Manifold Asymptotics of Quadratic-Form-Based Inference in Repeated Measures Designs

Paavo Sattler*

March 11, 2020

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 $\infty$ 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. Studying the limit distributions in detail, we follow Sattler and Pauly (2018) and propose an approximation to obtain critical values. The resulting test together with their approximation approach is investigated in an extensive simulation study with a focus on the exceptional asymptotic frameworks which are the main focus of this work.

Keywords: Multivariate Data, equal covariance matrices, high-dimensions, repeated measures.

* Dortmund University, Faculty of Statistics, Germany
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 $X_{ik} \sim N_d(\mu_i, \Sigma_i)$ null hypotheses regarding $\mu = (\mu_1, ..., \mu_a)^\top$ are investigated, where popular hypotheses are the existence of a group effect, a time effect as well as a interaction effect between time and group. For a classical repeated measure ANOVA design with $d \leq n_i$, this was treated for example in Brunner et al. (2012). 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 special case with just two groups but with a general distributional setting and without restriction on the dimension $d$ this was treated in Chen and Qin (2010). For more groups and a more general setting regarding hypotheses, Happ et al. (2016) uses a classical ANOVA F test statistic, which has just an exact $F$-distribution for very special covariances 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 Harrar and Kong (2016) 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, Pauly et al. (2015) investigated the case with just one normal distributed group, but fewer assumptions on the covariance matrix and no necessary relation between sample size and dimension.

Sattler and Pauly (2018) 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. Hereby, 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 Bathke (2002) or Bathke and Lankowski (2005) for fixed dimensions $d$ and balanced designs $n_i \equiv n$.

Therefore in addition to the large $a$ small $n$ case, we include the large $d$ small $n$...
case, and further the combination of both, and develop a technique that can be used in each of these settings. To this end, we follow the same approach as Kong and Harrar (2019) and assume homogenous covariance matrices with \( \Sigma_i = \Sigma > 0 \), again with no further assumptions on the structure of the covariance matrix \( \Sigma \). 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 remaining paper. 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 power of the tests, 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

\[
X_{ij} = (X_{ij,1}, \ldots, X_{ij,d})^\top \overset{\text{ind}}{\sim} N_d(\mu_i, \Sigma) \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, a,
\]

whereby each vector represents the measurement of one independent subject. It is assumed that mean vectors

\[
E(X_{i,1}) = \mu_i = (\mu_{i,t})_{t=1}^d \in \mathbb{R}^d
\]

and one positive definite covariance matrix \( \text{Cov}(X_{i,1}) = \Sigma > 0 \) 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 Konietschke et al. (2015) for example.

Within this model linear hypotheses of repeated measures ANOVA, formulated as

\[
\mathcal{H}_0(H): H\mu = 0 \quad \mu = (\mu_1^\top, \ldots, \mu_a^\top)^\top,
\]

are investigated. Here, \( H = H_W \otimes H_S \) denotes a proper hypothesis matrix, where \( H_W \) and \( H_S \) refer to whole-plot (group) and/or subplot (time) effects, while \( \otimes \) denotes the Kronecker product.

For theoretical considerations it is often more convenient to reformulate \( \mathcal{H}_0(H) \) through a corresponding projection matrix \( T = H^\top [HH^\top]^{-1} - H \), see e.g. Pauly et al.
Here $(\cdot)^{-}$ denotes some generalized inverse of the matrix and $\mathcal{H}_0(H)$ can equivalently be written as $\mathcal{H}_0(T) : T\mu = 0$. As discussed in Sattler and Pauly (2018), $T$ has the form $T = T_W \otimes T_S$ for projection matrices $T_W$ and $T_S$. Now hypotheses of interest are for example given by

(a) No group effect:
$$\mathcal{H}_0^a : (P_a \otimes \frac{1}{d}J_d) \mu = 0,$$

(b) No time effect:
$$\mathcal{H}_0^b : (\frac{1}{a}J_a \otimes P_d) \mu = 0,$$

(c) No interaction effect between time and group:
$$\mathcal{H}_0^{ab} : (P_a \otimes P_d) \mu = 0.$$

Here, $J_d$ is the d-dimensional matrix only containing 1s and $P_d := I_d - 1/d \cdot J_d$ is the centring 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 statistic and their asymptotic

In this work we consider the following 5 different asymptotic frameworks, which are:

$$\begin{align*}
a &\rightarrow \infty, \\
(a, d) &\rightarrow \infty, \\
(a, n_{\max}) &\rightarrow \infty, \\
(d, n_{\max}) &\rightarrow \infty, \\
(a, d, n_{\max}) &\rightarrow \infty.\
\end{align*}$$

(II) (III) (IV) (V)
This great diversity is exceptional and distinguishes from nearly all other approaches. Most of the existing procedures just consider special cases of one of these cases (for example Chen and Qin (2010) (IV) with $a = 1$ or Pauly et al. (2015) (IV) with $a = 2$). Other allow for only one as Happ et al. (2016) for (IV) or Bathke (2002) 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 $n_{\text{max}} = \max(n_1, \ldots, n_a) \to \infty$ with fixed $a$ and $d$ has already been studied in detail in the literature and is, thus excluded here, see, e.g., Friedrich et al. (2017) or Bathke et al. (2018) and the references cited therein.

**It is apparent that in contrast to Sattler and Pauly (2018) and other papers, the common conditions as $n_i \to \kappa_i \in (0, 1)$ are missing.** This is significant, 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 $\infty$.

