

Testing the hypothesis of a block compound symmetric covariance matrix for elliptically contoured distributions

Carlos A. Coelho¹  · Anuradha Roy²

Received: 12 August 2015 / Accepted: 11 November 2016 / Published online: 1 December 2016
© Sociedad de Estadística e Investigación Operativa 2016

Abstract In this paper, the authors study the problem of testing the hypothesis of a block compound symmetry covariance matrix with two-level multivariate observations, taken for m variables over u sites or time points. Through the use of a suitable block-diagonalization of the hypothesis matrix, it is possible to obtain a decomposition of the main hypothesis into two sub-hypotheses. Using this decomposition, it is then possible to obtain the likelihood ratio test statistic as well as its exact moments in a much simpler way. The exact distribution of the likelihood ratio test statistic is then analyzed. Because this distribution is quite elaborate, yielding a non-manageable distribution function, a manageable but very precise near-exact distribution is developed. Numerical studies conducted to evaluate the closeness between this near-exact distribution and the exact distribution show the very good performance of this approximation even for very small sample sizes and the approach followed allows us to extend its validity to situations where the population distributions are elliptically contoured. A real-data example is presented and a simulation study is also conducted.

This research was partially supported by FCT–Fundação para a Ciência e Tecnologia (Portuguese Foundation for Science and Technology), Project UID/MAT/00297/2013, through Centro de Matemática e Aplicações (CMA/FCT/UNL). A. Roy also thanks the support for the summer research grant from the College of Business at the University of Texas at San Antonio.

✉ Carlos A. Coelho
cmac@fct.unl.pt

¹ Departamento de Matemática and Centro de Matemática e Aplicações, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal

² Department of Management Science and Statistics, The University of Texas at San Antonio, San Antonio, TX, USA

Keywords Beta random variables · Characteristic function · Composition of hypotheses · Likelihood ratio statistic · Near-exact distributions · Product distribution

Mathematics Subject Classification 62H15 · 62H10 · 62E15 · 62E20 · 62E10 · 60E10

1 Introduction

The $u \times u$ covariance matrix Θ has a compound symmetric structure, with diagonal elements σ_0 and off-diagonal elements σ_1 , if it is a positive-definite matrix that can be written as

$$\Theta = \begin{bmatrix} \sigma_0 & \sigma_1 & \sigma_1 & \dots & \sigma_1 \\ \sigma_1 & \sigma_0 & \sigma_1 & \dots & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_0 & \dots & \sigma_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_1 & \sigma_1 & \sigma_1 & \dots & \sigma_0 \end{bmatrix} = \sigma_0 \mathbf{I}_u + \sigma_1 (\mathbf{J}_u - \mathbf{I}_u) = (\sigma_0 - \sigma_1) \mathbf{I}_u + \sigma_1 \mathbf{J}_u \quad (1)$$

where $-\sigma_0/(u-1) < \sigma_1 < \sigma_0$ and $\sigma_0 > 0$, and where \mathbf{I}_u denotes the identity matrix of order u and \mathbf{J}_u denotes a $u \times u$ matrix of 1's.

Compound symmetry (CS) is a widely used or assumed covariance structure (Timm 2002, Sect. 3.8). As a result of its wide application in many different statistical models, it is also known under a few other designations. Right from (1), the CS structure is also called *equivariance-equivariance* or *equivariance-euicorrelation* (Vonesh and Chinchilli 1997, Sect. 3.2) and sometimes just referred to as *euicorrelation* structure, which may be a somewhat misleading designation since euicorrelation may occur without CS occurring, since we may have euicorrelation without having equal variances. Morrison (1976) addresses, mainly in Chapter 8, a number of examples based on real data for which euicorrelation may be or may seem to be a plausible model for covariances, or at least one that one might be interested in testing. However, for some of these examples, CS may be not a plausible model given that the variances are not equal. As such, the alternative designation of *exchangeable* covariance or *exchangeable* correlation for the CS structure is indeed more adequate (Demidenko 2004, Sect. 2.4, 7.10).

Verbeke and Molenberghs (2000) do a quite thorough assessment of the application and usefulness of the CS as a covariance structure in linear mixed models and repeated measures or longitudinal data models. In Chapter 1 they stress the fact that “the marginal model corresponding to a random-intercept model” is a model with CS covariance structure [see also Demidenko (2004, Sects. 2.4, 7.2)], and in Sect. 3.3 they show how the general linear mixed-effects model is indeed a model with a CS covariance structure, where

$$\rho = \frac{\sigma_1}{\sigma_0} \quad (2)$$

is commonly called the intraclass correlation—see also Timm (2002, Sect. 3.9.d) and Kutner et al. (2005, Sect. 25.5). The CS covariance structure is thus also sometimes

called the intraclass correlation structure, since Θ in (1) may be written as

$$\Theta = \sigma_0 \{(1 - \rho)\mathbf{I}_u + \rho\mathbf{J}_u\} = \sigma_0 \{\mathbf{I}_u + \rho(\mathbf{J}_u - \mathbf{I}_u)\}.$$

Other names given to ρ in (2) are familial, intrablock or intracluster correlation (Rao 1945, 1953; King and Evans 1986).

Vonesh and Chinchilli (1997, Sect. 3.2) establish the equivalence between the univariate random effects model and a multivariate (MANOVA) model with a CS covariance matrix and state in Section 7.3 that CS “arises naturally from split-plot type designs”.

Timm (2002, Sect. 3.9.d) also brings to our attention that the CS structure for the covariance matrix in a univariate mixed ANOVA model is also a sufficient condition for the existence of an exact F test for testing the null hypothesis of equality of the treatment effects in this model. This is so because the CS structure is indeed a particular case of the type H matrices introduced by Huynh and Feldt (1970) who proved that this type H structure is the necessary covariance structure for the existence of such exact F tests in univariate repeated measures designs.

Two other interesting models where the CS covariance structure is used are brought to us by Matos et al. (2016) and Zimmerman and Núñez-Antón (2001). Matos et al. (2016) use a CS covariance structure to model censored data collected irregularly over time with mixed-effects models. These authors also bring to our attention that the CS structure is a particular case of the damping exponential correlation structure proposed by Muñoz et al. (1992). Zimmerman and Núñez-Antón (2001) use CS as a plausible covariance structure in “models for unbalanced data having some kind of dependence structure, all within the context of having a real continuous response variable and real explanatory variables”.

Qu and Li (2006) also use the CS structure as a model for the so-called “working correlation”, introduced by Liang and Zeger (1986) for longitudinal data models and the importance of testing this structure is also brought to us by these authors (Qu and Li 2006, pg. 381), when they state that “If the working correlation R is misspecified, the estimator of the regression parameter is still consistent, but is not efficient within the same class of estimating functions”. Li and Wong (2010) also refer, in the realm of longitudinal data models, that CS is one of the most commonly used covariance structures “to model the correlations among the repeated observations from the same subject”, given its simple form and good interpretability, and they suggest that the likelihood ratio testing procedure would be an adequate testing procedure for this covariance structure, even in the domain of semi-parametric models.

CS is a very parsimonious covariance structure which, as seen from (1), describes the whole covariance structure with only two parameters. The assumption of this structure may improve estimation and the power of tests as stated by Vonesh and Chinchilli (1997, Sect. 3.2, 7.3), besides allowing for the estimation to be adequately done with smaller sample sizes, given the fact that the covariance matrix is being modeled by a smaller number of parameters. Even in nonlinear models, as Malott (1990) states, “by incorporating the compound symmetric structure into the model, substantial improvements in the estimation of the covariance matrix for the parameter estimates are obtained”.

King and Evans (1986) bring up the importance of testing for CS covariance structures when they cite Scott and Holt (1982) as having proved that ignoring such correlation structures may lead “to seriously misleading confidence intervals and hypothesis tests based on inefficient ordinary least squares estimates”. A similar issue is also brought up by Vonesh and Chinchilli (1997, Sect. 7.3) who use the CS structure for linear and also nonlinear models, when these authors say that “ignoring compound symmetry in favor of a general covariance structure leads to significantly inflated Type I errors while correctly assuming compound symmetry leads to improved Type I errors”.

While all these strengthen the need for adequate testing procedures for the CS covariance structure, also in all these cases, the assumption of the CS structure always goes along with the assumption of normality—see also Jones (1993, Sect. 1.5). The likelihood ratio test (LRT) for the CS structure in (1), under the normality assumption, was developed by Wilks (1946).

