.008	.004	.003	.002	.002	.002	.002	.001	.001	.001	.001

All tests $\psi_z = 1\!\!1(W_N > z_{1-\alpha})$, $\psi_\chi = 1\!\!1(W_N > \chi^2_{1;1-\alpha})$ and $\varphi_N^\star = 1\!\!1\{W_N > K_{\hat{f}_P;1-\alpha}\}$ are used while $\chi^2_{1;1-\alpha}$ denotes the $1-\alpha$ quantile of a χ^2_1 distribution and $K_{\hat{f}_P;1-\alpha}$ the $1-\alpha$ quantile of $K_{\hat{f}_P}$. It must be noted that in the following figures, we use different axes for each setting to make them as detailed as possible.

In Figure 1 it can be seen that for $\beta_1 \to 0$, the usage of ψ_χ results in too conservative test decisions, especially for larger dimension. So, in this case, a rate that is in most cases lower than 0.04 would lead to a raised number of rejections when the null hypothesis is true. However, ψ_z has too high type-I error rates, especially in the case of small d=5. But, this improves for a higher dimension as well as a larger number of groups. For all dimensions, $\varphi_N^\star$ shows by far the best type-I error control rates and performs well with comparatively low dimensions or just a few groups. It can be seen that the error rates have less fluctuation for higher numbers of groups. The reason for this is that for fixed comparatively small sample sizes, an increasing number of groups not only improves the approximation but also is necessary to get reliable estimators.

In contrast, there is nearly no difference between ψ_χ and $\varphi_N^\star$ in Figure 2. This similarity is not surprising because from Figure 1 we know that f_P always has the value one. Furthermore, the small difference between both curves shows once more the good performance of the used estimators. Apart from that, again,

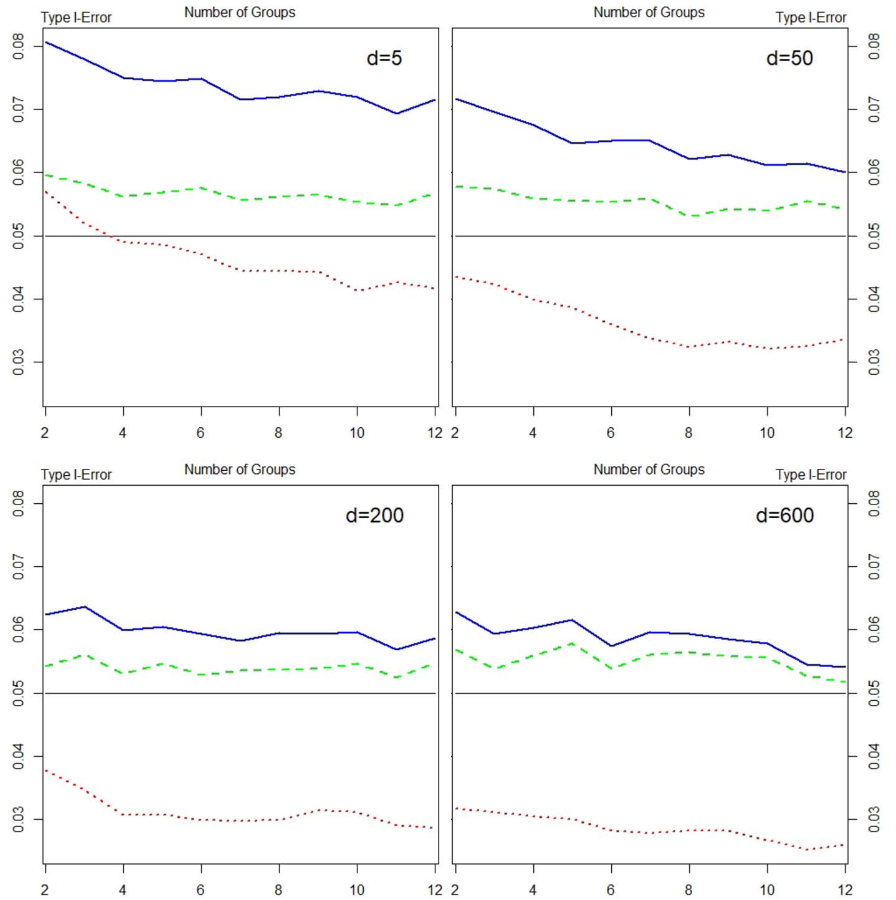


Fig 1: Simulated Type-I error rates ($\alpha = 5\%$) for ψ_z(—), ψ_χ ($\cdots$) and $\varphi_N^\star$(- -) under the null hypothesis $H_0^a : (\boldsymbol{P}_a \otimes \boldsymbol{P}_d)\,\boldsymbol{\mu} = \mathbf{0}$ for increasing dimension.

the performance of $\varphi_N^\star$ is quite good, particularly for a higher number of groups. Using the test φ_z that is based on the wrong limit distribution under $\mathcal{H}_0^b$ results in considerably larger type-I error rates between 0.065 and 0.085.

To sum up, $\varphi_N^\star$ shows really good type-I error rates, overall settings, dimensions, and group numbers, even for substantially unbalanced sample sizes, containing groups with just a few observations.

4.1.1. Power

The property to detect deviations from the nullhypothesis is investigated by considering the same distributional setting as for the type-I error rate, with

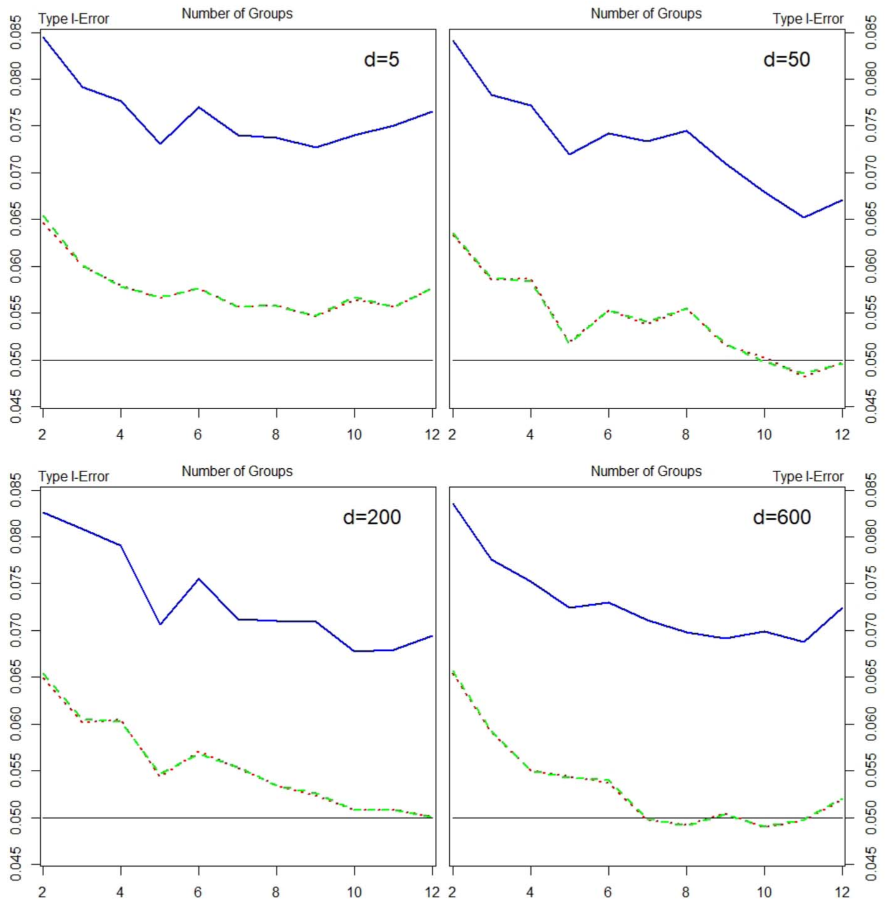


Fig 2: Simulated Type-I error rates ($\alpha = 5\%$) for ψ_z(—), ψ_χ (· · ·) and $\varphi_N^\star$(- -) under the null hypothesis $H_0^b : \left(\frac{1}{a}\boldsymbol{J}_a \otimes \frac{1}{d}\boldsymbol{J}_d\right)\boldsymbol{\mu} = \mathbf{0}$ for increasing dimension.

the same hypotheses. For this analysis we choose $d = 50$ and small($a = 2$), moderate($a = 4$) and large($a = 8$, $a = 10$) number of factor levels.

