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

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

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

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 $j = 1, \ldots, n_i$ denotes the individual subjects or units in group $i = 1, \ldots, a$, $a, n_i \in \mathbb{N}$, 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

$$\mathcal{H}_0(\boldsymbol{H}) : \boldsymbol{H}\boldsymbol{\mu} = \boldsymbol{0} \quad \boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \ldots, \boldsymbol{\mu}_a^\top)^\top, \qquad (2)$$

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:
$\mathcal{H}_0^{ab} : \left(\boldsymbol{P}_a \otimes \boldsymbol{P}_d \right) \boldsymbol{\mu} = \boldsymbol{0}$.

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 $n_{\max} = \max(n_1, \ldots, n_a) \to \infty$ 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 $\overline{\boldsymbol{X}} = (\overline{\boldsymbol{X}}_1^\top, \ldots \overline{\boldsymbol{X}}_a^\top)^\top$ with $\overline{\boldsymbol{X}}_i = n_i^{-1} \sum_{j=1}^{n_i} \boldsymbol{X}_{i,j}, i = 1, \ldots, a$, 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

$$\mathbb{E}(Q_N) \quad = \operatorname{tr}(\boldsymbol{T}_S \boldsymbol{\Sigma}) \cdot \sum_{i=1}^a \frac{N}{n_i}(\boldsymbol{T}_W)_{ii}$$

$$Var(Q_N) \quad = 2 \cdot \operatorname{tr}((\boldsymbol{T}_S \boldsymbol{\Sigma})^2) \cdot \sum_{i=1}^a \sum_{r=1}^a \frac{N^2}{n_i n_r}(\boldsymbol{T}_W)_{ir}^2.$$

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 $s = 1, \ldots, ad$. 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* $\sum_{s=1}^{\infty} b_s \left(C_s - 1\right)/\sqrt{2}$, *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* $C_i \stackrel{i.i.d.}{\sim} \chi_1^2$.

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 $r \in \{0, 1\}$. 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 $\widehat{\theta}_{n,d}/\theta \stackrel{\mathcal{P}}{\to} 1$. To get such estimators, we define

$$A_1 = \tfrac{1}{\sum_{i=1}^{a}(n_i-1)n_i} \sum_{i=1}^{a} \sum_{\ell_1 < \ell_2 = 1}^{n_i} (\boldsymbol{X}_{i,\ell_1} - \boldsymbol{X}_{i,\ell_2})^{\top} \boldsymbol{T}_S (\boldsymbol{X}_{i,\ell_1} - \boldsymbol{X}_{i,\ell_2}$$

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 $\text{tr}(\boldsymbol{T}_S \boldsymbol{\Sigma})$ and $\text{tr}\left((\boldsymbol{T}_S \boldsymbol{\Sigma})^2\right)$, 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

$$C_{i,1} := \frac{1}{8} \sum_{\ell_1 \neq ... \neq \ell_6 = 1}^{n_i} \boldsymbol{Y}_{i,\ell_1,\ell_2}^{\top} \boldsymbol{Y}_{i,\ell_3,\ell_4} \boldsymbol{Y}_{i,\ell_3,\ell_4}^{\top} \boldsymbol{Y}_{i,\ell_5,\ell_6} \boldsymbol{Y}_{i,\ell_5,\ell_6}^{\top} \boldsymbol{Y}_{i,\ell_1,\ell_2},$$

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 $v \in (0, \infty)$ is chosen and used to define $w_i = \lceil v \cdot \binom{n_i}{6} \rceil$, $i = 1, ..., a$ as the number of subsampling repetitions done for the i-th group. It is clear that the choice of v 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

$$\Lambda_1(\ell_1, \ell_2, \ell_3, \ell_4, \ell_5, \ell_6) = \boldsymbol{Y}_{i,\ell_1,\ell_2}^{\top} \boldsymbol{Y}_{i,\ell_3,\ell_4},$$

$$\Lambda_2(\ell_1, \ell_2, \ell_3, \ell_4, \ell_5, \ell_6) = \boldsymbol{Y}_{i,\ell_3,\ell_4}^{\top} \boldsymbol{Y}_{i,\ell_5,\ell_6},$$

$$\Lambda_3(\ell_1, \ell_2, \ell_3, \ell_4, \ell_5, \ell_6) = \boldsymbol{Y}_{i,\ell_5,\ell_6}^{\top} \boldsymbol{Y}_{i,\ell_1,\ell_2}.$$

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 $\mathrm{tr}\left((\boldsymbol{T}_S \boldsymbol{\Sigma})^3\right)$.