In the present paper, the authors will address the multivariate or block CS (BCS) structure, where a set of m variables is measured at u time points, and where Θ may be written as

$$\Theta = \begin{bmatrix} \Sigma_0 & \Sigma_1 & \dots & \Sigma_1 \\ \Sigma_1 & \Sigma_0 & \dots & \Sigma_1 \\ \vdots & & \ddots & \vdots \\ \Sigma_1 & \Sigma_1 & \dots & \Sigma_0 \end{bmatrix} = \mathbf{I}_u \otimes (\Sigma_0 - \Sigma_1) + \mathbf{J}_u \otimes \Sigma_1, \tag{3}$$

where Σ_0 is a positive-definite symmetric $m \times m$ matrix, and Σ_1 is a symmetric $m \times m$ matrix, subject to the constraints $-\frac{1}{u-1}\Sigma_0 < \Sigma_1$ and $\Sigma_1 < \Sigma_0$, which mean that $\Sigma_0 - \Sigma_1$ and $\Sigma_0 + (u - 1)\Sigma_1$ are positive-definite matrices, so that the $mu \times mu$ matrix Θ is also positive-definite (for a proof, see Lemma 2.1 by Roy and Leiva (2011)). A BCS structure as the one in (3) arises whenever m response variables are measured and modeled at any given site or time point and we would use for each single response variable a CS covariance matrix.

The $m \times m$ diagonal blocks Σ_0 in Θ represent the variance–covariance matrix of the m response variables at any given site or time point, whereas the $m \times m$ off diagonal blocks Σ_1 in Θ represent the covariance matrix of the m response variables between any two different sites or time points. Σ_0 is assumed to be constant for all sites or time points, and Σ_1 is also assumed to be the same for any two different sites or time points.

If Y_{tj} ($t = 1, \dots, u; j = 1, \dots, m$) denotes the j -th variable measured on site or time t , once the BCS structure is assumed, we will have

$$\text{Var}(Y_{tj}) = \text{Var}(Y_{sj}) \quad \text{and} \quad \text{Cov}(Y_{tj}, Y_{tk}) = \text{Cov}(Y_{sj}, Y_{sk})$$

for all $t, s \in \{1, \dots, u\}$ and $j, k \in \{1, \dots, m\}$, that is,

$$\text{Var}(\underline{Y}_t) = \text{Var}(\underline{Y}_s) = \Sigma_0 \quad \text{for all } t, s \in \{1, \dots, u\}$$

where $\text{Var}(\underline{Y}_t)$ denotes the covariance matrix for the subvector $\underline{Y}_t = [Y_{t1}, \dots, Y_{tm}]'$ ($t = 1, \dots, u$), and also

$$\text{Cov}(Y_{tj}, Y_{sk}) = \text{Cov}(Y_{tk}, Y_{sj}) \quad \text{for any } t, s \in \{1, \dots, u\} \quad \text{and } j, k \in \{1, \dots, m\},$$

or, equivalently,

$$\text{Cov}(\underline{Y}_t, \underline{Y}_s) = \Sigma_1, \quad \text{for any } t, s \in \{1, \dots, u\}, \quad \text{with } t \neq s,$$

where Σ_1 is a symmetric matrix.

Examples of multivariate models with this covariance structure are the multivariate repeated measurement or growth curve models used by [Reinsel \(1982\)](#) and the models used by [Arnold \(1979\)](#), [Timm \(1980\)](#) and [Timm \(2002, Sect. 6.5\)](#), [Roy \(2006\)](#) and [Roy and Fonseca \(2012\)](#).

As such, the need to test for BCS structure arises in many situations, namely those in which it is assumed as a structure for the covariance matrices involved in further analyses such as in many biomedical and medical researches. Indeed, one has to be very careful when assuming this structure for two-level multivariate data, since an incorrect assumption may result in wrong conclusions. Thus, testing the validity of this BCS structure is of vital importance before assuming it, for any statistical analysis and a few authors have marginally addressed this topic. [Timm \(2002, Sect. 6.5; 6.6\)](#), following [Krishnaiah and Lee \(1974, 1980\)](#), takes BCS as a particular case of the so-called linear structure, where Θ can be written as

$$\Theta = \sum_{i=1}^k \mathbf{G}_i \otimes \Sigma_i$$

where $\mathbf{G}_1, \dots, \mathbf{G}_k$ are known $u \times u$ matrices which commute, and $\Sigma_1, \dots, \Sigma_k$ are unknown $m \times m$ matrices. Then, he follows the testing procedure outlined by [Krishnaiah and Lee \(1974, 1980\)](#) and ends up recommending a chi-square approximation for the distribution of the LRT statistic. However, although this is a valid result in terms of convergence in distribution, it is indeed of no practical use, mainly when the sample sizes are not huge. As shown for example by [Coelho et al. \(2016\)](#) the chi-square approximation only works for quite large sample sizes when the overall number of variables involved is rather small. Since in the present situation, although the BCS covariance structure is a quite parsimonious one in terms of the number of parameters used to model the whole covariance matrix, we will anyway be dealing with quite large numbers of variables, the chi-square approximation would only work for extremely large sample sizes, and even in these cases would give a much worse approximation than the one that is obtained in the present paper. [Krishnaiah and Lee \(1974, 1980\)](#) address the test for BCS structure in general terms, encompassed in a general testing scheme for the linear structure and recommend the use of [Box \(1949\)](#) approximation for the distribution of the LRT statistic, anyway without addressing specifically the test for BCS structure. But also [Coelho and Marques \(2012\)](#) show how in situations

where the number of variables is moderately large to large, the asymptotic distributions obtained using Box's approximation may give quantiles and p -values which may fall quite far from the exact ones, since in these situations such asymptotic distributions commonly are not even real distributions, since both the p.d.f.'s and c.d.f.'s may assume values below zero.

Thus, our goal is to develop an approach which is not only able to allow for an easy way to obtain the LRT statistic to test BCS and the full characterization of its exact distribution, but which is furthermore able to allow for an easy way to obtain very sharp, but very manageable, near-exact approximations for the distribution of the LRT statistic. All this in order to make this test easy to implement in practice, since its practical application has been hindered by the complexity of the exact distribution of its LRT statistic. Moreover, the approach followed, allowing to express the overall LRT statistic as the product of the LRT statistic to test independence of groups of variables by the LRT statistic to test equality of covariance matrices, also allows for the immediate extension of the results obtained to populations with elliptically contoured distributions. In Chapters 8–10 of Anderson (2003), it is shown that, under the corresponding null hypotheses, the distributions of these two LRT statistics remain the same either for normally distributed or elliptically contoured distributions. As such, although the distribution obtained for the BCS LRT statistic is derived under the multivariate normality assumption, based on these results, both the exact as well as the near-exact distributions obtained remain valid for elliptically contoured distributions, thus widening much the scope of the results obtained.

Further sections in this paper are: Sect. 2, where the null hypothesis is formulated in two equivalent ways, the second of which will open the way for an easy means to obtain the LRT statistic to test BCS and also for two equivalent ways to characterize its exact distribution, the second of which will then lead the way to Sect. 3 where sharp near-exact distributions are obtained for this statistic; then in Sect. 4, some numerical studies are carried out to show how sharp the near-exact distributions developed are, even for very small sample sizes and also for large numbers of variables involved, situations in which the chi-square approximation is shown to not perform well. In Sect. 5, a simple real-data example is used to illustrate how the near-exact approximations developed may be used and a simulation study is carried out to show that if one thinks that by simulating the Beta random variables involved in the exact distribution of the LRT statistic quite sharp p -values and quantiles may be obtained, this will not be the case, even for quite long simulations, given the quite large number and variety of parameters of those Beta random variables. Finally in Sect. 6, some final conclusions are drawn.

2 Formulation of the hypothesis and the likelihood ratio test

Let us assume that $\underline{Y} = [\underline{Y}'_1, \dots, \underline{Y}'_u] \sim N(\underline{\mu}, \underline{\Sigma})$ and that we are interested in testing the hypothesis

$$H_0: \underline{\Sigma} = \underline{\Theta}, \quad (4)$$

where $\underline{\Theta}$ is defined in (3), versus the alternative hypothesis that $\underline{\Sigma}$ is only positive-definite.