We are interested in three kinds of alternatives:

- a *trend-alternative* with $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_3 = ..., \boldsymbol{\mu}_9 = \mathbf{0}_d$ and $(\boldsymbol{\mu}_2)_k = (\boldsymbol{\mu}_4)_k, ..., (\boldsymbol{\mu}_{10})_k = \delta \cdot k/d,\ k = 1, ..., d, \quad \delta \in [0, 2]$,
- a *one-point-alternative* with $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_3 = ..., \boldsymbol{\mu}_9 = \mathbf{0}_d$ and $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_4, ..., \boldsymbol{\mu}_{10} = \delta \cdot \boldsymbol{e}_1, \quad \delta \in [0, 3.5]$ and
- a *shift-alternative* with $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_3 = ..., \boldsymbol{\mu}_9 = \mathbf{0}$ and $(\boldsymbol{\mu}_2) = (\boldsymbol{\mu}_4), ..., (\boldsymbol{\mu}_{10}) = \delta \cdot \mathbf{1}_d$ for $\mathcal{H}_0^b, \quad \delta \in [0, 2]$

Here $\boldsymbol{e}_\ell$ denotes the vector containing 1 in the $\ell - th$ component, and 0 elsewhere and $\mathbf{1}_d$ contains just 1's in each component.

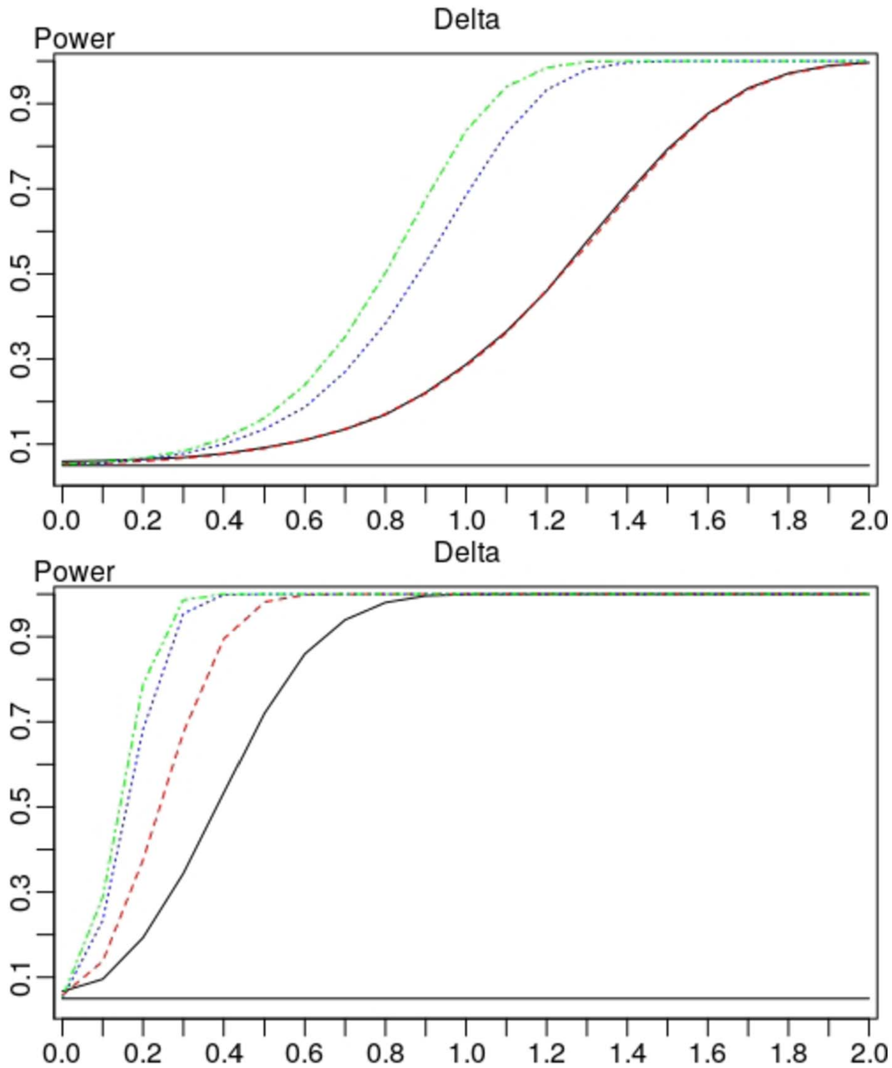


Fig 3: Simulated power curves of $\varphi_N^\star$ for a trend alternative with $d = 50$, 10000 simulation runs and an autoregressive structure($(\mathbf{\Sigma})_{i,j} = 0.6^{|i-j|}$). The sample size is $\boldsymbol{n} = (15, 15, 20, 35, 25, 20, 30, 30, 35, 20)$ and different numbers of groups were considered, namely $a = 2$(—), $a = 4$(- -), $a = 8$(···) and $a = 10$(· — ·—).

From the simulation result given in [18], it directly follows that it is challenging to detect the one-point alternative for $d = 50$ depending on the hypothesis. For this reason, we here consider a much larger value for δ.

For the trend alternative(Figure 3), $\varphi_N^\star$ has a high power for both null hypotheses where the power is essential higher for $\mathcal{H}_0^b$. Increasing the number of

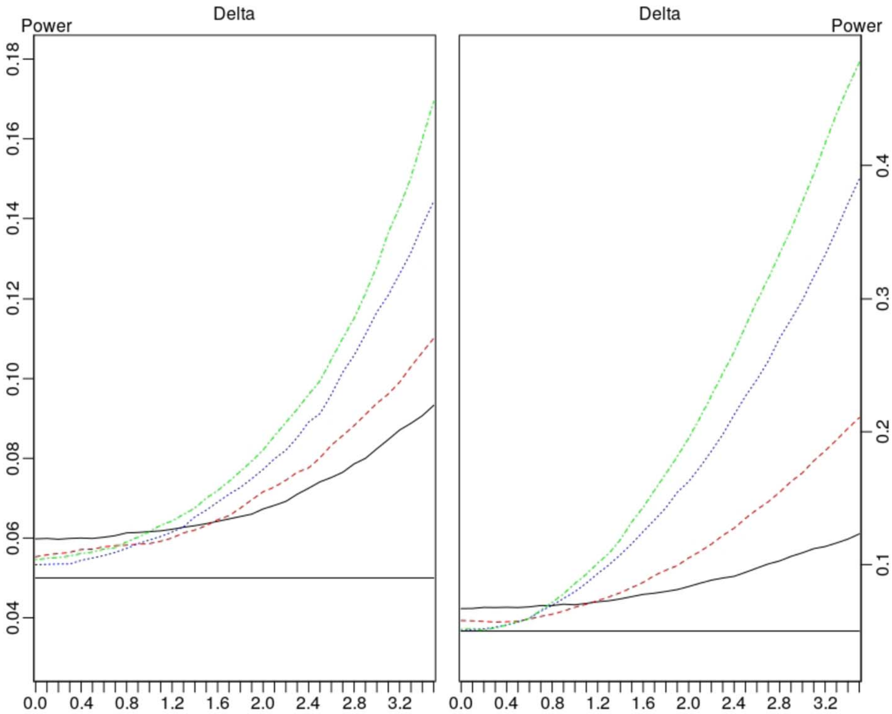


Fig 4: Simulated power curves of $\varphi_N^\star$ for a one-point alternative with $d = 50$, 10000 simulation runs and an autoregressive structure($(\mathbf{\Sigma})_{i,j} = 0.6^{|i-j|}$). The sample size is $\boldsymbol{n} = (15, 15, 20, 35, 25, 20, 30, 30, 35, 20)$and different numbers of groups were considered, namely $a = 2$(—), $a = 4$(- -), $a = 8$($\cdots$) and $a = 10$($\cdot - \cdot -$).

groups also increases the power in both hypotheses. It is noticeable that for $\mathcal{H}_0^a$ increasing the number from 8 to 10 groups has substantially more effect than from 2 to 4 groups while for $\mathcal{H}_0^b$ it's vice versa.