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 $\text{tr}((\boldsymbol{T}_S\boldsymbol{\Sigma}_1)^j) = \text{tr}((\boldsymbol{T}_S\boldsymbol{\Sigma}_2)^j) = ... = \text{tr}((\boldsymbol{T}_S\boldsymbol{\Sigma}_a)^j)$ 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 consequences 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 $\text{tr}((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^j)/\text{tr}(\boldsymbol{T}_S\boldsymbol{\Sigma})^j)$ 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 $\psi_z = \mathbb{1}(W_N > z_{1-\alpha})$, $\psi_\chi = \mathbb{1}(W_N > \chi^2_{1;1-\alpha})$ and $\varphi_N^\star = \mathbb{1}\{W_N > K_{\hat{f}_P;1-\alpha}\}$ are used while $\chi^2_{1;1-\alpha}$ denotes the $1 - \alpha$ quantile of a χ_1^2 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,
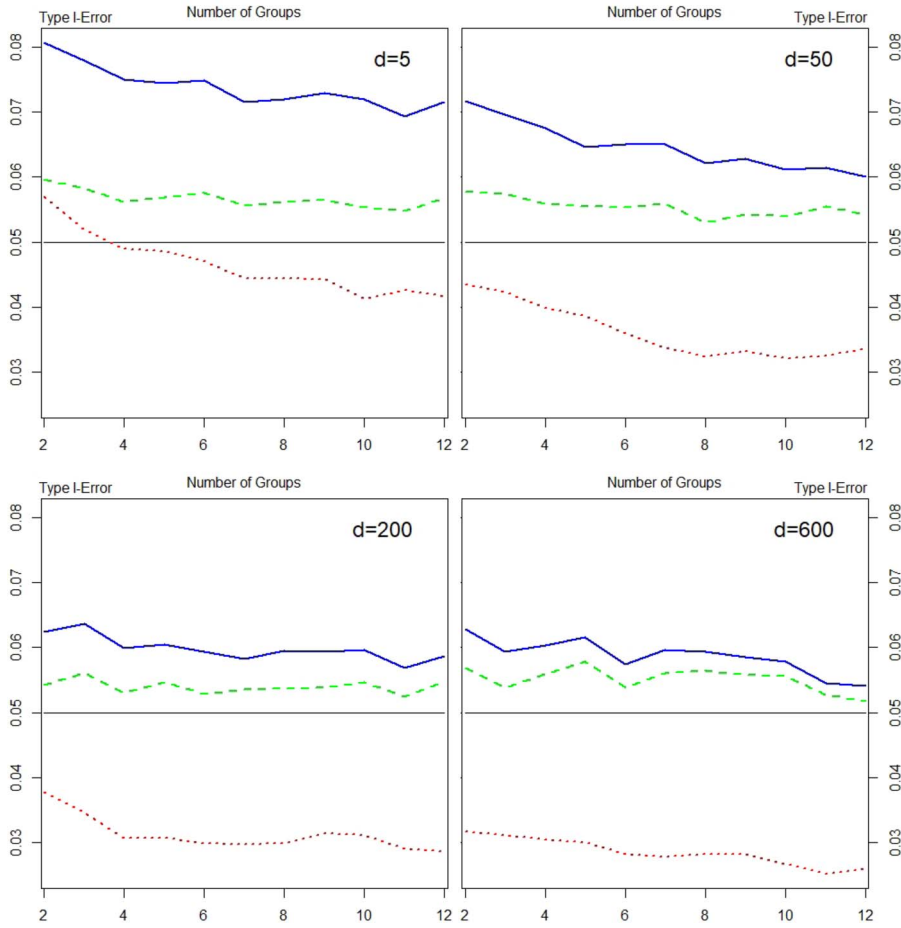
Fig 1: Simulated Type-I error rates ($\alpha = 5\%$) for $\psi_z(\text{—})$, ψ_χ ($\cdots$) and $\varphi_N^\star(\text{- -})$ under the null hypothesis $H_0^a : (\boldsymbol{P}_a \otimes \boldsymbol{P}_d)\,\boldsymbol{\mu} = \boldsymbol{0}$ for increasing dimension.

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
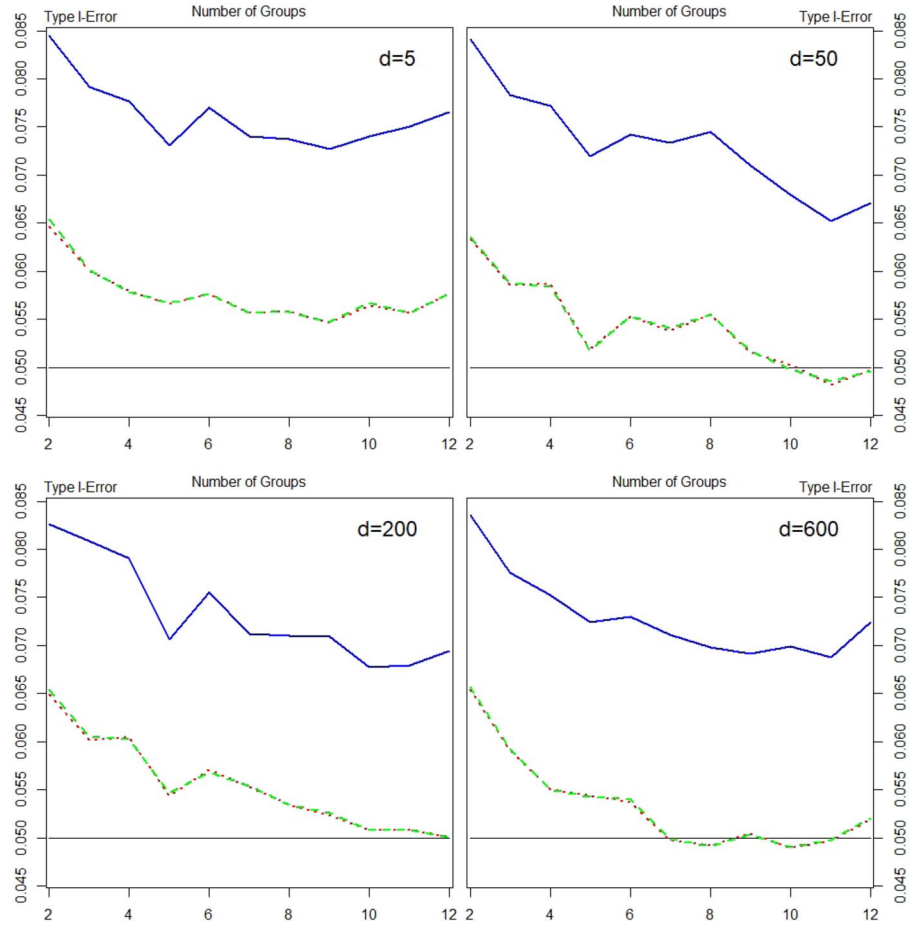
Fig 2: Simulated Type-I error rates ($\alpha = 5\%$) for $\psi_z(\text{---})$, ψ_χ $(\cdots)$ and $\varphi_N^\star(\text{- -})$ under the null hypothesis $H_0^b : \left(\frac{1}{a}\boldsymbol{J}_a \otimes \frac{1}{d}\boldsymbol{J}_d\right)\boldsymbol{\mu} = \boldsymbol{0}$ for increasing dimension.

