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

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

Paavo Sattler and Markus Pauly

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

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

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

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

Received October 2017.

Contents

1. Introduction

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

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

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

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

2. Statistical model and hypotheses

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

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

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

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

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

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

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

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

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

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

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

3. The test statistic and its asymptotics

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

$$a \in \mathbb{N} \text{ fixed and } \min(d, n_1, \ldots, n_a) \to \infty, \tag{3}$$

$$d \in \mathbb{N} \text{ fixed and } \min(a, n_1, \ldots, n_a) \to \infty, \tag{4}$$

$$\text{or } \min(a, d, n_1, \ldots, n_a) \to \infty, \tag{5}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

a) a standard normal distribution if and only if

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

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

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

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

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

Theorem 3.2.

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

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

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

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

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

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

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

3.1. Symmetrized trace estimators

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

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

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

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

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

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

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

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

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

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

It follows from Lemma 3.3 that

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

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

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

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

Theorem 3.5.

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

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

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

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

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

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

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

4. Better approximations

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

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

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

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

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

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

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

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

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

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

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

where

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

5. Simulations

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

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

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

and the variance estimator

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

using

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

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

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

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

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

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

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

5.1. Asymptotic distribution and type-I error control

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

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

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

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

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

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

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

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

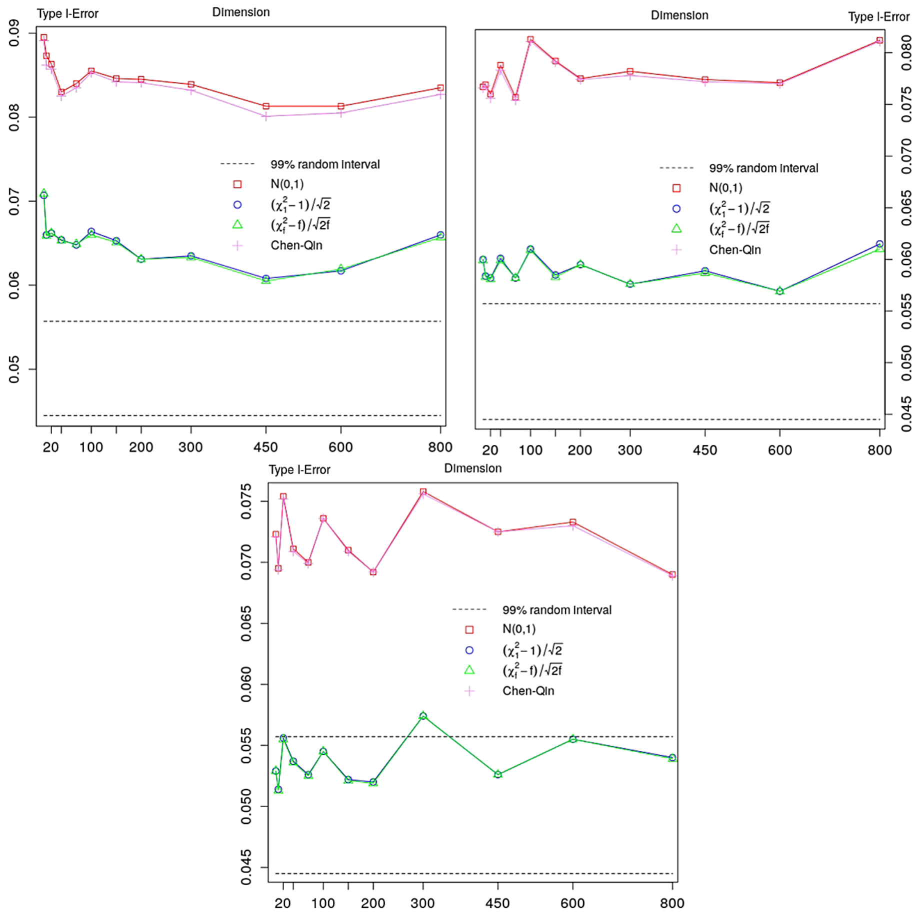


FIG 1. *Simulated type-I error rates ($\alpha = 5\%$) for the statistic $W_N \cdot \sqrt{N/(N-1)}$ compared with the critical values of a standard normal, standardized χ_1^2 and K_f-distribution and the test ψ_{CQ} of Chen & Qin under the null hypothesis $H_0^a : \left(\boldsymbol{P}_2 \otimes \frac{1}{d}\boldsymbol{J}_d\right)\boldsymbol{\mu}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

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

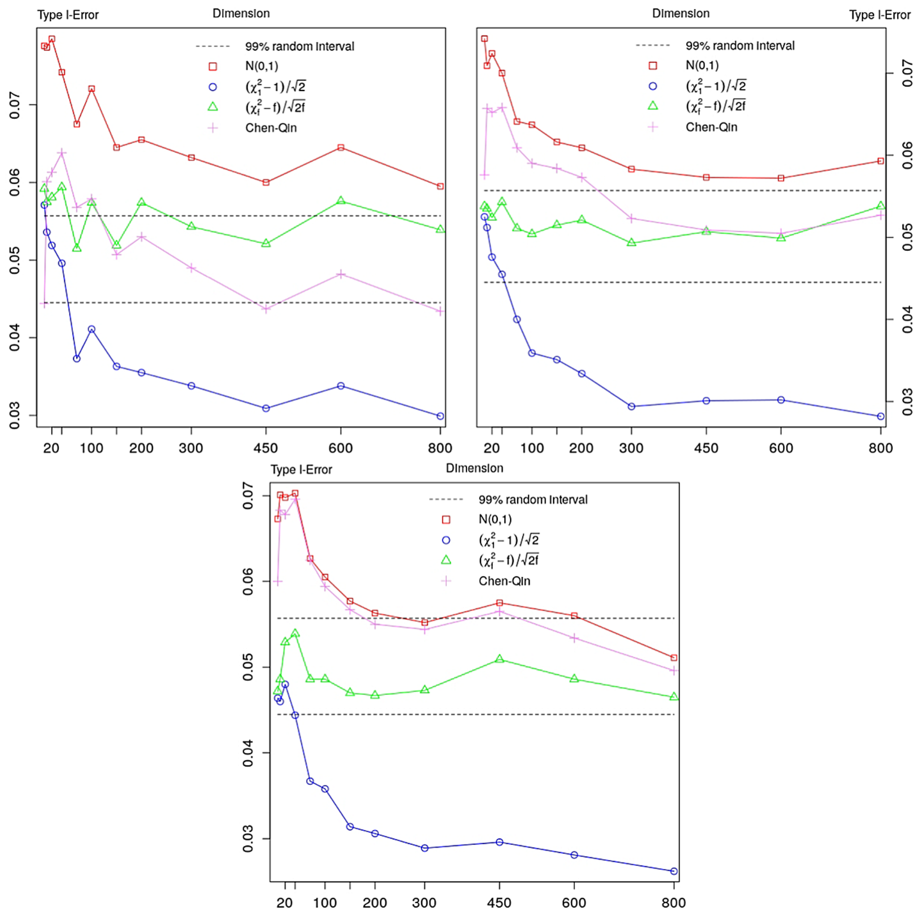


FIG 2. *Simulated type-I error rates ($\alpha = 5\%$) for the statistic $W_N \cdot \sqrt{N/(N-1)}$ compared with the critical values of a standard normal, standardized χ_1^2 and K_f-distribution and the test ψ_{CQ} of Chen & Qin under the null hypothesis $H_0^b : \left(\frac{1}{2}\boldsymbol{J}_2 \otimes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

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

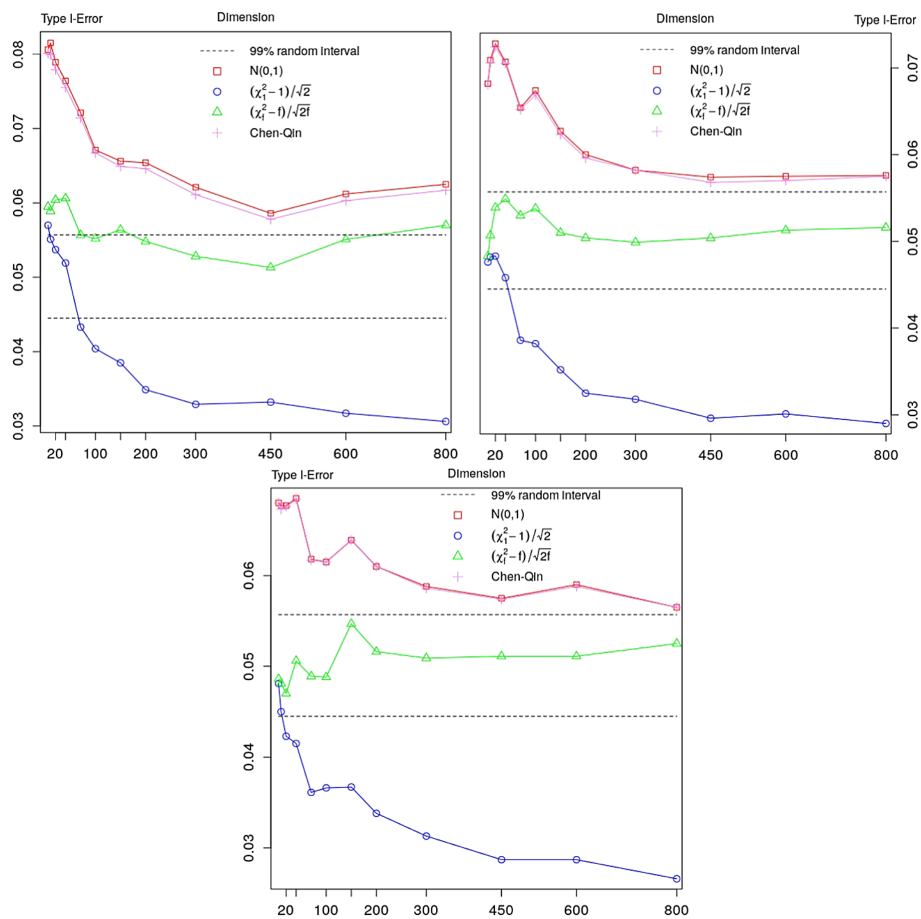


FIG 3. *Simulated type-I error rates ($\alpha = 5\%$) for the statistic $W_N \cdot \sqrt{N/(N-1)}$ compared with the critical values of a standard normal, standardized χ_1^2 and K_f-distribution and the test ψ_{CQ} of Chen & Qin, under the null hypothesis $H_0^{ab} : \left(\frac{1}{2}\boldsymbol{J}_2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

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

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

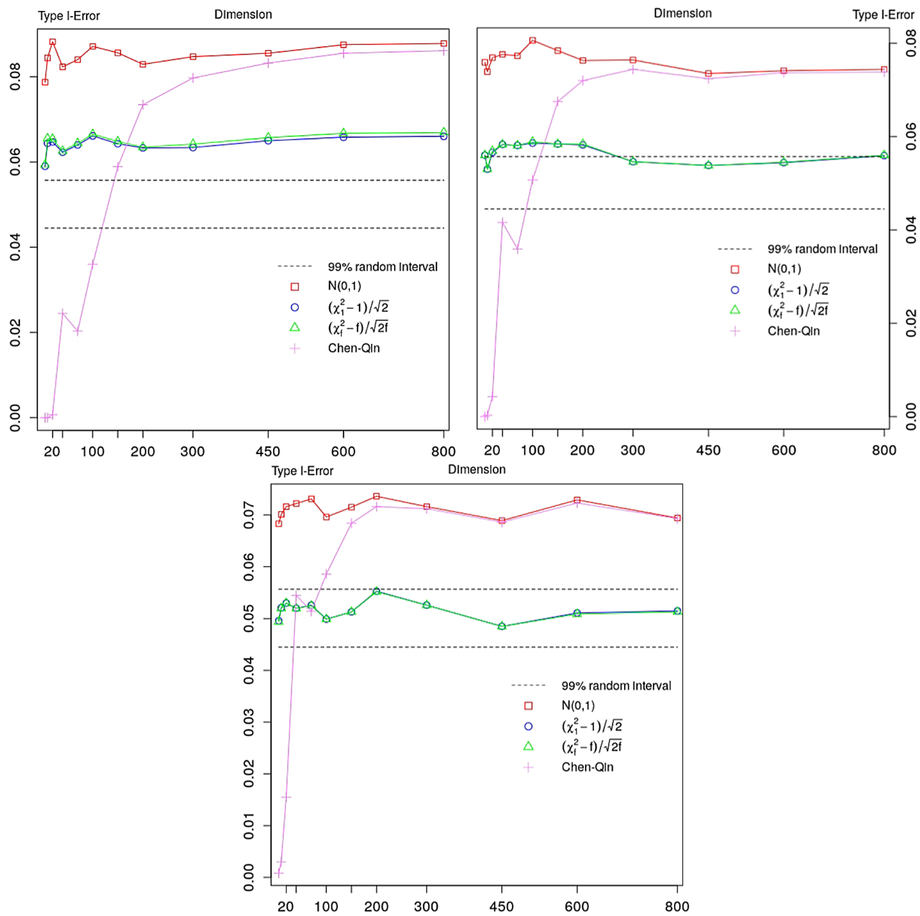


FIG 4. *Simulated type-I error rates ($\alpha = 5\%$) for the statistic $W_N \cdot \sqrt{N/(N-1)}$ compared with the critical values of a standard normal, standardized χ_1^2 and K_f-distribution and the test ψ_{CQ} of Chen & Qin, under the null hypothesis $H_0^b : \left(\frac{1}{2}\boldsymbol{J}_2 \otimes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

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

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

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

5.2. Power performance

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

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

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

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

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

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

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

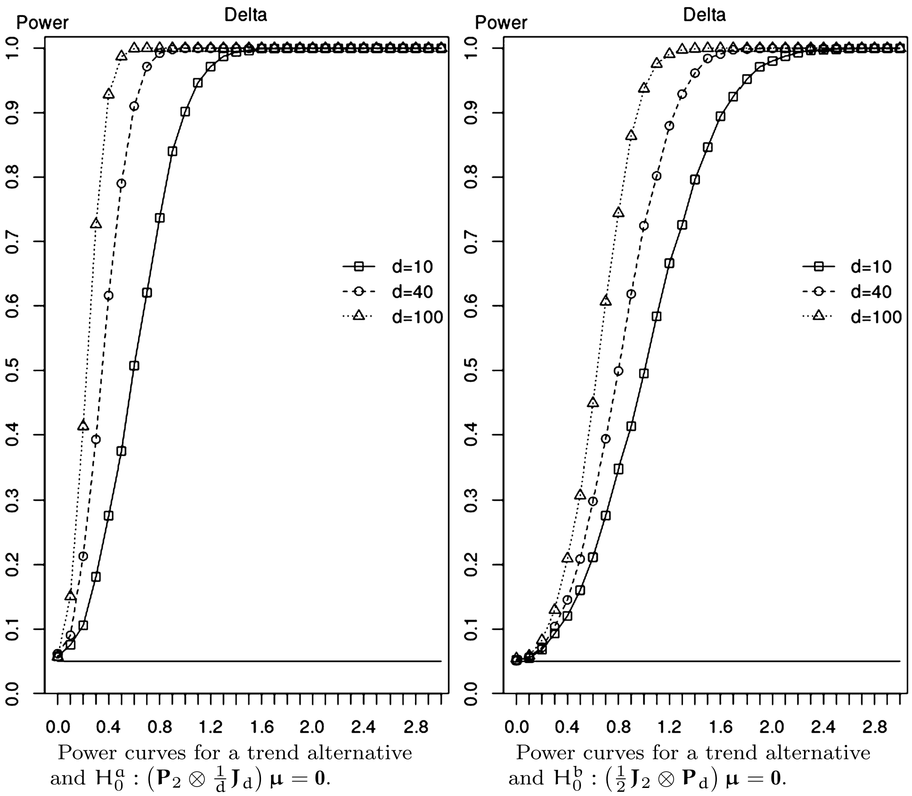


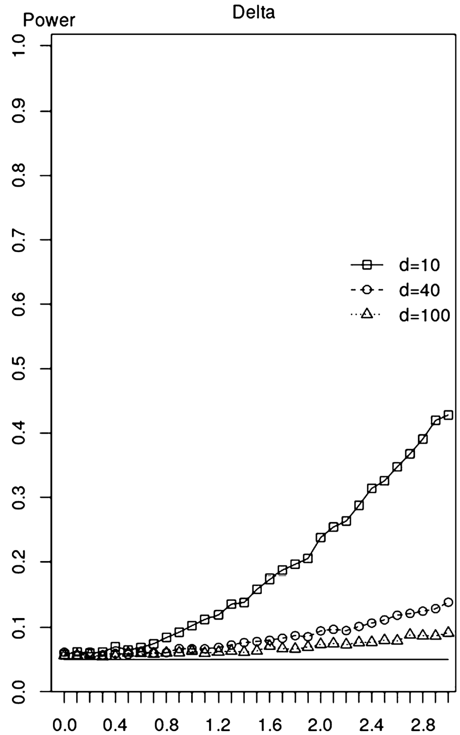
FIG 5. *Simulated power curves for the statistic* $W_N \cdot \sqrt{(N-1)/N}$ *in* 10^4 *simulation runs for different dimensions with* $n_1 = 20, n_2 = 30$ *and an autoregressive structure*($(\boldsymbol{\Sigma}_1)_{i,j} = 0.6^{|i-j|}$ *and* $(\boldsymbol{\Sigma}_2)_{i,j} = 0.65^{|i-j|}$).