In Lemma 3.1 by Roy and Fonseca (2012), it is shown that we may write

$$\Gamma \Theta \Gamma' = \begin{bmatrix} \Delta_2 & \mathbf{0} \\ \mathbf{0} & I_{u-1} \otimes \Delta_1 \end{bmatrix},$$

where

$$\begin{aligned} \Delta_1 &= \Sigma_0 - \Sigma_1, \\ \Delta_2 &= \Sigma_0 + (u - 1) \Sigma_1, \end{aligned}$$

and $\Gamma = C^{*'} \otimes I_m$, with C^* an orthogonal Helmert matrix whose first column is proportional to a vector of 1's.

Since Γ is not a function of either Σ_0 , or Σ_1 , to test H_0 in (4) is equivalent to test

$$H_0: \Sigma^* = \Omega \tag{5}$$

where

$$\Sigma^* = \Gamma \Sigma \Gamma' \quad \text{and} \quad \Omega = \Gamma \Theta \Gamma'.$$

The null hypothesis in (5) may be split as

$$H_0 \equiv H_{0b|a} \circ H_{0a},$$

where ‘ \circ ’ means ‘after’, and where

$$H_{0a}: \Sigma^* = \text{block-diag}(\Sigma_i^*, i = 1, \dots, u), \tag{6}$$

is the hypothesis of independence of the u diagonal blocks of size $m \times m$ of Σ^* , and

$$\begin{aligned} H_{0b|a}: \Sigma_2^* = \dots = \Sigma_u^*, \\ \text{assuming } H_{0a} \end{aligned} \tag{7}$$

is the null hypothesis corresponding to the test of equality of the $u - 1$ covariance matrices $\Sigma_2^*, \dots, \Sigma_u^*$, assuming H_{0a} .

The LRT statistic to test H_{0a} in (6) is, for a sample of size n , given by Anderson (2003, Sect. 9.2) as

$$\Lambda_a = \left(\frac{|A|}{\prod_{j=1}^u |A_j|} \right)^{n/2},$$

where $A = \Gamma \hat{\Sigma} \Gamma'$ is the maximum likelihood estimator of Σ^* , and A_j its $m \times m$ j -th diagonal block ($\hat{\Sigma}$ being the maximum likelihood estimator of Σ).

The LRT statistic to test $H_{0b|a}$ in (7) is (Anderson 2003, Section 10.2)

$$\Lambda_b = \left((u - 1)^{m(u-1)} \frac{\prod_{j=2}^u |A_j|}{|A^*|^{u-1}} \right)^{n/2}, \tag{8}$$

where

$$A^* = \sum_{j=2}^u A_j.$$

Then, the LRT statistic to test H_0 in (5) will be

$$\Lambda = \Lambda_a \Lambda_b = \left((u - 1)^{m(u-1)} \frac{|A|}{|A_1| |A^*|^{u-1}} \right)^{n/2}, \tag{9}$$

with the h -th moment of Λ , under H_0 in (4) or (5), given by

$$E \left(\Lambda^h \right) = E \left(\Lambda_a^h \right) E \left(\Lambda_b^h \right), \tag{10}$$

since, under H_{0a} , Λ_a is independent of A_1, \dots, A_u (Marques and Coelho 2012; Coelho and Marques 2013), which makes Λ_a independent of Λ_b , given that this latter one is only function of A_2, \dots, A_u . Since the range of Λ is delimited, from this expression for the h -th moment of Λ under H_0 in (4) or (5), we will then be able to obtain the characterization of the distribution of Λ under this null hypothesis, the second version of which, obtained at the end of this section, will then enable us to obtain in the next section very sharp near-exact distributions for Λ .

In (10) we have, under H_{0a} in (6), (Marques et al. 2011)

$$\begin{aligned} E \left(\Lambda_a^h \right) &= \prod_{k=1}^{u-1} \prod_{j=1}^m \frac{\Gamma \left(\frac{n-j}{2} \right) \Gamma \left(\frac{n-(uv-k)m-j}{2} + \frac{n}{2} h \right)}{\Gamma \left(\frac{n-(uv-k)m-j}{2} \right) \Gamma \left(\frac{n-j}{2} + \frac{n}{2} h \right)} \\ &= \underbrace{\left\{ \prod_{j=3}^{mu} \binom{n-j}{n}^{r_j} \left(\frac{n-j}{n} + h \right)^{-r_j} \right\}}_{\Phi_{a,1}(h)} \underbrace{\left\{ \frac{\Gamma \left(\frac{n-1}{2} \right) \Gamma \left(\frac{n-2}{2} + \frac{n}{2} h \right)}{\Gamma \left(\frac{n-1}{2} + \frac{n}{2} h \right) \Gamma \left(\frac{n-2}{2} \right)} \right\}^{k^*}}_{\Phi_{a,2}(h)} \end{aligned} \tag{11}$$

where

$$k^* = \begin{cases} \lfloor u/2 \rfloor, & m \text{ odd} \\ 0, & m \text{ even,} \end{cases}$$

and

$$r_j = \begin{cases} h_{j-2} + (-1)^j k^*, & j = 3, 4 \\ r_{j-2} + h_{j-2}, & j = 5, \dots, mu \end{cases} \tag{12}$$

with

$$h_j = \begin{cases} uv - 1, & j = 1, \dots, m \\ -1, & j = m + 1, \dots, mu - 2, \end{cases} \tag{13}$$

while for Λ_b we have, under $H_{0b|a}$ in (7),

$$\begin{aligned} E\left(\Lambda_b^h\right) &= \prod_{j=1}^m \prod_{k=1}^{u-1} \frac{\Gamma\left(\frac{n-1}{2} - \frac{j-1}{2(u-1)} + \frac{k-1}{u-1}\right) \Gamma\left(\frac{n-j}{2} + \frac{n}{2}h\right)}{\Gamma\left(\frac{n-1}{2} - \frac{j-1}{2(u-1)} + \frac{k-1}{u-1} + \frac{n}{2}h\right) \Gamma\left(\frac{n-j}{2}\right)} \\ &= \underbrace{\left\{ \prod_{j=2}^m \left(\frac{n-j}{n}\right)^{s_j} \left(\frac{n-j}{n} + h\right)^{-s_j} \right\}}_{\Phi_{b,1}(h)} \\ &\quad \times \left\{ \prod_{j=1}^{\lfloor m/2 \rfloor} \prod_{k=1}^{u-1} \frac{\Gamma\left(n-1 + \frac{k-2j}{u-1}\right) \Gamma\left(n + \left\lfloor \frac{k-2j}{u-1} - 1 \right\rfloor + nh\right)}{\Gamma\left(n-1 + \frac{k-2j}{u-1} + nh\right) \Gamma\left(n + \left\lfloor \frac{k-2j}{u-1} - 1 \right\rfloor\right)} \right\} \\ &\quad \times \left\{ \prod_{k=1}^{u-1} \frac{\Gamma\left(\frac{n-m}{2} + \frac{m(u-1)-u-m+2k}{2(u-1)}\right)}{\Gamma\left(\frac{n-m}{2} + \left\lfloor \frac{m(u-1)-u-m+2k}{2(u-1)} \right\rfloor\right)} \right. \\ &\quad \left. \times \frac{\Gamma\left(\frac{n-m}{2} + \left\lfloor \frac{m(u-1)-u-m+2k}{2(u-1)} \right\rfloor + \frac{n}{2}h\right)}{\Gamma\left(\frac{n-m}{2} + \frac{m(u-1)-u-m+2k}{2(u-1)} + \frac{n}{2}h\right)} \right\}^{m \perp 2} \tag{14} \\ &\quad \underbrace{\hspace{10em}}_{\Phi_{b,2}(h)} \end{aligned}$$

where s_j ($j = 2, \dots, m$) are given in Appendix 1 and where $m \perp 2$ is the remainder of the integer division of m by 2.

Since the supports of Λ_a and Λ_b are delimited, their distributions are defined by their moments and, as such, from the first expression in (11) we may write, under H_{0a} ,

$$\Lambda_a \sim \prod_{k=1}^{u-1} \prod_{j=1}^m (X_{jk})^{\frac{n}{2}}, \quad \text{with } X_{jk} \sim \text{Beta}\left(\frac{n - (u-k)m - j}{2}, \frac{(u-k)m}{2}\right), \tag{15}$$

where X_{jk} ($j = 1, \dots, m; k = 1, \dots, u - 1$) are independent, while from the first expression in (14), under $H_{0b|a}$,

$$\Lambda_b \sim \prod_{j=1}^m \prod_{k=1}^{u-1} (X_{jk}^*)^{\frac{n}{2}}, \quad \text{with } X_{jk}^* \sim \text{Beta}\left(\frac{n-j}{2}, \frac{2k + (u-2)j - u}{2}\right), \tag{16}$$

where X_{jk}^* ($j = 1, \dots, m; k = 1, \dots, u - 1$) are independent, so that, under H_0 in (4) or (5),

$$\Lambda \sim \prod_{j=1}^m \left\{ \left(\prod_{k=1}^{u-1} X_{jk} \right) \left(\prod_{k=1}^{u-1} X_{jk}^* \right) \right\}, \tag{17}$$

where all random variables are independent.

