TEST (2024) 33:496–516
https://doi.org/10.1007/s11749-023-00906-6

ORIGINAL PAPER

Testing hypotheses about correlation matrices in general MANOVA designs

Paavo Sattler[1] · **Markus Pauly**[1,2]

Received: 15 September 2022 / Accepted: 7 November 2023 / Published online: 12 December 2023

Abstract
Correlation matrices are an essential tool for investigating the dependency structures of random vectors or comparing them. We introduce an approach for testing a variety of null hypotheses that can be formulated based upon the correlation matrix. Examples cover MANOVA-type hypothesis of equal correlation matrices as well as testing for special correlation structures such as sphericity. Apart from existing fourth moments, our approach requires no other assumptions, allowing applications in various settings. To improve the small sample performance, a bootstrap technique is proposed and theoretically justified. Based on this, we also present a procedure to simultaneously test the hypotheses of equal correlation and equal covariance matrices. The performance of all new test statistics is compared with existing procedures through extensive simulations.

Keywords Bootstrap · Multivariate data · Nonparametric testing · Resampling · Correlation matrices · Quadratic forms

Mathematics Subject Classification 62H15 · 62G09 · 62G10

1 Motivation and introduction

Covariance matrices contain a multitude of information about a random vector. Therefore, they were the topic of manifold investigations. For testing, an important hypothesis is the equality of covariance matrices from different groups. This

✉ Paavo Sattler
paavo.sattler@tu-dortmund.de

Markus Pauly
markus.pauly@tu-dortmund.de

1 Department of Statistics, Institute for Mathematical Statistics and Industrial Applications, TU Dortmund University, Joseph-von-Fraunhofer-Straße 2-4, 44227 Dortmund, Germany

2 Research Center Trustworthy Data Science and Security, UA Ruhr, Joseph-von-Fraunhofer-Straße 25, 44227 Dortmund, Germany

was investigated in Bartlett and Rajalakshman (1953) as well as Boos and Brownie (2004). Moreover, Gupta and Xu (2006) proposed tests for a given covariance matrix. Extending on both, Sattler et al. (2022) proposed a unifying framework that allows for investigations of various hypotheses about the covariance. Additional examples cover testing equality of traces or comparing diagonal elements of the covariance matrix.

However, covariance matrices are not scale-invariant. This entails some disadvantages using them in the analysis of random vectors' dependency structure. For example, a simple change of a measuring unit can completely change the matrix. For this reason alone, it is more useful to consider the correlation matrix instead when inferring dependency structures.

This already starts in the bivariate case, i.e. the investigation of correlations. For ordinal data, rank-based correlation measures are most common. For example, Perreault et al. (2022) provides an approach for general hypotheses testing, while Nowak and Konietschke (2021) focused on simultaneous confidence intervals and multiple contrast tests for Kendall's τ. Spearman's rank-correlation coefficient ρ was investigated in Gaißer and Schmid (2010) for the hypothesis of the correlation matrix being an equicorrelation matrix. For metric data, the most common measure of correlation is the Pearson correlation coefficient. Here, tests and confidence intervals for investigating or comparing correlation coefficients have been discussed for the case of one, two or multiple group(s), see (Fisher 1921; Efron 1988; Sakaori 2002; Gupta and Xu 2006; Tian and Wilding 2008; Omelka and Pauly 2012; Welz et al. 2022) and the references cited therein.

For larger dimensions, equality of correlation matrices was investigated, for example, in Jennrich (1970). Moreover, different hypotheses regarding the structure of correlation matrices were treated in Joereskog (1978), Steiger (1980) and Wu et al. (2018). However, the above approaches usually require strong prerequisites on the distribution (such as multivariate normal distribution or particular properties of the moments), the components or the setting (such as bivariate or special structures). Moreover, most can be used only for a few specific hypotheses. Thus, to obtain an approach with fewer assumptions, which is at the same time applicable for a multitude of hypotheses, we expand the approach of Sattler et al. (2022) to the treatment of correlation matrices.

In the following section, the statistical model will be introduced together with examples of different null hypotheses that can be investigated using the proposed approach. Afterwards, the asymptotic distributions of the proposed test statistics are derived (Sect. 3). In Sects. 4 and 5, a resampling strategy and a Taylor-based Monte Carlo approach are used to generate critical values and improve our tests' small sample behaviour. A combined testing procedure which simultaneously checks the hypothesis of equal correlation matrices and covariance matrices is presented in Sect. 6. The simulation results regarding type-I-error control and power are discussed in Sect. 7, while an illustrative data analysis of EEG data is conducted in Sect. 8. All proofs are deferred to a technical supplement, where also more simulations can be found.

2 Statistical model and hypotheses

To allow for broad applicability, we use a general semiparametric model, given by independent d-dimensional random vectors

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

where the index $i = 1, \ldots, a$ refers to the treatment group and $k = 1, \ldots, n_i$ to the individual, on which d-variate observations are measured. As analysing a scalar's correlation is useless, we assume $d \geq 2$. This model as well as all subsequent results can in principle be expanded for different dimensions of the groups as done in Friedrich et al. (2017). However, as the notation would be a bit cumbersome, we did not follow this approach to increase the readability. The residuals $\boldsymbol{\epsilon}_{i1}, \ldots, \boldsymbol{\epsilon}_{in_i}$ are assumed to be centred $\mathbb{E}(\boldsymbol{\epsilon}_{i1}) = \mathbf{0}_d$ and i.i.d. within each group, with finite fourth moment $\mathbb{E}(||\boldsymbol{\epsilon}_{i1}||^4) < \infty$, while across groups, they are only independent. Thus, different distributions per group are possible. For our asymptotic considerations, we additionally assume

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

with $N = \sum_{i=1}^{a} n_i$, to preclude that a single group dominates the setting. In the following, we use $\xrightarrow{\mathcal{P}}$ to denote convergence in probability and $\xrightarrow{\mathcal{D}}$ for convergence in distribution as sample sizes increase. It follows from the context whether this means $N \to \infty$ or $n_i \to \infty$, respectively.

Moreover, to define correlation matrices, we additionally assume that the covariance matrices $\operatorname{Cov}(\boldsymbol{\epsilon}_{i1}) = \boldsymbol{V}_i = (V_{ijk})_{j,k\leq d}$ have positive diagonals $V_{i11}, ..., V_{idd} > 0$. Throughout, the so-called half-vectorization operation vech is used for the covariance matrices, which puts the upper triangular elements of a $d \times d$ matrix into a $p = d(d+1)/2$ dimensional vector. But, for a correlation matrix, the diagonal elements are always one and therefore contain no information, so this is not the best choice here. Hence, a new vectorization operation vech^- is defined, which we will call the upper-half-vectorization. With $\boldsymbol{R}_i$ as the correlation matrix for the i-th group, this vectorization operation allows us to define

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

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

$$\mathcal{H}_0^{\boldsymbol{r}} : \boldsymbol{C}\boldsymbol{r} = \boldsymbol{\zeta}, \tag{1}$$

with a proper hypothesis matrix $\boldsymbol{C} \in \mathbb{R}^{m \times ap_u}$ and a vector $\boldsymbol{\zeta} \in \mathbb{R}^m$. Hereby, we allow m to be much smaller than ap_u, which is often useful (e.g. for computational reasons) for hypotheses matrices without full rank. As explained in Sattler et al. (2022) for the case of covariance matrices, $\boldsymbol{C}$ does not have to be a projection matrix as $\boldsymbol{\zeta}$ is allowed to be different from the zero vector, see, for example, hypothesis (c).

Hypotheses which are part of the setting (1) are, among others:

(a) **Testing Homogeneity of correlation matrices**

$$\mathcal{H}_0^{\boldsymbol{r}} : \boldsymbol{R}_1 = \cdots = \boldsymbol{R}_a, \text{ resp. } \mathcal{H}_0^{\boldsymbol{r}} : (\boldsymbol{P}_a \otimes \boldsymbol{I}_{p_u})\boldsymbol{r} = \boldsymbol{0}_{ap_u},$$

with $\boldsymbol{P}_a := \boldsymbol{1}_a\boldsymbol{1}_a^\top - \boldsymbol{I}_a/a$ and $\otimes$ denoting the Kronecker product. This hypothesis was, for example, investigated in Jennrich (1970), while for $d = 2$, this includes the problem of testing the null hypothesis

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

of equal correlations $\rho_i = \text{Corr}(X_{i11}, X_{i12}), i = 1, \ldots, a$ within (1). This again contains testing equality of correlations between two groups, see, e.g. Sakaori (2002), Gupta and Xu (2006) or Omelka and Pauly (2012).

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

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

A test procedure for zero correlations was introduced by Bartlett (1951) for the one-group setting. More hypotheses on the structure of the covariance matrix can be found, for example, in Joereskog (1978), Steiger (1980) and Wu et al. (2018).