6. Analysis of a sleep laboratory data set

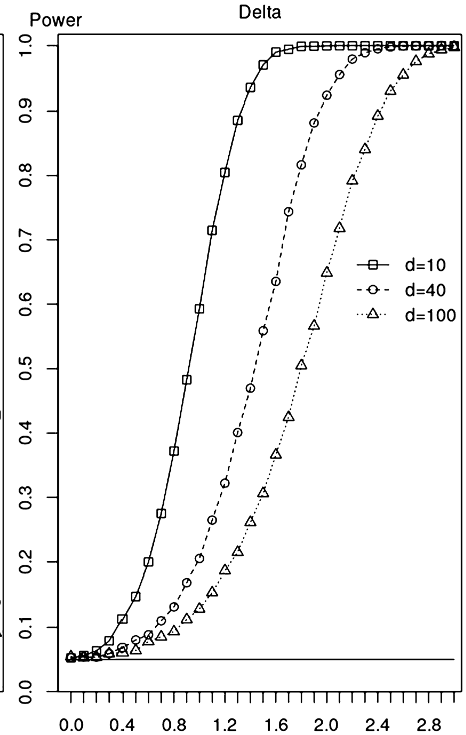
Finally, the new methods are exemplified on the sleep laboratory trial reported in Jordan et al. [25]. In this two-armed repeated measures trial, the activity of prostaglandin-D-synthase (β-trace) was measured every 4 hours over a period of 4 days. The grouping factor was gender and the above $d = 24$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

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

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



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

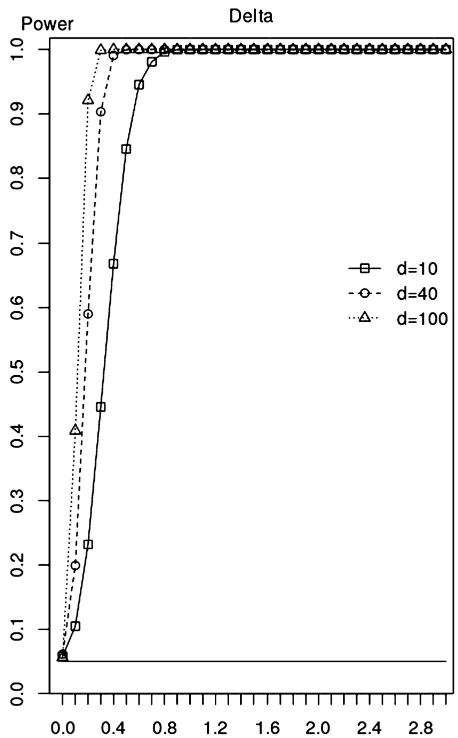


Power curves for an one-point alternative and $\mathrm{H}_0^{\mathrm{b}} : \left(\frac{1}{2}\mathbf{J}_2 \otimes \mathbf{P}_{\mathrm{d}}\right)\boldsymbol{\mu} = \mathbf{0}$.

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

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

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



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

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

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

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

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

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

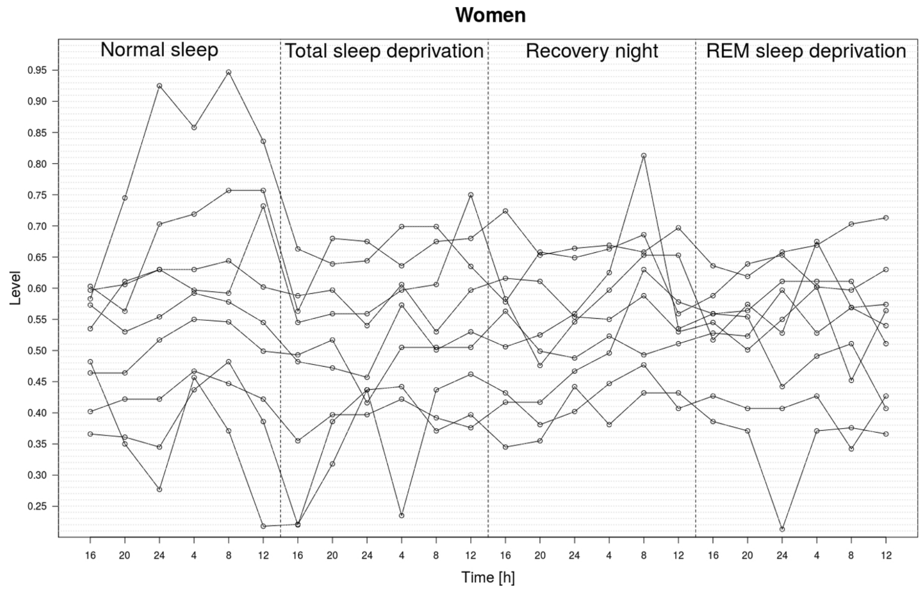


Fig 8. *Prostaglandin-D-synthase (ß-trace) of 10 young women during 4 days under different sleep conditions.*

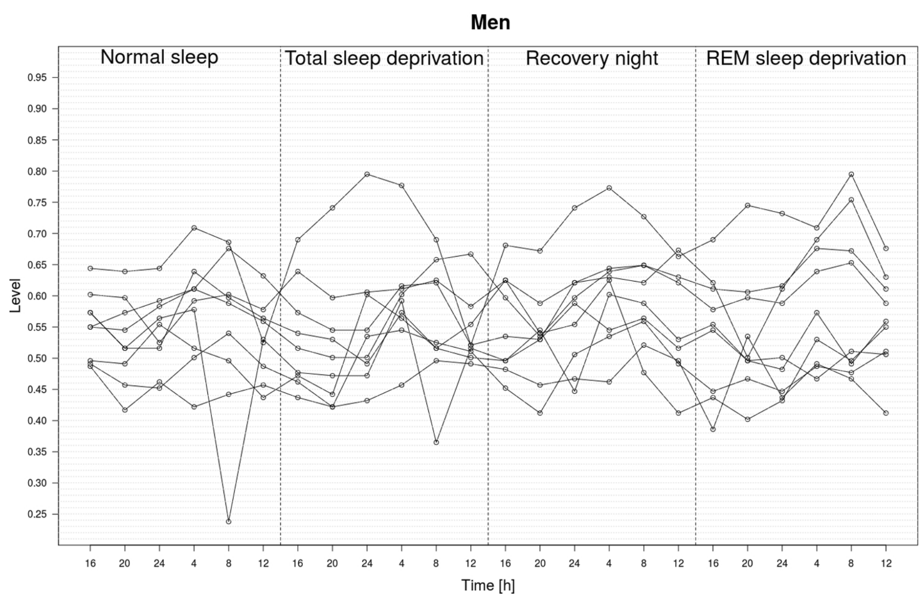


Fig 9. *Prostaglandin-D-synthase (ß-trace) of 10 young men during 4 days under different sleep conditions.*

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

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

7. Conclusion & outlook

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

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

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

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

Acknowledgement

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

Appendix A: Basics

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

all equations follows with the previous theorem. □

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

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

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

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

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

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

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

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

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

and so for X, Y *identically distributed*

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

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

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

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

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

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

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

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

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

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

Proof. With the rules for conditional expectations we calculate

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

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

The result follows by sum up this both parts. □

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

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

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

and

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

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

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

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

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

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

$$\mathbb{P}\left(\boldsymbol{\sigma}(k,m) \neq! \boldsymbol{\sigma}(\ell,m)\right) = \mathbb{P}\left(\bigcap_{i=1}^{a}\left(\boldsymbol{\sigma}_i(k,m) \neq! \boldsymbol{\sigma}_i(\ell,m)\right)\right)$$

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

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

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

All in all we calculate

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

$$= 1 - \left(1 - \frac{1}{B}\right) \cdot \prod_{i=1}^{a} \frac{\binom{n_i-m}{m}}{\binom{n_i}{m}}.$$

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

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

Appendix B: Proofs of Section 3

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

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

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

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

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

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

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

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

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

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

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

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

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

In the special case $t = 1$ this means

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

But we can size up the product by

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

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

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

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

For part a) we calculate

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

The second summand fulfills

$$\mathbb{E}\left(\frac{2\sqrt{N}\boldsymbol{\mu}^\top \boldsymbol{T}\boldsymbol{Z}}{\sqrt{2\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}}\right) = 0,$$

$$Var\left(\frac{2\sqrt{N}\boldsymbol{\mu}^\top \boldsymbol{T}\boldsymbol{Z}}{\sqrt{2\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}}\right) = 2\frac{N\boldsymbol{\mu}^\top \boldsymbol{T}\boldsymbol{V}_N\boldsymbol{T}\boldsymbol{\mu}}{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} \in \mathcal{O}(1)$$

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

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

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

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

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

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

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

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

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

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

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

as before. This evidently leads to

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

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

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

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

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

$$A_{i,3} = \frac{1}{4 \cdot 6\binom{n_i}{4}} \sum_{\substack{\ell_1,\ell_2=1 \\ \ell_1 > \ell_2}}^{n_i} \sum_{\substack{k_2=1 \\ k_2 \neq \ell_1,\ell_2}}^{n_i-1} \sum_{\substack{k_1=k_2+1 \\ k_1 \neq \ell_1,\ell_2}}^{n_i} \times \left[(\boldsymbol{X}_{i,\ell_1} - \boldsymbol{X}_{i,\ell_2})^\top \boldsymbol{T}_S (\boldsymbol{X}_{i,k_1} - \boldsymbol{X}_{i,k_2})\right]^2,$$

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

For $\ell \neq k$ we know $\boldsymbol{Y}_{i,\ell,k} \sim \mathcal{N}(\boldsymbol{0}_d, 2\boldsymbol{T}_S\boldsymbol{\Sigma}_i\boldsymbol{T}_S)$ and for totally different indices the $\boldsymbol{Y}_{i,\ell,k}$ are statistically independent. So the previous lemmata can be used to calculate the moments. The unbiasedness can be shown by calculating the expectation values for each estimator

$$\mathbb{E}(A_{i,1}) = \frac{1}{2 \cdot \binom{n_i}{2}} \sum_{\substack{\ell_1,\ell_2=1 \\ \ell_1 > \ell_2}}^{n_i} \mathbb{E}\left[\boldsymbol{Y}_{i,\ell_1,\ell_2}{}^\top \boldsymbol{Y}_{i,\ell_1,\ell_2}\right] \stackrel{\text{A.5}}{=} \operatorname{tr}(\boldsymbol{T}_S\boldsymbol{\Sigma}_i).$$

The following argument will be used several times in this work with small differences, so incidentally it will be more detailed.

We recognize first that $Cov\left[\boldsymbol{Y}_{i,\ell_1,\ell_2}{}^\top \boldsymbol{Y}_{i,\ell_1,\ell_2}; \boldsymbol{Y}_{i,\ell_1',\ell_2'}{}^\top \boldsymbol{Y}_{i,\ell_1',\ell_2'}\right]$ is 0 if all indices are totally different, so just $\binom{n_i}{2}\left(\binom{n_i}{2} - \binom{n_i-2}{2}\right)$ combinations remain. Instead of calculating the covariances of the remaining quadratic forms it is easier to use lemmata from above. By using the fact that all quadratic forms are identically distributed, we can calculate the variances which are all the same so it is just the number of remaining combinations multiplied with the variances. This leads to:

$$
\begin{aligned}
Var\left(A_{i,1}\right) &= \frac{1}{4\cdot\binom{n_i}{2}^2}\sum_{\substack{\ell_1,\ell_2=1\\ \ell_1>\ell_2}}^{n_i}\sum_{\substack{\ell_1',\ell_2'=1\\ \ell_1'>\ell_2'}}^{n_i} Cov\left[\boldsymbol{Y}_{i,\ell_1,\ell_2}{}^\top\boldsymbol{Y}_{i,\ell_1,\ell_2}\boldsymbol{;}\boldsymbol{Y}_{i,\ell_1',\ell_2'}{}^\top\boldsymbol{Y}_{i,\ell_1',\ell_2'}\right]\\
&\overset{\text{A.7}}{\leq} \frac{\binom{n_i}{2}-\binom{n_i-2}{2}}{4\binom{n_i}{2}}\; Var\left[\boldsymbol{Y}_{i,1,2}{}^\top\boldsymbol{Y}_{i,1,2}\right] + \frac{\binom{n_i-2}{2}}{4\binom{n_i}{2}}\cdot 0\\
&\overset{\text{A.5}}{=} \frac{\binom{n_i}{2}-\binom{n_i-2}{2}}{4\binom{n_i}{2}}\mathcal{O}\left(\operatorname{tr}^2\left(2\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)\right)\\
&= \mathcal{O}\left(n_i^{-1}\right)\cdot\mathcal{O}\left(\operatorname{tr}^2\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)\right).
\end{aligned}
$$

With these values we know for $\boldsymbol{V}_N = \bigoplus_{i=1}^a \frac{N}{n_i}\boldsymbol{\Sigma}_i$ that

$$
\mathbb{E}\left(\sum_{i=1}^a \frac{N}{n_i}(\boldsymbol{T}_W)_{ii}A_{i,1}\right) = \sum_{i=1}^a \frac{N}{n_i}(\boldsymbol{T}_W)_{ii}\mathbb{E}\left(A_{i,1}\right) = \operatorname{tr}\left(\boldsymbol{T}\boldsymbol{V}_N\right)
$$

and

$$
\begin{aligned}
Var\left(\frac{\sum\limits_{i=1}^a \frac{N}{n_i}(\boldsymbol{T}_W)_{ii}A_{i,1}}{\mathbb{E}\left(\sum\limits_{i=1}^a \frac{N}{n_i}(\boldsymbol{T}_W)_{ii}A_{i,1}\right)}\right) &= \frac{\sum\limits_{i=1}^a \frac{N^2}{n_i^2}(\boldsymbol{T}_W)_{ii}{}^2\; Var(A_{i,1})}{\operatorname{tr}^2\left(\boldsymbol{T}\boldsymbol{V}_N\right)}\\
&\leq \frac{\sum\limits_{i=1}^a \mathcal{O}\left(n_i^{-1}\right)\cdot\mathcal{O}\left(\operatorname{tr}^2\left(\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)\right)}{\operatorname{tr}^2\left(\boldsymbol{T}\boldsymbol{V}_N\right)}\\
&\leq \frac{\mathcal{O}\left(\frac{1}{n_{\min}}\right)\cdot\mathcal{O}\left(\sum\limits_{i=1}^a \operatorname{tr}^2\left(\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)\right)}{\operatorname{tr}^2\left(\boldsymbol{T}\boldsymbol{V}_N\right)}\\
&\leq \frac{\mathcal{O}\left(\frac{1}{n_{\min}}\right)\cdot\mathcal{O}\left(\operatorname{tr}^2\left(\sum\limits_{i=1}^a \frac{N}{n_i}(\boldsymbol{T}_W)_{ii}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)\right)}{\operatorname{tr}^2\left(\boldsymbol{T}\boldsymbol{V}_N\right)}\\
&= \mathcal{O}\left(\frac{1}{n_{\min}}\right).
\end{aligned}
$$

So the conditions for an unbiased and ratio-consistent estimator are fulfilled.

The same steps with a different number of remaining combinations leads to

$$
\begin{aligned}
\mathbb{E}\left(A_{i,3}\right) &= \frac{1}{4\cdot 6\binom{n_i}{4}}\sum_{\substack{\ell_1,\ell_2=1\\ \ell_1>\ell_2}}^{n_i}\sum_{\substack{k_2=1\\ k_2\neq\ell_1,\ell_2}}^{n_i-1}\sum_{\substack{k_1=k_2+1\\ k_1\neq\ell_1,\ell_2}}^{n_i}\mathbb{E}\left(\left[\boldsymbol{Y}_{i,\ell_1,\ell_2}{}^\top\boldsymbol{Y}_{i,k_1,k_2}\right]^2\right)\\
&\overset{\text{A.5}}{=} \frac{1}{4\cdot 6\binom{n_i}{4}}\cdot 6\binom{n_i}{4}\cdot\operatorname{tr}\left(4\cdot\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)^2\right) = \operatorname{tr}\left(\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)^2\right),
\end{aligned}
$$

$$\begin{aligned}
Var\left(A_{i,3}\right) &= \sum_{\substack{\ell_1,\ell_2=1\\ \ell_1>\ell_2}}^{n_i} \sum_{\substack{k_1,k_2=1\ k_1>k_2\\ \ell_2,\ell_1\neq k_1,k_2}}^{n_i} \sum_{\substack{\ell_1',\ell_2'=1\\ \ell_1'>\ell_2'}}^{n_i} \sum_{\substack{k_1',k_2'=1\ k_1'>k_2'\\ \ell_2',\ell_1'\neq k_1',k_2'}}^{n_i} \\
&\quad \times \frac{Cov\left([\boldsymbol{Y}_{i,\ell_1,\ell_2}{}^\top \boldsymbol{Y}_{i,k_1,k_2}]^2 ; [\boldsymbol{Y}_{i,\ell_1',\ell_2'}{}^\top \boldsymbol{Y}_{i,k_1',k_2'}]^2\right)}{4^2\cdot 6^2 \cdot \binom{n_i}{4}^2} \\
&\overset{A.7}{\leq} \frac{6\binom{n_i}{4} - 6\binom{n_i-4}{4}}{4^2\cdot 6\cdot \binom{n_i}{4}}\ Var\left(\left[\boldsymbol{Y}_{i,1,2}{}^\top \boldsymbol{Y}_{i,3,4}\right]^2\right) \\
&\overset{A.5}{=} \frac{\binom{n_i}{4} - \binom{n_i-4}{4}}{16\binom{n_i}{4}} \mathcal{O}\left(\operatorname{tr}^2\left((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^2\right)\right) \\
&= \mathcal{O}\left(n_i^{-1}\right)\cdot \mathcal{O}\left(\operatorname{tr}^2\left((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^2\right)\right), \\
\mathbb{E}\left(A_{i,r,2}\right) &= \frac{1}{4\cdot\binom{n_i}{2}\binom{n_r}{2}} \sum_{\substack{\ell_1,\ell_2=1\\ \ell_1>\ell_2}}^{n_i} \sum_{\substack{k_1,k_2=1\\ k_1>k_2}}^{n_r} \mathbb{E}\left(\left[\boldsymbol{Y}_{i,\ell_1,\ell_2}{}^\top \boldsymbol{Y}_{r,k_1,k_2}\right]^2\right) \\
&\overset{A.5}{=} \frac{1}{4\cdot\binom{n_i}{2}\binom{n_r}{2}} \cdot \binom{n_i}{2}\cdot\binom{n_r}{2}\cdot \operatorname{tr}\left(4\cdot \boldsymbol{\Sigma}_i\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right) \\
&= \operatorname{tr}\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right), \\
Var\left(\frac{2N^2}{n_in_r}A_{i,r,2}\right) &= \frac{4N^4}{n_i^2n_r^2} \sum_{\substack{\ell_1,\ell_2=1\\ \ell_1>\ell_2}}^{n_1} \sum_{\substack{k_1,k_2=1\\ k_1>k_2}}^{n_2} \sum_{\substack{\ell_1',\ell_2'=1\\ \ell_1'>\ell_2'}}^{n_i} \sum_{\substack{k_1',k_2'=1\\ k_1'>k_2'}}^{n_r} \\
&\quad \times \frac{Cov\left(\left[\boldsymbol{Y}_{i,\ell_1,\ell_2}{}^\top \boldsymbol{Y}_{r,k_1,k_2}\right]^2 ; \left[\boldsymbol{Y}_{i,\ell_1',\ell_2'}{}^\top \boldsymbol{Y}_{r,k_1',k_2'}\right]^2\right)}{16\cdot\binom{n_i}{2}^2\binom{n_r}{2}^2} \\
&\overset{A.7}{\leq} \frac{4N^4}{n_i^2n_r^2}\frac{\binom{n_i}{2}\binom{n_r}{2} - \binom{n_i-2}{2}\binom{n_r-2}{2}}{16\cdot\binom{n_i}{2}\binom{n_r}{2}}\ Var\left(\left[\boldsymbol{Y}_{i,1,2}{}^\top \boldsymbol{Y}_{r,1,2}\right]^2\right) \\
&\overset{A.5}{\leq} \frac{\binom{n_i}{2}\binom{n_r}{2} - \binom{n_i-2}{2}\binom{n_r-2}{2}}{\binom{n_i}{2}\binom{n_r}{2}} \cdot \mathcal{O}\left(\operatorname{tr}^2\left(\frac{N}{n_i}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\frac{N}{n_r}\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right)\right) \\
&\leq \mathcal{O}\left(\frac{1}{n_{\min}}\right)\cdot \mathcal{O}\left(\operatorname{tr}^2\left(\frac{N}{n_i}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\frac{N}{n_r}\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right)\right).
\end{aligned}$$