On the other hand, based on the results in Appendix 2 and from the second expressions in (11) and (14) we may write, for Λ_a ,

$$\Lambda_a \sim \left(\prod_{j=3}^{mu} e^{-Z_j} \right) \left(\prod_{j=1}^{k^*} (W_j)^{\frac{n}{2}} \right) \tag{18}$$

where

$$Z_j \sim \Gamma \left(r_j, \frac{n-j}{n} \right) \text{ and } W_j \sim \text{Beta} \left(\frac{n-2}{2}, \frac{1}{2} \right)$$

are all independent r.v.'s (random variables), while for Λ_b it is possible to write

$$\Lambda_b \sim \left(\prod_{j=2}^m e^{-Z_j^*} \right) \left(\prod_{j=1}^{\lfloor m/2 \rfloor} \prod_{k=1}^{u-1} (W_{1jk}^*)^n \right) \left(\prod_{k=1}^{u-1} (W_{2k}^*)^{\frac{n}{2}} \right)^{m \cdot \lfloor \lfloor 2 \rfloor} \tag{19}$$

where

$$Z_j^* \sim \Gamma \left(s_j, \frac{n-j}{n} \right), \quad W_{1jk}^* \sim \text{Beta} \left(n + \left\lfloor \frac{k-2j}{u-1} - 1 \right\rfloor, \frac{k-2j}{u-1} - \left\lfloor \frac{k-2j}{u-1} \right\rfloor \right),$$

and

$$W_{2k}^* \sim \text{Beta} \left(\frac{n-m}{2} + \left\lfloor \frac{m(u-1) - u - m + 2k}{2(u-1)} \right\rfloor, \frac{m(u-1) - u - m + 2k}{2(u-1)} - \left\lfloor \frac{m(u-1) - u - m + 2k}{2(u-1)} \right\rfloor \right)$$

are all independent r.v.'s.

From (18) and (19), one may thus write, under H_0 in (4) or (5),

$$\Lambda \sim \left(\prod_{j=2}^{mu} e^{-T_j} \right) \left(\prod_{j=1}^{k^*} (W_j)^{\frac{n}{2}} \right) \left(\prod_{j=1}^{\lfloor m/2 \rfloor} \prod_{k=1}^{u-1} (W_{1jk}^*)^n \right) \left(\prod_{k=1}^{u-1} (W_{2k}^*)^{\frac{n}{2}} \right)^{m \cdot \lfloor \lfloor 2 \rfloor} \tag{20}$$

where

$$T_j \sim \Gamma \left(\gamma_j, \frac{n-j}{n} \right), \quad (j = 2, \dots, mu)$$

with

$$\gamma_j = \sum_{j=2}^{mu} (r_j^+ + s_j^+) \quad (21)$$

where

$$r_j^+ = \begin{cases} 0 & j = 2 \\ r_j & j = 3, \dots, mu \end{cases} \quad \text{and} \quad s_j^+ = \begin{cases} s_j & j = 2, \dots, m \\ 0 & j = m + 1, \dots, mu \end{cases} \quad (22)$$

where r_j are given by (12) and (13), s_j are given by (28)–(32) in Appendix 1, and all the other variables are defined as above.

The form of the distribution of Λ in (20), although it may look more complicated than the one in (17), is more useful for the development of the near-exact distributions, as it will be shown in the next section.

It should also be brought to the attention of the reader that, given the results stated at the end of Chapters 8–10 of Anderson (2003), the form of the distribution of Λ in (20) remains valid in case we consider for \underline{Y} any elliptically contoured distribution.

3 The characteristic function of $W = -\log \Lambda$ and the near-exact approximation

From the developments in the previous section and the expression for $E(\Lambda^h)$, the characteristic function (c.f.) of $W = -\log \Lambda$ may be written as

$$\begin{aligned} \Phi_W(t) &= E(e^{itW}) = E(\Lambda^{-it}) \\ &= \underbrace{\left\{ \prod_{j=2}^{mu} \left(\frac{n-j}{n} \right)^{\gamma_j} \left(\frac{n-j}{n} - it \right)^{-\gamma_j} \right\}}_{\Phi_{W,1}(t)} \underbrace{\Phi_{a,2}(-it)\Phi_{b,2}(-it)}_{\Phi_{W,2}(t)} \quad (23) \end{aligned}$$

where γ_j is given by (21) and $\Phi_{a,2}(\cdot)$ and $\Phi_{b,2}(\cdot)$ are defined in (11) and (14), and $\Phi_{W,1}(t)$ is actually equal to $\Phi_{a,1}(-it)\Phi_{b,1}(-it)$, being these two functions also defined in (11) and (14).

Then, in building the near-exact distributions, $\Phi_{W,1}(t)$ will be kept untouched while $\Phi_{W,2}(t)$ will be asymptotically approximated by the c.f. of a finite mixture of Gamma distributions.

While $\Phi_{W,1}(t)$ is the c.f. of a GIG (Generalized Integer Gamma) distribution (Coelho 1998) of depth $mu - 1$, which is the distribution of the sum of mu independent Gamma distributed random variables, all with integer shape parameters, $\Phi_{W,2}(t)$ is the c.f. of the sum of $k^* + \lfloor m/2 \rfloor(u - 1) + (m \bmod 2)$ independent Logbeta distributed random variables. For $u = 2$ and even m , $\Phi_{W,1}(t)$ yields indeed the exact c.f. for W , which means that in this case we have the exact p.d.f. and c.d.f. of W and Λ in a simple closed form. This is, in the form of the p.d.f. and c.d.f. of a GIG distribution of depth $2m$, with shape parameters γ_j given by (21) and rate parameters $(n - j)/n$

($j = 1, \dots, 2m$) for W , or the form of the p.d.f. and c.d.f. of an EGIG (Exponentiated Generalized Integer Gamma) distribution (Arnold et al. 2013) for Λ .

It is based on the results in Sects. 5 and 6 of Tricomi and Erdélyi (1951), which show that the c.f. of a Logbeta(a, b) distribution may be asymptotically approximated by the c.f. of an infinite mixture of $\Gamma(b + j, a)$ ($j = 0, 1, \dots$) distributions that we will replace $\Phi_{W,2}(t)$ by

$$\Phi_2(t) = \sum_{k=0}^{m^*} \pi_k \lambda^{r+k} (\lambda - it)^{-(r+k)}, \tag{24}$$

which is the c.f. of a finite mixture of Gamma distributions, all with the same rate parameter λ . See Appendix 3 for further details on the approximation of $\Phi_{W,2}(t)$ by $\Phi_2(t)$. In (24), λ will be taken to be the rate parameter in

$$\Phi^*(t) = \theta \lambda^{\tau_1} (\lambda - it)^{-\tau_1} + (1 - \theta) \lambda^{\tau_2} (\lambda - it)^{-\tau_2}$$

where θ, λ, τ_1 and τ_2 are determined in such a way that

$$\left. \frac{\partial^h}{\partial t^h} \Phi^*(t) \right|_{t=0} = \left. \frac{\partial^h}{\partial t^h} \Phi_{W,2}(t) \right|_{t=0}, \quad h = 1, \dots, 4,$$

and

$$\begin{aligned} r &= \frac{k^*}{2} + \sum_{j=1}^{\lfloor m/2 \rfloor} \sum_{k=1}^{u-1} \frac{k-2j}{u-1} - \left\lfloor \frac{k-2j}{u-1} \right\rfloor \\ &+ \sum_{k=1}^{u-1} \frac{m(u-1) - u - m + 2k}{2(u-1)} - \left\lfloor \frac{m(u-1) - u - m + 2k}{2(u-1)} \right\rfloor \\ &= \begin{cases} \frac{m}{4}(u-2), & m \text{ even} \\ \lfloor \frac{u}{2} \rfloor + \frac{m+1}{4}(u-2), & m \text{ odd} \end{cases} \quad (u \geq 2), \end{aligned} \tag{25}$$

which is the sum of the second parameters of all the Beta r.v.'s in (20). Then, the weights $\pi_0, \dots, \pi_{m^*-1}$ in (24) will be determined in such a way that

$$\left. \frac{\partial^h}{\partial t^h} \Phi_2(t) \right|_{t=0} = \left. \frac{\partial^h}{\partial t^h} \Phi_{W,2}(t) \right|_{t=0}, \quad h = 1, \dots, m^*,$$

with $\pi_{m^*} = 1 - \sum_{k=0}^{m^*-1} \pi_k$.