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 $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_3 = ..., \boldsymbol{\mu}_9 = \boldsymbol{0}_d$ and $(\boldsymbol{\mu}_2)_k = (\boldsymbol{\mu}_4)_k, ...,$ $(\boldsymbol{\mu}_{10})_k = \delta \cdot k/d$, $k = 1, ..., d$, $\delta \in [0, 2]$,
- a *one-point-alternative* with $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_3 = ..., \boldsymbol{\mu}_9 = \boldsymbol{0}_d$ and $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_4, ...,$ $\boldsymbol{\mu}_{10} = \delta \cdot \boldsymbol{e}_1$, $\delta \in [0, 3.5]$ and
- a *shift-alternative* with $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_3 = ..., \boldsymbol{\mu}_9 = \boldsymbol{0}$ and $(\boldsymbol{\mu}_2) = (\boldsymbol{\mu}_4), ..., (\boldsymbol{\mu}_{10}) = \delta \cdot \boldsymbol{1}_d$ for $\mathcal{H}_0^b$, $\delta \in [0, 2]$

Here $\boldsymbol{e}_\ell$ denotes the vector containing 1 in the $\ell - th$ component, and 0 elsewhere and $\boldsymbol{1}_d$ contains just 1's in each component.
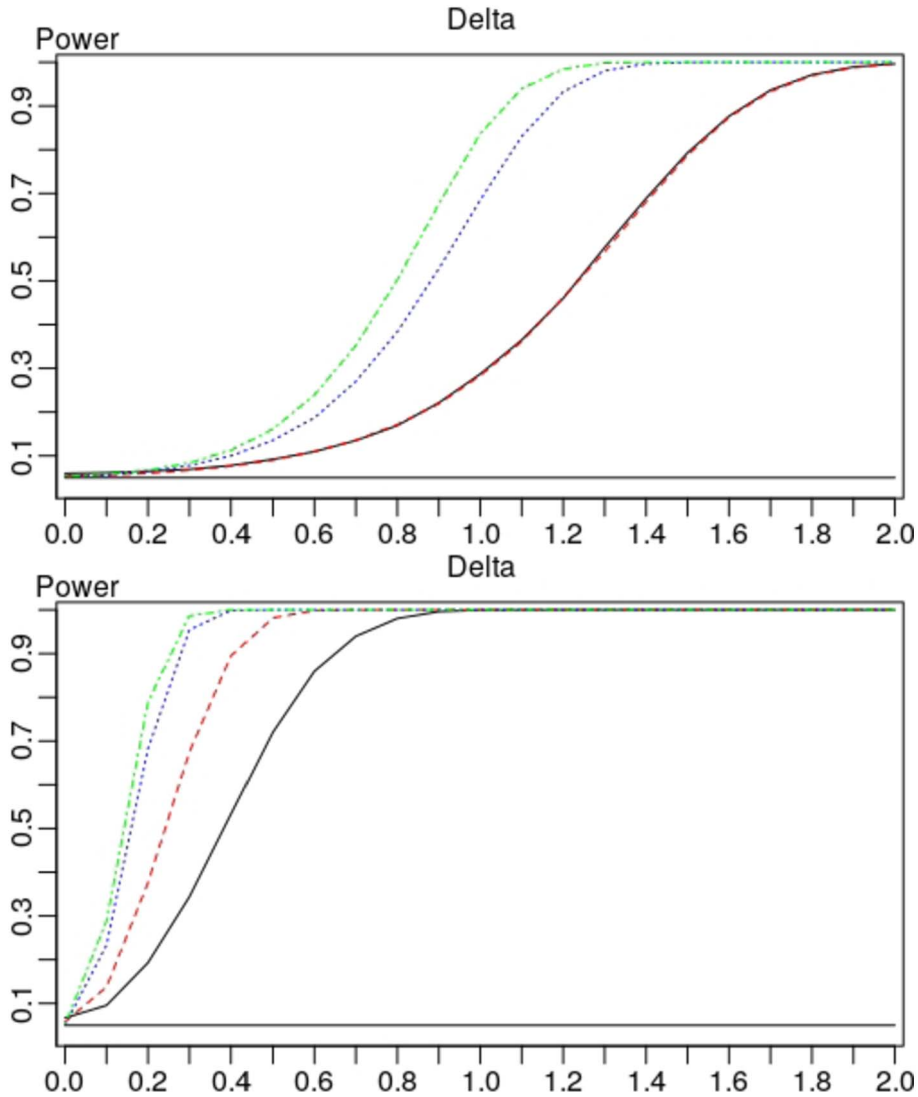
Fig 3: Simulated power curves of $\varphi_N^\star$ for a trend alternative with $d = 50$, 10000 simulation runs and an autoregressive structure($(\boldsymbol{\Sigma})_{i,j} = 0.6^{|i-j|}$). The sample size is $\boldsymbol{n} = (15, 15, 20, 35, 25, 20, 30, 30, 35, 20)$ and different numbers of groups were considered, namely $a = 2(—)$, $a = 4(\text{- -})$, $a = 8(\cdots)$ and $a = 10(\cdot — \cdot—)$.