As expected, detecting the one-point alternative(Figure 4) is challenging for both hypotheses, so the power is low in both cases, even for larger δ- values in particular for $\mathcal{H}_0^a$. This observation coincides with the power calculations from [18]. But it can be seen that an increasing number of groups increase the power essentially.

Finally, we considered a shift alternative(Figure 5), but just for $\mathcal{H}_0^b$. As in other cases([17],[18]), this alternative is comparatively easy to detect. This holds in particular for an increasing number groups.

All in all, except for the one-point alternative, $\varphi_N^\star$ has very high power even for these small sample sizes, especially $n_1 = n_2 = 15$. Moreover, H_0^b is much easier to detect in all settings.

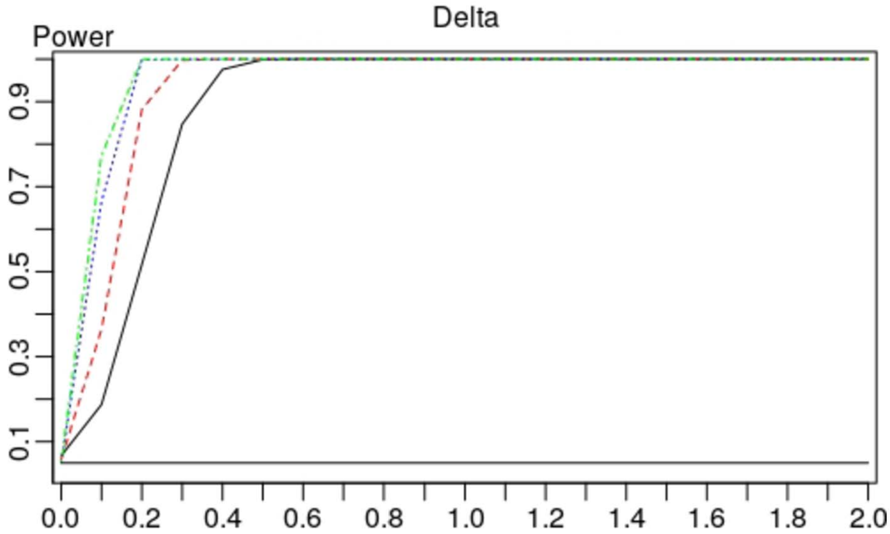


Fig 5: Simulated power curves of $\varphi_N^\star$ for a shift alternative with $d = 50$, 10000 simulation runs and an autoregressive structure($(\boldsymbol{\Sigma})_{i,j} = 0.6^{|i-j|}$). The sample size is $\boldsymbol{n} = (15, 15, 20, 35, 25, 20, 30, 30, 35, 20)$and different numbers of groups were considered, namely $a = 2$(—), $a = 4$(- -), $a = 8$(· · ·) and $a = 10$(· — ·—).

5. Conclusion

The present paper investigated a procedure for homoscedastic split-plot designs under various settings containing different kinds of potential high-dimensionality. Under equal covariance matrices or similar conditions (as mentioned in Section 2), results for settings with, for example, a large number of small independent groups are found. These kinds of data sets nowadays get more important because there is a trend to divide data sets more, e.g., in the context of personalized medicine or personalized insurance. Different from existing approaches, we take this development into account by considering a variety of different frameworks.

We were able to expand the central theorem of [18] also to cover this case for the price of the additional assumption of equal covariance matrices. Moreover, we generalized it to some more cases, in some sense completing the scope of the theorem. For all settings, we approximate the critical value of the test statistic by a standardized χ_f^2 distribution with appropriate f. To use these results, we developed estimators that can be used unattached of the asymptotic framework.

We conducted simulations to investigate the level of the resulting test as well as its power. The outcomes were convincing, especially for a larger number of groups.

Unfortunately, it is not that easy to verify the assumption of equal covariance matrices or just equal powers of traces. The most popular test under normality, Box's M-test [5], has quite good results but doesn't take care of our asymptotic frameworks. High-dimensional tests of equal covariance matrices are a field of great interest, which was, for example, investigated in [14] and [15]. We plan to combine their techniques with the results obtained in [19] in the near future.

Finally, various adjustments of estimators are planned to improve their performance when the homogeneity is violated.

6. Appendix

Proof of Theorem 1. For this proof, it is helpful to present the theorem in a more detailed way.

Let $\beta_s = \lambda_s \Big/ \sqrt{\sum_{\ell=1}^{ad} \lambda_\ell^2}$ for $s = 1, \dots, ad$. Then $\widetilde{W}_N$ has, under $H_0(\boldsymbol{T})$, and one of the frameworks I-V asymptotically

a) a standard normal distribution if and only if

$$\beta_1 = \max_{s \le ad} \beta_s \to 0 \quad \text{as} \quad N \to \infty,$$

b) a standardized $\left(\chi_1^2 - 1\right)/\sqrt{2}$ distribution if and only if

$$\beta_1 \to 1 \quad \text{as} \quad N \to \infty,$$

c) a distribution of the shape $\sum_{s=1}^{r} b_s \left(C_s - 1\right)/\sqrt{2} + \sqrt{1 - \sum_{s=1}^{r} b_s^2} \cdot Z$, if and only if

$$\text{for all } s \in \mathbb{N} \quad \beta_s \to b_s \quad \text{as} \quad N \to \infty,$$

for a decreasing sequence $(b_s)_s$ in $[0,1)$ with $r \in \mathbb{N} \setminus \{1\}$ with $b_r > 0$ and $b_{r+1} = 0$ with $C_i \overset{i.i.d.}{\sim} \chi_1^2$, $Z \sim \mathcal{N}(0,1)$.

d) a distribution of the shape $\sum_{s=1}^{\infty} b_s \left(C_s - 1\right)/\sqrt{2}$, if

$$\text{for all } s \in \mathbb{N} \quad \beta_s \to b_s \quad \text{as} \quad N \to \infty,$$

for a decreasing sequence $(b_s)_s$ in $(0,1)$ with $\sum_{s=1}^{\infty} b_s^2 = 1$ and $C_i \overset{i.i.d.}{\sim} \chi_1^2$.

The first two parts as well as the last one were proved in [18].