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

$$\mathcal{H}_0^{\boldsymbol{r}} : \boldsymbol{R}_1 = \boldsymbol{R} \text{ resp. } \mathcal{H}_0^{\boldsymbol{r}} : \boldsymbol{I}_{p_u}\boldsymbol{r}_1 = \text{vech}^-(\boldsymbol{R}).$$

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

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

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

(d) **Testing for equal correlations** For $a = 1$, we are interested whether the correlation between all components is the same, i.e.

$$\mathcal{H}_0^{\boldsymbol{r}} : \boldsymbol{r}_{11} = \boldsymbol{r}_{12} = ... = \boldsymbol{r}_{1p_u} \text{ resp. } \mathcal{H}_0^{\boldsymbol{r}} : \boldsymbol{P}_{p_u}\boldsymbol{r}_1 = \boldsymbol{0}_{p_u}.$$

This kind of hypothesis is connected with a compound symmetry matrix and was investigated in Wilks (1946) and Box (1950) for special settings.

3 Asymptotics regarding the vectorized correlation

To infer null hypotheses of the kind $\mathcal{H}_0^r : \boldsymbol{Cr} = \boldsymbol{\zeta}$, it is necessary first to investigate the asymptotic distribution of $\boldsymbol{C}\widehat{\boldsymbol{r}}$, while $\widehat{\boldsymbol{r}}$ is the pooled vector of upper-half-vectorized empirical correlation matrices. Thereto, Theorem 3.1 from Sattler et al. (2022) is shortly repeated first. Therein, for $\boldsymbol{v} = (\boldsymbol{v}_1^\top, ..., \boldsymbol{v}_a^\top)^\top = (\text{vech}(\boldsymbol{V}_1)^\top, ..., \text{vech}(\boldsymbol{V}_a)^\top)^\top$ and with empirical covariance matrices $\widehat{\boldsymbol{V}}_i$ and $\widehat{\boldsymbol{v}} = (\widehat{\boldsymbol{v}}_1^\top, ..., \widehat{\boldsymbol{v}}_a^\top)^\top = (\text{vech}(\widehat{\boldsymbol{V}}_1)^\top, ..., \text{vech}(\widehat{\boldsymbol{V}}_a)^\top)^\top$, it holds

$$\sqrt{N}\boldsymbol{C}(\widehat{\boldsymbol{v}} - \boldsymbol{v}) \xrightarrow{\mathcal{D}} \mathcal{N}_{ap}\left(\boldsymbol{0}_{ap}, \boldsymbol{C}\boldsymbol{\Sigma}\boldsymbol{C}^\top\right). \tag{2}$$

Here, the covariance matrix is defined as $\boldsymbol{\Sigma} = \bigoplus_{i=1}^a \frac{1}{k_i}\boldsymbol{\Sigma}_i$, where $\bigoplus$ denotes the direct sum and $\boldsymbol{\Sigma}_i = \text{Cov}(\text{vech}(\boldsymbol{\epsilon}_{i1}\boldsymbol{\epsilon}_{i1}^\top)^\top)$ for $i = 1, \ldots, a$. First, some additional matrices have to be defined to use this result for correlation matrices. Let $\boldsymbol{e}_{k,p} = (\delta_{k\ell})_{\ell=1}^p$ define the p-dimensional vector, which contains a one in the k-th component and zeros elsewhere.

Moreover, we need a d-dimensional auxiliary vector $\boldsymbol{a} = (a_1, ..., a_d)$, given through $a_k = 1 + \sum_{j=1}^{k-1}(d + 1 - j)$, $k = 1, ..., d$. It contains the position of components in the half-vectorized matrix, which are the diagonal elements of the original matrix. In accordance with this, we define the p_u-dimensional vector $\boldsymbol{b}$, which contains the numbers from one to p in ascending order without the elements from $\boldsymbol{a}$. This vector $\boldsymbol{b}$ contains the position of components in the half-vectorized matrix, which are non-diagonal elements.

With these vectors, we are able to define a $d \times d$ matrix $\boldsymbol{H} = \boldsymbol{1}_d\boldsymbol{a}$ and the vectors $\boldsymbol{h}_1 = \text{vech}^-(\boldsymbol{H})$ and $\boldsymbol{h}_2 = \text{vech}^-(\boldsymbol{H}^\top)$. Finally, we can define the matrices

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

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

With all these matrices, a connection can be found between $\sqrt{n_i}(\widehat{\boldsymbol{v}}_i - \boldsymbol{v}_i)$ and $\sqrt{n_i}(\widehat{\boldsymbol{r}}_i - \boldsymbol{r}_i)$, which allows getting the requested result by applying (2). The approach to connect vectorized correlation and vectorized covariance is based on Browne and Shapiro (1986) and Nel (1985) and adapted to the current more general setting.

Theorem 3.1 *With the previously defined matrices $\boldsymbol{L}_p^u$, $\boldsymbol{M}_1$ and*

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

it holds

$$\sqrt{n_i}(\widehat{\boldsymbol{r}}_i - \boldsymbol{r}_i) = \boldsymbol{M}(\boldsymbol{v}_i, \boldsymbol{r}_i)\sqrt{n_i}(\widehat{\boldsymbol{v}}_i - \boldsymbol{v}_i) + \mathcal{O}_P(1).$$

Thus,

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

and if Assumption (A1) is fulfilled also

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

Remark 3.1 The asymptotic normality of $\sqrt{n_i}(\widehat{\boldsymbol{r}}_i - \boldsymbol{r}_i)$ was also shown for similar statistics in the past. But through the existing relation between $\sqrt{n_i}(\widehat{\boldsymbol{v}}_i - \boldsymbol{v}_i)$ and $\sqrt{n_i}(\widehat{\boldsymbol{r}}_i - \boldsymbol{r}_i)$ it is, for example, possible to express $\boldsymbol{\Upsilon}$ as a function of $\boldsymbol{\Sigma}$, which allows to construct an estimator $\widehat{\boldsymbol{\Upsilon}}$ using $\widehat{\boldsymbol{\Sigma}}$. In Sects. 5 and 6 we show that this relation provides further opportunities.

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

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

and $\widehat{\boldsymbol{\Upsilon}} := \bigoplus_{i=1}^{a} \frac{N}{n_i}\widehat{\boldsymbol{\Upsilon}}_i$. To this end, $\widehat{\boldsymbol{\Sigma}}_i$ has to be a consistent estimator for $\boldsymbol{\Sigma}_i$. This is fulfilled, e.g. for the estimator from Sattler et al. (2022), given through

$$\frac{1}{n_i - 1}\sum_{k=1}^{n_i}\left[\operatorname{vech}\left(\widetilde{\boldsymbol{X}}_{ik}\widetilde{\boldsymbol{X}}_{ik}^\top - \sum_{\ell=1}^{n_i}\frac{\widetilde{\boldsymbol{X}}_{i\ell}\widetilde{\boldsymbol{X}}_{i\ell}^\top}{n_i}\right)\right]\left[\operatorname{vech}\left(\widetilde{\boldsymbol{X}}_{ik}\widetilde{\boldsymbol{X}}_{ik}^\top - \sum_{\ell=1}^{n_i}\frac{\widetilde{\boldsymbol{X}}_{i\ell}\widetilde{\boldsymbol{X}}_{i\ell}^\top}{n_i}\right)\right]^\top,$$

with $\widetilde{\boldsymbol{X}}_{ik} := \boldsymbol{X}_{ik} - \overline{\boldsymbol{X}}_i$. As continuous functions of consistent estimators, the estimators $\widehat{\boldsymbol{\Upsilon}}_i$ are consistent. With this asymptotic result, test statistics based on quadratic forms can be formulated through:

Theorem 3.2 *Let $\boldsymbol{E}_N \in \mathbb{R}^{m\times m}$ fulfilling $\boldsymbol{E}_N \xrightarrow{\mathcal{P}} \boldsymbol{E}$, where $\boldsymbol{E} \in \mathbb{R}^{m\times m}$ is symmetric. Then, under the null hypothesis $\mathcal{H}_0^r : \boldsymbol{C}\boldsymbol{r} = \boldsymbol{\zeta}$, the quadratic form $\widehat{Q}_{\boldsymbol{r}}$, defined by*

$$\widehat{Q}_{\boldsymbol{r}} = N\left[\boldsymbol{C}\widehat{\boldsymbol{r}} - \boldsymbol{\zeta}\right]^\top \boldsymbol{E}_N\left[\boldsymbol{C}\widehat{\boldsymbol{r}} - \boldsymbol{\zeta}\right]$$