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
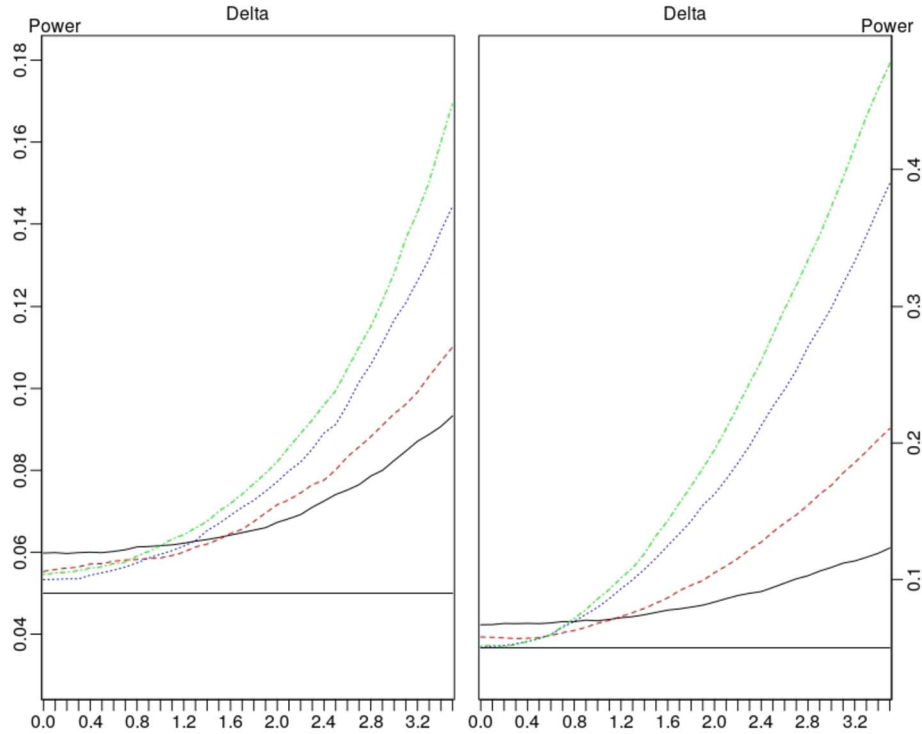
Fig 4: Simulated power curves of $\varphi_N^\star$ for a one-point alternative with $d = 50$, 10000 simulation runs and an autoregressive structure($(\boldsymbol{\Sigma})_{i,j} = 0.6^{|i-j|}$). The sample size is $\boldsymbol{n} = (15, 15, 20, 35, 25, 20, 30, 30, 35, 20)$and different numbers of groups were considered, namely $a = 2(—)$, $a = 4(\text{- -})$, $a = 8(\cdots)$ 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.
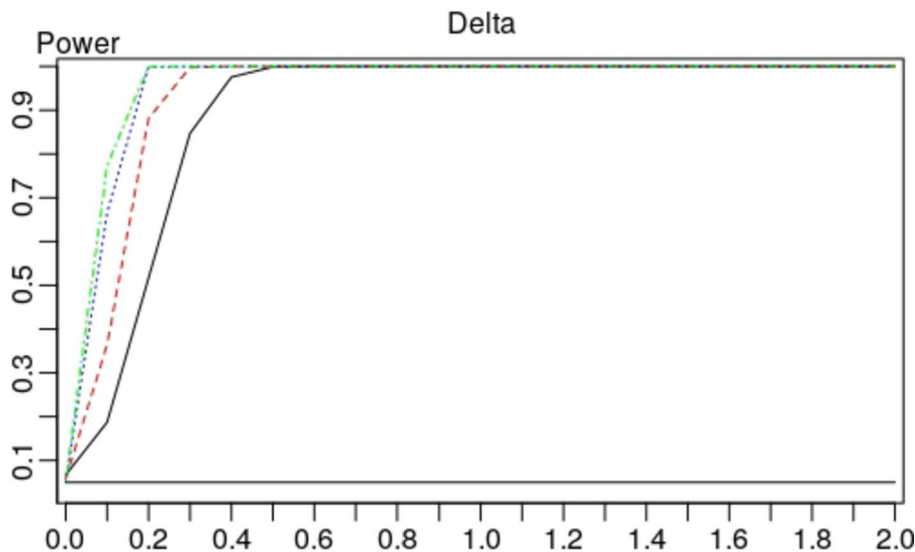
Fig 5: Simulated power curves of $\varphi_N^\star$ for a shift alternative with $d = 50$, 10000 simulation runs and an autoregressive structure$((\boldsymbol{\Sigma})_{i,j} = 0.6^{|i-j|})$. The sample size is $\boldsymbol{n} = (15, 15, 20, 35, 25, 20, 30, 30, 35, 20)$and different numbers of groups were considered, namely $a = 2(—)$, $a = 4(\text{- -})$, $a = 8(\cdots)$ and $a = 10(\cdot - \cdot-)$.