Finally, the conditions for A_4 have to be checked. With the expectation values from above we calculate

$$\begin{aligned}
&\mathbb{E}\left(A_4\right) \\
&= \sum_{i=1}^{a}\frac{N^2}{n_i^2}(\boldsymbol{T}_W)_{ii}{}^2\mathbb{E}(A_{i,3}) + 2\sum_{i=1}^{a-1}\sum_{r=i+1}^{a}\frac{N^2}{n_in_r}(\boldsymbol{T}_W)_{ir}{}^2\mathbb{E}\left(A_{i,r,2}\right) \\
&= \sum_{i=1}^{a}\frac{N^2}{n_i^2}(\boldsymbol{T}_W)_{ii}{}^2\operatorname{tr}\left((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^2\right) + \sum_{i=1}^{a-1}\sum_{r=i+1}^{a}\frac{2N^2}{n_in_r}(\boldsymbol{T}_W)_{ir}{}^2\operatorname{tr}\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right)
\end{aligned}$$

$$= \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right).$$

To calculate the variances the following additional inequalities are needed:

$$\frac{Var\left(\sum_{i=1}^{a}\left(\frac{N}{n_i}\right)^2(\boldsymbol{T}_W)_{ii}{}^2 A_{i,3}\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}$$

$$= \sum_{i=1}^{a}\frac{Var\left(\left(\frac{N}{n_i}\right)^2(\boldsymbol{T}_W)_{ii}{}^2 A_{i,3}\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}$$

$$\leq \sum_{i=1}^{a}\mathcal{O}\left(n_i^{-1}\right)\cdot\frac{\mathcal{O}\left((\boldsymbol{T}_W)_{ii}{}^4\operatorname{tr}^2\left(\left(\boldsymbol{T}_S\frac{N}{n_i}\boldsymbol{\Sigma}_i\right)^2\right)\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}$$

$$\leq \mathcal{O}\left(\frac{1}{n_{\min}}\right)\frac{\mathcal{O}\left(\operatorname{tr}^2\left(\sum_{i=1}^{a}(\boldsymbol{T}_W)_{ii}{}^2\left(\boldsymbol{T}_S\frac{N}{n_i}\boldsymbol{\Sigma}_i\right)^2\right)\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right) = \mathcal{O}\left(\frac{1}{n_{\min}}\right)$$

and

$$\frac{Var\left(2\sum_{r<i\in\mathbb{N}_a}\frac{N^2}{n_i n_r}(\boldsymbol{T}_W)_{ir}{}^2 A_{i,r,2}\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}$$

$$\overset{A.7}{\leq} 4\sum_{i<r\in\mathbb{N}_a}\sum_{h<g\in\mathbb{N}_a}\frac{\sqrt{Var\left(\frac{N^2}{n_i n_r}(\boldsymbol{T}_W)_{ir}A_{i,r,2}\right)}\sqrt{Var\left(\frac{N^2}{n_h n_g}(\boldsymbol{T}_W)_{gh}A_{h,g,2}\right)}}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}$$

$$\leq \left(\sum_{i\neq r\in\mathbb{N}_a}\frac{\sqrt{\mathcal{O}\left(\frac{1}{n_{\min}}\right)}(\boldsymbol{T}_W)_{ir}{}^2\operatorname{tr}\left(\boldsymbol{T}_S\frac{N}{n_i}\boldsymbol{\Sigma}_i\boldsymbol{T}_S\frac{N}{n_r}\boldsymbol{\Sigma}_r\right)}{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right)^2$$

$$\leq \mathcal{O}\left(\frac{1}{n_{\min}}\right)\left(\frac{\mathcal{O}\left(\sum_{i\neq r\in\mathbb{N}_a}(\boldsymbol{T}_W)_{ir}{}^2\operatorname{tr}\left(\boldsymbol{T}_S\frac{N}{n_i}\boldsymbol{\Sigma}_i\boldsymbol{T}_S\frac{N}{n_r}\boldsymbol{\Sigma}_r\right)\right)}{\sum_{i,r\in\mathbb{N}_a}(\boldsymbol{T}_W)_{ir}{}^2\operatorname{tr}\left(\boldsymbol{T}_S\frac{N}{n_i}\boldsymbol{\Sigma}_i\frac{N}{n_r}\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right)}\right)^2 \leq \mathcal{O}\left(\frac{1}{n_{\min}}\right).$$

Together this leads to

$$Var\left(\frac{A_4}{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right)$$

$$\overset{A.7}{\leq} \left[\sqrt{\frac{Var\left(2 \sum_{r<i\in\mathbb{N}_a} \frac{N^2}{n_i n_r} (\boldsymbol{T}_W)_{ir}{}^2 A_{i,r,2}\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}} + \sqrt{\frac{Var\left(\sum_{i=1}^{a} \frac{N}{n_i} (\boldsymbol{T}_W)_{ii}{}^2 A_{i,3}\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}} \right]^2$$

$$= \left[\sqrt{\mathcal{O}\left(\frac{1}{n_{\min}}\right)} + \sqrt{\mathcal{O}\left(\frac{1}{n_{\min}}\right)} \right]^2 = \mathcal{O}\left(\frac{1}{n_{\min}}\right)$$

and therefore A_4 is an unbiased and ratio-consistent estimator of $\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)$. Moreover, we want to stress that the zero sequences used as upper border for $\widehat{\mathbb{E}_{H_0}}(Q_N)$ and A_4 do not depend on the number of groups or dimensions, so this estimators can be also used for increasing number of groups.

With the expectation values and variances from the beginning it follows directly that $A_{i,1}, A_{i,r,2}, A_{i,3}, A_4$ are unbiased, ratio-consistent estimators of $\operatorname{tr}(\boldsymbol{T}_S\boldsymbol{\Sigma}_i)$,$\operatorname{tr}(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\boldsymbol{T}_S\boldsymbol{\Sigma}_r)$, $\operatorname{tr}\left((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^2\right)$ and $\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)$.

It is worth to note that all of this estimators also consistent estimators which are even dimension-stable in the sense of Brunner et al. [8]. □

For $A_{i,r,2}$ there exists an alternative form which can be implemented substantially more efficient and was considered in Brunner et al. [9]. It uses matrices of the form $\widehat{\boldsymbol{M}}_{i,r} = \boldsymbol{P}_{n_i}\left(\boldsymbol{T}_S\boldsymbol{X}_{i,1}, \ldots, \boldsymbol{T}_S\boldsymbol{X}_{i,n_i}\right)^\top \cdot \left(\boldsymbol{T}_S\boldsymbol{X}_{r,1}, \ldots, \boldsymbol{T}_S\boldsymbol{X}_{r,n_r}\right)\boldsymbol{P}_{n_r}^\top$. Recalling that $\mathbf{1}_n$ is the vector of ones and $\#$ denotes the Hadamard-Schur-Product, it can be seen that

$$A_{i,r,2} = \frac{\mathbf{1}_{n_i}{}^\top \left(\widehat{\boldsymbol{M}_{i,r}} \# \widehat{\boldsymbol{M}_{i,r}}\right) \mathbf{1}_{n_r}}{(n_i-1)(n_r-1)}.$$

For $A_{i,3}$ there also exists an alternative formula, which expands much longer, but is more efficient:

$$\begin{aligned}
A_{i,3} = &\sum_{\substack{\ell_1,\ell_2=1\\ \ell_1\neq\ell_2}}^{n_i} \frac{\left[\boldsymbol{X}_{i,\ell_1}{}^\top \boldsymbol{T}_S \boldsymbol{X}_{i,\ell_2}\right]^2}{n_i(n_i-1)} - \sum_{\substack{\ell_1,\ell_2,\ell_3=1\\ \ell_3\neq\ell_1,\ell_2}}^{n_i} \frac{\left[\boldsymbol{X}_{i,\ell_1}^\top \boldsymbol{T}_S \boldsymbol{X}_{i,\ell_3} \boldsymbol{X}_{i,\ell_2}^\top \boldsymbol{T}_S (\boldsymbol{X}_{i,\ell_3} + \boldsymbol{X}_{i,\ell_1})\right]}{n_i(n_i-1)(n_i-2)(n_i-3)} \\
&+ \sum_{\substack{\ell_1,\ell_2,\ell_3=1\\ \ell_1\neq\ell_2\neq\ell_3}}^{n_i} \frac{\left[\boldsymbol{X}_{i,\ell_1}^\top \boldsymbol{T}_S \boldsymbol{X}_{i,\ell_3} \boldsymbol{X}_{i,\ell_2}^\top \boldsymbol{T}_S \boldsymbol{X}_{i,\ell_2}\right] + (2n_i+5)\cdot\left[\boldsymbol{X}_{i,\ell_1}^\top \boldsymbol{T}_S \boldsymbol{X}_{i,\ell_2} \boldsymbol{X}_{i,\ell_1}^\top \boldsymbol{T}_S \boldsymbol{X}_{i,\ell_3}\right]}{n_i(n_i-1)(n_i-2)(n_i-3)} \\
&- \sum_{\substack{\ell_1,\ell_2,\ell_3=1\\ \ell_1\neq\ell_2}}^{n_i} \frac{\left[\boldsymbol{X}_{i,\ell_1}^\top \boldsymbol{T}_S \boldsymbol{X}_{i,\ell_2} \boldsymbol{X}_{i,\ell_2}^\top \boldsymbol{T}_S \boldsymbol{X}_{i,\ell_3}\right]}{n_i(n_i-1)(n_i-2)(n_i-3)} \\
&- \frac{n_i^2\left[\overline{\boldsymbol{X}}_i^\top \boldsymbol{T}_S \overline{\boldsymbol{X}}_i\right]\left(n_i^2 \overline{\boldsymbol{X}}_i^\top \boldsymbol{T}_S \overline{\boldsymbol{X}}_i - \sum_{\ell_1=1}^{n_i}\left[\boldsymbol{X}_{i,\ell_1}^\top \boldsymbol{T}_s \boldsymbol{X}_{i,\ell_1}\right]\right)}{n_i(n_i-1)(n_i-2)(n_i-3)}.
\end{aligned}$$

To finally prove Theorem 3.5 (p.2750) we need another lemma.

Lemma B.1. *For the previously defined estimators it holds for $n_{\min} \to \infty$ that*

$$\frac{\sum\limits_{i=1}^{a} \frac{N}{n_i} (\boldsymbol{T}_W)_{ii} A_{i,1} - \sum\limits_{i=1}^{a} \frac{N}{n_i} (\boldsymbol{T}_W)_{ii} \operatorname{tr}(\boldsymbol{T}_S \boldsymbol{\Sigma}_i)}{\sqrt{2 \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2 \right)}} \xrightarrow{\mathcal{P}} 0 \quad \textit{independent of } d \textit{ or } a.$$

Proof. We know that

$$\mathbb{E}\left(\sum_{i=1}^{a} \frac{\frac{N}{n_i} (\boldsymbol{T}_W)_{ii} ((A_{i,1}) - \operatorname{tr}(\boldsymbol{T}_S \boldsymbol{\Sigma}_i))}{\sqrt{2 \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2 \right)}} \right) = \sum_{i=1}^{a} \frac{\frac{N}{n_i} (\boldsymbol{T}_W)_{ii} \left(\mathbb{E}(A_{i,1}) - \operatorname{tr}(\boldsymbol{T}_S \boldsymbol{\Sigma}_i) \right)}{\sqrt{2 \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2 \right)}} = 0.$$

Thus,

$$\mathit{Var}\left(\sum_{i=1}^{a} \frac{\frac{N}{n_i} (\boldsymbol{T}_W)_{ii} \left(A_{i,1} - \operatorname{tr}(\boldsymbol{T}_S \boldsymbol{\Sigma}_i) \right)}{\sqrt{2 \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2 \right)}} \right) = \frac{\sum\limits_{i=1}^{a} \frac{N^2}{n_i^2} {(\boldsymbol{T}_W)_{ii}}^2 \, \mathit{Var}(A_{i,1})}{2 \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2 \right)} \overset{\text{Proof of } 3.3}{\leq} \mathcal{O}\left(\frac{1}{n_{\min}} \right) \frac{\sum\limits_{i=1}^{a} \frac{N^2}{n_i^2} {(\boldsymbol{T}_W)_{ii}}^2 \operatorname{tr}\left((2\boldsymbol{T}_S \boldsymbol{\Sigma}_i)^2 \right)}{2 \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2 \right)} = \mathcal{O}\left(\frac{1}{n_{\min}} \right).$$

In the last step we used the fact that all terms are non-negative and applied the binomial theorem in the last inequality. It is a zero sequence which only depends on $n_{\min}$, so again with Lemma A.6 (p.2770) the result is proved. □

Proof of Theorem 3.5 (p.2750). From Lemma A.6 it follows independent of a or d for $n_{\min} \to \infty$ that $A_4 / \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2 \right) \xrightarrow{\mathcal{P}} 1$ and therefore $\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2 \right) / A_4 \xrightarrow{\mathcal{P}} 1$. Moreover, it also follows that $\sqrt{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2 \right) / A_4} \xrightarrow{\mathcal{P}} 1$ and with Lemma B.1 we deduce $\frac{\sum_{i=1}^{a} \frac{N}{n_i} (\boldsymbol{T}_W)_{ii} A_{i,1} - \operatorname{tr}(\boldsymbol{T}\boldsymbol{V}_N)}{\sqrt{2 \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2 \right)}} \xrightarrow{\mathcal{P}} 0$.

Thus, we can finally calculate the standardized quadratic form as

$$W_N = \frac{Q_N - \sum_{i=1}^{a} \frac{N}{n_i} (\boldsymbol{T}_W)_{ii} A_{i,1}}{\sqrt{2A_4}}$$

$$= \left(\frac{Q_N - \operatorname{tr}(\boldsymbol{T}\boldsymbol{V}_N)}{\sqrt{2\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}} - \frac{\sum_{i=1}^a \frac{N}{n_i}(\boldsymbol{T}_W)_{ii} A_{i,1} - \operatorname{tr}(\boldsymbol{T}\boldsymbol{V}_N)}{\sqrt{2\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}} \right) \cdot \sqrt{\frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}{A_4}}$$

$$= \left(\frac{Q_N - \operatorname{tr}(\boldsymbol{T}\boldsymbol{V}_N)}{\sqrt{2\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}} - o_{\mathcal{P}}(1) \right) \cdot (1 + o_{\mathcal{P}}(1))$$

$$= \widetilde{W}_N + \widetilde{W}_N \cdot o_{\mathcal{P}}(1) - o_{\mathcal{P}}(1) - o_{\mathcal{P}}(1) \cdot o_{\mathcal{P}}(1).$$

The last two parts converge in probability to zero, so also in distribution and with Slutzky $\widetilde{W}_N \cdot o_{\mathcal{P}}(1)$ converges in distribution to zero if one of the conditions of Theorem 3.1 is fulfilled. Thereby W_N has asymptotical the same distribution as $\widetilde{W}_N$.

Replacing the traces by their estimators in the above calculation, it follows with the same arguments that the asymptotic distribution in both cases of local alternatives does not change, since the estimators are also consistent under the alternative. □

For large numbers of groups many estimators $A_{i,1}, A_{i,r,2}$ and $A_{i,3}$ and have to be calculated which leads to long computation time. In this cases it is better to again use subsamling-type estimators which leads to $A^\star_{i,1}, A^\star_{i,r,2}, A^\star_{i,3}$ and therefore to $A^\star_4$.