To examine the validity of the null hypothesis $H_0(T) : T\mu = 0$ unattached from the asymptotic framework, we use $Q_N = N \cdot \bar{X}^\top T \bar{X}$. Here $\bar{X} = (\bar{X}_1^\top, \ldots, \bar{X}_a^\top)^\top$ with $\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}, i = 1, \ldots, a$, denotes the vector of pooled group means. Unfortunately for many covariance matrices $\Sigma$, this random variable tends to converge to infinity, for $d \to \infty$ or $a \to \infty$. To avoid this behaviour the standardized quadratic form is given by

\[ \tilde{W}_N = \frac{Q_N - \mathbb{E}_{H_0}(Q_N)}{\sqrt{\text{Var}_{H_0}(Q_N)}}, \]

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

\[ \mathbb{E}(Q_N) = \text{tr}(T_S \Sigma) \cdot \sum_{i=1}^{a} \frac{N}{n_i} (T_W)_{ii}, \]

\[ \text{Var}(Q_N) = 2 \cdot \text{tr}((T_S \Sigma)^2) \cdot \sum_{i=1}^{a} \sum_{r=1}^{a} \frac{N_i^2}{n_i n_r} (T_W)_{ir}^2. \]

Observe, that for both values only the first factor $\text{tr}(T_S \Sigma)$ resp. $\text{tr}((T_S \Sigma)^2)$ de-
depends on the unknown covariance matrix, while all other quantities are known from the test setting.

Applying the representation theorem for quadratic forms in normally distributed random vectors from Mathai and Provost (1992) we can rewrite the standardized statistic \( \widetilde{W}_N \) as

\[
\widetilde{W}_N = \frac{Q_N - \mathbb{E}_{H_0}(Q_N)}{\text{Var}_{H_0}(Q_N)^{1/2}} = \sum_{s=1}^{ad} \frac{\lambda_s}{\sqrt{\sum_{\ell=1}^{ad} \lambda_\ell^2}} \left( \frac{C_s - 1}{\sqrt{2}} \right).
\]

Here \( \lambda_s \) are the eigenvalues of \( TV_N^T \) in decreasing order, \( V_N = \bigoplus_{i=1}^{a} \mathbb{N}_i \Sigma \) and \( (C_s)_s \) is a sequence of independent \( \chi^2_1 \)-distributed random variables. As a consequence, the asymptotic behavior of the eigenvalues, determine the asymptotic limit distribution of \( \widetilde{W}_N \). In fact, we obtain in generalization of Pauly et al. (2015) and Sattler and Pauly (2018):

**Theorem 3.1:**

Let \( \beta_s = \lambda_s / \sqrt{\sum_{\ell=1}^{ad} \lambda_\ell^2} \) for \( s = 1, \ldots, ad \). Then \( \widetilde{W}_N \) has, under \( H_0(T) \), 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

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

for a decreasing sequence \( (b_s)_s \) in \( [0, 1] \) with \( r := \#(b_i \neq 0) \), while \( C_i \overset{i.i.d.}{\sim} \chi^2_1 \), \( Z \sim \mathcal{N}(0, 1) \).

b) a distribution of the form \( \sum_{s=1}^{\infty} b_s (C_s - 1) / \sqrt{2} \), if

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

for a decreasing sequence \( (b_s)_s \) in \( (0, 1) \) with \( \sum_{s=1}^{\infty} b_s^2 = 1 \) and \( C_i \overset{i.i.d.}{\sim} \chi^2_1 \).

**Putting the results into context.** Chen and Qin (2010) only considered case a) with \( r = 0 \). Sattler and Pauly (2018) 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 enhances the continuum of limit distributions considerably through a mixture of normal distribution and finite sums of weighted standardized \( \chi^2_1 \)-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 which are ratio consistent in all our settings. To this end, define

\[ A_1 = \frac{1}{\sum_{i=1}^{n_i} (n_i-1)} \sum_{i=1}^{n_i} \sum_{\ell_1 < \ell_2 = 1} \sum_{a_i} \sum \left( X_{i,\ell_1} - X_{i,\ell_2} \right)^\top T_S \left( X_{i,\ell_1} - X_{i,\ell_2} \right) \]

and

\[ A_2 = \sum_{i=1}^{n_i} \sum_{\ell_1 < \ell_2} \sum_{k_1 = 1}^{n_i} \sum_{k_1 > k_2} \sum_{\ell_1 \neq \ell_2 \neq k_1 \neq k_2} \frac{\left( X_{i,\ell_1} - X_{i,\ell_2} \right)^\top T_S \left( X_{i,k_1} - X_{i,k_2} \right)^2}{4 \cdot \sum_{i=1}^{n_i} \left( \sum_{a_i} \right)} \]

Below we prove that they are unbiased and ratio consistent estimators for \( \text{tr}(T_S \Sigma) \) and \( \text{tr}(T_S \Sigma^2) \), respectively, under both, the null hypothesis and the alternative. This allows us to define the estimated version of our test statistic by

\[ W_N = \frac{Q_N - A_1 \cdot \sum_{i=1}^{n_i} (T_W)_{ii}}{\sqrt{2 \cdot A_2 \cdot \sum_{i=1}^{n_i} \sum_{r=1}^{a_i} \sum_{a_i} \left( \frac{N}{N_i} (T_W)_{ir} \right)^2}} \]

The following Lemma justifies the usage of the estimated version instead of the exact one.

**Theorem 3.2:**
Under \( H_0(T) : T \mu = 0 \) and one of the frameworks (I)-(V) the statistic \( W_N \) has the same asymptotic limit distributions as \( \bar{W}_N \), if the respective conditions (a)-(b) from Theorem 3.1 are fulfilled.