The near-exact distributions built in this way will match the first m^* exact moments of W and will have c.f.

$$\Phi_W^*(t) = \Phi_{W,1}(t) \Phi_2(t), \tag{26}$$

which, for non-integer r , is the c.f. of a finite mixture, with weights π_k ($k = 0, \dots, m^*$), of Generalized Near-Integer Gamma (GNIG) distributions of depth mu , with integer

shape parameters γ_j , given by (21) and (22) and non-integer shape parameter r given by (25) and corresponding rate parameters $(n - j)/n$ ($j = 2, \dots, mu$) and λ . See Coelho (2004) and Coelho and Marques (2012, Appendix 1) for the expressions for the p.d.f. and c.d.f. of the GNIG distribution. Using the notation from Appendix 1 in Coelho and Marques (2012), these near-exact distributions will yield for $W = -\log \Lambda$ p.d.f.'s and c.d.f.'s of the form

$$f_W(w) = \sum_{k=0}^{m^*} \pi_k f^{\text{GNIG}} \left(w \mid \gamma_2, \dots, \gamma_{mu}, r + k; \frac{n-2}{n}, \dots, \frac{n-mu}{n}, \lambda; mu \right),$$

$(w > 0)$

and

$$F_W(w) = \sum_{k=0}^{m^*} \pi_k F^{\text{GNIG}} \left(w \mid \gamma_2, \dots, \gamma_{mu}, r + k; \frac{n-2}{n}, \dots, \frac{n-mu}{n}, \lambda; mu \right),$$

$(w > 0),$

while the near-exact p.d.f. and c.d.f. for Λ are, respectively, given by

$$f_\Lambda(z) = \sum_{k=0}^{m^*} \pi_k f^{\text{GNIG}} \left(-\log z \mid r_2^*, \dots, r_p^*, r + k; \frac{n-2}{n}, \dots, \frac{n-mu}{n}, \lambda; mu \right) \frac{1}{z}$$

$(0 < z < 1)$

and

$$F_\Lambda(z) = \sum_{k=0}^{m^*} \pi_k \left(1 - F^{\text{GNIG}} \left(-\log z \mid r_2^*, \dots, r_p^*, r + k; \frac{n-2}{n}, \dots, \frac{n-mu}{n}, \lambda; mu \right) \right),$$

$(0 < z < 1).$

For integer r , the above GNIG distributions of depth mu become GIG distributions of depth mu (Coelho 1998; Arnold et al. 2013, App. B), which have even simpler and more manageable expressions, and in this case the near-exact distributions for Λ will be mixtures of what Arnold et al. (2013) call EGIG distributions.

From these near-exact distributions, one can easily compute near-exact p -values and quantiles, as it is illustrated in Sect. 5, which from the results in Sect. 4 are assured to lie extremely close to the exact ones, even for very small sample sizes and very large numbers of variables involved. As such, even in cases where one may want to compute the power of the LRT for a specific covariance matrix Σ that somehow violates the null hypothesis of BCS, one will preferably (i) use the near-exact quantile for the null distribution of the LRT statistic for the given values of n, m and u , and then simulate something like at least 10^5 or 10^6 pseudo-random samples from a multivariate normal distribution with that covariance matrix Σ , compute the value of the LRT statistic Λ , using (9), and take as the simulated value of power the proportion of cases where the

null hypothesis of BCS is rejected, rather than (ii) use the non-null distribution of Λ , which, given the already rather complicated facies of the null distribution of Λ , would be way too complicated to be computed.

It may be noted that for $m = 1$, this test yields the equivariance–equicorrelation or compound symmetry test in Wilks (1946), while, given the fact that, as stated at the end of Sect. 2, the form in (20) for the exact distribution of Λ remains valid when we assume for the underlying variables an elliptically contoured distribution; also the near-exact distributions developed in this section remain valid in this situation.

4 Numerical studies

To assess the performance of the near-exact distributions developed, that is, their closeness to the corresponding exact distribution, we use the measure

$$\Delta = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left| \frac{\Phi_W(t) - \Phi_W^*(t)}{t} \right| dt, \tag{27}$$

with

$$\max_{w>0} |F_W(w) - F_W^*(w)| = \max_{0<z<1} |F_\Lambda(z) - F_\Lambda^*(z)| \leq \Delta,$$

where $\Phi_W(t)$ is the exact c.f. of W in (23) and $\Phi_W^*(t)$ is the near-exact c.f. of W in (26) and $F_W(\cdot)$ and $F_W^*(\cdot)$ are the corresponding c.d.f.’s, that is, the exact and near-exact c.d.f. of W , being $F_\Lambda(\cdot)$ and $F_\Lambda^*(\cdot)$ the corresponding c.d.f.’s for Λ . That Δ in (27) always yields a finite value is shown in Appendix 4.

Table 1 shows values of the measure Δ for the common chi-square approximation to the distribution of the logarithm of the LRT statistic, which says that $-2 \log \Lambda \overset{a}{\sim} \chi_\nu^2$, with $\nu = mu(mu + 1)/2 - m(m + 1)$, and for the near-exact distributions developed in the previous section. In this table, different values of u (number of locations or time points), m (number of variables) and n (sample size) are used, and also different values of m^* , the number of exact moments of W matched by the near-exact distributions.

Values for Δ in Table 1 were computed using the numerical integration module *NIntegrate* from Mathematica[®], version 9, and using $\Phi_W(t)$ in (23) and $\Phi_W^*(t)$ in (26) for the near-exact distributions, and

$$\Phi_W^*(t) = \left(\frac{1}{2}\right)^{f/2} \left(\frac{1}{2} - i\frac{t}{2}\right)^{-f/2} = (1 - it)^{-f/2}$$

with $f = mu(mu + 1)/2 - m(m + 1)$ for the chi-square approximation for W . Because of numerical stability issues, usually Δ in (27) is computed by integrating between zero and plus infinity and then multiplying it by $1/\pi$. If in some cases the upper limit of plus infinity for the integral may still give some problems, a numerical limit like 3×10^4 or 5×10^4 is used, after checking for stability of the numerical value obtained for the integral.

Table 1 Values of Δ for the chi-square and near-exact distributions, for different values of m and u and sample sizes $n = mu + 2, 30, 100$