Lemma B.2. *With the definitions from above let be*

$$A^\star_{i,1}(B) = \frac{1}{2\cdot B}\sum_{b=1}^{B} \boldsymbol{Y}_{i,\sigma_{i1}(b),\sigma_{i2}(b)}{}^\top \boldsymbol{Y}_{i,\sigma_{i1}(b),\sigma_{i2}(b)},$$

$$A^\star_{i,r,2}(B) = \frac{1}{4\cdot B}\sum_{b=1}^{B} \left[\boldsymbol{Y}_{i,\sigma_{i1}(b),\sigma_{i2}(b)}{}^\top \boldsymbol{Y}_{r,\sigma_{r1}(b),\sigma_{r2}(b)}\right]^2,$$

$$A^\star_{i,3}(B) = \frac{1}{4\cdot B}\sum_{b=1}^{B} \left[\boldsymbol{Y}_{i,\sigma_{i1}(b),\sigma_{i2}(b)}{}^\top \boldsymbol{Y}_{i,\sigma_{i3}(b),\sigma_{i4}(b)}\right]^2,$$

$$A^\star_4(B) = \sum_{i=1}^{a} \frac{N^2}{n_i^2} {(\boldsymbol{T}_W)_{ii}}^2 A^\star_{i,3}(B) + 2\sum_{i=1}^{a}\sum_{r=1, r<i}^{a} \frac{N^2}{n_i n_r} {(\boldsymbol{T}_W)_{ir}}^2 A^\star_{i,r,2}(B).$$

If $B(N)\to\infty$, this estimators and $\sum_{i=1}^a A^\star_{i,1}$ have the same properties as $A_{i,1}, A_{i,r,2}, A_{i,3}$, A_4 and $\sum_{i=1}^a A_{i,1}$ which were defined in Lemma 3.3 (p.2750).

Proof. For $A^\star_{i,1}(B)$, this lemma will be proved in detail. For all other terms only the major steps are shown.

The unbiasedness is clear because the random variables $\sigma_{i1}(b), \sigma_{i2}(b)$ have no influence on the number of terms of the sum and also the terms are identically distributed. Hence,

$$\mathbb{E}\left(A^{\star}_{i,1}(B)\right) = \frac{1}{2\cdot B}\sum_{b=1}^{B}\mathbb{E}\left(\boldsymbol{Y}_{i,\sigma_{i1}(b),\sigma_{i2}(b)}{}^{\top}\boldsymbol{Y}_{i,\sigma_{i1}(b),\sigma_{i2}(b)}\right)$$
$$= \frac{1}{2\cdot B}\sum_{b=1}^{B}\mathbb{E}\left(\boldsymbol{Y}_{i,1,2}{}^{\top}\boldsymbol{Y}_{i,1,2}\right) \overset{\text{A.5}}{=} \operatorname{tr}(\boldsymbol{T}_S\boldsymbol{\Sigma}_i).$$

The second part is more complicated. Let $\mathcal{F}(\boldsymbol{\sigma}_i(B,m))$ be the smallest σ-field which contains $\boldsymbol{\sigma}_i(b,m)\ \forall b \in B$, so obvious $M(B,\boldsymbol{\sigma}_i(b))$ is $\mathcal{F}(\boldsymbol{\sigma}_i(B))$ -measurable. Identical for $\mathcal{F}(\boldsymbol{\sigma}_i(B,m),\boldsymbol{\sigma}_r(B,m))$ and $\mathcal{F}(\boldsymbol{\sigma}(B,m))$. Similar to the previous part, the distribution of the bilinear form does not depend on the index combination. Together with the independence of the normally distributed vectors and $\sigma_{i1}(b),\sigma_{i2}(b)$ this leads to

$$\mathit{Var}\left(\mathbb{E}\left(A^{\star}_{i,1}(B)\middle|\mathcal{F}(\boldsymbol{\sigma}_i(B,2))\right)\right) = \mathit{Var}\left(\operatorname{tr}\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)\right) = 0.$$

With Lemma A.9 (p.2771) we thus obtain

$$\mathit{Var}\left(A^{\star}_{i,1}(B)\right) = 0 + \mathbb{E}\left(\mathit{Var}\left(A^{\star}_{i,1}(B)|\mathcal{F}(\boldsymbol{\sigma}_i(B,2))\right)\right).$$

For the calculation of the conditional variance of the sum, it would be useful finding an upper bound that is based on the variance instead of calculate the covariances. To achieve this, we calculate the number of index combinations which leads to a covariance which is zero. This amount is non-deterministic and we recognize it contains the amount $M(B,\boldsymbol{\sigma}_i(b,2))$ which was considered before.

Again not the amount is important but the number of elements which are contained in $M(B,\boldsymbol{\sigma}_i(b,2))$ since the bilinear forms are identically distributed. Therefore the condition of the variance of the bilinear form disappears since the random indices have no influence on the variance. With the $\mathcal{F}(\boldsymbol{\sigma}_i(B,2))$-measurability of $M(B,\boldsymbol{\sigma}_i(b,2))$ it thus follows that

$$\mathit{Var}\left(A^{\star}_{i,1}(B)\right) = 0 + \mathbb{E}\left(\mathit{Var}\left(A^{\star}_{i,1}(B)|\mathcal{F}(\boldsymbol{\sigma}_i(B,2))\right)\right)$$
$$\overset{\text{A.7}}{\leq} \mathbb{E}\left(\sum_{(j,\ell)\in\mathbb{N}_B^2\setminus M(B,(\boldsymbol{\sigma}_i(b,2)))}\frac{\mathit{Var}\left(\boldsymbol{Y}_{i,\sigma_{i1}(j),\sigma_{i2}(j)}{}^{\top}\boldsymbol{Y}_{i,\sigma_{i1}(j),\sigma_{i2}(j)}\middle|\mathcal{F}(\boldsymbol{\sigma}_i(B,2))\right)}{4B^2}\right)$$
$$= \frac{1}{4B^2}\mathbb{E}\left(\sum_{(j,\ell)\in\mathbb{N}_B^2\setminus M(B,(\boldsymbol{\sigma}_i(b,2)))}\mathit{Var}\left(\boldsymbol{Y}_{i,1,2}{}^{\top}\boldsymbol{Y}_{i,1,2}\right)\right)$$
$$\overset{\text{A.5}}{=} \frac{\mathbb{E}\left(|\mathbb{N}_B^2\setminus M(B,(\boldsymbol{\sigma}_i(b,2)))|\right)}{B^2}\cdot\frac{\mathcal{O}\left(\operatorname{tr}^2\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)\right)}{4}$$
$$\overset{\text{A.10}}{=} \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_i-2}{2}}{\binom{n_i}{2}}\right)\cdot\mathcal{O}\left(\operatorname{tr}^2\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)\right).$$

The other values are calculated in a similar way.

$$\mathbb{E}\left(A^\star_{i,r,2}(B)\right) = \frac{1}{4\cdot B}\sum_{b=1}^{B}\mathbb{E}\left(\left[\boldsymbol{Y}_{i,\sigma_{i1}(b),\sigma_{i2}(b)}{}^\top \boldsymbol{Y}_{r,\sigma_{r1}(b),\sigma_{r2}(b)}\right]^2\right)$$
$$= \frac{1}{4\cdot B}\sum_{b=1}^{B}\mathbb{E}\left(\left[\boldsymbol{Y}_{i,1,2}{}^\top \boldsymbol{Y}_{r,1,2}\right]^2\right) \overset{A.5}{=} \operatorname{tr}(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\boldsymbol{T}_S\boldsymbol{\Sigma}_r).$$

$$Var\left(\mathbb{E}\left(A^\star_{i,r,2}(B)|\mathcal{F}(\boldsymbol{\sigma}_i(B,2),\boldsymbol{\sigma}_r(B,2))\right)\right) = Var\left(\operatorname{tr}\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right)\right) = 0.$$

$$Var\left(A^\star_{i,r,2}(B)\right) = 0 + \mathbb{E}\left(Var\left(A^\star_{i,r,2}(B)|\mathcal{F}(\boldsymbol{\sigma}_i(B),\boldsymbol{\sigma}_r(B,2))\right)\right)$$
$$\le \frac{\mathbb{E}\left(|\mathbb{N}_B^2\setminus M(B,\boldsymbol{\sigma}_i(b,2),\boldsymbol{\sigma}_r(b,2))|\right)}{B^2}\cdot Var\left(\left[\boldsymbol{Y}_{i,1,2}{}^\top \boldsymbol{Y}_{r,1,2}\right]^2\right)$$
$$\overset{A.5}{\le} \frac{\mathbb{E}\left(|\mathbb{N}_B^2\setminus M(B,\boldsymbol{\sigma}_i(b,2),\boldsymbol{\sigma}_r(b,2))|\right)}{B^2}\cdot \mathcal{O}\left(\operatorname{tr}^2\left(\frac{N}{n_i}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\frac{N}{n_r}\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right)\right)$$
$$\overset{A.10}{=} \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_i-2}{2}\cdot\binom{n_r-2}{2}}{\binom{n_i}{2}\cdot\binom{n_r}{2}}\right)\cdot \mathcal{O}\left(\operatorname{tr}^2\left(\frac{N}{n_i}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\frac{N}{n_r}\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right)\right)$$
$$\le \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_{\min}-2}{2}^2}{\binom{n_{\min}}{2}^2}\right)\cdot \mathcal{O}\left(\operatorname{tr}^2\left(\frac{N}{n_i}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\frac{N}{n_r}\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right)\right).$$

$$\mathbb{E}\left(A^\star_{i,3}(B)\right) = \frac{1}{4\cdot B}\sum_{b=1}^{B}\mathbb{E}\left(\left[\boldsymbol{Y}_{i,\sigma_{i1}(b),\sigma_{i2}(b)}{}^\top \boldsymbol{Y}_{i,\sigma_{i3}(b),\sigma_{i4}(b)}\right]^2\right)$$
$$= \frac{1}{4\cdot B}\sum_{b=1}^{B}\mathbb{E}\left(\left[\boldsymbol{Y}_{i,1,2}{}^\top \boldsymbol{Y}_{i,1,2}\right]^2\right) \overset{A.5}{=} \operatorname{tr}\left((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^2\right).$$
$$Var\left(\mathbb{E}\left(A^\star_{i,3}(B)|\mathcal{F}(\boldsymbol{\sigma}_i(B,4))\right)\right) = Var\left(\operatorname{tr}\left((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^2\right)\right) = 0.$$

$$Var\left(A^\star_{i,3}(B)\right)$$
$$= 0 + \mathbb{E}\left(Var\left(A^\star_{i,3}(B)|\mathcal{F}(\boldsymbol{\sigma}_i(B,4))\right)\right)$$
$$\overset{A.7}{\le} \mathbb{E}\left(\sum_{(j,\ell)\in\mathbb{N}_B^2\setminus M(B,\boldsymbol{\sigma}_i(b,4))}\frac{Var\left(\left[\boldsymbol{Y}_{i,\sigma_{i1}(j),\sigma_{i2}(j)}{}^\top \boldsymbol{Y}_{i,\sigma_{i3}(j),\sigma_{i4}(j)}\right]^2\Big|\mathcal{F}(\boldsymbol{\sigma}_i(B,4))\right)}{16B^2}\right)$$
$$\overset{A.5}{\le} \frac{\mathbb{E}\left(|\mathbb{N}_B^2\setminus M(B,\boldsymbol{\sigma}_i(b,4))|\right)}{B^2}\cdot\frac{\mathcal{O}\left(\operatorname{tr}^2\left((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^2\right)\right)}{16}$$
$$\overset{A.10}{=} \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_i-4}{4}}{\binom{n_i}{4}}\right)\cdot\mathcal{O}\left(\operatorname{tr}^2\left((\boldsymbol{T}_S\boldsymbol{\Sigma}_i)^2\right)\right).$$

$$\mathbb{E}\left(\sum_{i=1}^{a}\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}A^\star_{i,1}\right) = \sum_{i=1}^{a}\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}\mathbb{E}\left(A^\star_{i,1}\right) = \sum_{i=1}^{a}\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}\operatorname{tr}\left(\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right).$$

$$
\begin{aligned}
&Var\left(\frac{\sum\limits_{i=1}^{a}\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}A^\star_{i,1}}{\operatorname{tr}\left(\boldsymbol{T}\boldsymbol{V}_N\right)}\right)\\
&=\frac{\sum\limits_{i=1}^{a}\frac{N^2}{n_i^2}{(\boldsymbol{T}_W)_{ii}}^2\ Var\left(A^\star_{i,1}\right)}{\operatorname{tr}^2\left(\boldsymbol{T}\boldsymbol{V}_N\right)}\\
&=\frac{\sum\limits_{i=1}^{a}{(\boldsymbol{T}_W)_{ii}}^2\left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_i-2}{2}}{\binom{n_i}{2}}\right)\cdot\mathcal{O}\left(\operatorname{tr}^2\left(\boldsymbol{T}_S\frac{N}{n_i}\boldsymbol{\Sigma}_i\right)\right)}{\operatorname{tr}^2\left(\boldsymbol{T}\boldsymbol{V}_N\right)}\\
&\le\frac{\sum\limits_{i=1}^{a}{(\boldsymbol{T}_W)_{ii}}^2\left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_{\min}-2}{2}}{\binom{n_{\min}}{2}}\right)\cdot\mathcal{O}\left(\operatorname{tr}^2\left(\boldsymbol{T}_S\frac{N}{n_i}\boldsymbol{\Sigma}_i\right)\right)}{\operatorname{tr}^2\left(\boldsymbol{T}\boldsymbol{V}_N\right)}\\
&\le\left(1-\frac{\left(1-\frac{1}{B}\right)\cdot\binom{n_{\min}-2}{2}}{\binom{n_{\min}}{2}}\right)\cdot\frac{\mathcal{O}\left(\operatorname{tr}^2\left(\sum\limits_{i=1}^{a}\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)\right)}{\operatorname{tr}^2\left(\boldsymbol{T}\boldsymbol{V}_N\right)}\\
&=\left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_{\min}-2}{2}}{\binom{n_{\min}}{2}}\right)\cdot\mathcal{O}\left(1\right).
\end{aligned}
$$

For $B(N)\to\infty$ the first factor is a zero sequence and therefore $\sum\limits_{i=1}^{a}\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}A^\star_{i,1}$ a ratio-consistent, unbiased estimator of $\operatorname{tr}\left(\boldsymbol{T}\boldsymbol{V}_N\right)$.

$$
\begin{aligned}
&\mathbb{E}\left(\sum_{i=1}^{a}\frac{N^2}{n_i^2}{(\boldsymbol{T}_W)_{ii}}^2A^\star_{i,3}+\sum_{i\neq r\in\mathbb{N}_a}\frac{N^2}{n_in_r}{(\boldsymbol{T}_W)_{ir}}^2A^\star_{i,r,2}\right)\\
&=\sum_{i=1}^{a}\frac{N^2}{n_i^2}{(\boldsymbol{T}_W)_{ii}}^2\mathbb{E}\left(A^\star_{i,3}\right)+\sum_{i\neq r\in\mathbb{N}_a}\frac{N^2}{n_in_r}{(\boldsymbol{T}_W)_{ir}}^2\mathbb{E}\left(A^\star_{i,r,2}\right)=\operatorname{tr}\left(\left(\boldsymbol{T}\boldsymbol{V}_N\right)^2\right).
\end{aligned}
$$

$$
\begin{aligned}
&Var\left(\frac{\sum\limits_{i=1}^{a}\frac{N^2}{n_i^2}{(\boldsymbol{T}_W)_{ii}}^2A^\star_{i,3}}{\operatorname{tr}\left(\left(\boldsymbol{T}\boldsymbol{V}_N\right)^2\right)}\right)\\
&=\frac{\sum\limits_{i=1}^{a}Var\left(\frac{N^2}{n_i^2}{(\boldsymbol{T}_W)_{ii}}^2A^\star_{i,3}\right)}{\operatorname{tr}^2\left(\left(\boldsymbol{T}\boldsymbol{V}_N\right)^2\right)}\\
&\le\frac{\sum\limits_{i=1}^{a}{(\boldsymbol{T}_W)_{ii}}^4\left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_i-4}{4}}{\binom{n_i}{4}}\right)\cdot\mathcal{O}\left(\operatorname{tr}^2\left(\left(\boldsymbol{T}_S\frac{N}{n_i}\boldsymbol{\Sigma}_i\right)^2\right)\right)}{\operatorname{tr}^2\left(\left(\boldsymbol{T}\boldsymbol{V}_N\right)^2\right)}
\end{aligned}
$$

$$\leq \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_{\min}-4}{4}}{\binom{n_{\min}}{4}}\right)\cdot\frac{\sum_{i=1}^{a}(\boldsymbol{T}_W)_{ii}{}^4\mathcal{O}\left(\operatorname{tr}^2\left(\left(\boldsymbol{T}_S\frac{N}{n_i}\boldsymbol{\Sigma}_i\right)^2\right)\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}$$

$$\leq \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_{\min}-4}{4}}{\binom{n_{\min}}{4}}\right)\cdot\frac{\mathcal{O}\left(\operatorname{tr}^2\left(\left(\sum_{i=1}^{a}\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\right)^2\right)\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}$$

$$\leq \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_{\min}-4}{4}}{\binom{n_{\min}}{4}}\right)\cdot\mathcal{O}\left(1\right).$$

$$Var\left(\frac{\sum_{i\neq r\in\mathbb{N}_a}\frac{N^2}{n_i n_r}(\boldsymbol{T}_W)_{ir}{}^2A^{\star}_{i,r,2}}{\operatorname{tr}\left(\boldsymbol{T}\boldsymbol{V}_N\right)}\right)$$

$$\leq \left(\sum_{i\neq r\in\mathbb{N}_a}\frac{\sqrt{Var\left(\frac{N^2}{n_i n_j}(\boldsymbol{T}_W)_{ir}{}^2A^{\star}_{i,r,2}\right)}}{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right)^2$$

$$\leq \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_{\min}-2}{2}^2}{\binom{n_{\min}}{2}^2}\right)\cdot\left(\frac{\sum_{i\neq r\in\mathbb{N}_a}(\boldsymbol{T}_W)_{ir}{}^2\sqrt{\mathcal{O}\left(\operatorname{tr}^2\left(\frac{N}{n_i}\boldsymbol{T}_S\boldsymbol{\Sigma}_i\frac{N}{n_r}\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right)\right)}}{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right)^2$$

$$\leq \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_{\min}-2}{2}^2}{\binom{n_{\min}}{2}^2}\right)\cdot\left(\frac{\sum_{i\neq r\in\mathbb{N}_a}\mathcal{O}\left((\boldsymbol{T}_W)_{ir}{}^2\operatorname{tr}\left(\boldsymbol{T}_S\frac{N}{n_i}\boldsymbol{\Sigma}_i\boldsymbol{T}_S\frac{N}{n_r}\boldsymbol{\Sigma}_r\right)\right)}{\sum_{i,r\in\mathbb{N}_a}(\boldsymbol{T}_W)_{ir}{}^2\operatorname{tr}\left(\boldsymbol{T}_S\frac{N}{n_i}\boldsymbol{\Sigma}_i\frac{N}{n_r}\boldsymbol{T}_S\boldsymbol{\Sigma}_r\right)}\right)^2$$

$$\leq \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_{\min}-2}{2}^2}{\binom{n_{\min}}{2}^2}\right)\cdot\mathcal{O}(1).$$

$$Var\left(\frac{\sum_{i=1}^{a}\frac{N^2}{n_i^2}(\boldsymbol{T}_W)_{ii}{}^2A^{\star}_{i,3}+\sum_{i\neq r\in\mathbb{N}_a}\frac{N^2}{n_i n_r}(\boldsymbol{T}_W)_{ir}{}^2A^{\star}_{i,r,2}}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right)$$

$$\overset{\text{A.7}}{\leq}\left[\sqrt{\frac{Var\left(2\sum_{r<i\in\mathbb{N}_a}\frac{N^2}{n_i n_r}(\boldsymbol{T}_W)_{ir}{}^2A^{\star}_{i,r,2}\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}}+\sqrt{\frac{Var\left(\sum_{i=1}^{a}\frac{N}{n_i}(\boldsymbol{T}_W)_{ii}{}^2A^{\star}_{i,3}\right)}{\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}}\right]^2$$

$$\leq \left(1-\left(1-\frac{1}{B}\right)\cdot\frac{\binom{n_{\min}-2}{2}^2}{\binom{n_{\min}}{2}^2}\right)\cdot\mathcal{O}(1).$$