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

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

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

Thus, quadratic-form-type test statistics can be formulated, similar to Sattler et al. (2022) also for the vectorized correlation matrix. Here, we often consider symmetric matrices $\boldsymbol{E}_N$, which can be written as function of $\boldsymbol{C}$ and $\widehat{\boldsymbol{\Upsilon}}$ fullfilling $\boldsymbol{E}_N(\boldsymbol{C}, \widehat{\boldsymbol{\Upsilon}}) = \boldsymbol{E}_N \xrightarrow{\mathcal{P}} \boldsymbol{E} = \boldsymbol{E}(\boldsymbol{C}, \boldsymbol{\Upsilon})$. Examples cover:

$$\begin{aligned}\text{ATS}_{\boldsymbol{r}} &= N\,[\boldsymbol{C}(\widehat{\boldsymbol{r}} - \boldsymbol{r})]^\top\,[\boldsymbol{C}(\widehat{\boldsymbol{r}} - \boldsymbol{r})]\,/\,\operatorname{tr}(\boldsymbol{C}\widehat{\boldsymbol{\Upsilon}}\boldsymbol{C}^\top),\\ \text{WTS}_{\boldsymbol{r}} &= N\,[\boldsymbol{C}(\widehat{\boldsymbol{r}} - \boldsymbol{r})]^\top\,(\boldsymbol{C}\widehat{\boldsymbol{\Upsilon}}\boldsymbol{C}^\top)^+\,[\boldsymbol{C}(\widehat{\boldsymbol{r}} - \boldsymbol{r})]\,.\end{aligned}$$

This leads to an asymptotic valid test $\varphi_{WTS_{\boldsymbol{r}}} = 1\!\!1\{WTS_{\boldsymbol{r}} \notin (-\infty, \chi^2_{\operatorname{rank}(\boldsymbol{C});1-\alpha}]\}$, in case of $\boldsymbol{\Upsilon} > 0$. This is, however, hard to verify due to its complex structure. We therefore do not treat the WTS in the following.

Conversely, the ATS is no asymptotic pivot, but the simulation results from Sattler et al. (2022) suggest to use its Monte Carlo version. Hereto, the according matrices from Theorem 3.2 were estimated, and by this, we also get the estimated eigenvalues. By generating $B_1, \ldots B_{ap_u}$, the weighted sum can be calculated, and by repeating this frequently (e.g. W=10,000 times), the required α quantiles of the weighted χ_1^2 distribution can be estimated, denoted by q_α^{MC}. For example, this gives us the test $\varphi_{\text{ATS}_{\boldsymbol{r}}} := 1\!\!1\{\text{ATS}_{\boldsymbol{r}} \notin (-\infty, q_{1-\alpha}^{\text{MC}}]\}$, and a similar approach can be used for all quadratic forms fulfilling Theorem 3.2.

4 Resampling procedures

A resampling procedure may be useful, on the one hand, for a better small sample approximation and, on the other hand, for quadratic forms with critical values that are difficult to calculate. Since the simulations from Sattler et al. (2022) showed clear advantages of the parametric bootstrap, we focus on this approach but also present a wild bootstrap approach in the supplement. For every group $\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i}, i = 1, \ldots, a$, we calculate the covariance matrix $\widehat{\boldsymbol{\Upsilon}}_i$. With this covariance matrix, we generate i.i.d. random vectors $\boldsymbol{Y}_{i1}^\dagger, \ldots, \boldsymbol{Y}_{in_i}^\dagger \sim \mathcal{N}_{p_u}(\boldsymbol{0}_{p_u}, \widehat{\boldsymbol{\Upsilon}}_i)$ which are independent of the realizations and calculate their sample covariances $\widehat{\boldsymbol{\Upsilon}}_i^\dagger$, respectively, $\widehat{\boldsymbol{\Upsilon}}^\dagger := \bigoplus_{i=1}^a \frac{N}{n_i}\widehat{\boldsymbol{\Upsilon}}_i^\dagger$. For these random vectors, we now consider the asymptotic distribution.

Theorem 4.1 *If Assumption (A1) is fulfilled, it holds: Given the data, the conditional distribution of*

(a) $\sqrt{N}\,\overline{\boldsymbol{Y}}_i^\dagger$, *for* $i = 1, \ldots, a$, *converges weakly to* $\mathcal{N}_{p_u}\left(\boldsymbol{0}_{p_u}, \kappa_i{}^{-1}\boldsymbol{\Upsilon}_i\right)$ *in probability.*
(b) $\sqrt{N}\,\overline{\boldsymbol{Y}}^\dagger$ *converges weakly to* $\mathcal{N}_{ap_u}\left(\boldsymbol{0}_{ap_u}, \boldsymbol{\Upsilon}\right)$ *in probability.*

Moreover, we have $\widehat{\boldsymbol{\Upsilon}}_i^\dagger \xrightarrow{\mathcal{P}} \boldsymbol{\Upsilon}_i$ *and* $\widehat{\boldsymbol{\Upsilon}}^\dagger \xrightarrow{\mathcal{P}} \boldsymbol{\Upsilon}$. *Thus, the unknown covariance matrices can be estimated through these estimators.*

In consequence of Theorem 4.1, it is reasonable to calculate the bootstrap version of the previous quadratic forms,

$$Q_{\boldsymbol{r}}^\dagger = N[\boldsymbol{C}\overline{\boldsymbol{Y}}^\dagger]^\top \boldsymbol{E}[\boldsymbol{C}\,\overline{\boldsymbol{Y}}^\dagger] \quad \text{resp.} \quad Q_{\boldsymbol{r}}^\dagger = N[\boldsymbol{C}\overline{\boldsymbol{Y}}^\dagger]^\top \boldsymbol{E}_N(\boldsymbol{C}, \widehat{\boldsymbol{\Upsilon}}^\dagger)[\boldsymbol{C}\,\overline{\boldsymbol{Y}}^\dagger].$$

The first can be used if the limiting matrix $\boldsymbol{E} = \boldsymbol{E}(\boldsymbol{C}, \boldsymbol{\Upsilon})$ of $\boldsymbol{E}_N(\boldsymbol{C}, \widehat{\boldsymbol{\Upsilon}})$ is known. In the second version $\boldsymbol{E} = \boldsymbol{E}(\boldsymbol{C}, \boldsymbol{\Upsilon})$ is estimated via $\boldsymbol{E}_N = \boldsymbol{E}_N(\boldsymbol{C}, \widehat{\boldsymbol{\Upsilon}}^{\dagger})$ which is reasonable if from $\boldsymbol{A}_n \overset{\mathcal{P}}{\to} \boldsymbol{A}$, for random matrices $\boldsymbol{A}_n$ and $\boldsymbol{A}$ it follows $\boldsymbol{E}_N(\boldsymbol{C}, \boldsymbol{A}_n) \overset{\mathcal{P}}{\to} \boldsymbol{E}(\boldsymbol{C}, \boldsymbol{A})$ [(see e.g. van der Vaart and Wellner (1996)]. This is fulfilled, if the function is continuous in its second component which holds for the trace operator used in the ATS and which we assume in what follows.

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

Corollary 4.1 *For each parameter vector $\boldsymbol{r} \in \mathbb{R}^{ap_u}$ and $\boldsymbol{r}_0 \in \mathbb{R}^{ap_u}$ with $\boldsymbol{C}\boldsymbol{r}_0 = \boldsymbol{\zeta}$, under Assumption (A1), we have*

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

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

This motivates the definition of bootstrap tests as $\varphi_{Q_{\boldsymbol{r}}}^{\dagger} := 1\!\!1\{Q_{\boldsymbol{r}}^{\dagger} \notin (-\infty, c_{Q_{\boldsymbol{r}}^{\dagger},1-\alpha}]\}$, where $c_{Q_{\boldsymbol{r}}^{\dagger},1-\alpha}$ denotes the conditional quantile of $Q_{\boldsymbol{r}}^{\dagger}$. The above results ensure that these tests are of asymptotic level α, and we will use it for the ATS.

Fisher z-transformed vectors are often used in the analysis of correlation matrices instead of the original vectorized correlation matrices. Although the root of this approach is the distribution of the Fisher z-transformed correlation in the case of normally distributed observations, it is also used for tests without this distributional restriction, see e.g. Steiger (1980). In principle, we could also consider our tests together with the transformed vector. This approach assumes that all components of $\boldsymbol{\zeta}$ differ from one, which is always possible to ensure. We can define tests based on the transformation for each of our quadratic forms, including the tests based on bootstrap or Monte Carlo simulations. Our simulations showed that the tests based on this transformation have more liberal behaviour than the original one, which was also mentioned in Omelka and Pauly (2012) for the case of bivariate permutation tests. Since all of our test statistics were already a bit liberal, it is not useful to consider these versions further. However, some details on this are given in the supplement. Instead, we propose an additional Taylor approximation of higher order as described below.