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 $s = 1, \ldots, ad$. 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

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

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

$$\beta_1 \to 1 \quad \text{as } N \to \infty,$$

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

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

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

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

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 $t \in \mathbb{R}$

$$\prod_{\ell=1}^{r'} \left(1 - \frac{2b'_\ell t}{\sqrt{2}} \right)^{-1/2} \exp\left(-t \frac{b'_\ell}{\sqrt{2}} \right) = \prod_{\ell=1}^{r} \left(1 - \frac{2b_\ell t}{\sqrt{2}} \right)^{-1/2} \exp\left(-t \frac{b_\ell}{\sqrt{2}} \right).$$

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

$$\left(\prod_{\ell=1}^{r'} \left(1 - \frac{2b'_\ell t}{\sqrt{2}} \right)^{-1/2} \exp\left(-\frac{b'_\ell t}{\sqrt{2}} \right) \right)^{-2} = \left(\prod_{\ell=1}^{r} \left(1 - \frac{2b_\ell t}{\sqrt{2}} \right)^{-1/2} \exp\left(-\frac{b_\ell t}{\sqrt{2}} \right) \right)^{-2}$$

$$\Leftrightarrow \prod_{\ell=1}^{r'} \left(1 - \sqrt{2}b'_\ell t \right) \exp(-\sqrt{2}b'_\ell t) = \prod_{\ell=1}^{r} \left(1 - \sqrt{2}b_\ell t \right) \cdot \exp(-\sqrt{2}b_\ell t).$$

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. $\square$

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 $\mathrm{tr}(\boldsymbol{T}_S \boldsymbol{\Sigma})$, *in all of our frameworks.*

Proof. It is obvious that this is a unbiased estimator of $\mathrm{tr}(\boldsymbol{T}_S \boldsymbol{\Sigma})$. With well known rules and analogous to [18] we calculate

$$Var(A_1) \leq \frac{1}{\left[\sum_{i=1}^{a} \binom{n_i}{2} \right]^2} \sum_{i=1}^{a} \binom{n_i}{2} \left(\binom{n_i}{2} - \binom{n_i - 2}{2} \right) \cdot \mathcal{O}(\mathrm{tr}^2(\boldsymbol{T}_S \boldsymbol{\Sigma})).$$

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

$$
\begin{aligned}
Var(A_1) \ &\leq \ \frac{1}{\left[\sum_{i=1}^{a} \binom{n_i}{2}\right] \cdot \binom{n_{\max}}{2}} \sum_{i=1}^{a} \binom{n_i}{2} \left(\binom{n_i}{2} - \binom{n_i-2}{2}\right) \cdot \mathcal{O}(\operatorname{tr}^2(\boldsymbol{T}_S \boldsymbol{\Sigma})) \\
&\leq \ \frac{1}{\left[\sum_{i=1}^{a} \binom{n_i}{2}\right] \cdot \binom{n_{\max}}{2}} \sum_{i=1}^{a} \binom{n_i}{2} \left(\binom{n_{\max}}{2} - \binom{n_{\max}-2}{2}\right) \cdot \mathcal{O}(\operatorname{tr}^2(\boldsymbol{T}_S \boldsymbol{\Sigma})) \\
&= \ \frac{\left(\binom{n_{\max}}{2} - \binom{n_{\max}-2}{2}\right)}{\binom{n_{\max}}{2}} \cdot \mathcal{O}(\operatorname{tr}^2(\boldsymbol{T}_S \boldsymbol{\Sigma})) \\
&= \ \mathcal{O}\left(n_{\max}^{-1}\right) \cdot \mathcal{O}(\operatorname{tr}^2(\boldsymbol{T}_S \boldsymbol{\Sigma})).
\end{aligned}
$$