Unfortunately the calculation of the standardized eigenvalues \( \beta_s \) is in 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 dimension or number of groups are necessary for a good approximation, which make quantiles based on Theorem 3.1(a) difficult to apply. For similar reasons, in \[ \text{Pauly et al. (2015)} \] and \([\text{Sattler and Pauly (2018)}]\) they used the quantiles of a random variable of the kind

\[ K_f = \frac{(\chi^2_f - f) / \sqrt{2f}}{f} \]

in case of \( \beta_1 \to \{0, 1\} \). The choice of \( f_P = \text{tr}^3 \left( (TV_N)^2 \right) / \text{tr}^2 \left( (TV_N)^3 \right) \) for the degrees of freedom lead to a third moment approximation. In our homoscedastic model the usage of this random variable \( K_f \) is based on the following theorem.
Theorem 3.3:
Under the conditions of Theorem 3.1 and one of the frameworks (I)-(V) the random variable $K_{fp}$ has, under $H_0 : T \mu = 0_{a \times d}$, asymptotically

a) a standard normal distribution if $\beta_1 \to 0$ as $N \to \infty$,

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

With the well known rules for the kronecker product and traces we can decompose this number by

$$f_P = \frac{\text{tr}^3 \left( (T_S \Sigma)^2 \right)}{\text{tr}^2 \left( (T_S \Sigma)^3 \right)} \cdot \frac{\text{tr}^3 \left( \text{diag} \left( \frac{N}{n_1}, \ldots, \frac{N}{n_a} \right) \cdot T_W \right)^2}{\text{tr}^2 \left( \text{diag} \left( \frac{N}{n_1}, \ldots, \frac{N}{n_a} \right) \cdot T_W \right)^3} =: \frac{\text{tr}^3 \left( (T_S \Sigma)^2 \right)}{\text{tr}^2 \left( (T_S \Sigma)^3 \right)} \cdot \eta_{N,a}.$$

The connection between $f_P$ and $\beta_1$ in the two extreme cases, i.e. $\beta_1 \to 0$ if and only if $f_P \to \infty$ and $\beta_1 \to 1$ if and only if $f_P \to 1$, have been investigated in Pauly et al. (2015) 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 $T_W$.

Moreover, for $\eta_{N,a} \to \infty$, we also have $f_P \to \infty$, without estimation, because $\text{tr}^3 \left( (T_S \Sigma)^2 \right) / \text{tr}^2 \left( (T_S \Sigma)^3 \right) \geq 1$. Otherwise, however, the behaviour of $f_P$ is unclear and we have to find consistent estimators for $\text{tr} \left( (T_S \Sigma)^3 \right)$ in all our different frameworks. This achieved by considering the class of estimators

$$C_{i,1} := \frac{1}{8} \sum_{\ell_1 \neq \ldots \neq \ell_6 = 1}^{n_i} Y_{i,\ell_1,\ell_2} Y_{i,\ell_1,\ell_3} Y_{i,\ell_1,\ell_4} Y_{i,\ell_5,\ell_6} Y_{i,\ell_5,\ell_6} Y_{i,\ell_1,\ell_2},$$

which are based on suitable symmetrized U-statistics, while $\ell_1 \neq \ell_2 \neq \ldots \neq \ell_6$ 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

$$C_1 := \frac{1}{6!} \sum_{i=1}^{a} \frac{1}{(n_i)!} \sum_{i=1}^{a} C_{i,1}.$$

Together with the estimators from above, we can construct a consistent estimator for $f_P$ by $\hat{f}_P := A_2^3 / C_1 \cdot \eta_{N,a}$. 

8
Theorem 3.4:
In all our frameworks (I)–(V), it holds that

i) $C_1$ is an unbiased estimator for $\text{tr} \left( (T_S \Sigma)^3 \right)$,

ii) $\left( \hat{f}_p \right)^{-1} - (f_p)^{-1} \xrightarrow{P} 0$,

where $P$ denotes convergence in probability.

Through the usage of U-statistics with a kernel of order 6, for each estimator $C_1$, $6! \cdot \binom{n_i}{6}$ summations have to be done. In contrast, estimators based on observations from all groups would require much higher numbers. For example in Sattler and Pauly (2018) $\prod_{i=1}^a 6! \cdot \binom{n_i}{6}$ summations are necessary. Due to homogeneity, we don’t need this kind of estimator, but $C_{1,i}$ 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 Sattler and Pauly (2018), the usage of subsampling versions of our estimators is reasonable to make them applicable in practice. The underlying idea is instead of summing up all possible index combinations of one group we just 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 $\nu \in (0, \infty)$ is chosen and used to define $w_i = \lceil \nu \cdot \binom{n_i}{6} \rceil, i = 1, \ldots, a$ as the number of subsampling repetitions done for the $i$-th group. It is clear that the choice of $\nu$ has a great influence on the calculation time and accuracy, so it should be chosen suitable for the situation.

Then, random subsamples $\sigma_i(b) = \{\sigma_{i1}(b), \ldots, \sigma_{6i}(b)\}$ 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

$$C_{i,1}^* = C_{i,1}(w_i) = \sum_{b=1}^{w_i} \Lambda_1(\sigma_i(b)) \cdot \Lambda_2(\sigma_i(b)) \cdot \Lambda_3(\sigma_i(b)).$$

Here

$$\Lambda_1(\ell_1, \ell_2, \ell_3, \ell_4, \ell_5, \ell_6) = Y_{i,\ell_1,\ell_2}^T Y_{i,\ell_3,\ell_4},$$

$$\Lambda_2(\ell_1, \ell_2, \ell_3, \ell_4, \ell_5, \ell_6) = Y_{i,\ell_3,\ell_4}^T Y_{i,\ell_5,\ell_6},$$

$$\Lambda_3(\ell_1, \ell_2, \ell_3, \ell_4, \ell_5, \ell_6) = Y_{i,\ell_5,\ell_6}^T Y_{i,\ell_1,\ell_2}.$$
Combining them, allows to define the subsampling version of $C_1$ by

$$C_1^* := \frac{1}{8} \cdot \sum_{i=1}^{a} w_i \cdot \sum_{i=1}^{a} C_{i,1}(w_i).$$

Theorem 3.5:
For $\sum_{i=1}^{a} w_i \to \infty$, if $N \to \infty$ (which includes frameworks (I)-(V)) it holds:

- a) $C_1^*$ is unbiased for $\text{tr} \left( (T_S \Sigma)^3 \right)$.
- b) $\hat{f}_P^* := \frac{\Lambda_i^2}{(C_1^*)^3} \cdot \eta_{N,a}$ fullfiles $(\hat{f}_P^*)^{-1} - (f_P)^{-1} \overset{p}{\rightarrow} 0$.