For part c) from Cramers theorem it is well known that it needs an infinite number of summands to get a normal distribution as limit distribution. So it exists a infinite amount $M \subset \mathbb{N}$ with

$$\sum_{\ell \in M} \beta_\ell \left(\frac{C_\ell - 1}{\sqrt{2}} \right) \xrightarrow{\mathcal{D}} \sqrt{1 - \sum_{s=1}^{r} b_s^2} \cdot Z.$$

The proof of part a) shows, that $\beta_\ell \to 0$ for all $\ell \in M$, and because of the decreasing order there exists an $r' \in \mathbb{N}$ with $b_{r'} > 0$ and $b_{r'+1} = 0$. Assume now that $\beta_\ell \to b'_\ell$ for $\ell = 1, ..., r'$ otherwise consider the subsequence where this holds. It remains to show that from

$$\sum_{\ell=1}^{r'} \beta_\ell \left(\frac{C_\ell - 1}{\sqrt{2}}\right) \to \sum_{\ell=1}^{r'} b'_\ell \left(\frac{C_\ell - 1}{\sqrt{2}}\right) \overset{\mathcal{D}}{=} \sum_{\ell=1}^{r} b_\ell \left(\frac{C_\ell - 1}{\sqrt{2}}\right),$$

it follows $r = r'$ as well as $b_\ell = b'_\ell$. To this aim, we consider the Moment-generating functions, so we know, for all $t \in \mathbb{R}$

$$\prod_{\ell=1}^{r'} \left(1 - \tfrac{2b'_\ell t}{\sqrt{2}}\right)^{-1/2} \exp\left(-t\tfrac{b'_\ell}{\sqrt{2}}\right) = \prod_{\ell=1}^{r} \left(1 - \tfrac{2b_\ell t}{\sqrt{2}}\right)^{-1/2} \exp\left(-t\tfrac{b_\ell}{\sqrt{2}}\right).$$

Thus, applying the continous mapping theorem we have for all $t \in \mathbb{R}$

$$\begin{aligned} \left(\prod_{\ell=1}^{r'} \left(1 - \tfrac{2b'_\ell t}{\sqrt{2}}\right)^{-1/2} \exp\left(-\tfrac{b'_\ell t}{\sqrt{2}}\right)\right)^{-2} &= \left(\prod_{\ell=1}^{r} \left(1 - \tfrac{2b_\ell t}{\sqrt{2}}\right)^{-1/2} \exp\left(-\tfrac{b_\ell t}{\sqrt{2}}\right)\right)^{-2} \\ \Leftrightarrow \prod_{\ell=1}^{r'} \left(1 - \sqrt{2}b'_\ell t\right) \exp(-\sqrt{2}b'_\ell t) &= \prod_{\ell=1}^{r} \left(1 - \sqrt{2}b_\ell t\right) \cdot \exp(-\sqrt{2}b_\ell t). \end{aligned}$$

Now we consider the zero points of both sides, which are a consequence of the polynomial parts and can be written by $\frac{1}{\sqrt{2}b_\ell}$ resp. $\frac{1}{\sqrt{2}b'_\ell}$. It can be directly inferred from this that both polynomials has the same degree and therefore $r' = r$. Moreover, both of them have the same zero points with the same multiplicity. So the coefficients are the same on both sides, and because of the decreasing order, it follows $b_\ell = b'_\ell$ for $\ell = 1, ..., r$. Therefore the result follows. □

Given the fact that framework III is not really high-dimensional, and I just partwise, it would be possible to use other more classical estimators for the unknown traces. Nevertheless, our focus was to develop preferably general estimators that can be used in various settings.

Lemma 1. *With*

$$A_{i,1} = \frac{1}{2} \sum_{\ell_1 \neq \ell_2 = 1}^{n_i} (\boldsymbol{X}_{i,\ell_1} - \boldsymbol{X}_{i,\ell_2})^\top \boldsymbol{T}_S (\boldsymbol{X}_{i,\ell_1} - \boldsymbol{X}_{i,\ell_2})$$

we can define

$$A_1 = \frac{1}{\sum_{i=1}^{a}(n_i - 1)n_i} \sum_{i=1}^{a} A_{i,1},$$

which is an unbiased and ratio consistent estimator for $\operatorname{tr}(\boldsymbol{T}_S\boldsymbol{\Sigma})$*, in all of our frameworks.*

Proof. It is obvious that this is a unbiased estimator of $\operatorname{tr}(\boldsymbol{T}_S\boldsymbol{\Sigma})$. With well known rules and analogous to [18] we calculate

$$Var(A_1) \leq \frac{1}{\left[\sum_{i=1}^{a}\binom{n_i}{2}\right]^2} \sum_{i=1}^{a} \binom{n_i}{2} \left(\binom{n_i}{2} - \binom{n_i - 2}{2}\right) \cdot \mathcal{O}(\operatorname{tr}^2(\boldsymbol{T}_S\boldsymbol{\Sigma})).$$

Now we need a case analysis which is done for some of the following proofs. So the first one is in detail and the other proofs are shorter. At first we consider the case where $n_{\max} \to \infty$. Then

$$\begin{aligned} Var(A_1) &\leq \frac{1}{\left[\sum_{i=1}^a \binom{n_i}{2}\right]\cdot\binom{n_{\max}}{2}} \sum_{i=1}^a \binom{n_i}{2}\left(\binom{n_i}{2} - \binom{n_i-2}{2}\right)\cdot \mathcal{O}(\mathrm{tr}^2(\boldsymbol{T}_S\boldsymbol{\Sigma})) \\ &\leq \frac{1}{\left[\sum_{i=1}^a \binom{n_i}{2}\right]\cdot\binom{n_{\max}}{2}} \sum_{i=1}^a \binom{n_i}{2}\left(\binom{n_{\max}}{2} - \binom{n_{\max}-2}{2}\right)\cdot \mathcal{O}(\mathrm{tr}^2(\boldsymbol{T}_S\boldsymbol{\Sigma})) \\ &= \frac{\left(\binom{n_{\max}}{2} - \binom{n_{\max}-2}{2}\right)}{\binom{n_{\max}}{2}}\cdot \mathcal{O}(\mathrm{tr}^2(\boldsymbol{T}_S\boldsymbol{\Sigma})) \\ &= \mathcal{O}\left(n_{\max}^{-1}\right)\cdot \mathcal{O}(\mathrm{tr}^2(\boldsymbol{T}_S\boldsymbol{\Sigma})). \end{aligned}$$

For the other case $n_{\max}$ is bound and $a \to \infty$. In this situation it holds

$$\begin{aligned} Var(A_1) &\leq \frac{1}{\left[\sum_{i=1}^a \binom{n_i}{2}\right]\cdot a\cdot\binom{n_{\min}}{2}} \sum_{i=1}^a \binom{n_i}{2}\left(\binom{n_{\max}}{2} - \binom{n_{\max}-2}{2}\right)\cdot \mathcal{O}(\mathrm{tr}^2(\boldsymbol{T}_S\boldsymbol{\Sigma})) \\ &= \frac{\left(\binom{n_{\max}}{2} - \binom{n_{\max}-2}{2}\right)}{a\cdot\binom{n_{\min}}{2}}\cdot \mathcal{O}(\mathrm{tr}^2(\boldsymbol{T}_S\boldsymbol{\Sigma})) \\ &= \mathcal{O}\left(a^{-1}\right)\cdot \mathcal{O}(\mathrm{tr}^2(\boldsymbol{T}_S\boldsymbol{\Sigma})) \end{aligned}$$

So dividing by $\mathrm{tr}^2(\boldsymbol{T}_S\boldsymbol{\Sigma})$ and then using the Tschebyscheff inequality leads to the results in both cases. □

For the estimated version of the standardized quadratic form, one more estimator is needed.

Lemma 2. *The estimator, given by*

$$A_2 = \sum_{i=1}^a \sum_{\substack{\ell_1,\ell_2=1\\ \ell_1>\ell_2}}^{n_i} \sum_{\substack{k_2=1\\ k_2\neq\ell_1\neq\ell_2}}^{n_i} \sum_{\substack{k_1=1\\ \ell_2\neq\ell_1\neq k_1>k_2}}^{n_i} \frac{\left[\left(\boldsymbol{X}_{i,\ell_1} - \boldsymbol{X}_{i,\ell_2}\right)^\top \boldsymbol{T}_S\left(\boldsymbol{X}_{i,k_1} - \boldsymbol{X}_{i,k_2}\right)\right]^2}{4\cdot 6\sum_{i=1}^a \binom{n_i}{4}},$$

is an unbiased and ratio-consistent estimator of $\mathrm{tr}\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^2\right)$ *in all our asymptotic frameworks.*

Proof. Again the unbiasedness is clear, and we consider the variance.

