



Covariance structure tests for multivariate t -distribution

Katarzyna Filipiak¹ · Tõnu Kollo²

Received: 16 August 2023 / Revised: 24 April 2024
© The Author(s) 2024

Abstract

We derive an equation system for finding Maximum Likelihood Estimators (MLEs) for the parameters of a p -dimensional t -distribution with ν degrees of freedom, $t_{p,\nu}$, and use the MLEs for testing covariance structures for the $t_{p,\nu}$ -distributed population. The likelihood ratio test (LRT), Rao score test (RST) and Wald test (WT) statistics are derived under the general null-hypothesis $H_0 : \Sigma = \Sigma_0$, using a matrix derivative technique. Here the $p \times p$ -matrix Σ is a dispersion/scale parameter. Convergence to the asymptotic chi-square distribution under the null hypothesis is examined in extensive simulation experiments. Also the convergence to the chi-square distribution is studied empirically in the situation when the MLEs of a $t_{p,\nu}$ -distribution are changed to the corresponding estimators for a normal population. Type I errors and the power of the tests are also examined by simulation. In the simulation study the RST behaved more adequately than all remaining statistics in the situation when the dimensionality p was growing.

Keywords Multivariate t -distribution · Covariance structure testing · Likelihood ratio test · Rao score test · Wald test

Mathematics Subject Classification 62H10 · 62H15 · 62F03 · 62F05

1 Introduction

Classical multivariate statistics is based on a normally distributed population. In practice the normality assumption is often violated and more general classes of distributions are of interest. Continuous elliptical distributions and skew-elliptical distributions are

Katarzyna Filipiak and Tõnu Kollo have contributed equally to this work.

✉ Katarzyna Filipiak
katarzyna.filipiak@put.poznan.pl
Tõnu Kollo
tonu.kollo@ut.ee

¹ Institute of Mathematics, Poznan University of Technology, Piotrowo 3A, 60-965 Poznan, Poland

² Institute of Mathematics and Statistics, University of Tartu, Narva 18, 51009 Tartu, Estonia

usually the first choices for data modeling. The multivariate t -distribution, $t_{p,v}$, is of special interest because of two important properties—tail dependence and a heavier tail area than that of the normal distribution. When the number of degrees of freedom v tends to infinity, the $t_{p,v}$ -distribution converges to the normal distribution with the same parameters. Thus, we have a direct generalization of the normal population. A wide range of applications where a multivariate $t_{p,v}$ -distribution has been used can be found in (Kotz and Nadarajah 2004, Ch. 12), for instance. Among many recent reports we refer to Osorio et al. (2023) for applications in meteorology, Finegold and Drton (2011) for gene analysis, and Kan and Zhou (2017) and Galea et al. (2020) for economic and financial applications. For risk estimation in portfolio theory, a multivariate t -distribution is specially advocated in Lauprete et al. (2002) with reference to the t -copula.

Before applying specific methods of multivariate analysis it is important to examine first the structure of the covariance matrix Σ . Further analyses will be much simpler if one can take into account a specific covariance structure. Traditionally, tests have been developed for simple structures testing $\Sigma = \mathbf{I}_p$, sphericity $\Sigma = \sigma^2 \mathbf{I}_p$, uncorrelatedness $\Sigma = \Lambda$ with Λ diagonal, and intraclass correlation structure $\Sigma = \sigma^2[(1 - \rho)\mathbf{I}_p + \rho \mathbf{1}_p \mathbf{1}'_p]$, where $\mathbf{1}' = (1, 1, \dots, 1)$ is a p -dimensional row vector of ones and \mathbf{I}_p is the $p \times p$ identity matrix. In more complex situations, e.g. when analysing spatial-temporal data, the Kronecker product structure is present (Srivastava et al. 2008, 2009; Filipiak and Klein 2017, for example). When the population distribution is elliptical, the covariance matrix is a product of a univariate multiplier characterizing the distribution and a scale matrix, denoted also by Σ . We shall derive test statistics for the parameter Σ under the null hypothesis $\Sigma = \Sigma_0$, where Σ_0 can be specified.

Probably the most commonly used test is the likelihood ratio test (LRT) under the assumption of normality of the population (Anderson 2003; Bilodeau and Brenner 1999, for example). A profound study of basic LRTs concerning covariance structures under normality is presented in Muirhead (2005, Ch. 8). However, it is known that when the number of parameters to be tested is large, the LRT will almost always reject the null hypothesis. In order to overcome this problem, corrections to the test have been made so that it can be used in a high-dimensional setup as well; see Bai et al. (2009), for example. In practice the uncorrected test is still used. In Kollo et al. (2016) it is shown that instead of the LRT or Wald test (WT), the more consistent Rao score test (RST) should be used to test a particular covariance structure under normality. In this paper we examine these three tests for a multivariate $t_{p,v}$ -distributed population. Since one of our goals is to investigate possible differences in maximum likelihood estimates (MLEs) between t -distributed and normal populations, we focus on cases where the number of degrees of freedom, v , is known and as small as possible. Our second goal is to examine the speed of convergence of the considered test statistics when v is growing and to study the power of the tests (by simulations), the assumption of fixed v allows us to speed up the simulations, especially for higher values of dimension and sample size. Nevertheless, the formulas for test statistics under unknown v will be also presented, together with their distributions compared to the respective distributions under known v .

In Sect. 2 notation and required notions are explained. For fixed ν in Sect. 3 we derive equations for finding MLEs and present, in Proposition 1, the LRT statistic. In Sect. 4 we find a score vector and information matrix for a $t_{p,\nu}$ -distributed population, and in Proposition 2 the RST statistic is presented using trace functions. In Sect. 5 the WT statistic is derived and presented in Proposition 3. A modification, WT^* , of the Wald statistic is also introduced. In Sect. 6 convergence to the limiting chi-square distribution is examined in simulation experiments for all test statistics. Convergence is also examined in the situation when the MLEs of a $t_{p,\nu}$ -distribution are replaced by their approximations—the MLEs of the corresponding normal population. In the simulation studies, empirical type I errors and powers of the tests are also calculated. Assuming unknown ν in Sect. 7 we present relevant test statistics and we show, that their distributions are similar to the distributions of test statistics derived in Sects. 3–5. Finally we summarize the results in Sect. 8.

2 Notation and notions

Derivations in this paper utilize a matrix technique based on vectorization, Kronecker product, and matrix derivative. For deeper insight into this technique the interested reader is referred to Magnus and Neudecker (2019), Harville (1997), or Kollo and von Rosen (2005). The following properties of “vec” operator, which transforms a matrix into a vector by stacking the columns one under the other, are frequently used:

$$\begin{aligned}\text{vec}(\mathbf{ABC}) &= (\mathbf{C}' \otimes \mathbf{A})\text{vec} \mathbf{B}, \\ \text{vec}' \mathbf{A} \text{vec} \mathbf{B} &= \text{tr}(\mathbf{A}' \mathbf{B}), \\ \text{vec}(\mathbf{ab}') &= \mathbf{b} \otimes \mathbf{a},\end{aligned}$$

where “tr” denotes the trace operator. Matrices are denoted with capital letters and vectors with lowercase letters in bold.

Later we shall use matrix derivatives repeatedly, and the definition of Neudecker (1969) is applied.

Definition 1 Let the elements of $\mathbf{Y} \in \mathbb{R}^{r \times s}$ be functions of $\mathbf{X} \in \mathbb{R}^{p \times q}$ with non-constant and functionally independent elements x_{ij} . The matrix $\frac{d\mathbf{Y}}{d\mathbf{X}} \in \mathbb{R}^{rs \times pq}$ is called the matrix derivative of \mathbf{Y} by \mathbf{X} in a set A if the partial derivatives $\frac{\partial y_{kl}}{\partial x_{ij}}$ exist and are continuous in A , and

$$\frac{d\mathbf{Y}}{d\mathbf{X}} = \frac{d}{d\text{vec}' \mathbf{X}} \otimes \text{vec} \mathbf{Y}$$

where

$$\frac{d}{d\text{vec}' \mathbf{X}} = \left(\frac{\partial}{\partial x_{11}}, \dots, \frac{\partial}{\partial x_{p1}}, \frac{\partial}{\partial x_{12}}, \dots, \frac{\partial}{\partial x_{p2}}, \dots, \frac{\partial}{\partial x_{1q}}, \dots, \frac{\partial}{\partial x_{pq}} \right)$$

and $\text{vec}(\cdot)$ is the vectorization operator.

Definition of the derivative with respect to matrix \mathbf{X} of a scalar, vector and matrix function, can be also found in a recent paper by Liu et al. (2023), where an insightful overview of matrix-oriented results and their applications in statistics are presented.

From the basic properties of the matrix derivative (Magnus and Neudecker 2019) we obtain the differentiation rule for a composite function:

$$\text{when } \mathbf{Z} = \mathbf{Z}(\mathbf{Y}), \quad \mathbf{Y} = \mathbf{Y}(\mathbf{X}) \quad \text{then} \quad \frac{d\mathbf{Z}}{d\mathbf{X}} = \frac{d\mathbf{Z}}{d\mathbf{Y}} \frac{d\mathbf{Y}}{d\mathbf{X}},$$

and the derivatives of the determinant and the inverse of a square matrix \mathbf{X} :

$$\frac{d|\mathbf{X}|}{d\mathbf{X}} = |\mathbf{X}| \text{vec}'(\mathbf{X}^{-1})' \quad \text{and} \quad \frac{d\mathbf{X}^{-1}}{d\mathbf{X}} = -(\mathbf{X}^{-1})' \otimes \mathbf{X}^{-1},$$

where $|\cdot|$ denotes the determinant.

Note that for the symmetric matrix $\mathbf{X} = \mathbf{X}' : p \times p$, the derivative is computed with respect to the vectorized lower triangle of \mathbf{X} , denoted by $\text{vech } \mathbf{X}$, instead of $\text{vec } \mathbf{X}$, which results in $rs \times p(p+1)/2$ matrix. To avoid misunderstandings, the derivative with respect to $\text{vech } \mathbf{X}$ we denote by adding superscript to \mathbf{X} , i.e., $\frac{d\mathbf{Y}}{d\mathbf{X}^\Delta} = \frac{d\text{vec } \mathbf{Y}}{d\text{vech}' \mathbf{X}}$. Then, using the chain rule described above,

$$\frac{d\mathbf{Y}}{d\mathbf{X}^\Delta} = \frac{d\text{vec } \mathbf{Y}}{d\text{vec}' \mathbf{X}} \cdot \frac{d\text{vec } \mathbf{X}}{d\text{vech}' \mathbf{X}} = \frac{d\text{vec } \mathbf{Y}}{d\text{vec}' \mathbf{X}} \cdot \mathbf{D}_p,$$

where $\mathbf{D}_p : p^2 \times \frac{1}{2}p(p+1)$ is a duplication matrix that transforms $\text{vech } \mathbf{A}$ into $\text{vec } \mathbf{A}$, i.e.,

$$\mathbf{D}_p \text{vech } \mathbf{A} = \text{vec } \mathbf{A};$$

cf. Magnus and Neudecker (1986), Filipiak et al. (2016).

Let \mathbf{x} be a continuous random vector with distribution $P_{\mathbf{x}}(\boldsymbol{\theta})$ and density function $f_{\mathbf{x}}(\mathbf{x}, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of unknown parameters.

Definition 2 The score vector of a random vector \mathbf{x} is given by the matrix derivative

$$\mathbf{u}(\mathbf{x}, \boldsymbol{\theta}) = \left(\frac{d}{d\boldsymbol{\theta}} \ln f(\mathbf{x}, \boldsymbol{\theta}) \right)'.$$

Definition 3 The information matrix of a random vector \mathbf{x} is the covariance matrix of the score vector $\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})$:

$$\mathbf{I}(\mathbf{x}, \boldsymbol{\theta}) = \mathbb{D}(\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})) = \mathbb{E}(\mathbf{u}(\mathbf{x}, \boldsymbol{\theta})\mathbf{u}'(\mathbf{x}, \boldsymbol{\theta})).$$

Let $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ denote a random sample from the distribution $P_{\mathbf{x}}(\boldsymbol{\theta})$. Then the log-likelihood function of the sample is given by $\ell(\boldsymbol{\theta}, \mathbf{X}) = \sum_{i=1}^n \ln f(\mathbf{x}_i, \boldsymbol{\theta})$ and the

score function of the sample by $\mathbf{u}(\mathbf{X}, \boldsymbol{\theta}) = \sum_{i=1}^n \mathbf{u}(\mathbf{x}_i, \boldsymbol{\theta})$. The information matrix of the sample is given by $\mathbf{I}(\mathbf{X}, \boldsymbol{\theta}) = n \cdot \mathbf{I}(\mathbf{x}, \boldsymbol{\theta})$.

Definition 4 Let a random p -vector \mathbf{y} be normally distributed, $\mathbf{y} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$, and let $Z^2 \sim \chi_v^2$ be independent of \mathbf{y} . Then

$$\mathbf{x} = \frac{\sqrt{v}}{Z} \mathbf{y} + \boldsymbol{\mu}$$

is multivariate $t_{p,v}$ -distributed with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, i.e., $\mathbf{x} \sim t_{p,v}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Remark 1 When $\boldsymbol{\mu} = \mathbf{0}$ and instead of $\boldsymbol{\Sigma}$ we have a correlation matrix \mathbf{R} as the parameter of the normal distribution, i.e., $N_p(\mathbf{0}, \mathbf{R})$, we obtain in the univariate case a standard t -distribution. From Definition 4 we obtain a member of the location-scale family of the univariate t -distribution.