n	χ^2	m^*			χ^2	n	m^*			
		4	6	10			4	6	10	
		$u = 2, m = 3$								
8	7.35×10^{-1}	6.61×10^{-9}	1.12×10^{-10}	1.57×10^{-13}	11	9.35×10^{-1}	2.64×10^{-12}	4.88×10^{-15}	1.10×10^{-19}	
36	4.80×10^{-1}	3.86×10^{-12}	9.48×10^{-15}	2.71×10^{-19}	39	2.27×10^{-1}	2.34×10^{-15}	4.50×10^{-19}	1.85×10^{-25}	
106	3.01×10^{-2}	5.04×10^{-15}	1.38×10^{-18}	5.76×10^{-25}	109	7.74×10^{-2}	4.42×10^{-18}	1.05×10^{-22}	6.74×10^{-31}	
		$u = 5, m = 3$								
17	1.15×10^0	5.81×10^{-13}	7.29×10^{-17}	2.86×10^{-24}	23	1.28×10^0	4.31×10^{-14}	1.59×10^{-18}	5.32×10^{-27}	
45	4.67×10^{-1}	2.36×10^{-14}	8.10×10^{-19}	1.73×10^{-27}	51	7.02×10^{-1}	8.03×10^{-15}	1.43×10^{-19}	8.80×10^{-29}	
115	1.77×10^{-1}	2.26×10^{-16}	1.29×10^{-21}	6.84×10^{-32}	121	3.06×10^{-1}	1.14×10^{-16}	5.17×10^{-22}	1.27×10^{-32}	
		$u = 3, m = 5$								
17	1.17×10^0	5.16×10^{-15}	2.63×10^{-18}	5.81×10^{-25}	32	1.45×10^0	1.25×10^{-14}	4.17×10^{-19}	1.15×10^{-27}	
45	4.75×10^{-1}	5.02×10^{-16}	3.47×10^{-20}	5.50×10^{-28}	60	9.82×10^{-1}	3.85×10^{-15}	9.25×10^{-20}	1.14×10^{-28}	
115	1.80×10^{-1}	6.36×10^{-18}	5.70×10^{-23}	2.13×10^{-32}	130	5.49×10^{-1}	5.66×10^{-17}	3.42×10^{-22}	2.62×10^{-32}	
		$u = 5, m = 10$								
52	1.62×10^0	4.38×10^{-18}	6.51×10^{-24}	3.11×10^{-35}	32	1.42×10^0	1.53×10^{-15}	1.36×10^{-20}	1.75×10^{-30}	
80	1.29×10^0	6.57×10^{-18}	1.29×10^{-23}	9.27×10^{-35}	60	9.60×10^{-1}	9.15×10^{-16}	6.29×10^{-21}	1.66×10^{-31}	
150	9.51×10^{-1}	4.02×10^{-19}	3.31×10^{-25}	3.87×10^{-37}	130	5.29×10^{-1}	3.10×10^{-17}	5.63×10^{-23}	4.87×10^{-35}	
		$u = 10, m = 5$								
52	1.63×10^0	9.64×10^{-18}	1.26×10^{-23}	1.85×10^{-35}	102	1.91×10^0	1.13×10^{-21}	2.49×10^{-29}	2.52×10^{-42}	
80	1.28×10^0	1.88×10^{-17}	2.62×10^{-23}	5.87×10^{-35}	130	1.68×10^0	5.81×10^{-21}	2.33×10^{-28}	7.27×10^{-43}	
150	9.42×10^{-1}	1.67×10^{-18}	9.18×10^{-25}	2.87×10^{-37}	200	1.42×10^0	1.99×10^{-21}	5.16×10^{-29}	6.64×10^{-44}	
		$u = 10, m = 10$								
300	1.22×10^0	4.06×10^{-22}	5.54×10^{-30}	1.96×10^{-45}	400	1.08×10^0	1.16×10^{-22}	9.54×10^{-31}	1.23×10^{-46}	

As expected, as m^* increases the values of Δ for the near-exact distributions decrease clearly, showing an increasing closeness to the exact distribution. We may also see from Table 1 that the near-exact distributions developed exhibit a very good performance for very small sample sizes and also a very good asymptotic behavior not only for increasing sample sizes, but also for increasing values of both u and m , which is a much desirable feature. For all values of u and m , the values of Δ , upper bounds on the difference between the exact and the near-exact c.d.f., exhibit extremely low values. One may also note that, for larger values of u and m , the asymptotic behavior for increasing n becomes visible only for larger values of n .

From Table 1, it also becomes clear that indeed the chi-square asymptotic distribution may only yield somewhat sensible approximations for very large sample sizes and small numbers of variables involved, and that the performance of this approximation worsens much as the number of variables increases, that is, as either u or m increases. Indeed, the measure Δ in (27) gives very sharp upper-bounds on the difference between the exact and the approximate c.d.f.'s in case the approximation is rather good, while it may give too large values in the opposite case.

This is the reason why we get some values of Δ above one for the chi-square approximation for the smaller sample sizes for a number of the combinations of larger values of u and m , although indeed the values of Δ should always be between zero and one. This indicates that in these cases the classical chi-square approximation has a really very poor performance.

5 A real-data example and a simulation study

In this section, the authors show how to implement the new hypothesis testing procedure, using the block-diagonalization of the BCS structure, as a result of the application of Lemma 3.1 by Roy and Fonseca (2012), with a real data set taken from Johnson and Wichern (2007, p. 43). A researcher measured the mineral content of bones (radius, humerus and ulna) by photon absorptiometry to examine whether dietary supplements would slow bone loss in 25 older women. Measurements were recorded for the three bones on the dominant and non-dominant sides. As such, data have a two-level multivariate structure, with $u = 2$ and $m = 3$ and if we rearrange the variables in the data set by grouping together the mineral content of the dominant sides of radius, humerus and ulna as a first set of three variables, that is, the variables in the first location ($t = 1$ for the dominant side) and then the mineral contents for the non-dominant side of the same bones as the second set of three variables ($t = 2$ for the non-dominant side), the resulting maximum likelihood estimate of Σ is (rounded to five decimal places)

$$\hat{\Sigma} = \left[\begin{array}{ccc|ccc} 0.01248 & 0.02146 & 0.00876 & 0.00996 & 0.01928 & 0.00764 \\ 0.02146 & 0.07714 & 0.01616 & 0.01779 & 0.06411 & 0.01233 \\ 0.00876 & 0.01616 & 0.01111 & 0.00819 & 0.01703 & 0.00775 \\ \hline 0.00996 & 0.01779 & 0.00819 & 0.01096 & 0.02026 & 0.00855 \\ 0.01928 & 0.06411 & 0.01703 & 0.02026 & 0.06671 & 0.01612 \\ 0.00764 & 0.01233 & 0.00775 & 0.00855 & 0.01612 & 0.01018 \end{array} \right].$$

Not only the sample variance–covariance matrices of the three mineral contents for the dominant and non-dominant sides appear very similar, but also the covariance matrix of the mineral content for the three bones between the dominant and non-dominant sides suggests the possibility of an underlying symmetric population matrix. We may thus hypothesize that the population covariance matrix may have a BCS structure.

To carry out the test, according to the procedure outlined in Section 2, one needs to compute the matrix

$$A = \hat{\Sigma}^* = \Gamma \hat{\Sigma} \Gamma',$$

where

$$\Gamma = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \otimes I_3 = \left[\begin{array}{ccc|ccc} \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \\ \hline \frac{1}{\sqrt{2}} & 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & -\frac{1}{\sqrt{2}} \end{array} \right].$$

Then, from (9), the computed value for Λ is obtained as 0.0227794, for which, using the near-exact distributions developed in Sect. 3, we obtain the p -values in Table 2.

Table 2 gives the p -values for different values of m^* up to the decimal places which exactly match the decimal places of the p -value corresponding to the next m^* . If we just compare the p -values for $m^* = 1$ and $m^* = 2$, we see that the p -value for $m^* = 1$ is exact up to four decimal places. According to the way the near-exact distributions are built, the p -values have better precision for increasing values of m^* , the number of exact moments of W matched by the corresponding near-exact distribution. Thus, the null hypothesis that the covariance structure is of the BCS type should not be rejected, with a p -value = 0.2792, which is much lower than the p -value = 0.5786 obtained when the asymptotic χ^2_ν approximation for $-2 \log \Lambda$ with $\nu = mu(mu + 1)/2 - m(m + 1) = 9$ degrees of freedom is used.

Table 2 p -values from the near-exact approximations for different values of m^* (the number of exact moments matched) for the hypothesis test on bone mineral data

m^*	p -value
1	0.2792
2	0.2792168
4	0.2792168718
6	0.279216871862
10	0.279216871862222

The non-rejection of the BCS structure shows that the population covariance matrices for the mineral content of the three bones (radius, humerus and ulna) for the dominant and the non-dominant sides may be considered to be equal and that also the population covariance matrix for the mineral content between the dominant and the non-dominant sides should be considered to be a symmetric matrix, with

$$Cov(Y_{1j}, Y_{2k}) = Cov(Y_{2j}, Y_{1k}), \quad j, k \in \{1, 2, 3\}.$$

This means that, for example, the population covariance between the mineral content of the dominant side of the radius and the mineral content of the non-dominant side of humerus is the same as that of the mineral content of the non-dominant side of the radius and the mineral content of the dominant side of the humerus, and that this happens for any pair of two different bones.

In Fig. 1, we have, for $W = -\log \Lambda$, the plots of the p.d.f.'s and c.d.f.'s for the near-exact distribution for $m^* = 1$ and for the asymptotic Gamma distribution with shape parameter $9/2$ and rate parameter 1, which corresponds to the chi-square asymptotic distribution with nine degrees of freedom for $-2 \log \Lambda$.

That even p -values obtained from simulation may be not sharp enough was shown by a simulation study, where 100,000 pseudo-random samples with BCS structure for $u = 2, m = 3$ and $n = 25$ were generated. The p -value obtained from this simulation study for the computed value of $\Lambda = 0.0227794$ was 0.28163, which compared with the near-exact p -values in Table 2 shows that p -values obtained from simulation, even when using quite large simulations, may be not that precise.