We calculate, with $\boldsymbol{Y}_{i,\ell_1,\ell_2} := \boldsymbol{T}_S\left(\boldsymbol{X}_{i,k_1} - \boldsymbol{X}_{i,k_2}\right)$,

$$\begin{aligned} &Var(A_2) \\ =\ & \left[24\sum_{i=1}^a \binom{n_i}{4}\right]^{-2} \sum_{i=1}^a Var\left(\sum_{\substack{\ell_1,\ell_2=1\\ \ell_1>\ell_2}}^{n_i} \sum_{\substack{k_2=1\\ k_2\neq\ell_1\neq\ell_2}}^{n_i} \sum_{\substack{k_1=1\\ \ell_2\neq\ell_1\neq k_1>k_2}}^{n_i} \left[\boldsymbol{Y}_{i,\ell_1,\ell_2}^\top \boldsymbol{Y}_{i,k_1,k_2}\right]^2\right) \\ \leq\ & \frac{\sum_{i=1}^a \binom{n_i}{4}\left(\binom{n_i}{4} - \binom{n_i-4}{4}\right)}{\left[4\cdot\sum_{i=1}^a \binom{n_i}{4}\right]^2} \mathcal{O}\left(\mathrm{tr}^2\left((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^2\right)\right). \end{aligned}$$

Similar as before for $n_{\max} \to \infty$ we get

$$Var(A_2) \leq \mathcal{O}\left({n_{\max}}^{-1}\right)\cdot \mathcal{O}\left(\mathrm{tr}^2\left((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^2\right)\right)$$

and for $n_{\max}$ bound and $a \to \infty$

$$Var(A_2) \leq \mathcal{O}\left(a^{-1}\right) \cdot \mathcal{O}\left(\operatorname{tr}^2\left((\boldsymbol{T}_S \boldsymbol{\Sigma}_i)^2\right)\right).$$

Again the result follows by using Tschebyscheff's inequality. □

With these theorems, the usage of the estimated standardized quadratic form can be justified.

Proof of Theorem 2. The result follows directly by theorem 3.2 from [18]. □

For the proof of Theorem 4, we need to show different properties that combined lead to the result.

Proof of Theorem 4. We conduct this proof in several steps:

a) $\mathbb{E}(C_1) = \operatorname{tr}\left((\boldsymbol{T}_S \boldsymbol{\Sigma})^3\right)$,

b) $Var(C_1) = \frac{\sum_{j=1}^a \binom{n_j}{6}\left(\binom{n_j}{6} - \binom{n_j-6}{6}\right)}{\left(\sum_{i=1}^a \binom{n_i}{6}\right)^2} \cdot \mathcal{O}\left(\operatorname{tr}^3\left((\boldsymbol{T}_S \boldsymbol{\Sigma})\right)\right)$,

c) $\frac{C_1}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}_S \boldsymbol{\Sigma})^2\right)} - \frac{\operatorname{tr}\left((\boldsymbol{T}_S \boldsymbol{\Sigma})^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}_S \boldsymbol{\Sigma})^2\right)} \xrightarrow{\mathcal{P}} 0$ in our frameworks I-V,

d) $\frac{C_1^2}{A_2^4} - (f_P)^{-1} \xrightarrow{\mathcal{P}} 0$ in our frameworks I-V.

The results from [18] directly yield to

$$\mathbb{E}(C_1) = \operatorname{tr}\left((\boldsymbol{T}_S \boldsymbol{\Sigma})^3\right)$$

and

$$Var(C_1) = \sum_{i=1}^a \frac{Var(C_{i,1})}{6! \cdot \sum\limits_{j=1}^a \binom{n_j}{6}} \leq \frac{\sum_{j=1}^a \binom{n_j}{6}\left(\binom{n_j}{6} - \binom{n_j-6}{6}\right)}{\left(\sum_{i=1}^a \binom{n_i}{6}\right)^2} \cdot \mathcal{O}\left(\operatorname{tr}^3\left((\boldsymbol{T}_S \boldsymbol{\Sigma})\right)\right)$$

which proves a) and b). Together with Tschebychefs inequality this leads to an unbiased ratio consistent estimator for $\operatorname{tr}\left((\boldsymbol{T}_S \boldsymbol{\Sigma})^3\right)$.

For part c) we calculate

$$\mathbb{E}\left(\frac{C_1}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}_S \boldsymbol{\Sigma})^2\right)} - \frac{\operatorname{tr}\left((\boldsymbol{T}_S \boldsymbol{\Sigma})^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}_S \boldsymbol{\Sigma})^2\right)}\right) = 0$$

and

$$Var\left(\frac{C_1}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}_S \boldsymbol{\Sigma})^2\right)} - \frac{\operatorname{tr}\left((\boldsymbol{T}_S \boldsymbol{\Sigma})^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}_S \boldsymbol{\Sigma})^2\right)}\right)$$

$$= \frac{Var\left(C_1\right)}{\operatorname{tr}^3\left((\boldsymbol{T}_S \boldsymbol{\Sigma})^2\right)} \leq 27 \cdot \frac{\sum_{j=1}^a \binom{n_j}{6}\left(\binom{n_j}{6} - \binom{n_j-6}{6}\right)}{\left(\sum_{i=1}^a \binom{n_i}{6}\right)^2} \cdot \mathcal{O}(1)$$

Again this number is in $\mathcal{O}(n_{\max}^{-1})$ for $n_{\max} \to \infty$ and in $\mathcal{O}(a^{-1})$ for $a \to \infty$. So in both cases the result follows with the Tschebyscheff-inequality.

At last, the proof of part d) is done using the above results. A similar proof is part of [18], but we repeat it for better understanding.

With the last lemma it follows for both cases that

$$\begin{aligned}
&\frac{C_1^2}{\operatorname{tr}^3\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^2\right)} - \frac{1}{f_P}\\
&= \left(\frac{C_1}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^2\right)}\right)^2 - \left(\frac{\operatorname{tr}\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^2\right)}\right)^2\\
&= \left[\frac{C_1}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^2\right)} - \frac{\operatorname{tr}\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^2\right)}\right]\left[\frac{C_1}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^2\right)} + \frac{\operatorname{tr}\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^2\right)}\right]\\
&= o_P(1)\cdot\left[\tfrac{C_1}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^2\right)} - \tfrac{\operatorname{tr}\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^2\right)} + 2\tfrac{\operatorname{tr}\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^2\right)}\right] = o_P(1),
\end{aligned}$$

were for the last step the trace inquality was used together with Slutzky's theorem. With the ratio-consistency of A_2 it follows $A_2/\operatorname{tr}((\boldsymbol{T}_S\boldsymbol{\Sigma})) \xrightarrow{\mathcal{P}} 1$ and because of continous mapping $\operatorname{tr}^3((\boldsymbol{T}_S\boldsymbol{\Sigma}))/A_2^3 \xrightarrow{\mathcal{P}} 1$. This leads to

$$\begin{aligned}
\frac{C_1^2}{A_2^3} - (f_P)^{-1} &= \frac{\operatorname{tr}^3\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^2\right)}{A_2^3}\frac{C_1^2}{\operatorname{tr}^3\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^2\right)} - (f_P)^{-1}\\
&= (1 + o_P(1))\cdot\tfrac{1}{\widehat{f}_P} - \tfrac{1}{f_P}\\
&= \frac{1}{\widehat{f}_P} - \frac{1}{f_P} + o_P(1)\cdot\tfrac{1}{\widehat{f}_P} = o_P(1). \qquad \square
\end{aligned}$$

It is obvious that this estimator needs a sufficiently large amount of groups with at least six observations. Similar for the other estimators, which were introduced earlier. From a theoretical point of view, a scenario with $n_{\max} \leq 5$ is part of our model. In practice, however, this setting is rarely examined. In this case, it would be possible to define some estimators which combine observations from different groups, which would be much more complicated than our estimators.