The density function of $\mathbf{x} \sim t_{p,v}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given by

$$f_v(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = c_p |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \left[1 + \frac{1}{v} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]^{-\frac{v+p}{2}}, \tag{1}$$

where $c_p = \frac{\Gamma((v+p)/2)}{(\pi v)^{p/2} \Gamma(v/2)}$ (cf. Kotz and Nadarajah 2004, Ch. 5), and the first two moments of \mathbf{x} are

$$\begin{aligned} \mathbb{E}\mathbf{x} &= \boldsymbol{\mu}, \\ \mathbb{D}\mathbf{x} &= \frac{v}{v-2} \boldsymbol{\Sigma}, \quad v > 2; \end{aligned}$$

cf. (Muirhead 2005, p. 48).

Recall that for $v \rightarrow \infty$ the multivariate $t_{p,v}$ -distribution tends to the multivariate normal distribution, for which obviously

$$\mathbb{E}\mathbf{x} = \boldsymbol{\mu}, \quad \mathbb{D}\mathbf{x} = \boldsymbol{\Sigma},$$

and the maximum likelihood estimators (MLEs) of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i = \frac{1}{n} \mathbf{X} \mathbf{1}_n = \bar{\mathbf{x}}, \\ \hat{\boldsymbol{\Sigma}} &= \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' = \frac{1}{n} \mathbf{X} (\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n') \mathbf{X}' = \mathbf{S} \end{aligned} \tag{2}$$

with $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ being a sample from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{1}_n$ being an n -dimensional vector of ones.

It is worth noting that the results presented in Sects. 3–6 are obtained under the assumption of v to be known, while in Sect. 7 v is assumed to be unknown.

3 Likelihood ratio test

For testing the hypothesis

$$\begin{cases} H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \end{cases} \quad (3)$$

the LRT statistic in logarithmic form is

$$\begin{aligned} \text{LRT}(\mathbf{X}, \boldsymbol{\theta}_0) &= -2 \ln \left(\frac{L(\boldsymbol{\theta}_0, \mathbf{X})}{\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \mathbf{X})} \right) = -2[\ln L(\boldsymbol{\theta}_0, \mathbf{X}) - \ln L(\widehat{\boldsymbol{\theta}}, \mathbf{X})] \\ &= -2[\ell(\boldsymbol{\theta}_0, \mathbf{X}) - \ell(\widehat{\boldsymbol{\theta}}, \mathbf{X})], \end{aligned}$$

where $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ is a random sample from $P_{\mathbf{X}}(\boldsymbol{\theta})$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r)'$ is a vector of unknown parameters, $\widehat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$, and $L(\boldsymbol{\theta}_0, \mathbf{X})$ and $L(\widehat{\boldsymbol{\theta}}, \mathbf{X})$ ($\ell(\boldsymbol{\theta}_0, \mathbf{X})$ and $\ell(\widehat{\boldsymbol{\theta}}, \mathbf{X})$) are the likelihood (log-likelihood) functions of the vector of parameters under the null and alternative hypotheses respectively. When the sample size $n \rightarrow \infty$ and H_0 holds, the distribution of $\text{LRT}(\mathbf{X}, \boldsymbol{\theta})$ converges to χ_r^2 .

Assume now that $\boldsymbol{\theta} = (\boldsymbol{\theta}_{r-k}, \boldsymbol{\theta}_k)'$, where $\boldsymbol{\theta}_{r-k}$ is the set of $(r-k)$ fixed parameters under the null hypothesis, i.e., $H_0 : \boldsymbol{\theta}_{r-k} = \boldsymbol{\theta}_0$. Then the LRT in logarithmic form has the representation

$$\begin{aligned} \text{LRT}(\mathbf{X}, \boldsymbol{\theta}_0) &= -2 \ln \left(\frac{\max_{\boldsymbol{\theta}_k} L(\boldsymbol{\theta}_0, \mathbf{X})}{\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \mathbf{X})} \right) \\ &= -2[\ell((\boldsymbol{\theta}_0, \boldsymbol{\theta}_k)', \mathbf{X}) - \ell(\widehat{\boldsymbol{\theta}}, \mathbf{X})], \end{aligned}$$

where $\widehat{\boldsymbol{\theta}}_k$ is the MLE of k non-fixed parameters under the null hypothesis. The distribution of $\text{LRT}(\mathbf{X}, \boldsymbol{\theta}_0)$ converges to χ_{r-k}^2 when the sample size $n \rightarrow \infty$ and H_0 holds; c.f. (Rao 1973, §6e).

Let $\mathbf{x} \sim t_{p,v}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)'$, the scale parameter $\boldsymbol{\Sigma} > 0 : p \times p$ and let $\boldsymbol{\theta} = (\boldsymbol{\mu}', \text{vech}' \boldsymbol{\Sigma})'$. Due to (1), the likelihood function $L(\boldsymbol{\theta}, \mathbf{X})$ is

$$L(\boldsymbol{\theta}, \mathbf{X}) = (c_p)^n |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \prod_{i=1}^n \left[1 + \frac{1}{v} (\mathbf{x}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \right]^{-\frac{v+p}{2}},$$

and the log-likelihood function can be presented as

$$\ell(\boldsymbol{\theta}, \mathbf{X}) = n \ln c_p - \frac{n}{2} \ln |\boldsymbol{\Sigma}| - \frac{v+p}{2} \sum_{i=1}^n \ln \left[1 + \frac{1}{v} (\mathbf{x}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \right]. \quad (4)$$

Differentiating (4) with respect to μ and Σ gives the following partial derivatives:

$$\begin{aligned} \frac{\partial \ell}{\partial \mu} &= \frac{v+p}{v} \sum_{i=1}^n \frac{(\mathbf{x}_i - \mu)' \Sigma^{-1}}{1 + \frac{1}{v}(\mathbf{x}_i - \mu)' \Sigma^{-1}(\mathbf{x}_i - \mu)}, \\ \frac{\partial \ell}{\partial \Sigma} &= \left\{ -\frac{n}{2} \text{vec}' \Sigma^{-1} + \frac{v+p}{2v} \sum_{i=1}^n \frac{(\mathbf{x}_i - \mu)' \otimes (\mathbf{x}_i - \mu)'}{1 + \frac{1}{v}(\mathbf{x}_i - \mu)' \Sigma^{-1}(\mathbf{x}_i - \mu)} \Sigma^{-1} \otimes \Sigma^{-1} \right\} \mathbf{D}_p \end{aligned} \tag{5}$$

with \mathbf{D}_p being the respective duplication matrix. Thus, to obtain MLEs of μ and Σ , we have to solve the following system of equations:

$$\begin{cases} \mu = \sum_{i=1}^n \frac{\mathbf{x}_i}{t_i} / \sum_{i=1}^n \frac{1}{t_i} \\ \Sigma = \frac{v+p}{nv} \sum_{i=1}^n \frac{(\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)'}{t_i} \end{cases}$$

with $t_i = 1 + \frac{1}{v}(\mathbf{x}_i - \mu)' \Sigma^{-1}(\mathbf{x}_i - \mu)$. Noting that t_i depends on μ and Σ , the above system of equations can be solved numerically in the following way:

- fix starting values $\mu^{(0)}$ and $\Sigma^{(0)}$,
- for $k = 1, 2, \dots$, update the following system

$$\begin{aligned} t_i^{(k-1)} &= 1 + \frac{1}{v}(\mathbf{x}_i - \mu^{(k-1)})' (\Sigma^{(k-1)})^{-1} (\mathbf{x}_i - \mu^{(k-1)}), \quad i = 1, \dots, n \\ \mu^{(k)} &= \sum_{i=1}^n \frac{\mathbf{x}_i}{t_i^{(k-1)}} / \sum_{i=1}^n \frac{1}{t_i^{(k-1)}} \\ \Sigma^{(k)} &= \frac{v+p}{nv} \sum_{i=1}^n \frac{(\mathbf{x}_i - \mu^{(k)})(\mathbf{x}_i - \mu^{(k)})'}{t_i^{(k-1)}} \end{aligned} \tag{6}$$

until the convergence criterion is satisfied.

In our considerations the algorithm stops when $\|\mu^{(k)} - \mu^{(k-1)}\| \leq 10^{-6}$ and $\|\Sigma^{(k)} - \Sigma^{(k-1)}\|_F^2 \leq 10^{-6}$, where $\|\cdot\|_F$ denotes the Frobenius norm. To the best of our knowledge there are no recommendations regarding the choice of starting point. In our research we choose $\mu = \mathbf{0}$ and $\Sigma = \mathbf{I}_p$. Throughout the paper we denote the solutions obtained by $\hat{\mu}$ and $\hat{\Sigma}$.

Remark 2 In applications with missing data the system of Eq. (5) is solved using the EM algorithm; cf. Liu and Rubin (1995), McLachlan and Krishnan (1997), Finegold and Drton (2011). The procedure for solving (6) coincides with the EM algorithm.

In order to determine the LRT statistic to test

$$\begin{cases} H_0 : \Sigma = \Sigma_0 \\ H_1 : \Sigma \neq \Sigma_0 \end{cases} \tag{7}$$

when no constraints are imposed on $\boldsymbol{\mu}$, it is necessary to compute the MLE of $\boldsymbol{\mu}$ under H_0 , which will be denoted by $\widehat{\boldsymbol{\mu}}_0$. Using the same approach as above, $\widehat{\boldsymbol{\mu}}_0$ is obtained as the solution of

$$\boldsymbol{\mu} = \sum_{i=1}^n \frac{\mathbf{x}_i}{t_{i0}} / \sum_{i=1}^n \frac{1}{t_{i0}},$$

where $t_{i0} = 1 + \frac{1}{\nu}(\mathbf{x}_i - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_0^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_0)$. Applying the same algorithm as before, for a fixed $\boldsymbol{\Sigma}_0$ we choose a starting value $\boldsymbol{\mu}^{(0)}$, and then, for $k = 1, 2, \dots$, we solve the system

$$\begin{cases} t_{i0}^{(k-1)} = 1 + \frac{1}{\nu}(\mathbf{x}_i - \boldsymbol{\mu}^{(k-1)})' \boldsymbol{\Sigma}_0^{-1}(\mathbf{x}_i - \boldsymbol{\mu}^{(k-1)}), & i = 1, \dots, n \\ \boldsymbol{\mu}_0^{(k)} = \sum_{i=1}^n \frac{\mathbf{x}_i}{t_{i0}^{(k-1)}} / \sum_{i=1}^n \frac{1}{t_{i0}^{(k-1)}} \end{cases} \quad (8)$$

until the stopping rule holds.

The LRT statistic for testing (7) when no constraints are imposed on $\boldsymbol{\mu}$ has the following form:

$$\text{LRT}(\mathbf{X}, \boldsymbol{\Sigma}_0) = -2 \ln \left[\left(\frac{|\widehat{\boldsymbol{\Sigma}}|}{|\boldsymbol{\Sigma}_0|} \right)^{n/2} \prod_{i=1}^n \left(\frac{1 + \frac{1}{\nu}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_0)' \boldsymbol{\Sigma}_0^{-1}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_0)}{1 + \frac{1}{\nu}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}})' \widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}})} \right)^{-(\nu+p)/2} \right],$$

where $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ are the numerically determined MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ (solutions of (6)) and $\widehat{\boldsymbol{\mu}}_0$ is the solution of (8).

Due to Kollo and Valge (2020, formula (12.3)), their Proposition 12.1 should be rewritten as follows.

Proposition 1 *The LRT statistic for testing (7) when no constraints are imposed on $\boldsymbol{\mu}$ is given by*

$$\begin{aligned} & \text{LRT}(\mathbf{X}, \boldsymbol{\Sigma}_0) \\ &= -n \left[\ln |\widehat{\boldsymbol{\Sigma}}| - \ln |\boldsymbol{\Sigma}_0| \right] \\ &+ (\nu + p) \\ & \times \sum_{i=1}^n \left\{ \ln \left[1 + \frac{1}{\nu}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_0)' \boldsymbol{\Sigma}_0^{-1}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_0) \right] - \ln \left[1 + \frac{1}{\nu}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}})' \widehat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}) \right] \right\}, \end{aligned}$$

where $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ is a random sample from $t_{p,\nu}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\nu > 2$ is known, $\widehat{\boldsymbol{\mu}}$, $\widehat{\boldsymbol{\Sigma}}$ are the solutions of (6), and $\widehat{\boldsymbol{\mu}}_0$ is the solution of (8).

When $n \rightarrow \infty$ and H_0 holds, the distribution of $\text{LRT}(\mathbf{X}, \boldsymbol{\Sigma}_0)$ tends to the chi-square distribution with $p(p+1)/2$ degrees of freedom.

Note that the convergence to the chi-square distribution follows from e.g. Wilks (1938).

Remark 3 The LRT statistic in Proposition 1 contains elements of the form

$$(v + p) \ln \left(1 + \frac{u_i}{v} \right),$$

where u_i stands for $(\mathbf{x}_i - \bar{\mathbf{x}})' \Sigma_0^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})$ or/and $(\mathbf{x}_i - \bar{\mathbf{x}})' \widehat{\Sigma}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})$.