For the other case $n_{\max}$ is bound and $a \to \infty$. In this situation it holds

$$
\begin{aligned}
Var(A_1) \ &\leq \ \frac{1}{\left[\sum_{i=1}^{a} \binom{n_i}{2}\right] \cdot a \cdot \binom{n_{\min}}{2}} \sum_{i=1}^{a} \binom{n_i}{2} \left(\binom{n_{\max}}{2} - \binom{n_{\max}-2}{2}\right) \cdot \mathcal{O}(\operatorname{tr}^2(\boldsymbol{T}_S \boldsymbol{\Sigma})) \\
&= \ \frac{\left(\binom{n_{\max}}{2} - \binom{n_{\max}-2}{2}\right)}{a \cdot \binom{n_{\min}}{2}} \cdot \mathcal{O}(\operatorname{tr}^2(\boldsymbol{T}_S \boldsymbol{\Sigma})) \\
&= \ \mathcal{O}\left(a^{-1}\right) \cdot \mathcal{O}(\operatorname{tr}^2(\boldsymbol{T}_S \boldsymbol{\Sigma}))
\end{aligned}
$$

So dividing by $\operatorname{tr}^2(\boldsymbol{T}_S \boldsymbol{\Sigma})$ and then using the Tschebyscheff inequality leads to the results in both cases. $\qquad\square$

For the estimated version of the standardized quadratic form, one more estimator is needed.

Lemma 2. *The estimator, given by*

$$
A_2 = \sum_{i=1}^{a} \sum_{\substack{\ell_1,\ell_2=1 \\ \ell_1 > \ell_2}}^{n_i} \sum_{\substack{k_2=1 \\ k_2 \neq \ell_1 \neq \ell_2}}^{n_i} \sum_{\substack{k_1=1 \\ \ell_2 \neq \ell_1 \neq k_1 > k_2}}^{n_i} \frac{\left[(\boldsymbol{X}_{i,\ell_1} - \boldsymbol{X}_{i,\ell_2})^{\top} \boldsymbol{T}_S (\boldsymbol{X}_{i,k_1} - \boldsymbol{X}_{i,k_2})\right]^2}{4 \cdot 6 \sum_{i=1}^{a} \binom{n_i}{4}},
$$

is an unbiased and ratio-consistent estimator of $\operatorname{tr}\left((\boldsymbol{T}_S \boldsymbol{\Sigma})^2\right)$ *in all our asymptotic frameworks.*

Proof. Again the unbiasedness is clear, and we consider the variance.
We calculate, with $\boldsymbol{Y}_{i,\ell_1,\ell_2} := \boldsymbol{T}_S (\boldsymbol{X}_{i,k_1} - \boldsymbol{X}_{i,k_2})$,

$$
\begin{aligned}
&Var(A_2) \\
&= \ \left[24 \sum_{i=1}^{a} \binom{n_i}{4}\right]^{-2} \sum_{i=1}^{a} Var\left(\sum_{\substack{\ell_1,\ell_2=1 \\ \ell_1 > \ell_2}}^{n_i} \sum_{\substack{k_2=1 \\ k_2 \neq \ell_1 \neq \ell_2}}^{n_i} \sum_{\substack{k_1=1 \\ \ell_2 \neq \ell_1 \neq k_1 > k_2}}^{n_i} \left[\boldsymbol{Y}_{i,\ell_1,\ell_2}^{\top} \boldsymbol{Y}_{i,k_1,k_2}\right]^2\right) \\
&\leq \ \frac{\sum_{i=1}^{a} \binom{n_i}{4}\left(\binom{n_i}{4} - \binom{n_i-4}{4}\right)}{\left[4 \cdot \sum_{i=1}^{a} \binom{n_i}{4}\right]^2} \mathcal{O}\left(\operatorname{tr}^2\left((\boldsymbol{T}_S \boldsymbol{\Sigma}_i)^2\right)\right).
\end{aligned}
$$

Similar as before for $n_{\max} \to \infty$ we get

$$
Var(A_2) \leq \mathcal{O}\left(n_{\max}^{-1}\right) \cdot \mathcal{O}\left(\operatorname{tr}^2\left((\boldsymbol{T}_S \boldsymbol{\Sigma}_i)^2\right)\right)
$$

and for $n_{\max}$ bound and $a \to \infty$

$$Var(A_2) \leq \mathcal{O}\left(a^{-1}\right) \cdot \mathcal{O}\left(\mathrm{tr}^2\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)^2\right)\right).$$

Again the result follows by using Tschebyscheff's inequality. $\qquad\square$

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]. $\quad\square$

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) $\mathbb{E}(C_1) = \mathrm{tr}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^3\right)$,

b) $Var(C_1) = \dfrac{\sum_{j=1}^a \binom{n_j}{6}\left(\binom{n_j}{6}-\binom{n_j-6}{6}\right)}{\left(\sum_{i=1}^a \binom{n_i}{6}\right)^2} \cdot \mathcal{O}\left(\mathrm{tr}^3\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)\right)\right)$,

c) $\dfrac{C_1}{\mathrm{tr}^{3/2}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^2\right)} - \dfrac{\mathrm{tr}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^3\right)}{\mathrm{tr}^{3/2}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^2\right)} \xrightarrow{\mathcal{P}} 0$ in our frameworks I-V,

d) $\dfrac{C_1^2}{A_2^4} - (f_P)^{-1} \xrightarrow{\mathcal{P}} 0$ in our frameworks I-V.