Proof of Theorem 5. For this proof, some results of [18] are used and adapted. First the expactation value of the estimator, using the notation $w := \sum_{i=1}^a w_i$:

$$\begin{aligned}
\mathbb{E}\left(C_1^\star\right) &= \mathbb{E}\left(\frac{1}{8\cdot\sum_{i=1}^a w_i}\sum_{i=1}^a C_{i,1}^\star\right)\\
&= \frac{1}{8\cdot w}\sum_{i=1}^a \mathbb{E}\left(C_{i,1}^\star\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{8\cdot w}\sum_{i=1}^{a}\mathbb{E}\left(\sum_{b=1}^{w_i}\Lambda_1(\boldsymbol{\sigma}_i(b))\cdot\Lambda_2(\boldsymbol{\sigma}_i(b))\cdot\Lambda_3(\boldsymbol{\sigma}_i(b))\right)\\
&= \frac{1}{8\cdot w}\sum_{i=1}^{a}\sum_{b=1}^{w_i}\mathbb{E}\left(\Lambda_1(\boldsymbol{\sigma}_i(b))\cdot\Lambda_2(\boldsymbol{\sigma}_i(b))\cdot\Lambda_3(\boldsymbol{\sigma}_i(b))\right)\\
&= \frac{1}{8\cdot w}\sum_{i=1}^{a} w_i\cdot\mathbb{E}\left(\Lambda_1(1,2,3,4,5,6)\cdot\Lambda_2(1,2,3,4,5,6)\cdot\Lambda_3(1,2,3,4,5,6)\right)\\
&= \frac{1}{8\cdot w}\sum_{i=1}^{a} w_i\cdot 8\,\mathrm{tr}\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^3\right)\\
&= \mathrm{tr}\left((\boldsymbol{T}_S\boldsymbol{\Sigma})^3\right).
\end{aligned}
$$

With Theorem A.9 Theorem A.10 and Theorem A.16 from [18] for the variance we get

$$
\begin{aligned}
\mathit{Var}\left(C_1^\star\right) &= \tfrac{1}{(8\cdot w)^2}\sum_{i=1}^{a}\mathit{Var}\left(C_{i,1}^\star\right)\\
&\le \tfrac{1}{(8\cdot w)^2}\sum_{i=1}^{a} w_i^2\cdot\left[0+1-\left(1-\tfrac{1}{w_i}\right)\cdot\frac{\binom{n_i-6}{6}}{\binom{n_i}{6}}\right].
\end{aligned}
$$

Again there the same two cases. If $n_{\max}$ is bound and therefore $\max\limits_{i=1,\ldots,a}(w_i)$ is bound, it follows $a\to\infty$ and hereby

$$
\begin{aligned}
&\tfrac{1}{(8\cdot w)^2}\sum_{i=1}^{a} w_i^2\cdot\left[0+1-\left(1-\tfrac{1}{w_i}\right)\cdot\frac{\binom{n_i-6}{6}}{\binom{n_i}{6}}\right]\\
\le\;& \frac{1}{(8\cdot w)\cdot a\cdot\min\limits_{i=1,\ldots,a}(w_i)}\cdot\max_{i=1,\ldots,a}(w_i)\sum_{i=1}^{a} w_i\cdot 1\\
=\;& \mathcal{O}\left(a^{-1}\right)\cdot\frac{\max\limits_{i=1,\ldots,a}(w_i)}{\min\limits_{i=1,\ldots,a}(w_i)}\\
=\;& \mathcal{O}\left(a^{-1}\right)
\end{aligned}
$$

while for $n_{\max}\to\infty$ which implies $\max\limits_{i=1,\ldots,a}(w_i)\to\infty$ we calculate first

$$
\begin{aligned}
&w_i^2\cdot\left[0+1-\left(1-\frac{1}{w_i}\right)\cdot\frac{\binom{n_i-6}{6}}{\binom{n_i}{6}}\right]\\
=\;& w_i\cdot\left[w_i\cdot\left(1-\frac{\binom{n_i-6}{6}}{\binom{n_i}{6}}\right)+\frac{\binom{n_i-6}{6}}{\binom{n_i}{6}}\right]\\
\le\;& w_i\cdot\left[\left(\upsilon\cdot\binom{n_i}{6}+1\right)\left(1-\frac{\binom{n_i-6}{6}}{\binom{n_i}{6}}\right)+\frac{\binom{n_i-6}{6}}{\binom{n_i}{6}}\right]\\
=\;& w_i\cdot\left[\upsilon\left(\binom{n_i}{6}-\binom{n_i-6}{6}\right)+1\right]
\end{aligned}
$$

$$\leq \quad w_i \cdot \left[\upsilon \left(\binom{n_{\min}}{6} - \binom{n_{\min} - 6}{6} \right) + 1 \right]$$

and therefore

$$\begin{aligned}
&\leq \frac{1}{(8 \cdot w)^2} \sum_{i=1}^{a} w_i^2 \cdot \left[0 + 1 - \left(1 - \frac{1}{w_i}\right) \cdot \frac{\binom{n_i - 6}{6}}{\binom{n_i}{6}} \right] \\
&\leq \frac{1}{(8 \cdot w)^2} \sum_{i=1}^{a} w_i \cdot \left[\upsilon \left(\binom{n_{\min}}{6} - \binom{n_{\min} - 6}{6} \right) + 1 \right] \\
&\leq \frac{1}{(64 \cdot w) \cdot \max\limits_{i=1,\dots,a} (w_i)} \sum_{i=1}^{a} w_i \cdot \left[\upsilon \left(\binom{n_{\min}}{6} - \binom{n_{\min} - 6}{6} \right) + 1 \right] \\
&\leq \frac{1}{(64 \cdot w) \cdot \left(\upsilon \cdot \binom{n_{\max}}{6} - 1\right)} \sum_{i=1}^{a} w_i \cdot \left[\upsilon \left(\binom{n_{\min}}{6} - \binom{n_{\min} - 6}{6} \right) + 1 \right] \\
&= \frac{\left[\upsilon \left(\binom{n_{\min}}{6} - \binom{n_{\min} - 6}{6} \right) + 1 \right]}{64 \cdot \left(\upsilon \cdot \binom{n_{\max}}{6} - 1\right)} = \mathcal{O}\left(n_{\max}^{-1}\right).
\end{aligned}$$

Combining these results, the remainder of the proof follows analogously to the proof of Theorem 4. □

Acknowledgments

The author would like to thank Markus Pauly for helpful discussions and many valuable suggestions.

References

[1] Annadurai, Kalaivani; Danasekaran, Raja; Mani, Geetha: Personalized medicine: A paradigm shift towards promising health care. In: *Journal of pharmacy and bioallied sciences* 8 (2016), Nr. 1, S. 77–78.

[2] Bathke, Arne C.: ANOVA for a large number of treatments. In: *Mathematical Methods of Statistics* 11 (2002), 01. MR1900976

[3] Bathke, Arne C.; Lankowski, Dörte: Rank procedures for a large number of treatments. In: *Journal of Statistical Planning and Inference* 133 (2005), 08, S. 223–238. http://dx.doi.org/10.1016/j.jspi.2004.03.020. – DOI 10.1016/j.jspi.2004.03.020. MR2194475

[4] Bathke, Arne C.; Friedrich, Sarah; Pauly, Markus; Konietschke, Frank; Staffen, Wolfgang; Strobl, Nicolas; Höller, Yvonne: Testing Mean Differences among Groups: Multivariate and Repeated Measures Analysis with Minimal Assumptions. In: *Multivariate Behavioral Research* 53 (2018), 03, S. 348–359. http://dx.doi.org/10.1080/00273171.2018.1446320. – DOI 10.1080/00273171.2018.1446320.