5 Higher order of Taylor approximation

The main result of Sect. 3 is the connection between the vectorized empirical covariance matrix and the vectorized empirical correlation matrix given through

$$\sqrt{n_i}(\widehat{\boldsymbol{r}}_i - \boldsymbol{r}_i) = \boldsymbol{M}(\boldsymbol{v}_i, \boldsymbol{r}_i)\sqrt{n_i}(\widehat{\boldsymbol{v}}_i - \boldsymbol{v}_i) + \mathcal{O}_P(1).$$

Since this is based on a Taylor approximation, it leads to the question of whether a higher order of approximation could be useful here. In this way, we get

$$\sqrt{n_i}(\widehat{\boldsymbol{r}}_i - \boldsymbol{r}_i) = \boldsymbol{M}(\widehat{\boldsymbol{v}}_i, \widehat{\boldsymbol{r}}_i)\boldsymbol{Y}_i + f_{\widehat{\boldsymbol{v}}_i, \widehat{\boldsymbol{R}}_i}(\boldsymbol{Y}_i)/\sqrt{n_i} + \mathcal{O}_P\left(\sqrt{n_i^{-1}}\right) \tag{3}$$

with $\boldsymbol{Y}_i \sim \mathcal{N}_p\left(\boldsymbol{0}_p, \boldsymbol{\Sigma}_i\right)$ and a function $f_{\widehat{\boldsymbol{v}}_i, \widehat{\boldsymbol{R}}_i} : \mathbb{R}^p \to \mathbb{R}^{p_u}$. This equation's derivation, together with the complex concrete form of the function $f_{\widehat{\boldsymbol{v}}_i, \widehat{\boldsymbol{R}}_i}$, can be found in the supplementary material. Although the part $f_{\widehat{\boldsymbol{v}}_i, \widehat{\boldsymbol{R}}_i}(\boldsymbol{Y}_i)/\sqrt{n_i}$ is asymptotically negligible, for small sample sizes, it can affect the performance of the corresponding test. For this purpose, we propose an additional Monte Carlo-based approach with this Taylor approximation. Thereto for each group, we generate $\boldsymbol{Y}_i \sim \mathcal{N}_p(\boldsymbol{0}_p, \widehat{\boldsymbol{\Sigma}}_i)$ and transform it to $\boldsymbol{Y}_i^{\text{Tay}} := \boldsymbol{M}(\widehat{\boldsymbol{v}}_i, \widehat{\boldsymbol{r}}_i)\boldsymbol{Y}_i + f_{\widehat{\boldsymbol{v}}_i, \widehat{\boldsymbol{R}}_i}(\boldsymbol{Y}_i)/\sqrt{n_i}$. So with the pooled vector $\boldsymbol{Y}^{\text{Tay}} = (n_1^{-1/2}{\boldsymbol{Y}_1^{\text{Tay}}}^\top, ..., n_a^{-1/2}{\boldsymbol{Y}_a^{\text{Tay}}}^\top)^\top$, we get a version of this quadratic form by calculating

$$Q_{\boldsymbol{r}}^{\text{Tay}} = N[\boldsymbol{C}\boldsymbol{Y}^{\text{Tay}}]^\top \boldsymbol{E}_N(\boldsymbol{C}, \widehat{\boldsymbol{\Upsilon}})[\boldsymbol{C}\boldsymbol{Y}^{\text{Tay}}].$$

If we repeat this frequently enough, an asymptotic correct level α test can be constructed by comparing $\widehat{Q}_{\boldsymbol{r}}$ with the empirical $(1-\alpha)$ quantile of $Q_{\boldsymbol{r}}^{\text{Tay}}$. Since there it is only necessary to calculate $\widehat{\boldsymbol{\Upsilon}}$ one time, this approach is clearly less time-consuming than the corresponding bootstrap approaches.

It would also be possible to use a Taylor approximation of higher order. Since the transformation gets more complex and our simulations suggest that this only affects very small sample sizes, we renounce this here.

6 Combined tests for variances and correlations

The equality of two covariance matrices and the equality of two correlation matrices are interesting null hypotheses that are closely related. In this section, we exemplify how our methodology extends to a simultaneous comparison of variances and correlations. To this end, we combine the above results with those from Sattler et al. (2022) to develop a so-called multiple contrast test. For this, we define the statistic

$$\boldsymbol{T} = \begin{pmatrix} T_1 \\ \vdots \\ T_p \end{pmatrix} = \sqrt{N}\left(\begin{pmatrix} \widehat{V}_{111} \\ \widehat{V}_{122} \\ \vdots \\ \widehat{V}_{1dd} \\ \widehat{\boldsymbol{r}}_1 \end{pmatrix} - \begin{pmatrix} \widehat{V}_{211} \\ \widehat{V}_{222} \\ \vdots \\ \widehat{V}_{2dd} \\ \widehat{\boldsymbol{r}}_2 \end{pmatrix}\right)$$

which fulfils $\boldsymbol{T} \xrightarrow{\mathcal{D}} \mathcal{N}_p(\boldsymbol{0}_p, \boldsymbol{\Gamma})$ under the null hypothesis $\mathcal{H}_0^{\boldsymbol{v}} : \boldsymbol{V}_1 = \boldsymbol{V}_2$. The concrete form of the asymptotic covariance matrix $\boldsymbol{\Gamma}$ and its derivation can be found in the online supplement. In case its diagonal elements are positive, i.e.

$\Gamma_{11}, ..., \Gamma_{pp} > 0$, we consider the vector of studentized test statistics $\widetilde{\boldsymbol{T}} := \operatorname{diag}(\Gamma_{11}, ..., \Gamma_{pp})^{-1/2}\boldsymbol{T}$. Under the null hypothesis of equal covariances, $\widetilde{\boldsymbol{T}}$ is asymptotically normal distributed with expectation vector $\mathbf{0}_p$ and covariance/correlation matrix $\widetilde{\boldsymbol{\Gamma}} := \operatorname{diag}(\Gamma_{11}, ..., \Gamma_{pp})^{-1/2}\boldsymbol{\Gamma}\operatorname{diag}(\Gamma_{11}, ..., \Gamma_{pp})^{-1/2}$. Following the concept of multiple contrast tests (see e.g. Bretz et al. 2001 or Konietschke et al. 2013 for more details), we now consider $\max(|\widetilde{T}_1|, ..., |\widetilde{T}_p|)$ as test statistic and use the two-sided $(1-\alpha)$ equicoordinate quantile of the above limiting distribution $\mathcal{N}_p(\mathbf{0}_p, \widetilde{\boldsymbol{\Gamma}})$ as the critical value. Denoting this as $z_{1-\alpha,2,\widetilde{\boldsymbol{\Gamma}}}$, we obtain an asymptotic level α multiple contrast test by $\max(|\widetilde{T}_1|, ..., |\widetilde{T}_p|) \geq z_{1-\alpha,2,\widetilde{\boldsymbol{\Gamma}}}$. This multiple contrast test fulfils the desired properties, i.e. it allows us to simultaneously infer the hypotheses of equal covariances and equal correlations. In fact, if $\max(|\widetilde{T}_{d+1}|, ..., |\widetilde{T}_p|) \geq z_{1-\alpha,2,\widetilde{\boldsymbol{\Gamma}}}$, we can conclude that both groups have significantly different dependence structures and, therefore, neither equal covariance matrices nor equal correlation matrices. However, if $\max(|\widetilde{T}_{d+1}|, ..., |\widetilde{T}_p|) < z_{1-\alpha,2,\widetilde{\boldsymbol{\Gamma}}}$ but $\max(|\widetilde{T}_1|, ..., |\widetilde{T}_d|) \geq z_{1-\alpha,2,\widetilde{\boldsymbol{\Gamma}}}$, this indicates that both groups have the same dependence structure but different variances, i.e. we can (only) conclude that they have significantly different covariance matrices. Moreover, the information which components of $\widetilde{\boldsymbol{T}}$ exceed $z_{1-\alpha,2,\widetilde{\boldsymbol{\Gamma}}}$, in absolute value, helps to clarify where the departures from $\mathcal{H}_0$ might come from. We note that in all of these considerations, we use the same critical value $z_{1-\alpha,2,\widetilde{\boldsymbol{\Gamma}}}$ as common with multiple contrast tests.