This way of defining the number of subsampling repetitions $w_i$, guarantees that the relation between the subsampled parts $C_{1,i}^*$ 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 the influence of single groups is not too big.

These results allow to formulate a more useable version of $K_{f_P}$ through the following theorem.

Theorem 3.6:
The results of Theorem [3.3] remains valid if $f_P$ is replaced by $\hat{f}_P$ or $\hat{f}_P^*$.

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 to be reasonable but in practice, we would again need subsampling versions of these estimators, which take care of the structure of the dataset. 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 $X_{ij} \sim N_d(\mu_i, \Sigma_i)$ with $T_S \Sigma_1 = T_S \Sigma_2 = ... = T_S \Sigma_a$, 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 $\text{tr} \left( (T_S \Sigma_1)^j \right) = \text{tr} \left( (T_S \Sigma_2)^j \right) = ... = \text{tr} \left( (T_S \Sigma_a)^j \right)$ for $j \in \{1, 2, 3\}$, but this is nearly impossible to justify in practice.
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 unique ones, because they don’t require the usual condition of increasing sample sizes, and therefore they are a strict expansion of the settings considered in Sattler and Pauly (2018).

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 $n = (n_1, ..., n_{12}) = (15, 15, 20, 35, 25, 30, 30, 35, 20, 15, 25)$. We used 10,000 simulation runs and chose $\nu = 0.05$ for our subsampling type estimators. Thereby, the number of subsampling draws are between 251 and 81,158, one basis of the quite unbalanced setting. Higher values for $\nu$ would increase the accuracy but noticeable extend the computation time.

Two different nullhypotheses are investigated, to have a situation with $\beta_1 \to 0$ as well as with $\beta_1 \to 1$. These hypotheses are

- $\mathcal{H}_0^a : (P_a \otimes P_d) \mu = 0$,

- $\mathcal{H}_0^b : (\frac{1}{d} I_a \otimes \frac{1}{d} I_d) \mu = 0$.

For both hypotheses the same distributional setting is choosen, with $\Sigma$ as a autoregressive covariance matrix with parameter 0.6 e.g. $(\Sigma)_{i,j} = 0.6^{\lvert i - j \rvert}$ and $\mu_i = \mathbf{0}_d$ for $i = 1, ..., a$, to achieve better comparability. 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: $\tau_p$ for $T = \frac{1}{\alpha} J_a \otimes \frac{1}{\alpha} J_d$ and $(\Sigma)_{ij} = 0.6^{|j-i|}$ with different dimension and numbers of groups.

| $\tau_p$ | $a=2$ | $a=3$ | $a=4$ | $a=5$ | $a=6$ | $a=7$ | $a=8$ | $a=9$ | $a=10$ | $a=11$ | $a=12$ |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $d=5$    | .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  | .004  | .004  | .004  | .004  | .003  |
| $d=600$  | .008  | .004  | .003  | .002  | .002  | .002  | .002  | .001  | .001  | .001  | .001  |

Figure 1: Simulated Type I-Error rates ($\alpha = 5\%$) for $\psi_z (\text{---})$, $\psi_\chi (\cdot \cdot \cdot)$ and $\varphi_N^* (\text{--})$ under the null hypothesis $H_0^a: (P_a \otimes P_d) \mu = 0$ for increasing dimension.
All tests $\psi_z = \mathbb{I}(W_N > z_{1-\alpha})$, $\psi_X = \mathbb{I}(W_N > \chi^2_{1,1-\alpha})$ and $\varphi_N = \mathbb{I}(W_N > K_{f_N,1-\alpha})$ are used while $\chi^2_{1,1-\alpha}$ denotes the $1-\alpha$ quantile of a $\chi^2$ distribution and $K_{f_N,1-\alpha}$ the $1-\alpha$ quantile of $K_{f_N}$. 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 $\psi_X$ results in too conservative test decisions, especially for larger dimension. So, in this case, a rate which is in most cases lower than 0.04 would lead to a raised number of rejections when the null-hypothesis is true. However, $\psi_z$ 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 $\phi_\star^N$ 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 $\psi^X$ and $\phi_\star^N$ in Figure 2. This is not surprising, because from Figure 1 we know that $f_p$ has always the value one. Furthermore, the small difference between both curves shows once more the good performance of the used estimators. Apart from that again the performance of $\phi_\star^N$ is quite good, in particular for a higher number of groups. Using the test $\phi_\star$ that is based on the wrong limit distribution under $H_{0b}$ results in considerably larger type-I error rates between 0.065 and 0.085.

To sum up $\phi_\star^N$ shows really good type-I error rates, overall settings, dimensions, and group numbers, even for substantially unbalanced sample sizes, which moreover contains 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 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 $\mu_1 = \mu_3 = \ldots, \mu_9 = 0_d$ and $(\mu_2)_k = (\mu_4)_k, \ldots, (\mu_{10})_k = \delta \cdot k/d$, $k = 1, \ldots, d$, $\delta \in [0, 2]$,

- a *one_point_alternative* with $\mu_1 = \mu_3 = \ldots, \mu_9 = 0_d$ and $\mu_2 = \mu_4, \ldots, \mu_{10} = \delta \cdot e_1$, $\delta \in [0, 3.5]$ and

- a *shift_alternative* with $\mu_1 = \mu_3 = \ldots, \mu_9 = 0$ and $(\mu_2) = (\mu_4), \ldots, (\mu_{10}) = \delta \cdot 1_d$ for $H_{0b}^b$, $\delta \in [0, 2]$

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

From the simulation result given in Sattler and Pauly (2018), 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 $\delta$.