[5] Box, G. E. P.: NON-NORMALITY AND TESTS ON VARIANCES. In: *Biometrika* 40 (1953), Nr. 3-4, 318–335. http://dx.doi.org/10.1093/biomet/40.3-4.318. – DOI 10.1093/biomet/40.3–4.318. MR0058937

[6] Brunner, Edgar; Bathke, Arne C.; Placzek, Marius: Estimation of Box's e for low- and high-dimensional repeated measures designs with unequal covariance matrices. In: *Biometrical journal. Biometrische Zeitschrift* 54 (2012), 05, S. 301–16. http://dx.doi.org/10.1002/bimj.201100160. – DOI 10.1002/bimj.201100160. MR2931076

[7] Chen, Song X.; Qin, Ying-Li: A two-sample test for high-dimensional data with applications to gene-set testing. In: *Ann. Statist.* 38 (2010), 04, Nr. 2, 808–835. http://dx.doi.org/10.1214/09-AOS716. – DOI 10.1214/09–AOS716. MR2604697

[8] Friedrich, Sarah; Brunner, Edgar; Pauly, Markus: Permuting longitudinal data in spite of the dependencies. In: *Journal of Multivariate Analysis* 153 (2017), S. 255–265. MR3578850

[9] Happ, Martin; Harrar, Solomon W.; Bathke, Arne C.: Inference for low- and high-dimensional multigroup repeated measures designs with unequal covariance matrices. In: *Biometrical journal. Biometrische Zeitschrift* 58 4 (2016), S. 810–30. MR3527417

[10] Harrar, Solomon W.; Kong, Xiaoli: High-dimensional multivariate repeated measures analysis with unequal covariance matrices. In: *Journal of Multivariate Analysis* 145 (2016), 1–21. http://dx.doi.org/10.1016/j.jmva.2015.11.012. – DOI https://doi.org/10.1016/j.jmva.2015.11.012. – ISSN 0047–259X. MR3459934

[11] In: Huppertz, Berthold; Holzinger, Andreas: *Biobanks – A Source of Large Biological Data Sets: Open Problems and Future Challenges*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. – ISBN 978–3–662–43968–5, 317–330.

[12] Kong, Xiaoli; Harrar, Solomon W.: Accurate inference for repeated measures in high dimensions. In: *Electron. J. Statist.* 13 (2019), Nr. 2, 4916–4944. http://dx.doi.org/10.1214/19-EJS1644. – DOI 10.1214/19–EJS1644. MR4040758

[13] Konietschke, F.; Bathke, Arne C.; Harrar, S.W.; Pauly, M.: Parametric and nonparametric bootstrap methods for general MANOVA. In: *Journal of Multivariate Analysis* 140 (2015), S. 291–301. MR3372569

[14] Li, Jun; Chen, Song X.: Two sample tests for high-dimensional covariance matrices. In: *Ann. Statist.* 40 (2012), 04, Nr. 2, 908–940. http://dx.doi.org/10.1214/12-AOS993. – DOI 10.1214/12–AOS993. MR2985938

[15] Li, Weiming; Qin, Yingli: Hypothesis testing for high-dimensional covariance matrices. In: *Journal of Multivariate Analysis* 128 (2014), 108–119. http://dx.doi.org/https://doi.org/10.1016/j.jmva.2014.03.013. – DOI https://doi.org/10.1016/j.jmva.2014.03.013. – ISSN 0047–259X. MR3199831

[16] Mathai, Arakaparampil M.; Provost, Serge B.: *Quadratic forms in random variables, vol. 126 of Statistics: Textbooks and Monographs*, (1992).

MR1192786

[17] Pauly, Markus; Ellenberger, David; Brunner, Edgar: Analysis of high-dimensional one group repeated measures designs. In: *Statistics* 49 (2015), S. 1243–1261. MR3415218

[18] Sattler, Paavo; Pauly, Markus: Inference for high-dimensional split-plot-designs: A unified approach for small to large numbers of factor levels. In: *Electron. J. Statist.* 12 (2018), Nr. 2, 2743–2805. http://dx.doi.org/10.1214/18-EJS1465. – DOI 10.1214/18–EJS1465. MR3854906

[19] Sattler, Paavo; Bathke, Arne C.; Pauly, Markus: Testing Hypotheses about Covariance Matrices in General MANOVA Designs, (2021). arXiv:1909.06205

Article 3

Sattler, P., Bathke, A. C. and Pauly, M. (2022). Testing hypotheses about covariance matrices in general MANOVA designs. Journal of Statistical Planning and Inference. 219:134-146. doi.org/10.1016/j.jspi.2021.12.001

213

225

Bibliography

Aitkin, M. A., Nelson, W. C., and Reinfurt, K. H. (1968). Tests for correlation matrices. *Biometrika*, 55(2):327–334.

Akritas, M. G. and Brunner, E. (1997). A unified approach to rank tests for mixed models. *Journal of Statistical Planning and Inference*, 61(2):249–277.

Anderson, T. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics. Wiley.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis.* 2nd ed. Wiley, New York, USA.

Bai, Z., Hu, J., Wang, C., and Zhang, C. (2019). Test on the linear combinations of covariance matrices in high-dimensional data. pages 701–719.

Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statistica Sinica*, 6(2):311–329.

Bartlett, M. S. (1951). The effect of standardization on a chi square approximation in factor analysis. *Biometrika*, 38(3/4):337–344.

Bathke, A. (2002). Anova for a large number of treatments. *Mathematical Methods of Statistics*, 11(1):118–132.

Bathke, A. C., Friedrich, S., Pauly, M., Konietschke, F., Staffen, W., Strobl, N., and Höller, Y. (2018). Testing mean differences among groups: Multivariate and repeated measures analysis with minimal assumptions. *Multivariate Behavioral Research*, 53:348–359.

Bathke, A. C. and Lankowski, D. (2005). Rank procedures for a large number of treatments. *Journal of Statistical Planning and Inference*, 133(2):223–238.

Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.*, 36(1):199–227.

Bjerhammar, A. (1951). Rectangular reciprocal matrices, with special reference to geodetic calculations. *Bulletin Géodésique (1946-1975)*, 20(1):188–220.

Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrika*, 40(3-4):318–335.

Box, G. E. P. et al. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics*, 25(2):290–302.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31(2):144–152.

Browne, M. and Shapiro, A. (1986). The asymptotic covariance matrix of sample correlation coefficients under general conditions. *Linear Algebra and its Applications*, 82:169–176.

Brunner, E. (2001). Asymptotic and approximate analysis of repeated measures designs under heteroscedasticity.

Brunner, E., Bathke, A. C., and Konietschke, F. (2019). *Rank and Pseudo-Rank Procedures for Independent Observations in Factorial Designs*. Springer.

Brunner, E., Dette, H., and Munk, A. (1997). Box-type approximations in nonparametric factorial designs. *Journal of the American Statistical Association*, 92(440):1494–1502.

Brunner, E. and Puri, M. L. (2001). Nonparametric methods in factorial designs. *Statistical papers*, 42(1):1–52.

Cai, T. T., Ren, Z., and Zhou, H. H. (2016). Rejoinder of "estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation". *Electron. J. Statist.*, 10(1):81–89.

Chen, S. X. and Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.*, 38(2):808–835.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3):297–334.

Ditzhaus, M., Dobler, D., and Pauly, M. (2021). Inferring median survival differences in general factorial designs via permutation tests. *Statistical Methods in Medical Research*, 30(3):875–891.

Dobler, D. and Pauly, M. (2020). Factorial analyses of treatment effects under independent right-censoring. *Statistical Methods in Medical Research*, 29(2):325–343. PMID: 30834811.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1–26.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd.

Friedrich, S., Brunner, E., and Pauly, M. (2017). Permuting longitudinal data in spite of the dependencies. *Journal of Multivariate Analysis*, 153:255–265.

Friedrich, S., Konietschke, F., and Pauly, M. (2019). *MANOVA.RM: Analysis of Multivariate Data and Repeated Measures Designs*. R package version 0.3.2.

Friedrich, S. and Pauly, M. (2017). Mats: Inference for potentially singular and heteroscedastic manova. *Journal of Multivariate Analysis*, 165:166–179.

Google Scholar (2021). `https://scholar.google.com/`. Last accessed on 04.01.2021.