Unfortunately, the condition $\Gamma_{11}, ..., \Gamma_{pp} > 0$ is difficult to verify. A bootstrap approach can again resolve this, as was done in Friedrich and Pauly (2017) or Umlauft et al. (2019). However, we introduce an alternative Monte Carlo approach because it is easier to explain and already shows good small sample approximations as will be demonstrated in Sect. 7. To this end, we use that with matrix $\boldsymbol{A} \in \mathbb{R}^{d\times p}$ given by $\boldsymbol{A} = \sum_{\ell=1}^{d} \boldsymbol{e}_{\ell,d}\, \boldsymbol{e}_{\boldsymbol{a}_\ell,p}^\top$ it holds that

$$\sqrt{n_i}\left(\begin{pmatrix}\widehat{V}_{i11}\\ \widehat{V}_{i22}\\ \vdots\\ \widehat{V}_{idd}\\ \widehat{\boldsymbol{r}}_i\end{pmatrix} - \begin{pmatrix}V_{i11}\\ V_{i22}\\ \vdots\\ V_{idd}\\ \boldsymbol{r}_i\end{pmatrix}\right) = \begin{pmatrix}\boldsymbol{A}\\ \boldsymbol{M}(\boldsymbol{v}_i, \boldsymbol{r}_i)\end{pmatrix}\boldsymbol{Y}_i + \frac{1}{\sqrt{n_i}}\begin{pmatrix}\mathbf{0}_d\\ f_{\widehat{\boldsymbol{v}}_i,\widehat{\boldsymbol{R}}_i}(\boldsymbol{Y}_i)\end{pmatrix} + \mathcal{O}_P\left(\sqrt{n_i^{-1}}\right)$$

with $\boldsymbol{Y}_i \sim \mathcal{N}_p(\mathbf{0}_p, \boldsymbol{\Sigma}_i)$. Based on the above expansion, we generate $\boldsymbol{Y}_i \sim \mathcal{N}_p(\mathbf{0}_p, \widehat{\boldsymbol{\Sigma}}_i)$ similarly to Sect. 5 and transform it to

$$\boldsymbol{Y}_i^{\mathrm{Tay}} := \begin{pmatrix}\boldsymbol{A}\\ \boldsymbol{M}(\widehat{\boldsymbol{v}}_i, \widehat{\boldsymbol{r}}_i)\end{pmatrix}\boldsymbol{Y}_i + \frac{1}{\sqrt{n_i}}\begin{pmatrix}\mathbf{0}_d\\ f_{\widehat{\boldsymbol{v}}_i,\widehat{\boldsymbol{R}}_i}(\boldsymbol{Y}_i)\end{pmatrix}.$$

With this transformed vectors, we calculate $\boldsymbol{T}^{1,\mathrm{Tay}} = \sqrt{N}(n_1^{-1/2}\boldsymbol{Y}_1^{\mathrm{Tay}} - n_2^{-1/2}\boldsymbol{Y}_2^{\mathrm{Tay}})$, which has the same asymptotic distribution as $\boldsymbol{T}$, and repeat this B times to get $\boldsymbol{T}^{1,\mathrm{Tay}}, ..., \boldsymbol{T}^{B,\mathrm{Tay}}$. To obtain an appropriate multiple contrast test based on these values, we follow Munko et al. (2023b) resp. Munko et al. (2023a), where a corresponding procedure for functional data was developed based on Buhlmann (1998). Thus, for

$\ell = 1, ..., p$, we denote by $q_{\ell,\beta}$ the empirical β quantile of $T_\ell^{1,\mathrm{Tay}}, ..., T_\ell^{B,\mathrm{Tay}}$. To control the family-wise type-I-error rate by α, the appropriate β can be found through

$$\widetilde{\beta} = \max\left(\beta \in \mathcal{B} \middle| \sum_{b=1}^{B} \max_{\ell=1,\dots,p} \frac{\left(1 - 1\!\!1\left(q_{\ell,\frac{\beta}{2}} \leq T_\ell^{b,\mathrm{Tay}} \leq q_{\ell,1-\frac{\beta}{2}}\right)\right)}{B} \leq \alpha\right)$$

with $\mathcal{B} = \left\{0, \frac{1}{B}, ..., \frac{B-1}{B}\right\}$. An asymptotic level-$\alpha$ test of the null hypothesis of equal correlation matrices then rejects the null if and only if

$$\max_{\ell=1,\dots,p}\left(1 - 1\!\!1\left(q_{\ell,\frac{\widetilde{\beta}}{2}} \leq T_\ell \leq q_{\ell,1-\frac{\widetilde{\beta}}{2}}\right)\right) = 1$$

or equivalently if

$$\max_{\ell=1,\dots,p}\left(\max\left(\frac{T_\ell}{q_{\ell,1-\frac{\widetilde{\beta}}{2}}}, \frac{q_{\ell,\frac{\widetilde{\beta}}{2}}}{T_\ell}\right)\right) > 1.$$

Each component of the vector-valued test statistic $\boldsymbol{T}$ is treated in the same way since the same $\widetilde{\beta} < \alpha$ is used. This procedure allows for the same conclusions about the reason for the rejection as the above approach based on the equicoordinate quantile.

Of course, this combined test could be generalized to compare more than two groups by using an appropriate Tukey-type contrast matrix, see e.g. Konietschke et al. (2013).

7 Simulations

We analyse the type-I-error rate and power for different hypotheses to investigate the performance. Here, we focus on hypotheses with an implemented algorithm to have an appropriate competitor. In the R-package *psych* by Revelle (2019) several of these tests are included, like $\varphi_{Jennrich}$ from Jennrich (1970) and $\varphi_{\mathrm{Steiger}_{\mathrm{Fz}}}$ from Steiger (1980) for equality of correlation matrices. Testing whether the correlation matrix is equal to the identity matrix can be investigated with $\varphi_{\mathrm{Bartlett}}$ from Bartlett (1951) and again with $\varphi_{\mathrm{Steiger}_{\mathrm{Fz}}}$ resp. $\varphi_{\mathrm{Steiger}}$. Hereby, $\varphi_{\mathrm{Steiger}_{\mathrm{Fz}}}$ is the same test statistic as $\varphi_{\mathrm{Steiger}}$ but uses a Fisher z-transformation on the vectorized correlation matrices. Therefore, we consider the following hypotheses:

$A_{\boldsymbol{r}}$: Homogeneity of correlation matrices: $\mathcal{H}_0^{\boldsymbol{r}} : \boldsymbol{R}_1 = \boldsymbol{R}_2$,
$B_{\boldsymbol{r}}$: Diagonal structure of the covariance matrix $\mathcal{H}_0^{\boldsymbol{r}} : \boldsymbol{R}_1 = \boldsymbol{I}_p$ resp. $\boldsymbol{r}_1 = \boldsymbol{0}_{p_u}$,

with $\alpha = 0.05$. Further hypotheses are investigated in the supplementary material together with other settings and dimensions. The hypothesis matrices are chosen as the projection matrices $\boldsymbol{C}(A_{\boldsymbol{r}}) = \boldsymbol{P}_2 \otimes \boldsymbol{I}_{p_u}$ and $\boldsymbol{C}(B_{\boldsymbol{r}}) = \boldsymbol{I}_{p_u}$, while $\boldsymbol{\zeta}$ is in both cases a zero vector with appropriate dimension. Based on this the results of Sattler and Zimmermann (2023) can also be simplified to be computational more efficient without affecting the results.

We use 1000 bootstrap steps for our parametric bootstrap, 10,000 simulation steps for the Monte Carlo approach and 10,000 runs for all tests to get reliable results. For φ_{Steiger} and $\varphi_{\text{Steiger}_{\text{Fz}}}$ from Revelle (2019), the actual test statistic is multiplied with the factor $(N-3)/N$. This approach is based on a specific result of the Fisher z-transformation of the correlation vector of normally distributed random vectors for small sample sizes. Asymptotically, this factor has no influence, but it is also used for φ_{Steiger}, where no Fisher z-transformation was done.

To get a better impression of the impact of such a multiplication, we also include our ATS with parametric bootstrap using such multiplication and denote this with an m for multiplication. This also simplifies the comparison of the tests under equal conditions. At last, we simulated Monte Carlo-based ATS using our Fisher z-transformation. We denote it by ATSFz, while an additional version ATSFz-m is simulated, which is formed by multiplication with the factor $(N-3)/N$.

To have a comparative setting to Sattler et al. (2022), we used d=5 and therefore $p_u = 10$, while for one group, we have $n \in \{25, 50, 125, 250\}$, and for two groups, we have $n_1 = 0.6 \cdot N$ and $n_2 = 0.4 \cdot N$ with $N \in \{50, 100, 250, 500\}$. We considered 5-dimensional observations generated independently according to the model $\boldsymbol{X}_{ik} = \vec{\mu}_i + \boldsymbol{V}^{1/2}\boldsymbol{Z}_{ik}, i = 1, \ldots, a, k = 1, \ldots, n_i$ with $(\vec{\mu}_1 = (1^2, 2^2, ..., 5^2)/4^\top)$, $(\vec{\mu}_2 = \mathbf{0}_5)$ and error terms based on

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

For $A_{\boldsymbol{r}}$, we use $(\boldsymbol{V}_1)_{ij} = 1 - |i-j|/2d$ for the first group, and for the second group, we multiply this covariance matrix with diag(1, 1.2, ..., 1.8). Thus, we have a setting where the covariance matrices are different, but the correlation matrices are equal. To investigate $B_{\boldsymbol{r}}$, we use the matrix, $\boldsymbol{V} = \text{diag}(1, 1.2, ..., 1.8)$.

7.1 Type-I-error

The results of hypothesis $A_{\boldsymbol{r}}$ can be seen in Table 1 for the Toeplitz covariance matrix. Here, the values in the 95% binomial interval [0.0458, 0.0543] are printed in bold. It is interesting to note that the type-I-error rate of $\varphi_{Steifer_{Fz}}$ and $\varphi_{Jennrich}$ differs more and more from the 5% rate for increasing sample sizes. Therefore, these tests should not be used, at least for our setting. In contrast, all of our tests are too liberal but show a substantially better type-I-error rate. Moreover, these tests fulfil Bradley's liberal criterion (from Bradley 1978) for N larger than 50. This criterion is often consulted by applied statisticians, for example, in quantitative psychology. It states that a procedure is 'acceptable' if the type-I-error rate is between 0.5α and 1.5α.