6 Conclusions

As described in the Introduction of the paper, the Block-Compound Symmetric (BCS) covariance structure may be used as or may arise as the underlying covariance structure for many multivariate models, making the test for a BCS covariance structure a much necessary and desirable one. However, testing for a BCS structure of the covariance matrix may seem at first sight to be a not so easy task, given the facts that not only the likelihood ratio statistic is expected to have a rather complicated derivation but also and mainly because its exact distribution is expected to have an extremely complicated

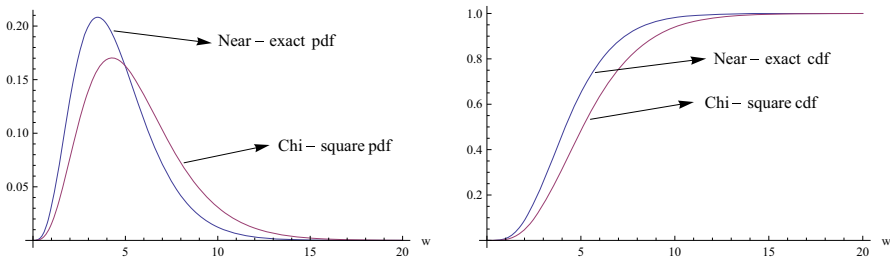


Fig. 1 Plots of the p.d.f.'s and c.d.f.'s of the near-exact distribution (for $m^* = 1$) and the asymptotic $\Gamma(9/2, 1)$ distribution, for $W = -\log \Lambda$ (this later one corresponding to the χ^2_9 asymptotic distribution for $-2 \log \Lambda$)

structure and expression. In this paper, the authors show how to use an adequate decomposition of the BCS null hypothesis; based on Lemma 3.1 by Roy and Fonseca (2012), it is possible to easily derive the expression for the likelihood ratio statistic and also to obtain the expression for its moments, under the BCS hypothesis. The approach followed also enabled the derivation of simple expressions for the p.d.f. and c.d.f. of the likelihood ratio statistic for some simpler particular cases, as well as, and most important, the development of very sharp but highly manageable near-exact distributions for the test statistic, which in turn enable an easy computation of quantiles and p -values. These near-exact distributions exhibit a sharp closeness to the exact distribution for very small samples and also very good asymptotic behaviors not only in terms of increasing sample sizes, but also in terms of increasing values of the number of variables, and number of locations or time points. This asymptotic behavior for increasing number of variables is a much desirable feature which common asymptotic distributions do not have. The authors also show that the common chi-square asymptotic approximation for $-2 \log \Lambda$ may only work in practice for very large sample sizes and when the number of variables involved is quite small, and that it may indeed not even work at all when the number of variables involved is rather large.

The approach followed in this paper may be extended to address more complicated covariance structures arising for multi-level multivariate data, and it also allows for an immediate extension of the results obtained, in terms of both exact and near-exact distributions, to underlying elliptically contoured distributions.

Acknowledgements Both authors would also like to thank the comments and remarks from two anonymous referees and the Editor-in-Chief of the Journal who helped in clarifying a number of small issues and improving the readability of the paper, namely in giving it a more adequate introduction to the usefulness and applicability of the CS and BCS covariance structures.

Appendix 1: Shape parameters in the moment expressions for A_b

According to Coelho and Marques (2012) and Marques et al. (2011), the shape parameters s_j in (14) are given by

$$s_j = \begin{cases} s_{j-1}^* & \text{for } j = 2, \dots, m, \\ s_{j-1}^* + (m - 2)(\alpha_2 - \alpha_1) \left((u - 1) - \frac{m-1}{2} + (u - 1) \left\lfloor \frac{m}{2(u-1)} \right\rfloor \right) & \text{except } j = m - 2\alpha_1 \\ & \text{for } j = m - 2\alpha_1 \end{cases} \tag{28}$$

with

$$s_j^* = \begin{cases} \gamma_j & \text{for } j = 1, \dots, \alpha + 1 \\ (u - 1) \left(\left\lfloor \frac{m}{2} \right\rfloor - \left\lfloor \frac{j}{2} \right\rfloor \right) & \text{for } j = \alpha + 2, \dots, \min(m - 2\alpha_1, m - 1) \\ (u - 1) \left(\left\lfloor \frac{m+1}{2} \right\rfloor - \left\lfloor \frac{j}{2} \right\rfloor \right) & \text{and } j = 2 + m - 2\alpha_1, \dots, 2 \left\lfloor \frac{m}{2} \right\rfloor - 1, \text{ by steps of } 2 \\ & \text{for } j = 1 + m - 2\alpha_1, \dots, m - 1, \text{ by steps of } 2, \end{cases} \tag{29}$$

and

$$\alpha = \left\lfloor \frac{m - 1}{u - 1} \right\rfloor, \quad \alpha_1 = \left\lfloor \frac{u - 2}{u - 1} \frac{m - 1}{2} \right\rfloor, \quad \alpha_2 = \left\lfloor \frac{u - 2}{u - 1} \frac{m + 1}{2} \right\rfloor, \tag{30}$$

where, for $j = 1, \dots, \alpha$,

$$\gamma_j = \left\lfloor \frac{u-1}{2} \right\rfloor \left((j-1)(u-1) - 2(u \perp 2) \left\lfloor \frac{j}{2} \right\rfloor \right) + \left\lfloor \frac{u-1}{2} \right\rfloor \left\lfloor \frac{u-1+j \perp 2}{2} \right\rfloor \tag{31}$$

and

$$\begin{aligned} \gamma_{\alpha+1} = & - \left(\left\lfloor \frac{m}{2} \right\rfloor - \alpha \left\lfloor \frac{u-1}{2} \right\rfloor \right)^2 + (u-1) \left(\left\lfloor \frac{m}{2} \right\rfloor - \left\lfloor \frac{\alpha+1}{2} \right\rfloor \right) \\ & + ((u-1) \perp 2) \left(\alpha \left\lfloor \frac{m}{2} \right\rfloor + \frac{\alpha \perp 2}{4} - \frac{\alpha^2}{4} - \alpha^2 \left\lfloor \frac{u-1}{2} \right\rfloor \right). \end{aligned} \tag{32}$$

Appendix 2: Gamma distribution and related results

We say that the r.v. X follows a Gamma distribution with shape parameter $r > 0$ and rate parameter $\lambda > 0$, if the p.d.f. of X is

$$f_X(x) = \frac{\lambda^r}{\Gamma(r)} e^{-\lambda x} x^{r-1}, \quad (x > 0)$$

and this fact is denoted by $X \sim \Gamma(r, \lambda)$. Then, the moment generating function of X is

$$M_X(t) = \lambda^r (\lambda - t)^{-r} \quad (t < \lambda),$$

so that if we define $Z = e^{-X}$ we have

$$E(Z^h) = E(e^{-hX}) = M_X(-h) = \lambda^r (\lambda + h)^{-r} \quad (h > -\lambda).$$

Appendix 3: The reasoning behind the use of $\Phi_2(t)$ in (24) to approximate $\Phi_{W,2}(t)$

From the two first expressions in Sect. 5 on [Tricomi and Erdélyi \(1951\)](#) and also expressions (11) and (14), this last one already in Sect. 6 of this same reference, we may write

$$\frac{\Gamma(a - it)}{\Gamma(a + b - it)} \approx \sum_{k=0}^{\infty} p_k(b) (a - it)^{-(b+k)} \tag{33}$$

where

$$p_k(b) = \frac{1}{k} \sum_{\ell=0}^{k-1} \left(\frac{\Gamma(1 - b - \ell)}{\Gamma(-b - k)(k - \ell + 1)!} + (-1)^{k+\ell} b^{k-\ell+1} \right) p_\ell(b), \quad k = 1, 2, \dots, \tag{34}$$

with $p_0(b) = 1$, and where the approximation in (33) gets sharper for larger values of a .

Then, since the c.f. of $Y = -\log X$, where $X \sim \text{Beta}(a, b)$, is given by

$$\Phi_Y(t) = \frac{\Gamma(a + b)}{\Gamma(a)} \frac{\Gamma(a - it)}{\Gamma(a + b - it)},$$

using (33), one may write

$$\Phi_Y(t) \approx \sum_{k=0}^{\infty} \underbrace{\frac{\Gamma(a + b)}{\Gamma(a)} \frac{p_k(b)}{a^{b+k}}}_{p_k^*(a, b)} a^{b+k} (a - it)^{-(b+k)}$$

whose right hand side is the c.f. of an infinite mixture of $\Gamma(b + k, a)$ distributions, with weights $p_k^*(a, b)$, with $p_k(b)$ given by (34).