Then, if $v \rightarrow \infty$, using the l'Hôpital rule we obtain

$$\lim_{v \rightarrow \infty} (v + p) \ln \left(1 + \frac{u_i}{v} \right) = \lim_{v \rightarrow \infty} \frac{\ln \left(1 + \frac{u_i}{v} \right)}{\frac{1}{v+p}} = \lim_{v \rightarrow \infty} \frac{u_i (v + p)^2}{v(v + u_i)} = u_i.$$

This means that the LRT converges to

$$\begin{aligned} & -n [\ln |\widehat{\Sigma}| - \ln |\Sigma_0|] + \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})' \left(\Sigma_0^{-1} - \widehat{\Sigma}^{-1} \right) (\mathbf{x}_i - \bar{\mathbf{x}}) \\ & = -n \ln |\widehat{\Sigma} \Sigma_0^{-1}| + \text{tr} \left[\left(\Sigma_0^{-1} - \widehat{\Sigma}^{-1} \right) \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})' (\mathbf{x}_i - \bar{\mathbf{x}}) \right]. \end{aligned}$$

From the definition of \mathbf{S} given in (2) we have

$$\text{LRT} \rightarrow -n \left[\ln |\mathbf{S} \Sigma_0^{-1}| + \text{tr} \left(\mathbf{S} \Sigma_0^{-1} \right) - p \right],$$

which is the LRT statistic for testing (7) under normality; cf. Kollo et al. (2016).

4 Rao score test

The RST statistic is a function of the score vector and the Fisher information matrix.

Definition 5 The RST statistic for testing hypothesis (3) is of the form

$$\text{RST}(\mathbf{X}, \boldsymbol{\theta}_0) = \mathbf{u}'(\mathbf{X}, \boldsymbol{\theta}_0) \mathbf{I}(\mathbf{X}, \boldsymbol{\theta}_0)^{-1} \mathbf{u}(\mathbf{X}, \boldsymbol{\theta}_0),$$

where $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ is a random sample from $P_{\mathbf{x}}(\boldsymbol{\theta})$, $\mathbf{u}(\mathbf{X}, \boldsymbol{\theta}_0)$ is the score function, $\mathbf{I}(\mathbf{X}, \boldsymbol{\theta}_0)$ is the information matrix, and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r)'$.

Following Rao (1948), when the sample size $n \rightarrow \infty$ and H_0 holds, the distribution of $\text{RST}(\mathbf{X}, \boldsymbol{\theta}_0)$ converges to χ_r^2 .

Assume now that $\boldsymbol{\theta} = (\boldsymbol{\theta}_{r-k}, \boldsymbol{\theta}_k)'$, where $\boldsymbol{\theta}_{r-k}$ is the set of $(r - k)$ fixed parameters under the null hypothesis, i.e., $H_0 : \boldsymbol{\theta}_{r-k} = \boldsymbol{\theta}_0$. Then the RST has the representation

$$\text{RST}(\mathbf{X}, \boldsymbol{\theta}_0) = \mathbf{u}'_1(\mathbf{X}, \boldsymbol{\theta}_0, \widehat{\boldsymbol{\theta}}_k) \mathbf{I}_{1,2}^{-1}(\mathbf{X}, \boldsymbol{\theta}_0, \widehat{\boldsymbol{\theta}}_k) \mathbf{u}_1(\mathbf{X}, \boldsymbol{\theta}_0, \widehat{\boldsymbol{\theta}}_k), \tag{9}$$

where the score vector $\mathbf{u}'(\mathbf{X}, \boldsymbol{\theta}_0, \boldsymbol{\theta}_k)$ has been divided into two parts, related to partial derivatives of $\ell(\boldsymbol{\theta}, \mathbf{X})$ with respect to $\boldsymbol{\theta}_{r-k}$ ($= \mathbf{u}_1(\mathbf{X}, \boldsymbol{\theta}_0, \boldsymbol{\theta}_k)$) and $\boldsymbol{\theta}_k$

(= $\mathbf{u}_2(\mathbf{X}, \boldsymbol{\theta}_0, \boldsymbol{\theta}_k)$), $I_{1,2}(\mathbf{X}, \boldsymbol{\theta}_0, \boldsymbol{\theta}_k)$ is the Schur complement of I_{22} in the Fisher information matrix $\mathbf{I}(\mathbf{X}, \boldsymbol{\theta}_0, \boldsymbol{\theta}_k) = (I_{ij})$ $i, j = 1, 2$, i.e., $I_{1,2} = I_{11} - I_{12}I_{22}^{-1}I_{21}$, and where $\widehat{\boldsymbol{\theta}}_k$ is the MLE of k non-fixed parameters under the null hypothesis; cf. Rao (2005). Moreover, when under the null hypothesis there are k parameters not fixed, the sample size $n \rightarrow \infty$ and H_0 holds, the distribution of $RST(\mathbf{X}, \boldsymbol{\theta})$ tends to χ_{r-k}^2 ; cf. (Rao 1973, Sect. 6e).

In our case $\boldsymbol{\theta} = (\boldsymbol{\mu}', \text{vech}'\boldsymbol{\Sigma})'$ and the $(p + \frac{p(p+1)}{2})$ -dimensional score vector of partial derivatives has the form

$$\mathbf{u}(\mathbf{X}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\mathbf{u}_1(\mathbf{X}, \boldsymbol{\mu}, \boldsymbol{\Sigma}), \mathbf{u}_2(\mathbf{X}, \boldsymbol{\mu}, \boldsymbol{\Sigma}))' = \left(\frac{\partial \ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{X})}{\partial \boldsymbol{\mu}}, \frac{\partial \ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{X})}{\partial \boldsymbol{\Sigma}^\Delta} \right)' \tag{10}$$

The components of the score vector are given by (5). Expressions for the elements of the $(p + p(p+1)/2) \times (p + p(p+1)/2)$ Fisher information matrix of a $t_{p,v}$ -distributed random vector can be computed using the formulae from Mitchell (1989), which in the matrix representation have the form

$$\mathbf{I}(\mathbf{X}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \begin{pmatrix} \frac{(v+p)n}{v+p+2} \boldsymbol{\Sigma}^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{n}{2(v+p+2)} \mathbf{D}'_p [(v+p)(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) - \text{vec } \boldsymbol{\Sigma}^{-1} \text{vec}' \boldsymbol{\Sigma}^{-1}] \mathbf{D}_p \end{pmatrix} \tag{11}$$

By inserting the MLE of $\boldsymbol{\mu}$ under the null hypothesis, $\widehat{\boldsymbol{\mu}}_0$, and $\boldsymbol{\Sigma}_0$ into the score vector (10) and information matrix (11), it can be seen that since there are no restrictions on $\boldsymbol{\mu}$ in the hypothesis (7), it is enough to take into account the second component of the score vector and the lower diagonal entry of the Fisher information matrix.

Proposition 2 *The RST statistic for testing (7) when no constraints are imposed on $\boldsymbol{\mu}$ is given by*

$$RST(\mathbf{X}, \boldsymbol{\Sigma}_0) = \frac{n(v+p+2)}{2v(v+p)} \left\{ v \cdot \text{tr}[(\mathbf{V}\boldsymbol{\Sigma}_0^{-1})^2] + \text{tr}^2(\mathbf{V}\boldsymbol{\Sigma}_0^{-1}) \right\}$$

with

$$\mathbf{V} = \mathbf{V}(\mathbf{X}, \boldsymbol{\Sigma}_0) = \boldsymbol{\Sigma}_0 - \frac{v+p}{nv} \sum_{i=1}^n \frac{(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_0)(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_0)'}{1 + \frac{1}{v}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_0)' \boldsymbol{\Sigma}_0^{-1} (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_0)} \tag{12}$$

where $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ is a sample from $t_{v,p}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $v > 2$ is known, and $\widehat{\boldsymbol{\mu}}_0$ is the solution of (8).

When $n \rightarrow \infty$ and H_0 holds, the distribution of $RST(\mathbf{X}, \boldsymbol{\Sigma}_0)$ tends to the chi-square distribution with $p(p+1)/2$ degrees of freedom.

Proof Since there are no restrictions on μ in (7), to determine the form of the RST we use (9), with the role of fixed/non-fixed parameters interchanged, i.e.,

$$\text{RST}(\mathbf{X}, \Sigma_0) = \mathbf{u}'_2(\mathbf{X}, \hat{\mu}, \Sigma_0) \mathbf{I}_{2,1}^{-1}(\mathbf{X}, \hat{\mu}, \Sigma_0) \mathbf{u}_2(\mathbf{X}, \hat{\mu}, \Sigma_0).$$

From (11) it follows that $\mathbf{I}_{2,1}(\mathbf{X}, \hat{\mu}, \Sigma_0)$ reduces to $\mathbf{I}_{22}(\mathbf{X}, \hat{\mu}, \Sigma_0)$.

Due to the second equality of (5)

$$\mathbf{u}_2(\mathbf{X}, \hat{\mu}, \Sigma_0) = -\frac{n}{2} \left[\text{vec } \Sigma_0^{-1} - \frac{v+p}{nv} (\Sigma_0^{-1} \otimes \Sigma_0^{-1}) \sum_{i=1}^n \frac{(\mathbf{x}_i - \hat{\mu}) \otimes (\mathbf{x}_i - \hat{\mu})}{1 + \frac{1}{v} (\mathbf{x}_i - \hat{\mu})' \Sigma_0^{-1} (\mathbf{x}_i - \hat{\mu})} \right] \mathbf{D}_p.$$

Using the properties of the vec-operator we can rewrite the above as

$$\mathbf{u}_2(\mathbf{X}, \hat{\mu}, \Sigma_0) = -\frac{n}{2} \left[(\Sigma_0^{-1} \otimes \Sigma_0^{-1}) \left(\text{vec } \Sigma_0 - \frac{v+p}{nv} \sum_{i=1}^n \frac{\text{vec } ((\mathbf{x}_i - \hat{\mu})(\mathbf{x}_i - \hat{\mu})')}{1 + \frac{1}{v} (\mathbf{x}_i - \hat{\mu})' \Sigma_0^{-1} (\mathbf{x}_i - \hat{\mu})} \right) \right] \mathbf{D}_p,$$

and, using the expression for \mathbf{V} from (12) we obtain

$$\mathbf{u}_2(\mathbf{X}, \hat{\mu}_0, \Sigma_0) = -\frac{n}{2} (\Sigma_0^{-1} \otimes \Sigma_0^{-1}) \text{vec } \mathbf{V} \cdot \mathbf{D}_p. \tag{13}$$

Observe now that the diagonal block \mathbf{I}_{22} of (11) can be represented as

$$\begin{aligned} \mathbf{I}_{22}(\Sigma_0) &= \frac{n}{2(v+p+2)} \mathbf{D}'_p (\Sigma_0^{-1} \otimes \Sigma_0^{-1}) \\ &\quad \times \left[(v+p)(\Sigma_0 \otimes \Sigma_0) - \text{vec } \Sigma_0 \text{vec}' \Sigma_0 \right] (\Sigma_0^{-1} \otimes \Sigma_0^{-1}) \mathbf{D}_p, \end{aligned}$$

and since the duplication matrix \mathbf{D}_p is of full rank, \mathbf{I}_{22}^{-1} can be expressed using the Moore-Penrose inverse \mathbf{D}_p^+

$$\begin{aligned} \mathbf{I}_{22}^{-1}(\Sigma_0) &= \frac{2(v+p+2)}{n} \mathbf{D}_p^+ (\Sigma_0 \otimes \Sigma_0) \\ &\quad \times \left[(v+p)(\Sigma_0 \otimes \Sigma_0) - \text{vec } \Sigma_0 \text{vec}' \Sigma_0 \right]^{-1} (\Sigma_0 \otimes \Sigma_0) \mathbf{D}_p^{+'}. \end{aligned}$$

The inverse matrix in square brackets we obtain using the inverse binomial theorem, i.e., assuming that all included inverses exist,

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{DA}^{-1} \mathbf{B} + \mathbf{C}^{-1})^{-1} \mathbf{DA}^{-1};$$

cf. (Kollo and von Rosen 2005, p. 75). Setting $\mathbf{A} = (v+p)(\Sigma_0 \otimes \Sigma_0)$, $\mathbf{B} = \text{vec } \Sigma_0$, $\mathbf{C} = -1$ and $\mathbf{D} = \text{vec}' \Sigma_0$, we have

$$\begin{aligned} &\left[(v+p)(\Sigma_0 \otimes \Sigma_0) - \text{vec } \Sigma_0 \text{vec}' \Sigma_0 \right]^{-1} \\ &= \frac{1}{v+p} (\Sigma_0^{-1} \otimes \Sigma_0^{-1}) \\ &\quad - \frac{1}{(v+p)^2} (\Sigma_0^{-1} \otimes \Sigma_0^{-1}) \text{vec } \Sigma_0 \left[\text{vec}' \Sigma_0 \frac{1}{v+p} (\Sigma_0^{-1} \otimes \Sigma_0^{-1}) \text{vec } \Sigma_0 - 1 \right]^{-1} \\ &\quad \times \text{vec}' \Sigma_0 (\Sigma_0^{-1} \otimes \Sigma_0^{-1}), \end{aligned}$$

and since the sum in the square brackets on right-hand side equals $\frac{p}{v+p} - 1$ we have

$$[(v+p)(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0) - \text{vec } \boldsymbol{\Sigma}_0 \text{vec}' \boldsymbol{\Sigma}_0]^{-1} = \frac{1}{v+p} \left((\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1}) + \frac{1}{v} \text{vec } \boldsymbol{\Sigma}_0^{-1} \text{vec}' \boldsymbol{\Sigma}_0^{-1} \right).$$