Figure 4: Simulated power curves of $\varphi^*_N$ for a one-point alternative with $d = 50$, 10000 simulation runs and an autoregressive structure($\Sigma_{i,j} = 0.6^{|i-j|}$). The sample size is $n = (15, 15, 20, 35, 20, 30, 30, 35, 20)$ and different numbers of groups were considered, namely $a = 2(—)$, $a = 4(\cdots)$, $a = 8(\cdots)$ and $a = 10(\cdots)$.

For the trend alternative(Figure 3) $\varphi^*_N$ has a high power for both nullhypotheses, the power is essential higher for $H^b_0$. Increasing the number of groups also increases the power in both hypothesis. It is noticeable that for $H^a_0$ increasing the number from 8 to 10 groups has substantial more effect than from 2 to 4 groups while for $H^b_0$ 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 $\delta$- values in particular for $H^a_0$. This coincides with the power calculations from Sattler and Pauly (2018). But it can be seen that an increasing number of groups increase the power essentially.
Figure 5: Simulated power curves of $\phi^*_N$ for a shift alternative with $d = 50$, 10000 simulation runs and an autoregressive structure $(\Sigma)_{i,j} = 0.6^{|i-j|}$. The sample size is $n = (15, 15, 20, 35, 25, 20, 30, 30, 35, 20)$ and different numbers of groups were considered, namely $a = 2(\cdots)$, $a = 4(\cdots)$, $a = 8(\cdots)$ and $a = 10(\cdots)$.

Finally, we considered a shift alternative (Figure 5), but just for $\mathcal{H}^b_0$. As in other cases (Pauly et al. (2015), Sattler and Pauly (2018)), 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, $\phi^*_N$ has very high power even for these small sample sizes, especially $n_1 = n_2 = 15$. Moreover $\mathcal{H}^b_0$ is much easier to detect, in all settings.

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. Different to existing approaches we take this development into account by con-
sidering a variety of different frameworks. We were able to expand the central theorem of \cite{Sattler2018} to also cover this case, for the price of the additional assumption of equal covariance matrices. Moreover, we generalized it to some more cases that kind of completes the theorem. For all settings, we approximate the critical value of the test statistic, by a standardized $\chi^2_f$ 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 \cite{Box1953}, 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 \cite{Li2012} and \cite{Li2014} and we plan to combine their techniques with the results obtained in \cite{Sattler2019} in the near future.

**Acknowledgement**

The author would like to thank Markus Pauly for helpful discussions and many valuable suggestions. This work was supported by the German Research Foundation project DFG-PA2409/4-1.
Appendix

Proof of Theorem 3.1: For this proof, it is helpful to present the theorem in a more detailed way.

Let \( \beta_s = \lambda_s / \sqrt{\sum_{\ell=1}^{a_d} \lambda_{\ell}^2} \) for \( s = 1, \ldots, a_d \). Then \( \tilde{W}_N \) has, under \( H_0(T) \), and one of the frameworks [I-V] asymptotically

a) a standard normal distribution if and only if
\[
\beta_1 = \max_{s \leq a_d} \beta_s \to 0 \quad \text{as} \quad N \to \infty,
\]

b) a standardized \((\chi_1^2 - 1) / \sqrt{2}\) distribution if and only if
\[
\beta_1 \to 1 \quad \text{as} \quad N \to \infty,
\]

c) a distribution of the shape \( \sum_{s=1}^r b_s (C_s - 1) / \sqrt{2} + \sqrt{1 - \sum_{s=1}^r b_s^2} \cdot Z \), if and only if

for all \( s \in \mathbb{N} \) \( \beta_s \to b_s \) 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 \( C_i \overset{\text{i.i.d.}}{\sim} \chi_1^2 \), \( Z \sim \mathcal{N}(0, 1) \).

d) a distribution of the shape \( \sum_{s=1}^\infty b_s (C_s - 1) / \sqrt{2} \), if

for all \( s \in \mathbb{N} \) \( \beta_s \to b_s \) as \( N \to \infty \),

for a decreasing sequence \((b_s)_s\) in \((0, 1)\) with \( \sum_{s=1}^\infty b_s^2 = 1 \) and \( C_i \overset{\text{i.i.d.}}{\sim} \chi_1^2 \).

The first two parts as well as the last one were proved in Sattler and Pauly (2018). 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 an infinite amount \( M \subseteq \mathbb{N} \) with

\[
\sum_{\ell \in M} \beta_\ell \left( \frac{C_\ell - 1}{\sqrt{2}} \right) \overset{\text{d}}{\to} \sqrt{1 - \sum_{s=1}^r b_s^2} \cdot Z.
\]

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 \( \beta_{r'} > 0 \) and \( \beta_{r'+1} = 0 \). Assume now that \( \beta_\ell \to b'_\ell \) for \( \ell = 1, \ldots, r' \) otherwise consider the subsequence where this holds. It remains to show that from

\[
\sum_{\ell=1}^{r'} \beta_\ell \left( \frac{C_\ell - 1}{\sqrt{2}} \right) \to \sum_{\ell=1}^{r'} b'_\ell \left( \frac{C_\ell - 1}{\sqrt{2}} \right) \overset{\text{d}}{=} \sum_{\ell=1}^r b_\ell \left( \frac{C_\ell - 1}{\sqrt{2}} \right),
\]

19
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 continuous mapping theorem we have for all \( t \in \mathbb{R} \)

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

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

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 [II] just part-wise, it would be possible to use other more classical estimators for the unknown traces. Nevertheless our focus was to develop preferably general estimators which can be used in a variety of settings.