So again this is a zero sequence, and $A_4^\star$ is an unbiased and dimensional stable (i.e. also ratio consistent) estimator of $\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)$. ☐

Appendix C: Proofs of Section 4

Lemma C.1. *For*

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

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

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

we define

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

With this notation it follows that

$$\mathbb{E}(C_5) = \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right), \quad Var(C_5) \le \frac{\left(\prod\limits_{i=1}^a \binom{n_i}{6} - \prod\limits_{i=1}^a \binom{n_i-6}{6}\right)}{\prod\limits_{i=1}^a \binom{n_i}{6}} \cdot 27\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right).$$

Proof. Set

$$\widetilde{\boldsymbol{Z}}_{(\ell_{3,1},\ell_{4,1},\dots,\ell_{3,a},\ell_{4,a})} := \left(\sqrt{2}\boldsymbol{V}_N^{1/2}\right)^{-1} \boldsymbol{Z}_{(\ell_{3,1},\ell_{4,1},\dots,\ell_{3,a},\ell_{4,a})} \sim \mathcal{N}_{ad}(\boldsymbol{0}_{ad}, \boldsymbol{I}_{ad}).$$

It then follows that

$$\begin{aligned}
&\mathbb{E}\left(\boldsymbol{T}\boldsymbol{Z}_{(\ell_{3,1},\ell_{4,1},\dots,\ell_{3,a},\ell_{4,a})} \cdot \boldsymbol{Z}_{(\ell_{3,1},\ell_{4,1},\dots,\ell_{3,a},\ell_{4,a})}{}^\top \boldsymbol{T}^\top\right)\\
&= \mathbb{E}\left(\left(\sqrt{2}\boldsymbol{T}\boldsymbol{V}_N^{1/2}\widetilde{Z}_{(\ell_{3,1},\ell_{4,1},\dots,\ell_{3,a},\ell_{4,a})}\right)\left(\sqrt{2}\boldsymbol{T}\boldsymbol{V}_N^{1/2}\widetilde{\boldsymbol{Z}}_{(\ell_{3,1},\ell_{4,1},\dots,\ell_{3,a},\ell_{4,a})}\right)^\top\right)\\
&= 2\boldsymbol{T}\boldsymbol{V}_N^{1/2}\mathbb{E}\left(\widetilde{\boldsymbol{Z}}_{(\ell_{3,1},\ell_{4,1},\dots,\ell_{3,a},\ell_{4,a})}\widetilde{\boldsymbol{Z}}^\top_{(\ell_{3,1},\ell_{4,1},\dots,\ell_{3,a},\ell_{4,a})}\right)\boldsymbol{V}_N^{1/2\,\top}\boldsymbol{T}\\
&= 2\boldsymbol{T}\boldsymbol{V}_N^{1/2}\boldsymbol{I}_{ad}\boldsymbol{V}_N^{1/2\,\top}\boldsymbol{T} = 2\boldsymbol{T}\boldsymbol{V}_N\boldsymbol{T}.
\end{aligned}$$

With the rules for conditional expectation and the involved independence it follows that

$$\mathbb{E}(C_5) = \sum_{\substack{\ell_{1,1},\dots,\ell_{6,1}=1\\ \ell_{1,1}\neq\cdots\neq\ell_{6,1}}}^{n_1} \cdots \sum_{\substack{\ell_{1,a},\dots,\ell_{6,a}=1\\ \ell_{1,a}\neq\cdots\neq\ell_{6,a}}}^{n_a}$$

$$\times \frac{\mathbb{E}\left(\Lambda_1(\ell_{1,1},\ldots,\ell_{6,a})\cdot\Lambda_2(\ell_{1,1},\ldots,\ell_{6,a})\cdot\Lambda_3(\ell_{1,1},\ldots,\ell_{6,a})\right)}{8\cdot\prod_{i=1}^{a}\frac{n_i!}{(n_i-6)!}}$$

$$= \sum_{\substack{\ell_{1,1},\ldots,\ell_{6,1}=1\\ \ell_{1,1}\neq\cdots\neq\ell_{6,1}}}^{n_1}\cdots\sum_{\substack{\ell_{1,a},\ldots,\ell_{6,a}=1\\ \ell_{1,a}\neq\cdots\neq\ell_{6,a}}}^{n_a}$$

$$\times \frac{\mathbb{E}\left(\boldsymbol{Z}_{(1,2)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\cdot\boldsymbol{Z}_{(3,4)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(5,6)}\cdot\boldsymbol{Z}_{(5,6)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(1,2)}\right)}{8\cdot\prod_{i=1}^{a}\frac{n_i!}{(n_i-6)!}}$$

$$= \frac{1}{8}\mathbb{E}\left(\boldsymbol{Z}_{(1,2)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\cdot\boldsymbol{Z}_{(3,4)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(5,6)}\cdot\boldsymbol{Z}_{(5,6)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(1,2)}\right)$$

$$= \frac{1}{8}\mathbb{E}\left(\mathbb{E}\left(\boldsymbol{Z}_{(1,2)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\cdot\boldsymbol{Z}_{(3,4)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(5,6)}\cdot\boldsymbol{Z}_{(5,6)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(1,2)}\;\middle|\;\boldsymbol{Z}_{(1,2)}\right)\right)$$

$$= \frac{1}{8}\mathbb{E}\left(\boldsymbol{Z}_{(1,2)}{}^{\top}\mathbb{E}\left(\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\cdot\boldsymbol{Z}_{(3,4)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(5,6)}\cdot\boldsymbol{Z}_{(5,6)}{}^{\top}\boldsymbol{T}\right)\boldsymbol{Z}_{(1,2)}\right)$$

$$= \frac{4}{8}\mathbb{E}\left(\boldsymbol{Z}_{(1,2)}{}^{\top}\boldsymbol{T}\boldsymbol{V}_N\boldsymbol{T}\boldsymbol{T}\boldsymbol{V}_N\boldsymbol{T}\boldsymbol{Z}_{(1,2)}\right)$$

$$= \frac{1}{2}\operatorname{tr}((\boldsymbol{T}\boldsymbol{V}_N\boldsymbol{T}\boldsymbol{T}\boldsymbol{V}_N\boldsymbol{T})2\boldsymbol{V}_N) = \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right).$$

Due to the fact that all $\boldsymbol{X}_{i,j}$ are identically distributed we can neglect the concrete indices, as long as we maintain the structure of dependence of the bilinear forms. The last term fulfills the requirements from Korollar A.5 (p.2769) with $\boldsymbol{Z}_{(1,2)} \sim \mathcal{N}(\boldsymbol{0}_{ad}, 2\boldsymbol{V}_N)$ and the matrix $\boldsymbol{T}\boldsymbol{V}_N\boldsymbol{T}\boldsymbol{T}\boldsymbol{V}_N\boldsymbol{T}$.

For the calculation of the variance it is useful to diagonalize the matrix ${\boldsymbol{V}_N^{1/2}}^{\top}\boldsymbol{T}\boldsymbol{V}_N^{1/2}$: There exists an orthogonal matrix $\boldsymbol{P}$ with $\boldsymbol{P}{\boldsymbol{V}_N^{1/2}}^{\top}\boldsymbol{T}\boldsymbol{V}_N^{1/2}\boldsymbol{P}^{\top} = \boldsymbol{D} = diag(\lambda_1,\ldots,\lambda_{ad})$, where λ_i are the eigenvalues of ${\boldsymbol{V}_N^{1/2}}^{\top}\boldsymbol{T}\boldsymbol{V}_N^{1/2}$. We define $\boldsymbol{J}_i := \boldsymbol{P}\widetilde{\boldsymbol{Z}}_{(i,j)}$ so with the properties of the standard normal distribution $\boldsymbol{J}_i \sim \mathcal{N}_{ad}(\boldsymbol{0}_{ad}, \boldsymbol{I}_{ad})$, where the $\boldsymbol{J}_i$ are independent for different indices. Thus, we can rewrite

$$\begin{aligned}\boldsymbol{Z}_{(1,2)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(3,4)} &= \widetilde{\boldsymbol{Z}}_{(1,2)}^{\top}2{\boldsymbol{V}_N^{1/2}}^{\top}\boldsymbol{T}\boldsymbol{V}_N^{1/2}\widetilde{\boldsymbol{Z}}_{(3,4)}\\ &= 2\widetilde{\boldsymbol{Z}}_{(1,2)}^{\top}\boldsymbol{P}^{\top}\boldsymbol{D}\boldsymbol{P}\widetilde{\boldsymbol{Z}}_{(3,4)} = 2\boldsymbol{J}_1^{\top}\boldsymbol{D}\boldsymbol{J}_3.\end{aligned}$$

With this argument for all three random variables it follows for the second moment that

$$\mathbb{E}\left(\left[\boldsymbol{J}_1^{\top}\boldsymbol{D}\boldsymbol{J}_3\boldsymbol{J}_3^{\top}\boldsymbol{D}\boldsymbol{J}_5\boldsymbol{J}_5^{\top}\boldsymbol{D}\boldsymbol{J}_1\right]^2\right)$$

$$= \mathbb{E}\left(\left[\sum_{i=1}^{ad}\lambda_i J_{1i}J_{3i}\right]^2\left[\sum_{j=1}^{ad}\lambda_j J_{3j}J_{5j}\right]^2\left[\sum_{\ell=1}^{ad}\lambda_\ell J_{5\ell}J_{1\ell}\right]^2\right)$$

$$= \sum_{\substack{i_1,j_1,\ell_1=1\\ i_2,j_2,\ell_2=1}}^{ad} \lambda_{i_1}\lambda_{i_2}\lambda_{j_1}\lambda_{j_2}\lambda_{\ell_1}\lambda_{\ell_2}$$
$$\times \mathbb{E}\left(J_{1i_1}J_{3i_1}J_{1i_2}J_{3i_2}J_{3j_1}J_{5j_1}J_{3j_2}J_{5j_2}J_{5\ell_1}J_{1\ell_1}J_{5\ell_2}J_{1\ell_2}\right).$$

Now we consider the expectation value for the different combinations. If all indices are equal, it is given by

$$\mathbb{E}\left(J_{11}^4 J_{31}^4 J_{51}^4\right) = 3^3 = 27.$$

Moreover, for $i_1 = i_2 \neq \ell_1 = \ell_2$ and $\ell_2 \neq j_1 = j_2 \neq i_1$ it holds that

$$\mathbb{E}\left(J_{11}^2 J_{31}^2 J_{32}^2 J_{52}^2 J_{13}^2 J_{53}^2\right) = 1^6 = 1.$$

Next, the case $i_1 = i_2 = j_1 = j_2 \neq \ell_1 = \ell_2$ is considered (noting this result can also be used for both analogue combinations):

$$\mathbb{E}\left(J_{11}^2 J_{31}^4 J_{51}^2 J_{12}^2 J_{52}^2\right) = 3^1 \cdot 1^4 = 3.$$

Finally, we consider the combination $i_1 = j_1 = \ell_1 \neq i_2 = j_2 = \ell_2$ and obtain

$$\mathbb{E}\left(\left[J_{11}J_{31}J_{12}J_{32}J_{51}J_{52}\right]^2\right) = \prod_{i=1}^{2} \mathbb{E}\left(J_{1i}^2\right)\mathbb{E}\left(J_{3i}^2\right)\mathbb{E}\left(J_{5i}^2\right) = 1^{3^2}.$$

This is also true for $i_1 = j_2 = \ell_1 \neq i_2 = j_1 = \ell_2$ and the analogue combinations, so, all in all, we have 4 combinations of this kind. All other index combinations lead to expectation zero because in this combinations at least one index appears just one time in the product. Thus, due to independence and the fact that all random variables J_i are centered, it follows that

$$\begin{aligned}
&\mathbb{E}\left(\left[\boldsymbol{J}_1^\top \boldsymbol{D}\boldsymbol{J}_3\boldsymbol{J}_3^\top \boldsymbol{D}\boldsymbol{J}_5\boldsymbol{J}_5^\top \boldsymbol{D}\boldsymbol{J}_1\right]^2\right)\\
&= \sum_{i=1}^{ad}\lambda_i^6 \cdot 27 + \sum_{\substack{i,j=1\\ i\neq j}}^{ad}\lambda_i^3\lambda_j^3 \cdot 1 \cdot 4 + \sum_{\substack{i,j=1\\ i\neq j}}^{d}\lambda_i^2\lambda_j^4 \cdot 9 + \sum_{\substack{i,j,\ell=1\\ i\neq j\neq \ell}}^{ad}\lambda_i^2\lambda_j^2\lambda_\ell^2\\
&= 23\sum_{i=1}^{ad}\lambda_i^6 + 4\left(\sum_{\substack{i,j=1\\ i\neq j}}^{ad}\lambda_i^3\lambda_j^3 + \sum_{i=j=1}^{ad}\lambda_i^3\lambda_j^3\right) + 9\sum_{\substack{i,j=1\\ i\neq j}}^{ad}\lambda_i^2\lambda_j^4 + \sum_{\substack{i,j,\ell=1\\ i\neq j\neq \ell}}^{ad}\lambda_i^2\lambda_j^2\lambda_\ell^2\\
&= 17\sum_{i=1}^{ad}\lambda_i^6 + 4\sum_{i,j=1}^{ad}\lambda_i^3\lambda_j^3 + 3\sum_{\substack{i,j=1\\ i\neq j}}^{ad}\lambda_i^2\lambda_j^4\\
&\quad + 6\left(\sum_{\substack{i,j=1\\ i\neq j}}^{ad}\lambda_i^2\lambda_j^4 + \sum_{i=1}^{ad}\lambda_i^6\right) + \sum_{\substack{i,j,\ell=1\\ i\neq j\neq \ell}}^{ad}\lambda_i^2\lambda_j^2\lambda_\ell^2
\end{aligned}$$

$$
\begin{aligned}
&= 17\sum_{i=1}^{ad}\lambda_i^6 + 4\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right) + 3\sum_{\substack{i,j=1\\ i\neq j}}^{ad}\lambda_i^2\lambda_j^4 + 6\sum_{i,j=1}^{ad}\lambda_i^2\lambda_j^4 + \sum_{\substack{i,j,\ell=1\\ i\neq j\neq \ell}}^{ad}\lambda_i^2\lambda_j^2\lambda_\ell^2\\
&\overset{A.2}{\leq} 21\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right) + 3\sum_{\substack{i,j=1\\ i\neq j}}^{ad}\lambda_i^2\lambda_j^4 + 6\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right)\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)\\
&\quad + \sum_{\substack{i,j,\ell=1\\ i\neq j\neq \ell}}^{ad}\lambda_i^2\lambda_j^2\lambda_\ell^2\\
&\overset{A.2}{\leq} 21\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right) + 3\sum_{\substack{i,j=1\\ i\neq j}}^{ad}\lambda_i^2\lambda_j^4 + 6\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right) + \sum_{\substack{i,j,\ell=1\\ i\neq j\neq \ell}}^{ad}\lambda_i^2\lambda_j^2\lambda_\ell^2\\
&\overset{A.2}{\leq} 20\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right) + 6\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)\\
&\quad + \left(\sum_{\substack{i,j,\ell=1\\ i\neq j\neq \ell}}^{ad}\lambda_i^2\lambda_j^2\lambda_\ell^2 + 3\sum_{\substack{i,j=1\\ i\neq j}}^{ad}\lambda_i^2\lambda_j^4 + \sum_{i=1}^{ad}\lambda_i^6\right)\\
&= 20\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right) + 7\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)\\
&\overset{A.2}{\leq} 20\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right)\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right) + 7\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)\\
&\overset{A.2}{\leq} 27\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right).
\end{aligned}
$$

So we can control the variance by

$$
\begin{aligned}
&Var(C_5)\\
&\overset{A.7}{\leq} \frac{Var\left(\Lambda_1(1,2,3,4,5,6,\dots,5,6)\cdot\Lambda_2(1,2,3,4,5,6,\dots,5,6)\cdot\Lambda_3(1,2,3,4,5,6,\dots,5,6)\right)}{64\cdot\prod_{i=1}^{a}\binom{n_i}{6}\cdot\left(\prod_{i=1}^{a}\binom{n_i}{6}-\prod_{i=1}^{a}\binom{n_i-6}{6}\right)^{-1}}\\
&\leq \frac{\mathbb{E}\left(\left[\Lambda_1(1,2,3,4,5,6,\dots,5,6)\cdot\Lambda_2(1,2,3,4,5,6,\dots,5,6)\cdot\Lambda_3(1,2,3,4,5,6,\dots,5,6)\right]^2\right)}{64\cdot\prod_{i=1}^{a}\binom{n_i}{6}\cdot\left(\prod_{i=1}^{a}\binom{n_i}{6}-\prod_{i=1}^{a}\binom{n_i-6}{6}\right)^{-1}}\\
&\qquad\qquad = \frac{\mathbb{E}\left(\left[2^3\cdot\boldsymbol{J}_1^\top\boldsymbol{D}\boldsymbol{J}_3\boldsymbol{J}_3^\top\boldsymbol{D}\boldsymbol{J}_5\boldsymbol{J}_5^\top\boldsymbol{D}\boldsymbol{J}_1\right]^2\right)}{64\cdot\prod_{i=1}^{a}\binom{n_i}{6}\cdot\left(\prod_{i=1}^{a}\binom{n_i}{6}-\prod_{i=1}^{a}\binom{n_i-6}{6}\right)^{-1}}\\
&\qquad\qquad \leq \frac{\left(\prod_{i=1}^{a}\binom{n_i}{6}-\prod_{i=1}^{a}\binom{n_i-6}{6}\right)}{\prod_{i=1}^{a}\binom{n_i}{6}}\cdot 27\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right). \qquad \square
\end{aligned}
$$

With this result, we can construct an estimator for τ_P step by step:

Lemma C.2. *For C_5 as previously defined, it holds for fixed a that*

$$\frac{C_5}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} - \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} \xrightarrow{\mathcal{P}} 0 \qquad \min(d, n_{\min}) \to \infty.$$

It even holds in the asymptotic frameworks (4)–(5) if $q > 1$ exists with $n_{\min} = \mathcal{O}(a^q)$.