Hence

$$\begin{aligned} \mathbf{I}_{22}^{-1}(\boldsymbol{\Sigma}_0) &= \frac{2(v+p+2)}{n(v+p)} \\ &\quad \times \mathbf{D}_p^+(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0) \left[(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1}) + \frac{1}{v} \text{vec } \boldsymbol{\Sigma}_0^{-1} \text{vec}' \boldsymbol{\Sigma}_0^{-1} \right] (\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0) \mathbf{D}_p^{+\prime} \\ &= \frac{2(v+p+2)}{nv(v+p)} \mathbf{D}_p^+ \left[v(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0) + \text{vec } \boldsymbol{\Sigma}_0 \text{vec}' \boldsymbol{\Sigma}_0 \right] \mathbf{D}_p^{+\prime}. \end{aligned} \quad (14)$$

Combining (13) and (14),

$$\begin{aligned} \text{RST}(\mathbf{X}, \boldsymbol{\Sigma}_0) &= \mathbf{u}_2'(\hat{\boldsymbol{\mu}}_0, \boldsymbol{\Sigma}_0) \mathbf{I}_{22}^{-1}(\boldsymbol{\Sigma}_0) \mathbf{u}_2(\hat{\boldsymbol{\mu}}_0, \boldsymbol{\Sigma}_0) \\ &= \frac{n(v+p+2)}{2v(v+p)} \text{vec}' \mathbf{V} (\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1}) \mathbf{D}_p \mathbf{D}_p^+ \\ &\quad \times \left[v(\boldsymbol{\Sigma}_0 \otimes \boldsymbol{\Sigma}_0) + \text{vec } \boldsymbol{\Sigma}_0 \text{vec}' \boldsymbol{\Sigma}_0 \right] \mathbf{D}_p^{+\prime} \mathbf{D}_p' (\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1}) \text{vec } \mathbf{V}. \end{aligned}$$

From (Magnus and Neudecker 1986, formula (54)) it is known that $\mathbf{D}_p \mathbf{D}_p^+ = \mathbf{N}_p$, where $\mathbf{N}_p = \frac{1}{2}(\mathbf{I}_{p^2} + \mathbf{K}_{p,p})$ with $\mathbf{K}_{p,p}$ being the commutation matrix, i.e., the matrix for which $\mathbf{K}_{p,p} \text{vec } \mathbf{A} = \text{vec } \mathbf{A}'$. Thus, for a symmetric matrix \mathbf{V} we have $\mathbf{N}_p \text{vec } \mathbf{V} = \text{vec } \mathbf{V}$. Moreover, since $(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1}) \mathbf{N}_p = \mathbf{N}_p (\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})$, we obtain

$$\begin{aligned} \text{RST}(\mathbf{X}, \boldsymbol{\Sigma}_0) &= \frac{n(v+p+2)}{2v(v+p)} \text{vec}' \mathbf{V} \left[v(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1}) + \text{vec } \boldsymbol{\Sigma}_0^{-1} \text{vec}' \boldsymbol{\Sigma}_0^{-1} \right] \text{vec } \mathbf{V} \\ &= \frac{n(v+p+2)}{2v(v+p)} \left[v \cdot \text{vec}' (\boldsymbol{\Sigma}_0^{-1} \mathbf{V} \boldsymbol{\Sigma}_0^{-1}) \text{vec } \mathbf{V} + \text{vec}' \mathbf{V} \text{vec } \boldsymbol{\Sigma}_0^{-1} \text{vec}' \boldsymbol{\Sigma}_0^{-1} \text{vec } \mathbf{V} \right]. \end{aligned}$$

Finally, again using the relation between the vec-operator and the trace function, we obtain

$$\text{RST}(\mathbf{X}, \boldsymbol{\Sigma}_0) = \frac{n(v+p+2)}{2v(v+p)} \left\{ v \cdot \text{tr}[(\mathbf{V} \boldsymbol{\Sigma}_0^{-1})^2] + \text{tr}^2(\mathbf{V} \boldsymbol{\Sigma}_0^{-1}) \right\}.$$

The limiting distribution follows from e.g. Rao (2005). \square

Note that if $v \rightarrow \infty$, the $\hat{\boldsymbol{\mu}}_0$ becomes the average from the observations, $\bar{\mathbf{x}}$, and

$$\mathbf{V} = \boldsymbol{\Sigma}_0 - \frac{1}{n} \mathbf{X} \mathbf{Q}_n \mathbf{X}' = \boldsymbol{\Sigma}_0 - \mathbf{S},$$

where $\mathbf{S} = \hat{\boldsymbol{\Sigma}}$ under normality. Thus, the formula for RST reduces to the formula for RST under normality

$$\text{RST} \rightarrow \frac{n}{2} \text{tr} \left[(\mathbf{I}_p - \mathbf{S} \boldsymbol{\Sigma}_0^{-1})^2 \right];$$

cf. Kollo et al. (2016).

Finally we mention that the RST statistic can also be determined using the formulae stated in Sutradhar (1993), where Neyman's score test (Neyman (1959)) for testing covariance structure when n is large and $\boldsymbol{\mu}$ and v are unknown is proposed.

5 Wald test statistic

The WT statistic is a function of the MLE and the Fisher information matrix.

Definition 6 The WT statistic for testing hypothesis (3) is given by

$$WT(\mathbf{X}, \boldsymbol{\theta}_0) = (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' \mathbf{I}(\mathbf{X}, \widehat{\boldsymbol{\theta}}) (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_r)'$, $\widehat{\boldsymbol{\theta}}$ is the MLE of $\boldsymbol{\theta}$ and \mathbf{X} is a random sample from a $P_{\boldsymbol{\theta}}$ -distributed population.

When sample size $n \rightarrow \infty$ and H_0 holds then, similarly to the previous cases, the distribution of $WT(\mathbf{X}, \boldsymbol{\theta}_0)$ converges to χ_r^2 ; cf. Wald (1943). As in the case of the RST, if the restrictions in the null hypothesis are imposed only on $r - k$ parameters, the WT can be represented as

$$WT(\mathbf{X}, \boldsymbol{\theta}_0) = (\widehat{\boldsymbol{\theta}}_{r-k} - \boldsymbol{\theta}_0)' \mathbf{I}_{1,2}(\mathbf{X}, \widehat{\boldsymbol{\theta}}) (\widehat{\boldsymbol{\theta}}_{r-k} - \boldsymbol{\theta}_0), \tag{15}$$

where $\widehat{\boldsymbol{\theta}}_{r-k}$ is the MLE of $\boldsymbol{\theta}_{r-k}$ under the alternative and $\mathbf{I}_{1,2}(\mathbf{X}, \widehat{\boldsymbol{\theta}})$ is the Schur complement of \mathbf{I}_{22} in the information matrix $\mathbf{I}(\mathbf{X}, \widehat{\boldsymbol{\theta}})$; cf. Rao (2005). When the null hypothesis $H_0 : \boldsymbol{\theta}_{r-k} = \boldsymbol{\theta}_0$ holds, the distribution of the WT statistic converges to χ_{r-k}^2 .

Let us rewrite the hypothesis (7) in equivalent form as

$$\begin{cases} H_0 : \text{vech } \boldsymbol{\Sigma} = \text{vech } \boldsymbol{\Sigma}_0, \\ H_1 : \text{vech } \boldsymbol{\Sigma} \neq \text{vech } \boldsymbol{\Sigma}_0, \end{cases} \tag{16}$$

since the vector of unknown parameters consists of $\boldsymbol{\mu}$ and $\text{vech } \boldsymbol{\Sigma}$. Then the following proposition for the WT statistic can be formulated.

Proposition 3 *The WT statistic for testing (7) when no constraints are imposed on $\boldsymbol{\mu}$ is given by*

$$WT(\mathbf{X}, \boldsymbol{\Sigma}_0) = \frac{n}{2(v+p+2)} \left\{ (v+p) \cdot \text{tr} \left[\left(\mathbf{I}_p - \boldsymbol{\Sigma}_0 \widehat{\boldsymbol{\Sigma}}^{-1} \right)^2 \right] - \text{tr}^2 \left(\mathbf{I}_p - \boldsymbol{\Sigma}_0 \widehat{\boldsymbol{\Sigma}}^{-1} \right) \right\},$$

where $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ is a sample from $t_{v,p}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $v > 2$ is known, and $\widehat{\boldsymbol{\Sigma}}$ is the solution of (5).

When $n \rightarrow \infty$ and H_0 holds, the distribution of $WT(\mathbf{X}, \boldsymbol{\Sigma}_0)$ tends to the chi-square distribution with $p(p + 1)/2$ degrees of freedom.

Proof For testing (16), due to (15) the WT statistic can be expressed as

$$WT(\mathbf{X}, \boldsymbol{\Sigma}_0) = \text{vech}'(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0) \mathbf{I}_{22}(\widehat{\boldsymbol{\Sigma}}) \text{vech}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0)$$

since $I_{2,1}(\mathbf{X}, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) = I_{22}(\widehat{\boldsymbol{\Sigma}})$ is the lower diagonal block of the Fisher information matrix in (11) with $\boldsymbol{\Sigma}$ replaced by its MLE, that is

$$I_{22}(\widehat{\boldsymbol{\Sigma}}) = \frac{n}{2(v+p+2)} \mathbf{D}'_p \left[(v+p)(\widehat{\boldsymbol{\Sigma}}^{-1} \otimes \widehat{\boldsymbol{\Sigma}}^{-1}) - \text{vec } \widehat{\boldsymbol{\Sigma}}^{-1} \text{vec}' \widehat{\boldsymbol{\Sigma}}^{-1} \right] \mathbf{D}_p.$$

We obtain

$$\begin{aligned} \text{WT}(\mathbf{X}, \boldsymbol{\Sigma}_0) &= \frac{n}{2(v+p+2)} \text{vech}'(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0) \mathbf{D}'_p \\ &\quad \times \left[(v+p)(\widehat{\boldsymbol{\Sigma}}^{-1} \otimes \widehat{\boldsymbol{\Sigma}}^{-1}) - \text{vec } \widehat{\boldsymbol{\Sigma}}^{-1} \text{vec}' \widehat{\boldsymbol{\Sigma}}^{-1} \right] \mathbf{D}_p \text{vech}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0), \end{aligned}$$

and since $\mathbf{D}'_p \text{vech}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0) = \text{vec}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0)$ (cf. Magnus and Neudecker 1986, formula (49)) we obtain

$$\begin{aligned} \text{WT}(\mathbf{X}, \boldsymbol{\Sigma}_0) &= \frac{n}{2(v+p+2)} \text{vec}'(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0) \\ &\quad \times \left[(v+p)(\widehat{\boldsymbol{\Sigma}}^{-1} \otimes \widehat{\boldsymbol{\Sigma}}^{-1}) - \text{vec } \widehat{\boldsymbol{\Sigma}}^{-1} \text{vec}' \widehat{\boldsymbol{\Sigma}}^{-1} \right] \text{vec}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0) \\ &= \frac{n}{2(v+p+2)} \left[\text{vec}'(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0)(v+p)(\widehat{\boldsymbol{\Sigma}}^{-1} \otimes \widehat{\boldsymbol{\Sigma}}^{-1}) \text{vec}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0) \right. \\ &\quad \left. - \text{vec}'(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0) \text{vec } \widehat{\boldsymbol{\Sigma}}^{-1} \text{vec}' \widehat{\boldsymbol{\Sigma}}^{-1} \text{vec}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0) \right] \\ &= \frac{n(v+p)}{2(v+p+2)} \text{vec}'(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0) \text{vec}(\widehat{\boldsymbol{\Sigma}}^{-1}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0)\widehat{\boldsymbol{\Sigma}}^{-1}) \\ &\quad - \frac{n}{2(v+p+2)} \text{vec}'(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0) \text{vec } \widehat{\boldsymbol{\Sigma}}^{-1} \text{vec}' \widehat{\boldsymbol{\Sigma}}^{-1} \text{vec}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0). \end{aligned}$$

Using the relation between the vec-operator and the trace function, we obtain

$$\text{WT}(\mathbf{X}, \boldsymbol{\Sigma}_0) = \frac{n(v+p)}{2(v+p+2)} \text{tr} \left\{ \left[(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0) \widehat{\boldsymbol{\Sigma}}^{-1} \right]^2 \right\} - \frac{n}{2(v+p+2)} \text{tr}^2 \left[(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}_0) \widehat{\boldsymbol{\Sigma}}^{-1} \right].$$

which is the same as stated in the proposition. The limiting distribution follows from e.g. Rao (2005). \square

Note that when $v \rightarrow \infty$, $\widehat{\boldsymbol{\Sigma}}$ tends to the MLE of $\boldsymbol{\Sigma}$ under normality, \mathbf{S} , and WT converges to the corresponding expression of the WT statistic under normality

$$\text{WT}(\mathbf{X}, \boldsymbol{\Sigma}_0) \rightarrow \frac{n}{2} \text{tr} \left[\left(\mathbf{I}_p - \boldsymbol{\Sigma}_0 \mathbf{S}^{-1} \right)^2 \right];$$

cf. Kollo et al. (2016).

In addition, let us consider the test statistic WT^* associated with WT in such a way that in the information matrix in Definition 6, $\widehat{\boldsymbol{\theta}}$ is replaced by $\boldsymbol{\theta}_0$. Then, using the same arguments as in the proof of Proposition 3, we obtain

$$\text{WT}^*(\mathbf{X}, \boldsymbol{\Sigma}_0) = \frac{n}{2(v+p+2)} \left\{ (v+p) \cdot \text{tr} \left[\left(\mathbf{I}_p - \widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}_0^{-1} \right)^2 \right] - \text{tr}^2 \left(\mathbf{I}_p - \widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}_0^{-1} \right) \right\},$$

which is simply WT with the roles of $\widehat{\boldsymbol{\Sigma}}$ and $\boldsymbol{\Sigma}_0$ exchanged. In Sect. 6 we show using Monte Carlo simulations that under the null hypothesis and increasing sample size

($n \rightarrow \infty$), the distribution of WT^* tends to the chi-square distribution with $p(p+1)/2$ degrees of freedom.