Similar to the results for covariance matrices, the bootstrap version has slightly better results than Monte Carlo-based tests, while the error rates get closer to the nominal level for larger sample sizes. In each setting, the Monte Carlo-based ATS formulated for Fisher z-transformed vectors has a better performance than the classical one, especially for smaller sample sizes.

It can be seen that a correction factor, like $(N-3)/N$, which was used for ATS-Par-m, clearly improves the small sample performance. In contrast to the results from Sattler et al. (2022), for testing correlation matrices, the small sample approximation seems a bit worse, making such a correction factor useful. The best performance for smaller sample sizes can be seen for our Taylor-based Monte Carlo approach, which has for $N > 50$ an error rate within the 95% interval. Therefore, for this setting, we recommend ATS-Tay, while for larger sample size, ATS-Par-m and ATSFz-m also show good results.

For hypothesis $\boldsymbol{B_r}$, the results are included in Table 2. Here, our Taylor-based approach has slightly conservative values, and therefore ATS-Para-m and ATSFz-m have the best results of our test statistics, while the results are slightly better than for hypothesis $\boldsymbol{A_r}$. For example, the number of values in the 95% binomial interval $[0.0458, 0.0543]$, printed bold, clearly increases. With the correction factor, these tests have a better type-I-error rate than φ_{Steiger} in all settings and are comparable to $\varphi_{\text{Steiger}_{\text{Fz}}}$.

Nevertheless, $\varphi_{\text{Bartlett}}$ is a test only developed for this special hypothesis and therefore has an excellent error rate through all distributions. But in addition to the type-I error rate, also other properties are highly relevant.

7.2 Power

The ability to detect deviations from the null hypothesis is also an important test criterion. To this aim, we also investigate the power of some of the tests mentioned above. We choose a quite simple kind of alternative suitable for our situation. As covariance matrix, we consider $\boldsymbol{V}_1 + \delta \cdot \boldsymbol{J}_d$ for $\delta \in [0, 3.5]$ in hypothesis $\boldsymbol{A_r}$ and for $\delta \in [0, 0.75]$ in hypothesis $\boldsymbol{B_r}$. The reason for this considerable difference in the δ range is that for hypothesis $\boldsymbol{B_r}$, the second summand changes the setting from uncorrelated to correlated. For hypothesis $\boldsymbol{A_r}$, it just increases the correlations, which is clearly more challenging to detect. Due to computational reasons, we simulate only one sample size, which is $N = 100$ resp. $n_1 = 50$ and consider error terms based on skewed normal distribution, while results for the gamma distribution can be found in the supplementary material. We simulate only the tests with good results for their type-I-error rate, which were for $\boldsymbol{A_r}$ $\varphi_{\text{ATS}_r^\dagger}$ and $\varphi_{\text{ATS}_r-\text{Tay}}$ as well as $\varphi_{\text{SteigerFZ}}$ as competitor, despite its performance in Table 1. Because of the similarity of the results from the parametric bootstrap and the more classical Monte Carlo-based approach, we do not consider further test statistics here. For hypothesis $\boldsymbol{A_r}$, Fig. 1 shows that the Taylor approximation makes the test slightly less liberal, and therefore reduces the power. This effect can be seen as a shift and does not influence the slope. In general, the power of both approaches is quite good, also $\varphi_{\text{SteigerFZ}}$ has even higher power, which furthermore increases faster. But since this test becomes even more liberal for increasing sample size, this is not surprising and makes it for this setting not recommendable. Based on the results from Table 2 for hypothesis $\boldsymbol{B_r}$, we consider$\varphi_{\text{ATS}_r^\dagger}$ and $\varphi_{\text{ATS}_r-\text{Tay}}$ as well as $\varphi_{\text{SteigerFz}}$ and $\varphi_{\text{Bartlett}}$, while the setting is the same. In Fig. 1 for hypothesis

Table 1 Simulated type-I-error rates ($\alpha = 5\%$) in scenario A_r ($\mathcal{H}_0^{\boldsymbol{r}} : \boldsymbol{R}_1 = \boldsymbol{R}_2$) for ATS and Steiger's test.

	t_9				Normal				Skew normal			
N	50	100	250	500	50	100	250	500	50	100	250	500
ATS-Par	0.0755	0.0657	0.0568	**0.0538**	0.0703	0.0627	**0.0524**	**0.0519**	0.0734	0.0635	**0.0530**	0.0567
ATS-Par-m	0.0653	0.0600	0.0553	**0.0528**	0.0609	0.0579	**0.0502**	**0.0511**	0.0626	0.0602	**0.0518**	0.0557
ATS	0.0818	0.0652	0.0581	**0.0542**	0.0776	0.0640	**0.0527**	**0.0511**	0.0798	0.0666	**0.0536**	0.0567
ATS-Tay	0.0603	**0.0528**	**0.0523**	**0.0511**	0.0567	**0.0536**	**0.0482**	**0.0495**	0.0579	**0.0536**	**0.0486**	**0.0539**
ATSFz	0.0781	0.0634	0.0568	**0.0540**	0.0726	0.0628	**0.0516**	**0.0508**	0.0757	0.0659	**0.0528**	0.0562
ATSFz-m	0.0683	0.0592	0.0556	**0.0530**	0.0623	0.0582	**0.0503**	**0.0501**	0.0642	0.0604	**0.0512**	0.0554
SteigerFz	0.0931	0.1106	0.1303	0.1296	0.0908	0.1130	0.1249	0.1229	0.0902	0.1158	0.1177	0.1272

The observation vectors have dimension 5, covariance matrix $(\boldsymbol{V}_1)_{ij} = 1 - |i - j|/2d$ resp. $\boldsymbol{V}_2 = \mathrm{diag}(1, 1.2, ..., 1.8)\boldsymbol{V}_1$, and there is always the same relation between group sample size with $n_1 := 0.6 \cdot N$ resp. $n_2 := 0.4 \cdot N$

Table 2 Simulated type-I-error rates ($\alpha = 5\%$) in scenario $B_{\boldsymbol{r}}$ ($\mathcal{H}_0^{\boldsymbol{r}} : \boldsymbol{r}_1 = \boldsymbol{0}_{10}$) for ATS, Steiger's and Bartlett's test.

	t_9				Normal				Skew normal			
N	25	50	125	250	25	50	125	250	25	50	125	250
ATS-Par	0.0863	0.0620	**0.0537**	**0.0501**	0.0835	0.0605	**0.0493**	0.0559	0.0846	0.0612	**0.0543**	**0.0509**
ATS-Par-m	**0.0507**	0.0448	**0.0473**	**0.0471**	0.0455	0.0422	0.0429	**0.0517**	**0.0484**	0.0440	**0.0470**	**0.0481**
ATS	0.0974	0.0646	**0.0533**	**0.0492**	0.0940	0.0621	**0.0506**	**0.0541**	0.0961	0.0630	**0.0531**	**0.0507**
ATS-Tay	**0.0512**	0.0394	0.0417	0.0431	0.0556	0.0436	0.0417	**0.0500**	**0.0504**	0.0408	0.0433	0.0456
ATSFz	0.0916	0.0607	**0.0522**	**0.0490**	0.0901	0.0605	**0.0495**	**0.0541**	0.0914	0.0606	**0.0523**	**0.0499**
ATSFz-m	**0.0523**	**0.0459**	0.0453	**0.0458**	**0.0508**	0.0436	0.0420	**0.0512**	**0.0526**	0.0442	**0.0462**	**0.0462**
Steiger	0.0220	0.0326	0.0419	**0.0457**	0.0255	0.0355	0.0421	**0.0507**	0.0252	0.0333	**0.0458**	**0.0479**
SteigerFz	**0.0515**	**0.0471**	**0.0489**	**0.0491**	0.0558	**0.0517**	**0.0492**	**0.0536**	0.0548	**0.0486**	**0.0542**	**0.0499**
Bartlett	**0.0470**	**0.0480**	**0.0488**	**0.0500**	**0.0525**	**0.0535**	**0.0479**	**0.0539**	**0.0531**	0.0456	**0.0516**	**0.0506**

The observation vectors have dimension 5 and covariance matrix $(\boldsymbol{V}) = diag(1, 1.2, 1.4, 1.6, 1.8)$