The results from [18] directly yield to

$$\mathbb{E}(C_1) = \mathrm{tr}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^3\right)$$

and

$$Var(C_1) = \sum_{i=1}^a \frac{Var(C_{i,1})}{6! \cdot \sum_{j=1}^a \binom{n_j}{6}} \leq \frac{\sum_{j=1}^a \binom{n_j}{6}\left(\binom{n_j}{6}-\binom{n_j-6}{6}\right)}{\left(\sum_{i=1}^a \binom{n_i}{6}\right)^2} \cdot \mathcal{O}\left(\mathrm{tr}^3\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)\right)\right)$$

which proves a) and b). Together with Tschebychefs inequality this leads to an unbiased ratio consistent estimator for $\mathrm{tr}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^3\right)$.

For part c) we calculate

$$\mathbb{E}\left(\frac{C_1}{\mathrm{tr}^{3/2}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^2\right)} - \frac{\mathrm{tr}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^3\right)}{\mathrm{tr}^{3/2}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^2\right)}\right) = 0$$

and

$$Var\left(\frac{C_1}{\mathrm{tr}^{3/2}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^2\right)} - \frac{\mathrm{tr}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^3\right)}{\mathrm{tr}^{3/2}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^2\right)}\right)$$

$$= \frac{Var\left(C_1\right)}{\mathrm{tr}^3\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^2\right)} \leq 27 \cdot \frac{\sum_{j=1}^a \binom{n_j}{6}\left(\binom{n_j}{6}-\binom{n_j-6}{6}\right)}{\left(\sum_{i=1}^a \binom{n_i}{6}\right)^2} \cdot \mathcal{O}(1)$$

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

$$
\frac{C_1^2}{\mathrm{tr}^3\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^2\right)} - \frac{1}{f_P}
$$

$$
= \left(\frac{C_1}{\mathrm{tr}^{3/2}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^2\right)}\right)^2 - \left(\frac{\mathrm{tr}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^3\right)}{\mathrm{tr}^{3/2}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^2\right)}\right)^2
$$

$$
= \left[\frac{C_1}{\mathrm{tr}^{3/2}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^2\right)} - \frac{\mathrm{tr}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^3\right)}{\mathrm{tr}^{3/2}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^2\right)}\right]\left[\frac{C_1}{\mathrm{tr}^{3/2}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^2\right)} + \frac{\mathrm{tr}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^3\right)}{\mathrm{tr}^{3/2}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^2\right)}\right]
$$

$$
= \mathcal{o}_P(1) \cdot \left[\frac{C_1}{\mathrm{tr}^{3/2}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^2\right)} - \frac{\mathrm{tr}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^3\right)}{\mathrm{tr}^{3/2}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^2\right)} + 2\frac{\mathrm{tr}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^3\right)}{\mathrm{tr}^{3/2}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^2\right)}\right] = \mathcal{o}_P(1),
$$

were for the last step the trace inquality was used together with Slutzky's theorem. With the ratio-consistency of A_2 it follows $A_2/\mathrm{tr}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)\right) \overset{\mathcal{P}}{\to} 1$ and because of continous mapping $\mathrm{tr}^3\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)\right)/A_2^3 \overset{\mathcal{P}}{\to} 1$. This leads to

$$
\frac{C_1^2}{A_2^3} - (f_P)^{-1} = \frac{\mathrm{tr}^3\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^2\right)}{A_2^3}\frac{C_1^2}{\mathrm{tr}^3\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}\right)^2\right)} - (f_P)^{-1}
$$

$$
= (1 + \mathcal{o}_P(1)) \cdot \frac{1}{\widehat{f_P}} - \frac{1}{f_P}
$$

$$
= \frac{1}{\widehat{f_P}} - \frac{1}{f_P} + \mathcal{o}_P(1) \cdot \frac{1}{\widehat{f_P}} = \mathcal{o}_P(1). \qquad \square
$$

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 expectation value of the estimator, using the notation $w := \sum_{i=1}^{a} w_i$:

$$
\mathbb{E}\left(C_1^\star\right) = \mathbb{E}\left(\frac{1}{8 \cdot \sum_{i=1}^{a} w_i}\sum_{i=1}^{a} C_{i,1}^\star\right)
$$

$$
= \frac{1}{8 \cdot w}\sum_{i=1}^{a}\mathbb{E}\left(C_{i,1}^\star\right)
$$

$$= \frac{1}{8 \cdot w} \sum_{i=1}^{a} \mathbb{E} \left(\sum_{b=1}^{w_i} \Lambda_1(\boldsymbol{\sigma}_i(b)) \cdot \Lambda_2(\boldsymbol{\sigma}_i(b)) \cdot \Lambda_3(\boldsymbol{\sigma}_i(b)) \right)$$

$$= \frac{1}{8 \cdot w} \sum_{i=1}^{a} \sum_{b=1}^{w_i} \mathbb{E} \left(\Lambda_1(\boldsymbol{\sigma}_i(b)) \cdot \Lambda_2(\boldsymbol{\sigma}_i(b)) \cdot \Lambda_3(\boldsymbol{\sigma}_i(b)) \right)$$