Finally note that when $\nu \rightarrow \infty$, i.e., when the multivariate t -distribution converges to multivariate normal, WT^* becomes the RST.

6 Simulations

In the simulation studies we examine the convergence to the asymptotic chi-square distributions of the test statistics given in Propositions 1–3 as well as the WT^* statistic. We are interested in several problems:

1. How does the speed of convergence depend on the number of degrees of freedom ν and the sample size n ?
2. How does the behavior of the statistics change when ν is growing?
3. What happens when we replace the MLEs of the $t_{p,\nu}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ -distribution by the corresponding MLEs of the normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$? From a practical point of view it is important to know whether we need to calculate MLEs numerically for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ for a $t_{p,\nu}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ -distribution or whether we can just plug in the sample mean $\bar{\mathbf{x}}$ and the sample covariance matrix \mathbf{S} instead.
4. Which of the derived statistics, LRT, RST, WT or WT^* , behaves best and can be recommended for use in data analysis?
5. What is the empirical type I error of the considered statistics?
6. Is it possible to indicate the test statistic with the highest power?

Note that a linear transformation of a multivariate $t_{p,\nu}$ -distributed vector, $\boldsymbol{\Sigma}_0^{-1/2}\mathbf{x}$, where $\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Sigma}_0^{-1/2} = \boldsymbol{\Sigma}_0^{-1}$ is a known nonsingular positive definite (p.d.) matrix, is still multivariate $t_{p,\nu}$ -distributed with the same number of degrees of freedom, location parameter $\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\mu}$ and scale matrix $\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1/2}$; cf. Kotz and Nadarajah (2004). Such a transformation allows us to simplify the null hypothesis and test $H_0 : \boldsymbol{\Sigma} = \mathbf{I}_p$ instead of $H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$. Thus, without loss of generality, in the simulation study we use the simplest possible values of the parameters, $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_p$. We fix the dimension to $p = 3, 9$, the sample size to $n = 10, 25, 50$, and the number of degrees of freedom to $\nu = 3, 10$; however, for $p = 9$ we increase the smallest sample size to $n = 12$, to avoid possible ill-conditioned matrices which may appear in the algorithm and cause perturbation of the empirical distribution of the test statistics. The number of simulation runs is 10,000. The results are obtained using *Mathematica* software.

6.1 Convergence of test statistics to the limiting distribution

In the first row of Figs. 1, 2 and 3 we can see the empirical null distribution of LRT, RST, WT and WT^* and the limiting χ^2 distribution for 3-dimensional data generated from a multivariate t -distribution with $\nu = 3$ degrees of freedom, and, respectively, sample sizes $n = 10, 25, 50$. We can see that the distribution of RST best matches the limiting distribution, while the fit of LRT and WT^* is somewhat worse, especially when the 95th quantiles are compared; cf. Table 1. In addition it can be observed

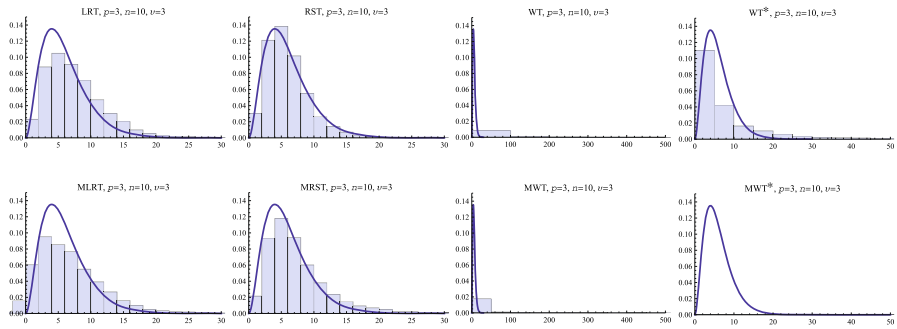


Fig. 1 Empirical null distributions of LRT, RST, WT and WT* (first row) as well as MLRT, MRST, MWT and MWT* (second row) for $\nu = p = 3$, $n = 10$, with limiting χ_6^2

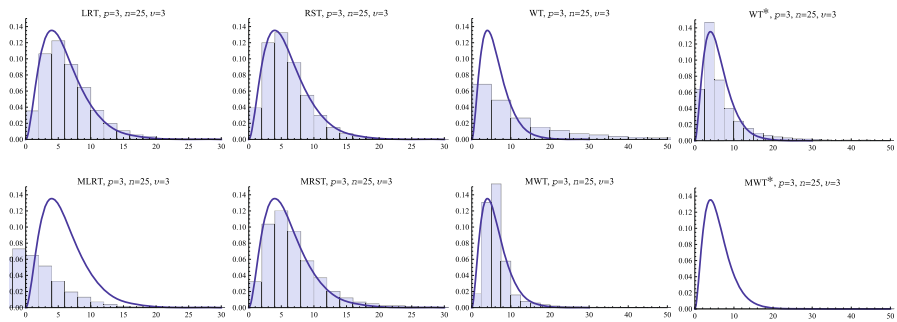


Fig. 2 Empirical null distributions of LRT, RST, WT and WT* (first row) as well as MLRT, MRST, MWT and MWT* (second row) for $\nu = p = 3$, $n = 25$, with limiting χ_6^2

that the convergence of the distribution of WT is the slowest, and even for $n = 50$ the 95th quantile is far from the corresponding chi-square quantile. It should also be noted that the quantile values of WT statistics are much higher than the values of other test statistics. This fact can be observed in the graphs as well as in Table 1. The same conclusions can be drawn from Figs. 4, 5 and 6, where respective distributions for $\nu = 10$ are presented.

Since the MLEs under a $t_{p,\nu}$ -distribution are not available in explicit form, we also performed simulations of the distributions of all four test statistics when the MLEs are replaced by MLEs of μ and Σ coming from the normal distribution $N_p(\mu, \Sigma)$. If the solutions of (5), $\hat{\mu}$ and $\hat{\Sigma}$, are replaced by \bar{x} and S , respectively, these modified test statistics will be denoted accordingly by MLRT, MRST, MWT and MWT*. The empirical null distributions of these modified test statistics are presented in the second rows of Figs. 1, 2 and 3 for $\nu = 3$, and Figs. 4, 5 and 6 for $\nu = 10$. In all these figures and in Table 1 we see that in all considered cases the MRST distribution does not differ significantly from the distribution of the original RST. Moreover, for larger ν the differences are almost not perceptible. This phenomenon does not hold in the case of MLRT, as increasing sample size destroys the convergence to the chi-square distribution (higher sample size causes a shift to the left). Surprisingly, with (M)WT the

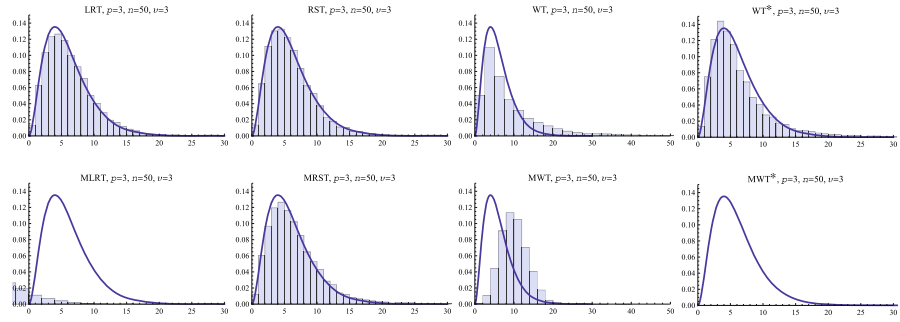


Fig. 3 Empirical null distributions of LRT, RST, WT and WT* (first row) as well as MLRT, MRST, MWT and MWT* (second row) for $\nu = p = 3, n = 50$, with limiting χ^2_6

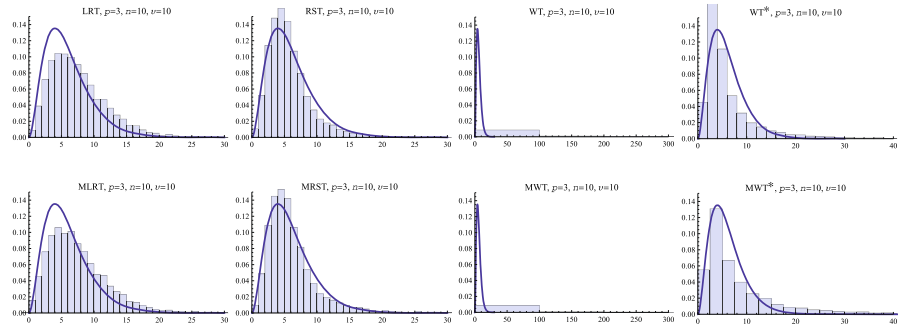


Fig. 4 Empirical null distributions of LRT, RST, WT and WT* (first row) as well as MLRT, MRST, MWT and MWT* (second row) for $\nu = 10, p = 3, n = 10$, with limiting χ^2_6

convergence improved in most of the cases, while the opposite occurred with (M)WT*. Observe, however, that for increasing ν the convergence of the distributions of modified test statistics improves compared with $\nu = 3$, since the $t_{p,\nu}$ -distribution is closer to normal. The discrepancy between normal and $t_{p,\nu}$ distributions can be measured by, for example, negentropy (cf. Osorio et al. (2023)), which for fixed degrees of freedom can be calculated as

$$H(p, \nu) = \frac{p}{2} (1 + \ln(2\pi)) + \ln K_p(\nu) - \frac{\nu+p}{2} (\psi(\frac{\nu+p}{2}) - \psi(\frac{\nu}{2})),$$

where $K_p(\nu) = ((\nu - 2)\pi)^{-(p/2)} \Gamma((\nu + p)/2) / \Gamma(\nu/2)$ and $\psi(z)$ is the digamma function. It can be seen in Fig. 13 that for $\nu = 10$ and $p = 3$ the negentropy is already very low, around 0.041, while for $p = 9$ it is around 0.20.

Figures 7, 8, 9, 10, 11 and 12 present the empirical null distributions of the considered statistics for $p = 9$ with the limiting χ^2_{45} -distribution. In addition, in Table 2 the 95th empirical quantiles of the respective distributions are given. It can be seen that the null distribution of RST still matches well the limiting distribution even for very small sample sizes. The convergence of the remaining distributions is now slower, and larger differences between them can be observed; in this case LRT outperforms WT*,

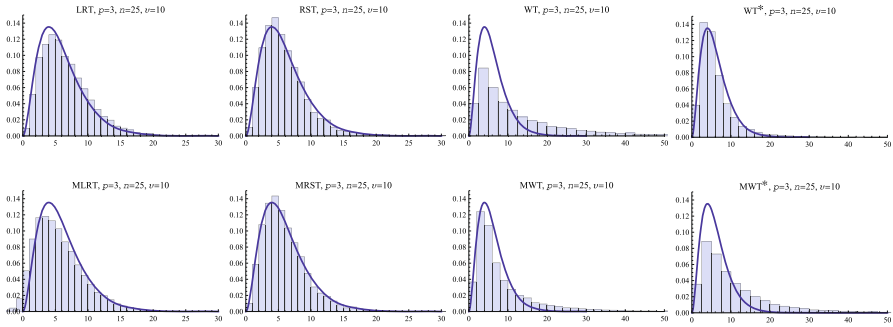


Fig. 5 Empirical null distributions of LRT, RST, WT and WT^* (first row) as well as MLRT, MRST, MWT and MWT^* (second row) for $\nu = 10$, $p = 3$, $n = 25$, with limiting χ_6^2

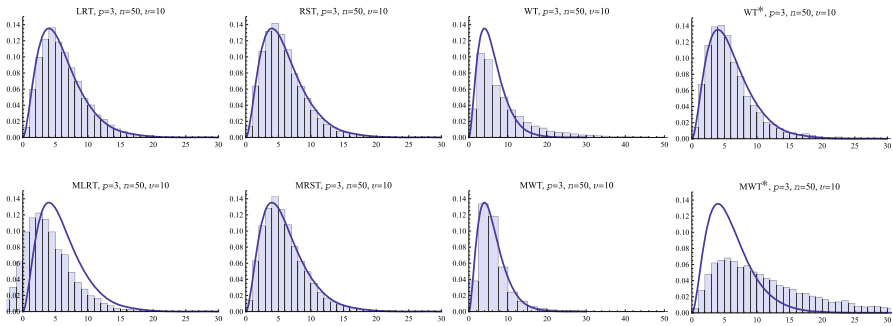


Fig. 6 Empirical null distributions of LRT, RST, WT and WT^* (first row) as well as MLRT, MRST, MWT and MWT^* (second row) for $\nu = 10$, $p = 3$, $n = 50$, with limiting χ_6^2

and the values of WT are huge, especially for small sample sizes. The conclusions concerning modified versions of the test statistics are the same as in the case $p = 3$; still, the MRST distribution outperforms all other modified test statistics.