**Lemma 6.1:**
With

\[
A_{i,1} = \frac{1}{2} \sum_{\ell_1 \neq \ell_2 = 1}^{n_i} (X_{i,\ell_1} - X_{i,\ell_2})^\top T_S (X_{i,\ell_1} - X_{i,\ell_2})
\]

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 \( \text{tr}(T_S \Sigma) \), in all of our frameworks.

**Proof:** It is obvious that this is a unbiased estimator of \( \text{tr}(T_S \Sigma) \). With well known rules and analogous to Sattler and Pauly (2018) we calculate

\[
\text{Var}(A_1) \leq \frac{1}{\left[ \sum_{i=1}^{a} \left( \begin{array}{c} n_i \\ 2 \end{array} \right) \right]^2} \sum_{i=1}^{a} \left( \begin{array}{c} n_i \\ 2 \end{array} \right) \left( \left( \begin{array}{c} n_i \\ 2 \end{array} \right) - \left( \begin{array}{c} n_i - 2 \\ 2 \end{array} \right) \right) \cdot o(\text{tr}^2(T_S \Sigma)).
\]
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 \( n_{\text{max}} \to \infty \). Then

\[
\text{Var}(A_1) \leq \frac{1}{\sum_{i=1}^{a} \left( \begin{array}{c} \frac{n_i}{2} \\ \frac{n_{\text{max}}}{2} \end{array} \right)} \sum_{i=1}^{a} \left( \begin{array}{c} n_i \\ \frac{n_{\text{max}}}{2} \end{array} \right) \left( \binom{\frac{n_i}{2}}{2} \right) \cdot o\left(\text{tr}^2(T_S \Sigma)\right)
\]

\[
\leq \frac{1}{\sum_{i=1}^{a} \left( \begin{array}{c} \frac{n_i}{2} \\ \frac{n_{\text{max}}}{2} \end{array} \right)} \sum_{i=1}^{a} \left( \begin{array}{c} n_i \\ \frac{n_{\text{max}}}{2} \end{array} \right) \left( \binom{\frac{n_{\text{max}}}{2}}{2} - \binom{\frac{n_{\text{max}}}{2}}{2} \right) \cdot o\left(\text{tr}^2(T_S \Sigma)\right)
\]

\[
= \frac{\left( \binom{\frac{n_{\text{max}}}{2}}{2} - \binom{\frac{n_{\text{max}}}{2}}{2} \right)}{\frac{n_{\text{max}}}{2}} \cdot o\left(\text{tr}^2(T_S \Sigma)\right)
\]

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

\[
\text{Var}(A_1) \leq \frac{1}{\sum_{i=1}^{a} \left( \begin{array}{c} \frac{n_i}{2} \\ \frac{n_{\text{min}}}{2} \end{array} \right)} \sum_{i=1}^{a} \left( \begin{array}{c} n_i \\ \frac{n_{\text{min}}}{2} \end{array} \right) \left( \binom{\frac{n_{\text{max}}}{2}}{2} - \binom{\frac{n_{\text{max}}}{2}}{2} \right) \cdot o\left(\text{tr}^2(T_S \Sigma)\right)
\]

\[
= \frac{\left( \binom{\frac{n_{\text{max}}}{2}}{2} - \binom{\frac{n_{\text{max}}}{2}}{2} \right)}{\frac{n_{\text{min}}}{2}} \cdot o\left(\text{tr}^2(T_S \Sigma)\right)
\]

So dividing by \( \text{tr}^2(T_S \Sigma) \) and then use 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 6.2:**

The estimator given by

\[
A_2 = \sum_{i=1}^{a} \sum_{\ell_1 = 1}^{n_i} \sum_{\ell_2 = 1}^{n_i} \sum_{k_1 = 1}^{n_i} \sum_{k_2 = 1}^{n_i} \frac{\left( X_{i,\ell_1} - X_{i,\ell_2} \right)^T T_S \left( X_{i,k_1} - X_{i,k_2} \right)}{4 \cdot 6 \sum_{i=1}^{a} \left( \begin{array}{c} n_i \\ 4 \end{array} \right)}
\]

is a unbiased and ratio-consistent estimator of \( \text{tr} \left( \left( T_S \Sigma \right)^2 \right) \) in all our asymptotic frameworks.

**Proof:** Again the unbiasedness is clear and we consider the variance.
We calculate

$$Var(A_2) = \sum_{i=1}^{a} \frac{Var \left( \sum_{e_1>l_2 \land e_2>k_2 \land e_1 \neq e_2 \land e_1 \neq k_2}^{n_i} \left( X_{i,e_1} - X_{i,e_2} \right)^{T} T_{S} \left( X_{i,k_1} - X_{i,k_2} \right)^2 \right)}{\left[ 4 \cdot \sum_{i=1}^{a} \binom{n_i}{4} \right]^2}$$

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

$$Var(A_2) \leq O \left( n_{\text{max}}^{-1} \right) \cdot O \left( \text{tr}^2 \left( (T_{S} \Sigma)^2 \right) \right)$$

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

$$Var(A_2) \leq O \left( a^{-1} \right) \cdot O \left( \text{tr}^2 \left( (T_{S} \Sigma)^2 \right) \right).$$

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 3.2**: The result follows directly by theorem 3.2 from Sattler and Pauly (2018).

For the proof of Theorem 3.4 we need to show different properties that combined leads to the result.