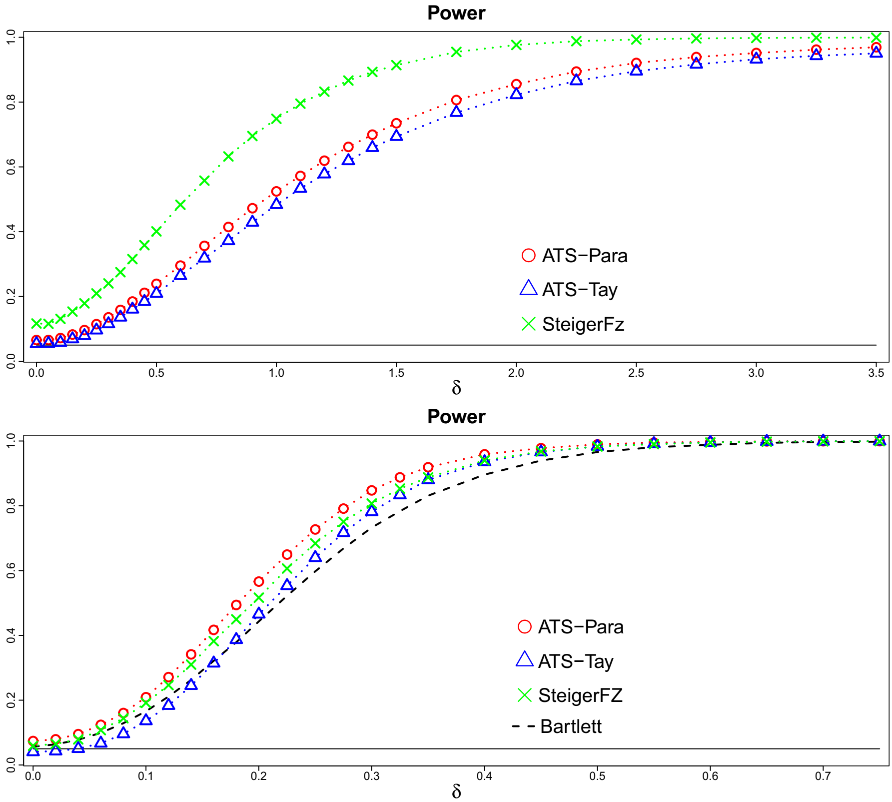

Fig. 1 Simulated power curves of different tests for the hypothesis $A_{\boldsymbol{r}}$ ($\mathcal{H}_0^{\boldsymbol{r}} : \boldsymbol{R}_1 = \boldsymbol{R}_2$) above, and hypothesis $B_{\boldsymbol{r}}$ ($\mathcal{H}_0^{\boldsymbol{r}} : \boldsymbol{r}_1 = \boldsymbol{0}_{10}$) below, for a five-dimensional skewed normal distribution. For hypothesis $A_{\boldsymbol{r}}$, it holds $n_1 = 60, n_2 = 40$ and the covariance matrix is $(\boldsymbol{V}_2)_{ij} = 1 - |i-j|/2d$ resp. $\boldsymbol{V}_1 = \boldsymbol{V}_2 + \delta \boldsymbol{J}_5$. The covariance matrix for hypothesis $B_{\boldsymbol{r}}$ is $\boldsymbol{V} = \boldsymbol{I}_5 + \delta \boldsymbol{J}_5$ and $n_1 = 50$

$B_{\boldsymbol{r}}$, it can be seen that $\varphi_{\text{ATS}_{\boldsymbol{r}}-\text{Tay}}$ has for smaller δ similar power as $\varphi_{\text{Bartlett}}$ which was specially developed for this hypothesis, while for larger deviation, the power of the Monte Carlo approach increases. Here, $\varphi_{\text{SteigerFZ}}$ has slightly less power than the parametric bootstrap due to the slightly liberal behaviour of the bootstrap approach.

The type-I error rate and the power for both hypotheses show that our developed tests are useful in many situations, although partially large sample sizes are necessary for good results. This is a known fact for testing hypotheses regarding correlation matrices, which was, for example, mentioned in Steiger (1980). Therefore, the results of ATS-Tay investigating $A_{\boldsymbol{r}}$ for smaller sample sizes are even more convincing.

All in all, the results of this section, together with the additional results from the supplement, allow us to give some general recommendations. Since SteigerFz has a type-I-error rate of more than 9% in Table 1, which also grows for increasing sample sizes, it is not useful apart from single hypotheses like $\mathcal{H}_0^{\boldsymbol{r}} : \boldsymbol{r}_1 = \boldsymbol{0}_{10}$. This hypothesis can be checked with Bartlett, which had good results but only allows this hypothesis.

Table 3 Number of observations for the different factor level combinations of sex and diagnosis

	AD	MCI	SCC+	SCC−
Male	12	27	14	6
Female	24	30	31	16

On the contrary, for all considered hypotheses, ATS-Par-m and ATS lead to good results for moderate to large sample sizes, while they are liberal for small sample sizes. Such liberality was, for example, also mentioned in Perreault et al. (2022) for tests based on Kendall's τ. As intended, the bootstrap improves the behaviour for smaller sample sizes, but not enough. In case of small sample sizes, the Taylor-based approach is recommended as it exhibited good small sample performance in all settings. Only for larger sample sizes, it is outperformed by ATS-Par-m.

8 Illustrative data analysis

To illustrate the method, we take a closer look at the EEG data set from the R-package *manova.rm* by Friedrich et al. (2019). In this study from Staffen et al. (2014), conducted at the University Clinic of Salzburg (Department of Neurology), electroencephalography (EEG) data were measured from 160 patients with different diagnoses of impairments. These are Alzheimer's disease (AD), mild cognitive impairment (MCI), and subjective cognitive complaints (SCC). Thereby, the last diagnosis can be differentiated between subjective cognitive complaints with minimal cognitive dysfunction (SCC+) and without (SCC−).

MANOVA-based comparisons were already made in Bathke et al. (2018). However, covariance and correlation comparisons were only made in a descriptive way. We now complement their analyses. In Table 3, the number of patients divided by sex and diagnosis can be found. Since in Bathke et al. (2018) and Sattler et al. (2022) no distinction between SCC+ and SCC− was made, we consider both together as diagnosis SCC.

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

(a) Homogeneity of correlation matrices between different diagnoses,
(b) Homogeneity of correlation matrices between different sexes,

while we will denote the corresponding hypothesis regarding the covariance matrix with $\mathcal{H}_0^{\boldsymbol{v}}$.

In Sattler et al. (2022), homogeneity of covariance matrices between different diagnoses as well as different sexes were investigated. Here, we consider the more general hypothesis of equal correlation matrices between the diagnoses and the sexes. Thereby, it is interesting to compare the results from the homogeneity of covariance matrices with those from testing the homogeneity of correlation matrices. We expect higher p

Table 4 P-values of different ATS for testing equality of correlation matrices and equality of covariance matrices

		ATS-Par for $\mathcal{H}_0^{\boldsymbol{v}}$ p value	ATS-Par for $\mathcal{H}_0^{\boldsymbol{r}}$ p value	ATS-Tay for $\mathcal{H}_0^{\boldsymbol{r}}$ p value
Male	AD versus MCI	0.1000	**0.0389**	0.0866
Male	AD versus SCC	**0.0452**	**0.0077**	**0.0363**
Male	MCI versus SCC	**0.0289**	0.0712	0.1279
Female	AD versus MCI	0.0613	0.3626	0.3749
Female	AD versus SCC	**0.0128**	0.4937	0.4892
Female	MCI versus SCC	0.5656	0.8788	0.8693
AD	male versus female	0.1008	**0.0290**	0.0658
MCI	male versus female	0.2455	0.6702	0.6577
SCC	male versus female	0.2066	0.1744	0.2238

values for equality of correlation through the larger hypothesis, but each rejection of equal correlation matrices directly allows us to reject the corresponding equality of covariance matrices. In Table 4 for both hypotheses, the p values for the ATS with parametric bootstrap are displayed, while for equality of correlations, we additionally use our test based on the Taylor-based Monte Carlo approach. For all considered bootstrap tests, 10,000 bootstrap runs are done, as well as 10,000 Monte Carlo steps.

Interestingly, for two hypotheses, the p value ATS-Par for equal correlation matrices is rejected at level 5%, while we could not reject the smaller hypothesis of equal covariance matrices. But for both hypotheses, the sample sizes are rather small with $N < 40$. Our simulation results for $d = 5$ showed that the ATS with parametric bootstrap is too liberal for small sample sizes, which might be the reason why the larger hypotheses can be rejected, and the smaller ones can not. But also ATS-Tay, which had a better small sample performance, rejected $\mathcal{H}_0^{\boldsymbol{r}}$ one time, while $\mathcal{H}_0^{\boldsymbol{v}}$ was not rejected. Moreover, it can be seen that the difference between some hypotheses is relatively small, like for the first three hypotheses, but it can also be quite large as for the comparison of women with AD and with SCC. This shows that from a rejection of $\mathcal{H}_0^{\boldsymbol{v}}$, no conclusion on $\mathcal{H}_0^{\boldsymbol{r}}$ can be drawn.

Due to the small sample size in relation to the dimension of the vectorized correlation matrix $p_u = 15$, even the results of ATS-Tay are a bit liberal. But nevertheless, the corresponding rejections allow various ideas for further investigations.