Summing up, we can formulate the following conclusions:

- for LRT, RST, WT and WT^* :
 - the convergence of RST to the limiting chi-square distribution is quicker than with the remaining test statistics; moreover, the distribution of WT does not fit to the limiting distribution, even for large sample size;
 - usually there is no significant difference between the empirical distributions of all four test statistics for $\nu = 3$ and $\nu = 10$;
- for MLRT, MRST, MWT and MWT^* :
 - except for MRST, the distributions of the modified test statistics do not fit well to the theoretical chi-square distribution; moreover, for MLRT and MWT the fit becomes worse with increasing sample size, and MWT^* produces extremely high values, especially for small numbers of degrees of freedom;
 - comparing MWT and WT, it seems that the replacement of the original MLEs by the MLEs of parameters of a normal distribution improves the fit of the test

Table 1 The 95th quantiles of the empirical null distributions of LRT, RST, WT, WT* and their modified versions for $\nu = 3, 10, p = 3$ and $n = 10, 25, 50$, as well as the 95th quantile of χ_6^2

n	$\nu = 3$			$\nu = 10$		
	10	25	50	10	25	50
LRT	15.765	13.687	13.306	15.904	13.691	13.028
RST	12.237	12.375	12.505	12.628	12.286	12.366
WT	326.196	47.805	26.413	252.731	44.842	25.315
WT*	34.047	19.930	16.216	18.895	15.013	13.605
MLRT	16.387	8.504	0.162	15.773	12.804	10.945
MRST	18.411	15.315	14.298	13.389	12.568	12.470
MWT	107.431	14.156	15.982	190.128	28.136	14.427
MWT*	638.37	1023.15	1726.11	29.733	31.761	34.652
χ_6^2	12.592	12.592	12.592	12.592	12.592	12.592

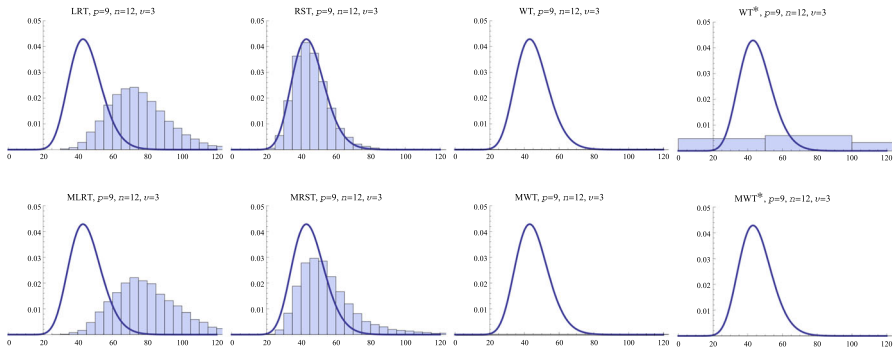


Fig. 7 Empirical null distributions of LRT, RST, WT and WT* (first row) as well as MLRT, MRST, MWT and MWT* (second row) for $\nu = 3, p = 9, n = 12$, with limiting χ_{45}^2

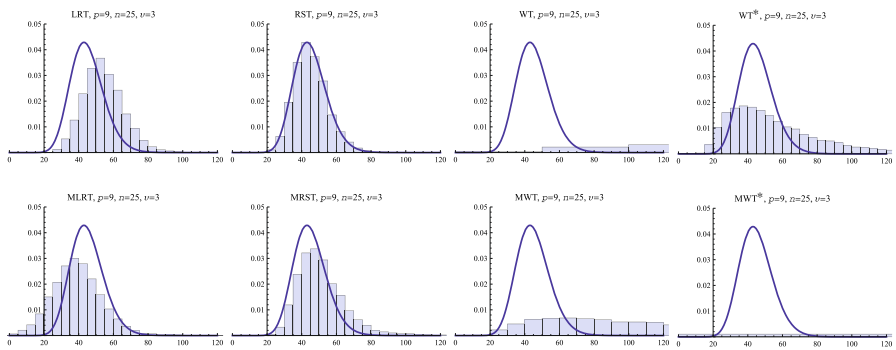


Fig. 8 Empirical null distributions of LRT, RST, WT and WT* (first row) as well as MLRT, MRST, MWT and MWT* (second row) for $\nu = 3, p = 9, n = 25$, with limiting χ_{45}^2

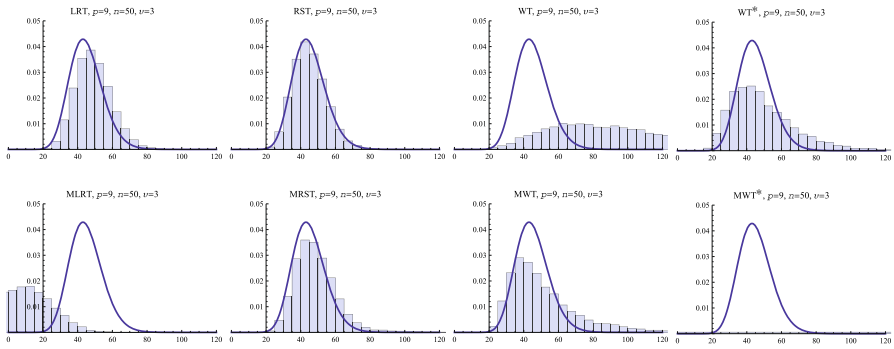


Fig. 9 Empirical null distributions of LRT, RST, WT and WT* (first row) as well as MLRT, MRST, MWT and MWT* (second row) for $v = 3, p = 9, n = 50$, with limiting χ^2_{45}

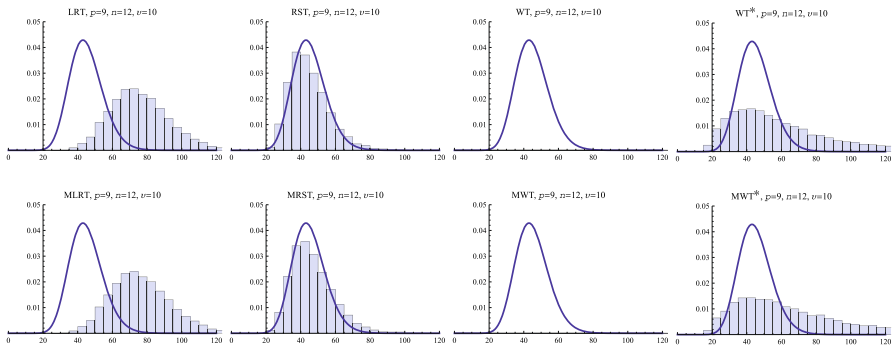


Fig. 10 Empirical null distributions of LRT, RST, WT and WT* (first row) as well as MLRT, MRST, MWT and MWT* (second row) for $v = 10, p = 9, n = 12$, with limiting χ^2_{45}

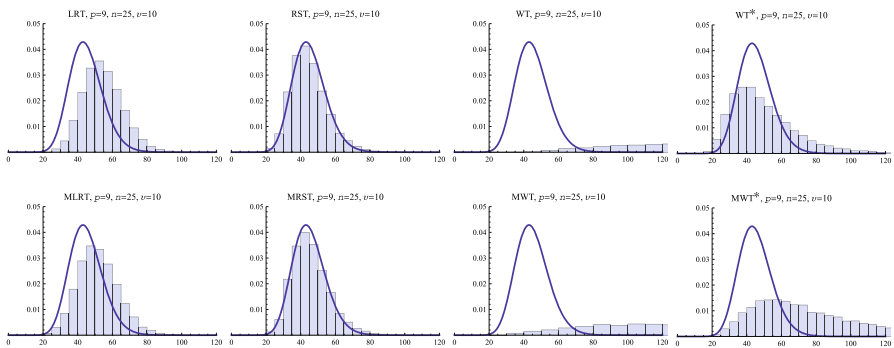


Fig. 11 Empirical null distributions of LRT, RST, WT and WT* (first row) as well as MLRT, MRST, MWT and MWT* (second row) for $v = 10, p = 9, n = 25$, with limiting χ^2_{45}

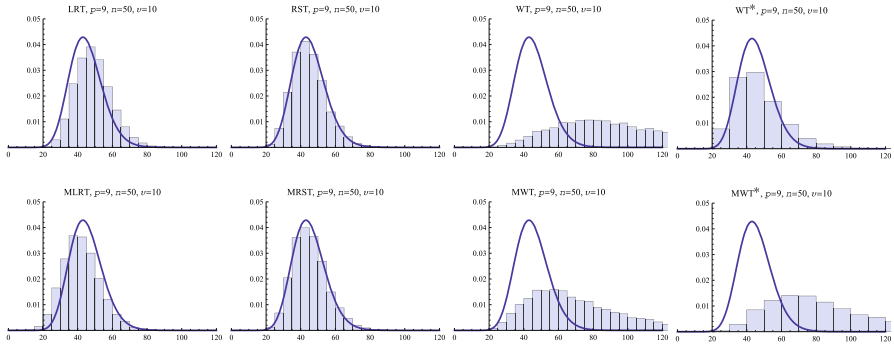


Fig. 12 Empirical null distributions of LRT, RST, WT and WT* (first row) as well as MLRT, MRST, MWT and MWT* (second row) for $\nu = 10, p = 9, n = 50$, with limiting χ^2_{45}

Table 2 The 95th quantiles of the empirical null distributions of LRT, RST, WT, WT* and their modified versions for $\nu = 3, 10, p = 9$ and $n = 12, 25, 50$, as well as the 95th quantile of χ^2_{45}

n	$\nu = 3$			$\nu = 10$		
	12	25	50	12	25	50
LRT	105.860	74.447	67.421	106.792	74.341	67.4085
RST	64.248	62.924	61.935	65.957	63.972	62.970
WT	473,798.000	869.174	217.420	283,888.906	802.834	211.011
WT*	524.012	120.539	87.755	142.372	88.657	74.438
MLRT	119.623	64.084	33.104	107.356	71.728	61.248
MRST	99.923	79.304	70.872	69.336	65.649	63.805
MWT	223589.647	404.04	95.171	266970.342	615.176	146.551
MWT*	6064.220	8552.336	11713.578	174.415	167.833	162.319
χ^2_{45}	61.656	61.656	61.656	61.656	61.656	61.656

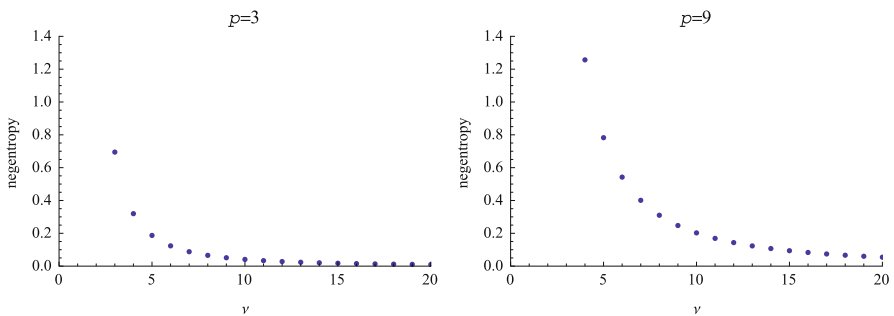


Fig. 13 Values of negentropy function between normal and t -distribution for $p = 3$ and $p = 9$

Table 3 The empirical Type I error for LRT, RST, WT and WT* for $\nu = 3, 10$, $p = 3, 9$ and various values of n

n	$\nu = 3$				$\nu = 10$			
	LRT	RST	WT	WT*	LRT	RST	WT	WT*
	$p = 3$							
10	0.125	0.045	0.612	0.188	0.128	0.051	0.623	0.108
25	0.072	0.047	0.337	0.124	0.074	0.045	0.351	0.079
50	0.063	0.048	0.219	0.095	0.058	0.046	0.210	0.064
100	0.056	0.049	0.137	0.078	0.054	0.048	0.136	0.059
500	0.051	0.050	0.070	0.055	0.053	0.053	0.069	0.058
	$p = 9$							
12	0.771	0.071	0.971	0.672	0.780	0.083	0.981	0.405
25	0.244	0.061	0.966	0.337	0.253	0.067	0.985	0.221
50	0.118	0.053	0.791	0.223	0.117	0.062	0.831	0.142
100	0.080	0.053	0.512	0.146	0.076	0.051	0.541	0.091
500	0.055	0.050	0.142	0.076	0.057	0.051	0.142	0.062

statistic distribution to the theoretical chi-square; this is not the case for the remaining pairs of statistics.

It is also worth noting that similar conclusions were reached by Kollo et al. (2016), where multivariate normality of the distribution of the observation matrix was assumed, i.e., the convergence of RST to the limiting chi-square distribution is quicker than with LRT and WT.

6.2 Type I error

In this section we start with a comparison of the convergence of the empirical Type I errors to the nominal significance level $\alpha = 0.05$. To reject the true null hypothesis we used the quantiles of the limiting chi-square distribution. The results are given in Table 3.

We can observe that the empirical Type I errors converge to the nominal significance level with increasing sample size for all statistics; however, the convergence of RST is very quick and, even for small sample size, the Type I error is close to 0.05. We may also notice that in this comparison LRT usually outperforms WT*, except for small sample sizes, especially when $\nu = 10$ and p is growing. Finally, note that the convergence of the type I error of the WT statistic is extremely slow, and even a sample size of $n = 500$ is not enough to reach the nominal significance level. This behavior follows from the poor convergence of the distribution of WT to the limiting distribution, which was indicated in Sect. 6.1.