**Proof of Theorem 3.4**: We conduct this proof in several steps:

a) $\mathbb{E}(C_1) = \text{tr} \left( (T_{S} \Sigma)^3 \right)$,

b) $Var(C_1) = \frac{\sum_{i=1}^{a} \binom{n_i}{4} \left( \binom{n_i}{6} - \binom{n_i-6}{6} \right)}{\left( \sum_{i=1}^{a} \binom{n_i}{6} \right)^2} \cdot O \left( \text{tr}^3 \left( (T_{S} \Sigma) \right) \right)$,

c) $\frac{C_1}{\text{tr}^{3/2}(T_{S} \Sigma)^2} - \frac{\text{tr}(T_{S} \Sigma)^3}{\text{tr}^{3/2}(T_{S} \Sigma)^2} \xrightarrow{p} 0$ in our frameworks,

d) $\frac{C_1}{A_2} - (f_P)^{-1} \xrightarrow{p} 0$ in our frameworks.
The results from Sattler and Pauly (2018) directly yield to

$$\mathbb{E}(C_1) = \text{tr} \left( (T_S \Sigma)^3 \right)$$

and

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

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

For part c) we calculate

$$\mathbb{E} \left( \frac{C_1}{\text{tr}^{3/2} ((T_S \Sigma)^2)} - \frac{\text{tr} \left( (T_S \Sigma)^3 \right)}{\text{tr}^{3/2} ((T_S \Sigma)^2)} \right) = 0$$

and

$$\text{Var} \left( \frac{C_1}{\text{tr}^{3/2} ((T_S \Sigma)^2)} - \frac{\text{tr} \left( (T_S \Sigma)^3 \right)}{\text{tr}^{3/2} ((T_S \Sigma)^2)} \right)$$

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

Again this number is in $O(n_{\text{max}}^{-1})$ for $n_{\text{max}} \to \infty$ and in $O(a^{-1})$ 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 Sattler and Pauly (2018) but we repeat it for better understanding.

With the last lemma it follows for both cases that

$$\frac{C_i^2}{\text{tr}^3 ((T_S \Sigma)^2)} - \frac{1}{f_{T}} = \left( \frac{C_i}{\text{tr}^{3/2} ((T_S \Sigma)^2)} \right)^2 - \left( \frac{\text{tr}( (T_S \Sigma)^3 )}{\text{tr}^{3/2} ((T_S \Sigma)^2)} \right)^2$$

$$= \left[ \frac{C_i}{\text{tr}^{3/2} ((T_S \Sigma)^2)} - \frac{\text{tr}( (T_S \Sigma)^3 )}{\text{tr}^{3/2} ((T_S \Sigma)^2)} \right] \left[ \frac{C_i}{\text{tr}^{3/2} ((T_S \Sigma)^2)} + \frac{\text{tr}( (T_S \Sigma)^3 )}{\text{tr}^{3/2} ((T_S \Sigma)^2)} \right]$$

$$= O_p(1) \cdot \left[ \frac{C_i}{\text{tr}^{3/2} ((T_S \Sigma)^2)} - \frac{\text{tr}( (T_S \Sigma)^3 )}{\text{tr}^{3/2} ((T_S \Sigma)^2)} + 2 \frac{\text{tr}( (T_S \Sigma)^3 )}{\text{tr}^{3/2} ((T_S \Sigma)^2)} \right] = O_p(1),$$
were for the last step the trace inequality was used together with Slutzky’s theorem. With the ratio-consistency of \( A_2 \) it follows
\[
\frac{C_2^2}{A_2^2} - \left( \frac{f_p}{1} \right)^{-1} = \frac{\text{tr}^2((T_S \Sigma)^3)}{\text{tr}^2((T_S \Sigma)^2)} \frac{C_2^2}{\text{tr}^3((T_S \Sigma)^2)} - \left( \frac{f_p}{1} \right)^{-1}
\]
\[
= (1 + o_p(1)) \cdot \frac{1}{f_p} - \frac{1}{f_p}
\]
\[
= \frac{1}{f_p} - \frac{1}{f_p} + o_p(1) \cdot \frac{1}{f_p} = o_p(1).
\]

\[\square\]

It is obvious that this estimator needs a sufficient big amount of groups with at least 6 observations. Similar for the other estimators, which were introduced earlier. From a theoretical point of view a scenario with \( n_{\text{max}} \leq 5 \) 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 3.5**: For this proof, some results of [Sattler and Pauly (2018)] are used and adapted. First the expectation value of the estimator, using the notation \( w := \sum_i w_i \):
\[
\mathbb{E} \left( C_1^* \right) = \mathbb{E} \left( \frac{1}{8 \sum_{i=1}^a w_i} \sum_{i=1}^a C_{i,1}^* \right)
\]
\[
= \frac{1}{8 \cdot w} \sum_{i=1}^a \mathbb{E} \left( C_{i,1}^* \right)
\]
\[
= \frac{1}{8 \cdot w} \sum_{i=1}^a \mathbb{E} \left( \sum_{b=1}^5 \Lambda_1(\sigma_i(b)) \cdot \Lambda_2(\sigma_i(b)) \cdot \Lambda_3(\sigma_i(b)) \right)
\]
\[
= \frac{1}{8 \cdot w} \sum_{i=1}^a \sum_{b=1}^5 \mathbb{E} \left( \Lambda_1(\sigma_i(b)) \cdot \Lambda_2(\sigma_i(b)) \cdot \Lambda_3(\sigma_i(b)) \right)
\]
\[
= \frac{1}{8 \cdot w} \sum_{i=1}^a w_i \cdot \mathbb{E} \left( \Lambda_1(1, 2, 3, 4, 5, 6) \cdot \Lambda_2(1, 2, 3, 4, 5, 6) \cdot \Lambda_3(1, 2, 3, 4, 5, 6) \right)
\]
\[
= \frac{1}{8 \cdot w} \sum_{i=1}^a w_i \cdot \frac{\text{tr} \left( (T_S \Sigma)^3 \right)}{8 \cdot \text{tr} \left( (T_S \Sigma)^2 \right)}
\]
\[
= \text{tr} \left( (T_S \Sigma)^3 \right).
\]
With Theorem A.9 Theorem A.10 and Theorem A.16 from [Sattler and Pauly (2018)] for the variance we get
\[ \text{Var} \left( C_i \right) = \frac{1}{(8-w)^2} \sum_{i=1}^{a} \text{Var} \left( C_{i,i} \right) \]
\[ \leq \frac{1}{(8-w)^2} \sum_{i=1}^{a} w_i^2 \cdot \left[ 0 + 1 - \left( 1 - \frac{1}{w_i} \right) \cdot \frac{(n_i - 6)}{n_i^6} \right] . \]