Proof. From the previous lemma, we know that

$$\begin{aligned}
&\mathbb{E}\left(\frac{C_5}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} - \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right)\\
&\quad = \mathbb{E}\left(\frac{C_5}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right) - \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} = 0,\\
&\mathit{Var}\left(\frac{C_5}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} - \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right)\\
&\quad = \frac{\mathit{Var}(C_5)}{\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} \overset{\text{C.1}}{\leq} 27 \cdot \frac{\left(\prod\limits_{i=1}^{a}\binom{n_i}{6} - \prod\limits_{i=1}^{a}\binom{n_i-6}{6}\right)}{\prod\limits_{i=1}^{a}\binom{n_i}{6}}.
\end{aligned}$$

For fixed a this is a zero sequence. If we consider $a \to \infty$ we need the existence of $q > 1$ and $n_{\min} = \mathcal{O}(a^q)$ to guarantee that the upper border is a zero sequence. So in both cases Lemma A.6 (p.2770) can be used. □

Lemma C.3. *Moreover C_5 holds for fixed a*

$$\frac{C_5^2}{\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} - \tau_P \xrightarrow{\mathcal{P}} 0 \qquad d, n_{\min} \to \infty.$$

If $q > 1$ exists with $n_{\min} = \mathcal{O}(a^q)$, the convergence even holds in the asymptotic frameworks (4)–(5).

Proof. With the last lemma it follows for both cases that

$$\begin{aligned}
\frac{C_5^2}{\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} - \tau_P &= \left(\frac{C_5}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right)^2 - \left(\frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right)^2\\
&= \left[\frac{C_5}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} - \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}\right]
\end{aligned}$$

$$\times \left[\frac{C_5}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} + \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} \right]$$
$$= \mathcal{O}_{\mathcal{P}}(1) \cdot \left[\frac{C_5}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} - \frac{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} + 2\sqrt{\tau_P} \right]$$
$$= \mathcal{O}_{\mathcal{P}}(1) \cdot \left[\mathcal{O}_{\mathcal{P}}(1) + 2\sqrt{\tau_P}\right] = \mathcal{O}_{\mathcal{P}}(1).$$

For the last step we used that $\tau_P \in [0,1]$ which is known from Lemma A.8 (p.2770) and hence $\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right) \big/ \operatorname{tr}^{3/2}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right) = \sqrt{\tau_P} \in [-1,1]$. As a product of a bound term and a term which converges to zero in probability, it also converges to zero in probability and with Slutzky's Lemma the result follows. □

Proof of Lemma 4.2. From Lemma 3.3 (p.2750) together with Lemma A.6 (p.2770) it follows

$$\frac{A_4}{\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} \xrightarrow{\mathcal{P}} 1 \quad \text{and therefore} \quad \frac{\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}{A_4^3} \xrightarrow{\mathcal{P}} 1 \quad \text{for} \quad n_{\min} \to \infty,$$

independent of d or a. With Lemma C.3 (p.2791) it follows

$$\frac{C_5^2}{\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} - \tau_P \xrightarrow{\mathcal{P}} 0 \qquad \text{for} \qquad d, n_{\min} \to \infty$$

or under the additional condition also in the asymptotic frameworks (4)–(5).

With these limits in both cases we can calculate

$$\frac{C_5^2}{A_4^3} - \tau_P = \frac{C_5^2}{\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} \cdot \frac{\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)}{A_4^3} - \tau_P$$
$$= \frac{C_5^2}{\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} \cdot (1 + \mathcal{O}_{\mathcal{P}}(1)) - \tau_P$$
$$= \frac{C_5^2}{\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} - \tau_P + \left(\frac{C_5^2}{\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)} - \tau_P + \tau_P \right) \cdot \mathcal{O}_{\mathcal{P}}(1)$$
$$= \mathcal{O}_{\mathcal{P}}(1) + \mathcal{O}_{\mathcal{P}}(1) \cdot \mathcal{O}_{\mathcal{P}}(1) + \tau_P \cdot \mathcal{O}_{\mathcal{P}}(1) = \mathcal{O}_{\mathcal{P}}(1).$$

As in the previous lemma we used $\tau_P \in [0,1]$ and Slutzky. □

For $C_5^\star$ the properties are shown in a similar way as in Lemma B.2 (p.2782).

Lemma C.4. *For*

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

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

define

$$C_5^{\star}(B) = \frac{1}{8\cdot B}\sum_{b=1}^{B}\Lambda_1(\boldsymbol{\sigma}(b,6))\cdot\Lambda_2(\boldsymbol{\sigma}(b,6))\cdot\Lambda_3(\boldsymbol{\sigma}(b,6)).$$

Then it holds

$$\mathbb{E}\left(C_5^{\star}(B)\right) = \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right),$$

$$Var\left(C_5^{\star}(B)\right) \leq \left(1-\left(1-\frac{1}{B}\right)\cdot\prod_{i=1}^{a}\frac{\binom{n_i-6}{6}}{\binom{n_i}{6}}\right)\cdot 27\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right).$$

Proof. With the same steps as in the previous lemma and by using the fact that expectation and variance do not depend on the concrete indices but rather on the structure of independences we get

$$\begin{aligned}
\mathbb{E}\left(C_5^{\star}(B)\right) &= \frac{1}{8B}\sum_{b=1}^{B}\mathbb{E}\left(\Lambda_1(\boldsymbol{\sigma}(b,6))\cdot\Lambda_2(\boldsymbol{\sigma}(b,6))\cdot\Lambda_3(\boldsymbol{\sigma}(b,6))\right)\\
&= \frac{1}{8B}\sum_{b=1}^{B}\mathbb{E}\left(\Lambda_1(\ell_{1,1},\dots,\ell_{6,a})\cdot\Lambda_2(\ell_{1,1},\dots,\ell_{6,a})\cdot\Lambda_3(\ell_{1,1},\dots,\ell_{6,a})\right).\\
&\overset{\text{C.1}}{=} \frac{1}{8B}\sum_{b=1}^{B}\operatorname{tr}\left((2\boldsymbol{T}\boldsymbol{V}_N)^3\right) = \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right).
\end{aligned}$$

$$Var(\mathbb{E}(C_5^{\star}(B)|\mathcal{F}(\boldsymbol{\sigma}(B,6)))) = Var\left(\operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)\right) = 0.$$

$$\begin{aligned}
Var\left(C_5^{\star}(B)\right) &= 0 + \mathbb{E}\left(Var\left(C_5^{\star}(B)|\mathcal{F}(\boldsymbol{\sigma}(B,6))\right)\right)\\
&\overset{\text{A.7}}{\leq} \mathbb{E}\Bigg(\sum_{(j,\ell)\in\mathbb{N}_B^2\setminus M(B,\boldsymbol{\sigma}(b,6))}\\
&\qquad\times\frac{Var\left(\Lambda_1(\boldsymbol{\sigma}(j,6))\Lambda_2(\boldsymbol{\sigma}(j,6))\Lambda_3(\boldsymbol{\sigma}(j,6))|\mathcal{F}(\boldsymbol{\sigma}(B,6))\right)}{64B^2}\Bigg)\\
&= \frac{\mathbb{E}\left(|\mathbb{N}_B^2\setminus M(B,\boldsymbol{\sigma}(b,6))|\right)}{B^2}\\
&\quad\cdot\frac{Var\left(\boldsymbol{Z}_{(1,2)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\cdot\boldsymbol{Z}_{(3,4)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(5,6)}\cdot\boldsymbol{Z}_{(5,6)}{}^{\top}\boldsymbol{T}\boldsymbol{Z}_{(1,2)}\right)}{64}\\
&\overset{\text{C.1}}{\leq} \left(1-\left(1-\frac{1}{B}\right)\cdot\prod_{i=1}^{a}\frac{\binom{n_i-6}{6}}{\binom{n_i}{6}}\right)\cdot 27\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right).
\end{aligned}$$

□

Proof of Theorem 4.3 (p.2753). With Lemma C.4 we recognize $\tau_P\to 1 \Leftrightarrow \widehat{\tau_P}\xrightarrow{\mathcal{P}} 1$ and $\tau_P\to 0 \Leftrightarrow \widehat{\tau_P}\xrightarrow{\mathcal{P}} 0$. Therefore $f_P\to 1\Leftrightarrow \widehat{f_P}\xrightarrow{\mathcal{P}} 1$ and $f_P\to\infty\Leftrightarrow\widehat{f}_P\xrightarrow{\mathcal{P}}\infty$. This is the only condition needed for the proof of Pauly et al. [34][Theorem 3.1], so the result follows. □

Although $n_{\min} = \mathcal{O}(a^q)$ with $q > 1$ is not too critical in most settings we additionally developed an estimator which can be used without any restrictions.

For this estimator another random vector has to be introduced: The random vector $\pi_{j,i}$ represents a random permutation of the numbers $1, \dots, n_i$, where $\pi_{j,i}$ are independent for different i or j and $\pi_{j,i}(l)$ denotes its l-th element. Then we define

$$C_7(w) = \frac{1}{w}\sum_{j=1}^{w}\sum_{\ell_1 \neq \cdots \neq \ell_6 = 1}^{n_{\min}} \frac{\Lambda_4(j;\ell_1,\dots,\ell_6)\cdot\Lambda_5(j;\ell_1,\dots,\ell_6)\cdot\Lambda_6(j;\ell_1,\dots,\ell_6)}{8\cdot\frac{n_{\min}!}{(n_{\min}-6)!}}$$

with

$$\begin{aligned}
\Lambda_4(j;\ell_1,\dots,\ell_6) &= {\boldsymbol{Z}^{\boldsymbol{\pi}_j}_{(\ell_1,\ell_2)}}^\top \boldsymbol{T}\boldsymbol{Z}^{\boldsymbol{\pi}_j}_{(\ell_3,\ell_4)},\\
\Lambda_5(j;\ell_1,\dots,\ell_6) &= {\boldsymbol{Z}^{\boldsymbol{\pi}_j}_{(\ell_3,\ell_4)}}^\top \boldsymbol{T}\boldsymbol{Z}^{\boldsymbol{\pi}_j}_{(\ell_5,\ell_6)},\\
\Lambda_6(j;\ell_1,\dots,\ell_6) &= {\boldsymbol{Z}^{\boldsymbol{\pi}_j}_{(\ell_5,\ell_6)}}^\top \boldsymbol{T}\boldsymbol{Z}^{\boldsymbol{\pi}_j}_{(\ell_1,\ell_2)}.
\end{aligned}$$

and

$$\boldsymbol{Z}^{\boldsymbol{\pi}_j}_{(\ell_1,\ell_2)} := \boldsymbol{Z}_{(\pi_{j,1}(\ell_1),\pi_{j,1}(\ell_2),\pi_{j,2}(\ell_1),\dots,\pi_{j,a}(\ell_1),\pi_{j,a}(\ell_2))}$$

This estimator again uses $\boldsymbol{Z}$, but different to C_5 the indices are the same for all groups. However the highest index is $n_{\min}$ and some index combinations are unachievable. For this reason, the above random permutations were used. So first the observations in each group were rearranged randomly and with this rearranged samples we calculated the sum of the used terms. Thereafter, we again rearrange the observations and the same terms as before are calculated. If these values were summed up and divided by the number of rearrangements we get an alternative for C_5 which is shown in the following lemma.

Lemma C.5. *For C_7 as defined before it holds*

$$\begin{aligned}
\mathbb{E}(C_7(w)) &= \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)\\
Var(C_7(w)) &\leq \left(\frac{\frac{n_{\min}!}{(n_{\min}-6)!} - \frac{(n_{\min}-6)!}{(n_{\min}-12)!}}{\frac{n_{\min}!}{(n_{\min}-6)!}}\right)\cdot\mathcal{O}\left(\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)\right).
\end{aligned}$$

Proof. Again we calculate

$$\begin{aligned}
\mathbb{E}(C_7(w)) &= \frac{1}{w}\sum_{j=1}^{w}\sum_{\ell_1 \neq \cdots \neq \ell_6 = 1}^{n_{\min}} \frac{\mathbb{E}\left(\prod_{m=4}^{6}\Lambda_m(j;\ell_1,\dots,\ell_6)\right)}{8\cdot\frac{n_{\min}!}{(n_{\min}-6)!}}\\
&= \frac{1}{w}\sum_{j=1}^{w}\sum_{\ell_1 \neq \cdots \neq \ell_6 = 1}^{n_{\min}} \frac{\mathbb{E}\left(\prod_{m=4}^{6}\Lambda_m(j;1,\dots,6)\right)}{8\cdot\frac{n_{\min}!}{(n_{\min}-6)!}} = \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right).
\end{aligned}$$

Because of the fact that all groups use the same indices, the number of remaining index combinations simplifies and we receive

$$\begin{aligned}
&Var\left(\sum_{\ell_1\neq\cdots\neq\ell_6=1}^{n_{\min}}\frac{\prod_{m=4}\Lambda_m\left(j;\ell_1,\ldots,\ell_6\right)}{8\cdot\frac{n_{\min}!}{(n_{\min}-6)!}}\right)\\
&\leq\frac{\frac{n_{\min}!}{(n_{\min}-6)!}-\frac{(n_{\min}-6)!}{(n_{\min}-12)!}}{\frac{n_{\min}!}{(n_{\min}-6)!}}\cdot Var\left(\prod_{m=4}^{6}\Lambda_m\left(j;\ell_1,\ldots,\ell_6\right)\right)\\
&\leq\frac{\frac{n_{\min}!}{(n_{\min}-6)!}-\frac{(n_{\min}-6)!}{(n_{\min}-12)!}}{\frac{n_{\min}!}{(n_{\min}-6)!}}\cdot\mathcal{O}\left(\mathrm{tr}^3\left(\left(\boldsymbol{TV}_N\right)^2\right)\right).
\end{aligned}$$

For the sum this leads to

$$\begin{aligned}
&Var\left(C_7\left(w\right)\right)\\
&=Var\left(\frac{1}{w}\sum_{j=1}^{w}\sum_{\ell_1\neq\cdots\neq\ell_6=1}^{n_{\min}}\frac{\prod_{m=4}^{6}\Lambda_m\left(j;\ell_1,\ldots,\ell_6\right)}{8\cdot\frac{n_{\min}!}{(n_{\min}-6)!}}\right)\\
&\overset{A.7}{\leq}\frac{1}{w^2}\sum_{j_1,j_2=1}^{w}Var\left(\sum_{\ell_1\neq\cdots\neq\ell_6=1}^{n_{\min}}\frac{\prod_{m=4}^{6}\Lambda_m\left(j;\ell_1,\ldots,\ell_6\right)}{8\cdot\frac{n_{\min}!}{(n_{\min}-6)!}}\right)\\
&\leq\frac{1}{w^2}\sum_{j_1,j_2=1}^{w}\left(\frac{\frac{n_{\min}!}{(n_{\min}-6)!}-\frac{(n_{\min}-6)!}{(n_{\min}-12)!}}{\frac{n_{\min}!}{(n_{\min}-6)!}}\right)\cdot\mathcal{O}\left(\mathrm{tr}^3\left(\left(\boldsymbol{TV}_N\right)^2\right)\right)\\
&=\left(\frac{\frac{n_{\min}!}{(n_{\min}-6)!}-\frac{(n_{\min}-6)!}{(n_{\min}-12)!}}{\frac{n_{\min}!}{(n_{\min}-6)!}}\right)\cdot\mathcal{O}\left(\mathrm{tr}^3\left(\left(\boldsymbol{TV}_N\right)^2\right)\right).
\end{aligned}$$

□

Simulations (not shown here) show that higher values for w lead to better estimations.