6.3 Power comparison

In this section we use a Monte Carlo simulation study to examine the behavior of the power functions of RST, LRT, WT and WT* with respect to the discrepancy between the null and alternative hypotheses in (7), and with respect to the sample size. As a measure of discrepancy between two distributions with different covariance matrices, we use Stein's loss function (cf. Stein (1956)) in the form

$$\zeta(\Sigma, \Sigma_0) = \text{tr}(\Sigma^{-1}\Sigma_0) - \ln|\Sigma^{-1}\Sigma_0| - p,$$

where Σ is the scale matrix under the alternative hypothesis and Σ_0 is the scale matrix under the null hypothesis. Since for the simulations we assumed $\Sigma_0 = \mathbf{I}_p$ in (7), the above function reduces to

$$\zeta(\Sigma) = \text{tr}\Sigma^{-1} - \ln|\Sigma^{-1}| - p. \quad (17)$$

Observing that Stein's loss function is not upper bounded, we use as a discrepancy measure $\eta(\Sigma) = 1 - 1/(1 - \zeta(\Sigma))$, to restrict the possible discrepancy to the interval $[0,1)$.

In power comparison we first set the parameters of the experiment as $p = 3$, $\nu = 3, 10$ and $n = 10, 25, 50$. We generate 100 p.d. matrices Σ , for which the discrepancies $\eta(\Sigma)$ are computed. Note, that since $\eta(\Sigma)$ is a function of Σ^{-1} , and since obviously $\Sigma_0^{-1} = \Sigma_0$ is diagonal matrix, we randomly choose tridiagonal Toeplitz matrices with positive entries on the diagonal, bigger than the off-diagonal elements, to guarantee positive definiteness, and we obtain Σ by inverting generated matrices. Then, for each Σ we generate 10,000 observation matrices from $t_{p,\nu}(\mu, \Sigma)$, and for every generated matrix we test the hypothesis (7) using the quantiles of the empirical null distribution of the relevant statistic. In all comparisons the significance level 0.05 is used. In this way we obtain 100 values of the power of each test, computed as the ratio between the number of rejected null hypotheses and the number of simulation runs (10,000). The results are presented in Figs. 14 and 15. Since there are no significant differences between the powers of tests when $\nu = 3$ and $\nu = 10$, we repeat the above procedure for $p = 9$ only for $\nu = 3$. The results for $n = 12, 25, 50$ are presented in Fig. 16.

In all these graphs the power shows an upward trend when the discrepancy increases. Comparing LRT and RST, slightly higher deviations between powers are noted for RST than for LRT. A similar phenomenon was observed by Filipiak et al. (2024), where various discrepancy measures were studied in the context of testing separability under doubly multivariate models. Moreover, in Fig. 17 RST has lower power than LRT, especially for small sample sizes; nevertheless, there are still alternatives for which the power of RST exceeds the power of LRT.

In the case of WT and WT* it is noted that for two equally distant matrices Σ the power differs significantly, and is often below the nominal significance level, even for large samples. This means that both tests are biased. A similar observation for testing of independence under a block compound symmetry structure in a doubly multivariate

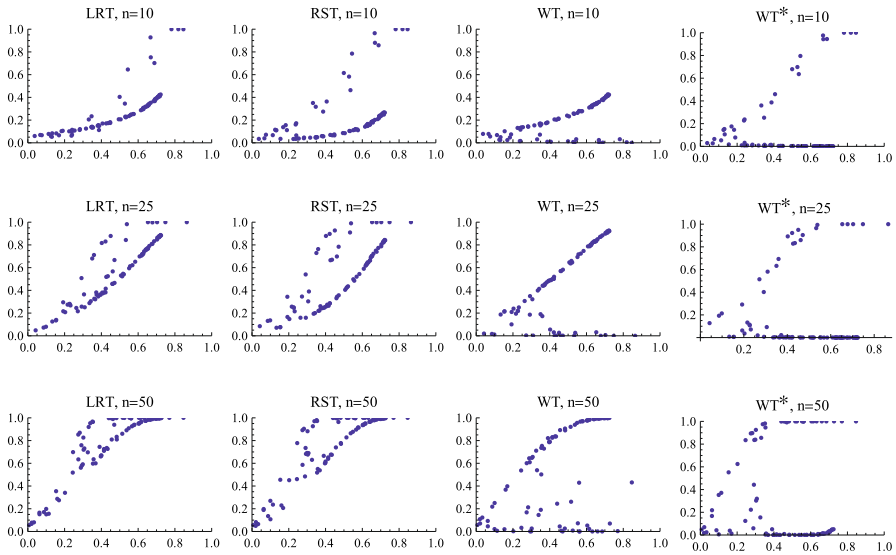


Fig. 14 Empirical power of LRT, RST, WT and WT^* for $\nu = 3$, $p = 3$, $n = 10, 25, 50$ (in columns)

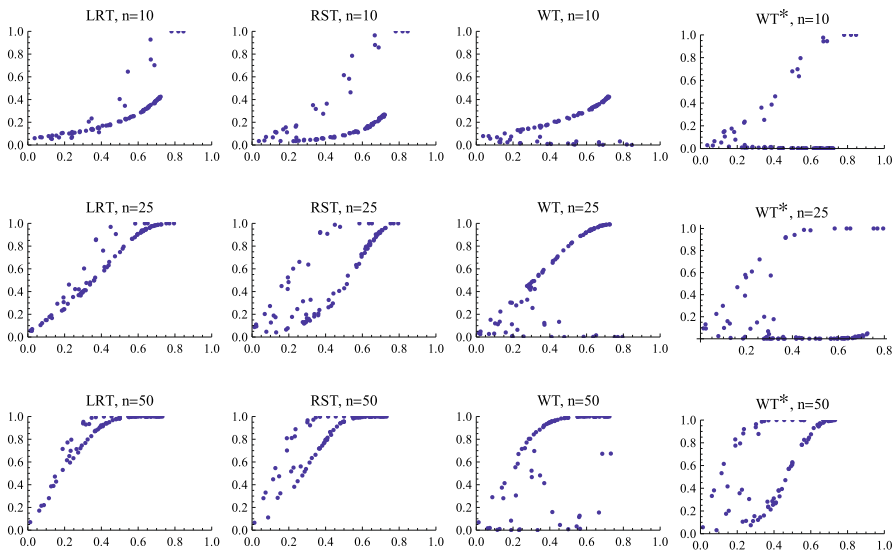


Fig. 15 Empirical power of LRT, RST, WT and WT^* for $\nu = 10$, $p = 3$, $n = 10, 25, 50$ (in columns)

normal model was made in Filipiak et al. (2023). It should also be mentioned that the power of WT increases very slowly with the discrepancy, and for small sample sizes it is below 0.5 even if the discrepancy becomes large.

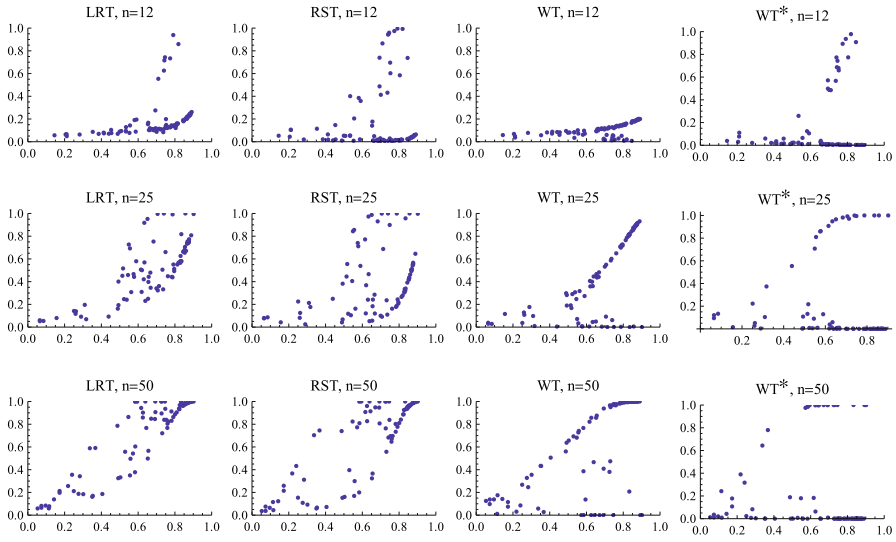


Fig. 16 Empirical power of LRT, RST, WT and WT* for $\nu = 3, p = 9, n = 12, 25, 50$ (in columns)

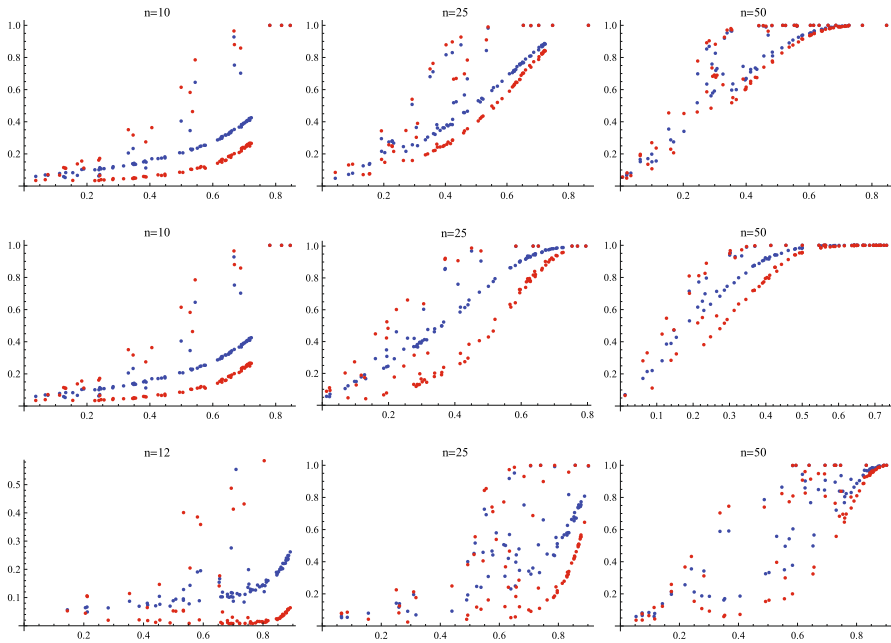


Fig. 17 Empirical power of LRT (blue) and RST (red) for $p = 3, \nu = 3$ (first row), $p = 3, \nu = 10$ (second row), $p = 9, \nu = 3$ (third row) for various values of sample size

7 Test statistics under unknown degrees of freedom

Following (McLachlan and Krishnan 1997, Sect. 5.8) to estimate ν in multivariate t distribution it is enough to compute $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ as the solution of (6), with ν replaced in each step by $\nu^{(k)}$ obtained as the solution of

$$\begin{aligned} \Psi\left(\frac{\nu}{2}\right) - \ln \frac{\nu}{2} &= \\ &= 1 + \frac{1}{n} \sum_{i=1}^n \left[\ln \frac{\nu^{(k-1)+p}}{\nu^{(k-1)+\delta_i^{(k)}}} - \frac{\nu^{(k-1)+p}}{\nu^{(k-1)+\delta_i^{(k)}}} \right] + \Psi\left(\frac{\nu^{(k-1)+p}}{2}\right) - \ln \frac{\nu^{(k-1)+p}}{2} \end{aligned} \quad (18)$$

with $\Psi(\cdot)$ being digamma function and $\delta_i^{(k)} = (\mathbf{x}_i - \boldsymbol{\mu}^{(k)})' \boldsymbol{\Sigma}^{(k)-1} (\mathbf{x}_i - \boldsymbol{\mu}^{(k)})$. Similarly, the ML estimator of ν under null hypothesis is simply the solution of (8) with ν replaced by $\nu^{(k)}$ obtained as the solution of (18).

Denoting by $\widehat{\boldsymbol{\mu}}$, $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\nu}$ the ML estimators under H_1 and by $\widehat{\boldsymbol{\mu}}_0$, and $\widehat{\nu}_0$ respective ML estimators under H_0 , we can formulate the following test statistics for testing (7) with no constraints imposed on $\boldsymbol{\mu}$:

(a)

$$\begin{aligned} \text{LRT}(\mathbf{X}, \boldsymbol{\Sigma}_0) &= n \left[p \ln \frac{\widehat{\nu}_0}{\widehat{\nu}} + 2 \ln \frac{\Gamma(\frac{\widehat{\nu}+p}{2}) \Gamma(\frac{\widehat{\nu}_0}{2})}{\Gamma(\frac{\widehat{\nu}_0+p}{2}) \Gamma(\frac{\widehat{\nu}}{2})} + \ln \frac{|\boldsymbol{\Sigma}_0|}{|\widehat{\boldsymbol{\Sigma}}|} \right] \\ &\quad + (\widehat{\nu}_0 + p) \sum_{i=1}^n \ln \left[1 + \frac{1}{\widehat{\nu}_0} (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_0)' \boldsymbol{\Sigma}_0^{-1} (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_0) \right] \\ &\quad - (\widehat{\nu} + p) \sum_{i=1}^n \ln \left[1 + \frac{1}{\widehat{\nu}} (\mathbf{x}_i - \widehat{\boldsymbol{\mu}})' \widehat{\boldsymbol{\Sigma}}^{-1} (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}) \right], \end{aligned}$$

(b)

$$\text{RST}(\mathbf{X}, \boldsymbol{\Sigma}_0) = \frac{n(\widehat{\nu}_0+p+2)}{2\widehat{\nu}_0(\widehat{\nu}_0+p)} \left\{ \widehat{\nu}_0 \cdot \text{tr}[(\mathbf{V}\boldsymbol{\Sigma}_0^{-1})^2] + \text{tr}^2(\mathbf{V}\boldsymbol{\Sigma}_0^{-1}) \right\}$$

with

$$\mathbf{V} = \mathbf{V}(\mathbf{X}, \boldsymbol{\Sigma}_0) = \boldsymbol{\Sigma}_0 - \frac{\widehat{\nu}_0+p}{n\widehat{\nu}_0} \sum_{i=1}^n \frac{(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_0)(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_0)'}{1 + \frac{1}{\widehat{\nu}_0} (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_0)' \boldsymbol{\Sigma}_0^{-1} (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_0)}, \quad (c)$$

$$\text{WT}(\mathbf{X}, \boldsymbol{\Sigma}_0) = \frac{n}{2(\widehat{\nu}+p+2)} \left\{ (\widehat{\nu} + p) \cdot \text{tr} \left[(\mathbf{I}_p - \boldsymbol{\Sigma}_0 \widehat{\boldsymbol{\Sigma}}^{-1})^2 \right] - \text{tr}^2(\mathbf{I}_p - \boldsymbol{\Sigma}_0 \widehat{\boldsymbol{\Sigma}}^{-1}) \right\}, \quad (d)$$

$$\text{WT}^*(\mathbf{X}, \boldsymbol{\Sigma}_0) = \frac{n}{2(\widehat{\nu}_0+p+2)} \left\{ (\widehat{\nu}_0 + p) \cdot \text{tr} \left[(\mathbf{I}_p - \widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}_0^{-1})^2 \right] - \text{tr}^2(\mathbf{I}_p - \widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}_0^{-1}) \right\},$$

where $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ is a random sample from $t_{p,\nu}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\nu > 2$ being unknown.