Again there the same to cases. If \( n_{\max} \) is bound and therefore \( \max_{i=1,...,a} w_i \) is bound, it follows \( a \to \infty \) and hereby
\[ \frac{1}{(8-w)^2} \sum_{i=1}^{a} w_i^2 \cdot \left[ 0 + 1 - \left( 1 - \frac{1}{w_i} \right) \cdot \frac{(n_i - 6)}{n_i^6} \right] \]
\[ \leq \frac{1}{(8-w) \cdot a \cdot \min_{i=1,...,a} \left( w_i \right)} \cdot \max_{i=1,...,a} \left( w_i \right) \sum_{i=1}^{a} w_i \cdot 1 \]
\[ = \Theta \left( a^{-1} \right) \cdot \frac{\max_{i=1,...,a} \left( w_i \right)}{\min_{i=1,...,a} \left( w_i \right)} \]
\[ = \Theta \left( a^{-1} \right) \]

while for \( n_{\max} \to \infty \) which implies \( \max_{i=1,...,a} \left( w_i \right) \to \infty \) we calculate first
\[ w_i^2 \cdot \left[ 0 + 1 - \left( 1 - \frac{1}{w_i} \right) \cdot \frac{(n_i - 6)}{n_i^6} \right] \]
\[ = w_i \cdot \left[ 1 - \left( \frac{(n_i - 6)}{n_i^6} \right) \right] + \left( \frac{(n_i - 6)}{n_i^6} \right) \]
\[ \leq w_i \cdot \left[ \left( \left( \frac{n_i}{6} \right) - (n_i - 6) \right) + 1 \right] \]
\[ \leq w_i \cdot \left[ \left( \frac{n_{\min}}{6} \right) - (n_{\min} - 6) \right] + 1 \]

and therefore
\[ \leq \frac{1}{(8-w)^2} \sum_{i=1}^{a} w_i^2 \cdot \left[ 0 + 1 - \left( 1 - \frac{1}{w_i} \right) \cdot \frac{(n_i - 6)}{n_i^6} \right] \]
\[ \leq \frac{1}{(8-w)^2} \sum_{i=1}^{a} w_i \cdot \left[ \left( \frac{n_{\min}}{6} \right) - (n_{\min} - 6) \right] + 1 \]
\[ \leq \frac{1}{(64-w) \cdot \max_{i=1,...,a} \left( w_i \right)} \sum_{i=1}^{a} w_i \cdot \left[ \left( \frac{n_{\min}}{6} \right) - (n_{\min} - 6) \right] + 1 \]
\[ \leq \frac{1}{(64-w) \cdot \left( \frac{n_{\max}}{6} - 1 \right)} \sum_{i=1}^{a} w_i \cdot \left[ \left( \frac{n_{\min}}{6} \right) - (n_{\min} - 6) \right] + 1 \]
\[ = \Theta \left( \frac{n_{\min} - 6}{64 \cdot \frac{n_{\max}}{6} - 1} \right) = \Theta \left( n_{\max}^{-1} \right) . \]

From this both values the results follows analogous to the proof of Theorem \ref{thm:3.4}. \qed
References

Bathke, A. (2002). Anova for a large number of treatments. *Mathematical Methods of Statistics*, 11.

Bathke, A., 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. and Lankowski, D. (2005). Rank procedures for a large number of treatments. *Journal of Statistical Planning and Inference*, 133:223–238.

Box, G. E. P. (1953). Non-normality and tests on variances. *Biometrika*, 40(3-4):318–335.

Brunner, E., Bathke, A., and Placzek, M. (2012). Estimation of box’s e for low- and high-dimensional repeated measures designs with unequal covariance matrices. *Biometrical journal. Biometrische Zeitschrift*, 54:301–16.

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.

Friedrich, S., Brunner, E., and Pauly, M. (2017). Permuting longitudinal data in spite of the dependencies. *Journal of Multivariate Analysis*, 153:255–265.

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.

Kong, X. and Harrar, S. W. (2019). Accurate inference for repeated measures in high dimensions. *Electron. J. Statist.*, 13(2):4916–4944.

Konietschke, F., Bathke, A., Harrar, S., and Pauly, M. (2015). Parametric and non-parametric bootstrap methods for general MANOVA. *Journal of Multivariate Analysis*, 140:291–301.

Li, J. and Chen, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist.*, 40(2):908–940.

Li, W. and Qin, Y. (2014). Hypothesis testing for high-dimensional covariance matrices. *Journal of Multivariate Analysis*, 128:108 – 119.
Mathai, A. M. and Provost, S. B. (1992). Quadratic forms in random variables, vol. 126 of statistics: Textbooks and monographs.

Pauly, M., Ellenberger, D., and Brunner, E. (2015). Analysis of high-dimensional one group repeated measures designs. *Statistics*, 49:1243–1261.

Sattler, P., Bathke, A. C., and Pauly, M. (2019). Testing hypotheses about covariance matrices in general manova designs.

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.