$$= \frac{1}{8 \cdot w} \sum_{i=1}^{a} w_i \cdot \mathbb{E} \left(\Lambda_1(1,2,3,4,5,6) \cdot \Lambda_2(1,2,3,4,5,6) \cdot \Lambda_3(1,2,3,4,5,6) \right)$$

$$= \frac{1}{8 \cdot w} \sum_{i=1}^{a} w_i \cdot 8 \operatorname{tr} \left((\boldsymbol{T}_S \boldsymbol{\Sigma})^3 \right)$$

$$= \operatorname{tr} \left((\boldsymbol{T}_S \boldsymbol{\Sigma})^3 \right).$$

With Theorem A.9 Theorem A.10 and Theorem A.16 from [18] for the variance we get

$$
\begin{aligned}
Var\left(C_1^{\star}\right) \quad &= \frac{1}{(8 \cdot w)^2} \sum_{i=1}^{a} Var\left(C_{i,1}^{\star}\right) \\
&\leq \frac{1}{(8 \cdot w)^2} \sum_{i=1}^{a} w_i^2 \cdot \left[0 + 1 - \left(1 - \frac{1}{w_i}\right) \cdot \frac{\binom{n_i-6}{6}}{\binom{n_i}{6}} \right].
\end{aligned}
$$

Again there the same two cases. If $n_{\max}$ is bound and therefore $\max_{i=1,\dots,a}(w_i)$ is bound, it follows $a \to \infty$ and hereby

$$
\begin{aligned}
&\frac{1}{(8 \cdot w)^2} \sum_{i=1}^{a} w_i^2 \cdot \left[0 + 1 - \left(1 - \frac{1}{w_i}\right) \cdot \frac{\binom{n_i-6}{6}}{\binom{n_i}{6}} \right] \\
\leq \quad &\frac{1}{(8 \cdot w) \cdot a \cdot \min_{i=1,\dots,a}(w_i)} \cdot \max_{i=1,\dots,a}(w_i) \sum_{i=1}^{a} w_i \cdot 1 \\
= \quad &\mathcal{O}\left(a^{-1}\right) \cdot \frac{\max_{i=1,\dots,a}(w_i)}{\min_{i=1,\dots,a}(w_i)} \\
= \quad &\mathcal{O}\left(a^{-1}\right)
\end{aligned}
$$

while for $n_{\max} \to \infty$ which implies $\max_{i=1,\dots,a}(w_i) \to \infty$ we calculate first

$$
\begin{aligned}
&w_i^2 \cdot \left[0 + 1 - \left(1 - \frac{1}{w_i}\right) \cdot \frac{\binom{n_i-6}{6}}{\binom{n_i}{6}} \right] \\
= \quad &w_i \cdot \left[w_i \cdot \left(1 - \frac{\binom{n_i-6}{6}}{\binom{n_i}{6}} \right) + \frac{\binom{n_i-6}{6}}{\binom{n_i}{6}} \right] \\
\leq \quad &w_i \cdot \left[\left(v \cdot \binom{n_i}{6} + 1 \right) \left(1 - \frac{\binom{n_i-6}{6}}{\binom{n_i}{6}} \right) + \frac{\binom{n_i-6}{6}}{\binom{n_i}{6}} \right] \\
= \quad &w_i \cdot \left[v \left(\binom{n_i}{6} - \binom{n_i-6}{6} \right) + 1 \right]
\end{aligned}
$$

$$\leq \quad w_i \cdot \left[v \left(\binom{n_{\min}}{6} - \binom{n_{\min} - 6}{6} \right) + 1 \right]$$

and therefore

$$\leq \quad \frac{1}{(8 \cdot w)^2} \sum_{i=1}^{a} w_i^2 \cdot \left[0 + 1 - \left(1 - \frac{1}{w_i} \right) \cdot \frac{\binom{n_i - 6}{6}}{\binom{n_i}{6}} \right]$$

$$\leq \quad \frac{1}{(8 \cdot w)^2} \sum_{i=1}^{a} w_i \cdot \left[v \left(\binom{n_{\min}}{6} - \binom{n_{\min} - 6}{6} \right) + 1 \right]$$

$$\leq \quad \frac{1}{(64 \cdot w) \cdot \max\limits_{i=1,\ldots,a} (w_i)} \sum_{i=1}^{a} w_i \cdot \left[v \left(\binom{n_{\min}}{6} - \binom{n_{\min} - 6}{6} \right) + 1 \right]$$

$$\leq \quad \frac{1}{(64 \cdot w) \cdot \left(v \cdot \binom{n_{\max}}{6} - 1 \right)} \sum_{i=1}^{a} w_i \cdot \left[v \left(\binom{n_{\min}}{6} - \binom{n_{\min} - 6}{6} \right) + 1 \right]$$

$$= \quad \frac{\left[v \left(\binom{n_{\min}}{6} - \binom{n_{\min} - 6}{6} \right) + 1 \right]}{64 \cdot \left(v \cdot \binom{n_{\max}}{6} - 1 \right)} = \mathcal{O} \left(n_{\max}^{-1} \right).$$

Combining these results, the remainder of the proof follows analogously to the proof of Theorem 4. □

Acknowledgments

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