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(1):101–114.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10(4):255–282.

Hallal, P., Victora, C., Wells, J., Costa, R., and Valle, N. (2004). Comparison of short and full-length international physical activity questionnaires. *Journal of Physical Activity and Health*, 1(3):227–234.

Happ, M., Harrar, S. W., and Bathke, A. C. (2016). Inference for low- and high-dimensional multigroup repeated measures designs with unequal covariance matrices. *Biometrical Journal. Biometrische Zeitschrift*, 58(4):810–30.

Harrar, S. W. and Kong, X. (2016). High-dimensional multivariate repeated measures analysis with unequal covariance matrices. *Journal of Multivariate Analysis*, 145:1–21.

Herzog, W., Boomsma, A., and Reinecke, S. (2007). The model-size effect on traditional and modified tests of covariance structures. *Structural Equation Modeling A Multidisciplinary Journal*, 14(3):361–390.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.*, 19(3):293–325.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and monte carlo. *Ann. Statist.*, 1(5):799–821.

Jennrich, R. I. (1970). An asymptotic chi square test for the equality of two correlation matrices. *Journal of the American Statistical Association*, 65(330):904–912.

Joereskog, K. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, 43(4):443–477.

Jordan, W., Tumani, H., Cohrs, S., Eggert, S., Rodenbeck, A., Brunner, E., Rüther, E., and Hajak, G. (2004). Prostaglandin-d-synthase (beta-trace) in healthy human sleep. *Sleep*, 27:867–874.

Kincaid, C. (2005). Guidelines for selecting the covariance structure in mixed model analysis. *Statistics and Data Analysis*, 30.

Kong, X. and Harrar, S. W. (2019). Accurate inference for repeated measures in high dimensions. *Electron. J. Statist.*, 13(2):4916–4944.

Magnus, J. and Neudecker, H. (1980). The elimination matrix: Some lemmas and applications. Other publications tisem, Tilburg University, School of Economics and Management.

Mathai, A. M. and Provost, S. B. (1992). Quadratic forms in random variables, vol. 126 of statistics: Textbooks and monographs.

Moore, E. H. (1920). On the reciprocal of the general algebraic matrix. *Bull. Am. Math. Soc.*, 26:394–395.

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

Nel, D. (1985). A matrix derivation of the asymptotic covariance matrix of sample correlation coefficients. *Linear Algebra and its Applications*, 67:137–145.

Nordhausen, K., Sirkia, Seija Oja, H., and Tyler, D. E. (2018). *MANOVA.RM: Analysis of Multivariate Data and Repeated Measures Designs*.

Omelka, M. and Pauly, M. (2012). Testing equality of correlation coefficients in two populations via permutation methods. *Journal of Statistical Planning and Inference*, 142(6):1396–1406.

Omer, M. A., Johnson, D. A., and Rowe, R. C. (2000). Recovery of verticillium dahliae from north american certified seed potatoes and characterization of strains by vegetative compatibility and aggressiveness. *American Journal of Potato Research*, 77(5):325–331.

Pauly, M., Ellenberger, D., and Brunner, E. (2015). Analysis of high-dimensional one group repeated measures designs. *Statistics*, 49(6):1243–1261.

Pauly, M., Umlauft, M., and Ünlü, A. (2016). Resampling-based inference methods for comparing two coefficients alpha. *Psychometrika*, 83:203–222.

Penrose, R. (1955). A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51(3):406–413.

Rao, C. and Mitra, S. (1971). *Generalized inverse of matrices and its applications*. Wiley New York.

Revelle, W. (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 1.9.12.

Roszkowski, M. and Bean, A. (1990). Believe it or not! longer questionnaires have lower response rates. *Journal of Business and Psychology*, 4(4):495–509.

Ryan, T. and Woodall, W. (2005). The most-cited statistical papers. *Journal of Applied Statistics*, 32(5):461–474.

Sattler, P. (2021). A comprehensive treatment of quadratic-form-based inference in repeated measures designs under diverse asymptotics. *Electronic Journal of Statistics*, 15(1):3611 – 3634.

Sattler, P., Bathke, A. C., and Pauly, M. (2022). Testing hypotheses about covariance matrices in general manova designs. *Journal of Statistical Planning and Inference*, 219:134–146.

Sattler, P. and Pauly, M. (2018). Inference for high-dimensional split-plot-designs: A unified approach for small to large numbers of factor levels. *Electron. J. Statist.*, 12(2):2743–2805.

Sproule, R. N. (1974). Asymptotic properties of u-statistics. *Transactions of the American Mathematical Society*, 199:55–64.

Srivastava, M. S. and Kubokawa, T. (2013). Tests for multivariate analysis of variance in high dimension under non-normality. *Journal of Multivariate Analysis*, 115:204–216.

Srivastava, M. S., Yanagihara, H., and Kubokawa, T. (2014). Tests for covariance matrices in high dimension with less sample size. *Journal of Multivariate Analysis*, 130:289–309.

Staffen, W., Strobl, N., Zauner, H., Höller, Y., Dobesberger, J., and Trinka, E. (2014). Combining spect and eeg analysis for assessment of disorders with amnestic symptoms to enhance accuracy in early diagnostics. *Poster A19 Presented at the 11th Annual Meeting of the Austrian Society of Neurology, 26–29 March, Salzburg, Austria*.

Steiger, J. H. (1980). Testing pattern hypotheses on correlation matrices: Alternative statistics and some empirical results. *Multivariate Behavioral Research*, 15(3):335–352. PMID: 26794186.

Vallejo, G., Fernández, M., and Livacic-Rojas, P. E. (2010). Analysis of unbalanced factorial designs with heteroscedastic data. *Journal of Statistical Computation and Simulation*, 80(1):75–88.

Wakaki, H., Eguchi, S., and Fujikoshi, Y. (1990). A class of tests for a general covariance structure. *Journal of Multivariate Analysis*, 32(2):313–325.

Web of Science (2021). `https://apps.webofknowledge.com`. Last accessed on 2021-01-04.

Werner, C. (2004). Dimensionsstabile approximation für verteilungen von quadratischen formen im repeated-measures-design. *Technical Report*.

Wu, L., Weng, C., Wang, X., Wang, K., and Liu, X. (2018). Test of covariance and correlation matrices. *arXiv: Methodology*.

Yuan, K.-H. (2005). Fit indices versus test statistics. *Multivariate Behavioral Research*, 40:115–148.

Zhang, J. and Boos, D. D. (1992). Bootstrap critical values for testing homogeneity of covariance matrices. *Journal of the American Statistical Association*, 87(418):425–429.

Zhang, J. and Boos, D. D. (1993). Testing hypotheses about covariance matrices using bootstrap methods. *Communications in Statistics - Theory and Methods*, 22(3):723–739.

Zhong, P.-S., Lan, W., Song, P. X. K., and Tsai, C.-L. (2017). Tests for covariance structures with high-dimensional repeated measurements. *Ann. Statist.*, 45(3):1185–1213.

Zhu, L.-X., Ng, K., and Jing, P. (2002). Resampling methods for homogeneity tests of covariance matrices. *Statistica Sinica*, 12:769–783.

Eidestattliche Versicherung

Ich, Paavo Aljoscha Nanosch Sattler, erkläre, dass ich die vorliegende Dissertation mit dem Titel

„Inference for multivariate and high-dimensional data
in heterogeneous designs“

selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Alle wörtlich oder inhaltlich übernommenen Stellen habe ich als solche gekennzeichnet. Ich versichere außerdem, dass ich die beigefügte Dissertation nur in diesem und keinem anderen Promotionsverfahren eingereicht habe und, dass diesem Promotionsverfahren keine endgültig gescheiterten Promotionsverfahren vorausgegangen sind. Ferner erkäre ich, dass keine Aberkennung eines bereits erworbenen Doktorgrades vorliegt. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

Paavo Sattler

Dortmund den ....

Paavo Aljoscha Nanosch Sattler