Similarly to the previous cases, due to Wilks (1938), when $n \rightarrow \infty$ and H_0 holds, the distributions of $\text{LRT}(\mathbf{X}, \boldsymbol{\Sigma}_0)$, $\text{RST}(\mathbf{X}, \boldsymbol{\Sigma}_0)$ and $\text{WT}(\mathbf{X}, \boldsymbol{\Sigma}_0)$ tend to the chi-square

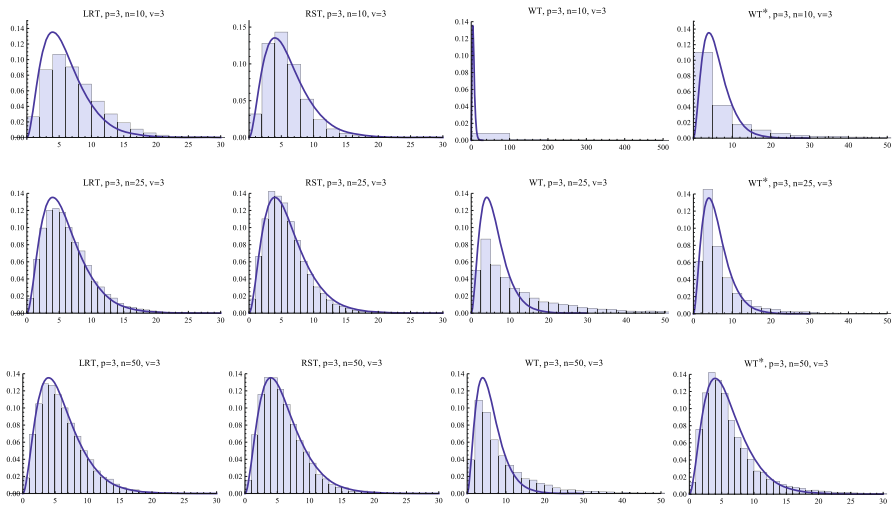


Fig. 18 Empirical null distributions of LRT, RST, WT and WT^* (in columns) for $p = 3$, $n = 10, 25, 50$ (in rows), along with limiting χ_6^2 , under the assumption of unknown ν

distribution with $p(p + 1)/2$ degrees of freedom. The same convergence holds for $WT^*(\mathbf{X}, \Sigma_0)$. The distributions of the above test statistics for data generated from $t_{3,3}(\mathbf{0}, \mathbf{I}_3)$, together with limiting chi-square distribution, are presented in Fig. 18. Comparing to corresponding distributions under known ν , similar behavior of the test statistics can be observed.

Finally note, that all simulations have been performed using *Mathematica*, however, the algorithm for ML estimators of multivariate t distribution with unknown degrees of freedom is also available in *fitHeavyTail* package of R.

8 Discussion and conclusions

In this paper we have determined the LRT, RST, WT and WT^* statistics for testing the covariance structure under the null hypothesis $H_0 : \Sigma = \Sigma_0$ for a multivariate $t_{p,\nu}$ -distribution. Our main interest was focused on the situation when the number of degrees of freedom ν is as small as possible ($\nu > 2$ to guarantee existence of the covariance matrix) to examine possible differences between the ML estimates of the $t_{p,\nu}$ -distribution and the corresponding normal distribution.

In the definition of the density of the $t_{p,\nu}$ -distribution we followed the classical definition where Σ is the scale matrix (cf. Kotz and Nadarajah (2004)). In some papers the density function of the $t_{p,\nu}$ -distribution is defined in a different form, where Σ is the covariance matrix (Sutradhar (1993), Osorio et al. (2023), for example). Note, however, that when testing covariance structures the behavior of the test statistics depends on the scale matrix Σ , and the univariate multiplier in the expression for the covariance matrix in our definition does not influence the convergence properties of the statistics. Note also that for a fixed number of degrees of freedom, there is no need

to estimate ν , and thus the reparameterization $\eta = 1/\nu$ does not influence our results at all, and for unknown ν we followed the EM algorithm given in McLachlan and Krishnan (1997), where direct estimation of degrees of freedom is presented.

From the simulation studies concerning the convergence of test statistics to the limiting distribution, we note that all four test statistics converge to the limiting chi-square distribution; however, RST outperforms all the remaining tests. Moreover, note that the forms of all considered test statistics for a $t_{p,\nu}$ -distribution differ significantly from the corresponding expressions in the case of a multivariate normal distribution. This is caused by the different form of the likelihood function and also the different MLEs of unknown parameters. Thus, one has to be careful with assumptions about the population distribution when testing the structure of the covariance matrix. Nevertheless, apart from the difference between RST and MRST, their empirical distributions are quite similar and close to the theoretical distribution even for small sample sizes. Thus, even if one was mistaken with regard to the density of the observation matrix, MRST can be seen as another test, more conservative than RST. Indeed, MRST can be found to be a special case of Neyman's score test statistic (Neyman 1959; Sutradhar 1993). Taking into account the quick convergence of the empirical type I error of RST to the nominal significance level, we would recommend it to be used by practitioners, pointing out that the power of RST does not seem to be significantly lower than the power of LRT.

In our considerations presented in Sect. 6 we assumed a fixed number of degrees of freedom; however, simulation studies show that increasing the degrees of freedom does not have much influence on the speed of convergence to the limiting chi-square distribution, or on power. This allows us to conclude that our findings remain valid even for unknown ν .

Finally, recall that we assumed a fixed dimension of the data, p , while the sample size tends to infinity. Studying the behavior of the test statistics under a high-dimensional setup, when p and n both tend to infinity, may also be of interest. Note that this problem has been addressed by many authors, usually under normality (see, e.g. Bai and Silverstein 2004; Ledoit and Wolf 2002; Srivastava 2005; Yao et al. 2015), but some references can also be found for non-Gaussian variables (Bai et al. 2009; Jiang 2016, for example). Usually in such cases the central limit theorem is involved and additional assumptions related to the moments of the distributions must be taken into account. Nevertheless, the problem of growing dimension under the multivariate t -distribution will be the topic of future research.

Acknowledgements The authors are grateful to the Referees and Associated Editor for suggestions and comments which have improved the presentation of the material in the paper. We also thank the Mathematical Research and Conference Center of the Polish Academy of Sciences, Będlewo, Poland, for providing the opportunity and support for this paper. The research of K. Filipiak was funded by the Poznan University of Technology under Grant No. 0213/SBAD/0118. T. Kollo is grateful for financial support from Grant PRG1197 of the Estonian Research Council.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If

material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Anderson TW (2003) An introduction to multivariate statistical analysis. Wiley, Hoboken
- Bai ZD, Silverstein JW (2004) CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann Probab* 32:553–605
- Bai Z, Jiang D, Yao J, Zheng S (2009) Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann Stat* 37:3822–3840
- Bilodeau M, Brenner D (1999) Theory of multivariate statistics. Springer, New York
- Filipiak K, Klein D (2017) Estimation of parameters under a generalized growth curve model. *J Multivariate Anal* 158:73–86
- Filipiak K, Klein D, Roy A (2016) Score test for a separable covariance structure with the first component as compound symmetric correlation matrix. *J Multivariate Anal* 150:105–124
- Filipiak K, John M, Klein D (2023) Testing independence under a block compound symmetry covariance structure. *Stat Pap* 64:677–704
- Filipiak K, Klein D, Mokrzycka M (2024) Discrepancy between structured matrices in the power analysis of a separability test. *Comput Stat Data Anal* 192:107907
- Finegold M, Drton M (2011) Robust graphical modeling of gene networks using classical and alternative t -distributions. *Ann Appl Stat* 5:1057–1080
- Galea M, Cademartori D, Curci R, Molina A (2020) Robust inference in the capital asset pricing model using the multivariate t -distribution. *J Risk Financ Manag* 13:123
- Harville DA (1997) Matrix algebra from a statistician's perspective. Springer, New York
- Jiang D (2016) Tests for large-dimensional covariance structure based on Rao's score test. *J Multivariate Anal* 152:28–39
- Kan R, Zhou G (2017) Modeling non-normality using multivariate t : implications for asset pricing. *China Finance Rev Int* 7:2–32
- Kollo T, von Rosen D (2005) Advanced multivariate statistics with matrices. Springer, Dordrecht
- Kollo T, Valge M (2020) Covariance structure tests for t -distribution. In: Holgersson T, Singull M (eds) Recent developments in multivariate and random matrix analysis. Springer, Cham, pp 199–217
- Kollo T, von Rosen D, Valge M (2016) Hypotheses testing on covariance structures: comparison of likelihood ratio test, Rao's score test and Wald's score test. In: Bozeman JR, Oliveira T, Skiadas CH (eds) Stochastic and data analysis methods and applications in statistics and demography. ISAST, Groningen, pp 423–435
- Kotz S, Nadarajah S (2004) Multivariate t distributions and their applications. Cambridge University Press, Cambridge
- Lauprete G, Samarov WR (2002) Robust portfolio optimization. *Metrika* 55:139–149
- Ledoit O, Wolf M (2002) Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann Stat* 30:1081–1102
- Liu C, Rubin DB (1995) ML estimation of the t distribution using EM and its extensions, ECM and ECME. *Stat Sin* 5:19–39
- Liu S, Trenkler G, Kollo T, von Rosen D, Baksalary OM (2023) Professor Heinz Neudecker and matrix differential calculus. *Stat Papers*. <https://doi.org/10.1007/s00362-023-01499-w>
- Magnus J, Neudecker H (1986) Symmetry, 0–1 matrices and Jacobians, a review. *Econom Theory* 2:157–190
- Magnus J, Neudecker H (2019) Matrix differential calculus with applications in statistics and econometrics, 3rd edn. Wiley, Hoboken
- McLachlan GJ, Krishnan T (1997) The EM algorithm and extensions. Wiley, Hoboken
- Mitchell AFS (1989) The information matrix skewness tensor and α -connections for the general multivariate elliptic distribution. *Ann Inst Stat Math* 41:289–304
- Muirhead RJ (2005) Aspects of multivariate statistical theory. Wiley, Hoboken
- Neudecker H (1969) Some theorems on matrix differentiations with special reference to Kronecker matrix products. *J Am Stat Assoc* 64:953–963

- Neyman J (1959) Optimal asymptotic tests of composite statistical hypotheses. In: Grenander V (ed) Probability and statistics. Wiley, Hoboken, pp 13–34
- Osorio F, Galea M, Henriquez C, Arellano-Valle R (2023) Addressing non-normality in multivariate analysis using the t -distribution. *AStA Adv Stat Anal*. <https://doi.org/10.1007/s10182-022-00468-2>
- Rao CR (1948) Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Math Proc Cambridge Philos Soc* 44:50–57
- Rao CR (1973) Linear statistical inference and its applications. Wiley, Hoboken
- Rao CR (2005) Score test: historical review and recent developments. In: Balakrishnan N, Nagaraja HN, Kannan N (eds) Advances in ranking and selection, multiple comparisons, and reliability. Springer, Berlin, pp 3–20
- Srivastava MS (2005) Some tests concerning the covariance matrix in high dimensional data. *J Jpn Stat Soc* 35:251–272
- Srivastava M, von Rosen T, von Rosen D (2008) Models with a Kronecker product covariance structure: estimation and testing. *Math Methods Stat* 17:357–370
- Srivastava M, von Rosen T, von Rosen D (2009) Estimation and testing in general multivariate linear models with Kronecker product covariance structure. *Sankhyā Ser A* 71:137–163
- Stein C (1956) Some problems in multivariate analysis. Technical Report, Department of Statistics, Stanford University, Part I
- Sutradhar B (1993) Score tests for the covariance matrix of the elliptical t -distribution. *J Multivariate Anal* 46:1–12
- Wald A (1943) Tests of statistical hypotheses concerning several parameters, when the number of observations is large. *Trans Am Math Soc* 54:426–482
- Wilks SS (1938) The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann Math Stat* 9(1):60–62
- Yao J, Zheng S, Bai Z (2015) Large sample covariance matrices and high-dimensional data analysis. Cambridge University Press, Cambridge

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.