Lemma C.6. *For C_7 as previously defined, it holds*

$$\frac{C_7^2}{\mathrm{tr}^3\left(\left(\boldsymbol{TV}_N\right)^2\right)}-\tau_P\xrightarrow{\mathcal{P}}0\qquad \textit{for}\quad n_{\min}\to\infty,$$

independent of a or d. Therefore this holds for the asymptotic frameworks (3)–(5).

Proof. With the previous lemma we know

$$\mathbb{E}\left(\frac{C_7(w)-\mathrm{tr}\left(\left(\boldsymbol{TV}_N\right)^3\right)}{\mathrm{tr}^{3/2}\left(\left(\boldsymbol{TV}_N\right)^2\right)}\right)$$

$$= \mathbb{E}\left(\frac{C_7(w)}{\operatorname{tr}^{3/2}\left((\boldsymbol{TV}_N)^2\right)}\right) - \frac{\operatorname{tr}\left((\boldsymbol{TV}_N)^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{TV}_N)^2\right)} = 0,$$

$$\begin{aligned} &Var\left(\frac{C_7(w) - \operatorname{tr}\left((\boldsymbol{TV}_N)^3\right)}{\operatorname{tr}^{3/2}\left((\boldsymbol{TV}_N)^2\right)}\right) \\ &= \frac{Var\left(C_7(w)\right)}{\operatorname{tr}^3\left((\boldsymbol{TV}_N)^2\right)} \leq \left(\frac{\frac{n_{\min}!}{(n_{\min}-6)!} - \frac{(n_{\min}-6)!}{(n_{\min}-12)!}}{\frac{n_{\min}!}{(n_{\min}-6)!}}\right) \cdot \mathcal{O}(1). \end{aligned}$$

So exactly the same steps as in the proof of Lemma 4.2, which in this case uses that the zero sequence not depends on a or d, leads to the result. □

But for the calculation of this estimator we need $w \cdot n_{\min}!/(n_{\min}-6)!$ summations. Thus, a subsampling-type version of C_7 is necessary which is now defined.

Lemma C.7. *For each $b = 1, \dots, B$ we independently draw random subsamples $\boldsymbol{\sigma}_0(b, 6)$ of length 6 from $\{1, \dots, n_{\min}\}$ and define*

$$C_7^\star(w, B) = \sum_{j=1}^{w} \sum_{b=1}^{B} \frac{\Lambda_4\left(j; \boldsymbol{\sigma}_0(b,6)\right) \Lambda_5\left(j; \boldsymbol{\sigma}_0(b,6)\right) \Lambda_6\left(j; \boldsymbol{\sigma}_0(b,6)\right)}{8wB}$$

which holds

$$\begin{aligned} \mathbb{E}\left(C_7^\star(w,B)\right) &= \operatorname{tr}\left((\boldsymbol{TV}_N)\right), \\ Var\left(C_7^\star(w,B)\right) &= \left(1 - \left(1 - \frac{1}{B}\right)\frac{\binom{n_{\min}-6}{6}}{\binom{n_{\min}}{6}}\right) 27 \operatorname{tr}^3\left((\boldsymbol{TV}_N)^2\right). \end{aligned}$$

Proof. The proof for this subsampling-type estimator takes the same steps as before, with another amount $M(B, \boldsymbol{\sigma}_0(b,6))$. At the beginning we calculate expectation value and an upper bound for the variance of the inner sum. We get

$$\begin{aligned} \mathbb{E}\left(\sum_{b=1}^{B} \frac{\prod_{m=4}^{6} \Lambda_m\left(j; \boldsymbol{\sigma}_0(b,6)\right)}{8B}\right) &= \sum_{b=1}^{B} \frac{\mathbb{E}\left(\prod_{m=4}^{6} \Lambda_m\left(j; 1, \dots, 6\right)\right)}{8B} \\ &= \operatorname{tr}\left((\boldsymbol{TV}_N)^3\right). \end{aligned}$$

$$Var\left(\mathbb{E}\left(\sum_{b=1}^{B} \frac{\prod_{m=4}^{6} \Lambda_m\left(j; \boldsymbol{\sigma}_0(b,6)\right)}{8B} \Big| \mathcal{F}\left(\boldsymbol{\sigma}_0(B)\right)\right)\right) = Var\left(\operatorname{tr}\left((\boldsymbol{TV}_N)^3\right)\right) = 0.$$

$$Var\left(\sum_{b=1}^{B} \frac{\prod_{m=4}^{6} \Lambda_m\left(j; \boldsymbol{\sigma}_0(b,6)\right)}{8B}\right)$$

$$
\begin{aligned}
&= 0 + \mathbb{E}\left(Var\left(\sum_{b=1}^{B} \frac{\prod_{m=4}^{6} \Lambda_m\left(j; \boldsymbol{\sigma}_0(b,6)\right)}{8B} \middle| \mathcal{F}\left(\boldsymbol{\sigma}_0(B)\right) \right) \right) \\
&\overset{A.7}{\leq} \mathbb{E}\left(\sum_{(b_1,b_2)\in\mathbb{N}_B^2 \setminus M(B,\boldsymbol{\sigma}_0(b,6))} Var\left(\prod_{m=4}^{6} \Lambda_m\left(j; \boldsymbol{\sigma}_0(b_1,6)\right) \middle| \mathcal{F}\left(\boldsymbol{\sigma}_0(B)\right) \right) \cdot \frac{1}{64B^2} \right) \\
&= \frac{\mathbb{E}\left(|\mathbb{N}_B^2 \setminus M\left(B, \boldsymbol{\sigma}_0(b,6)\right)| \right)}{B^2} \\
&\quad \cdot \frac{Var\left(\Lambda_4\left(j;1,\ldots,6\right) \cdot \Lambda_5\left(j;1,\ldots,6\right) \cdot \Lambda_6\left(j;1,\ldots,6\right)\right)}{64} \\
&\overset{C.1}{\leq} \left(1 - \left(1 - \frac{1}{B}\right) \cdot \frac{\binom{n_{\min}-6}{6}}{\binom{n_{\min}}{6}} \right) \cdot 27 \operatorname{tr}^3\left(\left(\boldsymbol{T}\boldsymbol{V}_N\right)^2 \right).
\end{aligned}
$$

With these values we can consider the whole estimator

$$
\begin{aligned}
\mathbb{E}\left(C_7^{\star}\left(w,B\right)\right) &= \frac{1}{w}\sum_{j=1}^{w} \mathbb{E}\left(\sum_{b=1}^{B} \frac{\prod_{m=4}^{6} \Lambda_m\left(j; \boldsymbol{\sigma}_0(b,6)\right)}{8B} \middle| \mathcal{F}\left(\boldsymbol{\sigma}_0(B)\right) \right) \\
&= \operatorname{tr}\left(\left(\boldsymbol{T}\boldsymbol{V}_N\right)^3 \right), \\
Var\left(C_7^{\star}(w,B)\right) &\leq \frac{1}{w^2} \left(\sum_{j=1}^{w} \sqrt{ Var\left(\sum_{b=1}^{B} \frac{\prod_{m=4}^{6} \Lambda_m\left(j; \boldsymbol{\sigma}_0(b,6)\right)}{8B} \right) } \right)^2 \\
&\leq \frac{1}{w^2} \left(\sum_{j=1}^{w} \sqrt{ \left(1 - \left(1 - \frac{1}{B}\right) \cdot \frac{\binom{n_{\min}-6}{6}}{\binom{n_{\min}}{6}} \right) \cdot 27 \operatorname{tr}^3\left(\left(\boldsymbol{T}\boldsymbol{V}_N\right)^2 \right) } \right)^2 \\
&= \left(1 - \left(1 - \frac{1}{B}\right) \cdot \frac{\binom{n_{\min}-6}{6}}{\binom{n_{\min}}{6}} \right) \cdot 27 \operatorname{tr}^3\left(\left(\boldsymbol{T}\boldsymbol{V}_N\right)^2 \right). \qquad \square
\end{aligned}
$$

The next lemma shows that the version of the estimators with random indices has all the properties the classical ones possess.

Lemma C.8. *The statements of Lemma B.1, Lemma C.2, Lemma C.3, Lemma 4.2 and Lemma C.6 are also true, if all or only a part of the estimators are replaced by the subsampling-type estimators.*

Moreover, Theorem 3.1, Theorem 3.5 and Theorem 4.3 hold, if all or only a part of the estimators are replaced by the subsampling-type estimators.

Proof. For the proofs of the classical estimators from the first paragraph, only the expectation values are used together with upper bounds for the variances which are zero sequences. With random indices, the expectation is the same and for the variance, all traces are the same but the zero sequence changes. So the proofs of the subsampling-type estimators work identically.

For the second paragraph, only some convergences are necessary, which the subsampling-type estimators also fulfills. □

Appendix D: On the asymptotic distribution in our simulation designs

To determine the asymptotic distribution of our test statistic (corresponding to validity of the different tests) in our simulation settings, the asymptotic behaviour of β_1 has to be investigated. Due to equivalence we calculate the value of $\tau_P = \operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^3\right)\Big/\operatorname{tr}^3\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)$. This is sufficient since $\boldsymbol{V}_N$ is known, i.e. no estimation is needed. The ratio n_1/N and n_2/N are the same for all our sample sizes, so the different numbers n_1, n_2 have no influence on the values of τ_P. Results for different choices of $\boldsymbol{T}$ and $\boldsymbol{\Sigma}_i, i = 1, 2$, corresponding to the simulation settings from Section 5 are displayed in Tables 3–5. It can be seen that for H_0^a (Table 3) we have $\tau_P \to 1$ and thus $\beta_1 \to 1$ by Lemma A.8. For H_0^b (Table 4) we have $\tau_P \to 0$ and thus $\beta_1 \to 0$; and in case of the autoregressive covariance matrices with correlation factor depending on the the dimension, we seem to have $\beta_1 \to b_1 \approx 0.7589$.

TABLE 3
τ_P for $\boldsymbol{T} = \left(\boldsymbol{P}_2 \otimes \frac{1}{d}\boldsymbol{J}_d\right)$ with $(\boldsymbol{\Sigma}_1)_{i,j} = 0.6^{|i-j|}$ and $(\boldsymbol{\Sigma}_2)_{i,j} = 0.65^{|i-j|}$

d	5	10	20	40	70	100	150	200	300	450	600	800
τ_P	1	1	1	1	1	1	1	1	1	1	1	1

TABLE 4
τ_P for $\boldsymbol{T} = \left(\frac{1}{2}\boldsymbol{J}_2 \otimes \boldsymbol{P}_d\right)$ and $\boldsymbol{T} = (\boldsymbol{P}_2 \otimes \boldsymbol{P}_d)$ with $(\boldsymbol{\Sigma}_1)_{i,j} = 0.6^{|i-j|}$ and $(\boldsymbol{\Sigma}_2)_{i,j} = 0.65^{|i-j|}$

d	5	10	20	40	70	100	150	200	300	450	600	800
τ_P	.49	.35	.21	.11	.061	.043	.029	.021	.014	.0095	.0071	.0053

TABLE 5
τ_P and β_1 for $\boldsymbol{T} = \left(\frac{1}{2}\boldsymbol{J}_2 \otimes \boldsymbol{P}_d\right)$ with $(\boldsymbol{\Sigma}_1)_{i,j} = 0.6^{|i-j|/d}$ and $(\boldsymbol{\Sigma}_2)_{i,j} = 0.65^{|i-j|/d}$

d	5	10	20	40	70	100	150	200	300	450	800
τ_P	.9311	.9408	.9444	.9454	.9457	.9457	.9458	.9458	.9458	.9458	.9458
β_1	.7082	.7392	.7534	.7575	.7584	.7587	.7588	.7588	.7589	.7589	.7589

Appendix E: On the Chen-Qin-Condition

We can also develop an estimator for $\tau_{CQ} = \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right)/\operatorname{tr}^2\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right) = 1/f_{CQ}$ on an analogical way as before. This leads to:

Lemma E.1. *Let be*

$$C_6 = \sum_{\substack{\ell_{1,1},\dots,\ell_{8,1}=1 \\ \ell_{1,1}\neq\dots\neq\ell_{8,1}}}^{n_1} \cdots \sum_{\substack{\ell_{1,a},\dots,\ell_{8,a}=1 \\ \ell_{1,a}\neq\dots\neq\ell_{8,a}}}^{n_a} \left[\frac{1}{6}\frac{\Lambda_7(\ell_{1,1},\dots,\ell_{8,a})}{16\cdot\prod_{i=1}^{a}\frac{n_i!}{(n_i-8)!}} - \frac{1}{2}\frac{\Lambda_8(\ell_{1,1},\dots,\ell_{8,a})}{16\cdot\prod_{i=1}^{a}\frac{n_i!}{(n_i-8)!}}\right]$$

with

$$\Lambda_7(\ell_{1,1},\ldots,\ell_{8,a}) = \left[\boldsymbol{Z}^\top_{(\ell_{1,1},\ell_{2,1},\ldots,\ell_{2,a})}\boldsymbol{T}\boldsymbol{Z}_{(\ell_{3,1},\ell_{4,1},\ldots\ell_{4,a})}\right]^4,$$
$$\Lambda_8(\ell_{1,1},\ldots,\ell_{8,a}) = \left[\sqrt{\Lambda_7(\ell_{1,1},\ldots,\ell_{8,a})}\cdot\boldsymbol{Z}^\top_{(\ell_{5,1},\ell_{6,1},\ldots,\ell_{6,a})}\boldsymbol{T}\boldsymbol{Z}_{(\ell_{7,1},\ell_{8,1},\ldots,\ell_{8,a})}\right]^2.$$

Then we know

$$\mathbb{E}(C_6) = \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right) \qquad Var(C_6) \le \frac{\prod_{i=1}^a \binom{n_i}{8} - \prod_{i=1}^a \binom{n_i-8}{8}}{16^2\cdot\prod_{i=1}^a\binom{n_i}{8}}\mathcal{O}\left(\operatorname{tr}^4\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)\right).$$

Proof.

$$\begin{aligned}\mathbb{E}(C_6) &= \frac{\mathbb{E}\left(\left[\boldsymbol{Z}^\top_{(1,2)}\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\right]^4\right)}{6\cdot 16} - \frac{\mathbb{E}\left(\left[\boldsymbol{Z}^\top_{(1,2)}\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\right]^2\left[\boldsymbol{Z}^\top_{(5,6)}\boldsymbol{T}\boldsymbol{Z}_{(7,8)}\right]^2\right)}{2\cdot 16}\\ &\overset{A.4}{=} \frac{1}{6\cdot 16}\left(6\operatorname{tr}\left((2\boldsymbol{T}\boldsymbol{V}_N)^4\right) + 3\operatorname{tr}^2\left((2\boldsymbol{T}\boldsymbol{V}_N)^2\right)\right) - \frac{1}{2\cdot 16}\operatorname{tr}^2\left((2\boldsymbol{T}\boldsymbol{V}_N)^2\right)\\ &= \operatorname{tr}\left((\boldsymbol{T}\boldsymbol{V}_N)^4\right)\end{aligned}$$

For the second inequality, the variance of parts is calculated. Like before with Lemma A.2 (p.2767) and Theorem A.4 (p.2768) we calculate

$$Var\left(\frac{1}{6}\left[\boldsymbol{Z}_{(1,2)}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\right]^4\right) = \mathcal{O}\left(\operatorname{tr}^4\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)\right)$$

and

$$\begin{aligned}&Var\left(\frac{1}{2}\left[\boldsymbol{Z}_{(1,2)}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\right]^2\left[\boldsymbol{Z}_{(5,6)}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(7,8)}\right]^2\right)\\ &\le \frac{1}{4}\cdot\mathbb{E}\left(\left[\boldsymbol{Z}_{(1,2)}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(3,4)}\right]^4\left[\boldsymbol{Z}_{(5,6)}{}^\top\boldsymbol{T}\boldsymbol{Z}_{(7,8)}\right]^4\right)\\ &= \frac{1}{4}\left(6\operatorname{tr}\left((2\boldsymbol{T}\boldsymbol{V}_N)^4\right) + 3\operatorname{tr}^2\left((2\boldsymbol{T}\boldsymbol{V}_N)^2\right)\right)^2 = \mathcal{O}\left(\operatorname{tr}^4\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)\right).\end{aligned}$$

With Lemma A.7 (p.2770) it is known

$$Var(B) \le Var(A) + Var(B) + 2|\,Cov(A,B)| \le \left(\sqrt{Var(A)} + \sqrt{Var(B)}\right)^2$$

and therefore

$$Var(C_6) \le \frac{\prod_{i=1}^a \binom{n_i}{8} - \prod_{i=1}^a \binom{n_i-8}{8}}{16^2\cdot\prod_{i=1}^a\binom{n_i}{8}}\,Var\left(\frac{1}{6}\Lambda_7(1,\ldots,8) - \frac{1}{2}\Lambda_8(1,\ldots,8)\right)$$

$$\leq \frac{\prod_{i=1}^{a}\binom{n_i}{8} - \prod_{i=1}^{a}\binom{n_i-8}{8}}{16^2 \cdot \prod_{i=1}^{a}\binom{n_i}{8}}$$

$$\times \left(\sqrt{\mathcal{O}\left(\mathrm{tr}^4\left((\boldsymbol{TV}_N)^2\right)\right)} + \sqrt{\mathcal{O}\left(\mathrm{tr}^4\left((\boldsymbol{TV}_N)^2\right)\right)}\right)^2$$

$$= \frac{\prod_{i=1}^{a}\binom{n_i}{8} - \prod_{i=1}^{a}\binom{n_i-8}{8}}{16^2 \cdot \prod_{i=1}^{a}\binom{n_i}{8}} \mathcal{O}\left(\mathrm{tr}^4\left((\boldsymbol{TV}_N)^2\right)\right). \qquad \square$$

Lemma E.2. *With the estimators introduced in the previous lemmata it holds for fixed* a

$$\frac{C_6}{A_4^2} - \frac{\mathrm{tr}\left((\boldsymbol{TV}_N)^4\right)}{\mathrm{tr}^2\left((\boldsymbol{TV}_N)^2\right)} \xrightarrow{\mathcal{P}} 0 \qquad \text{for} \quad d, n_{\min} \to \infty.$$

If $q > 1$ *exists with* $n_{\min} = \mathcal{O}(a^q)$*, the convergence even holds in the asymptotic frameworks* (4)–(5).