9 Conclusion & Outlook

In the present paper, a series of new test statistics was developed to check general null hypotheses formulated in terms of correlation matrices. The proposed method can be used for many popular quadratic forms, and the low restrictions allow their application for a variety of settings. In fact, existing procedures have more restrictive assumptions or could only be used for special hypotheses or settings. The diversity of possible null hypotheses, these low restrictions and the easy possibility of expansion, like using a Fisher z-transformation, make our approach attractive.

We proved the asymptotic normality of the estimation error of the vectorized empirical correlation matrix under the assumption of finite fourth moments of all components. Based on this, test statistics from a quite general group of quadratic forms were presented, and a bootstrap technique was developed to match their asymptotic distribution. To investigate the properties of the corresponding bootstrap test, an extensive simulation study was done. This also allows checking our test statistic based on a Fisher z-transformation. Here, hypotheses for one and two groups were considered, and the type-I-error control and the power to detect deviations from the null hypothesis were compared to existing test procedures. The developed tests outperform existing procedures for some hypotheses, while they offer good and interesting alternatives for others. Also, it is a known fact that a large sample size is required for testing correlations. Here, for group sample sizes larger than 50, Bradley's liberal criterion was often fulfilled. Especially our Taylor-based approach was convincing for small sample sizes with multiple groups. An illustrative data analysis completed our investigations.

In future research, we will take a closer look at our newly proposed combined test for simultaneously testing the hypotheses of equal covariance matrices and correlation matrices. Thereto extensive simulations will be done, to, among other things, examine the performance of the corresponding Taylor approach. This relates to studying the large number of possible null hypotheses included in our model. For example, tests for given covariance structures (such as compound symmetry or autoregressive) or structures of correlation matrices with unknown parameters are of great interest. Since there are heterogeneous versions of many popular structures, testing such structures or other patterns can be seen as a combination of testing hypotheses regarding covariance matrices and hypotheses regarding correlation matrices. Moreover, an investigation of Monte Carlo approaches for a higher order of Taylor approximation for real small sample sizes could be interesting.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11749-023-00906-6.

Acknowledgements We like to thank two anonymous referees and the editor. A special thank goes to the Associate Editor for suggesting the idea of simultaneously studying variances and correlations. Moreover, we would like to thank the German Research Foundation for the support received within project PA 2409/3-2.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Competing interests The authors report there are no competing interests to declare.

References

Aitkin MA, Nelson WC, Reinfurt KH (1968) Tests for correlation matrices. Biometrika 55(2):327–334. https://doi.org/10.1093/biomet/55.2.327

Bartlett MS (1951) The effect of standardization on a chi square approximation in factor analysis. Biometrika 38(3/4):337–344

Bartlett MS, Rajalakshman DV (1953) Goodness of fit tests for simultaneous autoregressive series. J R Stat Soc Ser B (Methodol) 15(1):107–124

Bathke AC, Friedrich S, Pauly M, Konietschke F, Staffen W, Strobl N, Höller Y (2018) 03). Testing mean differences among groups: multivariate and repeated measures analysis with minimal assumptions. Multivar Behav Res 53:348–359

Boos DD, Brownie C (2004) Comparing variances and other measures of dispersion. Statist Sci 19(4):571–578. https://doi.org/10.1214/088342304000000503

Box GEP (1950) Problems in the analysis of growth and wear curves. Biometrics 6(4):362–389

Bradley JV (1978) Robustness? Br J Math Stat Psychol 31(2):144–152. https://doi.org/10.1111/j.2044-8317.1978.tb00581.x

Bretz F, Genz A, Hothorn LA (2001) On the numerical availability of multiple comparison procedures. Biom J 43(5):645–656

Browne M, Shapiro A (1986) The asymptotic covariance matrix of sample correlation coefficients under general conditions. Linear Algebra Appl 82:169–176

Buhlmann P (1998) Sieve bootstrap for smoothing in nonstationary time series. Ann Stat 26(1):48–83

Efron B (1988) Bootstrap confidence intervals: Good or bad? Psychol Bull 104(2):293

Fisher R (1921) On the “probable error" of a correlation coefficient deduced from a small sample. Metron 1(4):3–32

Friedrich S, Brunner E, Pauly M (2017) Permuting longitudinal data in spite of the dependencies. J Multivar Anal 153:255–265

Friedrich S, Konietschke F, Pauly M (2019) Manova.rm: analysis of multivariate data and repeated measures designs [Computer software manual]. Retrieved from https://cran.r-project.org/package=MANOVA.RM (R package version 0.3.2)

Friedrich S, Pauly M (2017) 04). MATS: inference for potentially singular and heteroscedastic MANOVA. J Multivar Anal 165:166–179

Gaißer S, Schmid F (2010) On testing equality of pairwise rank correlations in a multivariate random vector. J Multivar Anal 101(10):2598–2615

Gupta AK, Xu J (2006) On some tests of the covariance matrix under general conditions. Ann Inst Stat Math 58(1):101–114. https://doi.org/10.1007/s10463-005-0010-z

Jennrich RI (1970) An asymptotic chi square test for the equality of two correlation matrices. J Am Stat Assoc 65(330):904–912. https://doi.org/10.1080/01621459.1970.10481133

Joereskog K (1978) Structural analysis of covariance and correlation matrices. Psychometrika 43(4):443–477

Konietschke F, Bösiger S, Brunner B, Hothorn LA (2013) Are multiple contrast tests superior to the ANOVA? Int J Biostat 9(1):63–73. https://doi.org/10.1515/ijb-2012-0020

Magnus J, Neudecker H (1980) The elimination matrix: some lemmas and applications (Other publications TiSEM). Tilburg University, School of Economics and Management. Retrieved from https://EconPapers.repec.org/RePEc:tiu:tiutis:0e3315d3-846c-4bc5-928e-f9f025fa05b5

Munko M, Ditzhaus M, Dobler D, Genuneit J (2023) Rmst-based multiple contrast tests in general factorial designs. arXiv:2308.08346

Munko M, Ditzhaus M, Pauly M, Smaga L, Zhang J-T (2023) General multiple tests for functional data. arXiv:2306.15259

Nel D (1985) A matrix derivation of the asymptotic covariance matrix of sample correlation coefficients. Linear Algebra Appl 67:137–145

Nowak CP, Konietschke F (2021) Simultaneous inference for Kendall’s tau. J Multivar Anal 185:104767

Omelka M, Pauly M (2012) Testing equality of correlation coefficients in two populations via permutation methods. J Stat Plan Inference 142(6):1396–1406

Perreault S, Nešlehová JG, Duchesne T (2022) Hypothesis tests for structured rank correlation matrices. J Am Stat Assoc. https://doi.org/10.1080/01621459.2022.2096619

Revelle W (2019) psych: procedures for psychological, psychometric, and personality research [Computer software manual]. Evanston, Illinois. Retrieved from https://CRAN.R-project.org/package=psych (R package version 1.9.12)

Sakaori F (2002) Permutation test for equality of correlation coefficients in two populations. Commun Stat Simul Comput 31:641–651

Sattler P, Bathke AC, Pauly M (2022) Testing hypotheses about covariance matrices in general MANOVA designs. J Stat Plan Inference 219:134–146

Sattler P, Zimmermann G (2023) Choice of the hypothesis matrix for using the Wald-type-statistic. arXiv:2310.05562

Staffen W, Strobl N, Zauner H, Höller Y, Dobesberger J, Trinka E (2014) Combining SPECT and EEG analysis for assessment of disorders with amnestic symptoms to enhance accuracy in early diagnostics. In: Poster A19 Presented at the 11th Annual Meeting of the Austrian Society of Neurology, 26–29 March, Salzburg, Austria

Steiger JH (1980) Testing pattern hypotheses on correlation matrices: alternative statistics and some empirical results. Multivar Behav Res 15(3):335–352 (**PMID: 26794186**)

Tian L, Wilding GE (2008) Confidence interval estimation of a common correlation coefficient. Comput Stat Data Anal 52(10):4872–4877

Umlauft M, Placzek M, Konietschke F, Pauly M (2019) Wild bootstrapping rank-based procedures: multiple testing in nonparametric factorial repeated measures designs. J Multivar Anal 171:176–192

van der Vaart AW, Wellner JA (1996) Weak convergence. In: Weak convergence and empirical processes: with applications to statistics. Springer New York, New York, NY, pp 16-28. Retrieved from https://doi.org/10.1007/978-1-4757-2545-2_3

Welz T, Doebler P, Pauly M (2022) Fisher transformation based confidence intervals of correlations in fixed-and random-effects meta-analysis. Br J Math Stat Psychol 75(1):1–22

Wilks SS (1946) Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. Ann Math Stat 17(3):257–281

Wu L, Weng C, Wang X, Wang K, Liu X (2018) Test of covariance and correlation matrices. arXiv:1812.01172