Then, since $\Phi_{W,2}(t)$ is the c.f. of a sum of independent Logbeta r.v.'s with different parameters, namely different first parameters, it would be approximated by a c.f. of an infinite mixture of sums of independent Gamma r.v.'s, with different rate parameters, which themselves are mixtures of Gamma r.v.'s. Thus, using a somewhat heuristic approach, one would use as a first simplification of this approximating c.f. a c.f. of an infinite mixture of Gamma distributions, all with the same rate parameter and with shape parameters $r + k$ for $k = 0, 1, \dots$, where r is equal to the sum of all the second parameters of the Logbeta r.v.'s in $\Phi_{W,2}(t)$, which will then be further simplified to the c.f. $\Phi_2(t)$, which is the c.f. of a finite mixture of Gamma distributions with shape parameters $r + k$ ($k = 0, 1, \dots$) and rate parameter λ , and with weights π_k which will be determined as it is explained in the body of the paper, after the call to this Appendix. The rate parameter λ will be defined in a somewhat heuristic way which has proven in practice to work very well, while the first m^* weights, $\pi_0, \dots, \pi_{m^*-1}$, are determined by equating the first m^* derivatives of $\Phi_{W,2}(t)$ and $\Phi_2(t)$, which will lead to near-exact distributions that match the first m^* exact moments of $W = -\log \Lambda$. Then, by taking $\pi_{m^*} = 1 - \sum_{k=0}^{m^*-1} \pi_k$, we will assure, in practice, that $\Phi_2(t)$ corresponds to a true c.f., and that the corresponding c.d.f. reaches the value of 1 as the running value of W goes to infinity. We may note that some of the weights π_k may be non-positive, indeed as already some of the weights $p_k(b)$ in (34) are also non-positive.

Appendix 4: Rational that shows that Δ in (27) always yields a finite value

The tails of $\left| \frac{\Phi_W(t) - \Phi_W^*(t)}{t} \right|$ for any two c.f.'s $\Phi_W(t)$ and $\Phi_W^*(t)$ are always dominated by the tails of $e^{-b|t|}$, for some $b > 0$, that is, there exists always some $\delta > 0$ such that for $|t| > \delta$,

$$\left| \frac{\Phi_W(t) - \Phi_W^*(t)}{t} \right| < e^{-b|t|}$$

while $\left| \frac{\Phi_W(t) - \Phi_W^*(t)}{t} \right|$ is a continuous function for which the limit when t tends towards zero always exists and is finite, being equal to the difference of the expected values corresponding to $\Phi_W(t)$ and $\Phi_W^*(t)$, in case both of these exist, so that since $\int_{-\infty}^{+\infty} e^{-b|t|} dt$ is finite, also Δ in (27) is.

References

- Anderson TW (2003) An introduction to multivariate statistical analysis, 3rd edn. Wiley, New Jersey
- Arnold SF (1979) Linear models with exchangeably distributed errors. *J Am Stat Assoc* 74:194–199
- Arnold BC, Coelho CA, Marques FJ (2013) The distribution of the product of powers of independent uniform random variables—a simple but useful tool to address and better understand the structure of some distributions. *J Multivar Anal* 113:19–36
- Box GEP (1949) A general distribution theory for a class of likelihood criteria. *Biometrika* 36:317–346
- Coelho CA (1998) The generalized integer gamma distribution—a basis for distribution in multivariate statistics. *J Multivar Anal* 64:86–102
- Coelho CA (2004) The generalized near-integer gamma distribution: a basis for ‘near-exact’ approximations to the distribution of statistics which are the product of an odd number of independent Beta random variables. *J Multivar Anal* 89:191–218
- Coelho CA, Marques FJ (2012) Near-exact distributions for the likelihood ratio test statistic to test equality of several variance-covariance matrices in elliptically contoured distributions. *Comp Stat* 27:627–659
- Coelho CA, Marques FJ (2013) The multi-sample block-scalar sphericity test - exact and near-exact distributions for its likelihood ratio test statistic. *Comm Statist Theory Methods* 42:1153–1175
- Coelho CA, Marques FJ, Oliveira S (2016) Near-exact distributions for likelihood ratio statistics used in the simultaneous test of conditions on mean vectors and patterns of covariance matrices. *Math Probl Eng*. doi:10.1155/2016/8975902
- Demidenko E (2004) Mixed models—theory and applications. Wiley, Hoboken
- Huynh H, Feldt LS (1970) Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. *J Am Stat Assoc* 65:1582–1589
- Johnson RA, Wichern DW (2007) Applied multivariate statistical analysis, 6th edn. Pearson Prentice Hall, Englewood Cliffs
- Jones RH (1993) Longitudinal data with serial correlation: a state-space approach. *Monographs on Statistics and Applied Probability* 47, Springer-Science and Business Media, B.V
- King ML, Evans MA (1986) Testing for block effects in regression models based on survey data. *J Am Stat Assoc* 81:677–679
- Krishnaiah PR, Lee JC (1974) On covariance structures. *Sankhyā* 38:357–371
- Krishnaiah PR, Lee JC (1980) Likelihood ratio tests for mean vectors and covariance matrices. In: Krishnaiah PR (ed) *Handbook of Statistics*. Elsevier, Amsterdam, vol 1, pp 513–570
- Kutner MH, Nachtsheim CJ, Neter J, Li W (2005) Applied linear statistical models, 5th edn. McGraw-Hill/Irwin, New York
- Li J, Wong WK (2010) Selection of covariance patterns for longitudinal data in semi-parametric models. *Stat Methods Med Res* 19:183–196
- Liang KY, Zeger SL (1986) Longitudinal data analysis using generalised linear models. *Biometrika* 73:12–22
- Malott C (1990) Maximum likelihood methods for nonlinear regression models with compound-symmetric error covariance. Ph.D. Thesis, The University of North Carolina, Chapel Hill
- Marques FJ, Coelho CA, Arnold BC (2011) A general near-exact distribution theory for the most common likelihood ratio test statistics used in multivariate analysis. *Test* 20:180–203
- Marques FJ, Coelho CA (2012) Near-exact distributions for the likelihood ratio test statistic of the multi-sample block-matrix sphericity test. *Appl Math Comput* 219:2861–2874
- Matos LA, Castro LM, Lachos VH (2016) Censored mixed-effects models for irregularly observed repeated measures with applications to HIV viral loads. *TEST* 25:627–653
- Morrison DF (1976) Multivariate statistical methods, 3rd edn. McGraw-Hill Inc., New York
- Muñoz A, Carey V, Schouten JP, Segal M, Rosner B (1992) A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics* 48:733–742

- Qu A, Li R (2006) Quadratic inference functions for varying-coefficient models with longitudinal data. *Biometrics* 62:379–391
- Rao CR (1945) Familial correlations or the multivariate generalizations of the intraclass correlation. *Curr Sci* 14:66–67
- Rao CR (1953) Discriminant functions for genetic differentiation and selection. *Sankhyā* 12:229–246
- Reinsel G (1982) Multivariate repeated-measurement or growth curve models with multivariate random-effects covariance structure. *J Am Stat Assoc* 77:190–195
- Roy A (2006) A new classification rule for incomplete doubly multivariate data using mixed effects model with performance comparisons on the imputed data. *Stat Med* 25:1715–1728
- Roy A, Fonseca M (2012) Linear models with doubly exchangeable distributed errors. *Comm Stat Theory Methods* 41:2545–2569
- Roy A, Leiva R (2011) Estimating and testing a structured covariance matrix for three-level multivariate data. *Comm Stat Theory Methods* 40:1945–1963
- Scott AJ, Holt D (1982) The effect of two-stage sampling on ordinary least squares methods. *J Am Stat Assoc* 77:848–854
- Timm NH (1980) Multivariate analysis of variance of repeated measurements. In: Krishnaiah PR (ed) *Handbook of Statistics*. Elsevier, North Holland, vol 1, pp 41–87
- Timm NH (2002) *Applied multivariate analysis*. Springer, New York
- Tricomi FG, Erdélyi A (1951) The asymptotic expansion of a ratio of Gamma functions. *Pac J Math* 1:133–142
- Verbeke G, Molenberghs G (2000) *Linear mixed models for longitudinal data*. Springer, New York
- Vonesh E, Chinchilli VM (1997) *Linear and nonlinear models for the analysis of repeated measurements*. CRC, Marcel Dekker, New York
- 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* 17:257–281
- Zimmerman DL, Núñez-Antón V (2001) Parametric modelling of growth curve data: an overview. *TEST* 10:1–73