Proof. Again we first consider the parts:

$$\mathbb{E}\left(\frac{C_6}{\mathrm{tr}^2\left((\boldsymbol{TV}_N)^2\right)} - \frac{\mathrm{tr}\left((\boldsymbol{TV}_N)^4\right)}{\mathrm{tr}^2\left((\boldsymbol{TV}_N)^2\right)}\right) = \frac{\mathbb{E}\left(C_6\right)}{\mathrm{tr}^2\left((\boldsymbol{TV}_N)^2\right)} - \frac{\mathrm{tr}\left((\boldsymbol{TV}_N)^4\right)}{\mathrm{tr}^2\left((\boldsymbol{TV}_N)^2\right)} = 0.$$

$$\mathit{Var}\left(\frac{C_6}{\mathrm{tr}^2\left((\boldsymbol{TV}_N)^2\right)} - \frac{\mathrm{tr}\left((\boldsymbol{TV}_N)^4\right)}{\mathrm{tr}^2\left((\boldsymbol{TV}_N)^2\right)}\right)$$

$$\leq \frac{\prod_{i=1}^{a}\binom{n_i}{8} - \prod_{i=1}^{a}\binom{n_i-8}{8}}{16^2 \cdot \prod_{i=1}^{a}\binom{n_i}{8}} \frac{\mathcal{O}\left(\mathrm{tr}^4\left((\boldsymbol{TV}_N)^2\right)\right)}{\mathrm{tr}^4\left((\boldsymbol{TV}_N)^2\right)}$$

$$\leq \frac{\prod_{i=1}^{a}\binom{n_i}{8} - \prod_{i=1}^{a}\binom{n_i-8}{8}}{\prod_{i=1}^{a}\binom{n_i}{8}} \cdot \mathcal{O}(1).$$

So with Lemma A.6 (p.2770) for fixed a and $d, n_{\min} \to \infty$ and moreover if the additional condition is fulfilled even for the asymptotic frameworks (4)–(5), it follows

$$\frac{C_6}{\mathrm{tr}^2\left((\boldsymbol{TV}_N)^2\right)} - \frac{\mathrm{tr}\left((\boldsymbol{TV}_N)^4\right)}{\mathrm{tr}^2\left((\boldsymbol{TV}_N)^2\right)} \xrightarrow{\mathcal{P}} 0.$$

Analogue to the proof of Lemma 4.2 it follows $\mathrm{tr}^2\left((\boldsymbol{TV}_N)^2\right) \big/ A_4^2 \xrightarrow{\mathcal{P}} 1$.

Together this leads to

$$\frac{C_6}{A_4^2} - \frac{\operatorname{tr}\left((\boldsymbol{TV}_N)^4\right)}{\operatorname{tr}^2\left((\boldsymbol{TV}_N)^2\right)} = \frac{C_6}{\operatorname{tr}^2\left((\boldsymbol{TV}_N)^2\right)} \cdot \frac{\operatorname{tr}^2\left((\boldsymbol{TV}_N)^2\right)}{A_4^2} - \frac{\operatorname{tr}\left((\boldsymbol{TV}_N)^4\right)}{\operatorname{tr}^2\left((\boldsymbol{TV}_N)^2\right)}$$

$$= \frac{C_6}{\operatorname{tr}^2\left((\boldsymbol{TV}_N)^2\right)} \cdot (1 + \mathcal{O}_{\mathcal{P}}(1)) - \frac{\operatorname{tr}\left((\boldsymbol{TV}_N)^4\right)}{\operatorname{tr}^2\left((\boldsymbol{TV}_N)^2\right)} = \mathcal{O}_{\mathcal{P}}(1) + \mathcal{O}_{\mathcal{P}}(1) = \mathcal{O}_{\mathcal{P}}(1). \qquad \square$$

Again in most cases, the subsampling-type version of this estimator should be used.

Lemma E.3. *Let be*

$$C_6^\star(B) = \frac{1}{16B} \sum_{b=1}^{B} \left(\frac{\Lambda_7(\boldsymbol{\sigma}(b,8))}{6} - \frac{\Lambda_8(\boldsymbol{\sigma}(b,8))}{2} \right).$$

Then it holds

$$\mathbb{E}\left(C_6^\star(B)\right) = \operatorname{tr}\left((\boldsymbol{TV}_N)^4\right),$$

$$Var\left(C_6^\star(B)\right) \leq \left(1 - \left(1 - \frac{1}{B}\right) \cdot \prod_{i=1}^{a} \frac{\binom{n_i - 8}{8}}{\binom{n_i}{6}}\right) \cdot \mathcal{O}\left(\operatorname{tr}^4\left((\boldsymbol{TV}_N)^2\right)\right).$$

Proof. By using the same steps as before it holds

$$\begin{aligned}
&\mathbb{E}\left(C_6^\star(B)\right) \\
&= \frac{1}{16B} \sum_{b=1}^{B} \mathbb{E}\left(\frac{\Lambda_7(\ell_{1,1},\dots,\ell_{8,a})}{6} - \frac{\Lambda_8(\ell_{1,1},\dots,\ell_{8,a})}{2} \right) \\
&= \frac{1}{16B} \sum_{b=1}^{B} \mathbb{E} \\
&\quad \times \left(\left[\boldsymbol{Z}_{(1,2)}{}^\top \boldsymbol{TZ}_{(3,4)}\right]^2 \cdot \left(\frac{\left[\boldsymbol{Z}_{(1,2)}{}^\top \boldsymbol{TZ}_{(3,4)}\right]^2}{6} - \frac{\left[\boldsymbol{Z}_{(5,6)}{}^\top \boldsymbol{TZ}_{(7,8)}\right]^2}{2} \right) \right) \\
&\overset{\text{E.1}}{=} \frac{1}{16B} \sum_{b=1}^{B} \operatorname{tr}\left((2\boldsymbol{TV}_N)^4\right) = \operatorname{tr}\left((\boldsymbol{TV}_N)^4\right).
\end{aligned}$$

$$Var\left(\mathbb{E}\left(C_6^\star(B) | \mathcal{F}(\boldsymbol{\sigma}(B,8))\right)\right) = Var\left(\operatorname{tr}\left((\boldsymbol{TV}_N)^4\right)\right) = 0.$$

$$\begin{aligned}
&Var\left(C_6^\star(B)\right) \\
&= 0 + \mathbb{E}\left(Var\left(C_6^\star(B) | \mathcal{F}(\boldsymbol{\sigma}(B,8))\right)\right) \\
&\overset{\text{A.7}}{\leq} \frac{1}{16^2 B^2} \mathbb{E}
\end{aligned}$$

$$\times \left(\sum_{(j,\ell)\in\mathbb{N}_B^2\setminus M(B,\boldsymbol{\sigma}(b,8))} Var\left(\frac{\Lambda_7(\boldsymbol{\sigma}(j,8))}{6} - \frac{\Lambda_8(\boldsymbol{\sigma}(j,8))}{2} \Big| \mathcal{F}(\boldsymbol{\sigma}(B,8)) \right) \right)$$

$$= \frac{Var\left(\frac{\Lambda_7(\ell_{1,1},\dots,\ell_{8,a})}{6} - \frac{\Lambda_8(\ell_{1,1},\dots,\ell_{8,a})}{2}\right)}{16^2 B \cdot \left(\mathbb{E}\left(|\mathbb{N}_B^2 \setminus M(B,\boldsymbol{\sigma}(b,8))|\right)\right)^{-1}}$$

$$\overset{\text{E.1}}{\leq} \left(1 - \left(1 - \frac{1}{B}\right) \cdot \prod_{i=1}^{a} \frac{\binom{n_i-8}{8}}{\binom{n_i}{8}}\right) \cdot \mathcal{O}\left(\operatorname{tr}^4\left((\boldsymbol{T}\boldsymbol{V}_N)^2\right)\right). \qquad \square$$

With Lemma C.7 we get an estimator for τ_{CQ} with $\widehat{\tau_{CQ}}(C_6^\star, A_4) = C_6^\star/A_4^2$ and once more for a large number of groups $A_4^\star$ should be used.

Lemma E.4. *Theorem 4.1 is also valid if f_P is replaced by f_{CQ} or by $(\widehat{\tau_{CQ}}(C_6, A_4))^{-1}$. Using $C_6^\star$ or $A_4^\star$ also doesn't change the result. Identical the result of Lemma E.2 remains true if one or all estimators are replaced by their subsampling version.*

Proof. With Lemma A.8 we know $f_P \to 1 \Leftrightarrow f_{CQ} \to 1$ and $f_P \to 0 \Leftrightarrow f_{CQ} \to 0$ so in both cases K_{f_P} is asymptotically identic with $K_{f_{CQ}}$.

From Lemma E.2 we know that $\widehat{\tau_{CQ}} - \tau_{CQ}$ converges in probability to zero so this result follows identically to Theorem 4.1. At last the subsampling versions have the same properties as the standard estimators. □

Therefore this is a second way to test the hypotheses and moreover, it provides an indicator for the choice of the limit distribution, because of Lemma A.8. For situation c) from Theorem 3.1 there is no proof that this approach can be used but in the case of just one group, it leads to good results.

References

[1] Ahmad, M. R., Werner, C. and Brunner, E. (2008). Analysis of High Dimensional Repeated Measures Designs: The One Sample Case. *Computational Statistics and Data Analysis* **53** 416–427. MR2649096

[2] Bai, Z. and Saranadasa, H. (1996). Effect of highdimension: by an example of a two sample problem. *Statistica Sinica* **6** 311–329. MR1399305

[3] Bathke, A. C. and Harrar, S. W. (2008). Nonparametric methods in multivariate factorial designs for large number of factor levels. *Journal of Statistical Planning and Inference* **138** 588–610. MR2382556

[4] Bathke, A. C., Harrar, S. W. and Madden, L. V. (2008). How to compare small multivariate samples using nonparametric tests. *Computational Statistics and Data Analysis* **52** 4951–4965. MR2526205

[5] Billingsley, P. (1968). *Convergence of probability measures*. John Wiley & Sons, New York. MR0233396

[6] Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics* **25** 290–302. MR0061787

[7] Brunner, E. (2009). Repeated measures under non-sphericity. Proceedings of the 6th St. Petersburg Workshop on Simulation.

[8] Brunner, E., Becker, B. and Werner, C. (2010). Approximate distributions of quadratic forms in high-dimensional repeated-measures designs. Technical Report, Department Medizinische Statistik Georg-August-Universität Göttingen

[9] Brunner, E., Bathke, A. C. and Placzek, M. (2012). Estimation of Box's ϵ for low- and high-dimensional repeated measures designs with unequal covariance matrices. *Biometrical Journal* **54** 301–316. MR2931076

[10] Cai, T., Liu, W. and Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B* **7**6 349–372. MR3164870

[11] Chen, S. X. and Qin, Y.-L. (2010). A Two-Sample Test for High-Dimensional Data with Applications to Gene-Set Testing. *The Annals of Statistics* **38** 808–835. MR2604697

[12] Cramér, H. (1936). Ueber eine Eigenschaft der normalen Verteilungsfunktion. *Mathematische Zeitschrift* **41** 405–414. MR1545629

[13] Croux, C., Rousseeuw, P. J. and Hössjer, O. (1994). Generalized S-estimators. *Journal of the American Statistical Association* **89** 1271–1281. MR1310221

[14] Davis, C. S. (2002). *Statistical Methods for the Analysis of Repeated Measurements*. Springer, New York. MR1883764

[15] Dümbgen, L. (1998). On Tyler's M-Functional of Scatter in High Dimension. *Annals of the Institute of Statistical Mathematics* **50** 471–491. MR1664575

[16] Friedrich, S., Brunner, E. and Pauly, M. (2017). Permuting longitudinal data in spite of the dependencies. *Journal of Multivariate Analysis* **153** 255–265. MR3578850

[17] Geisser, S. and Greenhouse, S. W. (1958). An Extension of Box's Result on the Use of the F Distribution in Multivariate Analysis. *Annals of Mathematical Statistics* **29** 885–891. MR0096331

[18] Greenhouse, S. W. and Geisser, S. (1959). On Methods in the Analysis of Profile Data. *Psychometrika* **24** 95–112. MR0103783

[19] Hájek, J., Šidak, Z. and Sen, P. K. (1999). *Theory of Rank Tests*. Academic Press, San Diego. MR1680991

[20] Happ, M., Harrar, S. W. and Bathke, A. C. (2016). Inference for low- and high-dimensional multigroup repeated measures designs with unequal covariance matrices. *Biometrical Journal* **58** 810–830. MR3527417

[21] Harden, M. (2012). Das Behrens-Fisher-Problem für hochdimensionale Split-Plot-Designs. Master thesis, University Göttingen. http://www.ams.med.uni-goettingen.de/download/diplom/Master-Harden.pdf

[22] Harrar, S. W. and Kong, X. (2016). High-dimensional multivariate repeated measures analysis with unequal covariance matrices. *Journal of Multivariate Analysis* **145** 1–21. MR3459934

[23] Huynh, H. and Feldt, L. S. (1976). Estimation of the Box Correction for Degrees of Freedom From Sample Data in Randomized Block and Split-Plot Designs. *Journal of Educational Statistics* **1** 69–82.

[24] Johnson, R. and Wichern, D. (2007). *Applied multivariate statistical analysis*. 6th Edition, Prentice Hall. MR2372475

[25] Jordan, W., Tumani, H., Cohrs, S., Eggert, S., Rodenbeck, A., Brunner, E., Rüther, E., Hajak, G. (2004). Prostaglandin-D-synthase (beta-trace) in healthy human sleep. *Sleep* **27** 867–874.

[26] Katayama, S., Kano, Y. and Srivastava, M. S. (2013). Asymptotic distributions of some test criteria for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis* **116** 410–421. MR3049913

[27] Kenward, M. G. and Roger, J. H. (2009). An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics & Data Analysis* **53** 2583–2595. MR2665909

[28] Keselman, H. J., Algina, J. and Kowalchuk, R. K. (2001). The analysis of repeated measures designs. *British Journal of Mathematical and Statistical Psychology* **54** 1–20.

[29] Konietschke, F., Bathke, A. C., Harrar, S. W. and Pauly, M. (2015). Parametric and Nonparametric Bootstrap Methods for General MANOVA. *Journal of Multivariate Analysis* **140** 291–301. MR3372569

[30] Lecoutre, B. (1991). A Correction for the $\widetilde{\epsilon}$: Approximative Test in Repeated Measures Designs With Two or More Independent Groups. *Journal of Educational Statistics* **16**, 371–372.

[31] Liu, Z., Liu, B., Zheng, S. and Shi, N.-Z. (2017). Simultaneous testing of mean vector and covariance matrix for high-dimensional data. *Journal of Statistical Planning and Inference* **188** 82–93. MR3648319

[32] Mathai, A. M. and Provost, S. B. (1992). *Quadratic forms in random variables*. Marcel Dekker Inc., New York. MR1192786

[33] Nishiyama, T., Hyodo, M., Seo, T. and Pavlenko, T. (2013). Testing linear hypotheses of mean vectors for high-dimension data with unequal covariance matrices. *Journal of Statistical Planning and Inference* **143** 1898–1911. MR3095080

[34] Pauly, M., Ellenberger, D. and Brunner, E. (2015). Analysis of high-dimensional one group repeated measures designs. *Statistics* **49**(6) 1243–1261. MR3415218

[35] Pesarin, F. and Salmaso, L. (2012). A review and some new results on permutation testing for multivariate problems. *Statistics and Computing* **22** 639–646. MR2865041

[36] Rao, C. R. and Mitra, S. K. (1971). *Generalized Inverse of Matrices and Its Applications*. Wiley, New York. MR0338013

[37] Secchi, P., Stamm, A. and Vantini, S. (2013). Inference for the mean of large p small n data: A finite-sample high-dimensional generalization of Hotelling's theorem. *Electronic Journal of Statistics* **7** 2005–2031. MR3085016

[38] Skene, S. S. and Kenward, M. G. (2010). The analysis of very small samples of repeated measurements II: A modified Box correction. *Statistics in Medicine* **29** 2838–2856. MR2758474

[39] Tyler, D. E., Critchley, F., Dümbgen L. and Oja, H. (2009). Invariant Co-ordinate Selection (with discussion). *Journal of the Royal Statistical Society, Series B* **71** 549–592. MR2749907

[40] Vallejo, G. and Ato, M. (2006). Modified Brown-Forsythe procedure for testing interaction effects in split-plot designs. *Multivariate Behavioral Research* **41** 549–578.

[41] Werner, C. (2004). Dimensionsstabile Approximation für Verteilungen von quadratischen Formen im Repeated-Measures-Design. Diploma Thesis, University of Göttingen. http://www.ams.med.uni-goettingen.de/download/TECHREPORT.PDF

[42] Witting, H. and Müller-Funke, U. (1995). Asymptotische Statistik: Parametrische Modelle und nicht-parametrische Funktionale, Stuttgart: Teubner.

[43] Zhan, D. and Hart, J. D. (2014). Testing equality of a large number of densities. *Biometrika* **101.2** 449–464. MR3215359

[44] Zhang, J. T. (2005). Approximate and asymptotic distributions of chi-squared–type mixtures with applications. *Journal of the American Statistical Association* **100** 273–285. MR2156837